Probability Theory: STAT310/MATH230; June 7, 2012

Amir Dembo

E-mail address: amir@math.stanford.edu

Department of Mathematics, Stanford University, Stanford, CA 94305.

Contents

Preface

These are the lecture notes for a year long, PhD level course in Probability Theory that I taught at Stanford University in 2004, 2006 and 2009. The goal of this course is to prepare incoming PhD students in Stanford's mathematics and statistics departments to do research in probability theory. More broadly, the goal of the text is to help the reader master the mathematical foundations of probability theory and the techniques most commonly used in proving theorems in this area. This is then applied to the rigorous study of the most fundamental classes of stochastic processes.

Towards this goal, we introduce in Chapter 1 the relevant elements from measure and integration theory, namely, the probability space and the σ-algebras of events in it, random variables viewed as measurable functions, their expectation as the corresponding Lebesgue integral, and the important concept of independence.

Utilizing these elements, we study in Chapter 2 the various notions of convergence of random variables and derive the weak and strong laws of large numbers.

Chapter 3 is devoted to the theory of weak convergence, the related concepts of distribution and characteristic functions and two important special cases: the Central Limit Theorem (in short CLT) and the Poisson approximation.

Drawing upon the framework of Chapter 1, we devote Chapter 4 to the definition, existence and properties of the conditional expectation and the associated regular conditional probability distribution.

Chapter 5 deals with filtrations, the mathematical notion of information progression in time, and with the corresponding stopping times. Results about the latter are obtained as a by product of the study of a collection of stochastic processes called martingales. Martingale representations are explored, as well as maximal inequalities, convergence theorems and various applications thereof. Aiming for a clearer and easier presentation, we focus here on the discrete time settings deferring the continuous time counterpart to Chapter 8.

Chapter 6 provides a brief introduction to the theory of Markov chains, a vast subject at the core of probability theory, to which many text books are devoted. We illustrate some of the interesting mathematical properties of such processes by examining a few special cases of interest.

Chapter 7 sets the framework for studying right-continuous stochastic processes indexed by a continuous time parameter, introduces the family of Gaussian processes and rigorously constructs the Brownian motion as a Gaussian process of continuous sample path and zero-mean, stationary independent increments.

Chapter 8 expands our earlier treatment of martingales and strong Markov processes to the continuous time setting, emphasizing the role of right-continuous filtration. The mathematical structure of such processes is then illustrated both in the context of Brownian motion and that of Markov jump processes.

Building on this, in Chapter 9 we re-construct the Brownian motion via the invariance principle as the limit of certain rescaled random walks. We further delve into the rich properties of its sample path and the many applications of Brownian motion to the CLT and the Law of the Iterated Logarithm (in short, LIL).

The intended audience for this course should have prior exposure to stochastic processes, at an informal level. While students are assumed to have taken a real analysis class dealing with Riemann integration, and mastered well this material, prior knowledge of measure theory is not assumed.

It is quite clear that these notes are much influenced by the text books [**Bil95, Dur10, Wil91, KaS97**] I have been using.

I thank my students out of whose work this text materialized and my teaching assistants Su Chen, Kshitij Khare, Guoqiang Hu, Julia Salzman, Kevin Sun and Hua Zhou for their help in the assembly of the notes of more than eighty students into a coherent document. I am also much indebted to Kevin Ross, Andrea Montanari and Oana Mocioalca for their feedback on earlier drafts of these notes, to Kevin Ross for providing all the figures in this text, and to Andrea Montanari, David Siegmund and Tze Lai for contributing some of the exercises in these notes.

AMIR DEMBO

STANFORD, CALIFORNIA
APRIL 2010

CHAPTER 1

Probability, measure and integration

This chapter is devoted to the mathematical foundations of probability theory. Section 1.1 introduces the basic measure theory framework, namely, the probability space and the σ-algebras of events in it. The next building blocks are random variables, introduced in Section 1.2 as measurable functions $\omega \mapsto X(\omega)$ and their distribution.

This allows us to define in Section 1.3 the important concept of expectation as the corresponding Lebesgue integral, extending the horizon of our discussion beyond the special functions and variables with density to which elementary probability theory is limited. Section 1.4 concludes the chapter by considering independence, the most fundamental aspect that differentiates probability from (general) measure theory, and the associated product measures.

1.1. Probability spaces, measures and σ-algebras

We shall define here the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ using the terminology of measure theory.

The *sample space* Ω is a set of all possible outcomes $\omega \in \Omega$ of some random experiment. Probabilities are assigned by $A \mapsto \mathbf{P}(A)$ to A in a subset $\mathcal{F}$ of all possible sets of outcomes. The *event space* $\mathcal{F}$ represents both the amount of information available as a result of the experiment conducted and the collection of all subsets of possible interest to us, where we denote elements of $\mathcal{F}$ as *events*. A pleasant mathematical framework results by imposing on $\mathcal{F}$ the structural conditions of a σ-algebra, as done in Subsection 1.1.1. The most common and useful choices for this σ-algebra are then explored in Subsection 1.1.2. Subsection 1.1.3 provides fundamental supplements from measure theory, namely Dynkin's and Carathéodory's theorems and their application to the construction of Lebesgue measure.

1.1.1. The probability space $(\Omega, \mathcal{F}, \mathbf{P})$.

We use 2^Ω to denote the set of all possible subsets of Ω. The event space is thus a subset $\mathcal{F}$ of 2^Ω, consisting of all allowed events, that is, those subsets of Ω to which we shall assign probabilities. We next define the structural conditions imposed on $\mathcal{F}$.

Definition 1.1.1. *We say that $\mathcal{F} \subseteq 2^\Omega$ is a σ-*algebra *(or a σ-*field*), if*
(a) $\Omega \in \mathcal{F}$,
(b) If $A \in \mathcal{F}$ then $A^c \in \mathcal{F}$ as well (where $A^c = \Omega \setminus A$).
(c) If $A_i \in \mathcal{F}$ for $i = 1, 2, 3, \ldots$ then also $\bigcup_i A_i \in \mathcal{F}$.

Remark. Using *DeMorgan's law*, we know that $(\bigcup_i A_i^c)^c = \bigcap_i A_i$. Thus the following is equivalent to property (c) of Definition 1.1.1:
(c') If $A_i \in \mathcal{F}$ for $i = 1, 2, 3, \ldots$ then also $\bigcap_i A_i \in \mathcal{F}$.

Definition 1.1.2. *A pair* $(\Omega, \mathcal{F})$ *with* $\mathcal{F}$ *a* σ*-algebra of subsets of* Ω *is called a* measurable space*. Given a measurable space* $(\Omega, \mathcal{F})$*, a* measure μ *is any* countably additive *non-negative* set function *on this space. That is,* $\mu : \mathcal{F} \to [0, \infty]$*, having the properties:*
(a) $\mu(A) \geq \mu(\emptyset) = 0$ *for all* $A \in \mathcal{F}$*.*
(b) $\mu(\bigcup_n A_n) = \sum_n \mu(A_n)$ *for any countable collection of disjoint sets* $A_n \in \mathcal{F}$*.*
When in addition $\mu(\Omega) = 1$*, we call the measure* μ *a* probability measure*, and often label it by* $\mathbf{P}$ *(it is also easy to see that then* $\mathbf{P}(A) \leq 1$ *for all* $A \in \mathcal{F}$*).*

Remark. When (b) of Definition 1.1.2 is relaxed to involve only finite collections of disjoint sets A_n, we say that μ is a *finitely additive* non-negative set-function. In measure theory we sometimes consider *signed measures*, whereby μ is no longer non-negative, hence its range is $[-\infty, \infty]$, and say that such measure is *finite* when its range is $\mathbb{R}$ (i.e. no set in $\mathcal{F}$ is assigned an infinite measure).

Definition 1.1.3. *A* measure space *is a triplet* $(\Omega, \mathcal{F}, \mu)$*, with* μ *a measure on the measurable space* $(\Omega, \mathcal{F})$*. A measure space* $(\Omega, \mathcal{F}, \mathbf{P})$ *with* $\mathbf{P}$ *a probability measure is called a* probability space*.*

The next exercise collects some of the fundamental properties shared by all probability measures.

Exercise 1.1.4. *Let* $(\Omega, \mathcal{F}, \mathbf{P})$ *be a probability space and* A, B, A_i *events in* $\mathcal{F}$*. Prove the following properties of every probability measure.*

(a) Monotonicity. *If* $A \subseteq B$ *then* $\mathbf{P}(A) \leq \mathbf{P}(B)$*.*
(b) Sub-additivity. *If* $A \subseteq \cup_i A_i$ *then* $\mathbf{P}(A) \leq \sum_i \mathbf{P}(A_i)$*.*
(c) Continuity from below: *If* $A_i \uparrow A$*, that is,* $A_1 \subseteq A_2 \subseteq \ldots$ *and* $\cup_i A_i = A$*, then* $\mathbf{P}(A_i) \uparrow \mathbf{P}(A)$*.*
(d) Continuity from above: *If* $A_i \downarrow A$*, that is,* $A_1 \supseteq A_2 \supseteq \ldots$ *and* $\cap_i A_i = A$*, then* $\mathbf{P}(A_i) \downarrow \mathbf{P}(A)$*.*

Remark. In the more general context of measure theory, note that properties (a)-(c) of Exercise 1.1.4 hold for any measure μ, whereas the continuity from above holds whenever $\mu(A_i) < \infty$ for all i sufficiently large. Here is more on this:

Exercise 1.1.5. *Prove that a finitely additive non-negative set function* μ *on a measurable space* $(\Omega, \mathcal{F})$ *with the "continuity" property*

$$B_n \in \mathcal{F}, \quad B_n \downarrow \emptyset, \quad \mu(B_n) < \infty \qquad \Longrightarrow \quad \mu(B_n) \to 0$$

must be countably additive if $\mu(\Omega) < \infty$*. Give an example that it is not necessarily so when* $\mu(\Omega) = \infty$*.*

The σ-algebra $\mathcal{F}$ always contains at least the set Ω and its complement, the empty set $\emptyset$. Necessarily, $\mathbf{P}(\Omega) = 1$ and $\mathbf{P}(\emptyset) = 0$. So, if we take $\mathcal{F}_0 = \{\emptyset, \Omega\}$ as our σ-algebra, then we are left with no degrees of freedom in choice of $\mathbf{P}$. For this reason we call $\mathcal{F}_0$ the *trivial* σ*-algebra*. Fixing Ω, we may expect that the larger the σ-algebra we consider, the more freedom we have in choosing the probability measure. This indeed holds to some extent, that is, as long as we have no problem satisfying the requirements in the definition of a probability measure. A natural question is when should we expect the maximal possible σ-algebra $\mathcal{F} = 2^\Omega$ to be useful?

Example 1.1.6. *When the sample space* Ω *is countable we can and typically shall take* $\mathcal{F} = 2^\Omega$*. Indeed, in such situations we assign a probability* $p_\omega > 0$ *to each* $\omega \in \Omega$

making sure that $\sum_{\omega\in\Omega} p_\omega = 1$. *Then, it is easy to see that taking* $\mathbf{P}(A) = \sum_{\omega\in A} p_\omega$ *for any* $A \subseteq \Omega$ *results with a probability measure on* $(\Omega, 2^\Omega)$. *For instance, when* Ω *is finite, we can take* $p_\omega = \frac{1}{|\Omega|}$, *the uniform measure on* Ω, *whereby computing probabilities is the same as counting. Concrete examples are a single coin toss, for which we have* $\Omega_1 = \{\mathrm{H}, \mathrm{T}\}$ *(*$\omega = \mathrm{H}$ *if the coin lands on its head and* $\omega = \mathrm{T}$ *if it lands on its tail), and* $\mathcal{F}_1 = \{\emptyset, \Omega, \mathrm{H}, \mathrm{T}\}$, *or when we consider a finite number of coin tosses, say* n, *in which case* $\Omega_n = \{(\omega_1, \ldots, \omega_n) : \omega_i \in \{\mathrm{H}, \mathrm{T}\}, i = 1, \ldots, n\}$ *is the set of all possible* n*-tuples of coin tosses, while* $\mathcal{F}_n = 2^{\Omega_n}$ *is the collection of all possible sets of* n*-tuples of coin tosses. Another example pertains to the set of all non-negative integers* $\Omega = \{0, 1, 2, \ldots\}$ *and* $\mathcal{F} = 2^\Omega$, *where we get the Poisson probability measure of parameter* $\lambda > 0$ *when starting from* $p_k = \frac{\lambda^k}{k!} e^{-\lambda}$ *for* $k = 0, 1, 2, \ldots$.

When Ω is uncountable such a strategy as in Example 1.1.6 will no longer work. The problem is that if we take $p_\omega = \mathbf{P}(\{\omega\}) > 0$ for uncountably many values of ω, we shall end up with $\mathbf{P}(\Omega) = \infty$. Of course we may define everything as before on a countable subset $\widehat{\Omega}$ of Ω and demand that $\mathbf{P}(A) = \mathbf{P}(A \cap \widehat{\Omega})$ for each $A \subseteq \Omega$. Excluding such trivial cases, to genuinely use an uncountable sample space Ω we need to restrict our σ-algebra $\mathcal{F}$ to a *strict subset* of 2^Ω.

Definition 1.1.7. *We say that a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$ *is* non-atomic, *or alternatively call* $\mathbf{P}$ *non-atomic if* $\mathbf{P}(A) > 0$ *implies the existence of* $B \in \mathcal{F}$, $B \subset A$ *with* $0 < \mathbf{P}(B) < \mathbf{P}(A)$.

Indeed, in contrast to the case of countable Ω, the generic uncountable sample space results with a non-atomic probability space (c.f. Exercise 1.1.27). Here is an interesting property of such spaces (see also [**Bil95**, Problem 2.19]).

Exercise 1.1.8. *Suppose* $\mathbf{P}$ *is non-atomic and* $A \in \mathcal{F}$ *with* $\mathbf{P}(A) > 0$.

(a) *Show that for every* $\epsilon > 0$, *we have* $B \subseteq A$ *such that* $0 < \mathbf{P}(B) < \epsilon$.
(b) *Prove that if* $0 < a < \mathbf{P}(A)$ *then there exists* $B \subset A$ *with* $\mathbf{P}(B) = a$.

Hint: *Fix* $\epsilon_n \downarrow 0$ *and define inductively numbers* x_n *and sets* $G_n \in \mathcal{F}$ *with* $H_0 = \emptyset$, $H_n = \cup_{k<n} G_k$, $x_n = \sup\{\mathbf{P}(G) : G \subseteq A \backslash H_n, \mathbf{P}(H_n \cup G) \le a\}$ *and* $G_n \subseteq A \backslash H_n$ *such that* $\mathbf{P}(H_n \bigcup G_n) \le a$ *and* $\mathbf{P}(G_n) \ge (1 - \epsilon_n) x_n$. *Consider* $B = \cup_k G_k$.

As you show next, the collection of all measures on a given space is a *convex cone*.

Exercise 1.1.9. *Given any measures* $\{\mu_n, n \ge 1\}$ *on* $(\Omega, \mathcal{F})$, *verify that* $\mu = \sum_{n=1}^\infty c_n \mu_n$ *is also a measure on this space, for any finite constants* $c_n \ge 0$.

Here are few properties of probability measures for which the conclusions of Exercise 1.1.4 are useful.

Exercise 1.1.10. *A function* $d : \mathcal{X} \times \mathcal{X} \to [0, \infty)$ *is called a* semi-metric *on the set* $\mathcal{X}$ *if* $d(x, x) = 0$, $d(x, y) = d(y, x)$ *and the triangle inequality* $d(x, z) \le d(x, y) + d(y, z)$ *holds. With* $A\Delta B = (A \cap B^c) \cup (A^c \cap B)$ *denoting the symmetric difference of subsets* A *and* B *of* Ω, *show that for any probability space* $(\Omega, \mathcal{F}, \mathbf{P})$, *the function* $d(A, B) = \mathbf{P}(A\Delta B)$ *is a semi-metric on* $\mathcal{F}$.

Exercise 1.1.11. *Consider events* $\{A_n\}$ *in a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$ *that are almost disjoint in the sense that* $\mathbf{P}(A_n \cap A_m) = 0$ *for all* $n \ne m$. *Show that then* $\mathbf{P}(\cup_{n=1}^\infty A_n) = \sum_{n=1}^\infty \mathbf{P}(A_n)$.

Exercise 1.1.12. *Suppose a random outcome N follows the Poisson probability measure of parameter $\lambda > 0$. Find a simple expression for the probability that N is an even integer.*

1.1.2. Generated and Borel σ-algebras. Enumerating the sets in the σ-algebra $\mathcal{F}$ is not a realistic option for uncountable Ω. Instead, as we see next, the most common construction of σ-algebras is then by implicit means. That is, we demand that certain sets (called the *generators*) be in our σ-algebra, and take the smallest possible collection for which this holds.

Exercise 1.1.13.

(a) *Check that the intersection of (possibly uncountably many) σ-algebras is also a σ-algebra.*

(b) *Verify that for any σ-algebras $\mathcal{H} \subseteq \mathcal{G}$ and any $H \in \mathcal{H}$, the collection $\mathcal{H}^H = \{A \in \mathcal{G} : A \cap H \in \mathcal{H}\}$ is a σ-algebra.*

(c) *Show that $H \mapsto \mathcal{H}^H$ is non-increasing with respect to set inclusions, with $\mathcal{H}^\Omega = \mathcal{H}$ and $\mathcal{H}^\emptyset = \mathcal{G}$. Deduce that $\mathcal{H}^{H \cup H'} = \mathcal{H}^H \cap \mathcal{H}^{H'}$ for any pair $H, H' \in \mathcal{H}$.*

In view of part (a) of this exercise we have the following definition.

Definition 1.1.14. *Given a collection of subsets $A_\alpha \subseteq \Omega$ (not necessarily countable), we denote the smallest σ-algebra $\mathcal{F}$ such that $A_\alpha \in \mathcal{F}$ for all $\alpha \in \Gamma$ either by $\sigma(\{A_\alpha\})$ or by $\sigma(A_\alpha, \alpha \in \Gamma)$, and call $\sigma(\{A_\alpha\})$ the* σ-algebra generated *by the sets A_α. That is,*
$\sigma(\{A_\alpha\}) = \bigcap \{\mathcal{G} : \ \mathcal{G} \subseteq 2^\Omega \text{ is a } \sigma - \text{algebra}, \quad A_\alpha \in \mathcal{G} \quad \forall \alpha \in \Gamma\}.$

Example 1.1.15. *Suppose $\Omega = \mathbb{S}$ is a topological space (that is, $\mathbb{S}$ is equipped with a notion of open subsets, or topology). An example of a generated σ-algebra is the* Borel σ-algebra *on $\mathbb{S}$ defined as $\sigma(\{O \subseteq \mathbb{S} \text{ open}\})$ and denoted by $\mathcal{B}_\mathbb{S}$. Of special importance is $\mathcal{B}_\mathbb{R}$ which we also denote by $\mathcal{B}$.*

Different sets of generators may result with the same σ-algebra. For example, taking $\Omega = \{1, 2, 3\}$ it is easy to see that $\sigma(\{1\}) = \sigma(\{2, 3\}) = \{\emptyset, \{1\}, \{2, 3\}, \{1, 2, 3\}\}$.
A σ-algebra $\mathcal{F}$ is *countably generated* if there exists a countable collection of sets that generates it. Exercise 1.1.17 shows that $\mathcal{B}_\mathbb{R}$ is countably generated, but as you show next, there exist non countably generated σ-algebras even on $\Omega = \mathbb{R}$.

Exercise 1.1.16. *Let $\mathcal{F}$ consist of all $A \subseteq \Omega$ such that either A is a countable set or A^c is a countable set.*

(a) *Verify that $\mathcal{F}$ is a σ-algebra.*

(b) *Show that $\mathcal{F}$ is countably generated if and only if Ω is a countable set.*

Recall that if a collection of sets $\mathcal{A}$ is a subset of a σ-algebra $\mathcal{G}$, then also $\sigma(\mathcal{A}) \subseteq \mathcal{G}$. Consequently, to show that $\sigma(\{A_\alpha\}) = \sigma(\{B_\beta\})$ for two different sets of generators $\{A_\alpha\}$ and $\{B_\beta\}$, we only need to show that $A_\alpha \in \sigma(\{B_\beta\})$ for each α and that $B_\beta \in \sigma(\{A_\alpha\})$ for each β. For instance, considering $\mathcal{B}_\mathbb{Q} = \sigma(\{(a, b) : a < b \in \mathbb{Q}\})$, we have by this approach that $\mathcal{B}_\mathbb{Q} = \sigma(\{(a, b) : a < b \in \mathbb{R}\})$, as soon as we show that any interval (a, b) is in $\mathcal{B}_\mathbb{Q}$. To see this fact, note that for any real $a < b$ there are rational numbers $q_n < r_n$ such that $q_n \downarrow a$ and $r_n \uparrow b$, hence $(a, b) = \cup_n (q_n, r_n) \in \mathcal{B}_\mathbb{Q}$. Expanding on this, the next exercise provides useful alternative definitions of $\mathcal{B}$.

Exercise 1.1.17. *Verify the alternative definitions of the* Borel σ-algebra $\mathcal{B}$*:*

$$\sigma(\{(a,b) : a < b \in \mathbb{R}\}) = \sigma(\{[a,b] : a < b \in \mathbb{R}\}) = \sigma(\{(-\infty, b] : b \in \mathbb{R}\})$$
$$= \sigma(\{(-\infty, b] : b \in \mathbb{Q}\}) = \sigma(\{O \subseteq \mathbb{R} \text{ open }\})$$

If $A \subseteq \mathbb{R}$ is in $\mathcal{B}$ of Example 1.1.15, we say that A is a *Borel set*. In particular, all open (closed) subsets of $\mathbb{R}$ are Borel sets, as are many other sets. However,

Proposition 1.1.18. *There exists a subset of* $\mathbb{R}$ *that is not in* $\mathcal{B}$*. That is, not all subsets of* $\mathbb{R}$ *are Borel sets.*

PROOF. See [**Wil91**, A.1.1] or [**Bil95**, page 45]. □

Example 1.1.19. *Another classical example of an uncountable* Ω *is relevant for studying the experiment with an infinite number of coin tosses, that is,* $\Omega_\infty = \Omega_1^{\mathbb{N}}$ *for* $\Omega_1 = \{\mathrm{H}, \mathrm{T}\}$ *(indeed, setting* $\mathrm{H} = 1$ *and* $\mathrm{T} = 0$*, each infinite sequence* $\omega \in \Omega_\infty$ *is in correspondence with a unique real number* $x \in [0,1]$ *with* ω *being the binary expansion of* x*, see Exercise 1.2.13). The* σ*-algebra should at least allow us to consider any possible outcome of a finite number of coin tosses. The natural* σ*-algebra in this case is the minimal* σ*-algebra having this property, or put more formally* $\mathcal{F}_c = \sigma(\{A_{\theta,k}, \theta \in \Omega_1^k, k = 1, 2, \ldots\})$*, where* $A_{\theta,k} = \{\omega \in \Omega_\infty : \omega_i = \theta_i, i = 1\ldots, k\}$ *for* $\theta = (\theta_1, \ldots, \theta_k)$*.*

The preceding example is a special case of the construction of a product of measurable spaces, which we detail now.

Example 1.1.20. *The* product *of the measurable spaces* $(\Omega_i, \mathcal{F}_i)$*,* $i = 1, \ldots, n$ *is the set* $\Omega = \Omega_1 \times \cdots \times \Omega_n$ *with the* σ*-algebra generated by* $\{A_1 \times \cdots \times A_n : A_i \in \mathcal{F}_i\}$*, denoted by* $\mathcal{F}_1 \times \cdots \mathcal{F}_n$*.*

You are now to check that the Borel σ-algebra of $\mathbb{R}^d$ is the product of d-copies of that of $\mathbb{R}$. As we see later, this helps simplifying the study of random vectors.

Exercise 1.1.21. *Show that for any* $d < \infty$*,*

$$\mathcal{B}_{\mathbb{R}^d} = \mathcal{B} \times \cdots \times \mathcal{B} = \sigma(\{(a_1, b_1) \times \cdots \times (a_d, b_d) : a_i < b_i \in \mathbb{R}, i = 1, \ldots, d\})$$

(you need to prove both *identities, with the middle term defined as in Example 1.1.20).*

Exercise 1.1.22. *Let* $\mathcal{F} = \sigma(A_\alpha, \alpha \in \Gamma)$ *where the collection of sets* A_α*,* $\alpha \in \Gamma$ *is uncountable (i.e.,* Γ *is uncountable). Prove that for each* $B \in \mathcal{F}$ *there exists a countable sub-collection* $\{A_{\alpha_j}, j = 1, 2, \ldots\} \subset \{A_\alpha, \alpha \in \Gamma\}$*, such that* $B \in \sigma(\{A_{\alpha_j}, j = 1, 2, \ldots\})$*.*

Often there is no explicit enumerative description of the σ-algebra generated by an infinite collection of subsets, but a notable exception is

Exercise 1.1.23. *Show that the sets in* $\mathcal{G} = \sigma(\{[a,b] : a, b \in \mathbf{Z}\})$ *are all possible unions of elements from the countable collection* $\{\{b\}, (b, b+1), b \in \mathbf{Z}\}$*, and deduce that* $\mathcal{B} \neq \mathcal{G}$*.*

Probability measures on the Borel σ-algebra of $\mathbb{R}$ are examples of *regular measures,* namely:

Exercise 1.1.24. *Show that if* $\mathbf{P}$ *is a probability measure on* $(\mathbb{R}, \mathcal{B})$ *then for any* $A \in \mathcal{B}$ *and* $\epsilon > 0$*, there exists an open set* G *containing* A *such that* $\mathbf{P}(A) + \epsilon > \mathbf{P}(G)$.

Here is more information about $\mathcal{B}_{\mathbb{R}^d}$.

Exercise 1.1.25. *Show that if* μ *is a finitely additive non-negative set function on* $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$ *such that* $\mu(\mathbb{R}^d) = 1$ *and for any Borel set* A*,*

$$\mu(A) = \sup\{\mu(K) : K \subseteq A, \ K \ \textit{compact} \ \},$$

then μ *must be a probability measure.*
Hint: *Argue by contradiction using the conclusion of Exercise 1.1.5. To this end, recall the finite intersection property (if compact* $K_i \subset \mathbb{R}^d$ *are such that* $\bigcap_{i=1}^n K_i$ *are non-empty for finite* n*, then the countable intersection* $\bigcap_{i=1}^\infty K_i$ *is also non-empty).*

1.1.3. Lebesgue measure and Carathéodory's theorem. Perhaps the most important measure on $(\mathbb{R}, \mathcal{B})$ is the *Lebesgue measure*, λ. It is the unique measure that satisfies $\lambda(F) = \sum_{k=1}^r (b_k - a_k)$ whenever $F = \bigcup_{k=1}^r (a_k, b_k]$ for some $r < \infty$ and $a_1 < b_1 < a_2 < b_2 \cdots < b_r$. Since $\lambda(\mathbb{R}) = \infty$, this is not a probability measure. However, when we restrict Ω to be the interval $(0, 1]$ we get

Example 1.1.26. *The* uniform probability measure *on* $(0, 1]$*, is denoted* U *and defined as above, now with added restrictions that* $0 \le a_1$ *and* $b_r \le 1$*. Alternatively,* U *is the restriction of the measure* λ *to the sub-*σ*-algebra* $\mathcal{B}_{(0,1]}$ *of* $\mathcal{B}$.

Exercise 1.1.27. *Show that* $((0, 1], \mathcal{B}_{(0,1]}, U)$ *is a non-atomic probability space and deduce that* $(\mathbb{R}, \mathcal{B}, \lambda)$ *is a non-atomic measure space.*

Note that any countable union of sets of probability zero has probability zero, but this is not the case for an uncountable union. For example, $U(\{x\}) = 0$ for every $x \in \mathbb{R}$, but $U(\mathbb{R}) = 1$.

As we have seen in Example 1.1.26 it is often impossible to explicitly specify the value of a measure on all sets of the σ-algebra $\mathcal{F}$. Instead, we wish to specify its values on a much smaller and better behaved collection of generators $\mathcal{A}$ of $\mathcal{F}$ and use Carathéodory's theorem to guarantee the existence of a unique measure on $\mathcal{F}$ that coincides with our specified values. To this end, we require that $\mathcal{A}$ be an algebra, that is,

Definition 1.1.28. *A collection* $\mathcal{A}$ *of subsets of* Ω *is an* algebra *(or a* field*) if*

(a) $\Omega \in \mathcal{A}$,
(b) *If* $A \in \mathcal{A}$ *then* $A^c \in \mathcal{A}$ *as well,*
(c) *If* $A, B \in \mathcal{A}$ *then also* $A \cup B \in \mathcal{A}$.

Remark. In view of the closure of algebra with respect to complements, we could have replaced the requirement that $\Omega \in \mathcal{A}$ with the (more standard) requirement that $\emptyset \in \mathcal{A}$. As part (c) of Definition 1.1.28 amounts to closure of an algebra under finite unions (and by DeMorgan's law also finite intersections), the difference between an algebra and a σ-algebra is that a σ-algebra is also closed under countable unions.

We sometimes make use of the fact that unlike generated σ-algebras, the algebra generated by a collection of sets $\mathcal{A}$ can be explicitly presented.

Exercise 1.1.29. *The algebra generated by a given collection of subsets* $\mathcal{A}$*, denoted* $f(\mathcal{A})$*, is the intersection of all algebras of subsets of* Ω *containing* $\mathcal{A}$.

(a) *Verify that $f(\mathcal{A})$ is indeed an algebra and that $f(\mathcal{A})$ is minimal in the sense that if $\mathcal{G}$ is an algebra and $\mathcal{A} \subseteq \mathcal{G}$, then $f(\mathcal{A}) \subseteq \mathcal{G}$.*

(b) *Show that $f(\mathcal{A})$ is the collection of all finite disjoint unions of sets of the form $\bigcap_{j=1}^{n_i} A_{ij}$, where for each i and j either A_{ij} or A_{ij}^c are in $\mathcal{A}$.*

We next state Carathéodory's extension theorem, a key result from measure theory, and demonstrate how it applies in the context of Example 1.1.26.

Theorem 1.1.30 (Carathéodory's extension theorem)**.** *If $\mu_0 : \mathcal{A} \mapsto [0,\infty]$ is a countably additive set function on an algebra $\mathcal{A}$ then there exists a measure μ on $(\Omega, \sigma(\mathcal{A}))$ such that $\mu = \mu_0$ on $\mathcal{A}$. Furthermore, if $\mu_0(\Omega) < \infty$ then such a measure μ is unique.*

To construct the measure U on $\mathcal{B}_{(0,1]}$ let $\Omega = (0,1]$ and

$$\mathcal{A} = \{(a_1, b_1] \cup \cdots \cup (a_r, b_r] : 0 \le a_1 < b_1 < \cdots < a_r < b_r \le 1\,,\, r < \infty\}$$

be a collection of subsets of $(0,1]$. It is not hard to verify that $\mathcal{A}$ is an algebra, and further that $\sigma(\mathcal{A}) = \mathcal{B}_{(0,1]}$ (c.f. Exercise 1.1.17, for a similar issue, just with $(0,1]$ replaced by $\mathbb{R}$). With U_0 denoting the non-negative set function on $\mathcal{A}$ such that

$$U_0\Big(\bigcup_{k=1}^{r}(a_k, b_k]\Big) = \sum_{k=1}^{r}(b_k - a_k)\,, \tag{1.1.1}$$

note that $U_0((0,1]) = 1$, hence the existence of a unique probability measure U on $((0,1], \mathcal{B}_{(0,1]})$ such that $U(A) = U_0(A)$ for sets $A \in \mathcal{A}$ follows by Carathéodory's extension theorem, as soon as we verify that

Lemma 1.1.31. *The set function U_0 is countably additive on $\mathcal{A}$. That is, if A_k is a sequence of disjoint sets in $\mathcal{A}$ such that $\cup_k A_k = A \in \mathcal{A}$, then $U_0(A) = \sum_k U_0(A_k)$.*

The proof of Lemma 1.1.31 is based on

Exercise 1.1.32. *Show that U_0 is finitely additive on $\mathcal{A}$. That is, $U_0(\bigcup_{k=1}^n A_k) = \sum_{k=1}^n U_0(A_k)$ for any finite collection of disjoint sets $A_1, \ldots, A_n \in \mathcal{A}$.*

Proof. Let $G_n = \bigcup_{k=1}^n A_k$ and $H_n = A \setminus G_n$. Then, $H_n \downarrow \emptyset$ and since $A_k, A \in \mathcal{A}$ which is an algebra it follows that G_n and hence H_n are also in $\mathcal{A}$. By definition, U_0 is finitely additive on $\mathcal{A}$, so

$$U_0(A) = U_0(H_n) + U_0(G_n) = U_0(H_n) + \sum_{k=1}^{n} U_0(A_k)\,.$$

To prove that U_0 is countably additive, it suffices to show that $U_0(H_n) \downarrow 0$, for then

$$U_0(A) = \lim_{n\to\infty} U_0(G_n) = \lim_{n\to\infty} \sum_{k=1}^{n} U_0(A_k) = \sum_{k=1}^{\infty} U_0(A_k)\,.$$

To complete the proof, we argue by contradiction, assuming that $U_0(H_n) \ge 2\varepsilon$ for some $\varepsilon > 0$ and all n, where $H_n \downarrow \emptyset$ are elements of $\mathcal{A}$. By the definition of $\mathcal{A}$ and U_0, we can find for each ℓ a set $J_\ell \in \mathcal{A}$ whose closure $\overline{J}_\ell$ is a subset of H_ℓ and $U_0(H_\ell \setminus J_\ell) \le \varepsilon 2^{-\ell}$ (for example, add to each a_k in the representation of H_ℓ the minimum of $\varepsilon 2^{-\ell}/r$ and $(b_k - a_k)/2$). With U_0 finitely additive on the algebra $\mathcal{A}$ this implies that for each n,

$$U_0\Big(\bigcup_{\ell=1}^{n}(H_\ell \setminus J_\ell)\Big) \le \sum_{\ell=1}^{n} U_0(H_\ell \setminus J_\ell) \le \varepsilon\,.$$

As $H_n \subseteq H_\ell$ for all $\ell \le n$, we have that

$$H_n \setminus \bigcap_{\ell \le n} J_\ell = \bigcup_{\ell \le n} (H_n \setminus J_\ell) \subseteq \bigcup_{\ell \le n} (H_\ell \setminus J_\ell) \,.$$

Hence, by finite additivity of U_0 and our assumption that $U_0(H_n) \ge 2\varepsilon$, also

$$U_0(\bigcap_{\ell \le n} J_\ell) = U_0(H_n) - U_0(H_n \setminus \bigcap_{\ell \le n} J_\ell) \ge U_0(H_n) - U_0(\bigcup_{\ell \le n} (H_\ell \setminus J_\ell)) \ge \varepsilon \,.$$

In particular, for every n, the set $\bigcap_{\ell \le n} J_\ell$ is non-empty and therefore so are the decreasing sets $K_n = \bigcap_{\ell \le n} \overline{J}_\ell$. Since K_n are compact sets (by Heine-Borel theorem), the set $\cap_\ell \overline{J}_\ell$ is then non-empty as well, and since $\overline{J}_\ell$ is a subset of H_ℓ for all ℓ we arrive at $\cap_\ell H_\ell$ non-empty, contradicting our assumption that $H_n \downarrow \emptyset$. □

Remark. The proof of Lemma 1.1.31 is generic (for finite measures). Namely, any non-negative finitely additive set function μ_0 on an algebra $\mathcal{A}$ is countably additive if $\mu_0(H_n) \downarrow 0$ whenever $H_n \in \mathcal{A}$ and $H_n \downarrow \emptyset$. Further, as this proof shows, when Ω is a topological space it suffices for countable additivity of μ_0 to have for any $H \in \mathcal{A}$ a sequence $J_k \in \mathcal{A}$ such that $\overline{J}_k \subseteq H$ are compact and $\mu_0(H \setminus J_k) \to 0$ as $k \to \infty$.

Exercise 1.1.33. *Show the necessity of the assumption that $\mathcal{A}$ be an algebra in Carathéodory's extension theorem, by giving an example of two probability measures $\mu \ne \nu$ on a measurable space $(\Omega, \mathcal{F})$ such that $\mu(A) = \nu(A)$ for all $A \in \mathcal{A}$ and $\mathcal{F} = \sigma(\mathcal{A})$.*
Hint: *This can be done with $\Omega = \{1, 2, 3, 4\}$ and $\mathcal{F} = 2^\Omega$.*

It is often useful to assume that the probability space we have is complete, in the sense we make precise now.

Definition 1.1.34. *We say that a measure space $(\Omega, \mathcal{F}, \mu)$ is* complete *if any subset N of any $B \in \mathcal{F}$ with $\mu(B) = 0$ is also in $\mathcal{F}$. If further $\mu = \mathbf{P}$ is a probability measure, we say that the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ is a* complete probability space.

Our next theorem states that any measure space can be completed by adding to its σ-algebra all subsets of sets of zero measure (a procedure that depends on the measure in use).

Theorem 1.1.35. *Given a measure space $(\Omega, \mathcal{F}, \mu)$, let $\mathcal{N} = \{N : N \subseteq A$ for some $A \in \mathcal{F}$ with $\mu(A) = 0\}$ denote the collection of μ-*null sets*. Then, there exists a complete measure space $(\Omega, \overline{\mathcal{F}}, \overline{\mu})$, called the* completion *of the measure space $(\Omega, \mathcal{F}, \mu)$, such that $\overline{\mathcal{F}} = \{F \cup N : F \in \mathcal{F}, N \in \mathcal{N}\}$ and $\overline{\mu} = \mu$ on $\mathcal{F}$.*

Proof. This is beyond our scope, but see detailed proof in [**Dur10**, Theorem A.2.3]. In particular, $\overline{\mathcal{F}} = \sigma(\mathcal{F}, \mathcal{N})$ and $\overline{\mu}(A \cup N) = \mu(A)$ for any $N \in \mathcal{N}$ and $A \in \mathcal{F}$ (c.f. [**Bil95**, Problems 3.10 and 10.5]). □

The following collections of sets play an important role in proving the easy part of Carathéodory's theorem, the uniqueness of the extension μ.

Definition 1.1.36. *A* π-system *is a collection $\mathcal{P}$ of sets closed under finite intersections (i.e. if $I \in \mathcal{P}$ and $J \in \mathcal{P}$ then $I \cap J \in \mathcal{P}$).*
A λ-system *is a collection $\mathcal{L}$ of sets containing Ω and $B \setminus A$ for any $A \subseteq B$ $A, B \in \mathcal{L}$,*

which is also closed under monotone increasing limits (i.e. if $A_i \in \mathcal{L}$ and $A_i \uparrow A$, then $A \in \mathcal{L}$ as well).

Obviously, an algebra is a π-system. Though an algebra may not be a λ-system,

Proposition 1.1.37. *A collection $\mathcal{F}$ of sets is a σ-algebra if and only if it is both a π-system and a λ-system.*

Proof. The fact that a σ-algebra is a λ-system is a trivial consequence of Definition 1.1.1. To prove the converse direction, suppose that $\mathcal{F}$ is both a π-system and a λ-system. Then Ω is in the λ-system $\mathcal{F}$ and so is $A^c = \Omega \setminus A$ for any $A \in \mathcal{F}$. Further, with $\mathcal{F}$ also a π-system we have that

$$A \cup B = \Omega \setminus (A^c \cap B^c) \in \mathcal{F} ,$$

for any $A, B \in \mathcal{F}$. Consequently, if $A_i \in \mathcal{F}$ then so are also $G_n = A_1 \cup \cdots \cup A_n \in \mathcal{F}$. Since $\mathcal{F}$ is a λ-system and $G_n \uparrow \bigcup_i A_i$, it follows that $\bigcup_i A_i \in \mathcal{F}$ as well, completing the verification that $\mathcal{F}$ is a σ-algebra. $\square$

The main tool in proving the uniqueness of the extension is *Dynkin's $\pi-\lambda$ theorem*, stated next.

Theorem 1.1.38 (Dynkin's $\pi - \lambda$ theorem). *If $\mathcal{P} \subseteq \mathcal{L}$ with $\mathcal{P}$ a π-system and $\mathcal{L}$ a λ-system then $\sigma(\mathcal{P}) \subseteq \mathcal{L}$.*

Proof. A short though dense exercise in set manipulations shows that the smallest λ-system containing $\mathcal{P}$ is a π-system (for details see [**Wil91**, Section A.1.3] or the proof of [**Bil95**, Theorem 3.2]). By Proposition 1.1.37 it is a σ-algebra, hence contains $\sigma(\mathcal{P})$. Further, it is contained in the λ-system $\mathcal{L}$, as $\mathcal{L}$ also contains $\mathcal{P}$. $\square$

Remark. Proposition 1.1.37 remains valid even if in the definition of λ-system we relax the condition that $B \setminus A \in \mathcal{L}$ for any $A \subseteq B$ $A, B \in \mathcal{L}$, to the condition $A^c \in \mathcal{L}$ whenever $A \in \mathcal{L}$. However, Dynkin's theorem does not hold under the latter definition.

As we show next, the uniqueness part of Carathéodory's theorem, is an immediate consequence of the $\pi - \lambda$ theorem.

Proposition 1.1.39. *If two measures μ_1 and μ_2 on $(\Omega, \sigma(\mathcal{P}))$ agree on the π-system $\mathcal{P}$ and are such that $\mu_1(\Omega) = \mu_2(\Omega) < \infty$, then $\mu_1 = \mu_2$.*

Proof. Let $\mathcal{L} = \{A \in \sigma(\mathcal{P}) : \mu_1(A) = \mu_2(A)\}$. Our assumptions imply that $\mathcal{P} \subseteq \mathcal{L}$ and that $\Omega \in \mathcal{L}$. Further, $\sigma(\mathcal{P})$ is a λ-system (by Proposition 1.1.37), and if $A \subseteq B$, $A, B \in \mathcal{L}$, then by additivity of the finite measures μ_1 and μ_2,

$$\mu_1(B \setminus A) = \mu_1(B) - \mu_1(A) = \mu_2(B) - \mu_2(A) = \mu_2(B \setminus A),$$

that is, $B \setminus A \in \mathcal{L}$. Similarly, if $A_i \uparrow A$ and $A_i \in \mathcal{L}$, then by the continuity from below of μ_1 and μ_2 (see remark following Exercise 1.1.4),

$$\mu_1(A) = \lim_{n\to\infty} \mu_1(A_n) = \lim_{n\to\infty} \mu_2(A_n) = \mu_2(A) ,$$

so that $A \in \mathcal{L}$. We conclude that $\mathcal{L}$ is a λ-system, hence by Dynkin's $\pi-\lambda$ theorem, $\sigma(\mathcal{P}) \subseteq \mathcal{L}$, that is, $\mu_1 = \mu_2$. $\square$

Remark. With a somewhat more involved proof one can relax the condition $\mu_1(\Omega) = \mu_2(\Omega) < \infty$ to the existence of $A_n \in \mathcal{P}$ such that $A_n \uparrow \Omega$ and $\mu_1(A_n) < \infty$ (c.f. [**Bil95**, Theorem 10.3] for details). Accordingly, in Carathéodory's extension theorem we can relax $\mu_0(\Omega) < \infty$ to the assumption that μ_0 is a *σ-finite measure*, that is $\mu_0(A_n) < \infty$ for some $A_n \in \mathcal{A}$ such that $A_n \uparrow \Omega$, as is the case with *Lebesgue's measure* λ on $\mathbb{R}$.

We conclude this subsection with an outline the proof of Carathéodory's extension theorem, noting that since an algebra $\mathcal{A}$ is a π-system and $\Omega \in \mathcal{A}$, the uniqueness of the extension to $\sigma(\mathcal{A})$ follows from Proposition 1.1.39. Our outline of the existence of an extension follows [**Wil91**, Section A.1.8] (or see [**Bil95**, Theorem 11.3] for the proof of a somewhat stronger result). This outline centers on the construction of the appropriate outer measure, a relaxation of the concept of measure, which we now define.

Definition 1.1.40. *An increasing, countably sub-additive, non-negative set function μ^* on a measurable space $(\Omega, \mathcal{F})$ is called an* outer measure. *That is, $\mu^* : \mathcal{F} \mapsto [0, \infty]$, having the properties:*
(a) $\mu^*(\emptyset) = 0$ and $\mu^*(A_1) \le \mu^*(A_2)$ for any $A_1, A_2 \in \mathcal{F}$ with $A_1 \subseteq A_2$.
(b) $\mu^*(\bigcup_n A_n) \le \sum_n \mu^*(A_n)$ for any countable collection of sets $A_n \in \mathcal{F}$.

In the first step of the proof we define the increasing, non-negative set function

$$\mu^*(E) = \inf\{\sum_{n=1}^{\infty} \mu_0(A_n) : E \subseteq \bigcup_n A_n, A_n \in \mathcal{A}\},$$

for $E \in \mathcal{F} = 2^{\Omega}$, and prove that it is countably sub-additive, hence an outer measure on $\mathcal{F}$.

By definition, $\mu^*(A) \le \mu_0(A)$ for any $A \in \mathcal{A}$. In the second step we prove that if in addition $A \subseteq \bigcup_n A_n$ for $A_n \in \mathcal{A}$, then the countable additivity of μ_0 on $\mathcal{A}$ results with $\mu_0(A) \le \sum_n \mu_0(A_n)$. Consequently, $\mu^* = \mu_0$ on the algebra $\mathcal{A}$.

The third step uses the countable additivity of μ_0 on $\mathcal{A}$ to show that for any $A \in \mathcal{A}$ the outer measure μ^* is additive when splitting subsets of Ω by intersections with A and A^c. That is, we show that any element of $\mathcal{A}$ is a μ^*-measurable set, as defined next.

Definition 1.1.41. *Let λ be a non-negative set function on a measurable space $(\Omega, \mathcal{F})$, with $\lambda(\emptyset) = 0$. We say that $A \in \mathcal{F}$ is a λ-measurable set if $\lambda(F) = \lambda(F \cap A) + \lambda(F \cap A^c)$ for all $F \in \mathcal{F}$.*

The fourth step consists of proving the following general lemma.

Lemma 1.1.42 (Carathéodory's lemma)**.** *Let μ^* be an outer measure on a measurable space $(\Omega, \mathcal{F})$. Then the μ^*-measurable sets in $\mathcal{F}$ form a σ-algebra $\mathcal{G}$ on which μ^* is countably additive, so that $(\Omega, \mathcal{G}, \mu^*)$ is a measure space.*

In the current setting, with $\mathcal{A}$ contained in the σ-algebra $\mathcal{G}$, it follows that $\sigma(\mathcal{A}) \subseteq \mathcal{G}$ on which μ^* is a measure. Thus, the restriction μ of μ^* to $\sigma(\mathcal{A})$ is the stated measure that coincides with μ_0 on $\mathcal{A}$.

Remark. In the setting of Carathéodory's extension theorem for finite measures, we have that the σ-algebra $\mathcal{G}$ of all μ^*-measurable sets is the completion of $\sigma(\mathcal{A})$ with respect to μ (c.f. [**Bil95**, Page 45]). In the context of Lebesgue's measure U

on $\mathcal{B}_{(0,1]}$, this is the σ-algebra $\overline{\mathcal{B}}_{(0,1]}$ of all Lebesgue measurable subsets of $(0,1]$. Associated with it are the *Lebesgue measurable* functions $f:(0,1]\mapsto\mathbb{R}$ for which $f^{-1}(B)\in\overline{\mathcal{B}}_{(0,1]}$ for all $B\in\mathcal{B}$. However, as noted for example in [**Dur10**, Theorem A.2.4], the non Borel set constructed in the proof of Proposition 1.1.18 is also non Lebesgue measurable.

The following concept of a monotone class of sets is a considerable relaxation of that of a λ-system (hence also of a σ-algebra, see Proposition 1.1.37).

Definition 1.1.43. *A* monotone class *is a collection $\mathcal{M}$ of sets closed under both monotone increasing and monotone decreasing limits (i.e. if $A_i\in\mathcal{M}$ and either $A_i\uparrow A$ or $A_i\downarrow A$, then $A\in\mathcal{M}$).*

When starting from an algebra instead of a π-system, one may save effort by applying Halmos's monotone class theorem instead of Dynkin's $\pi-\lambda$ theorem.

Theorem 1.1.44 (Halmos's monotone class theorem)**.** *If $\mathcal{A}\subseteq\mathcal{M}$ with $\mathcal{A}$ an algebra and $\mathcal{M}$ a monotone class then $\sigma(\mathcal{A})\subseteq\mathcal{M}$.*

Proof. Clearly, any algebra which is a monotone class must be a σ-algebra. Another short though dense exercise in set manipulations shows that the intersection $m(\mathcal{A})$ of all monotone classes containing an algebra $\mathcal{A}$ is both an algebra and a monotone class (see the proof of [**Bil95**, Theorem 3.4]). Consequently, $m(\mathcal{A})$ is a σ-algebra. Since $\mathcal{A}\subseteq m(\mathcal{A})$ this implies that $\sigma(\mathcal{A})\subseteq m(\mathcal{A})$ and we complete the proof upon noting that $m(\mathcal{A})\subseteq\mathcal{M}$. $\square$

Exercise 1.1.45. *We say that a subset V of $\{1,2,3,\cdots\}$ has* Cesáro density $\gamma(V)$ *and write $V\in$ CES if the limit*

$$\gamma(V)=\lim_{n\to\infty}n^{-1}|V\cap\{1,2,3,\cdots,n\}|\,,$$

exists. Give an example of sets $V_1\in$ CES and $V_2\in$ CES for which $V_1\cap V_2\notin$ CES. Thus, CES *is not an algebra.*

Here is an alternative specification of the concept of algebra.

Exercise 1.1.46.

(a) *Suppose that $\Omega\in\mathcal{A}$ and that $A\cap B^c\in\mathcal{A}$ whenever $A,B\in\mathcal{A}$. Show that $\mathcal{A}$ is an algebra.*
(b) *Give an example of a collection $\mathcal{C}$ of subsets of Ω such that $\Omega\in\mathcal{C}$, if $A\in\mathcal{C}$ then $A^c\in\mathcal{C}$ and if $A,B\in\mathcal{C}$ are* disjoint *then also $A\cup B\in\mathcal{C}$, while $\mathcal{C}$ is not an algebra.*

As we already saw, the σ-algebra structure is preserved under intersections. However, whereas the increasing union of algebras is an algebra, it is not necessarily the case for σ-algebras.

Exercise 1.1.47. *Suppose that $\mathcal{A}_n$ are classes of sets such that $\mathcal{A}_n\subseteq\mathcal{A}_{n+1}$.*

(a) *Show that if $\mathcal{A}_n$ are algebras then so is $\bigcup_{n=1}^\infty\mathcal{A}_n$.*
(b) *Provide an example of σ-algebras $\mathcal{A}_n$ for which $\bigcup_{n=1}^\infty\mathcal{A}_n$ is not a σ-algebra.*

1.2. Random variables and their distribution

Random variables are numerical functions $\omega \mapsto X(\omega)$ of the outcome of our random experiment. However, in order to have a successful mathematical theory, we limit our interest to the subset of *measurable* functions (or more generally, measurable mappings), as defined in Subsection 1.2.1 and study the closure properties of this collection in Subsection 1.2.2. Subsection 1.2.3 is devoted to the characterization of the collection of distribution functions induced by random variables.

1.2.1. Indicators, simple functions and random variables. We start with the definition of random variables, first in the general case, and then restricted to $\mathbb{R}$-valued variables.

Definition 1.2.1. *A mapping $X : \Omega \mapsto \mathbb{S}$ between two measurable spaces $(\Omega, \mathcal{F})$ and $(\mathbb{S}, \mathcal{S})$ is called an $(\mathbb{S}, \mathcal{S})$-valued* Random Variable *(R.V.) if*

$$X^{-1}(B) := \{\omega : X(\omega) \in B\} \in \mathcal{F} \qquad \forall B \in \mathcal{S}.$$

Such a mapping is also called a measurable mapping.

Definition 1.2.2. *When we say that X is a random variable, or a* measurable function*, we mean an $(\mathbb{R}, \mathcal{B})$-valued random variable which is the most common type of R.V. we shall encounter. We let $m\mathcal{F}$ denote the collection of all $(\mathbb{R}, \mathcal{B})$-valued measurable mappings, so X is a R.V. if and only if $X \in m\mathcal{F}$. If in addition Ω is a topological space and $\mathcal{F} = \sigma(\{O \subseteq \Omega \text{ open }\})$ is the corresponding Borel σ-algebra, we say that $X : \Omega \mapsto \mathbb{R}$ is a* Borel *(measurable)* function*. More generally, a* random vector *is an $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$-valued R.V. for some $d < \infty$.*

The next exercise shows that a random vector is merely a finite collection of R.V. on the same probability space.

Exercise 1.2.3. *Relying on Exercise 1.1.21 and Theorem 1.2.9, show that $\underline{X} : \Omega \mapsto \mathbb{R}^d$ is a random vector if and only if $\underline{X}(\omega) = (X_1(\omega), \ldots, X_d(\omega))$ with each $X_i : \Omega \mapsto \mathbb{R}$ a R.V.*

Hint: *Note that $\underline{X}^{-1}(B_1 \times \ldots \times B_d) = \bigcap_{i=1}^{d} X_i^{-1}(B_i)$.*

We now provide two important generic examples of random variables.

Example 1.2.4. *For any $A \in \mathcal{F}$ the function $I_A(\omega) = \begin{cases} 1, \omega \in A \\ 0, \omega \notin A \end{cases}$ is a R.V. Indeed, $\{\omega : I_A(\omega) \in B\}$ is for any $B \subseteq \mathbb{R}$ one of the four sets $\emptyset$, A, A^c or Ω (depending on whether $0 \in B$ or not and whether $1 \in B$ or not), all of whom are in $\mathcal{F}$. We call such R.V. also an* indicator function*.*

Exercise 1.2.5. *By the same reasoning check that $X(\omega) = \sum_{n=1}^{N} c_n I_{A_n}(\omega)$ is a R.V. for any finite N, non-random $c_n \in \mathbb{R}$ and sets $A_n \in \mathcal{F}$. We call any such X a* simple function*, denoted by $X \in \mathrm{SF}$.*

Our next proposition explains why simple functions are quite useful in probability theory.

Proposition 1.2.6. *For every R.V. $X(\omega)$ there exists a sequence of simple functions $X_n(\omega)$ such that $X_n(\omega) \to X(\omega)$ as $n \to \infty$, for each fixed $\omega \in \Omega$.*

PROOF. Let

$$f_n(x) = n\mathbf{1}_{x>n} + \sum_{k=0}^{n2^n-1} k2^{-n}\mathbf{1}_{(k2^{-n},(k+1)2^{-n}]}(x)\,,$$

noting that for R.V. $X \geq 0$, we have that $X_n = f_n(X)$ are simple functions. Since $X \geq X_{n+1} \geq X_n$ and $X(\omega) - X_n(\omega) \leq 2^{-n}$ whenever $X(\omega) \leq n$, it follows that $X_n(\omega) \to X(\omega)$ as $n \to \infty$, for each ω.

We write a general R.V. as $X(\omega) = X_+(\omega) - X_-(\omega)$ where $X_+(\omega) = \max(X(\omega), 0)$ and $X_-(\omega) = -\min(X(\omega), 0)$ are non-negative R.V.-s. By the above argument the simple functions $X_n = f_n(X_+) - f_n(X_-)$ have the convergence property we claimed. □

Note that in case $\mathcal{F} = 2^\Omega$, every mapping $X : \Omega \mapsto \mathbb{S}$ is measurable (and therefore is an $(\mathbb{S}, \mathcal{S})$-valued R.V.). The choice of the σ-algebra $\mathcal{F}$ is very important in determining the class of all $(\mathbb{S}, \mathcal{S})$-valued R.V. For example, there are non-trivial σ-algebras $\mathcal{G}$ and $\mathcal{F}$ on $\Omega = \mathbb{R}$ such that $X(\omega) = \omega$ is a measurable function for $(\Omega, \mathcal{F})$, but is non-measurable for $(\Omega, \mathcal{G})$. Indeed, one such example is when $\mathcal{F}$ is the Borel σ-algebra $\mathcal{B}$ and $\mathcal{G} = \sigma(\{[a, b] : a, b \in \mathbf{Z}\})$ (for example, the set $\{\omega : \omega \leq \alpha\}$ is not in $\mathcal{G}$ whenever $\alpha \notin \mathbf{Z}$).

Building on Proposition 1.2.6 we have the following analog of Halmos's monotone class theorem. It allows us to deduce in the sequel general properties of (bounded) measurable functions upon verifying them only for indicators of elements of π-systems.

Theorem 1.2.7 (MONOTONE CLASS THEOREM). *Suppose $\mathcal{H}$ is a collection of $\mathbb{R}$-valued functions on Ω such that:*

(a) *The constant function 1 is an element of $\mathcal{H}$.*
(b) *$\mathcal{H}$ is a vector space over $\mathbb{R}$. That is, if $h_1, h_2 \in \mathcal{H}$ and $c_1, c_2 \in \mathbb{R}$ then $c_1h_1 + c_2h_2$ is in $\mathcal{H}$.*
(c) *If $h_n \in \mathcal{H}$ are non-negative and $h_n \uparrow h$ where h is a (bounded) real-valued function on Ω, then $h \in \mathcal{H}$.*

If $\mathcal{P}$ is a π-system and $I_A \in \mathcal{H}$ for all $A \in \mathcal{P}$, then $\mathcal{H}$ contains all (bounded) functions on Ω that are measurable with respect to $\sigma(\mathcal{P})$.

Remark. We stated here two versions of the monotone class theorem, with the less restrictive assumption that (c) holds only for bounded h yielding the weaker conclusion about bounded elements of $m\sigma(\mathcal{P})$. In the sequel we use both versions, which as we see next, are derived by essentially the same proof. Adapting this proof you can also show that any collection $\mathcal{H}$ of non-negative functions on Ω satisfying the conditions of Theorem 1.2.7 apart from requiring (b) to hold only when $c_1h_1 + c_2h_2 \geq 0$, must contain all non-negative elements of $m\sigma(\mathcal{P})$.

PROOF. Let $\mathcal{L} = \{A \subseteq \Omega : I_A \in \mathcal{H}\}$. From (a) we have that $\Omega \in \mathcal{L}$, while (b) implies that $B \setminus A$ is in $\mathcal{L}$ whenever $A \subseteq B$ are both in $\mathcal{L}$. Further, in view of (c) the collection $\mathcal{L}$ is closed under monotone increasing limits. Consequently, $\mathcal{L}$ is a λ-system, so by Dynkin's π-λ theorem, our assumption that $\mathcal{L}$ contains $\mathcal{P}$ results with $\sigma(\mathcal{P}) \subseteq \mathcal{L}$. With $\mathcal{H}$ a vector space over $\mathbb{R}$, this in turn implies that $\mathcal{H}$ contains all simple functions with respect to the measurable space $(\Omega, \sigma(\mathcal{P}))$. In the proof of Proposition 1.2.6 we saw that any (bounded) measurable function is a difference of two (bounded) non-negative functions each of which is a monotone increasing limit of certain non-negative simple functions. Thus, from (b) and (c) we conclude that $\mathcal{H}$ contains all (bounded) measurable functions with respect to $(\Omega, \sigma(\mathcal{P}))$. $\square$

The concept of almost sure prevails throughout probability theory.

Definition 1.2.8. *We say that two $(\mathbb{S}, \mathcal{S})$-valued R.V. X and Y defined on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$ are almost surely the same if $\mathbf{P}(\{\omega : X(\omega) \neq Y(\omega)\}) = 0$. This shall be denoted by $X \stackrel{a.s.}{=} Y$. More generally, same notation applies to any property of R.V. For example, $X(\omega) \geq 0$ a.s. means that $\mathbf{P}(\{\omega : X(\omega) < 0\}) = 0$. Hereafter, we shall consider X and Y such that $X \stackrel{a.s.}{=} Y$ to be the same $\mathbb{S}$-valued R.V. hence often omit the qualifier "a.s." when stating properties of R.V. We also use the terms* almost surely *(a.s.),* almost everywhere *(a.e.), and* with probability 1 *(w.p.1) interchangeably.*

Since the σ-algebra $\mathcal{S}$ might be huge, it is very important to note that we may verify that a given mapping is measurable without the need to check that the pre-image $X^{-1}(B)$ is in $\mathcal{F}$ for every $B \in \mathcal{S}$. Indeed, as shown next, it suffices to do this only for a collection (of our choice) of generators of $\mathcal{S}$.

Theorem 1.2.9. *If $\mathcal{S} = \sigma(\mathcal{A})$ and $X : \Omega \mapsto \mathbb{S}$ is such that $X^{-1}(A) \in \mathcal{F}$ for all $A \in \mathcal{A}$, then X is an $(\mathbb{S}, \mathcal{S})$-valued R.V.*

Proof. We first check that $\widehat{\mathcal{S}} = \{B \in \mathcal{S} : X^{-1}(B) \in \mathcal{F}\}$ is a σ-algebra. Indeed,
a). $\emptyset \in \widehat{\mathcal{S}}$ since $X^{-1}(\emptyset) = \emptyset$.
b). If $A \in \widehat{\mathcal{S}}$ then $X^{-1}(A) \in \mathcal{F}$. With $\mathcal{F}$ a σ-algebra, $X^{-1}(A^c) = \big(X^{-1}(A)\big)^c \in \mathcal{F}$. Consequently, $A^c \in \widehat{\mathcal{S}}$.
c). If $A_n \in \widehat{\mathcal{S}}$ for all n then $X^{-1}(A_n) \in \mathcal{F}$ for all n. With $\mathcal{F}$ a σ-algebra, then also $X^{-1}(\bigcup_n A_n) = \bigcup_n X^{-1}(A_n) \in \mathcal{F}$. Consequently, $\bigcup_n A_n \in \widehat{\mathcal{S}}$.
Our assumption that $\mathcal{A} \subseteq \widehat{\mathcal{S}}$, then translates to $\mathcal{S} = \sigma(\mathcal{A}) \subseteq \widehat{\mathcal{S}}$, as claimed. $\square$

The most important σ-algebras are those generated by ($(\mathbb{S}, \mathcal{S})$-valued) random variables, as defined next.

Exercise 1.2.10. *Adapting the proof of Theorem 1.2.9, show that for any mapping $X : \Omega \mapsto \mathbb{S}$ and any σ-algebra $\mathcal{S}$ of subsets of $\mathbb{S}$, the collection $\{X^{-1}(B) : B \in \mathcal{S}\}$ is a σ-algebra. Verify that X is an $(\mathbb{S}, \mathcal{S})$-valued R.V. if and only if $\{X^{-1}(B) : B \in \mathcal{S}\} \subseteq \mathcal{F}$, in which case we denote $\{X^{-1}(B) : B \in \mathcal{S}\}$ either by $\sigma(X)$ or by $\mathcal{F}^X$ and call it the* σ-algebra generated by X.

To practice your understanding of generated σ-algebras, solve the next exercise, providing a convenient collection of generators for $\sigma(X)$.

Exercise 1.2.11. *If X is an $(\mathbb{S}, \mathcal{S})$-valued R.V. and $\mathcal{S} = \sigma(\mathcal{A})$ then $\sigma(X)$ is generated by the collection of sets $X^{-1}(\mathcal{A}) := \{X^{-1}(A) : A \in \mathcal{A}\}$.*

An important example of use of Exercise 1.2.11 corresponds to $(\mathbb{R}, \mathcal{B})$-valued random variables and $\mathcal{A} = \{(-\infty, x] : x \in \mathbb{R}\}$ (or even $\mathcal{A} = \{(-\infty, x] : x \in \mathbb{Q}\}$) which generates $\mathcal{B}$ (see Exercise 1.1.17), leading to the following alternative definition of the σ-algebra generated by such R.V. X.

Definition 1.2.12. *Given a function $X : \Omega \mapsto \mathbb{R}$ we denote by $\sigma(X)$ or by $\mathcal{F}^X$ the* smallest *σ-algebra $\mathcal{F}$ such that $X(\omega)$ is a measurable mapping from $(\Omega, \mathcal{F})$ to $(\mathbb{R}, \mathcal{B})$. Alternatively,*

$$\sigma(X) = \sigma(\{\omega : X(\omega) \leq \alpha\}, \alpha \in \mathbb{R}) = \sigma(\{\omega : X(\omega) \leq q\}, q \in \mathbb{Q}) .$$

More generally, given a random vector $\underline{X} = (X_1, \ldots, X_n)$, that is, random variables $X_1, \ldots, X_n$ on the same probability space, let $\sigma(X_k, k \leq n)$ (or $\mathcal{F}_n^{\mathbf{X}}$), denote the smallest σ-algebra $\mathcal{F}$ such that $X_k(\omega)$, $k = 1, \ldots, n$ are measurable on $(\Omega, \mathcal{F})$. Alternatively,

$$\sigma(X_k, k \leq n) = \sigma(\{\omega : X_k(\omega) \leq \alpha\}, \alpha \in \mathbb{R}, k \leq n) .$$

Finally, given a possibly uncountable collection of functions $X_\gamma : \Omega \mapsto \mathbb{R}$, indexed by $\gamma \in \Gamma$, we denote by $\sigma(X_\gamma, \gamma \in \Gamma)$ (or simply by $\mathcal{F}^{\mathbf{X}}$), the smallest σ-algebra $\mathcal{F}$ such that $X_\gamma(\omega)$, $\gamma \in \Gamma$ are measurable on $(\Omega, \mathcal{F})$.

The concept of σ-algebra is needed in order to produce a rigorous mathematical theory. It further has the *crucial* role of quantifying the amount of information we have. For example, $\sigma(X)$ contains exactly those events A for which we can say whether $\omega \in A$ or not, based on the value of $X(\omega)$. Interpreting Example 1.1.19 as corresponding to sequentially tossing coins, the R.V. $X_n(\omega) = \omega_n$ gives the result of the n-th coin toss in our experiment Ω_∞ of infinitely many such tosses. The σ-algebra $\mathcal{F}_n = 2^{\Omega_n}$ of Example 1.1.6 then contains exactly the information we have upon observing the outcome of the first n coin tosses, whereas the larger σ-algebra $\mathcal{F}_c$ allows us to also study the limiting properties of this sequence (and as you show next, $\mathcal{F}_c$ is isomorphic, in the sense of Definition 1.4.24, to $\mathcal{B}_{[0,1]}$).

Exercise 1.2.13. *Let $\mathcal{F}_c$ denote the cylindrical σ-algebra for the set $\Omega_\infty = \{0, 1\}^{\mathbb{N}}$ of infinite binary sequences, as in Example 1.1.19.*

(a) *Show that $X(\omega) = \sum_{n=1}^{\infty} \omega_n 2^{-n}$ is a measurable map from $(\Omega_\infty, \mathcal{F}_c)$ to $([0, 1], \mathcal{B}_{[0,1]})$.*
(b) *Conversely, let $Y(x) = (\omega_1, \ldots, \omega_n, \ldots)$ where for each $n \geq 1$, $\omega_n(1) = 1$ while $\omega_n(x) = I(\lfloor 2^n x \rfloor$ is an odd number) when $x \in [0, 1)$. Show that $Y = X^{-1}$ is a measurable map from $([0, 1], \mathcal{B}_{[0,1]})$ to $(\Omega_\infty, \mathcal{F}_c)$.*

Here are some alternatives for Definition 1.2.12.

Exercise 1.2.14. *Verify the following relations and show that each generating collection of sets on the right hand side is a π-system.*

(a) $\sigma(X) = \sigma(\{\omega : X(\omega) \leq \alpha\}, \alpha \in \mathbb{R})$
(b) $\sigma(X_k, k \leq n) = \sigma(\{\omega : X_k(\omega) \leq \alpha_k, 1 \leq k \leq n\}, \alpha_1, \ldots, \alpha_n \in \mathbb{R})$
(c) $\sigma(X_1, X_2, \ldots) = \sigma(\{\omega : X_k(\omega) \leq \alpha_k, 1 \leq k \leq m\}, \alpha_1, \ldots, \alpha_m \in \mathbb{R}, m \in \mathbb{N})$
(d) $\sigma(X_1, X_2, \ldots) = \sigma(\bigcup_n \sigma(X_k, k \leq n))$

As you next show, when approximating a random variable by a simple function, one may also specify the latter to be based on sets in any generating algebra.

Exercise 1.2.15. *Suppose $(\Omega, \mathcal{F}, \mathbf{P})$ is a probability space, with $\mathcal{F} = \sigma(\mathcal{A})$ for an algebra $\mathcal{A}$.*

(a) *Show that* $\inf\{\mathbf{P}(A \Delta B) : A \in \mathcal{A}\} = 0$ *for any $B \in \mathcal{F}$ (recall that $A \Delta B = (A \cap B^c) \cup (A^c \cap B)$).*

(b) *Show that for any bounded random variable X and $\epsilon > 0$ there exists a simple function $Y = \sum_{n=1}^{N} c_n I_{A_n}$ with $A_n \in \mathcal{A}$ such that $\mathbf{P}(|X - Y| > \epsilon) < \epsilon$.*

Exercise 1.2.16. *Let $\mathcal{F} = \sigma(A_\alpha, \alpha \in \Gamma)$ and suppose there exist $\omega_1 \neq \omega_2 \in \Omega$ such that for any $\alpha \in \Gamma$, either $\{\omega_1, \omega_2\} \subseteq A_\alpha$ or $\{\omega_1, \omega_2\} \subseteq A_\alpha^c$.*

(a) *Show that if mapping X is measurable on $(\Omega, \mathcal{F})$ then $X(\omega_1) = X(\omega_2)$.*
(b) *Provide an explicit σ-algebra $\mathcal{F}$ of subsets of $\Omega = \{1, 2, 3\}$ and a mapping $X : \Omega \mapsto \mathbb{R}$ which is* not *a random variable on $(\Omega, \mathcal{F})$.*

We conclude with a glimpse of the canonical measurable space associated with a stochastic process $(X_t, t \in \mathbb{T})$ (for more on this, see Lemma 7.1.7).

Exercise 1.2.17. *Fixing a possibly uncountable collection of random variables X_t, indexed by $t \in \mathbb{T}$, let $\mathcal{F}^{\mathbf{X}}_{\mathbb{C}} = \sigma(X_t, t \in \mathbb{C})$ for each $\mathbb{C} \subseteq \mathbb{T}$. Show that*

$$\mathcal{F}^{\mathbf{X}}_{\mathbb{T}} = \bigcup_{\mathbb{C} \text{ countable}} \mathcal{F}^{\mathbf{X}}_{\mathbb{C}}$$

and that any R.V. Z on $(\Omega, \mathcal{F}^{\mathbf{X}}_{\mathbb{T}})$ is measurable on $\mathcal{F}^{\mathbf{X}}_{\mathbb{C}}$ for some countable $\mathbb{C} \subseteq \mathbb{T}$.

1.2.2. Closure properties of random variables. For the typical measurable space with uncountable Ω it is impractical to list all possible R.V. Instead, we state a few useful closure properties that often help us in showing that a given mapping $X(\omega)$ is indeed a R.V.

We start with closure with respect to the composition of a R.V. and a measurable mapping.

Proposition 1.2.18. *If $X : \Omega \mapsto \mathbb{S}$ is an $(\mathbb{S}, \mathcal{S})$-valued R.V. and f is a measurable mapping from $(\mathbb{S}, \mathcal{S})$ to $(\mathbb{T}, \mathcal{T})$, then the composition $f(X) : \Omega \mapsto \mathbb{T}$ is a $(\mathbb{T}, \mathcal{T})$-valued R.V.*

Proof. Considering an arbitrary $B \in \mathcal{T}$, we know that $f^{-1}(B) \in \mathcal{S}$ since f is a measurable mapping. Thus, as X is an $(\mathbb{S}, \mathcal{S})$-valued R.V. it follows that

$$[f(X)]^{-1}(B) = X^{-1}(f^{-1}(B)) \in \mathcal{F} .$$

This holds for any $B \in \mathcal{T}$, thus concluding the proof. □

In view of Exercise 1.2.3 we have the following special case of Proposition 1.2.18, corresponding to $\mathbb{S} = \mathbb{R}^n$ and $\mathbb{T} = \mathbb{R}$ equipped with the respective Borel σ-algebras.

Corollary 1.2.19. *Let X_i, $i = 1, \ldots, n$ be R.V. on the same measurable space $(\Omega, \mathcal{F})$ and $f : \mathbb{R}^n \mapsto \mathbb{R}$ a Borel function. Then, $f(X_1, \ldots, X_n)$ is also a R.V. on the same space.*

To appreciate the power of Corollary 1.2.19, consider the following exercise, in which you show that every continuous function is also a Borel function.

Exercise 1.2.20. *Suppose $(\mathbb{S}, \rho)$ is a metric space (for example, $\mathbb{S} = \mathbb{R}^n$). A function $g : \mathbb{S} \mapsto [-\infty, \infty]$ is called* lower semi-continuous *(l.s.c.) if $\liminf_{\rho(y,x) \downarrow 0} g(y) \geq g(x)$, for all $x \in \mathbb{S}$. A function g is said to be* upper semi-continuous*(u.s.c.) if $-g$ is l.s.c.*

(a) *Show that if g is l.s.c. then $\{x : g(x) \leq b\}$ is closed for each $b \in \mathbb{R}$.*
(b) *Conclude that semi-continuous functions are Borel measurable.*
(c) *Conclude that continuous functions are Borel measurable.*

A concrete application of Corollary 1.2.19 shows that any linear combination of finitely many R.V.-s is a R.V.

Example 1.2.21. *Suppose X_i are R.V.-s on the same measurable space and $c_i \in \mathbb{R}$. Then, $W_n(\omega) = \sum_{i=1}^n c_i X_i(\omega)$ are also R.V.-s. To see this, apply Corollary 1.2.19 for $f(x_1, \ldots, x_n) = \sum_{i=1}^n c_i x_i$ a continuous, hence Borel (measurable) function (by Exercise 1.2.20).*

We turn to explore the closure properties of $m\mathcal{F}$ with respect to operations of a limiting nature, starting with the following key theorem.

Theorem 1.2.22. *Let $\overline{\mathbb{R}} = [-\infty, \infty]$ equipped with its Borel σ-algebra*

$$\mathcal{B}_{\overline{\mathbb{R}}} = \sigma\left([-\infty, b) : \ b \in \mathbb{R}\right).$$

If X_i are $\overline{\mathbb{R}}$-valued R.V.-s on the same measurable space, then

$$\inf_n X_n, \quad \sup_n X_n, \quad \liminf_{n\to\infty} X_n, \quad \limsup_{n\to\infty} X_n \,,$$

are also $\overline{\mathbb{R}}$-valued random variables.

Proof. Pick an arbitrary $b \in \mathbb{R}$. Then,

$$\{\omega : \inf_n X_n(\omega) < b\} = \bigcup_{n=1}^{\infty} \{\omega : X_n(\omega) < b\} = \bigcup_{n=1}^{\infty} X_n^{-1}([-\infty, b)) \in \mathcal{F}.$$

Since $\mathcal{B}_{\overline{\mathbb{R}}}$ is generated by $\{[-\infty, b) : b \in \mathbb{R}\}$, it follows by Theorem 1.2.9 that $\inf_n X_n$ is an $\overline{\mathbb{R}}$-valued R.V.

Observing that $\sup_n X_n = -\inf_n(-X_n)$, we deduce from the above and Corollary 1.2.19 (for $f(x) = -x$), that $\sup_n X_n$ is also an $\overline{\mathbb{R}}$-valued R.V.

Next, recall that

$$W = \liminf_{n\to\infty} X_n = \sup_n \left[\inf_{l\geq n} X_l\right].$$

By the preceding proof we have that $Y_n = \inf_{l\geq n} X_l$ are $\overline{\mathbb{R}}$-valued R.V.-s and hence so is $W = \sup_n Y_n$.

Similarly to the arguments already used, we conclude the proof either by observing that

$$Z = \limsup_{n\to\infty} X_n = \inf_n \left[\sup_{l\geq n} X_l\right],$$

or by observing that $\limsup_n X_n = -\liminf_n(-X_n)$. $\square$

Remark. Since $\inf_n X_n$, $\sup_n X_n$, $\limsup_n X_n$ and $\liminf_n X_n$ may result in values $\pm\infty$ even when every X_n is $\mathbb{R}$-valued, hereafter we let $m\mathcal{F}$ also denote the collection of $\overline{\mathbb{R}}$-valued R.V.

An important corollary of this theorem deals with the existence of limits of sequences of R.V.

Corollary 1.2.23. *For any sequence $X_n \in m\mathcal{F}$, both*

$$\Omega_0 = \{\omega \in \Omega : \liminf_{n\to\infty} X_n(\omega) = \limsup_{n\to\infty} X_n(\omega)\}$$

and

$$\Omega_1 = \{\omega \in \Omega : \liminf_{n\to\infty} X_n(\omega) = \limsup_{n\to\infty} X_n(\omega) \in \mathbb{R}\}$$

are measurable sets, that is, $\Omega_0 \in \mathcal{F}$ and $\Omega_1 \in \mathcal{F}$.

Proof. By Theorem 1.2.22 we have that $Z = \limsup_n X_n$ and $W = \liminf_n X_n$ are two $\overline{\mathbb{R}}$-valued variables on the same space, with $Z(\omega) \geq W(\omega)$ for all ω. Hence, $\Omega_1 = \{\omega : Z(\omega) - W(\omega) = 0, Z(\omega) \in \mathbb{R}, W(\omega) \in \mathbb{R}\}$ is measurable (apply Corollary 1.2.19 for $f(z, w) = z - w$), as is $\Omega_0 = W^{-1}(\{\infty\}) \cup Z^{-1}(\{-\infty\}) \cup \Omega_1$. $\square$

The following structural result is yet another consequence of Theorem 1.2.22.

Corollary 1.2.24. *For any $d < \infty$ and R.V.-s $Y_1, \ldots, Y_d$ on the same measurable space $(\Omega, \mathcal{F})$ the collection $\mathcal{H} = \{h(Y_1, \ldots, Y_d); h : \mathbb{R}^d \mapsto \mathbb{R}$ Borel function$\}$ is a vector space over $\mathbb{R}$ containing the constant functions, such that if $X_n \in \mathcal{H}$ are non-negative and $X_n \uparrow X$, an $\mathbb{R}$-valued function on Ω, then $X \in \mathcal{H}$.*

Proof. By Example 1.2.21 the collection of all Borel functions is a vector space over $\mathbb{R}$ which evidently contains the constant functions. Consequently, the same applies for $\mathcal{H}$. Next, suppose $X_n = h_n(Y_1, \ldots, Y_d)$ for Borel functions h_n such that $0 \leq X_n(\omega) \uparrow X(\omega)$ for all $\omega \in \Omega$. Then, $\overline{h}(y) = \sup_n h_n(y)$ is by Theorem 1.2.22 an $\overline{\mathbb{R}}$-valued Borel function on $\mathbb{R}^d$, such that $X = \overline{h}(Y_1, \ldots, Y_d)$. Setting $h(y) = \overline{h}(y)$ when $\overline{h}(y) \in \mathbb{R}$ and $h(y) = 0$ otherwise, it is easy to check that h is a real-valued Borel function. Moreover, with $X : \Omega \mapsto \mathbb{R}$ (finite valued), necessarily $X = h(Y_1, \ldots, Y_d)$ as well, so $X \in \mathcal{H}$. $\square$

The point-wise convergence of R.V., that is $X_n(\omega) \to X(\omega)$, for every $\omega \in \Omega$ is often too strong of a requirement, as it may fail to hold as a result of the R.V. being ill-defined for a *negligible set* of values of ω (that is, a set of zero measure). We thus define the more useful, weaker notion of almost sure convergence of random variables.

Definition 1.2.25. *We say that a sequence of random variables X_n on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$* converges almost surely *if $\mathbf{P}(\Omega_0) = 1$. We then set $X_\infty = \limsup_{n\to\infty} X_n$, and say that X_n converges almost surely to X_∞, or use the notation $X_n \stackrel{a.s.}{\to} X_\infty$.*

Remark. Note that in Definition 1.2.25 we allow the limit $X_\infty(\omega)$ to take the values $\pm\infty$ with positive probability. So, we say that X_n converges almost surely to a *finite limit* if $\mathbf{P}(\Omega_1) = 1$, or alternatively, if $X_\infty \in \mathbb{R}$ with probability one.

We proceed with an explicit characterization of the functions measurable with respect to a σ-algebra of the form $\sigma(Y_k, k \leq n)$.

Theorem 1.2.26. *Let $\mathcal{G} = \sigma(Y_k, k \leq n)$ for some $n < \infty$ and R.V.-s $Y_1, \ldots, Y_n$ on the same measurable space $(\Omega, \mathcal{F})$. Then, $m\mathcal{G} = \{g(Y_1, \ldots, Y_n) : g : \mathbb{R}^n \mapsto \mathbb{R}$ is a Borel function$\}$.*

Proof. From Corollary 1.2.19 we know that $Z = g(Y_1, \ldots, Y_n)$ is in $m\mathcal{G}$ for each Borel function $g : \mathbb{R}^n \mapsto \mathbb{R}$. Turning to prove the converse result, recall part (b) of Exercise 1.2.14 that the σ-algebra $\mathcal{G}$ is generated by the π-system $\mathcal{P} = \{A_{\underline{\alpha}} : \underline{\alpha} = (\alpha_1, \ldots, \alpha_n) \in \mathbb{R}^n\}$ where $I_{A_{\underline{\alpha}}} = h_{\underline{\alpha}}(Y_1, \ldots, Y_n)$ for the Borel function $h_{\underline{\alpha}}(y_1, \ldots, y_n) = \prod_{k=1}^n \mathbf{1}_{y_k \leq \alpha_k}$. Thus, in view of Corollary 1.2.24, we have by the monotone class theorem that $\mathcal{H} = \{g(Y_1, \ldots, Y_n) : g : \mathbb{R}^n \mapsto \mathbb{R}$ is a Borel function$\}$ contains all elements of $m\mathcal{G}$. $\square$

We conclude this sub-section with a few exercises, starting with Borel measurability of monotone functions (regardless of their continuity properties).

Exercise 1.2.27. *Show that any monotone function $g : \mathbb{R} \mapsto \mathbb{R}$ is Borel measurable.*

Next, Exercise 1.2.20 implies that the set of points at which a given function g is discontinuous, is a Borel set.

Exercise 1.2.28. *Fix an arbitrary function $g : \mathbb{S} \mapsto \mathbb{R}$.*

(a) *Show that for any $\delta > 0$ the function $g_*(x, \delta) = \inf\{g(y) : \rho(x, y) < \delta\}$ is u.s.c. and the function $g^*(x, \delta) = \sup\{g(y) : \rho(x, y) < \delta\}$ is l.s.c.*
(b) *Show that $\mathbf{D}_g = \{x : \sup_k g_*(x, k^{-1}) < \inf_k g^*(x, k^{-1})\}$ is exactly the set of points at which g is discontinuous.*
(c) *Deduce that the set $\mathbf{D}_g$ of points of discontinuity of g is a Borel set.*

Here is an alternative characterization of $\mathcal{B}$ that complements Exercise 1.2.20.

Exercise 1.2.29. *Show that if $\mathcal{F}$ is a σ-algebra of subsets of $\mathbb{R}$ then $\mathcal{B} \subseteq \mathcal{F}$ if and only if every continuous function $f : \mathbb{R} \mapsto \mathbb{R}$ is in $m\mathcal{F}$ (i.e. $\mathcal{B}$ is the smallest σ-algebra on $\mathbb{R}$ with respect to which all continuous functions are measurable).*

Exercise 1.2.30. *Suppose X_n and X_∞ are real-valued random variables and*

$$\mathbf{P}(\{\omega : \limsup_{n\to\infty} X_n(\omega) \leq X_\infty(\omega)\}) = 1 .$$

Show that for any $\varepsilon > 0$, there exists an event A with $\mathbf{P}(A) < \varepsilon$ and a non-random $N = N(\epsilon)$, sufficiently large such that $X_n(\omega) < X_\infty(\omega) + \varepsilon$ for all $n \geq N$ and every $\omega \in A^c$.

Equipped with Theorem 1.2.22 you can also strengthen Proposition 1.2.6.

Exercise 1.2.31. *Show that the class $m\mathcal{F}$ of $\overline{\mathbb{R}}$-valued measurable functions, is the smallest class containing* SF *and closed under point-wise limits.*

Your next exercise also relies on Theorem 1.2.22.

Exercise 1.2.32. *Given a measurable space $(\Omega, \mathcal{F})$ and $\Gamma \subseteq \Omega$ (not necessarily in $\mathcal{F}$), let $\mathcal{F}_\Gamma = \{A \cap \Gamma : A \in \mathcal{F}\}$.*

(a) *Check that $(\Gamma, \mathcal{F}_\Gamma)$ is a measurable space.*
(b) *Show that any bounded, $\mathcal{F}_\Gamma$-measurable function (on Γ), is the restriction to Γ of some bounded, $\mathcal{F}$-measurable $f : \Omega \to \mathbb{R}$.*

Finally, relying on Theorem 1.2.26 it is easy to show that a Borel function can only reduce the amount of information quantified by the corresponding generated σ-algebras, whereas such information content is invariant under invertible Borel transformations, that is

Exercise 1.2.33. *Show that $\sigma(g(Y_1, \ldots, Y_n)) \subseteq \sigma(Y_k, k \leq n)$ for any Borel function $g : \mathbb{R}^n \mapsto \mathbb{R}$. Further, if $Y_1, \ldots, Y_n$ and $Z_1, \ldots, Z_m$ defined on the same probability space are such that $Z_k = g_k(Y_1, \ldots, Y_n)$, $k = 1, \ldots, m$ and $Y_i = h_i(Z_1, \ldots, Z_m)$, $i = 1, \ldots, n$ for some Borel functions $g_k : \mathbb{R}^n \mapsto \mathbb{R}$ and $h_i : \mathbb{R}^m \mapsto \mathbb{R}$, then $\sigma(Y_1, \ldots, Y_n) = \sigma(Z_1, \ldots, Z_m)$.*

1.2.3. Distribution, density and law.

As defined next, every random variable X induces a probability measure on its range which is called the law of X.

Definition 1.2.34. *The* law *of a real-valued R.V. X, denoted $\mathcal{P}_X$, is the probability measure on $(\mathbb{R}, \mathcal{B})$ such that $\mathcal{P}_X(B) = \mathbf{P}(\{\omega : X(\omega) \in B\})$ for any Borel set B.*

Remark. Since X is a R.V., it follows that $\mathcal{P}_X(B)$ is well defined for all $B \in \mathcal{B}$. Further, the non-negativity of $\mathbf{P}$ implies that $\mathcal{P}_X$ is a non-negative set function on $(\mathbb{R}, \mathcal{B})$, and since $X^{-1}(\mathbb{R}) = \Omega$, also $\mathcal{P}_X(\mathbb{R}) = 1$. Consider next disjoint Borel sets B_i, observing that $X^{-1}(B_i) \in \mathcal{F}$ are disjoint subsets of Ω such that

$$X^{-1}(\bigcup_i B_i) = \bigcup_i X^{-1}(B_i)\,.$$

Thus, by the countable additivity of $\mathbf{P}$ we have that

$$\mathcal{P}_X(\bigcup_i B_i) = \mathbf{P}(\bigcup_i X^{-1}(B_i)) = \sum_i \mathbf{P}(X^{-1}(B_i)) = \sum_i \mathcal{P}_X(B_i)\,.$$

This shows that $\mathcal{P}_X$ is also countably additive, hence a probability measure, as claimed in Definition 1.2.34.

Note that the law $\mathcal{P}_X$ of a R.V. $X : \Omega \longrightarrow \mathbb{R}$, determines the values of the probability measure $\mathbf{P}$ on $\sigma(X)$.

Definition 1.2.35. *We write $X \overset{\mathcal{D}}{=} Y$ and say that X equals Y in law (or in distribution), if and only if $\mathcal{P}_X = \mathcal{P}_Y$.*

A good way to practice your understanding of the Definitions 1.2.34 and 1.2.35 is by verifying that if $X \overset{a.s.}{=} Y$, then also $X \overset{\mathcal{D}}{=} Y$ (that is, any two random variables we consider to be the same would indeed have the same law).

The next concept we define, the distribution function, is closely associated with the law $\mathcal{P}_X$ of the R.V.

Definition 1.2.36. *The* distribution function F_X *of a real-valued R.V. X is*

$$F_X(\alpha) = \mathbf{P}(\{\omega : X(\omega) \le \alpha\}) = \mathcal{P}_X((-\infty, \alpha]) \qquad \forall \alpha \in \mathbb{R}$$

Our next result characterizes the set of all functions $F : \mathbb{R} \mapsto [0,1]$ that are distribution functions of some R.V.

Theorem 1.2.37. *A function $F : \mathbb{R} \mapsto [0,1]$ is a distribution function of some R.V. if and only if*

(a) *F is non-decreasing*
(b) $\lim_{x\to\infty} F(x) = 1$ *and* $\lim_{x\to-\infty} F(x) = 0$
(c) *F is right-continuous, i.e.* $\lim_{y \downarrow x} F(y) = F(x)$

Proof. First, assuming that $F = F_X$ is a distribution function, we show that it must have the stated properties (a)-(c). Indeed, if $x \le y$ then $(-\infty, x] \subseteq (-\infty, y]$, and by the monotonicity of the probability measure $\mathcal{P}_X$ (see part (a) of Exercise 1.1.4), we have that $F_X(x) \le F_X(y)$, proving that F_X is non-decreasing. Further, $(-\infty, x] \uparrow \mathbb{R}$ as $x \uparrow \infty$, while $(-\infty, x] \downarrow \emptyset$ as $x \downarrow -\infty$, resulting with property (b) of the theorem by the continuity from below and the continuity from above of the probability measure $\mathcal{P}_X$ on $\mathbb{R}$. Similarly, since $(-\infty, y] \downarrow (-\infty, x]$ as $y \downarrow x$ we get the right continuity of F_X by yet another application of continuity from above of $\mathcal{P}_X$.

We proceed to prove the converse result, that is, assuming F has the stated properties (a)-(c), we consider the random variable $X^-(\omega) = \sup\{y : F(y) < \omega\}$ on the probability space $((0,1], \mathcal{B}_{(0,1]}, U)$ and show that $F_{X^-} = F$. With F having property (b), we see that for any $\omega > 0$ the set $\{y : F(y) < \omega\}$ is non-empty and further if $\omega < 1$ then $X^-(\omega) < \infty$, so $X^- : (0,1) \mapsto \mathbb{R}$ is well defined. The identity

$$\{\omega : X^-(\omega) \leq x\} = \{\omega : \omega \leq F(x)\}\,, \tag{1.2.1}$$

implies that $F_{X^-}(x) = U((0, F(x)]) = F(x)$ for all $x \in \mathbb{R}$, and further, the sets $(0, F(x)]$ are all in $\mathcal{B}_{(0,1]}$, implying that X^- is a measurable function (i.e. a R.V.).
Turning to prove (1.2.1) note that if $\omega \leq F(x)$ then $x \notin \{y : F(y) < \omega\}$ and so by definition (and the monotonicity of F), $X^-(\omega) \leq x$. Now suppose that $\omega > F(x)$. Since F is right continuous, this implies that $F(x+\epsilon) < \omega$ for some $\epsilon > 0$, hence by definition of X^- also $X^-(\omega) \geq x + \epsilon > x$, completing the proof of (1.2.1) and with it the proof of the theorem. □

Check your understanding of the preceding proof by showing that the collection of distribution functions for $\overline{\mathbb{R}}$-valued random variables consist of all $F : \mathbb{R} \mapsto [0,1]$ that are non-decreasing and right-continuous.

Remark. The construction of the random variable $X^-(\omega)$ in Theorem 1.2.37 is called *Skorokhod's representation.* You can, and should, verify that the random variable $X^+(\omega) = \sup\{y : F(y) \leq \omega\}$ would have worked equally well for that purpose, since $X^+(\omega) \neq X^-(\omega)$ only if $X^+(\omega) > q \geq X^-(\omega)$ for some rational q, in which case by definition $\omega \geq F(q) \geq \omega$, so there are most countably many such values of ω (hence $\mathbf{P}(X^+ \neq X^-) = 0$). We shall return to this construction when dealing with convergence in distribution in Section 3.2. An alternative approach to Theorem 1.2.37 is to adapt the construction of the probability measure of Example 1.1.26, taking here $\Omega = \mathbb{R}$ with the corresponding change to $\mathcal{A}$ and replacing the right side of (1.1.1) with $\sum_{k=1}^{r}(F(b_k) - F(a_k))$, yielding a probability measure $\mathcal{P}$ on $(\mathbb{R}, \mathcal{B})$ such that $\mathcal{P}((-\infty, \alpha]) = F(\alpha)$ for all $\alpha \in \mathbb{R}$ (c.f. [**Bil95**, Theorem 12.4]).

Our next example highlights the possible shape of the distribution function.

Example 1.2.38. *Consider Example 1.1.6 of n coin tosses, with σ-algebra $\mathcal{F}_n = 2^{\Omega_n}$, sample space $\Omega_n = \{H, T\}^n$, and the probability measure $\mathbf{P}_n(A) = \sum_{\omega \in A} p_\omega$, where $p_\omega = 2^{-n}$ for each $\omega \in \Omega_n$ (that is, $\omega = \{\omega_1, \omega_2, \cdots, \omega_n\}$ for $\omega_i \in \{H, T\}$), corresponding to independent, fair, coin tosses. Let $Y(\omega) = I_{\{\omega_1 = H\}}$ measure the outcome of the first toss. The law of this random variable is,*

$$\mathcal{P}_Y(B) = \frac{1}{2}\mathbf{1}_{\{0 \in B\}} + \frac{1}{2}\mathbf{1}_{\{1 \in B\}}$$

and its distribution function is

$$F_Y(\alpha) = \mathcal{P}_Y((-\infty, \alpha]) = \mathbf{P}_n(Y(\omega) \leq \alpha) = \begin{cases} 1, & \alpha \geq 1 \\ \frac{1}{2}, & 0 \leq \alpha < 1 \\ 0, & \alpha < 0 \end{cases}\,. \tag{1.2.2}$$

Note that in general $\sigma(X)$ is a strict subset of the σ-algebra $\mathcal{F}$ (in Example 1.2.38 we have that $\sigma(Y)$ determines the probability measure for the first coin toss, but tells us nothing about the probability measure assigned to the remaining $n-1$ tosses). Consequently, though the law $\mathcal{P}_X$ determines the probability measure $\mathbf{P}$ on $\sigma(X)$ it usually does not completely determine $\mathbf{P}$.

Example 1.2.38 is somewhat generic. That is, if the R.V. X is a simple function (or more generally, when the set $\{X(\omega) : \omega \in \Omega\}$ is countable and has no accumulation points), then its distribution function F_X is piecewise constant with jumps at the possible values that X takes and jump sizes that are the corresponding probabilities. Indeed, note that $(-\infty, y] \uparrow (-\infty, x)$ as $y \uparrow x$, so by the continuity from below of $\mathcal{P}_X$ it follows that

$$F_X(x^-) := \lim_{y \uparrow x} F_X(y) = \mathbf{P}(\{\omega : X(\omega) < x\}) = F_X(x) - \mathbf{P}(\{\omega : X(\omega) = x\}),$$

for any R.V. X.

A direct corollary of Theorem 1.2.37 shows that any distribution function has a collection of continuity points that is dense in $\mathbb{R}$.

Exercise 1.2.39. *Show that a distribution function F has at most countably many points of discontinuity and consequently, that for any $x \in \mathbb{R}$ there exist y_k and z_k at which F is continuous such that $z_k \downarrow x$ and $y_k \uparrow x$.*

In contrast with Example 1.2.38 the distribution function of a R.V. with a density is continuous and almost everywhere differentiable, that is,

Definition 1.2.40. *We say that a R.V. $X(\omega)$ has a* probability density function *f_X if and only if its distribution function F_X can be expressed as*

$$F_X(\alpha) = \int_{-\infty}^{\alpha} f_X(x)dx\,, \qquad \forall \alpha \in \mathbb{R}\,. \tag{1.2.3}$$

By Theorem 1.2.37 a probability density function f_X must be an integrable, Lebesgue almost everywhere non-negative function, with $\int_{\mathbb{R}} f_X(x)dx = 1$. Such F_X is continuous with $\frac{dF_X}{dx}(x) = f_X(x)$ except possibly on a set of values of x of zero Lebesgue measure.

Remark. To make Definition 1.2.40 precise we temporarily assume that probability density functions f_X are Riemann integrable and interpret the integral in (1.2.3) in this sense. In Section 1.3 we construct Lebesgue's integral and extend the scope of Definition 1.2.40 to *Lebesgue integrable* density functions $f_X \geq 0$ (in particular, accommodating Borel functions f_X). This is the setting we assume thereafter, with the right-hand-side of (1.2.3) interpreted as the integral $\overline{\lambda}(f_X; (-\infty, \alpha])$ of f_X with respect to the restriction on $(-\infty, \alpha]$ of the *completion* $\overline{\lambda}$ of the *Lebesgue measure* on $\mathbb{R}$ (c.f. Definition 1.3.59 and Example 1.3.60). Further, the function f_X is uniquely defined only as a representative of an equivalence class. That is, in this context we consider f and g to be the same function when $\overline{\lambda}(\{x : f(x) \neq g(x)\}) = 0$.

Building on Example 1.1.26 we next detail a few classical examples of R.V. that have densities.

Example 1.2.41. *The distribution function F_U of the R.V. of Example 1.1.26 is*

$$F_U(\alpha) = \mathbf{P}(U \leq \alpha) = \mathbf{P}(U \in [0, \alpha]) = \begin{cases} 1, & \alpha > 1 \\ \alpha, & 0 \leq \alpha \leq 1 \\ 0, & \alpha < 0 \end{cases} \tag{1.2.4}$$

and its density is $f_U(u) = \begin{cases} 1, & 0 \leq u \leq 1 \\ 0, & \textit{otherwise} \end{cases}$.

The exponential distribution function *is*

$$F(x) = \begin{cases} 0, & x \leq 0 \\ 1 - e^{-x}, & x \geq 0 \end{cases},$$

corresponding to the density $f(x) = \begin{cases} 0, \ x \le 0 \\ e^{-x}, \ x > 0 \end{cases}$ *, whereas the* standard normal distribution *has the density*

$$\phi(x) = (2\pi)^{-1/2} e^{-\frac{x^2}{2}} ,$$

with no closed form expression for the corresponding distribution function $\Phi(x) = \int^x \phi(u)du$ *in terms of elementary functions.*

Every real-valued R.V. X has a distribution function but not necessarily a density. For example $X = 0$ w.p.1 has distribution function $F_X(\alpha) = \mathbf{1}_{\alpha \ge 0}$. Since F_X is discontinuous at 0, the R.V. X does not have a density.

Definition 1.2.42. *We say that a function* F *is a* Lebesgue singular function *if it has a zero derivative except on a set of zero Lebesgue measure.*

Since the distribution function of any R.V. is non-decreasing, from real analysis we know that it is almost everywhere differentiable. However, perhaps somewhat surprisingly, there are continuous distribution functions that are Lebesgue singular functions. Consequently, there are non-discrete random variables that do not have a density. We next provide one such example.

Example 1.2.43. *The* Cantor set C *is defined by removing* $(1/3, 2/3)$ *from* $[0, 1]$ *and then iteratively removing the middle third of each interval that remains. The* uniform distribution *on the (closed) set* C *corresponds to the distribution function obtained by setting* $F(x) = 0$ *for* $x \le 0$, $F(x) = 1$ *for* $x \ge 1$, $F(x) = 1/2$ *for* $x \in [1/3, 2/3]$, *then* $F(x) = 1/4$ *for* $x \in [1/9, 2/9]$, $F(x) = 3/4$ *for* $x \in [7/9, 8/9]$, *and so on (which as you should check, satisfies the properties (a)-(c) of Theorem 1.2.37). From the definition, we see that* $dF/dx = 0$ *for almost every* $x \notin C$ *and that the corresponding probability measure has* $\mathbf{P}(C^c) = 0$. *As the Lebesgue measure of* C *is zero, we see that the derivative of* F *is zero except on a set of zero Lebesgue measure, and consequently, there is no function* f *for which* $F(x) = \int_{-\infty}^x f(y)dy$ *holds. Though it is somewhat more involved, you may want to check that* F *is everywhere continuous (c.f.* [**Bil95**, Problem 31.2]*).*

Even discrete distribution functions can be quite complex. As the next example shows, the points of discontinuity of such a function might form a (countable) dense subset of $\mathbb{R}$ (which in a sense is extreme, per Exercise 1.2.39).

Example 1.2.44. *Let* $q_1, q_2, \ldots$ *be an enumeration of the rational numbers and set*

$$F(x) = \sum_{i=1}^{\infty} 2^{-i} \mathbf{1}_{[q_i,\infty)}(x)$$

(where $\mathbf{1}_{[q_i,\infty)}(x) = 1$ *if* $x \ge q_i$ *and zero otherwise). Clearly, such* F *is non-decreasing, with limits* 0 *and* 1 *as* $x \to -\infty$ *and* $x \to \infty$*, respectively. It is not hard to check that* F *is also right continuous, hence a distribution function, whereas by construction* F *is discontinuous at each rational number.*

As we have that $\mathbf{P}(\{\omega : X(\omega) \le \alpha\}) = F_X(\alpha)$ for the generators $\{\omega : X(\omega) \le \alpha\}$ of $\sigma(X)$, we are not at all surprised by the following proposition.

Proposition 1.2.45. *The distribution function* F_X *uniquely determines the law* $\mathcal{P}_X$ *of* X.

PROOF. Consider the collection $\pi(\mathbb{R}) = \{(-\infty, b] : b \in \mathbb{R}\}$ of subsets of $\mathbb{R}$. It is easy to see that $\pi(\mathbb{R})$ is a π-system, which generates $\mathcal{B}$ (see Exercise 1.1.17). Hence, by Proposition 1.1.39, any two probability measures on $(\mathbb{R}, \mathcal{B})$ that coincide on $\pi(\mathbb{R})$ are the same. Since the distribution function F_X specifies the restriction of such a probability measure $\mathcal{P}_X$ on $\pi(\mathbb{R})$ it thus uniquely determines the values of $\mathcal{P}_X(B)$ for all $B \in \mathcal{B}$. $\square$

Different probability measures $\mathbf{P}$ on the measurable space $(\Omega, \mathcal{F})$ may "trivialize" different σ-algebras. That is,

Definition 1.2.46. *If a σ-algebra $\mathcal{H} \subseteq \mathcal{F}$ and a probability measure $\mathbf{P}$ on $(\Omega, \mathcal{F})$ are such that $\mathbf{P}(H) \in \{0, 1\}$ for all $H \in \mathcal{H}$, we call $\mathcal{H}$ a* $\mathbf{P}$-trivial *σ-algebra. Similarly, a random variable X is called* $\mathbf{P}$-trivial *or* $\mathbf{P}$-degenerate*, if there exists a non-random constant c such that $\mathbf{P}(X \neq c) = 0$.*

Using distribution functions we show next that all random variables on a $\mathbf{P}$-trivial σ-algebra are $\mathbf{P}$-trivial.

Proposition 1.2.47. *If a random variable $X \in m\mathcal{H}$ for a $\mathbf{P}$-trivial σ-algebra $\mathcal{H}$, then X is $\mathbf{P}$-trivial.*

PROOF. By definition, the sets $\{\omega : X(\omega) \leq \alpha\}$ are in $\mathcal{H}$ for all $\alpha \in \mathbb{R}$. Since $\mathcal{H}$ is $\mathbf{P}$-trivial this implies that $F_X(\alpha) \in \{0, 1\}$ for all $\alpha \in \mathbb{R}$. In view of Theorem 1.2.37 this is possible only if $F_X(\alpha) = \mathbf{1}_{\alpha \geq c}$ for some non-random $c \in \mathbb{R}$ (for example, set $c = \inf\{\alpha : F_X(\alpha) = 1\}$). That is, $\mathbf{P}(X \neq c) = 0$, as claimed. $\square$

We conclude with few exercises about the support of measures on $(\mathbb{R}, \mathcal{B})$.

Exercise 1.2.48. *Let μ be a measure on $(\mathbb{R}, \mathcal{B})$. A point x is said to be in the* support *of μ if $\mu(O) > 0$ for every open neighborhood O of x. Prove that the support is a closed set whose complement is the maximal open set on which μ vanishes.*

Exercise 1.2.49. *Given an arbitrary closed set $C \subseteq \mathbb{R}$, construct a probability measure on $(\mathbb{R}, \mathcal{B})$ whose support is C.*
Hint: *Try a measure consisting of a countable collection of atoms (i.e. points of positive probability).*

As you are to check next, the discontinuity points of a distribution function are closely related to the support of the corresponding law.

Exercise 1.2.50. *The* support *of a distribution function F is the set $S_F = \{x \in \mathbb{R}$ such that $F(x + \epsilon) - F(x - \epsilon) > 0$ for all $\epsilon > 0\}$.*

(a) *Show that all points of discontinuity of $F(\cdot)$ belong to S_F, and that any isolated point of S_F (that is, $x \in S_F$ such that $(x - \delta, x + \delta) \cap S_F = \{x\}$ for some $\delta > 0$) must be a point of discontinuity of $F(\cdot)$.*
(b) *Show that the support of the law $\mathcal{P}_X$ of a random variable X, as defined in Exercise 1.2.48, is the same as the support of its distribution function F_X.*

1.3. Integration and the (mathematical) expectation

A key concept in probability theory is the mathematical expectation of random variables. In Subsection 1.3.1 we provide its definition via the framework of Lebesgue integration with respect to a measure and study properties such as monotonicity and linearity. In Subsection 1.3.2 we consider fundamental inequalities associated with the expectation. Subsection 1.3.3 is about the exchange of integration and limit operations, complemented by uniform integrability and its consequences in Subsection 1.3.4. Subsection 1.3.5 considers densities relative to arbitrary measures and relates our treatment of integration and expectation to Riemann's integral and the classical definition of the expectation for a R.V. with probability density. We conclude with Subsection 1.3.6 about moments of random variables, including their values for a few well known distributions.

1.3.1. Lebesgue integral, linearity and monotonicity. Let SF_+ denote the collection of non-negative simple functions with respect to the given measurable space $(\mathbb{S}, \mathcal{F})$ and $m\mathcal{F}_+$ denote the collection of $[0, \infty]$-valued measurable functions on this space. We next define Lebesgue's integral with respect to any measure μ on $(\mathbb{S}, \mathcal{F})$, first for $\varphi \in \mathrm{SF}_+$, then extending it to all $f \in m\mathcal{F}_+$. With the notation $\mu(f) := \int_{\mathbb{S}} f(s)d\mu(s)$ for this integral, we also denote by $\mu_0(\cdot)$ the more restrictive integral, defined only on SF_+, so as to clarify the role each of these plays in some of our proofs. We call an $\overline{\mathbb{R}}$-valued measurable function $f \in m\mathcal{F}$ for which $\mu(|f|) < \infty$, a *μ-integrable* function, and denote the collection of all μ-integrable functions by $L^1(\mathbb{S}, \mathcal{F}, \mu)$, extending the definition of the integral $\mu(f)$ to all $f \in L^1(\mathbb{S}, \mathcal{F}, \mu)$.

Definition 1.3.1. *Fix a measure space $(\mathbb{S}, \mathcal{F}, \mu)$ and define $\mu(f)$ by the following four step procedure:*

Step 1. *Define $\mu_0(I_A) := \mu(A)$ for each $A \in \mathcal{F}$.*

Step 2. *Any $\varphi \in \mathrm{SF}_+$ has a representation $\varphi = \sum_{l=1}^{n} c_l I_{A_l}$ for some finite $n < \infty$, non-random $c_l \in [0, \infty]$ and sets $A_l \in \mathcal{F}$, yielding the definition of the integral via*

$$\mu_0(\varphi) := \sum_{l=1}^{n} c_l \mu(A_l)\,,$$

where we adopt hereafter the convention that $\infty \times 0 = 0 \times \infty = 0$.

Step 3. *For $f \in m\mathcal{F}_+$ we define*

$$\mu(f) := \sup\{\mu_0(\varphi) : \varphi \in \mathrm{SF}_+, \varphi \leq f\}.$$

Step 4. *For $f \in m\mathcal{F}$ let $f_+ = \max(f, 0) \in m\mathcal{F}_+$ and $f_- = -\min(f, 0) \in m\mathcal{F}_+$. We then set $\mu(f) = \mu(f_+) - \mu(f_-)$ provided either $\mu(f_+) < \infty$ or $\mu(f_-) < \infty$. In particular, this applies whenever $f \in L^1(\mathbb{S}, \mathcal{F}, \mu)$, for then $\mu(f_+) + \mu(f_-) = \mu(|f|)$ is finite, hence $\mu(f)$ is well defined and finite valued.*

We use the notation $\int_{\mathbb{S}} f(s)d\mu(s)$ for $\mu(f)$ which we call Lebesgue integral *of f with respect to the measure μ.*

The *expectation* $\mathbf{E}[X]$ of a random variable X on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ is merely Lebesgue's integral $\int X(\omega)d\mathbf{P}(\omega)$ of X with respect to $\mathbf{P}$. That is,

Step 1. $\mathbf{E}[I_A] = \mathbf{P}(A)$ for any $A \in \mathcal{F}$.

Step 2. Any $\varphi \in \mathrm{SF}_+$ has a representation $\varphi = \sum_{l=1}^{n} c_l I_{A_l}$ for some non-random $n < \infty$, $c_l \geq 0$ and sets $A_l \in \mathcal{F}$, to which corresponds

$$\mathbf{E}[\varphi] = \sum_{l=1}^{n} c_l \mathbf{E}[I_{A_l}] = \sum_{l=1}^{n} c_l \mathbf{P}(A_l)\,.$$

Step 3. For $X \in m\mathcal{F}_+$ define

$$\mathbf{E}X = \sup\{\mathbf{E}Y : Y \in \mathrm{SF}_+, Y \leq X\}.$$

Step 4. Represent $X \in m\mathcal{F}$ as $X = X_+ - X_-$, where $X_+ = \max(X, 0) \in m\mathcal{F}_+$ and $X_- = -\min(X, 0) \in m\mathcal{F}_+$, with the corresponding definition

$$\mathbf{E}X = \mathbf{E}X_+ - \mathbf{E}X_- ,$$

provided either $\mathbf{E}X_+ < \infty$ or $\mathbf{E}X_- < \infty$.

Remark. Note that we may have $\mathbf{E}X = \infty$ while $X(\omega) < \infty$ for all ω. For instance, take the random variable $X(\omega) = \omega$ for $\Omega = \{1, 2, \ldots\}$ and $\mathcal{F} = 2^\Omega$. If $\mathbf{P}(\omega = k) = ck^{-2}$ with $c = [\sum_{k=1}^\infty k^{-2}]^{-1}$ a positive, finite normalization constant, then $\mathbf{E}X = c\sum_{k=1}^\infty k^{-1} = \infty$.

Similar to the notation of μ-integrable functions introduced in the last step of the definition of Lebesgue's integral, we have the following definition for random variables.

Definition 1.3.2. *We say that a random variable X is (absolutely)* integrable*, or X has finite expectation, if $\mathbf{E}|X| < \infty$, that is, both $\mathbf{E}X_+ < \infty$ and $\mathbf{E}X_- < \infty$. Fixing $1 \leq q < \infty$ we denote by $L^q(\Omega, \mathcal{F}, \mathbf{P})$ the collection of random variables X on $(\Omega, \mathcal{F})$ for which $||X||_q = [\mathbf{E}|X|^q]^{1/q} < \infty$. For example, $L^1(\Omega, \mathcal{F}, \mathbf{P})$ denotes the space of all (absolutely) integrable random-variables. We use the short notation L^q when the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ is clear from the context.*

We next verify that Lebesgue's integral of each function f is assigned a unique value in Definition 1.3.1. To this end, we focus on $\mu_0 : \mathrm{SF}_+ \mapsto [0, \infty]$ of Step 2 of our definition and derive its structural properties, such as monotonicity, linearity and invariance to a change of argument on a μ-negligible set.

Lemma 1.3.3. *$\mu_0(\varphi)$ assigns a unique value to each $\varphi \in \mathrm{SF}_+$. Further,*
a). $\mu_0(\varphi) = \mu_0(\psi)$ if $\varphi, \psi \in \mathrm{SF}_+$ are such that $\mu(\{s : \varphi(s) \neq \psi(s)\}) = 0$.
b). μ_0 is linear, that is

$$\mu_0(\varphi + \psi) = \mu_0(\varphi) + \mu_0(\psi) , \qquad \mu_0(c\varphi) = c\mu_0(\varphi) ,$$

for any $\varphi, \psi \in \mathrm{SF}_+$ and $c \geq 0$.
c). μ_0 is monotone, that is $\mu_0(\varphi) \leq \mu_0(\psi)$ if $\varphi(s) \leq \psi(s)$ for all $s \in \mathbb{S}$.

Proof. Note that a non-negative simple function $\varphi \in \mathrm{SF}_+$ has many different representations as weighted sums of indicator functions. Suppose for example that

$$\sum_{l=1}^n c_l I_{A_l}(s) = \sum_{k=1}^m d_k I_{B_k}(s) , \tag{1.3.1}$$

for some $c_l \geq 0$, $d_k \geq 0$, $A_l \in \mathcal{F}$, $B_k \in \mathcal{F}$ and all $s \in \mathbb{S}$. There exists a finite partition of $\mathbb{S}$ to at most 2^{n+m} disjoint sets C_i such that each of the sets A_l and B_k is a union of some C_i, $i = 1, \ldots, 2^{n+m}$. Expressing both sides of (1.3.1) as finite weighted sums of I_{C_i}, we necessarily have for each i the same weight on both sides. Due to the (finite) additivity of μ over unions of disjoint sets C_i, we thus get after some algebra that

$$\sum_{l=1}^n c_l \mu(A_l) = \sum_{k=1}^m d_k \mu(B_k) . \tag{1.3.2}$$

Consequently, $\mu_0(\varphi)$ is well-defined and independent of the chosen representation for φ. Further, the conclusion (1.3.2) applies also when the two sides of (1.3.1) differ for $s \in C$ as long as $\mu(C) = 0$, hence proving the first stated property of the lemma.

Choosing the representation of $\varphi + \psi$ based on the representations of φ and ψ immediately results with the stated linearity of μ_0. Given this, if $\varphi(s) \leq \psi(s)$ for all s, then $\psi = \varphi + \xi$ for some $\xi \in \mathrm{SF}_+$, implying that $\mu_0(\psi) = \mu_0(\varphi) + \mu_0(\xi) \geq \mu_0(\varphi)$, as claimed. $\square$

Remark. The stated monotonicity of μ_0 implies that $\mu(\cdot)$ coincides with $\mu_0(\cdot)$ on SF_+. As μ_0 is uniquely defined for each $f \in \mathrm{SF}_+$ and $f = f_+$ when $f \in m\mathcal{F}_+$, it follows that $\mu(f)$ is uniquely defined for each $f \in m\mathcal{F}_+ \cup L^1(\mathbb{S}, \mathcal{F}, \mu)$.

All three properties of μ_0 (hence μ) stated in Lemma 1.3.3 for functions in SF_+ extend to all of $m\mathcal{F}_+ \cup L^1$. Indeed, the facts that $\mu(cf) = c\mu(f)$, that $\mu(f) \leq \mu(g)$ whenever $0 \leq f \leq g$, and that $\mu(f) = \mu(g)$ whenever $\mu(\{s : f(s) \neq g(s)\}) = 0$ are immediate consequences of our definition (once we have these for $f, g \in \mathrm{SF}_+$). Since $f \leq g$ implies $f_+ \leq g_+$ and $f_- \geq g_-$, the monotonicity of $\mu(\cdot)$ extends to functions in L^1 (by Step 4 of our definition). To prove that $\mu(h + g) = \mu(h) + \mu(g)$ for all $h, g \in m\mathcal{F}_+ \cup L^1$ requires an application of the *monotone convergence theorem* (in short MON), which we now state, while deferring its proof to Subsection 1.3.3.

Theorem 1.3.4 (Monotone convergence theorem)**.** *If $0 \leq h_n(s) \uparrow h(s)$ for all $s \in \mathbb{S}$ and $h_n \in m\mathcal{F}_+$, then $\mu(h_n) \uparrow \mu(h) \leq \infty$.*

Indeed, recall that while proving Proposition 1.2.6 we constructed the sequence f_n such that for every $g \in m\mathcal{F}_+$ we have $f_n(g) \in \mathrm{SF}_+$ and $f_n(g) \uparrow g$. Specifying $g, h \in m\mathcal{F}_+$ we have that $f_n(h) + f_n(g) \in \mathrm{SF}_+$. So, by Lemma 1.3.3,

$$\mu(f_n(h)+f_n(g)) = \mu_0(f_n(h)+f_n(g)) = \mu_0(f_n(h))+\mu_0(f_n(g)) = \mu(f_n(h))+\mu(f_n(g)) \,.$$

Since $f_n(h) \uparrow h$ and $f_n(h) + f_n(g) \uparrow h + g$, by monotone convergence,

$$\mu(h + g) = \lim_{n\to\infty} \mu(f_n(h) + f_n(g)) = \lim_{n\to\infty} \mu(f_n(h)) + \lim_{n\to\infty} \mu(f_n(g)) = \mu(h) + \mu(g) \,.$$

To extend this result to $g, h \in m\mathcal{F}_+ \cup L^1$, note that $h_- + g_- = f + (h+g)_- \geq f$ for some $f \in m\mathcal{F}_+$ such that $h_+ + g_+ = f + (h+g)_+$. Since $\mu(h_-) < \infty$ and $\mu(g_-) < \infty$, by linearity and monotonicity of $\mu(\cdot)$ on $m\mathcal{F}_+$ necessarily also $\mu(f) < \infty$ and the linearity of $\mu(h+g)$ on $m\mathcal{F}_+ \cup L^1$ follows by elementary algebra. In conclusion, we have just proved that

Proposition 1.3.5. *The integral $\mu(f)$ assigns a unique value to each $f \in m\mathcal{F}_+ \cup L^1(\mathbb{S}, \mathcal{F}, \mu)$. Further,*
a). $\mu(f) = \mu(g)$ whenever $\mu(\{s : f(s) \neq g(s)\}) = 0$.
b). μ is linear, that is for any $f, h, g \in m\mathcal{F}_+ \cup L^1$ and $c \geq 0$,

$$\mu(h + g) = \mu(h) + \mu(g) \,, \qquad \mu(cf) = c\mu(f) \,.$$

c). μ is monotone, that is $\mu(f) \leq \mu(g)$ if $f(s) \leq g(s)$ for all $s \in \mathbb{S}$.

Our proof of the identity $\mu(h + g) = \mu(h) + \mu(g)$ is an example of the following general approach to proving that certain properties hold for all $h \in L^1$.

Definition 1.3.6 (Standard Machine)**.** *To prove the validity of a certain property for all $h \in L^1(\mathbb{S}, \mathcal{F}, \mu)$, break your proof to four easier steps, following those of Definition 1.3.1.*

Step 1. *Prove the property for h which is an indicator function.*
Step 2. *Using linearity, extend the property to all* SF_+.
Step 3. *Using MON extend the property to all* $h \in m\mathcal{F}_+$.
Step 4. *Extend the property in question to* $h \in L^1$ *by writing* $h = h_+ - h_-$ *and using linearity.*

Here is another application of the standard machine.

Exercise 1.3.7. *Suppose that a probability measure* $\mathcal{P}$ *on* $(\mathbb{R}, \mathcal{B})$ *is such that* $\mathcal{P}(B) = \lambda(fI_B)$ *for the Lebesgue measure* λ *on* $\mathbb{R}$, *some non-negative Borel function* $f(\cdot)$ *and all* $B \in \mathcal{B}$. *Using the standard machine, prove that then* $\mathcal{P}(h) = \lambda(fh)$ *for any Borel function* h *such that either* $h \geq 0$ *or* $\lambda(f|h|) < \infty$.
Hint: *See the proof of Proposition 1.3.56.*

We shall see more applications of the standard machine later (for example, when proving Proposition 1.3.56 and Theorem 1.3.61).

We next strengthen the non-negativity and monotonicity properties of Lebesgue's integral $\mu(\cdot)$ by showing that

Lemma 1.3.8. *If* $\mu(h) = 0$ *for* $h \in m\mathcal{F}_+$, *then* $\mu(\{s : h(s) > 0\}) = 0$. *Consequently, if for* $f, g \in L^1(\mathbb{S}, \mathcal{F}, \mu)$ *both* $\mu(f) = \mu(g)$ *and* $\mu(\{s : f(s) > g(s)\}) = 0$, *then* $\mu(\{s : f(s) \neq g(s)\}) = 0$.

Proof. By continuity below of the measure μ we have that

$$\mu(\{s : h(s) > 0\}) = \lim_{n\to\infty} \mu(\{s : h(s) > n^{-1}\})$$

(see Exercise 1.1.4). Hence, if $\mu(\{s : h(s) > 0\}) > 0$, then for some $n < \infty$,

$$0 < n^{-1}\mu(\{s : h(s) > n^{-1}\}) = \mu_0(n^{-1}I_{h>n^{-1}}) \leq \mu(h)\,,$$

where the right most inequality is a consequence of the definition of $\mu(h)$ and the fact that $h \geq n^{-1}I_{h>n^{-1}} \in \mathrm{SF}_+$. Thus, our assumption that $\mu(h) = 0$ must imply that $\mu(\{s : h(s) > 0\}) = 0$.

To prove the second part of the lemma, consider $\widetilde{h} = g - f$ which is non-negative outside a set $N \in \mathcal{F}$ such that $\mu(N) = 0$. Hence, $h = (g - f)I_{N^c} \in m\mathcal{F}_+$ and $0 = \mu(g) - \mu(f) = \mu(\widetilde{h}) = \mu(h)$ by Proposition 1.3.5, implying that $\mu(\{s : h(s) > 0\}) = 0$ by the preceding proof. The same applies for $\widetilde{h}$ and the statement of the lemma follows. □

We conclude this subsection by stating the results of Proposition 1.3.5 and Lemma 1.3.8 in terms of the expectation on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$.

Theorem 1.3.9. *The mathematical expectation* $\mathbf{E}[X]$ *is well defined for every R.V.* X *on* $(\Omega, \mathcal{F}, \mathbf{P})$ *provided either* $X \geq 0$ *almost surely, or* $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$. *Further,*
(a) $\mathbf{E}X = \mathbf{E}Y$ *whenever* $X \overset{a.s.}{=} Y$.
(b) The expectation is a linear *operation, for if* Y *and* Z *are integrable R.V. then for any constants* α, β *the R.V.* $\alpha Y + \beta Z$ *is integrable and* $\mathbf{E}(\alpha Y + \beta Z) = \alpha(\mathbf{E}Y) + \beta(\mathbf{E}Z)$. *The same applies when* $Y, Z \geq 0$ *almost surely and* $\alpha, \beta \geq 0$.
(c) The expectation is monotone. *That is, if* Y *and* Z *are either integrable or non-negative and* $Y \geq Z$ *almost surely, then* $\mathbf{E}Y \geq \mathbf{E}Z$. *Further, if* Y *and* Z *are integrable with* $Y \geq Z$ *a.s. and* $\mathbf{E}Y = \mathbf{E}Z$, *then* $Y \overset{a.s.}{=} Z$.
(d) Constants are invariant under the expectation. That is, if $X \overset{a.s.}{=} c$ *for non-random* $c \in (-\infty, \infty]$, *then* $\mathbf{E}X = c$.

Remark. Part (d) of the theorem relies on the fact that $\mathbf{P}$ is a probability measure, namely $\mathbf{P}(\Omega) = 1$. Indeed, it is obtained by considering the expectation of the simple function cI_Ω to which X equals with probability one.

The linearity of the expectation (i.e. part (b) of the preceding theorem), is often extremely helpful when looking for an explicit formula for it. We next provide a few examples of this.

Exercise 1.3.10. *Write $(\Omega, \mathcal{F}, \mathbf{P})$ for a random experiment whose outcome is a recording of the results of n independent rolls of a balanced six-sided dice (including their order). Compute the expectation of the random variable $D(\omega)$ which counts the number of different faces of the dice recorded in these n rolls.*

Exercise 1.3.11 (Matching)**.** *In a random* matching *experiment, we apply a random permutation π to the integers $\{1, 2, \ldots, n\}$, where each of the possible $n!$ permutations is equally likely. Let $Z_i = I_{\{\pi(i)=i\}}$ be the random variable indicating whether $i = 1, 2, \ldots, n$ is a fixed point of the random permutation, and $X_n = \sum_{i=1}^n Z_i$ count the number of fixed points of the random permutation (i.e. the number of self-matchings). Show that $\mathbf{E}[X_n(X_n - 1)\cdots(X_n - k + 1)] = 1$ for $k = 1, 2, \ldots, n$.*

Similarly, here is an elementary application of the monotonicity of the expectation (i.e. part (c) of the preceding theorem).

Exercise 1.3.12. *Suppose an integrable random variable X is such that $\mathbf{E}(XI_A) = 0$ for each $A \in \sigma(X)$. Show that necessarily $X = 0$ almost surely.*

1.3.2. Inequalities.

The linearity of the expectation often allows us to compute $\mathbf{E}X$ even when we cannot compute the distribution function F_X. In such cases the expectation can be used to bound tail probabilities, based on the following classical inequality.

Theorem 1.3.13 (Markov's inequality)**.** *Suppose $\psi : \mathbb{R} \mapsto [0, \infty]$ is a Borel function and let $\psi_*(A) = \inf\{\psi(y) : y \in A\}$ for any $A \in \mathcal{B}$. Then for any R.V. X,*

$$\psi_*(A)\mathbf{P}(X \in A) \leq \mathbf{E}(\psi(X)I_{X\in A}) \leq \mathbf{E}\psi(X).$$

Proof. By the definition of $\psi_*(A)$ and non-negativity of ψ we have that

$$\psi_*(A)I_{x\in A} \leq \psi(x)I_{x\in A} \leq \psi(x)\,,$$

for all $x \in \mathbb{R}$. Therefore, $\psi_*(A)I_{X\in A} \leq \psi(X)I_{X\in A} \leq \psi(X)$ for every $\omega \in \Omega$. We deduce the stated inequality by the monotonicity of the expectation and the identity $\mathbf{E}(\psi_*(A)I_{X\in A}) = \psi_*(A)\mathbf{P}(X \in A)$ (due to Step 2 of Definition 1.3.1). $\square$

We next specify three common instances of Markov's inequality.

Example 1.3.14. *(a). Taking $\psi(x) = x_+$ and $A = [a, \infty)$ for some $a > 0$ we have that $\psi_*(A) = a$. Markov's inequality is then*

$$\mathbf{P}(X \geq a) \leq \frac{\mathbf{E}X_+}{a}\,,$$

which is particularly appealing when $X \geq 0$, so $\mathbf{E}X_+ = \mathbf{E}X$.
(b). Taking $\psi(x) = |x|^q$ and $A = (-\infty, -a] \cup [a, \infty)$ for some $a > 0$, we get that $\psi_*(A) = a^q$. Markov's inequality is then $a^q\mathbf{P}(|X| \geq a) \leq \mathbf{E}|X|^q$. Considering $q = 2$

and $X = Y - \mathbf{E}Y$ *for* $Y \in L^2$*, this amounts to*

$$\mathbf{P}(|Y - \mathbf{E}Y| \geq a) \leq \frac{\mathsf{Var}(Y)}{a^2}\,,$$

which we call Chebyshev's inequality *(c.f. Definition 1.3.67 for the variance and moments of random variable* Y*).*
(c). Taking $\psi(x) = e^{\theta x}$ *for some* $\theta > 0$ *and* $A = [a, \infty)$ *for some* $a \in \mathbb{R}$ *we have that* $\psi_*(A) = e^{\theta a}$*. Markov's inequality is then*

$$\mathbf{P}(X \geq a) \leq e^{-\theta a}\mathbf{E}e^{\theta X}.$$

This bound provides an exponential decay in a*, at the cost of requiring* X *to have finite exponential moments.*

In general, we cannot compute $\mathbf{E}X$ explicitly from the Definition 1.3.1 except for discrete R.V.s and for R.V.s having a probability density function. We thus appeal to the properties of the expectation listed in Theorem 1.3.9, or use various inequalities to bound one expectation by another. To this end, we start with Jensen's inequality, dealing with the effect that a convex function makes on the expectation.

Proposition 1.3.15 (Jensen's inequality)**.** *Suppose* $g(\cdot)$ *is a convex function on an open interval* G *of* $\mathbb{R}$*, that is,*

$$\lambda g(x) + (1-\lambda)g(y) \geq g(\lambda x + (1-\lambda)y) \qquad \forall\ x, y \in G, \qquad 0 \leq \lambda \leq 1.$$

If X *is an integrable R.V. with* $\mathbf{P}(X \in G) = 1$ *and* $g(X)$ *is also integrable, then* $\mathbf{E}(g(X)) \geq g(\mathbf{E}X)$*.*

Proof. The convexity of $g(\cdot)$ on G implies that $g(\cdot)$ is continuous on G (hence $g(X)$ is a random variable) and the existence for each $c \in G$ of $b = b(c) \in \mathbb{R}$ such that

$$g(x) \geq g(c) + b(x - c), \qquad \forall x \in G\,. \tag{1.3.3}$$

Since G is an open interval of $\mathbb{R}$ with $\mathbf{P}(X \in G) = 1$ and X is integrable, it follows that $\mathbf{E}X \in G$. Assuming (1.3.3) holds for $c = \mathbf{E}X$, that $X \in G$ a.s., and that both X and $g(X)$ are integrable, we have by Theorem 1.3.9 that

$$\mathbf{E}(g(X)) = \mathbf{E}(g(X)I_{X\in G}) \geq \mathbf{E}[(g(c)+b(X-c))I_{X\in G}] = g(c)+b(\mathbf{E}X-c) = g(\mathbf{E}X)\,,$$

as stated. To derive (1.3.3) note that if $(c - h_2, c + h_1) \subseteq G$ for positive h_1 and h_2, then by convexity of $g(\cdot)$,

$$\frac{h_2}{h_1 + h_2} g(c + h_1) + \frac{h_1}{h_1 + h_2} g(c - h_2) \geq g(c)\,,$$

which amounts to $[g(c + h_1) - g(c)]/h_1 \geq [g(c) - g(c - h_2)]/h_2$. Considering the infimum over $h_1 > 0$ and the supremum over $h_2 > 0$ we deduce that

$$\inf_{h>0, c+h\in G} \frac{g(c + h) - g(c)}{h} := (D_+g)(c) \geq (D_-g)(c) := \sup_{h>0, c-h\in G} \frac{g(c) - g(c - h)}{h}\,.$$

With G an open set, obviously $(D_-g)(x) > -\infty$ and $(D_+g)(x) < \infty$ for any $x \in G$ (in particular, $g(\cdot)$ is continuous on G). Now for any $b \in [(D_-g)(c), (D_+g)(c)] \subset \mathbb{R}$ we get (1.3.3) out of the definition of D_+g and D_-g. $\square$

Remark. Since $g(\cdot)$ is convex if and only if $-g(\cdot)$ is concave, we may as well state Jensen's inequality for concave functions, just reversing the sign of the inequality in this case. A trivial instance of Jensen's inequality happens when $X(\omega) = xI_A(\omega) + yI_{A^c}(\omega)$ for some $x, y \in \mathbb{R}$ and $A \in \mathcal{F}$ such that $\mathbf{P}(A) = \lambda$. Then,

$$\mathbf{E}X = x\mathbf{P}(A) + y\mathbf{P}(A^c) = x\lambda + y(1-\lambda)\,,$$

whereas $g(X(\omega)) = g(x)I_A(\omega) + g(y)I_{A^c}(\omega)$. So,

$$\mathbf{E}g(X) = g(x)\lambda + g(y)(1-\lambda) \geq g(x\lambda + y(1-\lambda)) = g(\mathbf{E}X)\,,$$

as g is convex.

Applying Jensen's inequality, we show that the spaces $L^q(\Omega, \mathcal{F}, \mathbf{P})$ of Definition 1.3.2 are nested in terms of the parameter $q \geq 1$.

Lemma 1.3.16. *Fixing $Y \in m\mathcal{F}$, the mapping $q \mapsto ||Y||_q = [\mathbf{E}|Y|^q]^{1/q}$ is non-decreasing for $q > 0$. Hence, the space $L^q(\Omega, \mathcal{F}, \mathbf{P})$ is contained in $L^r(\Omega, \mathcal{F}, \mathbf{P})$ for any $r \leq q$.*

Proof. Fix $q > r > 0$ and consider the sequence of bounded R.V. $X_n(\omega) = \{\min(|Y(\omega)|, n)\}^r$. Obviously, X_n and $X_n^{q/r}$ are both in L^1. Apply Jensen's Inequality for the convex function $g(x) = |x|^{q/r}$ and the non-negative R.V. X_n, to get that

$$(\mathbf{E}X_n)^{\frac{q}{r}} \leq \mathbf{E}(X_n^{\frac{q}{r}}) = \mathbf{E}[\{\min(|Y|, n)\}^q] \leq \mathbf{E}(|Y|^q)\,.$$

For $n \uparrow \infty$ we have that $X_n \uparrow |Y|^r$, so by monotone convergence $\mathbf{E}\left(|Y|^r\right)^{\frac{q}{r}} \leq (\mathbf{E}|Y|^q)$. Taking the $1/q$-th power yields the stated result $||Y||_r \leq ||Y||_q \leq \infty$. $\square$

We next bound the expectation of the product of two R.V. while assuming nothing about the relation between them.

Proposition 1.3.17 (Hölder's inequality)**.** *Let X, Y be two random variables on the same probability space. If $p, q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$, then*

$$\mathbf{E}|XY| \leq ||X||_p ||Y||_q\,. \tag{1.3.4}$$

Remark. Recall that if XY is integrable then $\mathbf{E}|XY|$ is by itself an upper bound on $|[\mathbf{E}XY]|$. The special case of $p = q = 2$ in Hölder's inequality

$$\mathbf{E}|XY| \leq \sqrt{\mathbf{E}X^2}\sqrt{\mathbf{E}Y^2}\,,$$

is called the *Cauchy-Schwarz inequality.*

Proof. Fixing $p > 1$ and $q = p/(p-1)$ let $\lambda = ||X||_p$ and $\xi = ||Y||_q$. If $\lambda = 0$ then $|X|^p \overset{a.s.}{=} 0$ (see Theorem 1.3.9). Likewise, if $\xi = 0$ then $|Y|^q \overset{a.s.}{=} 0$. In either case, the inequality (1.3.4) trivially holds. As this inequality also trivially holds when either $\lambda = \infty$ or $\xi = \infty$, we may and shall assume hereafter that both λ and ξ are finite and strictly positive. Recall that

$$\frac{x^p}{p} + \frac{y^q}{q} - xy \geq 0, \qquad \forall x, y \geq 0$$

(c.f. [**Dur10**, Page 21] where it is proved by considering the first two derivatives in x). Taking $x = |X|/\lambda$ and $y = |Y|/\xi$, we have by linearity and monotonicity of the expectation that

$$1 = \frac{1}{p} + \frac{1}{q} = \frac{\mathbf{E}|X|^p}{\lambda^p p} + \frac{\mathbf{E}|Y|^q}{\xi^q q} \geq \frac{\mathbf{E}|XY|}{\lambda\xi}\,,$$

yielding the stated inequality (1.3.4). □

A direct consequence of Hölder's inequality is the triangle inequality for the norm $||X||_p$ in $L^p(\Omega, \mathcal{F}, \mathbf{P})$, that is,

Proposition 1.3.18 (Minkowski's inequality)**.** *If $X, Y \in L^p(\Omega, \mathcal{F}, \mathbf{P}), p \geq 1$, then $||X+Y||_p \leq ||X||_p + ||Y||_p$.*

Proof. With $|X+Y| \leq |X|+|Y|$, by monotonicity of the expectation we have the stated inequality in case $p=1$. Considering hereafter $p>1$, it follows from Hölder's inequality (Proposition 1.3.17) that

$$\begin{aligned}
\mathbf{E}|X+Y|^p &= \mathbf{E}(|X+Y||X+Y|^{p-1}) \\
&\leq \mathbf{E}(|X||X+Y|^{p-1}) + \mathbf{E}(|Y||X+Y|^{p-1}) \\
&\leq (\mathbf{E}|X|^p)^{\frac{1}{p}} (\mathbf{E}|X+Y|^{(p-1)q})^{\frac{1}{q}} + \ (\mathbf{E}|Y|^p)^{\frac{1}{p}} (\mathbf{E}|X+Y|^{(p-1)q})^{\frac{1}{q}} \\
&= (||X||_p + ||Y||_p)\, (\mathbf{E}|X+Y|^p)^{\frac{1}{q}}
\end{aligned}$$

(recall that $(p-1)q = p$). Since $X, Y \in L^p$ and

$$|x+y|^p \leq (|x|+|y|)^p \leq 2^{p-1}(|x|^p + |y|^p), \qquad \forall x, y \in \mathbb{R}, \ \ p>1,$$

if follows that $a_p = \mathbf{E}|X+Y|^p < \infty$. There is nothing to prove unless $a_p > 0$, in which case dividing by $(a_p)^{1/q}$ we get that

$$(\mathbf{E}|X+Y|^p)^{1-\frac{1}{q}} \leq ||X||_p + ||Y||_p\,,$$

giving the stated inequality (since $1 - \frac{1}{q} = \frac{1}{p}$). □

Remark. Jensen's inequality applies only for probability measures, while both Hölder's inequality $\mu(|fg|) \leq \mu(|f|^p)^{1/p}\mu(|g|^q)^{1/q}$ and Minkowski's inequality apply for any measure μ, with exactly the same proof we provided for probability measures.

To practice your understanding of Markov's inequality, solve the following exercise.

Exercise 1.3.19. *Let X be a non-negative random variable with $\mathsf{Var}(X) \leq 1/2$. Show that then $\mathbf{P}(-1+\mathbf{E}X \leq X \leq 2\mathbf{E}X) \geq 1/2$.*

To practice your understanding of the proof of Jensen's inequality, try to prove its extension to convex functions on $\mathbb{R}^n$.

Exercise 1.3.20. *Suppose $g : \mathbb{R}^n \to \mathbb{R}$ is a convex function and $X_1, X_2, \ldots, X_n$ are integrable random variables, defined on the same probability space and such that $g(X_1, \ldots, X_n)$ is integrable. Show that then $\mathbf{E}g(X_1, \ldots, X_n) \geq g(\mathbf{E}X_1, \ldots, \mathbf{E}X_n)$.*
Hint: *Use convex analysis to show that $g(\cdot)$ is continuous and further that for any $\underline{c} \in \mathbb{R}^n$ there exists $\underline{b} \in \mathbb{R}^n$ such that $g(\underline{x}) \geq g(\underline{c}) + \langle \underline{b}, \underline{x} - \underline{c} \rangle$ for all $\underline{x} \in \mathbb{R}^n$ (with $\langle \cdot, \cdot \rangle$ denoting the inner product of two vectors in $\mathbb{R}^n$).*

Exercise 1.3.21. *Let $Y \geq 0$ with $v = \mathbf{E}(Y^2) < \infty$.*

(a) *Show that for any $0 \leq a < \mathbf{E}Y$,*

$$\mathbf{P}(Y > a) \geq \frac{(\mathbf{E}Y - a)^2}{\mathbf{E}(Y^2)}$$

Hint: *Apply the Cauchy-Schwarz inequality to $YI_{Y>a}$.*

(b) *Show that $(\mathbf{E}|Y^2 - v|)^2 \leq 4v(v - (\mathbf{E}Y)^2)$.*

(c) *Derive the second* Bonferroni inequality*,*

$$\mathbf{P}(\bigcup_{i=1}^{n} A_i) \geq \sum_{i=1}^{n} \mathbf{P}(A_i) - \sum_{1\leq j<i\leq n} \mathbf{P}(A_i \cap A_j)\,.$$

How does it compare with the bound of part (a) for $Y = \sum_{i=1}^{n} I_{A_i}$*?*

1.3.3. Convergence, limits and expectation. Asymptotic behavior is a key issue in probability theory. We thus explore here various notions of convergence of random variables and the relations among them, focusing on the integrability conditions needed for exchanging the order of limit and expectation operations. Unless explicitly stated otherwise, throughout this section we assume that all R.V. are defined on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$.

In Definition 1.2.25 we have encountered the convergence almost surely of R.V. A weaker notion of convergence is *convergence in probability* as defined next.

Definition 1.3.22. *We say that R.V.* X_n *converge to a given R.V.* X_∞ *in probability, denoted* $X_n \stackrel{p}{\to} X_\infty$*, if* $\mathbf{P}(\{\omega : |X_n(\omega) - X_\infty(\omega)| > \varepsilon\}) \to 0$ *as* $n \to \infty$*, for any fixed* $\varepsilon > 0$*. This is equivalent to* $|X_n - X_\infty| \stackrel{p}{\to} 0$*, and is a special case of the* convergence in μ-measure *of* $f_n \in m\mathcal{F}$ *to* $f_\infty \in m\mathcal{F}$*, that is* $\mu(\{s : |f_n(s) - f_\infty(s)| > \varepsilon\}) \to 0$ *as* $n \to \infty$*, for any fixed* $\varepsilon > 0$*.*

Our next exercise and example clarify the relationship between convergence almost surely and convergence in probability.

Exercise 1.3.23. *Verify that convergence almost surely* to a finite limit *implies convergence in probability, that is if* $X_n \stackrel{a.s.}{\to} X_\infty \in \mathbb{R}$ *then* $X_n \stackrel{p}{\to} X_\infty$*.*

Remark 1.3.24. Generalizing Definition 1.3.22, for a separable metric space $(\mathbb{S}, \rho)$ we say that $(\mathbb{S}, \mathcal{B}_{\mathbb{S}})$-valued random variables X_n converge to X_∞ in probability if and only if for every $\varepsilon > 0$, $\mathbf{P}(\rho(X_n, X_\infty) > \varepsilon) \to 0$ as $n \to \infty$ (see [**Dud89**, Section 9.2] for more details). Equipping $\mathbb{S} = \overline{\mathbb{R}}$ with a suitable metric (for example, $\rho(x, y) = |\varphi(x) - \varphi(y)|$ with $\varphi(x) = x/(1 + |x|) : \overline{\mathbb{R}} \mapsto [-1, 1]$), this definition removes the restriction to X_∞ finite in Exercise 1.3.23.

In general, $X_n \stackrel{p}{\to} X_\infty$ does not imply that $X_n \stackrel{a.s.}{\to} X_\infty$.

Example 1.3.25. *Consider the probability space* $((0, 1], \mathcal{B}_{(0,1]}, U)$ *and* $X_n(\omega) = \mathbf{1}_{[t_n,t_n+s_n]}(\omega)$ *with* $s_n \downarrow 0$ *as* $n \to \infty$ *slowly enough and* $t_n \in [0, 1 - s_n]$ *are such that any* $\omega \in (0, 1]$ *is in infinitely many intervals* $[t_n, t_n + s_n]$*. The latter property applies if* $t_n = (i - 1)/k$ *and* $s_n = 1/k$ *when* $n = k(k - 1)/2 + i$*,* $i = 1, 2, \ldots, k$ *and* $k = 1, 2, \ldots$ *(plot the intervals* $[t_n, t_n + s_n]$ *to convince yourself). Then,* $X_n \stackrel{p}{\to} 0$ *(since* $s_n = U(X_n \neq 0) \to 0$*), whereas fixing each* $\omega \in (0, 1]$*, we have that* $X_n(\omega) = 1$ *for infinitely many values of* n*, hence* X_n *does not converge a.s. to zero.*

Associated with each space $L^q(\Omega, \mathcal{F}, \mathbf{P})$ is the notion of L^q convergence which we now define.

Definition 1.3.26. *We say that* X_n converges in L^q *to* X_∞*, denoted* $X_n \stackrel{L^q}{\to} X_\infty$*, if* $X_n, X_\infty \in L^q$ *and* $||X_n - X_\infty||_q \to 0$ *as* $n \to \infty$ *(i.e.,* $\mathbf{E}\left(|X_n - X_\infty|^q\right) \to 0$ *as* $n \to \infty$*.*

Remark. For $q = 2$ we have the explicit formula

$$||X_n - X||_2^2 = \mathbf{E}(X_n^2) - 2\mathbf{E}(X_n X) + \mathbf{E}(X^2).$$

Thus, it is often easiest to check convergence in L^2.

The following immediate corollary of Lemma 1.3.16 provides an ordering of L^q convergence in terms of the parameter q.

Corollary 1.3.27. *If $X_n \stackrel{L^q}{\to} X_\infty$ and $q \geq r$, then $X_n \stackrel{L^r}{\to} X_\infty$.*

Next note that the L^q convergence implies the convergence of the expectation of $|X_n|^q$.

Exercise 1.3.28. *Fixing $q \geq 1$, use Minkowski's inequality (Proposition 1.3.18), to show that if $X_n \stackrel{L^q}{\to} X_\infty$, then $\mathbf{E}|X_n|^q \to \mathbf{E}|X_\infty|^q$ and for $q = 1, 2, 3, \ldots$ also $\mathbf{E}X_n^q \to \mathbf{E}X_\infty^q$.*

Further, it follows from Markov's inequality that the convergence in L^q implies convergence in probability (for any value of q).

Proposition 1.3.29. *If $X_n \stackrel{L^q}{\to} X_\infty$, then $X_n \stackrel{p}{\to} X_\infty$.*

Proof. Fixing $q > 0$ recall that Markov's inequality results with

$$\mathbf{P}(|Y| > \varepsilon) \leq \varepsilon^{-q}\mathbf{E}[|Y|^q],$$

for any R.V. Y and any $\varepsilon > 0$ (c.f part (b) of Example 1.3.14). The assumed convergence in L^q means that $\mathbf{E}[|X_n - X_\infty|^q] \to 0$ as $n \to \infty$, so taking $Y = Y_n = X_n - X_\infty$, we necessarily have also $\mathbf{P}(|X_n - X_\infty| > \varepsilon) \to 0$ as $n \to \infty$. Since $\varepsilon > 0$ is arbitrary, we see that $X_n \stackrel{p}{\to} X_\infty$ as claimed. $\square$

The converse of Proposition 1.3.29 does not hold in general. As we next demonstrate, even the stronger almost surely convergence (see Exercise 1.3.23), and having a non-random constant limit are not enough to guarantee the L^q convergence, for any $q > 0$.

Example 1.3.30. *Fixing $q > 0$, consider the probability space $((0,1], \mathcal{B}_{(0,1]}, U)$ and the R.V. $Y_n(\omega) = n^{1/q} I_{[0,n^{-1}]}(\omega)$. Since $Y_n(\omega) = 0$ for all $n \geq n_0$ and some finite $n_0 = n_0(\omega)$, it follows that $Y_n(\omega) \stackrel{a.s.}{\to} 0$ as $n \to \infty$. However, $\mathbf{E}[|Y_n|^q] = nU([0, n^{-1}]) = 1$ for all n, so Y_n does not converge to zero in L^q (see Exercise 1.3.28).*

Considering Example 1.3.25, where $X_n \stackrel{L^q}{\to} 0$ while X_n does not converge a.s. to zero, and Example 1.3.30 which exhibits the converse phenomenon, we conclude that the convergence in L^q and the a.s. convergence are in general non comparable, and neither one is a consequence of convergence in probability.

Nevertheless, a sequence X_n can have at most one limit, regardless of which convergence mode is considered.

Exercise 1.3.31. *Check that if $X_n \stackrel{L^q}{\to} X$ and $X_n \stackrel{a.s.}{\to} Y$ then $X \stackrel{a.s.}{=} Y$.*

Though we have just seen that in general the order of the limit and expectation operations is non-interchangeable, we examine for the remainder of this subsection various conditions which do allow for such an interchange. Note in passing that upon proving any such result under certain point-wise convergence conditions, we may with no extra effort relax these to the corresponding almost sure convergence (and the same applies for integrals with respect to measures, see part (a) of Theorem 1.3.9, or that of Proposition 1.3.5).

Turning to do just that, we first outline the results which apply in the more general measure theory setting, starting with the proof of the *monotone convergence theorem.*

PROOF OF THEOREM 1.3.4. By part (c) of Proposition 1.3.5, the proof of which did not use Theorem 1.3.4, we know that $\mu(h_n)$ is a non-decreasing sequence that is bounded above by $\mu(h)$. It therefore suffices to show that

$$\begin{aligned}\lim_{n\to\infty}\mu(h_n) &= \sup_n\{\mu_0(\psi) : \psi \in \mathrm{SF}_+, \psi \le h_n\}\\ &\ge \sup\{\mu_0(\varphi) : \varphi \in \mathrm{SF}_+, \varphi \le h\} = \mu(h) \end{aligned} \tag{1.3.5}$$

(see Step 3 of Definition 1.3.1). That is, it suffices to find for each non-negative simple function $\varphi \le h$ a sequence of non-negative simple functions $\psi_n \le h_n$ such that $\mu_0(\psi_n) \to \mu_0(\varphi)$ as $n \to \infty$. To this end, fixing φ, we may and shall choose without loss of generality a representation $\varphi = \sum\limits_{l=1}^{m} c_l I_{A_l}$ such that $A_l \in \mathcal{F}$ are disjoint and further $c_l\mu(A_l) > 0$ for $l = 1,\dots,m$ (see proof of Lemma 1.3.3). Using hereafter the notation $f_*(A) = \inf\{f(s) : s \in A\}$ for $f \in m\mathcal{F}_+$ and $A \in \mathcal{F}$, the condition $\varphi(s) \le h(s)$ for all $s \in \mathbb{S}$ is equivalent to $c_l \le h_*(A_l)$ for all l, so

$$\mu_0(\varphi) \le \sum_{l=1}^{m} h_*(A_l)\mu(A_l) = V\,.$$

Suppose first that $V < \infty$, that is $0 < h_*(A_l)\mu(A_l) < \infty$ for all l. In this case, fixing $\lambda < 1$, consider for each n the disjoint sets $A_{l,\lambda,n} = \{s \in A_l : h_n(s) \ge \lambda h_*(A_l)\} \in \mathcal{F}$ and the corresponding

$$\psi_{\lambda,n}(s) = \sum_{l=1}^{m} \lambda h_*(A_l) I_{A_{l,\lambda,n}}(s) \in \mathrm{SF}_+\,,$$

where $\psi_{\lambda,n}(s) \le h_n(s)$ for all $s \in \mathbb{S}$. If $s \in A_l$ then $h(s) > \lambda h_*(A_l)$. Thus, $h_n \uparrow h$ implies that $A_{l,\lambda,n} \uparrow A_l$ as $n \to \infty$, for each l. Consequently, by definition of $\mu(h_n)$ and the continuity from below of μ,

$$\lim_{n\to\infty}\mu(h_n) \ge \lim_{n\to\infty}\mu_0(\psi_{\lambda,n}) = \lambda V\,.$$

Taking $\lambda \uparrow 1$ we deduce that $\lim_n \mu(h_n) \ge V \ge \mu_0(\varphi)$. Next suppose that $V = \infty$, so without loss of generality we may and shall assume that $h_*(A_1)\mu(A_1) = \infty$. Fixing $x \in (0, h_*(A_1))$ let $A_{1,x,n} = \{s \in A_1 : h_n(s) \ge x\} \in \mathcal{F}$ noting that $A_{1,x,n} \uparrow A_1$ as $n \to \infty$ and $\psi_{x,n}(s) = xI_{A_{1,x,n}}(s) \le h_n(s)$ for all n and $s \in \mathbb{S}$, is a non-negative simple function. Thus, again by continuity from below of μ we have that

$$\lim_{n\to\infty}\mu(h_n) \ge \lim_{n\to\infty}\mu_0(\psi_{x,n}) = x\mu(A_1)\,.$$

Taking $x \uparrow h_*(A_1)$ we deduce that $\lim_n \mu(h_n) \ge h_*(A_1)\mu(A_1) = \infty$, completing the proof of (1.3.5) and that of the theorem. □

Considering probability spaces, Theorem 1.3.4 tells us that we can exchange the order of the limit and the expectation in case of monotone upward a.s. convergence of non-negative R.V. (with the limit possibly infinite). That is,

Theorem 1.3.32 (MONOTONE CONVERGENCE THEOREM)**.** *If $X_n \ge 0$ and $X_n(\omega) \uparrow X_\infty(\omega)$ for almost every ω, then $\mathbf{E}X_n \uparrow \mathbf{E}X_\infty$.*

In Example 1.3.30 we have a point-wise convergent sequence of R.V. whose expectations exceed that of their limit. In a sense this is always the case, as stated next in Fatou's lemma (which is a direct consequence of the monotone convergence theorem).

Lemma 1.3.33 (Fatou's lemma)**.** *For any measure space* $(\mathbb{S}, \mathcal{F}, \mu)$ *and any* $f_n \in m\mathcal{F}$*, if* $f_n(s) \geq g(s)$ *for some* μ*-integrable function* g*, all* n *and* μ*-almost-every* $s \in \mathbb{S}$*, then*

$$\liminf_{n\to\infty} \mu(f_n) \geq \mu(\liminf_{n\to\infty} f_n)\,. \tag{1.3.6}$$

Alternatively, if $f_n(s) \leq g(s)$ *for all* n *and a.e.* s*, then*

$$\limsup_{n\to\infty} \mu(f_n) \leq \mu(\limsup_{n\to\infty} f_n)\,. \tag{1.3.7}$$

Proof. Assume first that $f_n \in m\mathcal{F}_+$ and let $h_n(s) = \inf_{k\geq n} f_k(s)$, noting that $h_n \in m\mathcal{F}_+$ is a non-decreasing sequence, whose point-wise limit is $h(s) := \liminf_{n\to\infty} f_n(s)$. By the monotone convergence theorem, $\mu(h_n) \uparrow \mu(h)$. Since $f_n(s) \geq h_n(s)$ for all $s \in \mathbb{S}$, the monotonicity of the integral (see Proposition 1.3.5) implies that $\mu(f_n) \geq \mu(h_n)$ for all n. Considering the $\liminf$ as $n \to \infty$ we arrive at (1.3.6).

Turning to extend this inequality to the more general setting of the lemma, note that our conditions imply that $f_n \overset{a.e.}{=} g + (f_n - g)_+$ for each n. Considering the countable union of the μ-negligible sets in which one of these identities is violated, we thus have that

$$h := \liminf_{n\to\infty} f_n \overset{a.e.}{=} g + \liminf_{n\to\infty} (f_n - g)_+\,.$$

Further, $\mu(f_n) = \mu(g) + \mu((f_n - g)_+)$ by the linearity of the integral in $m\mathcal{F}_+ \cup L^1$. Taking $n \to \infty$ and applying (1.3.6) for $(f_n - g)_+ \in m\mathcal{F}_+$ we deduce that

$$\liminf_{n\to\infty} \mu(f_n) \geq \mu(g) + \mu(\liminf_{n\to\infty} (f_n - g)_+) = \mu(g) + \mu(h - g) = \mu(h)$$

(where for the right most identity we used the linearity of the integral, as well as the fact that $-g$ is μ-integrable).

Finally, we get (1.3.7) for f_n by considering (1.3.6) for $-f_n$. □

Remark. In terms of the expectation, Fatou's lemma is the statement that if R.V. $X_n \geq X$, almost surely, for some $X \in L^1$ and all n, then

$$\liminf_{n\to\infty} \mathbf{E}(X_n) \geq \mathbf{E}(\liminf_{n\to\infty} X_n)\,, \tag{1.3.8}$$

whereas if $X_n \leq X$, almost surely, for some $X \in L^1$ and all n, then

$$\limsup_{n\to\infty} \mathbf{E}(X_n) \leq \mathbf{E}(\limsup_{n\to\infty} X_n)\,. \tag{1.3.9}$$

Some text books call (1.3.9) and (1.3.7) the *Reverse Fatou Lemma* (e.g. [**Wil91**, Section 5.4]).

Using Fatou's lemma, we can easily prove Lebesgue's dominated convergence theorem (in short DOM).

Theorem 1.3.34 (Dominated convergence theorem)**.** *For any measure space* $(\mathbb{S}, \mathcal{F}, \mu)$ *and any* $f_n \in m\mathcal{F}$*, if for some* μ*-integrable function* g *and* μ*-almost-every* $s \in \mathbb{S}$ *both* $f_n(s) \to f_\infty(s)$ *as* $n \to \infty$*, and* $|f_n(s)| \leq g(s)$ *for all* n*, then* f_∞ *is* μ*-integrable and further* $\mu(|f_n - f_\infty|) \to 0$ *as* $n \to \infty$*.*

PROOF. Up to a μ-negligible subset of $\mathbb{S}$, our assumption that $|f_n| \le g$ and $f_n \to f_\infty$, implies that $|f_\infty| \le g$, hence f_∞ is μ-integrable. Applying Fatou's lemma (1.3.7) for $|f_n - f_\infty| \le 2g$ such that $\limsup_n |f_n - f_\infty| = 0$, we conclude that

$$0 \le \limsup_{n\to\infty} \mu(|f_n - f_\infty|) \le \mu(\limsup_{n\to\infty} |f_n - f_\infty|) = \mu(0) = 0 \,,$$

as claimed. □

By Minkowski's inequality, $\mu(|f_n - f_\infty|) \to 0$ implies that $\mu(|f_n|) \to \mu(|f_\infty|)$. The dominated convergence theorem provides us with a simple sufficient condition for the converse implication in case also $f_n \to f_\infty$ a.e.

Lemma 1.3.35 (SCHEFFÉ'S LEMMA). *If $f_n \in m\mathcal{F}$ converges a.e. to $f_\infty \in m\mathcal{F}$ and $\mu(|f_n|) \to \mu(|f_\infty|) < \infty$ then $\mu(|f_n - f_\infty|) \to 0$ as $n \to \infty$.*

Remark. In terms of expectation, Scheffé's lemma states that if $X_n \overset{a.s.}{\to} X_\infty$ and $\mathbf{E}|X_n| \to \mathbf{E}|X_\infty| < \infty$, then $X_n \overset{L^1}{\to} X_\infty$ as well.

PROOF. As already noted, we may assume without loss of generality that $f_n(s) \to f_\infty(s)$ for *all* $s \in \mathbb{S}$, that is $g_n(s) = f_n(s) - f_\infty(s) \to 0$ as $n \to \infty$, for all $s \in \mathbb{S}$. Further, since $\mu(|f_n|) \to \mu(|f_\infty|) < \infty$, we may and shall assume also that f_n are $\mathbb{R}$-valued and μ-integrable for all $n \le \infty$, hence $g_n \in L^1(\mathbb{S}, \mathcal{F}, \mu)$ as well.

Suppose first that $f_n \in m\mathcal{F}_+$ for all $n \le \infty$. In this case, $0 \le (g_n)_- \le f_\infty$ for all n and s. As $(g_n)_-(s) \to 0$ for every $s \in \mathbb{S}$, applying the dominated convergence theorem we deduce that $\mu((g_n)_-) \to 0$. From the assumptions of the lemma (and the linearity of the integral on L^1), we get that $\mu(g_n) = \mu(f_n) - \mu(f_\infty) \to 0$ as $n \to \infty$. Since $|x| = x + 2x_-$ for any $x \in \mathbb{R}$, it thus follows by linearity of the integral on L^1 that $\mu(|g_n|) = \mu(g_n) + 2\mu((g_n)_-) \to 0$ for $n \to \infty$, as claimed.

In the general case of $f_n \in m\mathcal{F}$, we know that both $0 \le (f_n)_+(s) \to (f_\infty)_+(s)$ and $0 \le (f_n)_-(s) \to (f_\infty)_-(s)$ for every s, so by (1.3.6) of Fatou's lemma, we have that

$$\begin{aligned}\mu(|f_\infty|) = \mu((f_\infty)_+) + \mu((f_\infty)_-) &\le \liminf_{n\to\infty} \mu((f_n)_-) + \liminf_{n\to\infty} \mu((f_n)_+) \\ &\le \liminf_{n\to\infty} [\mu((f_n)_-) + \mu((f_n)_+)] = \lim_{n\to\infty} \mu(|f_n|) = \mu(|f_\infty|) \,.\end{aligned}$$

Hence, necessarily both $\mu((f_n)_+) \to \mu((f_\infty)_+)$ and $\mu((f_n)_-) \to \mu((f_\infty)_-)$. Since $|x - y| \le |x_+ - y_+| + |x_- - y_-|$ for all $x, y \in \mathbb{R}$ and we already proved the lemma for the non-negative $(f_n)_-$ and $(f_n)_+$, we see that

$$\lim_{n\to\infty} \mu(|f_n - f_\infty|) \le \lim_{n\to\infty} \mu(|(f_n)_+ - (f_\infty)_+|) + \lim_{n\to\infty} \mu(|(f_n)_- - (f_\infty)_-|) = 0 \,,$$

concluding the proof of the lemma. □

We conclude this sub-section with quite a few exercises, starting with an alternative characterization of convergence almost surely.

Exercise 1.3.36. *Show that $X_n \overset{a.s.}{\to} 0$ if and only if for each $\varepsilon > 0$ there is n so that for each* random *integer M with $M(\omega) \ge n$ for all $\omega \in \Omega$ we have that $\mathbf{P}(\{\omega : |X_{M(\omega)}(\omega)| > \varepsilon\}) < \varepsilon$.*

Exercise 1.3.37. *Let Y_n be (real-valued) random variables on $(\Omega, \mathcal{F}, \mathbf{P})$, and N_k positive integer valued random variables on the same probability space.*

(a) *Show that $Y_{N_k}(\omega) = Y_{N_k(\omega)}(\omega)$ are random variables on $(\Omega, \mathcal{F})$.*

(b) *Show that if $Y_n \overset{a.s}{\to} Y_\infty$ and $N_k \overset{a.s.}{\to} \infty$ then $Y_{N_k} \overset{a.s.}{\to} Y_\infty$.*

(c) *Provide an example of* $Y_n \xrightarrow{p} 0$ *and* $N_k \xrightarrow{a.s.} \infty$ *such that almost surely* $Y_{N_k} = 1$ *for all* k.
(d) *Show that if* $Y_n \xrightarrow{a.s.} Y_\infty$ *and* $\mathbf{P}(N_k > r) \to 1$ *as* $k \to \infty$, *for every fixed* $r < \infty$, *then* $Y_{N_k} \xrightarrow{p} Y_\infty$.

In the following four exercises you find some of the many applications of the monotone convergence theorem.

Exercise 1.3.38. *You are now to relax the non-negativity assumption in the monotone convergence theorem.*
(a) *Show that if* $\mathbf{E}[(X_1)_-] < \infty$ *and* $X_n(\omega) \uparrow X(\omega)$ *for almost every* ω, *then* $\mathbf{E}X_n \uparrow \mathbf{E}X$.
(b) *Show that if in addition* $\sup_n \mathbf{E}[(X_n)_+] < \infty$, *then* $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$.

Exercise 1.3.39. *In this exercise you are to show that for any R.V.* $X \geq 0$,

$$\mathbf{E}X = \lim_{\delta \downarrow 0} \mathbf{E}_\delta X \quad \text{for} \quad \mathbf{E}_\delta X = \sum_{j=0}^{\infty} j\delta \mathbf{P}(\{\omega : j\delta < X(\omega) \leq (j+1)\delta\}) \,. \tag{1.3.10}$$

First use monotone convergence to show that $\mathbf{E}_{\delta_k} X$ *converges to* $\mathbf{E}X$ *along the sequence* $\delta_k = 2^{-k}$. *Then, check that* $\mathbf{E}_\delta X \leq \mathbf{E}_\eta X + \eta$ *for any* $\delta, \eta > 0$ *and deduce from it the identity (1.3.10).*

Applying (1.3.10) verify that if X *takes at most countably many values* $\{x_1, x_2, \ldots\}$, *then* $\mathbf{E}X = \sum_i x_i \mathbf{P}(\{\omega : X(\omega) = x_i\})$ *(this applies to every R.V.* $X \geq 0$ *on a countable* Ω*). More generally, verify that such formula applies whenever the series is absolutely convergent (which amounts to* $X \in L^1$*).*

Exercise 1.3.40. *Use monotone convergence to show that for any sequence of non-negative R.V.* Y_n,

$$\mathbf{E}(\sum_{n=1}^{\infty} Y_n) = \sum_{n=1}^{\infty} \mathbf{E}Y_n \,.$$

Exercise 1.3.41. *Suppose* $X_n, X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ *are such that*
(a) $X_n \geq 0$ *almost surely,* $\quad \mathbf{E}[X_n] = 1, \quad \mathbf{E}[X_n \log X_n] \leq 1,$ *and*
(b) $\mathbf{E}[X_n Y] \to \mathbf{E}[XY]$ *as* $n \to \infty$, *for each bounded random variable* Y *on* $(\Omega, \mathcal{F})$.

Show that then $X \geq 0$ *almost surely,* $\mathbf{E}[X] = 1$ *and* $\mathbf{E}[X \log X] \leq 1$.
Hint: *Jensen's inequality is handy for showing that* $\mathbf{E}[X \log X] \leq 1$.

Next come few direct applications of the dominated convergence theorem.

Exercise 1.3.42.
(a) *Show that for any random variable* X, *the function* $t \mapsto \mathbf{E}[e^{-|t-X|}]$ *is continuous on* $\mathbb{R}$ *(this function is sometimes called the* bilateral exponential transform*).*
(b) *Suppose* $X \geq 0$ *is such that* $\mathbf{E}X^q < \infty$ *for some* $q > 0$. *Show that then* $q^{-1}(\mathbf{E}X^q - 1) \to \mathbf{E}\log X$ *as* $q \downarrow 0$ *and deduce that also* $q^{-1} \log \mathbf{E}X^q \to \mathbf{E}\log X$ *as* $q \downarrow 0$.

Hint: *Fixing* $x \geq 0$ *deduce from convexity of* $q \mapsto x^q$ *that* $q^{-1}(x^q - 1) \downarrow \log x$ *as* $q \downarrow 0$.

Exercise 1.3.43. *Suppose* X *is an integrable random variable.*

(a) *Show that* $\mathbf{E}(|X|I_{\{X>n\}}) \to 0$ *as* $n \to \infty$.
(b) *Deduce that for any* $\varepsilon > 0$ *there exists* $\delta > 0$ *such that*

$$\sup\{\mathbf{E}[|X|I_A] : \mathbf{P}(A) \leq \delta\} \leq \varepsilon\,.$$

(c) *Provide an example of* $X \geq 0$ *with* $\mathbf{E}X = \infty$ *for which the preceding fails, that is,* $\mathbf{P}(A_k) \to 0$ *as* $k \to \infty$ *while* $\mathbf{E}[XI_{A_k}]$ *is bounded away from zero.*

The following generalization of the dominated convergence theorem is also left as an exercise.

Exercise 1.3.44. *Suppose* $g_n(\cdot) \leq f_n(\cdot) \leq h_n(\cdot)$ *are* μ*-integrable functions in the same measure space* $(\mathbb{S}, \mathcal{F}, \mu)$ *such that for* μ*-almost-every* $s \in \mathbb{S}$ *both* $g_n(s) \to g_\infty(s)$, $f_n(s) \to f_\infty(s)$ *and* $h_n(s) \to h_\infty(s)$ *as* $n \to \infty$. *Show that if further* g_∞ *and* h_∞ *are* μ*-integrable functions such that* $\mu(g_n) \to \mu(g_\infty)$ *and* $\mu(h_n) \to \mu(h_\infty)$, *then* f_∞ *is* μ*-integrable and* $\mu(f_n) \to \mu(f_\infty)$.

Finally, here is a demonstration of one of the many issues that are particularly easy to resolve with respect to the $L^2(\Omega, \mathcal{F}, \mathbf{P})$ norm.

Exercise 1.3.45. *Let* $X = (X(t))_{t\in\mathbb{R}}$ *be a mapping from* $\mathbb{R}$ *into* $L^2(\Omega, \mathcal{F}, \mathbf{P})$. *Show that* $t \mapsto X(t)$ *is a continuous mapping (with respect to the norm* $\|\cdot\|_2$ *on* $L^2(\Omega, \mathcal{F}, \mathbf{P})$*), if and only if both*

$$\mu(t) = \mathbf{E}[X(t)] \quad \textit{and} \quad r(s,t) = \mathbf{E}[X(s)X(t)] - \mu(s)\mu(t)$$

are continuous real-valued functions ($r(s,t)$ *is continuous as a map from* $\mathbb{R}^2$ *to* $\mathbb{R}$*).*

1.3.4. L^1**-convergence and uniform integrability.** For probability theory, the dominated convergence theorem states that if random variables $X_n \overset{a.s.}{\to} X_\infty$ are such that $|X_n| \leq Y$ for all n and some random variable Y such that $\mathbf{E}Y < \infty$, then $X_\infty \in L^1$ and $X_n \overset{L^1}{\to} X_\infty$. Since constants have finite expectation (see part (d) of Theorem 1.3.9), we have as its corollary the *bounded convergence theorem*, that is,

Corollary 1.3.46 (Bounded Convergence)**.** *Suppose that a.s.* $|X_n(\omega)| \leq K$ *for some finite non-random constant* K *and all* n. *If* $X_n \overset{a.s.}{\to} X_\infty$, *then* $X_\infty \in L^1$ *and* $X_n \overset{L^1}{\to} X_\infty$.

We next state a *uniform integrability* condition that together with convergence in probability implies the convergence in L^1.

Definition 1.3.47. *A possibly uncountable collection of R.V.-s* $\{X_\alpha, \alpha \in \mathcal{I}\}$ *is called* uniformly integrable *(U.I.) if*

$$\lim_{M\to\infty} \sup_\alpha \mathbf{E}[|X_\alpha| I_{|X_\alpha|>M}] = 0\,. \tag{1.3.11}$$

Our next lemma shows that U.I. is a relaxation of the condition of dominated convergence, and that U.I. still implies the boundedness in L^1 of $\{X_\alpha, \alpha \in \mathcal{I}\}$.

Lemma 1.3.48. *If* $|X_\alpha| \leq Y$ *for all* α *and some R.V.* Y *such that* $\mathbf{E}Y < \infty$, *then the collection* $\{X_\alpha\}$ *is U.I. In particular, any finite collection of integrable R.V. is U.I.*

Further, if $\{X_\alpha\}$ *is U.I. then* $\sup_\alpha \mathbf{E}|X_\alpha| < \infty$.

Proof. By monotone convergence, $\mathbf{E}[YI_{Y\le M}] \uparrow \mathbf{E}Y$ as $M \uparrow \infty$, for any R.V. $Y \ge 0$. Hence, if in addition $\mathbf{E}Y < \infty$, then by linearity of the expectation, $\mathbf{E}[YI_{Y>M}] \downarrow 0$ as $M \uparrow \infty$. Now, if $|X_\alpha| \le Y$ then $|X_\alpha| I_{|X_\alpha|>M} \le YI_{Y>M}$, hence $\mathbf{E}[|X_\alpha| I_{|X_\alpha|>M}] \le \mathbf{E}[YI_{Y>M}]$, which does not depend on α, and for $Y \in L^1$ converges to zero when $M \to \infty$. We thus proved that if $|X_\alpha| \le Y$ for all α and some Y such that $\mathbf{E}Y < \infty$, then $\{X_\alpha\}$ is a U.I. collection of R.V.-s

For a finite collection of R.V.-s $X_i \in L^1$, $i = 1, \dots, k$, take $Y = |X_1| + |X_2| + \cdots + |X_k| \in L^1$ such that $|X_i| \le Y$ for $i = 1, \dots, k$, to see that any finite collection of integrable R.V.-s is U.I.

Finally, since

$$\mathbf{E}|X_\alpha| = \mathbf{E}[|X_\alpha| I_{|X_\alpha|\le M}] + \mathbf{E}[|X_\alpha| I_{|X_\alpha|>M}] \le M + \sup_\alpha \mathbf{E}[|X_\alpha| I_{|X_\alpha|>M}] \,,$$

we see that if $\{X_\alpha, \alpha \in \mathcal{I}\}$ is U.I. then $\sup_\alpha \mathbf{E}|X_\alpha| < \infty$. $\square$

We next state and prove Vitali's convergence theorem for probability measures, deferring the general case to Exercise 1.3.53.

Theorem 1.3.49 (Vitali's convergence theorem). *Suppose $X_n \xrightarrow{p} X_\infty$. Then, the collection $\{X_n\}$ is U.I. if and only if $X_n \xrightarrow{L^1} X_\infty$ which in turn is equivalent to X_n being integrable for all $n \le \infty$ and $\mathbf{E}|X_n| \to \mathbf{E}|X_\infty|$.*

Remark. In view of Lemma 1.3.48, Vitali's theorem relaxes the assumed a.s. convergence $X_n \to X_\infty$ of the dominated (or bounded) convergence theorem, and of Scheffé's lemma, to that of convergence in probability.

Proof. Suppose first that $|X_n| \le M$ for some non-random finite constant M and all n. For each $\varepsilon > 0$ let $B_{n,\varepsilon} = \{\omega : |X_n(\omega) - X_\infty(\omega)| > \varepsilon\}$. The assumed convergence in probability means that $\mathbf{P}(B_{n,\varepsilon}) \to 0$ as $n \to \infty$ (see Definition 1.3.22). Since $\mathbf{P}(|X_\infty| \ge M + \varepsilon) \le \mathbf{P}(B_{n,\varepsilon})$, taking $n \to \infty$ and considering $\varepsilon = \varepsilon_k \downarrow 0$, we get by continuity from below of $\mathbf{P}$ that almost surely $|X_\infty| \le M$. So, $|X_n - X_\infty| \le 2M$ and by linearity and monotonicity of the expectation, for any n and $\varepsilon > 0$,

$$\begin{aligned}\mathbf{E}|X_n - X_\infty| &= \mathbf{E}[|X_n - X_\infty| I_{B^c_{n,\varepsilon}}] + \mathbf{E}[|X_n - X_\infty| I_{B_{n,\varepsilon}}] \\ &\le \mathbf{E}[\varepsilon I_{B^c_{n,\varepsilon}}] + \mathbf{E}[2M I_{B_{n,\varepsilon}}] \le \varepsilon + 2M\mathbf{P}(B_{n,\varepsilon})\,.\end{aligned}$$

Since $\mathbf{P}(B_{n,\varepsilon}) \to 0$ as $n \to \infty$, it follows that $\limsup_{n\to\infty} \mathbf{E}|X_n - X_\infty| \le \varepsilon$. Taking $\varepsilon \downarrow 0$ we deduce that $\mathbf{E}|X_n - X_\infty| \to 0$ in this case.

Moving to deal now with the general case of a collection $\{X_n\}$ that is U.I., let $\varphi_M(x) = \max(\min(x, M), -M)$. As $|\varphi_M(x) - \varphi_M(y)| \le |x - y|$ for any $x, y \in \mathbb{R}$, our assumption $X_n \xrightarrow{p} X_\infty$ implies that $\varphi_M(X_n) \xrightarrow{p} \varphi_M(X_\infty)$ for any fixed $M < \infty$. With $|\varphi_M(\cdot)| \le M$, we then have by the preceding proof of bounded convergence that $\varphi_M(X_n) \xrightarrow{L^1} \varphi_M(X_\infty)$. Further, by Minkowski's inequality, also $\mathbf{E}|\varphi_M(X_n)| \to \mathbf{E}|\varphi_M(X_\infty)|$. By Lemma 1.3.48, our assumption that $\{X_n\}$ are U.I. implies their L^1 boundedness, and since $|\varphi_M(x)| \le |x|$ for all x, we deduce that for any M,

$$\infty > c := \sup_n \mathbf{E}|X_n| \ge \lim_{n\to\infty} \mathbf{E}|\varphi_M(X_n)| = \mathbf{E}|\varphi_M(X_\infty)|\,. \tag{1.3.12}$$

With $|\varphi_M(X_\infty)| \uparrow |X_\infty|$ as $M \uparrow \infty$, it follows from monotone convergence that $\mathbf{E}|\varphi_M(X_\infty)| \uparrow \mathbf{E}|X_\infty|$, hence $\mathbf{E}|X_\infty| \le c < \infty$ in view of (1.3.12). Fixing $\varepsilon > 0$, choose $M = M(\varepsilon) < \infty$ large enough for $\mathbf{E}[|X_\infty| I_{|X_\infty|>M}] < \varepsilon$, and further increasing M if needed, by the U.I. condition also $\mathbf{E}[|X_n|I_{|X_n|>M}] < \varepsilon$ for all n. Considering the expectation of the inequality $|x-\varphi_M(x)| \le |x|I_{|x|>M}$ (which holds for all $x \in \mathbb{R}$), with $x = X_n$ and $x = X_\infty$, we obtain that

$$\begin{aligned}\mathbf{E}|X_n - X_\infty| &\le \mathbf{E}|X_n - \varphi_M(X_n)| + \mathbf{E}|\varphi_M(X_n) - \varphi_M(X_\infty)| + \mathbf{E}|X_\infty - \varphi_M(X_\infty)| \\ &\le 2\varepsilon + \mathbf{E}|\varphi_M(X_n) - \varphi_M(X_\infty)|\,.\end{aligned}$$

Recall that $\varphi_M(X_n) \stackrel{L^1}{\to} \varphi_M(X_\infty)$, hence $\limsup_n \mathbf{E}|X_n - X_\infty| \le 2\varepsilon$. Taking $\varepsilon \to 0$ completes the proof of L^1 convergence of X_n to X_∞.

Suppose now that $X_n \stackrel{L^1}{\to} X_\infty$. Then, by Jensen's inequality (for the convex function $g(x) = |x|$),

$$|\mathbf{E}|X_n| - \mathbf{E}|X_\infty|| \le \mathbf{E}[|\,|X_n| - |X_\infty|\,|] \le \mathbf{E}|X_n - X_\infty| \to 0.$$

That is, $\mathbf{E}|X_n| \to \mathbf{E}|X_\infty|$ and X_n, $n \le \infty$ are integrable.

It thus remains only to show that if $X_n \stackrel{p}{\to} X_\infty$, all of which are integrable and $\mathbf{E}|X_n| \to \mathbf{E}|X_\infty|$ then the collection $\{X_n\}$ is U.I. To the end, for any $M > 1$, let

$$\psi_M(x) = |x|I_{|x|\le M-1} + (M-1)(M-|x|)I_{(M-1,M]}(|x|)\,,$$

a piecewise-linear, continuous, bounded function, such that $\psi_M(x) = |x|$ for $|x| \le M-1$ and $\psi_M(x) = 0$ for $|x| \ge M$. Fixing $\epsilon > 0$, with X_∞ integrable, by dominated convergence $\mathbf{E}|X_\infty| - \mathbf{E}\psi_m(X_\infty) \le \epsilon$ for some finite $m = m(\epsilon)$. Further, as $|\psi_m(x) - \psi_m(y)| \le (m-1)|x-y|$ for any $x, y \in \mathbb{R}$, our assumption $X_n \stackrel{p}{\to} X_\infty$ implies that $\psi_m(X_n) \stackrel{p}{\to} \psi_m(X_\infty)$. Hence, by the preceding proof of bounded convergence, followed by Minkowski's inequality, we deduce that $\mathbf{E}\psi_m(X_n) \to \mathbf{E}\psi_m(X_\infty)$ as $n \to \infty$. Since $|x|I_{|x|>m} \le |x| - \psi_m(x)$ for all $x \in \mathbb{R}$, our assumption $\mathbf{E}|X_n| \to \mathbf{E}|X_\infty|$ thus implies that for some $n_0 = n_0(\epsilon)$ finite and all $n \ge n_0$ and $M \ge m(\epsilon)$,

$$\begin{aligned}\mathbf{E}[|X_n|I_{|X_n|>M}] \le \mathbf{E}[|X_n|I_{|X_n|>m}] &\le \mathbf{E}|X_n| - \mathbf{E}\psi_m(X_n) \\ &\le \mathbf{E}|X_\infty| - \mathbf{E}\psi_m(X_\infty) + \epsilon \le 2\epsilon\,.\end{aligned}$$

As each X_n is integrable, $\mathbf{E}[|X_n|I_{|X_n|>M}] \le 2\epsilon$ for some $M \ge m$ finite and all n (including also $n < n_0(\epsilon)$). The fact that such finite $M = M(\epsilon)$ exists for any $\epsilon > 0$ amounts to the collection $\{X_n\}$ being U.I. $\square$

The following exercise builds upon the bounded convergence theorem.

Exercise 1.3.50. *Show that for any $X \ge 0$ (do not assume $\mathbf{E}(1/X) < \infty$), both*

(a) $\lim_{y\to\infty} y\mathbf{E}[X^{-1}I_{X>y}] = 0$ *and*

(b) $\lim_{y\downarrow 0} y\mathbf{E}[X^{-1}I_{X>y}] = 0$.

Next is an example of the advantage of Vitali's convergence theorem over the dominated convergence theorem.

Exercise 1.3.51. *On $((0,1], \mathcal{B}_{(0,1]}, U)$, let $X_n(\omega) = (n/\log n)I_{(0,n^{-1})}(\omega)$ for $n \ge 2$. Show that the collection $\{X_n\}$ is U.I. such that $X_n \stackrel{a.s.}{\to} 0$ and $\mathbf{E}X_n \to 0$, but there is no random variable Y with finite expectation such that $Y \ge X_n$ for all $n \ge 2$ and almost all $\omega \in (0,1]$.*

By a simple application of Vitali's convergence theorem you can derive a classical result of analysis, dealing with the convergence of Cesáro averages.

Exercise 1.3.52. *Let U_n denote a random variable whose law is the* uniform probability measure *on $(0,n]$, namely, Lebesgue measure restricted to the interval $(0,n]$ and normalized by n^{-1} to a probability measure. Show that $g(U_n) \xrightarrow{p} 0$ as $n \to \infty$, for any Borel function $g(\cdot)$ such that $|g(y)| \to 0$ as $y \to \infty$. Further, assuming that also $\sup_y |g(y)| < \infty$, deduce that $\mathbf{E}|g(U_n)| = n^{-1} \int_0^n |g(y)| dy \to 0$ as $n \to \infty$.*

Here is Vitali's convergence theorem for a general measure space.

Exercise 1.3.53. *Given a measure space $(\mathbb{S}, \mathcal{F}, \mu)$, suppose $f_n, f_\infty \in m\mathcal{F}$ with $\mu(|f_n|)$ finite and $\mu(|f_n - f_\infty| > \varepsilon) \to 0$ as $n \to \infty$, for each fixed $\varepsilon > 0$. Show that $\mu(|f_n - f_\infty|) \to 0$ as $n \to \infty$ if and only if both $\sup_n \mu(|f_n| I_{|f_n|>k}) \to 0$ and $\sup_n \mu(|f_n| I_{A_k}) \to 0$ for $k \to \infty$ and some $\{A_k\} \subseteq \mathcal{F}$ such that $\mu(A_k^c) < \infty$.*

We conclude this subsection with a useful sufficient criterion for uniform integrability and few of its consequences.

Exercise 1.3.54. *Let $f \geq 0$ be a Borel function such that $f(r)/r \to \infty$ as $r \to \infty$. Suppose $\mathbf{E} f(|X_\alpha|) \leq C$ for some finite non-random constant C and all $\alpha \in \mathcal{I}$. Show that then $\{X_\alpha : \alpha \in \mathcal{I}\}$ is a uniformly integrable collection of R.V.*

Exercise 1.3.55.

(a) *Construct random variables X_n such that $\sup_n \mathbf{E}(|X_n|) < \infty$, but the collection $\{X_n\}$ is not uniformly integrable.*
(b) *Show that if $\{X_n\}$ is a U.I. collection and $\{Y_n\}$ is a U.I. collection, then $\{X_n + Y_n\}$ is also U.I.*
(c) *Show that if $X_n \xrightarrow{p} X_\infty$ and the collection $\{X_n\}$ is uniformly integrable, then $\mathbf{E}(X_n I_A) \to \mathbf{E}(X_\infty I_A)$ as $n \to \infty$, for any measurable set A.*

1.3.5. Expectation, density and Riemann integral. Applying the *standard machine* we now show that fixing a measure space $(\mathbb{S}, \mathcal{F}, \mu)$, each non-negative measurable function f induces a measure $f\mu$ on $(\mathbb{S}, \mathcal{F})$, with f being the natural generalization of the concept of probability density function.

Proposition 1.3.56. *Fix a measure space $(\mathbb{S}, \mathcal{F}, \mu)$. Every $f \in m\mathcal{F}_+$ induces a measure $f\mu$ on $(\mathbb{S}, \mathcal{F})$ via $(f\mu)(A) = \mu(f I_A)$ for all $A \in \mathcal{F}$. These measures satisfy the composition relation $h(f\mu) = (hf)\mu$ for all $f, h \in m\mathcal{F}_+$. Further, $h \in L^1(\mathbb{S}, \mathcal{F}, f\mu)$ if and only if $fh \in L^1(\mathbb{S}, \mathcal{F}, \mu)$ and then $(f\mu)(h) = \mu(fh)$.*

PROOF. Fixing $f \in m\mathcal{F}_+$, obviously $f\mu$ is a non-negative set function on $(\mathbb{S}, \mathcal{F})$ with $(f\mu)(\emptyset) = \mu(f I_\emptyset) = \mu(0) = 0$. To check that $f\mu$ is countably additive, hence a measure, let $A = \cup_k A_k$ for a countable collection of disjoint sets $A_k \in \mathcal{F}$. Since $\sum_{k=1}^n f I_{A_k} \uparrow f I_A$, it follows by monotone convergence and linearity of the integral that,

$$\mu(f I_A) = \lim_{n \to \infty} \mu(\sum_{k=1}^n f I_{A_k}) = \lim_{n \to \infty} \sum_{k=1}^n \mu(f I_{A_k}) = \sum_k \mu(f I_{A_k})$$

Thus, $(f\mu)(A) = \sum_k (f\mu)(A_k)$ verifying that $f\mu$ is a measure.

Fixing $f \in m\mathcal{F}_+$, we turn to prove that the identity

$$(f\mu)(h I_A) = \mu(f h I_A) \qquad \forall A \in \mathcal{F}\,, \tag{1.3.13}$$

holds for any $h \in m\mathcal{F}_+$. Since the left side of (1.3.13) is the value assigned to A by the measure $h(f\mu)$ and the right side of this identity is the value assigned to

the same set by the measure $(hf)\mu$, this would verify the stated composition rule $h(f\mu) = (hf)\mu$. The proof of (1.3.13) proceeds by applying the standard machine:
Step 1. If $h = I_B$ for $B \in \mathcal{F}$ we have by the definition of the integral of an indicator function that

$$(f\mu)(I_B I_A) = (f\mu)(I_{A\cap B}) = (f\mu)(A \cap B) = \mu(fI_{A\cap B}) = \mu(fI_B I_A)\,,$$

which is (1.3.13).
Step 2. Take $h \in \mathrm{SF}_+$ represented as $h = \sum_{l=1}^n c_l I_{B_l}$ with $c_l \ge 0$ and $B_l \in \mathcal{F}$. Then, by Step 1 and the linearity of the integrals with respect to $f\mu$ and with respect to μ, we see that

$$(f\mu)(hI_A) = \sum_{l=1}^n c_l (f\mu)(I_{B_l} I_A) = \sum_{l=1}^n c_l \mu(fI_{B_l} I_A) = \mu(f\sum_{l=1}^n c_l I_{B_l} I_A) = \mu(fhI_A)\,,$$

again yielding (1.3.13).
Step 3. For any $h \in m\mathcal{F}_+$ there exist $h_n \in \mathrm{SF}_+$ such that $h_n \uparrow h$. By Step 2 we know that $(f\mu)(h_n I_A) = \mu(fh_n I_A)$ for any $A \in \mathcal{F}$ and all n. Further, $h_n I_A \uparrow hI_A$ and $fh_n I_A \uparrow fhI_A$, so by monotone convergence (for both integrals with respect to $f\mu$ and μ),

$$(f\mu)(hI_A) = \lim_{n\to\infty} (f\mu)(h_n I_A) = \lim_{n\to\infty} \mu(fh_n I_A) = \mu(fhI_A)\,,$$

completing the proof of (1.3.13) for all $h \in m\mathcal{F}_+$.
Writing $h \in m\mathcal{F}$ as $h = h_+ - h_-$ with $h_+ = \max(h, 0) \in m\mathcal{F}_+$ and $h_- = -\min(h, 0) \in m\mathcal{F}_+$, it follows from the composition rule that

$$\int h_\pm d(f\mu) = (f\mu)(h_\pm I_{\mathbb{S}}) = h_\pm(f\mu)(\mathbb{S}) = ((h_\pm f)\mu)(\mathbb{S}) = \mu(fh_\pm I_{\mathbb{S}}) = \int fh_\pm d\mu\,.$$

Observing that $fh_\pm = (fh)_\pm$ when $f \in m\mathcal{F}_+$, we thus deduce that h is $f\mu$-integrable if and only if fh is μ-integrable in which case $\int hd(f\mu) = \int fhd\mu$, as stated. ☐

Fixing a measure space $(\mathbb{S}, \mathcal{F}, \mu)$, every set $D \in \mathcal{F}$ induces a σ-algebra $\mathcal{F}_D = \{A \in \mathcal{F} : A \subseteq D\}$. Let μ_D denote the *restriction* of μ to $(D, \mathcal{F}_D)$. As a corollary of Proposition 1.3.56 we express the integral with respect to μ_D in terms of the original measure μ.

Corollary 1.3.57. *Fixing $D \in \mathcal{F}$ let h_D denote the restriction of $h \in m\mathcal{F}$ to $(D, \mathcal{F}_D)$. Then, $\mu_D(h_D) = \mu(hI_D)$ for any $h \in m\mathcal{F}_+$. Further, $h_D \in L^1(D, \mathcal{F}_D, \mu_D)$ if and only if $hI_D \in L^1(\mathbb{S}, \mathcal{F}, \mu)$, in which case also $\mu_D(h_D) = \mu(hI_D)$.*

Proof. Note that the measure $I_D\mu$ of Proposition 1.3.56 coincides with μ_D on the σ-algebra $\mathcal{F}_D$ and assigns to any set $A \in \mathcal{F}$ the same value it assigns to $A \cap D \in \mathcal{F}_D$. By Definition 1.3.1 this implies that $(I_D\mu)(h) = \mu_D(h_D)$ for any $h \in m\mathcal{F}_+$. The corollary is thus a re-statement of the composition and integrability relations of Proposition 1.3.56 for $f = I_D$. ☐

Remark 1.3.58. Corollary 1.3.57 justifies using hereafter the notation $\int_A fd\mu$ or $\mu(f; A)$ for $\mu(fI_A)$, or writing $\mathbf{E}(X; A) = \int_A X(\omega)dP(\omega)$ for $\mathbf{E}(XI_A)$. With this notation in place, Proposition 1.3.56 states that each $Z \ge 0$ such that $\mathbf{E}Z = 1$ induces a probability measure $\mathbf{Q} = Z\mathbf{P}$ such that $\mathbf{Q}(A) = \int_A Zd\mathbf{P}$ for all $A \in \mathcal{F}$, and then $\mathbf{E}_{\mathbf{Q}}(W) := \int Wd\mathbf{Q} = \mathbf{E}(ZW)$ whenever $W \ge 0$ or $ZW \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ (the assumption $\mathbf{E}Z = 1$ translates to $\mathbf{Q}(\Omega) = 1$).

Proposition 1.3.56 is closely related to the probability density function of Definition 1.2.40. En-route to showing this, we first define the collection of Lebesgue integrable functions.

Definition 1.3.59. *Consider Lebesgue's measure λ on $(\mathbb{R}, \mathcal{B})$ as in Section 1.1.3, and its* completion *$\overline{\lambda}$ on $(\mathbb{R}, \overline{\mathcal{B}})$ (see Theorem 1.1.35). A set $B \in \overline{\mathcal{B}}$ is called* Lebesgue measurable *and $f : \mathbb{R} \mapsto \mathbb{R}$ is called* Lebesgue integrable function *if $f \in m\overline{\mathcal{B}}$, and $\overline{\lambda}(|f|) < \infty$. As we show in Proposition 1.3.64, any non-negative Riemann integrable function is also Lebesgue integrable, and the integral values coincide, justifying the notation $\int_B f(x)dx$ for $\overline{\lambda}(f; B)$, where the function f and the set B are both Lebesgue measurable.*

Example 1.3.60. *Suppose f is a non-negative Lebesgue integrable function such that $\int_{\mathbb{R}} f(x)dx = 1$. Then, $\mathcal{P} = f\overline{\lambda}$ of Proposition 1.3.56 is a probability measure on $(\mathbb{R}, \overline{\mathcal{B}})$ such that $\mathcal{P}(B) = \overline{\lambda}(f; B) = \int_B f(x)dx$ for any Lebesgue measurable set B. By Theorem 1.2.37 it is easy to verify that $F(\alpha) = \mathcal{P}((-\infty, \alpha])$ is a distribution function, such that $F(\alpha) = \int_{-\infty}^{\alpha} f(x)dx$. That is, $\mathcal{P}$ is the law of a R.V. $X : \mathbb{R} \mapsto \mathbb{R}$ whose probability density function is f (c.f. Definition 1.2.40 and Proposition 1.2.45).*

Our next theorem allows us to compute expectations of functions of a R.V. X in the space $(\mathbb{R}, \mathcal{B}, \mathcal{P}_X)$, using the law of X (c.f. Definition 1.2.34) and calculus, instead of working on the original general probability space. One of its immediate consequences is the "obvious" fact that if $X \stackrel{\mathcal{D}}{=} Y$ then $\mathbf{E}h(X) = \mathbf{E}h(Y)$ for any non-negative Borel function h.

Theorem 1.3.61 (Change of variables formula)**.** *Let $X : \Omega \mapsto \mathbb{R}$ be a random variable on $(\Omega, \mathcal{F}, \mathbf{P})$ and h a Borel measurable function such that $\mathbf{E}h_+(X) < \infty$ or $\mathbf{E}h_-(X) < \infty$. Then,*

$$\int_\Omega h(X(\omega))d\mathbf{P}(\omega) = \int_{\mathbb{R}} h(x)d\mathcal{P}_X(x). \tag{1.3.14}$$

Proof. Apply the standard machine with respect to $h \in m\mathcal{B}$:
Step 1. Taking $h = I_B$ for $B \in \mathcal{B}$, note that by the definition of expectation of indicators

$$\mathbf{E}h(X) = \mathbf{E}[I_B(X(\omega))] = \mathbf{P}(\{\omega : X(\omega) \in B\}) = \mathcal{P}_X(B) = \int h(x)d\mathcal{P}_X(x).$$

Step 2. Representing $h \in \text{SF}_+$ as $h = \sum_{l=1}^m c_l I_{B_l}$ for $c_l \geq 0$, the identity (1.3.14) follows from Step 1 by the linearity of the expectation in both spaces.
Step 3. For $h \in m\mathcal{B}_+$, consider $h_n \in \text{SF}_+$ such that $h_n \uparrow h$. Since $h_n(X(\omega)) \uparrow h(X(\omega))$ for all ω, we get by monotone convergence on $(\Omega, \mathcal{F}, \mathbf{P})$, followed by applying Step 2 for h_n, and finally monotone convergence on $(\mathbb{R}, \mathcal{B}, \mathcal{P}_X)$, that

$$\begin{aligned}\int_\Omega h(X(\omega))d\mathbf{P}(\omega) &= \lim_{n\to\infty} \int_\Omega h_n(X(\omega))d\mathbf{P}(\omega) \\ &= \lim_{n\to\infty} \int_{\mathbb{R}} h_n(x)d\mathcal{P}_X(x) = \int_{\mathbb{R}} h(x)d\mathcal{P}_X(x),\end{aligned}$$

as claimed.
Step 4. Write a Borel function $h(x)$ as $h_+(x) - h_-(x)$. Then, by Step 3, (1.3.14) applies for both non-negative functions h_+ and h_-. Further, at least one of these

two identities involves finite quantities. So, taking their difference and using the linearity of the expectation (in both probability spaces), lead to the same result for h. □

Combining Theorem 1.3.61 with Example 1.3.60, we show that the expectation of a Borel function of a R.V. X having a density f_X can be computed by performing calculus type integration on the real line.

Corollary 1.3.62. *Suppose that the distribution function of a R.V. X is of the form (1.2.3) for some Lebesgue integrable function $f_X(x)$. Then, for any Borel measurable function $h : \mathbb{R} \mapsto \mathbb{R}$, the R.V. $h(X)$ is integrable if and only if $\int |h(x)|f_X(x)dx < \infty$, in which case $\mathbf{E}h(X) = \int h(x)f_X(x)dx$. The latter formula applies also for any non-negative Borel function $h(\cdot)$.*

Proof. Recall Example 1.3.60 that the law $\mathcal{P}_X$ of X equals to the probability measure $f_X\overline{\lambda}$. For $h \geq 0$ we thus deduce from Theorem 1.3.61 that $\mathbf{E}h(X) = f_X\overline{\lambda}(h)$, which by the composition rule of Proposition 1.3.56 is given by $\overline{\lambda}(f_X h) = \int h(x)f_X(x)dx$. The decomposition $h = h_+ - h_-$ then completes the proof of the general case. □

Our next task is to compare Lebesgue's integral (of Definition 1.3.1) with Riemann's integral. To this end recall,

Definition 1.3.63. *A function $f : (a, b] \mapsto [0, \infty]$ is* Riemann integrable *with integral $R(f) < \infty$ if for any $\varepsilon > 0$ there exists $\delta = \delta(\varepsilon) > 0$ such that $|\sum_l f(x_l)\lambda(J_l) - R(f)| < \varepsilon$, for any $x_l \in J_l$ and $\{J_l\}$ a finite partition of $(a, b]$ into disjoint subintervals whose length $\lambda(J_l) < \delta$.*

Lebesgue's integral of a function f is based on splitting its *range* to small intervals and approximating $f(s)$ by a constant on the subset of $\mathbb{S}$ for which $f(\cdot)$ falls into each such interval. As such, it accommodates an arbitrary domain $\mathbb{S}$ of the function, in contrast to Riemann's integral where the *domain* of integration is split into small rectangles – hence limited to $\mathbb{R}^d$. As we next show, even for $\mathbb{S} = (a, b]$, if $f \geq 0$ (or more generally, f bounded), is Riemann integrable, then it is also Lebesgue integrable, with the integrals coinciding in value.

Proposition 1.3.64. *If $f(x)$ is a non-negative Riemann integrable function on an interval $(a, b]$, then it is also Lebesgue integrable on $(a, b]$ and $\overline{\lambda}(f) = R(f)$.*

Proof. Let $f_*(J) = \inf\{f(x) : x \in J\}$ and $f^*(J) = \sup\{f(x) : x \in J\}$. Varying x_l over J_l we see that

$$R(f) - \varepsilon \leq \sum_l f_*(J_l)\lambda(J_l) \leq \sum_l f^*(J_l)\lambda(J_l) \leq R(f) + \varepsilon\,, \tag{1.3.15}$$

for any finite partition Π of $(a, b]$ into disjoint subintervals J_l such that $\sup_l \lambda(J_l) \leq \delta$. For any such partition, the non-negative simple functions $\ell(\Pi) = \sum_l f_*(J_l)I_{J_l}$ and $u(\Pi) = \sum_l f^*(J_l)I_{J_l}$ are such that $\ell(\Pi) \leq f \leq u(\Pi)$, whereas $R(f) - \varepsilon \leq \lambda(\ell(\Pi)) \leq \lambda(u(\Pi)) \leq R(f) + \varepsilon$, by (1.3.15). Consider the dyadic partitions Π_n of $(a, b]$ to 2^n intervals of length $(b - a)2^{-n}$ each, such that Π_{n+1} is a refinement of Π_n for each $n = 1, 2, \ldots$. Note that $u(\Pi_n)(x) \geq u(\Pi_{n+1})(x)$ for all $x \in (a, b]$ and any n, hence $u(\Pi_n))(x) \downarrow u_\infty(x)$ a Borel measurable $\overline{\mathbb{R}}$-valued function (see

Exercise 1.2.31). Similarly, $\ell(\Pi_n)(x) \uparrow \ell_\infty(x)$ for all $x \in (a,b]$, with ℓ_∞ also Borel measurable, and by the monotonicity of Lebesgue's integral,

$$R(f) \leq \lim_{n\infty} \lambda(\ell(\Pi_n)) \leq \lambda(\ell_\infty) \leq \lambda(u_\infty) \leq \lim_{n\to\infty} \lambda(u(\Pi_n)) \leq R(f)\,.$$

We deduce that $\lambda(u_\infty) = \lambda(\ell_\infty) = R(f)$ for $u_\infty \geq f \geq \ell_\infty$. The set $\{x \in (a,b] : f(x) \neq \ell_\infty(x)\}$ is a subset of the Borel set $\{x \in (a,b] : u_\infty(x) > \ell_\infty(x)\}$ whose Lebesgue measure is zero (see Lemma 1.3.8). Consequently, f is Lebesgue measurable on $(a,b]$ with $\overline{\lambda}(f) = \lambda(\ell_\infty) = R(f)$ as stated. $\square$

Here is an alternative, direct proof of the fact that $\mathbf{Q}$ in Remark 1.3.58 is a probability measure.

Exercise 1.3.65. *Suppose* $\mathbf{E}|X| < \infty$ *and* $A = \bigcup_n A_n$ *for some disjoint sets* $A_n \in \mathcal{F}$.

(a) *Show that then*

$$\sum_{n=0}^{\infty} \mathbf{E}(X; A_n) = \mathbf{E}(X; A)\,,$$

that is, the sum converges absolutely and has the value on the right.

(b) *Deduce from this that for* $Z \geq 0$ *with* $\mathbf{E}Z$ *positive and finite,* $\mathbf{Q}(A) := \mathbf{E}ZI_A/\mathbf{E}Z$ *is a probability measure.*

(c) *Suppose that* X *and* Y *are non-negative random variables on the same probability space* $(\Omega, \mathcal{F}, \mathbf{P})$ *such that* $\mathbf{E}X = \mathbf{E}Y < \infty$. *Deduce from the preceding that if* $\mathbf{E}XI_A = \mathbf{E}YI_A$ *for any* A *in a* π-*system* $\mathcal{A}$ *such that* $\mathcal{F} = \sigma(\mathcal{A})$, *then* $X \overset{a.s.}{=} Y$.

Exercise 1.3.66. *Suppose* $\mathcal{P}$ *is a probability measure on* $(\mathbb{R}, \mathcal{B})$ *and* $f \geq 0$ *is a Borel function such that* $\mathcal{P}(B) = \int_B f(x)dx$ *for* $B = (-\infty, b]$, $b \in \mathbb{R}$. *Using the* $\pi - \lambda$ *theorem show that this identity applies for all* $B \in \mathcal{B}$. *Building on this result, use the standard machine to directly prove Corollary 1.3.62 (without Proposition 1.3.56).*

1.3.6. Mean, variance and moments.

We start with the definition of moments of a random variable.

Definition 1.3.67. *If* k *is a positive integer then* $\mathbf{E}X^k$ *is called the* k*th* moment *of* X. *When it is well defined, the first moment* $m_X = \mathbf{E}X$ *is called the* mean. *If* $\mathbf{E}X^2 < \infty$, *then the* variance *of* X *is defined to be*

$$\mathsf{Var}(X) = \mathbf{E}(X - m_X)^2 = \mathbf{E}X^2 - m_X^2 \leq \mathbf{E}X^2\,. \tag{1.3.16}$$

Since $\mathbf{E}(aX + b) = a\mathbf{E}X + b$ (linearity of the expectation), it follows from the definition that

$$\mathsf{Var}(aX + b) = \mathbf{E}(aX + b - \mathbf{E}(aX + b))^2 = a^2\mathbf{E}(X - m_X)^2 = a^2\,\mathsf{Var}(X) \tag{1.3.17}$$

We turn to some examples, starting with R.V. having a density.

Example 1.3.68. *If* X *has the* exponential distribution *then*

$$\mathbf{E}X^k = \int_0^\infty x^k e^{-x} dx = k!$$

for any k *(see Example 1.2.41 for its density). The mean of* X *is* $m_X = 1$ *and its variance is* $\mathbf{E}X^2 - (\mathbf{E}X)^2 = 1$. *For any* $\lambda > 0$, *it is easy to see that* $T = X/\lambda$

has density $f_T(t) = \lambda e^{-\lambda t}\mathbf{1}_{t>0}$, *called the* exponential density *of parameter* λ. *By (1.3.17) it follows that* $m_T = 1/\lambda$ *and* $\mathsf{Var}(T) = 1/\lambda^2$.

Similarly, if X *has a standard normal distribution, then by symmetry, for* k *odd,*

$$\mathbf{E}X^k = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^k e^{-x^2/2}dx = 0\,,$$

whereas by integration by parts, the even moments satisfy the relation

$$\mathbf{E}X^{2\ell} = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} x^{2\ell-1}xe^{-x^2/2}dx = (2\ell-1)\mathbf{E}X^{2\ell-2}\,, \tag{1.3.18}$$

for $\ell = 1,2,\ldots$. *In particular,*

$$\mathsf{Var}(X) = \mathbf{E}X^2 = 1\,.$$

Consider $G = \sigma X + \mu$, *where* $\sigma > 0$ *and* $\mu \in \mathbb{R}$, *whose density is*

$$f_G(y) = \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y-\mu)^2}{2\sigma^2}}\,.$$

We call the law of G *the* normal distribution *of mean* μ *and variance* σ^2 *(as* $\mathbf{E}G = \mu$ *and* $\mathsf{Var}(G) = \sigma^2$*).*

Next are some examples of R.V. with finite or countable set of possible values.

Example 1.3.69. *We say that* B *has a* Bernoulli distribution *of parameter* $p \in [0,1]$ *if* $\mathbf{P}(B=1) = 1 - \mathbf{P}(B=0) = p$. *Clearly,*

$$\mathbf{E}B = p\cdot 1 + (1-p)\cdot 0 = p\,.$$

Further, $B^2 = B$ *so* $\mathbf{E}B^2 = \mathbf{E}B = p$ *and*

$$\mathsf{Var}(B) = \mathbf{E}B^2 - (\mathbf{E}B)^2 = p - p^2 = p(1-p)\,.$$

Recall that N *has a* Poisson distribution *with parameter* $\lambda \geq 0$ *if*

$$\mathbf{P}(N=k) = \frac{\lambda^k}{k!}e^{-\lambda} \quad \textit{for} \quad k = 0,1,2,\ldots$$

(where in case $\lambda = 0$, $\mathbf{P}(N=0) = 1$*). Observe that for* $k = 1,2,\ldots$,

$$\begin{aligned}\mathbf{E}(N(N-1)\cdots(N-k+1)) &= \sum_{n=k}^{\infty} n(n-1)\cdots(n-k+1)\frac{\lambda^n}{n!}e^{-\lambda} \\ &= \lambda^k \sum_{n=k}^{\infty}\frac{\lambda^{n-k}}{(n-k)!}e^{-\lambda} = \lambda^k\,.\end{aligned}$$

Using this formula, it follows that $\mathbf{E}N = \lambda$ *while*

$$\mathsf{Var}(N) = \mathbf{E}N^2 - (\mathbf{E}N)^2 = \lambda\,.$$

The random variable Z *is said to have a* Geometric distribution *of success probability* $p \in (0,1)$ *if*

$$\mathbf{P}(Z=k) = p(1-p)^{k-1} \quad \textit{for} \quad k = 1,2,\ldots$$

This is the distribution of the number of independent coin tosses needed till the first appearance of a Head, or more generally, the number of independent trials till the *first occurrence in this sequence of a specific event whose probability is p. Then,*

$$\mathbf{E}Z = \sum_{k=1}^{\infty} kp(1-p)^{k-1} = \frac{1}{p}$$
$$\mathbf{E}Z(Z-1) = \sum_{k=2}^{\infty} k(k-1)p(1-p)^{k-1} = \frac{2(1-p)}{p^2}$$
$$\mathsf{Var}(Z) = \mathbf{E}Z(Z-1) + \mathbf{E}Z - (\mathbf{E}Z)^2 = \frac{1-p}{p^2}\,.$$

Exercise 1.3.70. *Consider a counting random variable* $N_n = \sum_{i=1}^n I_{A_i}$.

(a) *Provide a formula for* $\mathsf{Var}(N_n)$ *in terms of* $\mathbf{P}(A_i)$ *and* $\mathbf{P}(A_i \cap A_j)$ *for* $i \neq j$.

(b) *Using your formula, find the variance of the number* N_n *of empty boxes when distributing at random* r *distinct balls among* n *distinct boxes, where each of the possible* n^r *assignments of balls to boxes is equally likely.*

Exercise 1.3.71. *Show that if* $\mathbf{P}(X \in [a,b]) = 1$, *then* $\mathsf{Var}(X) \leq (b-a)^2/4$.

1.4. Independence and product measures

In Subsection 1.4.1 we build-up the notion of independence, from events to random variables via σ-algebras, relating it to the structure of the joint distribution function. Subsection 1.4.2 considers finite product measures associated with the joint law of independent R.V.-s. This is followed by Kolmogorov's extension theorem which we use in order to construct infinitely many independent R.V.-s. Subsection 1.4.3 is about Fubini's theorem and its applications for computing the expectation of functions of independent R.V.

1.4.1. Definition and conditions for independence.

Recall the classical definition that two events $A, B \in \mathcal{F}$ are *independent* if $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$.

For example, suppose two fair dice are thrown (i.e. $\Omega = \{1,2,3,4,5,6\}^2$ with $\mathcal{F} = 2^{\Omega}$ and the uniform probability measure). Let $E_1 = \{\text{Sum of two is } 6\}$ and $E_2 = \{\text{first die is } 4\}$ then E_1 and E_2 are not independent since

$$\mathbf{P}(E_1) = \mathbf{P}(\{(1,5)\ (2,4)\ (3,3)\ (4,2)\ (5,1)\}) = \frac{5}{36}, \quad \mathbf{P}(E_2) = \mathbf{P}(\{\omega : \omega_1 = 4\}) = \frac{1}{6}$$

and

$$\mathbf{P}(E_1 \cap E_2) = \mathbf{P}(\{(4,2)\}) = \frac{1}{36} \neq \mathbf{P}(E_1)\mathbf{P}(E_2).$$

However one can check that E_2 and $E_3 = \{\text{sum of dice is } 7\}$ are independent.

In analogy with the independence of events we define the independence of two random vectors and more generally, that of two σ-algebras.

Definition 1.4.1. *Two σ-algebras $\mathcal{H}, \mathcal{G} \subseteq \mathcal{F}$ are independent (also denoted* **P**-independent*), if*

$$\mathbf{P}(G \cap H) = \mathbf{P}(G)\mathbf{P}(H), \qquad \forall G \in \mathcal{G},\ \forall H \in \mathcal{H}\ ,$$

that is, two σ-algebras are independent if every event in one of them is independent of every event in the other.

The random vectors $\underline{X} = (X_1, \ldots, X_n)$ *and* $\underline{Y} = (Y_1, \ldots, Y_m)$ *on the same probability space are independent if the corresponding* σ*-algebras* $\sigma(X_1, \ldots, X_n)$ *and* $\sigma(Y_1, \ldots, Y_m)$ *are independent.*

Remark. Our definition of independence of random variables is consistent with that of independence of events. For example, if the events $A, B \in \mathcal{F}$ are independent, then so are I_A and I_B. Indeed, we need to show that $\sigma(I_A) = \{\emptyset, \Omega, A, A^c\}$ and $\sigma(I_B) = \{\emptyset, \Omega, B, B^c\}$ are independent. Since $\mathbf{P}(\emptyset) = 0$ and $\emptyset$ is invariant under intersections, whereas $\mathbf{P}(\Omega) = 1$ and all events are invariant under intersection with Ω, it suffices to consider $G \in \{A, A^c\}$ and $H \in \{B, B^c\}$. We check independence first for $G = A$ and $H = B^c$. Noting that A is the union of the disjoint events $A \cap B$ and $A \cap B^c$ we have that

$$\mathbf{P}(A \cap B^c) = \mathbf{P}(A) - \mathbf{P}(A \cap B) = \mathbf{P}(A)[1 - \mathbf{P}(B)] = \mathbf{P}(A)\mathbf{P}(B^c)\,,$$

where the middle equality is due to the assumed independence of A and B. The proof for all other choices of G and H is very similar.

More generally we define the *mutual* independence of events as follows.

Definition 1.4.2. *Events* $A_i \in \mathcal{F}$ *are* $\mathbf{P}$-mutually independent *if for any* $L < \infty$ *and distinct indices* $i_1, i_2, \ldots, i_L$,

$$\mathbf{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_L}) = \prod_{k=1}^{L} \mathbf{P}(A_{i_k}).$$

We next generalize the definition of mutual independence to σ-algebras, random variables and beyond. This definition applies to the mutual independence of both finite and infinite number of such objects.

Definition 1.4.3. *We say that the collections of events* $\mathcal{A}_\alpha \subseteq \mathcal{F}$ *with* $\alpha \in \mathcal{I}$ *(possibly an infinite index set) are* $\mathbf{P}$*-mutually independent if for any* $L < \infty$ *and distinct* $\alpha_1, \alpha_2, \ldots, \alpha_L \in \mathcal{I}$,

$$\mathbf{P}(A_1 \cap A_2 \cap \cdots \cap A_L) = \prod_{k=1}^{L} \mathbf{P}(A_k), \qquad \forall A_k \in \mathcal{A}_{\alpha_k},\ k = 1, \ldots, L\,.$$

We say that random variables X_α, $\alpha \in \mathcal{I}$ *are* $\mathbf{P}$*-mutually independent if the* σ*-algebras* $\sigma(X_\alpha)$, $\alpha \in \mathcal{I}$ *are* $\mathbf{P}$*-mutually independent.*

When the probability measure $\mathbf{P}$ *in consideration is clear from the context, we say that random variables, or collections of events, are mutually independent.*

Our next theorem gives a sufficient condition for the mutual independence of a collection of σ-algebras which as we later show, greatly simplifies the task of checking independence.

Theorem 1.4.4. *Suppose* $\mathcal{G}_i = \sigma(\mathcal{A}_i) \subseteq \mathcal{F}$ *for* $i = 1, 2, \cdots, n$ *where* $\mathcal{A}_i$ *are* π*-systems. Then, a sufficient condition for the mutual independence of* $\mathcal{G}_i$ *is that* $\mathcal{A}_i$, $i = 1, \ldots, n$ *are mutually independent.*

Proof. Let $H = A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_L}$, where $i_1, i_2, \ldots, i_L$ are distinct elements from $\{1, 2, \ldots, n-1\}$ and $A_i \in \mathcal{A}_i$ for $i = 1, \ldots, n-1$. Consider the two finite measures $\mu_1(A) = \mathbf{P}(A \cap H)$ and $\mu_2(A) = \mathbf{P}(H)\mathbf{P}(A)$ on the measurable space $(\Omega, \mathcal{G}_n)$. Note that

$$\mu_1(\Omega) = \mathbf{P}(\Omega \cap H) = \mathbf{P}(H) = \mathbf{P}(H)\mathbf{P}(\Omega) = \mu_2(\Omega)\,.$$

If $A \in \mathcal{A}_n$, then by the mutual independence of $\mathcal{A}_i$, $i = 1, \ldots, n$, it follows that

$$\mu_1(A) = \mathbf{P}(A_{i_1} \cap A_{i_2} \cap A_{i_3} \cap \cdots \cap A_{i_L} \cap A) = (\prod_{k=1}^{L} \mathbf{P}(A_{i_k}))\mathbf{P}(A)$$
$$= \mathbf{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_L})\mathbf{P}(A) = \mu_2(A) \ .$$

Since the finite measures $\mu_1(\cdot)$ and $\mu_2(\cdot)$ agree on the π-system $\mathcal{A}_n$ and on Ω, it follows that $\mu_1 = \mu_2$ on $\mathcal{G}_n = \sigma(\mathcal{A}_n)$ (see Proposition 1.1.39). That is, $\mathbf{P}(G \cap H) = \mathbf{P}(G)\mathbf{P}(H)$ for any $G \in \mathcal{G}_n$.

Since this applies for arbitrary $A_i \in \mathcal{A}_i$, $i = 1, \ldots, n-1$, in view of Definition 1.4.3 we have just proved that if $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_n$ are mutually independent, then $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{G}_n$ are mutually independent.

Applying the latter relation for $\mathcal{G}_n, \mathcal{A}_1, \ldots, \mathcal{A}_{n-1}$ (which are mutually independent since Definition 1.4.3 is invariant to a permutation of the order of the collections) we get that $\mathcal{G}_n, \mathcal{A}_1, \ldots, \mathcal{A}_{n-2}, \mathcal{G}_{n-1}$ are mutually independent. After n such iterations we have the stated result. □

Because the mutual independence of the collections of events $\mathcal{A}_\alpha$, $\alpha \in \mathcal{I}$ amounts to the mutual independence of any finite number of these collections, we have the immediate consequence:

Corollary 1.4.5. *If π-systems of events $\mathcal{A}_\alpha$, $\alpha \in \mathcal{I}$, are mutually independent, then $\sigma(\mathcal{A}_\alpha)$, $\alpha \in \mathcal{I}$, are also mutually independent.*

Another immediate consequence deals with the closure of mutual independence under projections.

Corollary 1.4.6. *If the π-systems of events $\mathcal{H}_{\alpha,\beta}$, $(\alpha, \beta) \in \mathcal{J}$ are mutually independent, then the σ-algebras $\mathcal{G}_\alpha = \sigma\left(\cup_\beta \mathcal{H}_{\alpha,\beta}\right)$, are also mutually independent.*

Proof. Let $\mathcal{A}_\alpha$ be the collection of sets of the form $A = \cap_{j=1}^m H_j$ where $H_j \in \mathcal{H}_{\alpha,\beta_j}$ for some $m < \infty$ and distinct $\beta_1, \ldots, \beta_m$. Since $\mathcal{H}_{\alpha,\beta}$ are π-systems, it follows that so is $\mathcal{A}_\alpha$ for each α. Since a finite intersection of sets $A_k \in \mathcal{A}_{\alpha_k}$, $k = 1, \ldots, L$ is merely a finite intersection of sets from distinct collections $\mathcal{H}_{\alpha_k,\beta_j(k)}$, the assumed mutual independence of $\mathcal{H}_{\alpha,\beta}$ implies the mutual independence of $\mathcal{A}_\alpha$. By Corollary 1.4.5, this in turn implies the mutual independence of $\sigma(\mathcal{A}_\alpha)$. To complete the proof, simply note that for any β, each $H \in \mathcal{H}_{\alpha,\beta}$ is also an element of $\mathcal{A}_\alpha$, implying that $\mathcal{G}_\alpha \subseteq \sigma(\mathcal{A}_\alpha)$. □

Relying on the preceding corollary you can now establish the following characterization of independence (which is key to proving Kolmogorov's 0-1 law).

Exercise 1.4.7. *Show that if for each $n \geq 1$ the σ-algebras $\mathcal{F}_n^{\mathbf{X}} = \sigma(X_1, \ldots, X_n)$ and $\sigma(X_{n+1})$ are $\mathbf{P}$-mutually independent then the random variables $X_1, X_2, X_3, \ldots$ are $\mathbf{P}$-mutually independent. Conversely, show that if $X_1, X_2, X_3, \ldots$ are independent, then for each $n \geq 1$ the σ-algebras $\mathcal{F}_n^{\mathbf{X}}$ and $\mathcal{T}_n^{\mathbf{X}} = \sigma(X_r, r > n)$ are independent.*

It is easy to check that a $\mathbf{P}$-trivial σ-algebra $\mathcal{H}$ is $\mathbf{P}$-independent of any other σ-algebra $\mathcal{G} \subseteq \mathcal{F}$. Conversely, as we show next, independence is a great tool for proving that a σ-algebra is $\mathbf{P}$-trivial.

Lemma 1.4.8. *If each of the σ-algebras $\mathcal{G}_k \subseteq \mathcal{G}_{k+1}$ is $\mathbf{P}$-independent of a σ-algebra $\mathcal{H} \subseteq \sigma(\bigcup_{k \geq 1} \mathcal{G}_k)$ then $\mathcal{H}$ is $\mathbf{P}$-trivial.*

Remark. In particular, if $\mathcal{H}$ is $\mathbf{P}$-independent of itself, then $\mathcal{H}$ is $\mathbf{P}$-trivial.

PROOF. Since $\mathcal{G}_k \subseteq \mathcal{G}_{k+1}$ for all k and $\mathcal{G}_k$ are σ-algebras, it follows that $\mathcal{A} = \bigcup_{k\geq 1} \mathcal{G}_k$ is a π-system. The assumed $\mathbf{P}$-independence of $\mathcal{H}$ and $\mathcal{G}_k$ for each k yields the $\mathbf{P}$-independence of $\mathcal{H}$ and $\mathcal{A}$. Thus, by Theorem 1.4.4 we have that $\mathcal{H}$ and $\sigma(\mathcal{A})$ are $\mathbf{P}$-independent. Since $\mathcal{H} \subseteq \sigma(\mathcal{A})$ it follows that in particular $\mathbf{P}(H) = \mathbf{P}(H \cap H) = \mathbf{P}(H)\,\mathbf{P}(H)$ for each $H \in \mathcal{H}$. So, necessarily $\mathbf{P}(H) \in \{0, 1\}$ for all $H \in \mathcal{H}$. That is, $\mathcal{H}$ is $\mathbf{P}$-trivial. $\square$

We next define the tail σ-algebra of a stochastic process.

Definition 1.4.9. *For a stochastic process $\{X_k\}$ we set $\mathcal{T}_n^{\mathbf{X}} = \sigma(X_r, r > n)$ and call $\mathcal{T}^{\mathbf{X}} = \cap_n \mathcal{T}_n^{\mathbf{X}}$ the* tail σ-algebra *of the process $\{X_k\}$.*

As we next see, the $\mathbf{P}$-triviality of the tail σ-algebra of independent random variables is an immediate consequence of Lemma 1.4.8. This result, due to Kolmogorov, is just one of the many *0-1 laws* that exist in probability theory.

Corollary 1.4.10 (KOLMOGOROV'S 0-1 LAW). *If $\{X_k\}$ are $\mathbf{P}$-mutually independent then the corresponding tail σ-algebra $\mathcal{T}^{\mathbf{X}}$ is $\mathbf{P}$-trivial.*

PROOF. Note that $\mathcal{F}_k^{\mathbf{X}} \subseteq \mathcal{F}_{k+1}^{\mathbf{X}}$ and $\mathcal{T}^{\mathbf{X}} \subseteq \mathcal{F}^{\mathbf{X}} = \sigma(X_k, k \geq 1) = \sigma(\bigcup_{k\geq 1} \mathcal{F}_k^{\mathbf{X}})$ (see Exercise 1.2.14 for the latter identity). Further, recall Exercise 1.4.7 that for any $n \geq 1$, the σ-algebras $\mathcal{T}_n^{\mathbf{X}}$ and $\mathcal{F}_n^{\mathbf{X}}$ are $\mathbf{P}$-mutually independent. Hence, each of the σ-algebras $\mathcal{F}_k^{\mathbf{X}}$ is also $\mathbf{P}$-mutually independent of the tail σ-algebra $\mathcal{T}^{\mathbf{X}}$, which by Lemma 1.4.8 is thus $\mathbf{P}$-trivial. $\square$

Out of Corollary 1.4.6 we deduce that functions of disjoint collections of mutually independent random variables are mutually independent.

Corollary 1.4.11. *If R.V. $X_{k,j}$, $1 \leq k \leq m$, $1 \leq j \leq l(k)$ are mutually independent and $f_k : \mathbb{R}^{l(k)} \mapsto \mathbb{R}$ are Borel functions, then $Y_k = f_k(X_{k,1}, \ldots, X_{k,l(k)})$ are mutually independent random variables for $k = 1, \ldots, m$.*

PROOF. We apply Corollary 1.4.6 for the index set $\mathcal{J} = \{(k,j) : 1 \leq k \leq m, 1 \leq j \leq l(k)\}$, and mutually independent π-systems $\mathcal{H}_{k,j} = \sigma(X_{k,j})$, to deduce the mutual independence of $\mathcal{G}_k = \sigma(\cup_j \mathcal{H}_{k,j})$. Recall that $\mathcal{G}_k = \sigma(X_{k,j}, 1 \leq j \leq l(k))$ and $\sigma(Y_k) \subseteq \mathcal{G}_k$ (see Definition 1.2.12 and Exercise 1.2.33). We complete the proof by noting that Y_k are mutually independent if and only if $\sigma(Y_k)$ are mutually independent. $\square$

Our next result is an application of Theorem 1.4.4 to the independence of random variables.

Corollary 1.4.12. *Real-valued random variables $X_1, X_2, \ldots, X_m$ on the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$ are mutually independent if and only if*

$$\mathbf{P}(X_1 \leq x_1, \ldots, X_m \leq x_m) = \prod_{i=1}^{m} \mathbf{P}(X_i \leq x_i)\,,\ \forall x_1, \ldots, x_m \in \mathbb{R}. \tag{1.4.1}$$

PROOF. Let $\mathcal{A}_i$ denote the collection of subsets of Ω of the form $X_i^{-1}((-\infty, b])$ for $b \in \mathbb{R}$. Recall that $\mathcal{A}_i$ generates $\sigma(X_i)$ (see Exercise 1.2.11), whereas (1.4.1) states that the π-systems $\mathcal{A}_i$ are mutually independent (by continuity from below of $\mathbf{P}$, taking $x_i \uparrow \infty$ for $i \neq i_1, i \neq i_2, \ldots, i \neq i_L$, has the same effect as taking a subset of distinct indices $i_1, \ldots, i_L$ from $\{1, \ldots, m\}$). So, just apply Theorem 1.4.4 to conclude the proof. $\square$

The condition (1.4.1) for mutual independence of R.V.-s is further simplified when these variables are either discrete valued, or having a density.

Exercise 1.4.13. *Suppose $(X_1, \ldots, X_m)$ are random variables and $(\mathbb{S}_1, \ldots, \mathbb{S}_m)$ are countable sets such that $\mathbf{P}(X_i \in \mathbb{S}_i) = 1$ for $i = 1, \ldots, m$. Show that if*

$$\mathbf{P}(X_1 = x_1, \ldots, X_m = x_m) = \prod_{i=1}^{m} \mathbf{P}(X_i = x_i)$$

whenever $x_i \in \mathbb{S}_i$, $i = 1, \ldots, m$, then $X_1, \ldots, X_m$ are mutually independent.

Exercise 1.4.14. *Suppose the random vector $\underline{X} = (X_1, \ldots, X_m)$ has a* joint probability density function *$f_{\underline{X}}(\underline{x}) = g_1(x_1) \cdots g_m(x_m)$. That is,*

$$\mathbf{P}((X_1, \ldots, X_m) \in A) = \int_A g_1(x_1) \cdots g_m(x_m) dx_1 \ldots dx_m \,, \qquad \forall A \in \mathcal{B}_{\mathbb{R}^m} \,,$$

where g_i are non-negative, Lebesgue integrable functions. Show that then $X_1, \ldots, X_m$ are mutually independent.

Beware that pairwise independence (of each pair A_k, A_j for $k \neq j$), does not imply mutual independence of *all* the events in question and the same applies to three or more random variables. Here is an illustrating example.

Exercise 1.4.15. *Consider the sample space $\Omega = \{0, 1, 2\}^2$ with probability measure on $(\Omega, 2^\Omega)$ that assigns equal probability (i.e. $1/9$) to each possible value of $\omega = (\omega_1, \omega_2) \in \Omega$. Then, $X(\omega) = \omega_1$ and $Y(\omega) = \omega_2$ are independent R.V. each taking the values $\{0, 1, 2\}$ with equal (i.e. $1/3$) probability. Define $Z_0 = X$, $Z_1 = (X + Y)\text{mod}3$ and $Z_2 = (X + 2Y)\text{mod}3$.*

(a) *Show that Z_0 is independent of Z_1, Z_0 is independent of Z_2, Z_1 is independent of Z_2, but if we know the value of Z_0 and Z_1, then we also know Z_2.*

(b) *Construct four $\{-1, 1\}$-valued random variables such that any three of them are independent but all four are not.*
Hint: *Consider products of independent random variables.*

Here is a somewhat counter intuitive example about tail σ-algebras, followed by an elaboration on the theme of Corollary 1.4.11.

Exercise 1.4.16. *Let $\sigma(\mathcal{A}, \mathcal{A}')$ denote the smallest σ-algebra $\mathcal{G}$ such that any function measurable on $\mathcal{A}$ or on $\mathcal{A}'$ is also measurable on $\mathcal{G}$. Let $W_0, W_1, W_2, \ldots$ be independent random variables with $\mathbf{P}(W_n = +1) = \mathbf{P}(W_n = -1) = 1/2$ for all n. For each $n \geq 1$, define $X_n := W_0 W_1 \ldots W_n$.*
(a) Prove that the variables $X_1, X_2, \ldots$ are independent.
(b) Show that $\mathcal{S} = \sigma(\mathcal{T}_0^{\mathbf{W}}, \mathcal{T}^{\mathbf{X}})$ is a strict *subset of the σ-algebra $\mathcal{F} = \cap_n \sigma(\mathcal{T}_0^{\mathbf{W}}, \mathcal{T}_n^{\mathbf{X}})$.*
Hint: *Show that $W_0 \in m\mathcal{F}$ is independent of $\mathcal{S}$.*

Exercise 1.4.17. *Consider random variables $(X_{i,j}, 1 \leq i, j \leq n)$ on the same probability space. Suppose that the σ-algebras $\mathcal{R}_1, \ldots, \mathcal{R}_n$ are $\mathbf{P}$-mutually independent, where $\mathcal{R}_i = \sigma(X_{i,j}, 1 \leq j \leq n)$ for $i = 1, \ldots, n$. Suppose further that the σ-algebras $\mathcal{C}_1, \ldots, \mathcal{C}_n$ are $\mathbf{P}$-mutually independent, where $\mathcal{C}_j = \sigma(X_{i,j}, 1 \leq i \leq n)$. Prove that the random variables $(X_{i,j}, 1 \leq i, j \leq n)$ must then be $\mathbf{P}$-mutually independent.*

We conclude this subsection with an application in number theory.

Exercise 1.4.18. *Recall Euler's zeta-function which for real $s > 1$ is given by $\zeta(s) = \sum_{k=1}^{\infty} k^{-s}$. Fixing such s, let X and Y be independent random variables with $\mathbf{P}(X = k) = \mathbf{P}(Y = k) = k^{-s}/\zeta(s)$ for $k = 1, 2, \ldots$.*

(a) *Show that the events $D_p = \{X \text{ is divisible by } p\}$, with p a prime number, are $\mathbf{P}$-mutually independent.*
(b) *By considering the event $\{X = 1\}$, provide a* probabilistic *explanation of Euler's formula $1/\zeta(s) = \prod_p (1 - 1/p^s)$.*
(c) *Show that the probability that no perfect square other than 1 divides X is precisely $1/\zeta(2s)$.*
(d) *Show that $\mathbf{P}(G = k) = k^{-2s}/\zeta(2s)$, where G is the greatest common divisor of X and Y.*

1.4.2. Product measures and Kolmogorov's theorem. Recall Example 1.1.20 that given two measurable spaces $(\Omega_1, \mathcal{F}_1)$ and $(\Omega_2, \mathcal{F}_2)$ the product (measurable) space $(\Omega, \mathcal{F})$ consists of $\Omega = \Omega_1 \times \Omega_2$ and $\mathcal{F} = \mathcal{F}_1 \times \mathcal{F}_2$, which is the same as $\mathcal{F} = \sigma(\mathcal{A})$ for

$$\mathcal{A} = \Big\{ \biguplus_{j=1}^{m} A_j \times B_j : A_j \in \mathcal{F}_1, B_j \in \mathcal{F}_2, m < \infty \Big\},$$

where throughout, $\biguplus$ denotes the union of disjoint subsets of Ω.

We now construct product measures on such product spaces, first for two, then for finitely many, probability (or even σ-finite) measures. As we show thereafter, these product measures are associated with the joint law of independent R.V.-s.

Theorem 1.4.19. *Given two σ-finite measures ν_i on $(\Omega_i, \mathcal{F}_i)$, $i = 1, 2$, there exists a unique σ-finite measure μ_2 on the product space $(\Omega, \mathcal{F})$ such that*

$$\mu_2(\biguplus_{j=1}^{m} A_j \times B_j) = \sum_{j=1}^{m} \nu_1(A_j)\nu_2(B_j), \quad \forall A_j \in \mathcal{F}_1, B_j \in \mathcal{F}_2, m < \infty .$$

We denote $\mu_2 = \nu_1 \times \nu_2$ and call it the product *of the measures ν_1 and ν_2.*

Proof. By Carathéodory's extension theorem, it suffices to show that $\mathcal{A}$ is an algebra on which μ_2 is countably additive (see Theorem 1.1.30 for the case of finite measures). To this end, note that $\Omega = \Omega_1 \times \Omega_2 \in \mathcal{A}$. Further, $\mathcal{A}$ is closed under intersections, since

$$\begin{aligned}(\biguplus_{j=1}^{m} A_j \times B_j) \bigcap (\biguplus_{i=1}^{n} C_i \times D_i) &= \biguplus_{i,j} [(A_j \times B_j) \cap (C_i \times D_i)] \\ &= \biguplus_{i,j} (A_j \cap C_i) \times (B_j \cap D_i) .\end{aligned}$$

It is also closed under complementation, for

$$(\biguplus_{j=1}^{m} A_j \times B_j)^c = \bigcap_{j=1}^{m} [(A_j^c \times B_j) \cup (A_j \times B_j^c) \cup (A_j^c \times B_j^c)] .$$

By DeMorgan's law, $\mathcal{A}$ is an algebra.

Note that countable unions of disjoint elements of $\mathcal{A}$ are also countable unions of disjoint elements of the collection $\mathcal{R} = \{A \times B : A \in \mathcal{F}_1, B \in \mathcal{F}_2\}$ of *measurable* *rectangles*. Hence, if we show that

$$\sum_{j=1}^{m} \nu_1(A_j)\nu_2(B_j) = \sum_i \nu_1(C_i)\nu_2(D_i)\,, \tag{1.4.2}$$

whenever $\biguplus_{j=1}^m A_j \times B_j = \biguplus_i (C_i \times D_i)$ for some $m < \infty$, $A_j, C_i \in \mathcal{F}_1$ and $B_j, D_i \in \mathcal{F}_2$, then we deduce that the value of $\mu_2(E)$ is independent of the representation we choose for $E \in \mathcal{A}$ in terms of measurable rectangles, and further that μ_2 is countably additive on $\mathcal{A}$. To this end, note that the preceding set identity amounts to

$$\sum_{j=1}^{m} I_{A_j}(x) I_{B_j}(y) = \sum_i I_{C_i}(x) I_{D_i}(y) \qquad \forall x \in \Omega_1, y \in \Omega_2\,.$$

Hence, fixing $x \in \Omega_1$, we have that $\varphi(y) = \sum_{j=1}^m I_{A_j}(x) I_{B_j}(y) \in \text{SF}_+$ is the monotone increasing limit of $\psi_n(y) = \sum_{i=1}^n I_{C_i}(x) I_{D_i}(y) \in \text{SF}_+$ as $n \to \infty$. Thus, by linearity of the integral with respect to ν_2 and monotone convergence,

$$g(x) := \sum_{j=1}^{m} \nu_2(B_j) I_{A_j}(x) = \nu_2(\varphi) = \lim_{n\to\infty} \nu_2(\psi_n) = \lim_{n\to\infty} \sum_{i=1}^{n} I_{C_i}(x)\nu_2(D_i)\,.$$

We deduce that the non-negative $g(x) \in m\mathcal{F}_1$ is the monotone increasing limit of the non-negative measurable functions $h_n(x) = \sum_{i=1}^n \nu_2(D_i) I_{C_i}(x)$. Hence, by the same reasoning,

$$\sum_{j=1}^{m} \nu_2(B_j)\nu_1(A_j) = \nu_1(g) = \lim_{n\to\infty} \nu_1(h_n) = \sum_i \nu_2(D_i)\nu_1(C_i)\,,$$

proving (1.4.2) and the theorem. □

It follows from Theorem 1.4.19 by induction on n that given any finite collection of σ-finite measure spaces $(\Omega_i, \mathcal{F}_i, \nu_i)$, $i = 1, \ldots, n$, there exists a unique *product measure* $\mu_n = \nu_1 \times \cdots \times \nu_n$ on the product space $(\Omega, \mathcal{F})$ (i.e., $\Omega = \Omega_1 \times \cdots \times \Omega_n$ and $\mathcal{F} = \sigma(A_1 \times \cdots \times A_n; A_i \in \mathcal{F}_i, i = 1, \ldots, n)$), such that

$$\mu_n(A_1 \times \cdots \times A_n) = \prod_{i=1}^{n} \nu_i(A_i) \qquad \forall A_i \in \mathcal{F}_i, \;\; i = 1, \ldots, n. \tag{1.4.3}$$

Remark 1.4.20. A notable special case of this construction is when $\Omega_i = \mathbb{R}$ with the Borel σ-algebra and Lebesgue measure λ of Section 1.1.3. The product space is then $\mathbb{R}^n$ with its Borel σ-algebra and the product measure is λ^n, the Lebesgue measure on $\mathbb{R}^n$.

The notion of the *law* $\mathcal{P}_X$ of a real-valued random variable X as in Definition 1.2.34, naturally extends to the *joint law* $\mathcal{P}_{\underline{X}}$ of a random vector $\underline{X} = (X_1, \ldots, X_n)$ which is the probability measure $\mathcal{P}_{\underline{X}} = \mathbf{P} \circ \underline{X}^{-1}$ on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$.

We next characterize the joint law of independent random variables $X_1, \ldots, X_n$ as the product of the laws of X_i for $i = 1, \ldots, n$.

Proposition 1.4.21. *Random variables $X_1, \ldots, X_n$ on the same probability space, having laws $\nu_i = \mathcal{P}_{X_i}$, are mutually independent if and only if their joint law is $\mu_n = \nu_1 \times \cdots \times \nu_n$.*

Proof. By Definition 1.4.3 and the identity (1.4.3), if $X_1, \ldots, X_n$ are mutually independent then for $B_i \in \mathcal{B}$,

$$\begin{aligned}\mathcal{P}_{\underline{X}}(B_1 \times \cdots \times B_n) &= \mathbf{P}(X_1 \in B_1, \ldots, X_n \in B_n) \\ &= \prod_{i=1}^n \mathbf{P}(X_i \in B_i) = \prod_{i=1}^n \nu_i(B_i) = \nu_1 \times \cdots \times \nu_n(B_1 \times \cdots \times B_n)\,.\end{aligned}$$

This shows that the law of $(X_1, \ldots, X_n)$ and the product measure μ_n agree on the collection of all measurable rectangles $B_1 \times \cdots \times B_n$, a π-system that generates $\mathcal{B}_{\mathbb{R}^n}$ (see Exercise 1.1.21). Consequently, these two probability measures agree on $\mathcal{B}_{\mathbb{R}^n}$ (c.f. Proposition 1.1.39).

Conversely, if $\mathcal{P}_{\underline{X}} = \nu_1 \times \cdots \times \nu_n$, then by same reasoning, for Borel sets B_i,

$$\begin{aligned}\mathbf{P}(\bigcap_{i=1}^n \{\omega : X_i(\omega) \in B_i\}) &= \mathcal{P}_{\underline{X}}(B_1 \times \cdots \times B_n) = \nu_1 \times \cdots \times \nu_n(B_1 \times \cdots \times B_n) \\ &= \prod_{i=1}^n \nu_i(B_i) = \prod_{i=1}^n \mathbf{P}(\{\omega : X_i(\omega) \in B_i\})\,,\end{aligned}$$

which amounts to the mutual independence of $X_1, \ldots, X_n$. $\square$

We wish to extend the construction of product measures to that of an infinite collection of independent random variables. To this end, let $\mathbf{N} = \{1, 2, \ldots\}$ denote the set of natural numbers and $\mathbb{R}^{\mathbf{N}} = \{\mathbf{x} = (x_1, x_2, \ldots) : x_i \in \mathbb{R}\}$ denote the collection of all infinite sequences of real numbers. We equip $\mathbb{R}^{\mathbf{N}}$ with the product σ-algebra $\mathcal{B}_c = \sigma(\mathcal{R})$ generated by the collection $\mathcal{R}$ of all finite dimensional measurable rectangles (also called *cylinder sets*), that is sets of the form $\{\mathbf{x} : x_1 \in B_1, \ldots, x_n \in B_n\}$, where $B_i \in \mathcal{B}$, $i = 1, \ldots, n \in \mathbf{N}$ (e.g. see Example 1.1.19).

Kolmogorov's extension theorem provides the existence of a unique probability measure $\mathbf{P}$ on $(\mathbb{R}^{\mathbf{N}}, \mathcal{B}_c)$ whose projections coincide with a given consistent sequence of probability measures μ_n on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$.

Theorem 1.4.22 (Kolmogorov's extension theorem)**.** *Suppose we are given probability measures μ_n on $(\mathbb{R}^n, \mathcal{B}_{\mathbb{R}^n})$ that are* consistent*, that is,*

$$\mu_{n+1}(B_1 \times \cdots \times B_n \times \mathbb{R}) = \mu_n(B_1 \times \cdots \times B_n) \qquad \forall B_i \in \mathcal{B}, \quad i = 1, \ldots, n < \infty$$

Then, there is a unique probability measure $\mathbf{P}$ on $(\mathbb{R}^{\mathbf{N}}, \mathcal{B}_c)$ such that

$$\mathbf{P}(\{\omega : \omega_i \in B_i, i = 1, \ldots, n\}) = \mu_n(B_1 \times \cdots \times B_n)\ \forall B_i \in \mathcal{B},\ i \le n < \infty \tag{1.4.4}$$

Proof. (sketch only) We take a similar approach as in the proof of Theorem 1.4.19. That is, we use (1.4.4) to define the non-negative set function $\mathbf{P}_0$ on the collection $\mathcal{R}$ of all finite dimensional measurable rectangles, where by the consistency of $\{\mu_n\}$ the value of $\mathbf{P}_0$ is independent of the specific representation chosen for a set in $\mathcal{R}$. Then, we extend $\mathbf{P}_0$ to a finitely additive set function on the algebra

$$\mathcal{A} = \Big\{ \biguplus_{j=1}^m E_j : E_j \in \mathcal{R}, m < \infty \Big\}\,,$$

in the same linear manner we used when proving Theorem 1.4.19. Since $\mathcal{A}$ generates $\mathcal{B}_c$ and $\mathbf{P}_0(\mathbb{R}^{\mathbf{N}}) = \mu_n(\mathbb{R}^n) = 1$, by Carathéodory's extension theorem it suffices to check that $\mathbf{P}_0$ is countably additive on $\mathcal{A}$. The countable additivity of $\mathbf{P}_0$ is verified by the method we already employed when dealing with Lebesgue's measure. That is, by the remark after Lemma 1.1.31, it suffices to prove that $\mathbf{P}_0(H_n) \downarrow 0$ whenever $H_n \in \mathcal{A}$ and $H_n \downarrow \emptyset$. The proof by contradiction of the latter, adapting the argument of Lemma 1.1.31, is based on approximating each $H \in \mathcal{A}$ by a finite union $J_k \subseteq H$ of *compact* rectangles, such that $\mathbf{P}_0(H \setminus J_k) \to 0$ as $k \to \infty$. This is done for example in [**Bil95**, Page 490]. $\square$

Example 1.4.23. *To systematically construct an infinite sequence of independent random variables $\{X_i\}$ of prescribed laws $\mathcal{P}_{X_i} = \nu_i$, we apply Kolmogorov's extension theorem for the product measures $\mu_n = \nu_1 \times \cdots \times \nu_n$ constructed following Theorem 1.4.19 (where it is by definition that the sequence μ_n is consistent). Alternatively, for infinite product measures one can take arbitrary probability spaces $(\Omega_i, \mathcal{F}_i, \nu_i)$ and directly show by contradiction that $\mathbf{P}_0(H_n) \downarrow 0$ whenever $H_n \in \mathcal{A}$ and $H_n \downarrow \emptyset$ (for more details, see* [**Str93**, Exercise 1.1.14]*).*

Remark. As we shall find in Sections 6.1 and 7.1, Kolmogorov's extension theorem is the key to the study of *stochastic processes*, where it relates the law of the process to its finite dimensional distributions. Certain properties of $\mathbb{R}$ are key to the proof of Kolmogorov's extension theorem which indeed is false if $(\mathbb{R}, \mathcal{B})$ is replaced with an arbitrary measurable space $(\mathbb{S}, \mathcal{S})$ (see the discussions in [**Dur10**, Subsection 2.1.4] and [**Dud89**, notes for Section 12.1]). Nevertheless, as you show next, the conclusion of this theorem applies for any $\mathcal{B}$-*isomorphic* measurable space $(\mathbb{S}, \mathcal{S})$.

Definition 1.4.24. *Two measurable spaces $(\mathbb{S}, \mathcal{S})$ and $(\mathbb{T}, \mathcal{T})$ are isomorphic if there exists a one to one and onto measurable mapping between them whose inverse is also a measurable mapping. A measurable space $(\mathbb{S}, \mathcal{S})$ is $\mathcal{B}$-isomorphic if it is isomorphic to a Borel subset $\mathbb{T}$ of $\mathbb{R}$ equipped with the* induced Borel σ-algebra *$\mathcal{T} = \{B \cap \mathbb{T} : B \in \mathcal{B}\}$.*

Here is our generalized version of Kolmogorov's extension theorem.

Corollary 1.4.25. *Given a measurable space $(\mathbb{S}, \mathcal{S})$ let $\mathbb{S}^{\mathbf{N}}$ denote the collection of all infinite sequences of elements in $\mathbb{S}$ equipped the product σ-algebra $\mathcal{S}_c$ generated by the collection of all cylinder sets of the form $\{\mathbf{s} : s_1 \in A_1, \ldots, s_n \in A_n\}$, where $A_i \in \mathcal{S}$ for $i = 1, \ldots, n$. If $(\mathbb{S}, \mathcal{S})$ is $\mathcal{B}$-isomorphic then for any consistent sequence of probability measures ν_n on $(\mathbb{S}^n, \mathcal{S}^n)$ (that is, $\nu_{n+1}(A_1 \times \cdots \times A_n \times \mathbb{S}) = \nu_n(A_1 \times \cdots \times A_n)$ for all n and $A_i \in \mathcal{S}$), there exists a unique probability measure $\mathbf{Q}$ on $(\mathbb{S}^{\mathbf{N}}, \mathcal{S}_c)$ such that for all n and $A_i \in \mathcal{S}$,*

$$\mathbf{Q}(\{\mathbf{s} : s_i \in A_i, i = 1, \ldots, n\}) = \nu_n(A_1 \times \cdots \times A_n) \,. \tag{1.4.5}$$

Next comes a guided proof of Corollary 1.4.25 out of Theorem 1.4.22.

Exercise 1.4.26.

(a) *Verify that our proof of Theorem 1.4.22 applies in case $(\mathbb{R}, \mathcal{B})$ is replaced by $\mathbb{T} \in \mathcal{B}$ equipped with the induced Borel σ-algebra $\mathcal{T}$ (with $\mathbb{R}^{\mathbf{N}}$ and $\mathcal{B}_c$ replaced by $\mathbb{T}^{\mathbf{N}}$ and $\mathcal{T}_c$, respectively).*

(b) *Fixing such $(\mathbb{T}, \mathcal{T})$ and $(\mathbb{S}, \mathcal{S})$ isomorphic to it, let $g : \mathbb{S} \mapsto \mathbb{T}$ be one to one and onto such that both g and g^{-1} are measurable. Check that the one to one and onto mappings $g_n(\mathbf{s}) = (g(s_1), \ldots, g(s_n))$ are measurable and deduce that $\mu_n(B) = \nu_n(g_n^{-1}(B))$ are consistent probability measures on $(\mathbb{T}^n, \mathcal{T}^n)$.*

(c) *Consider the one to one and onto mapping $g_\infty(\mathbf{s}) = (g(s_1), \ldots, g(s_n), \ldots)$ from $\mathbb{S}^{\mathbf{N}}$ to $\mathbb{T}^{\mathbf{N}}$ and the unique probability measure $\mathbf{P}$ on $(\mathbb{T}^{\mathbf{N}}, \mathcal{T}_c)$ for* *which (1.4.4) holds. Verify that $\mathcal{S}_c$ is contained in the σ-algebra of subsets A of $\mathbb{S}^{\mathbf{N}}$ for which $g_\infty(A)$ is in $\mathcal{T}_c$ and deduce that $\mathbf{Q}(A) = \mathbf{P}(g_\infty(A))$ is a probability measure on $(\mathbb{S}^{\mathbf{N}}, \mathcal{S}_c)$.*

(d) *Conclude your proof of Corollary 1.4.25 by showing that this $\mathbf{Q}$ is the unique probability measure for which (1.4.5) holds.*

Remark. Recall that Carathéodory's extension theorem applies for any σ-finite measure. It follows that, by the same proof as in the preceding exercise, any consistent sequence of σ-finite measures ν_n uniquely determines a σ-finite measure $\mathbf{Q}$ on $(\mathbb{S}^{\mathbf{N}}, \mathcal{S}_c)$ for which (1.4.5) holds, a fact which we use in later parts of this text (for example, in the study of Markov chains in Section 6.1).

Our next proposition shows that in most applications one encounters $\mathcal{B}$-isomorphic measurable spaces (for which Kolmogorov's theorem applies).

Proposition 1.4.27. *If $\mathbb{S} \in \mathcal{B}_M$ for a complete separable metric space M and $\mathcal{S}$ is the restriction of $\mathcal{B}_M$ to $\mathbb{S}$ then $(\mathbb{S}, \mathcal{S})$ is $\mathcal{B}$-isomorphic.*

Remark. While we do not provide the proof of this proposition, we note in passing that it is an immediate consequence of [**Dud89**, Theorem 13.1.1].

1.4.3. Fubini's theorem and its application. Returning to $(\Omega, \mathcal{F}, \mu)$ which is the product of two σ-finite measure spaces, as in Theorem 1.4.19, we now prove that:

Theorem 1.4.28 (Fubini's theorem)**.** *Suppose $\mu = \mu_1 \times \mu_2$ is the product of the σ-finite measures μ_1 on $(\mathbb{X}, \mathfrak{X})$ and μ_2 on $(\mathbb{Y}, \mathcal{Y})$. If $h \in m\mathcal{F}$ for $\mathcal{F} = \mathfrak{X} \times \mathcal{Y}$ is such that $h \geq 0$ or $\int |h|\, d\mu < \infty$, then,*

$$\int_{\mathbb{X}\times\mathbb{Y}} h\, d\mu = \int_{\mathbb{X}} \left[\int_{\mathbb{Y}} h(x,y)\, d\mu_2(y) \right] d\mu_1(x) = \int_{\mathbb{Y}} \left[\int_{\mathbb{X}} h(x,y)\, d\mu_1(x) \right] d\mu_2(y) \tag{1.4.6}$$

Remark. The iterated integrals on the right side of (1.4.6) are finite and well defined whenever $\int |h| d\mu < \infty$. However, for $h \notin m\mathcal{F}_+$ the inner integrals might be well defined only in the almost everywhere sense.

Proof of Fubini's theorem. Clearly, it suffices to prove the first identity of (1.4.6), as the second immediately follows by exchanging the roles of the two measure spaces. We thus prove Fubini's theorem by showing that

$$y \mapsto h(x,y) \in m\mathcal{Y}, \quad \forall x \in \mathbb{X}, \tag{1.4.7}$$

$$x \mapsto f_h(x) := \int_{\mathbb{Y}} h(x,y)\, d\mu_2(y) \in m\mathfrak{X}, \tag{1.4.8}$$

so the double integral on the right side of (1.4.6) is well defined and

$$\int_{\mathbb{X}\times\mathbb{Y}} h\, d\mu = \int_{\mathbb{X}} f_h(x) d\mu_1(x)\,. \tag{1.4.9}$$

We do so in three steps, first proving (1.4.7)-(1.4.9) for finite measures and bounded h, proceeding to extend these results to non-negative h and σ-finite measures, and then showing that (1.4.6) holds whenever $h \in m\mathcal{F}$ and $\int |h| d\mu$ is finite.

Step 1. Let $\mathcal{H}$ denote the collection of bounded functions on $\mathbb{X}\times\mathbb{Y}$ for which (1.4.7)–(1.4.9) hold. Assuming that both $\mu_1(\mathbb{X})$ and $\mu_2(\mathbb{Y})$ are finite, we deduce that $\mathcal{H}$ contains all bounded $h \in m\mathcal{F}$ by verifying the assumptions of the monotone class theorem (i.e. Theorem 1.2.7) for $\mathcal{H}$ and the π-system $\mathcal{R} = \{A \times B : A \in \mathfrak{X}, B \in \mathcal{Y}\}$ of measurable rectangles (which generates $\mathcal{F}$).

To this end, note that if $h = I_E$ and $E = A\times B \in \mathcal{R}$, then either $h(x,\cdot) = I_B(\cdot)$ (in case $x \in A$), or $h(x,\cdot)$ is identically zero (when $x \notin A$). With $I_B \in m\mathcal{Y}$ we thus have (1.4.7) for any such h. Further, in this case the simple function $f_h(x) = \mu_2(B)I_A(x)$ on $(\mathbb{X},\mathfrak{X})$ is in $m\mathfrak{X}$ and

$$\int_{\mathbb{X}\times\mathbb{Y}} I_E d\mu = \mu_1 \times \mu_2(E) = \mu_2(B)\mu_1(A) = \int_{\mathbb{X}} f_h(x) d\mu_1(x) .$$

Consequently, $I_E \in \mathcal{H}$ for all $E \in \mathcal{R}$; in particular, the constant functions are in $\mathcal{H}$.

Next, with both $m\mathcal{Y}$ and $m\mathfrak{X}$ vector spaces over $\mathbb{R}$, by the linearity of $h \mapsto f_h$ over the vector space of bounded functions satisfying (1.4.7) and the linearity of $f_h \mapsto \mu_1(f_h)$ and $h \mapsto \mu(h)$ over the vector spaces of bounded measurable f_h and h, respectively, we deduce that $\mathcal{H}$ is also a vector space over $\mathbb{R}$.

Finally, if non-negative $h_n \in \mathcal{H}$ are such that $h_n \uparrow h$, then for each $x \in \mathbb{X}$ the mapping $y \mapsto h(x,y) = \sup_n h_n(x,y)$ is in $m\mathcal{Y}_+$ (by Theorem 1.2.22). Further, $f_{h_n} \in m\mathfrak{X}_+$ and by monotone convergence $f_{h_n} \uparrow f_h$ (for all $x \in \mathbb{X}$), so by the same reasoning $f_h \in m\mathfrak{X}_+$. Applying monotone convergence twice more, it thus follows that

$$\mu(h) = \sup_n \mu(h_n) = \sup_n \mu_1(f_{h_n}) = \mu_1(f_h) ,$$

so h satisfies (1.4.7)–(1.4.9). In particular, if h is bounded then also $h \in \mathcal{H}$.

Step 2. Suppose now that $h \in m\mathcal{F}_+$. If μ_1 and μ_2 are finite measures, then we have shown in Step 1 that (1.4.7)–(1.4.9) hold for the bounded non-negative functions $h_n = h \wedge n$. With $h_n \uparrow h$ we have further seen that (1.4.7)-(1.4.9) hold also for the possibly unbounded h. Further, the closure of (1.4.8) and (1.4.9) with respect to monotone increasing limits of non-negative functions has been shown by monotone convergence, and as such it extends to σ-finite measures μ_1 and μ_2. Turning now to σ-finite μ_1 and μ_2, recall that there exist $E_n = A_n \times B_n \in \mathcal{R}$ such that $A_n \uparrow \mathbb{X}$, $B_n \uparrow \mathbb{Y}$, $\mu_1(A_n) < \infty$ and $\mu_2(B_n) < \infty$. As h is the monotone increasing limit of $h_n = hI_{E_n} \in m\mathcal{F}_+$ it thus suffices to verify that for each n the non-negative $f_n(x) = \int_{\mathbb{Y}} h_n(x,y)d\mu_2(y)$ is measurable with $\mu(h_n) = \mu_1(f_n)$. Fixing n and simplifying our notations to $E = E_n$, $A = A_n$ and $B = B_n$, recall Corollary 1.3.57 that $\mu(h_n) = \mu_E(h_E)$ for the restrictions h_E and μ_E of h and μ to the measurable space $(E,\mathcal{F}_E)$. Also, as $E = A \times B$ we have that $\mathcal{F}_E = \mathfrak{X}_A \times \mathcal{Y}_B$ and $\mu_E = (\mu_1)_A \times (\mu_2)_B$ for the finite measures $(\mu_1)_A$ and $(\mu_2)_B$. Finally, as $f_n(x) = f_{h_E}(x) := \int_B h_E(x,y)d(\mu_2)_B(y)$ when $x \in A$ and zero otherwise, it follows that $\mu_1(f_n) = (\mu_1)_A(f_{h_E})$. We have thus reduced our problem (for h_n), to the case of finite measures $\mu_E = (\mu_1)_A \times (\mu_2)_B$ which we have already successfully resolved.

Step 3. Write $h \in m\mathcal{F}$ as $h = h_+ - h_-$, with $h_\pm \in m\mathcal{F}_+$. By Step 2 we know that $y \mapsto h_\pm(x,y) \in m\mathcal{Y}$ for each $x \in \mathbb{X}$, hence the same applies for $y \mapsto h(x,y)$. Let $\mathbb{X}_0$ denote the subset of $\mathbb{X}$ for which $\int_{\mathbb{Y}} |h(x,y)|d\mu_2(y) < \infty$. By linearity of the integral with respect to μ_2 we have that for all $x \in \mathbb{X}_0$

$$f_h(x) = f_{h_+}(x) - f_{h_-}(x) \tag{1.4.10}$$

is finite. By Step 2 we know that $f_{h_\pm} \in m\mathfrak{X}$, hence $\mathbb{X}_0 = \{x : f_{h_+}(x) + f_{h_-}(x) < \infty\}$ is in $\mathfrak{X}$. From Step 2 we further have that $\mu_1(f_{h_\pm}) = \mu(h_\pm)$ whereby our assumption that $\int |h|\, d\mu = \mu_1(f_{h_+} + f_{h_-}) < \infty$ implies that $\mu_1(\mathbb{X}_0^c) = 0$. Let $\widetilde{f}_h(x) = f_{h_+}(x) - f_{h_-}(x)$ on $\mathbb{X}_0$ and $\widetilde{f}_h(x) = 0$ for all $x \notin \mathbb{X}_0$. Clearly, $\widetilde{f}_h \in m\mathfrak{X}$ is μ_1-almost-everywhere the same as the inner integral on the right side of (1.4.6). Moreover, in view of (1.4.10) and linearity of the integrals with respect to μ_1 and μ we deduce that

$$\mu(h) = \mu(h_+) - \mu(h_-) = \mu_1(f_{h_+}) - \mu_1(f_{h_-}) = \mu_1(\widetilde{f}_h)\,,$$

which is exactly the identity (1.4.6). □

Equipped with Fubini's theorem, we have the following simpler formula for the expectation of a Borel function h of two independent R.V.

Theorem 1.4.29. *Suppose that X and Y are independent random variables of laws $\mu_1 = \mathcal{P}_X$ and $\mu_2 = \mathcal{P}_Y$. If $h : \mathbb{R}^2 \mapsto \mathbb{R}$ is a Borel measurable function such that $h \geq 0$ or $\mathbf{E}|h(X,Y)| < \infty$, then,*

$$\mathbf{E}h(X,Y) = \int \Big[\int h(x,y)\, d\mu_1(x) \Big]\, d\mu_2(y) \tag{1.4.11}$$

In particular, for Borel functions $f, g : \mathbb{R} \mapsto \mathbb{R}$ such that $f, g \geq 0$ or $\mathbf{E}|f(X)| < \infty$ and $\mathbf{E}|g(Y)| < \infty$,

$$\mathbf{E}(f(X)g(Y)) = \mathbf{E}f(X)\,\mathbf{E}g(Y) \tag{1.4.12}$$

Proof. Subject to minor changes of notations, the proof of Theorem 1.3.61 applies to any $(\mathbb{S}, \mathcal{S})$-valued R.V. Considering this theorem for the random vector (X, Y) whose joint law is $\mu_1 \times \mu_2$ (c.f. Proposition 1.4.21), together with Fubini's theorem, we see that

$$\mathbf{E}h(X,Y) = \int_{\mathbb{R}^2} h(x,y)\, d(\mu_1 \times \mu_2)(x,y) = \int \Big[\int h(x,y)\, d\mu_1(x) \Big]\, d\mu_2(y)\ ,$$

which is (1.4.11). Take now $h(x, y) = f(x)g(y)$ for non-negative Borel functions $f(x)$ and $g(y)$. In this case, the iterated integral on the right side of (1.4.11) can be further simplified to,

$$\begin{aligned}\mathbf{E}(f(X)g(Y)) &= \int \Big[\int f(x)g(y)\, d\mu_1(x) \Big] d\mu_2(y) = \int g(y) [\int f(x)\, d\mu_1(x)]\, d\mu_2(y) \\ &= \int [\mathbf{E}f(X)] g(y)\, d\mu_2(y) = \mathbf{E}f(X)\,\mathbf{E}g(Y)\end{aligned}$$

(with Theorem 1.3.61 applied twice here), which is the stated identity (1.4.12).

To deal with Borel functions f and g that are not necessarily non-negative, first apply (1.4.12) for the non-negative functions $|f|$ and $|g|$ to get that $\mathbf{E}(|f(X)g(Y)|) = \mathbf{E}|f(X)|\mathbf{E}|g(Y)| < \infty$. Thus, the assumed integrability of $f(X)$ and of $g(Y)$ allows us to apply again (1.4.11) for $h(x, y) = f(x)g(y)$. Now repeat the argument we used for deriving (1.4.12) in case of non-negative Borel functions. □

Another consequence of Fubini's theorem is the following *integration by parts* formula.

Lemma 1.4.30 (INTEGRATION BY PARTS). *Suppose $H(x) = \int_{-\infty}^{x} h(y)dy$ for a non-negative Borel function h and all $x \in \mathbb{R}$. Then, for any random variable X,*

$$\mathbf{E}H(X) = \int_{\mathbb{R}} h(y)\mathbf{P}(X > y)dy\,. \tag{1.4.13}$$

PROOF. Combining the change of variables formula (Theorem 1.3.61), with our assumption about $H(\cdot)$, we have that

$$\mathbf{E}H(X) = \int_{\mathbb{R}} H(x)d\mathcal{P}_X(x) = \int_{\mathbb{R}} \Big[\int_{\mathbb{R}} h(y)I_{x>y}\,d\lambda(y)\Big]d\mathcal{P}_X(x)\,,$$

where λ denotes Lebesgue's measure on $(\mathbb{R},\mathcal{B})$. For each $y \in \mathbb{R}$, the expectation of the simple function $x \mapsto h(x,y) = h(y)I_{x>y}$ with respect to $(\mathbb{R},\mathcal{B},\mathcal{P}_X)$ is merely $h(y)\mathbf{P}(X > y)$. Thus, applying Fubini's theorem for the non-negative measurable function $h(x,y)$ on the product space $\mathbb{R}\times\mathbb{R}$ equipped with its Borel σ-algebra $\mathcal{B}_{\mathbb{R}^2}$, and the σ-finite measures $\mu_1 = \mathcal{P}_X$ and $\mu_2 = \lambda$, we have that

$$\mathbf{E}H(X) = \int_{\mathbb{R}} \Big[\int_{\mathbb{R}} h(y)I_{x>y}\,d\mathcal{P}_X(x)\Big]d\lambda(y) = \int_{\mathbb{R}} h(y)\mathbf{P}(X > y)dy\,,$$

as claimed. □

Indeed, as we see next, by combining the integration by parts formula with Hölder's inequality we can convert bounds on tail probabilities to bounds on the moments of the corresponding random variables.

Lemma 1.4.31.

(a) *For any $r > p > 0$ and any random variable $Y \geq 0$,*

$$\begin{aligned}\mathbf{E}Y^p &= \int_0^\infty py^{p-1}\mathbf{P}(Y > y)dy = \int_0^\infty py^{p-1}\mathbf{P}(Y \geq y)dy \\ &= (1-\frac{p}{r})\int_0^\infty py^{p-1}\mathbf{E}[\min(Y/y,1)^r]dy\,.\end{aligned}$$

(b) *If $X, Y \geq 0$ are such that $\mathbf{P}(Y \geq y) \leq y^{-1}\mathbf{E}[XI_{Y\geq y}]$ for all $y > 0$, then $\|Y\|_p \leq q\|X\|_p$ for any $p > 1$ and $q = p/(p-1)$.*

(c) *Under the same hypothesis also $\mathbf{E}Y \leq 1 + \mathbf{E}[X(\log Y)_+]$.*

PROOF. (a) The first identity is merely the integration by parts formula for $h_p(y) = py^{p-1}\mathbf{1}_{y>0}$ and $H_p(x) = x^p\mathbf{1}_{x\geq 0}$ and the second identity follows by the fact that $\mathbf{P}(Y = y) = 0$ up to a (countable) set of zero Lebesgue measure. Finally, it is easy to check that $H_p(x) = \int_{\mathbb{R}} h_{p,r}(x,y)dy$ for the non-negative Borel function $h_{p,r}(x,y) = (1-p/r)py^{p-1}\min(x/y,1)^r\mathbf{1}_{x\geq 0}\mathbf{1}_{y>0}$ and any $r > p > 0$. Hence, replacing $h(y)I_{x>y}$ throughout the proof of Lemma 1.4.30 by $h_{p,r}(x,y)$ we find that $\mathbf{E}[H_p(X)] = \int_0^\infty \mathbf{E}[h_{p,r}(X,y)]dy$, which is exactly our third identity.
(b) In a similar manner it follows from Fubini's theorem that for $p > 1$ and any non-negative random variables X and Y

$$\mathbf{E}[XY^{p-1}] = \mathbf{E}[XH_{p-1}(Y)] = \mathbf{E}[\int_{\mathbb{R}} h_{p-1}(y)XI_{Y\geq y}dy] = \int_{\mathbb{R}} h_{p-1}(y)\mathbf{E}[XI_{Y\geq y}]dy\,.$$

Thus, with $y^{-1}h_p(y) = qh_{p-1}(y)$ our hypothesis implies that

$$\mathbf{E}Y^p = \int_{\mathbb{R}} h_p(y)\mathbf{P}(Y \geq y)dy \leq \int_{\mathbb{R}} qh_{p-1}(y)\mathbf{E}[XI_{Y\geq y}]dy = q\mathbf{E}[XY^{p-1}]\,.$$

Applying Hölder's inequality we deduce that

$$\mathbf{E}Y^p \leq q\mathbf{E}[XY^{p-1}] \leq q\|X\|_p\|Y^{p-1}\|_q = q\|X\|_p[\mathbf{E}Y^p]^{1/q}$$

where the right-most equality is due to the fact that $(p-1)q = p$. In case Y is bounded, dividing both sides of the preceding bound by $[\mathbf{E}Y^p]^{1/q}$ implies that $\|Y\|_p \leq q\|X\|_p$. To deal with the general case, let $Y_n = Y \wedge n$, $n = 1, 2, \ldots$ and note that either $\{Y_n \geq y\}$ is empty (for $n < y$) or $\{Y_n \geq y\} = \{Y \geq y\}$. Thus, our assumption implies that $\mathbf{P}(Y_n \geq y) \leq y^{-1}\mathbf{E}[XI_{Y_n \geq y}]$ for all $y > 0$ and $n \geq 1$. By the preceding argument $\|Y_n\|_p \leq q\|X\|_p$ for any n. Taking $n \to \infty$ it follows by monotone convergence that $\|Y\|_p \leq q\|X\|_p$.
(c) Considering part (a) with $p = 1$, we bound $\mathbf{P}(Y \geq y)$ by one for $y \in [0,1]$ and by $y^{-1}\mathbf{E}[XI_{Y \geq y}]$ for $y > 1$, to get by Fubini's theorem that

$$\begin{aligned}\mathbf{E}Y &= \int_0^\infty \mathbf{P}(Y \geq y)dy \leq 1 + \int_1^\infty y^{-1}\mathbf{E}[XI_{Y \geq y}]dy \\ &= 1 + \mathbf{E}[X \int_1^\infty y^{-1}I_{Y \geq y}dy] = 1 + \mathbf{E}[X(\log Y)_+]\,.\end{aligned}$$

□

We further have the following corollary of (1.4.12), dealing with the expectation of a product of mutually independent R.V.

Corollary 1.4.32. *Suppose that $X_1, \ldots, X_n$ are $\mathbf{P}$-mutually independent random variables such that either $X_i \geq 0$ for all i, or $\mathbf{E}|X_i| < \infty$ for all i. Then,*

$$\mathbf{E}\Big(\prod_{i=1}^n X_i\Big) = \prod_{i=1}^n \mathbf{E}X_i\,, \tag{1.4.14}$$

that is, the expectation on the left exists and has the value given on the right.

Proof. By Corollary 1.4.11 we know that $X = X_1$ and $Y = X_2 \cdots X_n$ are independent. Taking $f(x) = |x|$ and $g(y) = |y|$ in Theorem 1.4.29, we thus have that $\mathbf{E}|X_1 \cdots X_n| = \mathbf{E}|X_1|\mathbf{E}|X_2 \cdots X_n|$ for any $n \geq 2$. Applying this identity iteratively for $X_l, \ldots, X_n$, starting with $l = m$, then $l = m+1, m+2, \ldots, n-1$ leads to

$$\mathbf{E}|X_m \cdots X_n| = \prod_{k=m}^n \mathbf{E}|X_k|\,, \tag{1.4.15}$$

holding for any $1 \leq m \leq n$. If $X_i \geq 0$ for all i, then $|X_i| = X_i$ and we have (1.4.14) as the special case $m = 1$.
To deal with the proof in case $X_i \in L^1$ for all i, note that for $m = 2$ the identity (1.4.15) tells us that $\mathbf{E}|Y| = \mathbf{E}|X_2 \cdots X_n| < \infty$, so using Theorem 1.4.29 with $f(x) = x$ and $g(y) = y$ we have that $\mathbf{E}(X_1 \cdots X_n) = (\mathbf{E}X_1)\mathbf{E}(X_2 \cdots X_n)$. Iterating this identity for $X_l, \ldots, X_n$, starting with $l = 1$, then $l = 2, 3, \ldots, n-1$ leads to the desired result (1.4.14). □

Another application of Theorem 1.4.29 provides us with the familiar formula for the probability density function of the sum $X+Y$ of independent random variables X and Y, having densities f_X and f_Y respectively.

Corollary 1.4.33. *Suppose that R.V. X with a Borel measurable probability density function f_X and R.V. Y with a Borel measurable probability density function f_Y are independent. Then, the random variable $Z = X + Y$ has the probability density function*

$$f_Z(z) = \int_{\mathbb{R}} f_X(z-y) f_Y(y) dy \,.$$

PROOF. Fixing $z \in \mathbb{R}$, apply Theorem 1.4.29 for $h(x,y) = \mathbf{1}_{(x+y\le z)}$, to get that

$$F_Z(z) = \mathbf{P}(X + Y \le z) = \mathbf{E}h(X,Y) = \int_{\mathbb{R}} \Big[\int_{\mathbb{R}} h(x,y) d\mathcal{P}_X(x) \Big] d\mathcal{P}_Y(y) \,.$$

Considering the inner integral for a fixed value of y, we have that

$$\int_{\mathbb{R}} h(x,y) d\mathcal{P}_X(x) = \int_{\mathbb{R}} I_{(-\infty,z-y]}(x) d\mathcal{P}_X(x) = \mathcal{P}_X((-\infty, z-y]) = \int_{-\infty}^{z-y} f_X(x) dx \,,$$

where the right most equality is by the existence of a density $f_X(x)$ for X (c.f. Definition 1.2.40). Clearly, $\int_{-\infty}^{z-y} f_X(x)dx = \int_{-\infty}^{z} f_X(x-y)dx$. Thus, applying Fubini's theorem for the Borel measurable function $g(x,y) = f_X(x-y) \ge 0$ and the product of the σ-finite Lebesgue's measure on $(-\infty, z]$ and the probability measure $\mathcal{P}_Y$, we see that

$$F_Z(z) = \int_{\mathbb{R}} \Big[\int_{-\infty}^{z} f_X(x-y) dx \Big] d\mathcal{P}_Y(y) = \int_{-\infty}^{z} \Big[\int_{\mathbb{R}} f_X(x-y) d\mathcal{P}_Y(y) \Big] dx$$

(in this application of Fubini's theorem we replace one iterated integral by another, exchanging the order of integrations). Since this applies for any $z \in \mathbb{R}$, it follows by definition that Z has the probability density

$$f_Z(z) = \int_{\mathbb{R}} f_X(z-y) d\mathcal{P}_Y(y) = \mathbf{E} f_X(z-Y) \,.$$

With Y having density f_Y, the stated formula for f_Z is a consequence of Corollary 1.3.62. □

Definition 1.4.34. *The expression $\int f(z-y)g(y)dy$ is called the* convolution *of the non-negative Borel functions f and g, denoted by $f * g(z)$. The convolution of measures μ and ν on $(\mathbb{R}, \mathcal{B})$ is the measure $\mu * \nu$ on $(\mathbb{R}, \mathcal{B})$ such that $\mu * \nu(B) = \int \mu(B - x) d\nu(x)$ for any $B \in \mathcal{B}$ (where $B - x = \{y : x + y \in B\}$).*

Corollary 1.4.33 states that if two independent random variables X and Y have densities, then so does $Z = X + Y$, whose density is the convolution of the densities of X and Y. Without assuming the existence of densities, one can show by a similar argument that the law of $X + Y$ is the convolution of the law of X and the law of Y (c.f. [**Dur10**, Theorem 2.1.10] or [**Bil95**, Page 266]).

Convolution is often used in analysis to provide a more regular approximation to a given function. Here are few of the reasons for doing so.

Exercise 1.4.35. *Suppose Borel functions f, g are such that g is a probability density and $\int |f(x)|dx$ is finite. Consider the scaled densities $g_n(\cdot) = ng(n\cdot)$, $n \ge 1$.*

(a) *Show that $f * g(y)$ is a Borel function with $\int |f * g(y)|dy \le \int |f(x)|dx$ and if g is uniformly continuous, then so is $f * g$.*
(b) *Show that if $g(x) = 0$ whenever $|x| \ge 1$, then $f * g_n(y) \to f(y)$ as $n \to \infty$, for any continuous f and each $y \in \mathbb{R}$.*

Next you find two of the many applications of Fubini's theorem in real analysis.

Exercise 1.4.36. *Show that the set $G_f = \{(x,y) \in \mathbb{R}^2 : 0 \le y \le f(x)\}$ of points under the graph of a non-negative Borel function $f : \mathbb{R} \mapsto [0,\infty)$ is in $\mathcal{B}_{\mathbb{R}^2}$ and deduce the well-known formula $\lambda \times \lambda(G_f) = \int f(x) d\lambda(x)$, for its area.*

Exercise 1.4.37. *For $n \ge 2$, consider the unit sphere $S^{n-1} = \{\underline{x} \in \mathbb{R}^n : \|\underline{x}\| = 1\}$ equipped with the topology induced by $\mathbb{R}^n$. Let the* surface measure *of $A \in \mathcal{B}_{S^{n-1}}$ be $\nu(A) = n\lambda^n(C_{0,1}(A))$, for $C_{a,b}(A) = \{r\underline{x} : r \in (a,b],\ \underline{x} \in A\}$ and the n-fold product Lebesgue measure λ^n (as in Remark 1.4.20).*

(a) *Check that $C_{a,b}(A) \in \mathcal{B}_{\mathbb{R}^n}$ and deduce that $\nu(\cdot)$ is a finite measure on S^{n-1} (which is further invariant under orthogonal transformations).*

(b) *Verify that $\lambda^n(C_{a,b}(A)) = \frac{b^n - a^n}{n}\nu(A)$ and deduce that for any $B \in \mathcal{B}_{\mathbb{R}^n}$*

$$\lambda^n(B) = \int_0^\infty \Big[\int_{S^{n-1}} I_{r\underline{x} \in B}\, d\nu(\underline{x}) \Big] r^{n-1}\, d\lambda(r)\,.$$

Hint: *Recall that $\lambda^n(\gamma B) = \gamma^n \lambda^n(B)$ for any $\gamma \ge 0$ and $B \in \mathcal{B}_{\mathbb{R}^n}$.*

Combining (1.4.12) with Theorem 1.2.26 leads to the following characterization of the independence between two random vectors (compare with Definition 1.4.1).

Exercise 1.4.38. *Show that the $\mathbb{R}^n$-valued random variable $(X_1, \ldots, X_n)$ and the $\mathbb{R}^m$-valued random variable $(Y_1, \ldots, Y_m)$ are independent if and only if*

$$\mathbf{E}(h(X_1, \ldots, X_n) g(Y_1, \ldots, Y_m)) = \mathbf{E}(h(X_1, \ldots, X_n)) \mathbf{E}(g(Y_1, \ldots, Y_m)),$$

for all bounded, Borel measurable functions $g : \mathbb{R}^m \mapsto \mathbb{R}$ and $h : \mathbb{R}^n \mapsto \mathbb{R}$. Then show that the assumption of $h(\cdot)$ and $g(\cdot)$ bounded can be relaxed to both $h(X_1, \ldots, X_n)$ and $g(Y_1, \ldots, Y_m)$ being in $L^1(\Omega, \mathcal{F}, \mathbf{P})$.

Here is another application of (1.4.12):

Exercise 1.4.39. *Show that $\mathbf{E}(f(X)g(X)) \ge (\mathbf{E}f(X))(\mathbf{E}g(X))$ for every random variable X and any bounded non-decreasing functions $f, g : \mathbb{R} \mapsto \mathbb{R}$.*

In the following exercise you bound the exponential moments of certain random variables.

Exercise 1.4.40. *Suppose Y is an integrable random variable such that $\mathbf{E}[e^Y]$ is finite and $\mathbf{E}[Y] = 0$.*

(a) *Show that if $|Y| \le \kappa$ then*

$$\log \mathbf{E}[e^Y] \le \kappa^{-2}(e^\kappa - \kappa - 1)\mathbf{E}[Y^2]\,.$$

Hint: *Use the Taylor expansion of $e^Y - Y - 1$.*

(b) *Show that if $\mathbf{E}[Y^2 e^Y] \le \kappa^2 \mathbf{E}[e^Y]$ then*

$$\log \mathbf{E}[e^Y] \le \log\cosh(\kappa)\,.$$

Hint: *Note that $\varphi(u) = \log \mathbf{E}[e^{uY}]$ is convex, non-negative and finite on $[0,1]$ with $\varphi(0) = 0$ and $\varphi'(0) = 0$. Verify that $\varphi''(u) + \varphi'(u)^2 = \mathbf{E}[Y^2 e^{uY}]/\mathbf{E}[e^{uY}]$ is non-decreasing on $[0,1]$ and $\phi(u) = \log\cosh(\kappa u)$ satisfies the differential equation $\phi''(u) + \phi'(u)^2 = \kappa^2$.*

As demonstrated next, Fubini's theorem is also handy in proving the impossibility of certain constructions.

Exercise 1.4.41. *Explain why it is impossible to have* $\mathbf{P}$*-mutually independent random variables* $U_t(\omega)$, $t \in [0,1]$, *on the same probability space* $(\Omega, \mathcal{F}, \mathbf{P})$, *having each the uniform probability measure on* $[-1/2, 1/2]$, *such that* $t \mapsto U_t(\omega)$ *is a Borel function for almost every* $\omega \in \Omega$.
Hint: *Show that* $\mathbf{E}[(\int_0^r U_t(\omega)dt)^2] = 0$ *for all* $r \in [0,1]$.

Random variables X and Y such that $\mathbf{E}(X^2) < \infty$ and $\mathbf{E}(Y^2) < \infty$ are called *uncorrelated* if $\mathbf{E}(XY) = \mathbf{E}(X)\mathbf{E}(Y)$. It follows from (1.4.12) that independent random variables X, Y with finite second moment are uncorrelated. While the converse is not necessarily true, it does apply for pairs of random variables that take only two different values each.

Exercise 1.4.42. *Suppose* X *and* Y *are uncorrelated random variables.*

(a) *Show that if* $X = I_A$ *and* $Y = I_B$ *for some* $A, B \in \mathcal{F}$ *then* X *and* Y *are also independent.*
(b) *Using this, show that if* $\{a, b\}$*-valued R.V.* X *and* $\{c, d\}$*-valued R.V.* Y *are uncorrelated, then they are also independent.*
(c) *Give an example of a pair of R.V.* X *and* Y *that are uncorrelated but not independent.*

Next come a pair of exercises utilizing Corollary 1.4.32.

Exercise 1.4.43. *Suppose* X *and* Y *are random variables on the same probability space,* X *has a Poisson distribution with parameter* $\lambda > 0$, *and* Y *has a Poisson distribution with parameter* $\mu > \lambda$ *(see Example 1.3.69).*

(a) *Show that if* X *and* Y *are independent then* $\mathbf{P}(X \geq Y) \leq \exp(-(\sqrt{\mu} - \sqrt{\lambda})^2)$.
(b) *Taking* $\mu = \gamma\lambda$ *for* $\gamma > 1$, *find* $I(\gamma) > 0$ *such that* $\mathbf{P}(X \geq Y) \leq 2\exp(-\lambda I(\gamma))$ *even when* X *and* Y *are not independent.*

Exercise 1.4.44. *Suppose* X *and* Y *are independent random variables of identical distribution such that* $X > 0$ *and* $\mathbf{E}[X] < \infty$.

(a) *Show that* $\mathbf{E}[X^{-1}Y] > 1$ *unless* $X(\omega) = c$ *for some non-random* c *and almost every* $\omega \in \Omega$.
(b) *Provide an example in which* $\mathbf{E}[X^{-1}Y] = \infty$.

We conclude this section with a concrete application of Corollary 1.4.33, computing the density of the sum of mutually independent R.V., each having the same exponential density. To this end, recall

Definition 1.4.45. *The* gamma density *with parameters* $\alpha > 0$ *and* $\lambda > 0$ *is given by*

$$f_\Gamma(s) = \Gamma(\alpha)^{-1}\lambda^\alpha s^{\alpha-1}e^{-\lambda s}\mathbf{1}_{s>0}\,,$$

where $\Gamma(\alpha) = \int_0^\infty s^{\alpha-1}e^{-s}ds$ *is finite and positive. In particular,* $\alpha = 1$ *corresponds to the exponential density* f_T *of Example 1.3.68.*

Exercise 1.4.46. *Suppose* X *has a* gamma *density of parameters* α_1 *and* λ *and* Y *has a* gamma *density of parameters* α_2 *and* λ. *Show that if* X *and* Y *are independent then* $X + Y$ *has a* gamma *density of parameters* $\alpha_1 + \alpha_2$ *and* λ. *Deduce that if* $T_1, \ldots, T_n$ *are mutually independent R.V. each having the exponential density of parameter* λ, *then* $W_n = \sum_{i=1}^n T_i$ *has the* gamma *density of parameters* $\alpha = n$ *and* λ.

CHAPTER 2

Asymptotics: the law of large numbers

Building upon the foundations of Chapter 1 we turn to deal with asymptotic theory. To this end, this chapter is devoted to degenerate limit laws, that is, situations in which a sequence of random variables converges to a non-random (constant) limit. Though not exclusively dealing with it, our focus here is on the sequence of empirical averages $n^{-1}\sum_{i=1}^{n} X_i$ as $n \to \infty$.

Section 2.1 deals with the *weak law of large numbers*, where convergence in probability (or in L^q for some $q > 1$) is considered. This is strengthened in Section 2.3 to a *strong law of large numbers*, namely, to convergence almost surely. The key tools for this improvement are the Borel-Cantelli lemmas, to which Section 2.2 is devoted.

2.1. Weak laws of large numbers

A weak law of large numbers corresponds to the situation where the normalized sums of large number of random variables converge in probability to a non-random constant. Usually, the derivation of a weak low involves the computation of variances, on which we focus in Subsection 2.1.1. However, the L^2 convergence we obtain there is of a somewhat limited scope of applicability. To remedy this, we introduce the method of *truncation* in Subsection 2.1.2 and illustrate its power in a few representative examples.

2.1.1. L^2 limits for sums of uncorrelated variables. The key to our derivation of weak laws of large numbers is the computation of variances. As a preliminary step we define the covariance of two R.V. and extend the notion of a pair of *uncorrelated* random variables, to a (possibly infinite) family of R.V.

Definition 2.1.1. *The* covariance *of two random variables* $X, Y \in L^2(\Omega, \mathcal{F}, \mathbf{P})$ *is*

$$\mathsf{Cov}(X,Y) = \mathbf{E}[(X - \mathbf{E}X)(Y - \mathbf{E}Y)] = \mathbf{E}XY - \mathbf{E}X\mathbf{E}Y\,,$$

so in particular, $\mathsf{Cov}(X,X) = \mathsf{Var}(X)$.
We say that random variables $X_\alpha \in L^2(\Omega, \mathcal{F}, \mathbf{P})$ *are* uncorrelated *if*

$$\mathbf{E}(X_\alpha X_\beta) = \mathbf{E}(X_\alpha)\mathbf{E}(X_\beta) \qquad \forall \alpha \neq \beta\,,$$

or equivalently, if

$$\mathsf{Cov}(X_\alpha, X_\beta) = 0 \qquad \forall \alpha \neq \beta\,.$$

As we next show, the variance of the sum of finitely many uncorrelated random variables is the sum of the variances of the variables.

Lemma 2.1.2. *Suppose* $X_1, \ldots, X_n$ *are uncorrelated random variables (which necessarily are defined on the same probability space). Then,*

$$\mathsf{Var}(X_1 + \cdots + X_n) = \mathsf{Var}(X_1) + \cdots + \mathsf{Var}(X_n)\,. \tag{2.1.1}$$

PROOF. Let $S_n = \sum_{i=1}^{n} X_i$. By Definition 1.3.67 of the variance and linearity of the expectation we have that

$$\mathsf{Var}(S_n) = \mathbf{E}([S_n - \mathbf{E}S_n]^2) = \mathbf{E}\big([\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mathbf{E}X_i]^2\big) = \mathbf{E}\big([\sum_{i=1}^{n}(X_i - \mathbf{E}X_i)]^2\big) .$$

Writing the square of the sum as the sum of all possible cross-products, we get that

$$\begin{aligned}\mathsf{Var}(S_n) &= \sum_{i,j=1}^{n} \mathbf{E}[(X_i - \mathbf{E}X_i)(X_j - \mathbf{E}X_j)] \\ &= \sum_{i,j=1}^{n} \mathsf{Cov}(X_i, X_j) = \sum_{i=1}^{n} \mathsf{Cov}(X_i, X_i) = \sum_{i=1}^{n} \mathsf{Var}(X_i) ,\end{aligned}$$

where we use the fact that $\mathsf{Cov}(X_i, X_j) = 0$ for each $i \neq j$ since X_i and X_j are uncorrelated. $\square$

Equipped with this lemma we have our

Theorem 2.1.3 (L^2 WEAK LAW OF LARGE NUMBERS). *Consider $S_n = \sum_{i=1}^{n} X_i$ for uncorrelated random variables $X_1, \ldots, X_n, \ldots$. Suppose that $\mathsf{Var}(X_i) \leq C$ and $\mathbf{E}X_i = \overline{x}$ for some finite constants C, $\overline{x}$, and all $i = 1, 2, \ldots$. Then, $n^{-1}S_n \xrightarrow{L^2} \overline{x}$ as $n \to \infty$, and hence also $n^{-1}S_n \xrightarrow{p} \overline{x}$.*

PROOF. Our assumptions imply that $\mathbf{E}(n^{-1}S_n) = \overline{x}$, and further by Lemma 2.1.2 we have the bound $\mathsf{Var}(S_n) \leq nC$. Recall the scaling property (1.3.17) of the variance, implying that

$$\mathbf{E}\Big[(n^{-1}S_n - \overline{x})^2\Big] = \mathsf{Var}\left(n^{-1}S_n\right) = \frac{1}{n^2}\,\mathsf{Var}(S_n) \leq \frac{C}{n} \to 0$$

as $n \to \infty$. Thus, $n^{-1}S_n \xrightarrow{L^2} \overline{x}$ (recall Definition 1.3.26). By Proposition 1.3.29 this implies that also $n^{-1}S_n \xrightarrow{p} \overline{x}$. $\square$

The most important special case of Theorem 2.1.3 is,

Example 2.1.4. *Suppose that $X_1, \ldots, X_n$ are independent and identically distributed (or in short,* i.i.d.*), with $\mathbf{E}X_1^2 < \infty$. Then, $\mathbf{E}X_i^2 = C$ and $\mathbf{E}X_i = m_X$ are both finite and independent of i. So, the L^2 weak law of large numbers tells us that $n^{-1}S_n \xrightarrow{L^2} m_X$, and hence also $n^{-1}S_n \xrightarrow{p} m_X$.*

Remark. As we shall see, the weaker condition $\mathbf{E}|X_i| < \infty$ suffices for the convergence in probability of $n^{-1}S_n$ to m_X. In Section 2.3 we show that it even suffices for the convergence almost surely of $n^{-1}S_n$ to m_X, a statement called the strong law of large numbers.

Exercise 2.1.5. *Show that the conclusion of the L^2 weak law of large numbers holds even for correlated X_i, provided $\mathbf{E}X_i = \overline{x}$ and $\mathsf{Cov}(X_i, X_j) \leq r(|i - j|)$ for all i, j, and some bounded sequence $r(k) \to 0$ as $k \to \infty$.*

With an eye on generalizing the L^2 weak law of large numbers we observe that

Lemma 2.1.6. *If the random variables $Z_n \in L^2(\Omega, \mathcal{F}, \mathbf{P})$ and the non-random b_n are such that $b_n^{-2}\,\mathsf{Var}(Z_n) \to 0$ as $n \to \infty$, then $b_n^{-1}(Z_n - \mathbf{E}Z_n) \xrightarrow{L^2} 0$.*

PROOF. We have $\mathbf{E}[(b_n^{-1}(Z_n - \mathbf{E}Z_n))^2] = b_n^{-2}\,\mathsf{Var}(Z_n) \to 0$. $\square$

Example 2.1.7. *Let $Z_n = \sum_{k=1}^n X_k$ for uncorrelated random variables $\{X_k\}$. If $\mathsf{Var}(X_k)/k \to 0$ as $k \to \infty$, then Lemma 2.1.6 applies for Z_n and $b_n = n$, hence $n^{-1}(Z_n - \mathbf{E}Z_n) \to 0$ in L^2 (and in probability). Alternatively, if also $\mathsf{Var}(X_k) \to 0$, then Lemma 2.1.6 applies even for Z_n and $b_n = n^{-1/2}$.*

Many limit theorems involve random variables of the form $S_n = \sum_{k=1}^{n} X_{n,k}$, that is, the row sums of triangular arrays of random variables $\{X_{n,k} : k = 1, \ldots, n\}$. Here are two such examples, both relying on Lemma 2.1.6.

Example 2.1.8 (COUPON COLLECTOR'S PROBLEM)**.** *Consider i.i.d. random variables $U_1, U_2, \ldots$, each distributed uniformly on $\{1, 2, \ldots, n\}$. Let $|\{U_1, \ldots, U_l\}|$ denote the number of distinct elements among the first l variables, and $\tau_k^n = \inf\{l : |\{U_1, \ldots, U_l\}| = k\}$ be the first time one has k distinct values. We are interested in the asymptotic behavior as $n \to \infty$ of $T_n = \tau_n^n$, the time it takes to have at least one representative of each of the n possible values.*

To motivate the name assigned to this example, think of collecting a set of n different coupons, where independently of all previous choices, each item is chosen at random in such a way that each of the possible n outcomes is equally likely. Then, T_n is the number of items one has to collect till having the complete set.

Setting $\tau_0^n = 0$, let $X_{n,k} = \tau_k^n - \tau_{k-1}^n$ denote the additional time it takes to get an item different from the first $k-1$ distinct items collected. Note that $X_{n,k}$ has a geometric distribution of success probability $q_{n,k} = 1 - \frac{k-1}{n}$, hence $\mathbf{E}X_{n,k} = q_{n,k}^{-1}$ and $\mathsf{Var}(X_{n,k}) \le q_{n,k}^{-2}$ (see Example 1.3.69). Since

$$T_n = \tau_n^n - \tau_0^n = \sum_{k=1}^{n} (\tau_k^n - \tau_{k-1}^n) = \sum_{k=1}^{n} X_{n,k} \,,$$

we have by linearity of the expectation that

$$\mathbf{E}T_n = \sum_{k=1}^{n} \Big(1 - \frac{k-1}{n}\Big)^{-1} = n \sum_{\ell=1}^{n} \ell^{-1} = n(\log n + \gamma_n) \,,$$

where $\gamma_n = \sum_{\ell=1}^{n} \ell^{-1} - \int_1^n x^{-1} dx$ is between zero and one (by monotonicity of $x \mapsto x^{-1}$). Further, $X_{n,k}$ is independent of each earlier waiting time $X_{n,j}$, $j = 1, \ldots, k-1$, hence we have by Lemma 2.1.2 that

$$\mathsf{Var}(T_n) = \sum_{k=1}^{n} \mathsf{Var}(X_{n,k}) \le \sum_{k=1}^{n} \Big(1 - \frac{k-1}{n}\Big)^{-2} \le n^2 \sum_{\ell=1}^{\infty} \ell^{-2} = Cn^2 \,,$$

for some $C < \infty$. Applying Lemma 2.1.6 with $b_n = n \log n$, we deduce that

$$\frac{T_n - n(\log n + \gamma_n)}{n \log n} \stackrel{L^2}{\to} 0 \,.$$

Since $\gamma_n / \log n \to 0$, it follows that

$$\frac{T_n}{n \log n} \stackrel{L^2}{\to} 1 \,,$$

and $T_n/(n \log n) \to 1$ in probability as well.

One possible extension of Example 2.1.8 concerns infinitely many possible coupons. That is,

Exercise 2.1.9. *Suppose $\{\xi_k\}$ are i.i.d. positive integer valued random variables, with $\mathbf{P}(\xi_1 = i) = p_i > 0$ for $i = 1, 2, \dots$. Let $D_l = |\{\xi_1, \dots, \xi_l\}|$ denote the number of distinct elements among the first l variables.*

(a) *Show that $D_n \xrightarrow{a.s.} \infty$ as $n \to \infty$.*

(b) *Show that $n^{-1}\mathbf{E}D_n \to 0$ as $n \to \infty$ and deduce that $n^{-1}D_n \xrightarrow{p} 0$.*

Hint: *Recall that $(1-p)^n \geq 1 - np$ for any $p \in [0,1]$ and $n \geq 0$.*

Example 2.1.10 (An occupancy problem). *Suppose we distribute at random r distinct balls among n distinct boxes, where each of the possible n^r assignments of balls to boxes is equally likely. We are interested in the asymptotic behavior of the number N_n of empty boxes when $r/n \to \alpha \in [0, \infty]$, while $n \to \infty$. To this end, let A_i denote the event that the i-th box is empty, so $N_n = \sum_{i=1}^{n} I_{A_i}$. Since $\mathbf{P}(A_i) = (1 - 1/n)^r$ for each i, it follows that $\mathbf{E}(n^{-1}N_n) = (1 - 1/n)^r \to e^{-\alpha}$. Further, $\mathbf{E}N_n^2 = \sum_{i,j=1}^{n} \mathbf{P}(A_i \cap A_j)$ and $\mathbf{P}(A_i \cap A_j) = (1 - 2/n)^r$ for each $i \neq j$. Hence, splitting the sum according to $i = j$ or $i \neq j$, we see that*

$$\mathsf{Var}(n^{-1}N_n) = \frac{1}{n^2}\mathbf{E}N_n^2 - (1 - \frac{1}{n})^{2r} = \frac{1}{n}(1 - \frac{1}{n})^r + (1 - \frac{1}{n})(1 - \frac{2}{n})^r - (1 - \frac{1}{n})^{2r}.$$

As $n \to \infty$, the first term on the right side goes to zero, and with $r/n \to \alpha$, each of the other two terms converges to $e^{-2\alpha}$. Consequently, $\mathsf{Var}(n^{-1}N_n) \to 0$, so applying Lemma 2.1.6 for $b_n = n$ we deduce that

$$\frac{N_n}{n} \to e^{-\alpha}$$

in L^2 and in probability.

2.1.2. Weak laws and truncation. Our next order of business is to extend the weak law of large numbers for row sums S_n in triangular arrays of independent $X_{n,k}$ which lack a finite second moment. Of course, with S_n no longer in L^2, there is no way to establish convergence in L^2. So, we aim to retain only the convergence in probability, using *truncation*. That is, we consider the row sums $\overline{S}_n$ for the truncated array $\overline{X}_{n,k} = X_{n,k} I_{|X_{n,k}| \leq b_n}$, with $b_n \to \infty$ slowly enough to control the variance of $\overline{S}_n$ and fast enough for $\mathbf{P}(S_n \neq \overline{S}_n) \to 0$. As we next show, this gives the convergence in probability for $\overline{S}_n$ which translates to same convergence result for S_n.

Theorem 2.1.11 (Weak law for triangular arrays). *Suppose that for each n, the random variables $X_{n,k}$, $k = 1, \dots, n$ are pairwise independent. Let $\overline{X}_{n,k} = X_{n,k} I_{|X_{n,k}| \leq b_n}$ for non-random $b_n > 0$ such that as $n \to \infty$ both*

(a) $\sum_{k=1}^{n} \mathbf{P}(|X_{n,k}| > b_n) \to 0$,

and

(b) $b_n^{-2} \sum_{k=1}^{n} \mathsf{Var}(\overline{X}_{n,k}) \to 0$.

Then, $b_n^{-1}(S_n - a_n) \xrightarrow{p} 0$ as $n \to \infty$, where $S_n = \sum_{k=1}^{n} X_{n,k}$ and $a_n = \sum_{k=1}^{n} \mathbf{E}\overline{X}_{n,k}$.

PROOF. Let $\overline{S}_n = \sum_{k=1}^n \overline{X}_{n,k}$. Clearly, for any $\varepsilon > 0$,

$$\Big\{|\frac{S_n - a_n}{b_n}| > \varepsilon\Big\} \subseteq \Big\{S_n \neq \overline{S}_n\Big\} \bigcup \Big\{|\frac{\overline{S}_n - a_n}{b_n}| > \varepsilon\Big\} .$$

Consequently,

$$\mathbf{P}\Big(|\frac{S_n - a_n}{b_n}| > \varepsilon\Big) \le \mathbf{P}(S_n \neq \overline{S}_n) + \mathbf{P}\Big(|\frac{\overline{S}_n - a_n}{b_n}| > \varepsilon\Big) . \tag{2.1.2}$$

To bound the first term, note that our condition (a) implies that as $n \to \infty$,

$$\begin{aligned}\mathbf{P}(S_n \neq \overline{S}_n) &\le \mathbf{P}\Big(\bigcup_{k=1}^n \{X_{n,k} \neq \overline{X}_{n,k}\}\Big) \\ &\le \sum_{k=1}^n \mathbf{P}(X_{n,k} \neq \overline{X}_{n,k}) = \sum_{k=1}^n \mathbf{P}(|X_{n,k}| > b_n) \to 0 .\end{aligned}$$

Turning to bound the second term in (2.1.2), recall that pairwise independence is preserved under truncation, hence $\overline{X}_{n,k}$, $k = 1, \dots, n$ are uncorrelated random variables (to convince yourself, apply (1.4.12) for the appropriate functions). Thus, an application of Lemma 2.1.2 yields that as $n \to \infty$,

$$\mathsf{Var}(b_n^{-1}\overline{S}_n) = b_n^{-2} \sum_{k=1}^n \mathsf{Var}(\overline{X}_{n,k}) \to 0 ,$$

by our condition (b). Since $a_n = \mathbf{E}\overline{S}_n$, from Chebyshev's inequality we deduce that for any fixed $\varepsilon > 0$,

$$\mathbf{P}\Big(|\frac{\overline{S}_n - a_n}{b_n}| > \varepsilon\Big) \le \varepsilon^{-2}\,\mathsf{Var}(b_n^{-1}\overline{S}_n) \to 0 ,$$

as $n \to \infty$. In view of (2.1.2), this completes the proof of the theorem. $\square$

Specializing the weak law of Theorem 2.1.11 to a single sequence yields the following.

Proposition 2.1.12 (Weak law of large numbers)**.** *Consider i.i.d. random variables $\{X_i\}$, such that $x\mathbf{P}(|X_1| > x) \to 0$ as $x \to \infty$. Then, $n^{-1}S_n - \mu_n \xrightarrow{p} 0$, where $S_n = \sum\limits_{i=1}^n X_i$ and $\mu_n = \mathbf{E}[X_1 I_{\{|X_1| \le n\}}]$.*

PROOF. We get the result as an application of Theorem 2.1.11 for $X_{n,k} = X_k$ and $b_n = n$, in which case $a_n = n\mu_n$. Turning to verify condition (a) of this theorem, note that

$$\sum_{k=1}^n \mathbf{P}(|X_{n,k}| > n) = n\mathbf{P}(|X_1| > n) \to 0$$

as $n \to \infty$, by our assumption. Thus, all that remains to do is to verify that condition (b) of Theorem 2.1.11 holds here. This amounts to showing that as $n \to \infty$,

$$\Delta_n = n^{-2} \sum_{k=1}^n \mathsf{Var}(\overline{X}_{n,k}) = n^{-1}\,\mathsf{Var}(\overline{X}_{n,1}) \to 0 .$$

Recall that for any R.V. Z,

$$\mathsf{Var}(Z) = \mathbf{E}Z^2 - (\mathbf{E}Z)^2 \le \mathbf{E}|Z|^2 = \int_0^\infty 2y\mathbf{P}(|Z| > y)dy$$

(see part (a) of Lemma 1.4.31 for the right identity). Considering $Z = \overline{X}_{n,1} = X_1 I_{\{|X_1| \le n\}}$ for which $\mathbf{P}(|Z| > y) = \mathbf{P}(|X_1| > y) - \mathbf{P}(|X_1| > n) \le \mathbf{P}(|X_1| > y)$ when $0 < y < n$ and $\mathbf{P}(|Z| > y) = 0$ when $y \ge n$, we deduce that

$$\Delta_n = n^{-1}\mathsf{Var}(Z) \le n^{-1}\int_0^n g(y)dy\,,$$

where by our assumption, $g(y) = 2y\mathbf{P}(|X_1| > y) \to 0$ for $y \to \infty$. Further, the non-negative Borel function $g(y) \le 2y$ is then uniformly bounded on $[0,\infty)$, hence $n^{-1}\int_0^n g(y)dy \to 0$ as $n \to \infty$ (c.f. Exercise 1.3.52). Verifying that $\Delta_n \to 0$, we established condition (b) of Theorem 2.1.11 and thus completed the proof of the proposition. $\square$

Remark. The condition $x\mathbf{P}(|X_1| > x) \to 0$ for $x \to \infty$ is indeed necessary for the existence of non-random μ_n such that $n^{-1}S_n - \mu_n \xrightarrow{p} 0$ (c.f. [**Fel71**, Page 234-236] for a proof).

Exercise 2.1.13. *Let $\{X_i\}$ be i.i.d. with $\mathbf{P}(X_1 = (-1)^k k) = 1/(ck^2 \log k)$ for integers $k \ge 2$ and a normalization constant $c = \sum_k 1/(k^2 \log k)$. Show that $\mathbf{E}|X_1| = \infty$, but there is a non-random $\mu < \infty$ such that $n^{-1}S_n \xrightarrow{p} \mu$.*

As a corollary to Proposition 2.1.12 we next show that $n^{-1}S_n \xrightarrow{p} m_X$ as soon as the i.i.d. random variables X_i are in L^1.

Corollary 2.1.14. *Consider $S_n = \sum_{k=1}^{n} X_k$ for i.i.d. random variables $\{X_i\}$ such that $\mathbf{E}|X_1| < \infty$. Then, $n^{-1}S_n \xrightarrow{p} \mathbf{E}X_1$ as $n \to \infty$.*

Proof. In view of Proposition 2.1.12, it suffices to show that if $\mathbf{E}|X_1| < \infty$, then both $n\mathbf{P}(|X_1| > n) \to 0$ and $\mathbf{E}X_1 - \mu_n = \mathbf{E}[X_1 I_{\{|X_1|>n\}}] \to 0$ as $n \to \infty$. To this end, recall that $\mathbf{E}|X_1| < \infty$ implies that $\mathbf{P}(|X_1| < \infty) = 1$ and hence the sequence $X_1 I_{\{|X_1|>n\}}$ converges to zero a.s. and is bounded by the integrable $|X_1|$. Thus, by dominated convergence $\mathbf{E}[X_1 I_{\{|X_1|>n\}}] \to 0$ as $n \to \infty$. Applying dominated convergence for the sequence $nI_{\{|X_1|>n\}}$ (which also converges a.s. to zero and is bounded by the integrable $|X_1|$), we deduce that $n\mathbf{P}(|X_1| > n) = \mathbf{E}[nI_{\{|X_1|>n\}}] \to 0$ when $n \to \infty$, thus completing the proof of the corollary. $\square$

We conclude this section by considering an example for which $\mathbf{E}|X_1| = \infty$ and Proposition 2.1.12 does not apply, but nevertheless, Theorem 2.1.11 allows us to deduce that $c_n^{-1}S_n \xrightarrow{p} 1$ for some c_n such that $c_n/n \to \infty$.

Example 2.1.15. *Let $\{X_i\}$ be i.i.d. random variables such that $\mathbf{P}(X_1 = 2^j) = 2^{-j}$ for $j = 1, 2, \ldots$. This has the interpretation of a game, where in each of its independent rounds you win 2^j dollars if it takes exactly j tosses of a fair coin to get the first Head. This example is called the* St. Petersburg paradox, *since though $\mathbf{E}X_1 = \infty$, you clearly would not pay an infinite amount just in order to play this game. Applying Theorem 2.1.11 we find that one should be willing to pay roughly $n \log_2 n$ dollars for playing n rounds of this game, since $S_n/(n \log_2 n) \xrightarrow{p} 1$ as $n \to \infty$. Indeed, the conditions of Theorem 2.1.11 apply for $b_n = 2^{m_n}$ provided* *the integers m_n are such that $m_n - \log_2 n \to \infty$. Taking $m_n \le \log_2 n + \log_2(\log_2 n)$ implies that $b_n \le n \log_2 n$ and $a_n/(n \log_2 n) = m_n / \log_2 n \to 1$ as $n \to \infty$, with the consequence of $S_n/(n \log_2 n) \xrightarrow{p} 1$ (for details see* [**Dur10**, Example 2.2.7]*).*

2.2. The Borel-Cantelli lemmas

When dealing with asymptotic theory, we often wish to understand the relation between countably many events A_n in the same probability space. The two Borel-Cantelli lemmas of Subsection 2.2.1 provide information on the probability of the set of outcomes that are in infinitely many of these events, based only on $\mathbf{P}(A_n)$. There are numerous applications to these lemmas, few of which are given in Subsection 2.2.2 while many more appear in later sections of these notes.

2.2.1. Limit superior and the Borel-Cantelli lemmas.

We are often interested in the *limits superior* and *limits inferior* of a sequence of events A_n on the same measurable space $(\Omega, \mathcal{F})$.

Definition 2.2.1. *For a sequence of subsets $A_n \subseteq \Omega$, define*

$$\begin{aligned} A^\infty := \limsup A_n &= \bigcap_{m=1}^{\infty} \bigcup_{\ell=m}^{\infty} A_\ell \\ &= \{\omega : \omega \in A_n \text{ for infinitely many } n\text{'s }\} \\ &= \{\omega : \omega \in A_n \text{ infinitely often }\} = \{A_n \text{ i.o. }\} \end{aligned}$$

Similarly,

$$\begin{aligned} \liminf A_n &= \bigcup_{m=1}^{\infty} \bigcap_{\ell=m}^{\infty} A_\ell \\ &= \{\omega : \omega \in A_n \text{ for all but finitely many } n\text{'s }\} \\ &= \{\omega : \omega \in A_n \text{ eventually }\} = \{A_n \text{ ev. }\} \end{aligned}$$

Remark. Note that if $A_n \in \mathcal{F}$ are measurable, then so are $\limsup A_n$ and $\liminf A_n$. By DeMorgan's law, we have that $\{A_n \textit{ ev. }\} = \{A_n^c \textit{ i.o. }\}^c$, that is, $\omega \in A_n$ for all n large enough if and only if $\omega \in A_n^c$ for finitely many n's.

Also, if $\omega \in A_n$ eventually, then certainly $\omega \in A_n$ infinitely often, that is

$$\liminf A_n \subseteq \limsup A_n .$$

The notations $\limsup A_n$ and $\liminf A_n$ are due to the intimate connection of these sets to the lim sup and lim inf of the indicator functions on the sets A_n. For example,

$$\limsup_{n\to\infty} I_{A_n}(\omega) = I_{\limsup A_n}(\omega),$$

since for a given $\omega \in \Omega$, the lim sup on the left side equals 1 if and only if the sequence $n \mapsto I_{A_n}(\omega)$ contains an infinite subsequence of ones. In other words, if and only if the given ω is in infinitely many of the sets A_n. Similarly,

$$\liminf_{n\to\infty} I_{A_n}(\omega) = I_{\liminf A_n}(\omega),$$

since for a given $\omega \in \Omega$, the lim inf on the left side equals 1 if and only if there are only finitely many zeros in the sequence $n \mapsto I_{A_n}(\omega)$ (for otherwise, their limit inferior is zero). In other words, if and only if the given ω is in A_n for all n large enough.

In view of the preceding remark, Fatou's lemma yields the following relations.

Exercise 2.2.2. *Prove that for any sequence $A_n \in \mathcal{F}$,*

$$\mathbf{P}(\limsup A_n) \geq \limsup_{n\to\infty} \mathbf{P}(A_n) \geq \liminf_{n\to\infty} \mathbf{P}(A_n) \geq \mathbf{P}(\liminf A_n)\,.$$

Show that the right most inequality holds even when the probability measure is replaced by an arbitrary measure $\mu(\cdot)$, but the left most inequality may then fail unless $\mu(\bigcup_{k\geq n} A_k) < \infty$ for some n.

Practice your understanding of the concepts of lim sup and lim inf of sets by solving the following exercise.

Exercise 2.2.3. *Assume that $\mathbf{P}(\limsup A_n) = 1$ and $\mathbf{P}(\liminf B_n) = 1$. Prove that $\mathbf{P}(\limsup(A_n \cap B_n)) = 1$. What happens if the condition on $\{B_n\}$ is weakened to $\mathbf{P}(\limsup B_n) = 1$?*

Our next result, called the first Borel-Cantelli lemma, states that if the probabilities $\mathbf{P}(A_n)$ of the individual events A_n converge to zero fast enough, then almost surely, A_n occurs for only finitely many values of n, that is, $\mathbf{P}(A_n \text{ i.o.}) = 0$. This lemma is extremely useful, as the possibly complex relation between the different events A_n is irrelevant for its conclusion.

Lemma 2.2.4 (Borel-Cantelli I)**.** *Suppose $A_n \in \mathcal{F}$ and $\sum_{n=1}^{\infty} \mathbf{P}(A_n) < \infty$. Then, $\mathbf{P}(A_n \text{ i.o.}) = 0$.*

Proof. Define $N(\omega) = \sum_{k=1}^{\infty} I_{A_k}(\omega)$. By the monotone convergence theorem and our assumption,

$$\mathbf{E}[N(\omega)] = \mathbf{E}\Big[\sum_{k=1}^{\infty} I_{A_k}(\omega)\Big] = \sum_{k=1}^{\infty} \mathbf{P}(A_k) < \infty.$$

Since the expectation of N is finite, certainly $\mathbf{P}(\{\omega : N(\omega) = \infty\}) = 0$. Noting that the set $\{\omega : N(\omega) = \infty\}$ is merely $\{\omega : A_n \text{ i.o.}\}$, the conclusion $\mathbf{P}(A_n \text{ i.o.}) = 0$ of the lemma follows. □

Our next result, left for the reader to prove, relaxes somewhat the conditions of Lemma 2.2.4.

Exercise 2.2.5. *Suppose $A_n \in \mathcal{F}$ are such that $\sum_{n=1}^{\infty} \mathbf{P}(A_n \cap A_{n+1}^c) < \infty$ and $\mathbf{P}(A_n) \to 0$. Show that then $\mathbf{P}(A_n \text{ i.o.}) = 0$.*

The first Borel-Cantelli lemma states that if the series $\sum_n \mathbf{P}(A_n)$ converges then almost every ω is in finitely many sets A_n. If $\mathbf{P}(A_n) \to 0$, but the series $\sum_n \mathbf{P}(A_n)$ diverges, then the event $\{A_n \text{ i.o.}\}$ might or might not have positive probability. In this sense, the Borel-Cantelli I is not tight, as the following example demonstrates.

Example 2.2.6. *Consider the uniform probability measure U on $((0,1], \mathcal{B}_{(0,1]})$, and the events $A_n = (0, 1/n]$. Then $A_n \downarrow \emptyset$, so $\{A_n \text{ i.o.}\} = \emptyset$, but $U(A_n) = 1/n$, so $\sum_n U(A_n) = \infty$ and the Borel-Cantelli I does not apply.*

Recall also Example 1.3.25 showing the existence of $A_n = (t_n, t_n + 1/n]$ such that $U(A_n) = 1/n$ while $\{A_n \text{ i.o.}\} = (0,1]$. Thus, in general the probability of $\{A_n \text{ i.o.}\}$ depends on the relation between the different events A_n.

As seen in the preceding example, the divergence of the series $\sum_n \mathbf{P}(A_n)$ is not sufficient for the occurrence of a set of positive probability of ω values, each of which is in infinitely many events A_n. However, upon adding the assumption that the events A_n are mutually independent (flagrantly not the case in Example 2.2.6), we conclude that *almost all* ω must be in infinitely many of the events A_n:

Lemma 2.2.7 (Borel-Cantelli II). *Suppose $A_n \in \mathcal{F}$ are mutually independent and $\sum_{n=1}^{\infty} \mathbf{P}(A_n) = \infty$. Then, necessarily $\mathbf{P}(A_n \text{ i.o.}) = 1$.*

Proof. Fix $0 < m < n < \infty$. Use the mutual independence of the events A_ℓ and the inequality $1 - x \le e^{-x}$ for $x \ge 0$, to deduce that

$$\mathbf{P}\Big(\bigcap_{\ell=m}^{n} A_\ell^c\Big) = \prod_{\ell=m}^{n} \mathbf{P}(A_\ell^c) = \prod_{\ell=m}^{n} (1 - \mathbf{P}(A_\ell))$$
$$\le \prod_{\ell=m}^{n} e^{-\mathbf{P}(A_\ell)} = \exp(-\sum_{\ell=m}^{n} \mathbf{P}(A_\ell))\,.$$

As $n \to \infty$, the set $\bigcap_{\ell=m}^{n} A_\ell^c$ shrinks. With the series in the exponent diverging, by continuity from above of the probability measure $\mathbf{P}(\cdot)$ we see that for any m,

$$\mathbf{P}\Big(\bigcap_{\ell=m}^{\infty} A_\ell^c\Big) \le \exp(-\sum_{\ell=m}^{\infty} \mathbf{P}(A_\ell)) = 0\,.$$

Take the complement to see that $\mathbf{P}(B_m) = 1$ for $B_m = \bigcup_{\ell=m}^{\infty} A_\ell$ and all m. Since $B_m \downarrow \{A_n \text{ i.o. }\}$ when $m \uparrow \infty$, it follows by continuity from above of $\mathbf{P}(\cdot)$ that

$$\mathbf{P}(A_n \text{ i.o.}) = \lim_{m\to\infty} \mathbf{P}(B_m) = 1\,,$$

as stated. □

As an immediate corollary of the two Borel-Cantelli lemmas, we observe yet another 0-1 law.

Corollary 2.2.8. *If $A_n \in \mathcal{F}$ are $\mathbf{P}$-mutually independent then $\mathbf{P}(A_n \text{ i.o.})$ is either 0 or 1. In other words, for any given sequence of mutually independent events, either almost all outcomes are in infinitely many of these events, or almost all outcomes are in finitely many of them.*

The *Kochen-Stone lemma*, left as an exercise, generalizes Borel-Cantelli II to situations lacking independence.

Exercise 2.2.9. *Suppose A_k are events on the same probability space such that $\sum_k \mathbf{P}(A_k) = \infty$ and*

$$\limsup_{n\to\infty} \Big(\sum_{k=1}^{n} \mathbf{P}(A_k)\Big)^2 \Big/ \Big(\sum_{1\le j,k\le n} \mathbf{P}(A_j \cap A_k)\Big) = \alpha > 0\,.$$

Prove that then $\mathbf{P}(A_n \text{ i.o. }) \ge \alpha$.
Hint: *Consider part (a) of Exercise 1.3.21 for $Y_n = \sum_{k\le n} I_{A_k}$ and $a_n = \lambda \mathbf{E} Y_n$.*

2.2.2. Applications. In the sequel we explore various applications of the two Borel-Cantelli lemmas. In doing so, unless explicitly stated otherwise, all events and random variables are defined on the same probability space.

We know that the convergence a.s. of X_n to X_∞ implies the convergence in probability of X_n to X_∞, but not vice verse (see Exercise 1.3.23 and Example 1.3.25). As our first application of Borel-Cantelli I, we refine the relation between these two modes of convergence, showing that convergence in probability is equivalent to convergence almost surely along sub-sequences.

Theorem 2.2.10. *$X_n \stackrel{p}{\to} X_\infty$ if and only if for every subsequence $m \mapsto X_{n(m)}$ there exists a further sub-subsequence $X_{n(m_k)}$ such that $X_{n(m_k)} \stackrel{a.s.}{\to} X_\infty$ as $k \to \infty$.*

We start the proof of this theorem with a simple analysis lemma.

Lemma 2.2.11. *Let y_n be a sequence in a topological space. If every subsequence $y_{n(m)}$ has a further sub-subsequence $y_{n(m_k)}$ that converges to y, then $y_n \to y$.*

Proof. If y_n does not converge to y, then there exists an open set G containing y and a subsequence $y_{n(m)}$ such that $y_{n(m)} \notin G$ for all m. But clearly, then we cannot find a further subsequence of $y_{n(m)}$ that converges to y. $\square$

Remark. Applying Lemma 2.2.11 to $y_n = \mathbf{E}|X_n - X_\infty|$ we deduce that $X_n \stackrel{L^1}{\to} X_\infty$ if and only if any subsequence $n(m)$ has a further sub-subsequence $n(m_k)$ such that $X_{n(m_k)} \stackrel{L^1}{\to} X_\infty$ as $k \to \infty$.

Proof of Theorem 2.2.10. First, we show sufficiency, assuming $X_n \stackrel{p}{\to} X_\infty$. Fix a subsequence $n(m)$ and $\varepsilon_k \downarrow 0$. By the definition of convergence in probability, there exists a sub-subsequence $n(m_k) \uparrow \infty$ such that $\mathbf{P}\left(|X_{n(m_k)} - X_\infty| > \varepsilon_k\right) \le 2^{-k}$. Call this sequence of events $A_k = \left\{\omega : |X_{n(m_k)}(\omega) - X_\infty(\omega)| > \varepsilon_k\right\}$. Then the series $\sum_k \mathbf{P}(A_k)$ converges. Therefore, by Borel-Cantelli I, $\mathbf{P}(\limsup A_k) = 0$. For any $\omega \notin \limsup A_k$ there are only finitely many values of k such that $|X_{n(m_k)} - X_\infty| > \varepsilon_k$, or alternatively, $|X_{n(m_k)} - X_\infty| \le \varepsilon_k$ for all k large enough. Since $\varepsilon_k \downarrow 0$, it follows that $X_{n(m_k)}(\omega) \to X_\infty(\omega)$ when $\omega \notin \limsup A_k$, that is, with probability one.

Conversely, fix $\delta > 0$. Let $y_n = \mathbf{P}(|X_n - X_\infty| > \delta)$. By assumption, for every subsequence $n(m)$ there exists a further subsequence $n(m_k)$ so that $X_{n(m_k)}$ converges to X_∞ almost surely, hence in probability, and in particular, $y_{n(m_k)} \to 0$. Applying Lemma 2.2.11 we deduce that $y_n \to 0$, and since $\delta > 0$ is arbitrary it follows that $X_n \stackrel{p}{\to} X_\infty$. $\square$

It is not hard to check that convergence almost surely is invariant under application of an a.s. continuous mapping.

Exercise 2.2.12. *Let $g : \mathbb{R} \mapsto \mathbb{R}$ be a Borel function and denote by $\mathbf{D}_g$ its set of discontinuities. Show that if $X_n \stackrel{a.s.}{\to} X_\infty$ finite valued, and $\mathbf{P}(X_\infty \in \mathbf{D}_g) = 0$, then $g(X_n) \stackrel{a.s.}{\to} g(X_\infty)$ as well (recall Exercise 1.2.28 that $\mathbf{D}_g \in \mathcal{B}$). This applies for a continuous function g in which case $\mathbf{D}_g = \emptyset$.*

A direct consequence of Theorem 2.2.10 is that convergence in probability is also preserved under an a.s. *continuous mapping* (and if the mapping is also bounded, we even get L^1 convergence).

Corollary 2.2.13. *Suppose* $X_n \xrightarrow{p} X_\infty$, *$g$ is a Borel function and* $\mathbf{P}(X_\infty \in \mathbf{D}_g) = 0$. *Then,* $g(X_n) \xrightarrow{p} g(X_\infty)$. *If in addition g is bounded, then* $g(X_n) \xrightarrow{L^1} g(X_\infty)$ *(and* $\mathbf{E}g(X_n) \to \mathbf{E}g(X_\infty)$*).*

PROOF. Fix a subsequence $X_{n(m)}$. By Theorem 2.2.10 there exists a subsequence $X_{n(m_k)}$ such that $\mathbf{P}(A) = 1$ for $A = \{\omega : X_{n(m_k)}(\omega) \to X_\infty(\omega) \text{ as } k \to \infty\}$. Let $B = \{\omega : X_\infty(\omega) \notin \mathbf{D}_g\}$, noting that by assumption $\mathbf{P}(B) = 1$. For any $\omega \in A \cap B$ we have $g(X_{n(m_k)}(\omega)) \to g(X_\infty(\omega))$ by the continuity of g outside $\mathbf{D}_g$. Therefore, $g(X_{n(m_k)}) \xrightarrow{a.s.} g(X_\infty)$. Now apply Theorem 2.2.10 in the reverse direction: For any subsequence, we have just constructed a further subsequence with convergence a.s., hence $g(X_n) \xrightarrow{p} g(X_\infty)$.

Finally, if g is bounded, then the collection $\{g(X_n)\}$ is U.I. yielding, by Vitali's convergence theorem, its convergence in L^1 (and hence that $\mathbf{E}g(X_n) \to \mathbf{E}g(X_\infty)$). $\square$

You are next to extend the scope of Theorem 2.2.10 and the continuous mapping of Corollary 2.2.13 to random variables taking values in a separable metric space.

Exercise 2.2.14. *Recall the definition of convergence in probability in a separable metric space* $(\mathbb{S}, \rho)$ *as in Remark 1.3.24.*

(a) *Extend the proof of Theorem 2.2.10 to apply for any* $(\mathbb{S}, \mathcal{B}_\mathbb{S})$*-valued random variables* $\{X_n, n \le \infty\}$ *(and in particular for* $\overline{\mathbb{R}}$*-valued variables).*

(b) *Denote by* $\mathbf{D}_g$ *the set of discontinuities of a Borel measurable* $g : \mathbb{S} \mapsto \overline{\mathbb{R}}$ *(defined similarly to Exercise 1.2.28, where real-valued functions are considered). Suppose* $X_n \xrightarrow{p} X_\infty$ *and* $\mathbf{P}(X_\infty \in \mathbf{D}_g) = 0$. *Show that then* $g(X_n) \xrightarrow{p} g(X_\infty)$ *and if in addition g is bounded, then also* $g(X_n) \xrightarrow{L^1} g(X_\infty)$.

The following result in analysis is obtained by combining the continuous mapping of Corollary 2.2.13 with the weak law of large numbers.

Exercise 2.2.15 (INVERTING LAPLACE TRANSFORMS). *The* Laplace transform *of a bounded, continuous function $h(x)$ on* $[0, \infty)$ *is the function* $L_h(s) = \int_0^\infty e^{-sx} h(x)dx$ *on* $(0, \infty)$.

(a) *Show that for any $s > 0$ and positive integer k,*
$$(-1)^{k-1} \frac{s^k L_h^{(k-1)}(s)}{(k-1)!} = \int_0^\infty e^{-sx} \frac{s^k x^{k-1}}{(k-1)!} h(x)dx = \mathbf{E}[h(W_k)] \,,$$
where $L_h^{(k-1)}(\cdot)$ *denotes the* $(k-1)$*-th derivative of the function* $L_h(\cdot)$ *and* W_k *has the* gamma density *with parameters k and s.*

(b) *Recall Exercise 1.4.46 that for* $s = n/y$ *the law of* W_n *coincides with the law of* $n^{-1}\sum_{i=1}^n T_i$ *where* $T_i \ge 0$ *are i.i.d. random variables, each having the exponential distribution of parameter* $1/y$ *(with* $\mathbf{E}T_1 = y$ *and finite moments of all order, c.f. Example 1.3.68). Deduce that the inversion formula*
$$h(y) = \lim_{n\to\infty} (-1)^{n-1} \frac{(n/y)^n}{(n-1)!} L_h^{(n-1)}(n/y) \,,$$
holds for any $y > 0$.

The next application of Borel-Cantelli I provides our first strong law of large numbers.

Proposition 2.2.16. *Suppose* $\mathbf{E}[Z_n^2] \leq C$ *for some* $C < \infty$ *and all* n. *Then,* $n^{-1}Z_n \overset{a.s.}{\to} 0$ *as* $n \to \infty$.

PROOF. Fixing $\delta > 0$ let $A_k = \{\omega : |k^{-1}Z_k(\omega)| > \delta\}$ for $k = 1, 2, \ldots$. Then, by Chebyshev's inequality and our assumption,

$$\mathbf{P}(A_k) = \mathbf{P}(\{\omega : |Z_k(\omega)| \geq k\delta\}) \leq \frac{\mathbf{E}(Z_k^2)}{(k\delta)^2} \leq \frac{C}{\delta^2}k^{-2}\,.$$

Since $\sum_k k^{-2} < \infty$, it follows by Borel Cantelli I that $\mathbf{P}(A^\infty) = 0$, where $A^\infty = \{\omega : |k^{-1}Z_k(\omega)| > \delta$ for infinitely many values of $k\}$. Hence, for any fixed $\delta > 0$, with probability one $k^{-1}|Z_k(\omega)| \leq \delta$ for all large enough k, that is, $\limsup_{n\to\infty} n^{-1}|Z_n(\omega)| \leq \delta$ a.s. Considering a sequence $\delta_m \downarrow 0$ we conclude that $n^{-1}Z_n \to 0$ for $n \to \infty$ and a.e. ω. □

Exercise 2.2.17. *Let* $S_n = \sum\limits_{l=1}^{n} X_l$, *where* $\{X_i\}$ *are i.i.d. random variables with* $\mathbf{E}X_1 = 0$ *and* $\mathbf{E}X_1^4 < \infty$.

(a) *Show that* $n^{-1}S_n \overset{a.s.}{\to} 0$.
Hint: *Verify that Proposition 2.2.16 applies for* $Z_n = n^{-1}S_n^2$.

(b) *Show that* $n^{-1}D_n \overset{a.s.}{\to} 0$ *where* D_n *denotes the number of distinct integers among* $\{\xi_k, k \leq n\}$ *and* $\{\xi_k\}$ *are i.i.d. integer valued random variables.*
Hint: $D_n \leq 2M + \sum_{k=1}^n I_{|\xi_k|\geq M}$.

In contrast, here is an example where the empirical averages of integrable, zero mean independent variables do not converge to zero. Of course, the trick is to have non-identical distributions, with the bulk of the probability drifting to negative one.

Exercise 2.2.18. *Suppose* X_i *are mutually independent random variables such that* $\mathbf{P}(X_n = n^2 - 1) = 1 - \mathbf{P}(X_n = -1) = n^{-2}$ *for* $n = 1, 2, \ldots$. *Show that* $\mathbf{E}X_n = 0$, *for all* n, *while* $n^{-1}\sum_{i=1}^n X_i \overset{a.s.}{\to} -1$ *for* $n \to \infty$.

Next we have few other applications of Borel-Cantelli I, starting with some additional properties of convergence a.s.

Exercise 2.2.19. *Show that for any R.V.* X_n

(a) $X_n \overset{a.s.}{\to} 0$ *if and only if* $\mathbf{P}(|X_n| > \varepsilon$ *i.o.* $) = 0$ *for each* $\varepsilon > 0$.

(b) *There exist non-random constants* $b_n \uparrow \infty$ *such that* $X_n/b_n \overset{a.s.}{\to} 0$.

Exercise 2.2.20. *Show that if* $W_n > 0$ *and* $\mathbf{E}W_n \leq 1$ *for every* n, *then almost surely,*

$$\limsup_{n\to\infty} n^{-1}\log W_n \leq 0\,.$$

Our next example demonstrates how Borel-Cantelli I is typically applied in the study of the asymptotic growth of running maxima of random variables.

Example 2.2.21 (HEAD RUNS)**.** *Let* $\{X_k, k \in \mathbf{Z}\}$ *be a two-sided sequence of i.i.d.* $\{0,1\}$*-valued random variables, with* $\mathbf{P}(X_1 = 1) = \mathbf{P}(X_1 = 0) = 1/2$. *With* $\ell_m = \max\{i : X_{m-i+1} = \cdots = X_m = 1\}$ *denoting the length of the run of* 1*'s going* *backwards from time m, we are interested in the asymptotics of the longest such run during $1, 2, \ldots, n$, that is,*

$$\begin{aligned} L_n &= \max\{\ell_m : m = 1, \ldots, n\} \\ &= \max\{m - k : X_{k+1} = \cdots = X_m = 1 \text{ for some } m = 1, \ldots, n\} . \end{aligned}$$

Noting that $\ell_m + 1$ has a geometric distribution of success probability $p = 1/2$, we deduce by an application of Borel-Cantelli I that for each $\varepsilon > 0$, with probability one, $\ell_n \leq (1+\varepsilon)\log_2 n$ for all n large enough. Hence, on the same set of probability one, we have $N = N(\omega)$ finite such that $L_n \leq \max(L_N, (1+\varepsilon)\log_2 n)$ for all $n \geq N$. Dividing by $\log_2 n$ and considering $n \to \infty$ followed by $\varepsilon_k \downarrow 0$, this implies that

$$\limsup_{n\to\infty} \frac{L_n}{\log_2 n} \overset{a.s.}{\leq} 1 .$$

For each fixed $\varepsilon > 0$ let $A_n = \{L_n < k_n\}$ for $k_n = [(1-\varepsilon)\log_2 n]$. Noting that

$$A_n \subseteq \bigcap_{i=1}^{m_n} B_i^c ,$$

for $m_n = [n/k_n]$ and the independent events $B_i = \{X_{(i-1)k_n+1} = \cdots = X_{ik_n} = 1\}$, yields a bound of the form $\mathbf{P}(A_n) \leq \exp(-n^\varepsilon/(2\log_2 n))$ for all n large enough (c.f. [**Dur10**, Example 2.3.3] *for details). Since $\sum_n \mathbf{P}(A_n) < \infty$, we have that*

$$\liminf_{n\to\infty} \frac{L_n}{\log_2 n} \overset{a.s.}{\geq} 1$$

by yet another application of Borel-Cantelli I, followed by $\varepsilon_k \downarrow 0$. We thus conclude that

$$\frac{L_n}{\log_2 n} \overset{a.s.}{\to} 1 .$$

The next exercise combines both Borel-Cantelli lemmas to provide the 0-1 law for another problem about head runs.

Exercise 2.2.22. *Let $\{X_k\}$ be a sequence of i.i.d. $\{0,1\}$-valued random variables, with $\mathbf{P}(X_1 = 1) = p$ and $\mathbf{P}(X_1 = 0) = 1 - p$. Let A_k be the event that $X_m = \cdots = X_{m+k-1} = 1$ for some $2^k \leq m \leq 2^{k+1} - k$. Show that $\mathbf{P}(A_k \text{ i.o. }) = 1$ if $p \geq 1/2$ and $\mathbf{P}(A_k \text{ i.o. }) = 0$ if $p < 1/2$.*
Hint: *When $p \geq 1/2$ consider only $m = 2^k + (i-1)k$ for $i = 0, \ldots, [2^k/k]$.*

Here are a few direct applications of the second Borel-Cantelli lemma.

Exercise 2.2.23. *Suppose that $\{Z_k\}$ are i.i.d. random variables such that $\mathbf{P}(Z_1 = z) < 1$ for any $z \in \mathbb{R}$.*

(a) *Show that $\mathbf{P}(Z_k$ converges for $k \to \infty) = 0$.*
(b) *Determine the values of $\limsup_{n\to\infty}(Z_n/\log n)$ and $\liminf_{n\to\infty}(Z_n/\log n)$ in case Z_k has the* exponential distribution *(of parameter $\lambda = 1$).*

After deriving the classical bounds on the tail of the normal distribution, you use both Borel-Cantelli lemmas in bounding the fluctuations of the sums of i.i.d. standard normal variables.

Exercise 2.2.24. *Let $\{G_i\}$ be i.i.d. standard normal random variables.*

(a) *Show that for any* $x > 0$,

$$(x^{-1} - x^{-3})e^{-x^2/2} \le \int_x^\infty e^{-y^2/2} dy \le x^{-1} e^{-x^2/2}\,.$$

Many texts prove these estimates, for example see [**Dur10**, Theorem 1.2.3].

(b) *Show that, with probability one,*

$$\limsup_{n\to\infty} \frac{G_n}{\sqrt{2\log n}} = 1\,.$$

(c) *Let* $S_n = G_1 + \cdots + G_n$. *Recall that* $n^{-1/2} S_n$ *has the standard normal distribution. Show that*

$$\mathbf{P}(|S_n| < 2\sqrt{n \log n},\ \textit{ev.}\ ) = 1\,.$$

Remark. Ignoring the dependence between the elements of the sequence S_k, the bound in part (c) of the preceding exercise is not tight. The definite result here is the *law of the iterated logarithm* (in short LIL), which states that when the i.i.d. summands are of zero mean and variance one,

$$\mathbf{P}(\limsup_{n\to\infty} \frac{S_n}{\sqrt{2n \log\log n}} = 1) = 1\,. \tag{2.2.1}$$

We defer the derivation of (2.2.1) to Theorem 9.2.29, building on a similar LIL for the Brownian motion (but, see [**Bil95**, Theorem 9.5] for a direct proof of (2.2.1), using both Borel-Cantelli lemmas).

The next exercise relates explicit integrability conditions for i.i.d. random variables to the asymptotics of their running maxima.

Exercise 2.2.25. *Consider possibly* $\overline{\mathbb{R}}$*-valued, i.i.d. random variables* $\{Y_i\}$ *and their running maxima* $M_n = \max_{k\le n} Y_k$.

(a) *Using (2.3.4) if needed, show that* $\mathbf{P}(|Y_n| > n \;\; i.o.\ ) = 0$ *if and only if* $\mathbf{E}[|Y_1|] < \infty$.
(b) *Show that* $n^{-1} Y_n \overset{a.s.}{\to} 0$ *if and only if* $\mathbf{E}[|Y_1|] < \infty$.
(c) *Show that* $n^{-1} M_n \overset{a.s.}{\to} 0$ *if and only if* $\mathbf{E}[(Y_1)_+] < \infty$ *and* $\mathbf{P}(Y_1 > -\infty) > 0$.
(d) *Show that* $n^{-1} M_n \overset{p}{\to} 0$ *if and only if* $n\mathbf{P}(Y_1 > n) \to 0$ *and* $\mathbf{P}(Y_1 > -\infty) > 0$.
(e) *Show that* $n^{-1} Y_n \overset{p}{\to} 0$ *if and only if* $\mathbf{P}(|Y_1| < \infty) = 1$.

In the following exercise, you combine Borel Cantelli I and the variance computation of Lemma 2.1.2 to improve upon Borel Cantelli II.

Exercise 2.2.26. *Suppose* $\sum_{n=1}^\infty \mathbf{P}(A_n) = \infty$ *for pairwise independent events* $\{A_i\}$. *Let* $S_n = \sum_{i=1}^n I_{A_i}$ *be the number of events occurring among the first* n.

(a) *Prove that* $\mathsf{Var}(S_n) \le \mathbf{E}(S_n)$ *and deduce from it that* $S_n/\mathbf{E}(S_n) \overset{p}{\to} 1$.
(b) *Applying Borel-Cantelli I show that* $S_{n_k}/\mathbf{E}(S_{n_k}) \overset{a.s.}{\to} 1$ *as* $k \to \infty$, *where* $n_k = \inf\{n : \mathbf{E}(S_n) \ge k^2\}$.
(c) *Show that* $\mathbf{E}(S_{n_{k+1}})/\mathbf{E}(S_{n_k}) \to 1$ *and since* $n \mapsto S_n$ *is non-decreasing, deduce that* $S_n/\mathbf{E}(S_n) \overset{a.s.}{\to} 1$.

Remark. Borel-Cantelli II is the a.s. convergence $S_n \to \infty$ for $n \to \infty$, which is a consequence of part (c) of the preceding exercise (since $\mathbf{E}S_n \to \infty$).

We conclude this section with an example in which the asymptotic rate of growth of random variables of interest is obtained by an application of Exercise 2.2.26.

Example 2.2.27 (Record values)**.** *Let $\{X_i\}$ be a sequence of i.i.d. random variables with a continuous distribution function $F_X(x)$. The event $A_k = \{X_k > X_j, j = 1, \ldots, k-1\}$ represents the occurrence of a record at the k instance (for example, think of X_k as an athlete's kth distance jump). We are interested in the asymptotics of the count $R_n = \sum\limits_{i=1}^{n} I_{A_i}$ of record events during the first n instances. Because of the continuity of F_X we know that a.s. the values of X_i, $i = 1, 2, \ldots$ are distinct. Further, rearranging the random variables $X_1, X_2, \ldots, X_n$ in a decreasing order induces a random permutation π_n on $\{1, 2, \ldots, n\}$, where all $n!$ possible permutations are equally likely. From this it follows that $\mathbf{P}(A_k) = \mathbf{P}(\pi_k(k) = 1) = 1/k$, and though definitely not obvious at first sight, the events A_k are mutually independent (see* [**Dur10**, Example 2.3.2] *for details). So, $\mathbf{E}R_n = \log n + \gamma_n$ where γ_n is between zero and one, and from Exercise 2.2.26 we deduce that $(\log n)^{-1} R_n \overset{a.s.}{\to} 1$ as $n \to \infty$. Note that this result is independent of the law of X, as long as the distribution function F_X is continuous.*

2.3. Strong law of large numbers

In Corollary 2.1.14 we got the classical weak law of large numbers, namely, the convergence in probability of the empirical averages $n^{-1} \sum_{i=1}^{n} X_i$ of i.i.d. integrable random variables X_i to the mean $\mathbf{E}X_1$. Assuming in addition that $\mathbf{E}X_1^4 < \infty$, you used Borel-Cantelli I in Exercise 2.2.17 en-route to the corresponding strong law of large numbers, that is, replacing the convergence in probability with the stronger notion of convergence almost surely.

We provide here two approaches to the strong law of large numbers, both of which get rid of the unnecessary finite moment assumptions. Subsection 2.3.1 follows Etemadi's (1981) direct proof of this result via the subsequence method. Subsection 2.3.2 deals in a more systematic way with the convergence of random series, yielding the strong law of large numbers as one of its consequences.

2.3.1. The subsequence method.

Etemadi's key observation is that it essentially suffices to consider non-negative X_i, for which upon proving the a.s. convergence along a not too sparse subsequence n_l, the interpolation to the whole sequence can be done by the monotonicity of $n \mapsto \sum^n X_i$. This is an example of a general approach to a.s. convergence, called the *subsequence method*, which you have already encountered in Exercise 2.2.26.

We thus start with the strong law for integrable, non-negative variables.

Proposition 2.3.1. *Let $S_n = \sum_{i=1}^{n} X_i$ for non-negative, pairwise independent and identically distributed, integrable random variables $\{X_i\}$. Then, $n^{-1} S_n \overset{a.s.}{\to} \mathbf{E}X_1$ as $n \to \infty$.*

Proof. The proof progresses along the themes of Section 2.1, starting with the truncation $\overline{X}_k = X_k I_{|X_k| \le k}$ and its corresponding sums $\overline{S}_n = \sum_{i=1}^{n} \overline{X}_i$.

Since $\{X_i\}$ are identically distributed and $x \mapsto \mathbf{P}(|X_1| > x)$ is non-increasing, we have that

$$\sum_{k=1}^{\infty} \mathbf{P}(X_k \neq \overline{X}_k) = \sum_{k=1}^{\infty} \mathbf{P}(|X_1| > k) \leq \int_0^{\infty} \mathbf{P}(|X_1| > x)dx = \mathbf{E}|X_1| < \infty$$

(see part (a) of Lemma 1.4.31 for the rightmost identity and recall our assumption that X_1 is integrable). Thus, by Borel-Cantelli I, with probability one, $X_k(\omega) = \overline{X}_k(\omega)$ for all but finitely many k's, in which case necessarily $\sup_n |S_n(\omega) - \overline{S}_n(\omega)|$ is finite. This shows that $n^{-1}(S_n - \overline{S}_n) \stackrel{a.s.}{\to} 0$, whereby it suffices to prove that $n^{-1}\overline{S}_n \stackrel{a.s.}{\to} \mathbf{E}X_1$.

To this end, we next show that it suffices to prove the following lemma about almost sure convergence of $\overline{S}_n$ along suitably chosen subsequences.

Lemma 2.3.2. *Fixing $\alpha > 1$ let $n_l = [\alpha^l]$. Under the conditions of the proposition, $n_l^{-1}(\overline{S}_{n_l} - \mathbf{E}\overline{S}_{n_l}) \stackrel{a.s.}{\to} 0$ as $l \to \infty$.*

By dominated convergence, $\mathbf{E}[X_1 I_{|X_1| \leq k}] \to \mathbf{E}X_1$ as $k \to \infty$, and consequently, as $n \to \infty$,

$$\frac{1}{n}\mathbf{E}\overline{S}_n = \frac{1}{n}\sum_{k=1}^{n} \mathbf{E}\overline{X}_k = \frac{1}{n}\sum_{k=1}^{n} \mathbf{E}[X_1 I_{|X_1| \leq k}] \to \mathbf{E}X_1$$

(we have used here the consistency of Cesáro averages, c.f. Exercise 1.3.52 for an integral version). Thus, assuming that Lemma 2.3.2 holds, we have that $n_l^{-1}\overline{S}_{n_l} \stackrel{a.s.}{\to} \mathbf{E}X_1$ when $l \to \infty$, for each $\alpha > 1$.

We complete the proof of the proposition by interpolating from the subsequences $n_l = [\alpha^l]$ to the whole sequence. To this end, fix $\alpha > 1$. Since $n \mapsto \overline{S}_n$ is non-decreasing, we have for all $\omega \in \Omega$ and any $n \in [n_l, n_{l+1}]$,

$$\frac{n_l}{n_{l+1}} \frac{\overline{S}_{n_l}(\omega)}{n_l} \leq \frac{\overline{S}_n(\omega)}{n} \leq \frac{n_{l+1}}{n_l} \frac{\overline{S}_{n_{l+1}}(\omega)}{n_{l+1}}$$

With $n_l/n_{l+1} \to 1/\alpha$ for $l \to \infty$, the a.s. convergence of $m^{-1}\overline{S}_m$ along the subsequence $m = n_l$ implies that the event

$$A_\alpha := \{\omega : \frac{1}{\alpha}\mathbf{E}X_1 \leq \liminf_{n\to\infty} \frac{\overline{S}_n(\omega)}{n} \leq \limsup_{n\to\infty} \frac{\overline{S}_n(\omega)}{n} \leq \alpha\mathbf{E}X_1\}\,,$$

has probability one. Consequently, taking $\alpha_m \downarrow 1$, we deduce that the event $B := \bigcap_m A_{\alpha_m}$ also has probability one, and further, $n^{-1}\overline{S}_n(\omega) \to \mathbf{E}X_1$ for each $\omega \in B$. We thus deduce that $n^{-1}\overline{S}_n \stackrel{a.s.}{\to} \mathbf{E}X_1$, as needed to complete the proof of the proposition. □

Remark. The monotonicity of certain random variables (here $n \mapsto \overline{S}_n$), is crucial to the successful application of the subsequence method. The subsequence n_l for which we need a direct proof of convergence is completely determined by the scaling function b_n^{-1} applied to this monotone sequence (here $b_n = n$); we need $b_{n_{l+1}}/b_{n_l} \to \alpha$, which should be arbitrarily close to 1. For example, same subsequences $n_l = [\alpha^l]$ are to be used whenever b_n is roughly of a polynomial growth in n, while even $n_l = (l!)^c$ would work in case $b_n = \log n$.

Likewise, the truncation level is determined by the highest moment of the basic variables which is assumed to be finite. For example, we can take $\overline{X}_k = X_k I_{|X_k| \leq k^p}$ for any $p > 0$ such that $\mathbf{E}|X_1|^{1/p} < \infty$.

PROOF OF LEMMA 2.3.2. Note that $\mathbf{E}[\overline{X}_k^2]$ is non-decreasing in k. Further, $\overline{X}_k$ are pairwise independent, hence uncorrelated, so by Lemma 2.1.2,

$$\mathsf{Var}(\overline{S}_n) = \sum_{k=1}^n \mathsf{Var}(\overline{X}_k) \le \sum_{k=1}^n \mathbf{E}[\overline{X}_k^2] \le n\mathbf{E}[\overline{X}_n^2] = n\mathbf{E}[X_1^2 I_{|X_1|\le n}]\,.$$

Combining this with Chebychev's inequality yield the bound

$$\mathbf{P}(|\overline{S}_n - \mathbf{E}\overline{S}_n| \ge \varepsilon n) \le (\varepsilon n)^{-2}\,\mathsf{Var}(\overline{S}_n) \le \varepsilon^{-2} n^{-1}\mathbf{E}[X_1^2 I_{|X_1|\le n}]\,,$$

for any $\varepsilon > 0$. Applying Borel-Cantelli I for the events $A_l = \{|\overline{S}_{n_l} - \mathbf{E}\overline{S}_{n_l}| \ge \varepsilon n_l\}$, followed by $\varepsilon_m \downarrow 0$, we get the a.s. convergence to zero of $n^{-1}|\overline{S}_n - \mathbf{E}\overline{S}_n|$ along any subsequence n_l for which

$$\sum_{l=1}^\infty n_l^{-1}\mathbf{E}[X_1^2 I_{|X_1|\le n_l}] = \mathbf{E}[X_1^2 \sum_{l=1}^\infty n_l^{-1} I_{|X_1|\le n_l}] < \infty$$

(the latter identity is a special case of Exercise 1.3.40). Since $\mathbf{E}|X_1| < \infty$, it thus suffices to show that for $n_l = [\alpha^l]$ and any $x > 0$,

$$u(x) := \sum_{l=1}^\infty n_l^{-1} I_{x\le n_l} \le c x^{-1}\,, \tag{2.3.1}$$

where $c = 2\alpha/(\alpha - 1) < \infty$. To establish (2.3.1) fix $\alpha > 1$ and $x > 0$, setting $L = \min\{l \ge 1 : n_l \ge x\}$. Then, $\alpha^L \ge x$, and since $[y] \ge y/2$ for all $y \ge 1$,

$$u(x) = \sum_{l=L}^\infty n_l^{-1} \le 2\sum_{l=L}^\infty \alpha^{-l} = c\alpha^{-L} \le c x^{-1}\,.$$

So, we have established (2.3.1) and hence completed the proof of the lemma. □

As already promised, it is not hard to extend the scope of the strong law of large numbers beyond integrable and non-negative random variables.

Theorem 2.3.3 (STRONG LAW OF LARGE NUMBERS). *Let $S_n = \sum_{i=1}^n X_i$ for pairwise independent and identically distributed random variables $\{X_i\}$, such that either $\mathbf{E}[(X_1)_+]$ is finite or $\mathbf{E}[(X_1)_-]$ is finite. Then, $n^{-1}S_n \overset{a.s.}{\to} \mathbf{E}X_1$ as $n \to \infty$.*

PROOF. First consider non-negative X_i. The case of $\mathbf{E}X_1 < \infty$ has already been dealt with in Proposition 2.3.1. In case $\mathbf{E}X_1 = \infty$, consider $S_n^{(m)} = \sum_{i=1}^n X_i^{(m)}$ for the bounded, non-negative, pairwise independent and identically distributed random variables $X_i^{(m)} = \min(X_i, m) \le X_i$. Since Proposition 2.3.1 applies for $\{X_i^{(m)}\}$, it follows that a.s. for any fixed $m < \infty$,

$$\liminf_{n\to\infty} n^{-1}S_n \ge \liminf_{n\to\infty} n^{-1}S_n^{(m)} = \mathbf{E}X_1^{(m)} = \mathbf{E}\min(X_1, m)\,. \tag{2.3.2}$$

Taking $m \uparrow \infty$, by monotone convergence $\mathbf{E}\min(X_1, m) \uparrow \mathbf{E}X_1 = \infty$, so (2.3.2) results with $n^{-1}S_n \to \infty$ a.s.

Turning to the general case, we have the decomposition $X_i = (X_i)_+ - (X_i)_-$ of each random variable to its positive and negative parts, with

$$n^{-1}S_n = n^{-1}\sum_{i=1}^n (X_i)_+ - n^{-1}\sum_{i=1}^n (X_i)_- \tag{2.3.3}$$

Since $(X_i)_+$ are non-negative, pairwise independent and identically distributed, it follows that $n^{-1}\sum_{i=1}^n (X_i)_+ \overset{a.s.}{\to} \mathbf{E}[(X_1)_+]$ as $n \to \infty$. For the same reason, also $n^{-1}\sum_{i=1}^n (X_i)_- \stackrel{a.s.}{\to} \mathbf{E}[(X_1)_-]$. Our assumption that either $\mathbf{E}[(X_1)_+] < \infty$ or $\mathbf{E}[(X_1)_-] < \infty$ implies that $\mathbf{E}X_1 = \mathbf{E}[(X_1)_+] - \mathbf{E}[(X_1)_-]$ is well defined, and in view of (2.3.3) we have the stated a.s. convergence of $n^{-1}S_n$ to $\mathbf{E}X_1$. $\square$

Exercise 2.3.4. *You are to prove now a converse to the strong law of large numbers (for a more general result, due to Feller (1946), see* [**Dur10**, Theorem 2.5.9]*).*

(a) *Let Y denote the integer part of a random variable $Z \geq 0$. Show that $Y = \sum_{n=1}^\infty I_{\{Z \geq n\}}$, and deduce that*

$$\sum_{n=1}^\infty \mathbf{P}(Z \geq n) \leq \mathbf{E}Z \leq 1 + \sum_{n=1}^\infty \mathbf{P}(Z \geq n)\,. \tag{2.3.4}$$

(b) *Suppose $\{X_i\}$ are i.i.d R.V.s with $\mathbf{E}[|X_1|^\alpha] = \infty$ for some $\alpha > 0$. Show that for any $k > 0$,*

$$\sum_{n=1}^\infty \mathbf{P}(|X_n| > kn^{1/\alpha}) = \infty\,,$$

and deduce that a.s. $\limsup_{n\to\infty} n^{-1/\alpha}|X_n| = \infty$.

(c) *Conclude that if $S_n = X_1 + X_2 + \cdots + X_n$, then*

$$\limsup_{n\to\infty} n^{-1/\alpha}|S_n| = \infty, \qquad \text{a.s.}$$

We provide next two classical applications of the strong law of large numbers, the first of which deals with the large sample asymptotics of the empirical distribution function.

Example 2.3.5 (Empirical distribution function)**.** *Let*

$$F_n(x) = n^{-1}\sum_{i=1}^n I_{(-\infty,x]}(X_i)\,,$$

denote the observed fraction of values among the first n variables of the sequence $\{X_i\}$ which do not exceed x. The functions $F_n(\cdot)$ are thus called the empirical distribution functions of this sequence.

For i.i.d. $\{X_i\}$ with distribution function F_X our next result improves the strong law of large numbers by showing that F_n converges uniformly to F_X as $n \to \infty$.

Theorem 2.3.6 (Glivenko-Cantelli)**.** *For i.i.d. $\{X_i\}$ with arbitrary distribution function F_X, as $n \to \infty$,*

$$D_n = \sup_{x\in\mathbb{R}} |F_n(x) - F_X(x)| \stackrel{a.s.}{\to} 0\,.$$

Remark. While outside our scope, we note in passing the Dvoretzky-Kiefer-Wolfowitz inequality that $\mathbf{P}(D_n > \varepsilon) \leq 2\exp(-2n\varepsilon^2)$ for any n and all $\varepsilon > 0$, quantifying the rate of convergence of D_n to zero (see [**DKW56**], or [**Mas90**] for the optimal pre-exponential constant).

Proof. By the right continuity of both $x \mapsto F_n(x)$ and $x \mapsto F_X(x)$ (c.f. Theorem 1.2.37), the value of D_n is unchanged when the supremum over $x \in \mathbb{R}$ is replaced by the one over $x \in \mathbb{Q}$ (the rational numbers). In particular, this shows that each D_n is a random variable (c.f. Theorem 1.2.22).

Applying the strong law of large numbers for the i.i.d. non-negative $I_{(-\infty,x]}(X_i)$ whose expectation is $F_X(x)$, we deduce that $F_n(x) \stackrel{a.s.}{\to} F_X(x)$ for each fixed non-random $x \in \mathbb{R}$. Similarly, considering the strong law of large numbers for the i.i.d. non-negative $I_{(-\infty,x)}(X_i)$ whose expectation is $F_X(x^-)$, we have that $F_n(x^-) \stackrel{a.s.}{\to} F_X(x^-)$ for each fixed non-random $x \in \mathbb{R}$. Consequently, for any fixed $l < \infty$ and $x_{1,l}, \dots, x_{l,l}$ we have that

$$D_{n,l} = \max(\max_{k=1}^{l} |F_n(x_{k,l}) - F_X(x_{k,l})|, \max_{k=1}^{l} |F_n(x_{k,l}^-) - F_X(x_{k,l}^-)|) \stackrel{a.s.}{\to} 0\,,$$

as $n \to \infty$. Choosing $x_{k,l} = \inf\{x : F_X(x) \geq k/(l+1)\}$, we get out of the monotonicity of $x \mapsto F_n(x)$ and $x \mapsto F_X(x)$ that $D_n \leq D_{n,l} + l^{-1}$ (c.f. [**Bil95**, Proof of Theorem 20.6]). Therefore, taking $n \to \infty$ followed by $l \to \infty$ completes the proof of the theorem. □

We turn to our second example, which is about counting processes.

Example 2.3.7 (Renewal theory)**.** *Let $\{\tau_i\}$ be i.i.d. positive, finite random variables and $T_n = \sum_{k=1}^n \tau_k$. Here T_n is interpreted as the time of the n-th occurrence of a given event, with τ_k representing the length of the time interval between the $(k-1)$ occurrence and that of the k-th such occurrence. Associated with T_n is the dual process $N_t = \sup\{n : T_n \leq t\}$ counting the number of occurrences during the time interval $[0,t]$. In the next exercise you are to derive the strong law for the large t asymptotics of $t^{-1}N_t$.*

Exercise 2.3.8. *Consider the setting of Example 2.3.7.*

(a) *By the strong law of large numbers argue that $n^{-1}T_n \stackrel{a.s.}{\to} \mathbf{E}\tau_1$. Then, adopting the convention $\frac{1}{\infty} = 0$, deduce that $t^{-1}N_t \stackrel{a.s.}{\to} 1/\mathbf{E}\tau_1$ for $t \to \infty$.* Hint: *From the definition of N_t it follows that $T_{N_t} \leq t < T_{N_t+1}$ for all $t \geq 0$.*
(b) *Show that $t^{-1}N_t \stackrel{a.s.}{\to} 1/\mathbf{E}\tau_2$ as $t \to \infty$, even if the law of τ_1 is different from that of the i.i.d. $\{\tau_i,\, i \geq 2\}$.*

Here is a strengthening of the preceding result to convergence in L^1.

Exercise 2.3.9. *In the context of Example 2.3.7 fix $\delta > 0$ such that $\mathbf{P}(\tau_1 > \delta) > \delta$ and let $\widetilde{T}_n = \sum_{k=1}^n \widetilde{\tau}_k$ for the i.i.d. random variables $\widetilde{\tau}_i = \delta I_{\{\tau_i > \delta\}}$. Note that $\widetilde{T}_n \leq T_n$ and consequently $N_t \leq \widetilde{N}_t = \sup\{n : \widetilde{T}_n \leq t\}$.*

(a) *Show that $\limsup_{t\to\infty} t^{-2}\mathbf{E}\widetilde{N}_t^2 < \infty$.*
(b) *Deduce that $\{t^{-1}N_t : t \geq 1\}$ is uniformly integrable (see Exercise 1.3.54), and conclude that $t^{-1}\mathbf{E}N_t \to 1/\mathbf{E}\tau_1$ when $t \to \infty$.*

The next exercise deals with an elaboration over Example 2.3.7.

Exercise 2.3.10. *For $i = 1, 2, \dots$ the ith light bulb burns for an amount of time τ_i and then remains burned out for time s_i before being replaced by the $(i+1)$th bulb. Let R_t denote the fraction of time during $[0,t]$ in which we have a working light. Assuming that the two sequences $\{\tau_i\}$ and $\{s_i\}$ are independent, each consisting of i.i.d. positive and integrable random variables, show that $R_t \stackrel{a.s.}{\to} \mathbf{E}\tau_1/(\mathbf{E}\tau_1 + \mathbf{E}s_1)$.*

Here is another exercise, dealing with sampling "at times of heads" in independent fair coin tosses, from a non-random bounded sequence of weights $v(l)$, the averages of which converge.

Exercise 2.3.11. *For a sequence $\{B_i\}$ of i.i.d. Bernoulli random variables of parameter $p = 1/2$, let T_n be the time that the corresponding partial sums reach level n. That is, $T_n = \inf\{k : \sum_{i=1}^k B_i \ge n\}$, for $n = 1, 2, \ldots$.*

(a) *Show that $n^{-1} T_n \stackrel{a.s.}{\to} 2$ as $n \to \infty$.*
(b) *Given non-negative, non-random $\{v(k)\}$ show that $k^{-1} \sum_{i=1}^k v(T_i) \stackrel{a.s.}{\to} s$ as $k \to \infty$, for some non-random s, if and only if $n^{-1} \sum_{l=1}^n v(l) B_l \stackrel{a.s.}{\to} s/2$ as $n \to \infty$.*
(c) *Deduce that if $n^{-1} \sum_{l=1}^n v(l)^2$ is bounded and $n^{-1} \sum_{l=1}^n v(l) \to s$ as $n \to \infty$, then $k^{-1} \sum_{i=1}^k v(T_i) \stackrel{a.s.}{\to} s$ as $k \to \infty$.*

Hint: *For part (c) consider first the limit of $n^{-1} \sum_{l=1}^n v(l)(B_l - 0.5)$ as $n \to \infty$.*

We proceed with a few additional applications of the strong law of large numbers, first to a problem of *universal hypothesis testing*, then an application involving stochastic geometry, and finally one motivated by investment science.

Exercise 2.3.12. *Consider i.i.d. $[0,1]$-valued random variables $\{X_k\}$.*

(a) *Find Borel measurable functions $f_n : [0,1]^n \mapsto \{0,1\}$, which are independent of the law of X_k, such that $f_n(X_1, X_2, \ldots, X_n) \stackrel{a.s.}{\to} 0$ whenever $\mathbf{E}X_1 < 1/2$ and $f_n(X_1, X_2, \ldots, X_n) \stackrel{a.s.}{\to} 1$ whenever $\mathbf{E}X_1 > 1/2$.*
(b) *Modify your answer to assure that $f_n(X_1, X_2, \ldots, X_n) \stackrel{a.s.}{\to} 1$ also in case $\mathbf{E}X_1 = 1/2$.*

Exercise 2.3.13. *Let $\{U_n\}$ be i.i.d. random vectors, each uniformly distributed on the unit ball $\{u \in \mathbb{R}^2 : |u| \le 1\}$. Consider the $\mathbb{R}^2$-valued random vectors $X_n = |X_{n-1}| U_n$, $n = 1, 2, \ldots$ starting at a non-random, non-zero vector X_0 (that is, each point is uniformly chosen in a ball centered at the origin and whose radius is the distance from the origin to the previously chosen point). Show that $n^{-1} \log |X_n| \stackrel{a.s.}{\to} -1/2$ as $n \to \infty$.*

Exercise 2.3.14. *Let $\{V_n\}$ be i.i.d. non-negative random variables. Fixing $r > 0$ and $q \in (0,1]$, consider the sequence $W_0 = 1$ and $W_n = (qr + (1-q)V_n) W_{n-1}$, $n = 1, 2, \ldots$. A motivating example is of W_n recording the relative growth of a portfolio where a constant fraction q of one's wealth is re-invested each year in a risk-less asset that grows by r per year, with the remainder re-invested in a risky asset whose annual growth factors are the random V_n.*

(a) *Show that $n^{-1} \log W_n \stackrel{a.s.}{\to} w(q)$, for $w(q) = \mathbf{E} \log(qr + (1-q)V_1)$.*
(b) *Show that $q \mapsto w(q)$ is concave on $(0,1]$.*
(c) *Using Jensen's inequality show that $w(q) \le w(1)$ in case $\mathbf{E}V_1 \le r$. Further, show that if $\mathbf{E}V_1^{-1} \le r^{-1}$, then the almost sure convergence applies also for $q = 0$ and that $w(q) \le w(0)$.*
(d) *Assuming that $\mathbf{E}V_1^2 < \infty$ and $\mathbf{E}V_1^{-2} < \infty$ show that $\sup\{w(q) : q \in [0,1]\}$ is finite, and further that the maximum of $w(q)$ is obtained at some $q^* \in (0,1)$ when $\mathbf{E}V_1 > r > 1/\mathbf{E}V_1^{-1}$. Interpret your results in terms of the preceding investment example.*

Hint: *Consider small $q > 0$ and small $1 - q > 0$ and recall that $\log(1+x) \ge x - x^2/2$ for any $x \ge 0$.*

We conclude this subsection with another example where an almost sure convergence is derived by the subsequence method.

Exercise 2.3.15. *Show that almost surely* $\limsup_{n\to\infty} \log Z_n / \log \mathbf{E} Z_n \le 1$ *for any positive, non-decreasing random variables* Z_n *such that* $Z_n \xrightarrow{a.s.} \infty$.

2.3.2. Convergence of random series. A second approach to the strong law of large numbers is based on studying the convergence of random series. The key tool in this approach is Kolmogorov's maximal inequality, which we prove next.

Proposition 2.3.16 (Kolmogorov's maximal inequality)**.** *The random variables* $Y_1, \ldots, Y_n$ *are mutually independent, with* $\mathbf{E} Y_l^2 < \infty$ *and* $\mathbf{E} Y_l = 0$ *for* $l = 1, \ldots, n$. *Then, for* $Z_k = Y_1 + \cdots + Y_k$ *and any* $z > 0$,

$$z^2 \mathbf{P}(\max_{1\le k\le n} |Z_k| \ge z) \le \mathsf{Var}(Z_n) \,. \tag{2.3.5}$$

Remark. Chebyshev's inequality gives only $z^2 \mathbf{P}(|Z_n| \ge z) \le \mathsf{Var}(Z_n)$ which is significantly weaker and insufficient for our current goals.

Proof. Fixing $z > 0$ we decompose the event $A = \{\max_{1\le k\le n} |Z_k| \ge z\}$ according to the minimal index k for which $|Z_k| \ge z$. That is, A is the union of the disjoint events $A_k = \{|Z_k| \ge z > |Z_j|,\ j = 1, \ldots, k-1\}$ over $1 \le k \le n$. Obviously,

$$z^2 \mathbf{P}(A) = \sum_{k=1}^n z^2 \mathbf{P}(A_k) \le \sum_{k=1}^n \mathbf{E}[Z_k^2; A_k] \,, \tag{2.3.6}$$

since $Z_k^2 \ge z^2$ on A_k. Further, $\mathbf{E} Z_n = 0$ and A_k are disjoint, so

$$\mathsf{Var}(Z_n) = \mathbf{E} Z_n^2 \ge \sum_{k=1}^n \mathbf{E}[Z_n^2; A_k] \,. \tag{2.3.7}$$

It suffices to show that $\mathbf{E}[(Z_n - Z_k) Z_k; A_k] = 0$ for any $1 \le k \le n$, since then

$$\begin{aligned} \mathbf{E}[Z_n^2; A_k] - \mathbf{E}[Z_k^2; A_k] &= \mathbf{E}[(Z_n - Z_k)^2; A_k] + 2\mathbf{E}[(Z_n - Z_k) Z_k; A_k] \\ &= \mathbf{E}[(Z_n - Z_k)^2; A_k] \ge 0 \,, \end{aligned}$$

and (2.3.5) follows by comparing (2.3.6) and (2.3.7). Since $Z_k I_{A_k}$ can be represented as a non-random Borel function of $(Y_1, \ldots, Y_k)$, it follows that $Z_k I_{A_k}$ is measurable on $\sigma(Y_1, \ldots, Y_k)$. Consequently, for fixed k and $l > k$ the variables Y_l and $Z_k I_{A_k}$ are independent, hence uncorrelated. Further $\mathbf{E} Y_l = 0$, so

$$\mathbf{E}[(Z_n - Z_k) Z_k; A_k] = \sum_{l=k+1}^n \mathbf{E}[Y_l Z_k I_{A_k}] = \sum_{l=k+1}^n \mathbf{E}(Y_l) \mathbf{E}(Z_k I_{A_k}) = 0 \,,$$

completing the proof of Kolmogorov's inequality. □

Equipped with Kolmogorov's inequality, we provide an easy to check sufficient condition for the convergence of random series of independent R.V.

Theorem 2.3.17. *Suppose* $\{X_i\}$ *are independent random variables with* $\mathsf{Var}(X_i) < \infty$ *and* $\mathbf{E} X_i = 0$. *If* $\sum_n \mathsf{Var}(X_n) < \infty$ *then w.p.1. the random series* $\sum_n X_n(\omega)$ *converges (that is, the sequence* $S_n(\omega) = \sum_{k=1}^n X_k(\omega)$ *has a finite limit* $S_\infty(\omega)$*).*

Proof. Applying Kolmogorov's maximal inequality for the independent variables $Y_l = X_{l+r}$, we have that for any $\varepsilon > 0$ and positive integers r and n,

$$\mathbf{P}(\max_{r\le k\le r+n} |S_k - S_r| \ge \varepsilon) \le \varepsilon^{-2} \mathsf{Var}(S_{r+n} - S_r) = \varepsilon^{-2} \sum_{l=r+1}^{r+n} \mathsf{Var}(X_l) \,.$$

Taking $n \to \infty$, we get by continuity from below of $\mathbf{P}$ that

$$\mathbf{P}(\sup_{k\geq r} |S_k - S_r| \geq \varepsilon) \leq \varepsilon^{-2} \sum_{l=r+1}^{\infty} \mathsf{Var}(X_l)$$

By our assumption that $\sum_n \mathsf{Var}(X_n)$ is finite, it follows that $\sum_{l>r} \mathsf{Var}(X_l) \to 0$ as $r \to \infty$. Hence, if we let $T_r = \sup_{n,m\geq r} |S_n - S_m|$, then for any $\varepsilon > 0$,

$$\mathbf{P}(T_r \geq 2\varepsilon) \leq \mathbf{P}(\sup_{k\geq r} |S_k - S_r| \geq \varepsilon) \to 0$$

as $r \to \infty$. Further, $r \mapsto T_r(\omega)$ is non-increasing, hence,

$$\mathbf{P}(\limsup_{M\to\infty} T_M \geq 2\varepsilon) = \mathbf{P}(\inf_M T_M \geq 2\varepsilon) \leq \mathbf{P}(T_r \geq 2\varepsilon) \to 0\,.$$

That is, $T_M(\omega) \stackrel{a.s.}{\to} 0$ for $M \to \infty$. By definition, the convergence to zero of $T_M(\omega)$ is the statement that $S_n(\omega)$ is a Cauchy sequence. Since every Cauchy sequence in $\mathbb{R}$ converges to a finite limit, we have the stated a.s. convergence of $S_n(\omega)$. $\square$

We next provide some applications of Theorem 2.3.17.

Example 2.3.18. *Considering non-random a_n such that $\sum_n a_n^2 < \infty$ and independent Bernoulli variables B_n of parameter $p = 1/2$, Theorem 2.3.17 tells us that $\sum_n (-1)^{B_n} a_n$ converges with probability one. That is, when the signs in $\sum_n \pm a_n$ are chosen on the toss of a fair coin, the series almost always converges (though quite possibly $\sum_n |a_n| = \infty$).*

Exercise 2.3.19. *Consider the* record events *A_k of Example 2.2.27.*

(a) *Verify that the events A_k are mutually independent with $\mathbf{P}(A_k) = 1/k$.*
(b) *Show that the random series $\sum_{n\geq 2}(I_{A_n} - 1/n)/\log n$ converges almost surely and deduce that $(\log n)^{-1} R_n \stackrel{a.s.}{\to} 1$ as $n \to \infty$.*
(c) *Provide a counterexample to the preceding in case the distribution function $F_X(x)$ is not continuous.*

The link between convergence of random series and the strong law of large numbers is the following classical analysis lemma.

Lemma 2.3.20 (Kronecker's lemma)**.** *Consider two sequences of real numbers $\{x_n\}$ and $\{b_n\}$ where $b_n > 0$ and $b_n \uparrow \infty$. If $\sum_n x_n/b_n$ converges, then $s_n/b_n \to 0$ for $s_n = x_1 + \cdots + x_n$.*

Proof. Let $u_n = \sum_{k=1}^n (x_k/b_k)$ which by assumption converges to a finite limit denoted u_∞. Setting $u_0 = b_0 = 0$, "summation by parts" yields the identity,

$$s_n = \sum_{k=1}^n b_k(u_k - u_{k-1}) = b_n u_n - \sum_{k=1}^n (b_k - b_{k-1})u_{k-1}\,.$$

Since $u_n \to u_\infty$ and $b_n \uparrow \infty$, the Cesáro averages $b_n^{-1}\sum_{k=1}^n (b_k - b_{k-1})u_{k-1}$ also converge to u_∞. Consequently, $s_n/b_n \to u_\infty - u_\infty = 0$. $\square$

Theorem 2.3.17 provides an alternative proof for the strong law of large numbers of Theorem 2.3.3 in case $\{X_i\}$ are i.i.d. (that is, replacing pairwise independence by mutual independence). Indeed, applying the same truncation scheme as in the proof of Proposition 2.3.1, it suffices to prove the following alternative to Lemma 2.3.2.

Lemma 2.3.21. *For integrable i.i.d. random variables $\{X_k\}$, let $\overline{S}_m = \sum_{k=1}^m \overline{X}_k$ and $\overline{X}_k = X_k I_{|X_k|\le k}$. Then, $n^{-1}(\overline{S}_n - \mathbf{E}\overline{S}_n) \stackrel{a.s.}{\to} 0$ as $n \to \infty$.*

Lemma 2.3.21, in contrast to Lemma 2.3.2, does not require the restriction to a subsequence n_l. Consequently, in this proof of the strong law there is no need for an interpolation argument so it is carried directly for X_k, with no need to split each variable to its positive and negative parts.

Proof of Lemma 2.3.21. We will shortly show that

$$\sum_{k=1}^{\infty} k^{-2}\,\mathsf{Var}(\overline{X}_k) \le 2\mathbf{E}|X_1|\,. \tag{2.3.8}$$

With X_1 integrable, applying Theorem 2.3.17 for the independent variables $Y_k = k^{-1}(\overline{X}_k - \mathbf{E}\overline{X}_k)$ this implies that for some A with $\mathbf{P}(A) = 1$, the random series $\sum_n Y_n(\omega)$ converges for all $\omega \in A$. Using Kronecker's lemma for $b_n = n$ and $x_n = \overline{X}_n(\omega) - \mathbf{E}\overline{X}_n$ we get that $n^{-1}\sum_{k=1}^n(\overline{X}_k - \mathbf{E}\overline{X}_k) \to 0$ as $n \to \infty$, for every $\omega \in A$, as stated.

The proof of (2.3.8) is similar to the computation employed in the proof of Lemma 2.3.2. That is, $\mathsf{Var}(\overline{X}_k) \le \mathbf{E}\overline{X}_k^2 = \mathbf{E}X_1^2 I_{|X_1|\le k}$ and $k^{-2} \le 2/(k(k+1))$, yielding that

$$\sum_{k=1}^{\infty} k^{-2}\,\mathsf{Var}(\overline{X}_k) \le \sum_{k=1}^{\infty} \frac{2}{k(k+1)} \mathbf{E}X_1^2 I_{|X_1|\le k} = \mathbf{E}X_1^2 v(|X_1|)\ ,$$

where for any $x > 0$,

$$v(x) = 2\sum_{k=\lceil x\rceil}^{\infty} \frac{1}{k(k+1)} = 2\sum_{k=\lceil x\rceil}^{\infty}\big[\frac{1}{k} - \frac{1}{k+1}\big] = \frac{2}{\lceil x\rceil} \le 2x^{-1}\,.$$

Consequently, $\mathbf{E}X_1^2 v(|X_1|) \le 2\mathbf{E}|X_1|$, and (2.3.8) follows. $\square$

Many of the ingredients of this proof of the strong law of large numbers are also relevant for solving the following exercise.

Exercise 2.3.22. *Let c_n be a bounded sequence of non-random constants, and $\{X_i\}$ i.i.d. integrable R.V.-s of zero mean. Show that $n^{-1}\sum_{k=1}^n c_k X_k \stackrel{a.s.}{\to} 0$ for $n \to \infty$.*

Next you find few exercises that illustrate how useful Kronecker's lemma is when proving the strong law of large numbers in case of independent but not identically distributed summands.

Exercise 2.3.23. *Let $S_n = \sum_{k=1}^n Y_k$ for independent random variables $\{Y_i\}$ such that $\mathsf{Var}(Y_k) < B < \infty$ and $\mathbf{E}Y_k = 0$ for all k. Show that $[n(\log n)^{1+\epsilon}]^{-1/2} S_n \stackrel{a.s.}{\to} 0$ as $n \to \infty$ and $\epsilon > 0$ is fixed (this falls short of the law of the iterated logarithm of (2.2.1), but each Y_k is allowed here to have a different distribution).*

Exercise 2.3.24. *Suppose the independent random variables $\{X_i\}$ are such that $\mathsf{Var}(X_k) \le p_k < \infty$ and $\mathbf{E}X_k = 0$ for $k = 1, 2, \ldots$.*

(a) *Show that if $\sum_k p_k < \infty$ then $n^{-1}\sum_{k=1}^n kX_k \stackrel{a.s.}{\to} 0$.*
(b) *Conversely, assuming $\sum_k p_k = \infty$, give an example of independent random variables $\{X_i\}$, such that $\mathsf{Var}(X_k) \le p_k$, $\mathbf{E}X_k = 0$, for which almost surely $\limsup_n X_n(\omega) = 1$.*

(c) *Show that the example you just gave is such that with probability one, the sequence* $n^{-1}\sum_{k=1}^n kX_k(\omega)$ *does not converge to a finite limit.*

Exercise 2.3.25. *Consider independent, non-negative random variables* X_n*.*

(a) *Show that if*

$$\sum_{n=1}^{\infty}[\mathbf{P}(X_n \geq 1) + \mathbf{E}(X_n I_{X_n<1})] < \infty \tag{2.3.9}$$

then the random series $\sum_n X_n(\omega)$ *converges w.p.1.*

(b) *Prove the converse, namely, that if* $\sum_n X_n(\omega)$ *converges w.p.1. then (2.3.9) holds.*

(c) *Suppose* G_n *are mutually independent random variables, with* G_n *having the normal distribution* $\mathcal{N}(\mu_n, v_n)$*. Show that w.p.1. the random series* $\sum_n G_n^2(\omega)$ *converges if and only if* $e = \sum_n(\mu_n^2 + v_n)$ *is finite.*

(d) *Suppose* τ_n *are mutually independent random variables, with* τ_n *having the exponential distribution of parameter* $\lambda_n > 0$*. Show that w.p.1. the random series* $\sum_n \tau_n(\omega)$ *converges if and only if* $\sum_n 1/\lambda_n$ *is finite.*

Hint: *For part (b) recall that for any* $a_n \in [0,1)$*, the series* $\sum_n a_n$ *is finite if and only if* $\prod_n(1-a_n) > 0$*. For part (c) let* $f(y) = \sum_n \min((\mu_n + \sqrt{v_n}y)^2, 1)$ *and observe that if* $e = \infty$ *then* $f(y) + f(-y) = \infty$ *for all* $y \neq 0$*.*

You can now also show that for such strong law of large numbers (that is, with independent but not identically distributed summands), it suffices to strengthen the corresponding weak law (only) along the subsequence $n_r = 2^r$.

Exercise 2.3.26. *Let* $Z_k = \sum_{j=1}^k Y_j$ *where* Y_j *are mutually independent R.V.-s.*

(a) *Fixing* $\varepsilon > 0$ *show that if* $2^{-r}Z_{2^r} \xrightarrow{a.s.} 0$ *then* $\sum_r \mathbf{P}(|Z_{2^{r+1}} - Z_{2^r}| > 2^r\varepsilon)$ *is finite and if* $m^{-1}Z_m \xrightarrow{p} 0$ *then* $\max_{m<k\leq 2m}\mathbf{P}(|Z_{2m} - Z_k| \geq \varepsilon m) \to 0$*.*

(b) *Adapting the proof of Kolmogorov's maximal inequality show that for any* n *and* $z > 0$*,*

$$\mathbf{P}(\max_{1\leq k\leq n}|Z_k| \geq 2z)\min_{1\leq k\leq n}\mathbf{P}(|Z_n - Z_k| < z) \leq \mathbf{P}(|Z_n| > z)\,.$$

(c) *Deduce that if both* $m^{-1}Z_m \xrightarrow{p} 0$ *and* $2^{-r}Z_{2^r} \xrightarrow{a.s.} 0$ *then also* $n^{-1}Z_n \xrightarrow{a.s.} 0$*.*

Hint: *For part (c) combine parts (a) and (b) with* $z = n\varepsilon$*,* $n = 2^r$ *and the mutually independent* Y_{j+n}*,* $1 \leq j \leq n$*, to show that* $\sum_r \mathbf{P}(2^{-r}D_r \geq 2\varepsilon)$ *is finite for* $D_r = \max_{2^r<k\leq 2^{r+1}}|Z_k - Z_{2^r}|$ *and any fixed* $\varepsilon > 0$*.*

Finally, here is an interesting property of non-negative random variables, regardless of their level of dependence.

Exercise 2.3.27. *Suppose random variables* $Y_k \geq 0$ *are such that* $n^{-1}\sum_{k=}^n Y_k \xrightarrow{p} 1$*. Show that then* $n^{-1}\max_{k=1}^n Y_k \xrightarrow{p} 0$*, and conclude that* $n^{-r}\sum_{k=1}^n Y_k^r \xrightarrow{p} 0$*, for any fixed* $r > 1$*.*

CHAPTER 3

Weak convergence, CLT and Poisson approximation

After dealing in Chapter 2 with examples in which random variables converge to non-random constants, we focus here on the more general theory of weak convergence, that is situations in which the laws of random variables converge to a limiting law, typically of a non-constant random variable. To motivate this theory, we start with Section 3.1 where we derive the celebrated Central Limit Theorem (in short CLT), the most widely used example of weak convergence. This is followed by the exposition of the theory, to which Section 3.2 is devoted. Section 3.3 is about the key tool of characteristic functions and their role in establishing convergence results such as the CLT. This tool is used in Section 3.4 to derive the Poisson approximation and provide an introduction to the Poisson process. In Section 3.5 we generalize the characteristic function to the setting of random vectors and study their properties while deriving the multivariate CLT.

3.1. The Central Limit Theorem

We start this section with the property of the normal distribution that makes it the likely limit for properly scaled sums of independent random variables. This is followed by a bare-hands proof of the CLT for triangular arrays in Subsection 3.1.1. We then present in Subsection 3.1.2 some of the many examples and applications of the CLT.

Recall the *normal distribution* of mean $\mu \in \mathbb{R}$ and variance $v > 0$, denoted hereafter $\mathcal{N}(\mu, v)$, the density of which is

$$f(y) = \frac{1}{\sqrt{2\pi v}} \exp\Big(-\frac{(y-\mu)^2}{2v}\Big)\,. \tag{3.1.1}$$

As we show next, the normal distribution is preserved when the sum of independent variables is considered (which is the main reason for its role as the limiting law for the CLT).

Lemma 3.1.1. *Let $Y_{n,k}$ be mutually independent random variables, each having the normal distribution $\mathcal{N}(\mu_{n,k}, v_{n,k})$. Then, $G_n = \sum_{k=1}^n Y_{n,k}$ has the normal distribution $\mathcal{N}(\mu_n, v_n)$, with $\mu_n = \sum_{k=1}^n \mu_{n,k}$ and $v_n = \sum_{k=1}^n v_{n,k}$.*

PROOF. Recall that Y has a $\mathcal{N}(\mu, v)$ distribution if and only if $Y - \mu$ has the $\mathcal{N}(0, v)$ distribution. Therefore, we may and shall assume without loss of generality that $\mu_{n,k} = 0$ for all k and n. Further, it suffices to prove the lemma for $n = 2$, as the general case immediately follows by an induction argument. With $n = 2$ fixed, we simplify our notations by omitting it everywhere. Next recall the formula of Corollary 1.4.33 for the probability density function of $G = Y_1 + Y_2$, which for Y_i

of $\mathcal{N}(0, v_i)$ distribution, $i = 1, 2$, is

$$f_G(z) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi v_1}} \exp(-\frac{(z-y)^2}{2v_1}) \frac{1}{\sqrt{2\pi v_2}} \exp(-\frac{y^2}{2v_2}) dy \,.$$

Comparing this with the formula of (3.1.1) for $v = v_1 + v_2$, it just remains to show that for any $z \in \mathbb{R}$,

$$1 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi u}} \exp(\frac{z^2}{2v} - \frac{(z-y)^2}{2v_1} - \frac{y^2}{2v_2}) dy \,, \tag{3.1.2}$$

where $u = v_1 v_2/(v_1 + v_2)$. It is not hard to check that the argument of the exponential function in (3.1.2) is $-(y - cz)^2/(2u)$ for $c = v_2/(v_1 + v_2)$. Consequently, (3.1.2) is merely the obvious fact that the $\mathcal{N}(cz, u)$ density function integrates to one (as any density function should), no matter what the value of z is. $\square$

Considering Lemma 3.1.1 for $Y_{n,k} = (nv)^{-1/2}(Y_k - \mu)$ and i.i.d. random variables Y_k, each having a normal distribution of mean μ and variance v, we see that $\mu_{n,k} = 0$ and $v_{n,k} = 1/n$, so $G_n = (nv)^{-1/2}(\sum_{k=1}^n Y_k - n\mu)$ has the standard $\mathcal{N}(0, 1)$ distribution, regardless of n.

3.1.1. Lindeberg's CLT for triangular arrays.

Our next proposition, the celebrated CLT, states that the distribution of $\widehat{S}_n = (nv)^{-1/2}(\sum_{k=1}^n X_k - n\mu)$ approaches the standard normal distribution in the limit $n \to \infty$, even though X_k may well be non-normal random variables.

Proposition 3.1.2 (Central Limit Theorem). *Let*

$$\widehat{S}_n = \frac{1}{\sqrt{nv}} \Big(\sum_{k=1}^n X_k - n\mu \Big) \,,$$

where $\{X_k\}$ are i.i.d with $v = \mathsf{Var}(X_1) \in (0, \infty)$ and $\mu = \mathbf{E}(X_1)$. Then,

$$\lim_{n\to\infty} \mathbf{P}(\widehat{S}_n \le b) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{b} \exp(-\frac{y^2}{2}) dy \qquad \text{for every } \ b \in \mathbb{R} \,. \tag{3.1.3}$$

As we have seen in the context of the weak law of large numbers, it pays to extend the scope of consideration to triangular arrays in which the random variables $X_{n,k}$ are independent within each row, but not necessarily of identical distribution. This is the context of Lindeberg's CLT, which we state next.

Theorem 3.1.3 (Lindeberg's CLT). *Let $\widehat{S}_n = \sum_{k=1}^n X_{n,k}$ for $\mathbf{P}$-mutually independent random variables $X_{n,k}$, $k = 1, \dots, n$, such that $\mathbf{E}X_{n,k} = 0$ for all k and*

$$v_n = \sum_{k=1}^n \mathbf{E}X_{n,k}^2 \to 1 \qquad \textit{as } n \to \infty \,.$$

Then, the conclusion (3.1.3) applies if for each $\varepsilon > 0$,

$$g_n(\varepsilon) = \sum_{k=1}^n \mathbf{E}[X_{n,k}^2; |X_{n,k}| \ge \varepsilon] \to 0 \quad \text{as } n \to \infty \,. \tag{3.1.4}$$

Note that the variables in different rows need not be independent of each other and could even be defined on different probability spaces.

Remark 3.1.4. Under the assumptions of Proposition 3.1.2 the variables $X_{n,k} = (nv)^{-1/2}(X_k - \mu)$ are mutually independent and such that

$$\mathbf{E}X_{n,k} = (nv)^{-1/2}(\mathbf{E}X_k - \mu) = 0, \qquad v_n = \sum_{k=1}^n \mathbf{E}X_{n,k}^2 = \frac{1}{nv}\sum_{k=1}^n \mathsf{Var}(X_k) = 1\,.$$

Further, per fixed n these $X_{n,k}$ are identically distributed, so

$$g_n(\varepsilon) = n\mathbf{E}[X_{n,1}^2\,;\,|X_{n,1}| \ge \varepsilon] = v^{-1}\mathbf{E}[(X_1-\mu)^2 I_{|X_1-\mu|\ge\sqrt{nv}\varepsilon}]\,.$$

For each $\varepsilon > 0$ the sequence $(X_1-\mu)^2\mathbf{I}_{|X_1-\mu|\ge\sqrt{nv}\varepsilon}$ converges a.s. to zero for $n \to \infty$ and is dominated by the integrable random variable $(X_1-\mu)^2$. Thus, by dominated convergence, $g_n(\varepsilon) \to 0$ as $n \to \infty$. We conclude that all assumptions of Theorem 3.1.3 are satisfied for this choice of $X_{n,k}$, hence Proposition 3.1.2 is a special instance of Lindeberg's CLT, to which we turn our attention next.

Let $r_n = \max\{\sqrt{v_{n,k}} : k = 1,\ldots,n\}$ for $v_{n,k} = \mathbf{E}X_{n,k}^2$. Since for every n, k and $\varepsilon > 0$,

$$v_{n,k} = \mathbf{E}X_{n,k}^2 = \mathbf{E}[X_{n,k}^2; |X_{n,k}| < \varepsilon] + \mathbf{E}[X_{n,k}^2; |X_{n,k}| \ge \varepsilon] \le \varepsilon^2 + g_n(\varepsilon)\,,$$

it follows that

$$r_n^2 \le \varepsilon^2 + g_n(\varepsilon) \qquad \forall n, \varepsilon > 0\,,$$

hence Lindeberg's condition (3.1.4) implies that $r_n \to 0$ as $n \to \infty$.

Remark. Lindeberg proved Theorem 3.1.3, introducing the condition (3.1.4). Later, Feller proved that (3.1.3) plus $r_n \to 0$ implies that Lindeberg's condition holds. Together, these two results are known as the Feller-Lindeberg Theorem.

We see that the variables $X_{n,k}$ are of uniformly small variance for large n. So, considering independent random variables $Y_{n,k}$ that are also independent of the $X_{n,k}$ and such that each $Y_{n,k}$ has a $\mathcal{N}(0, v_{n,k})$ distribution, for a smooth function $h(\cdot)$ one may control $|\mathbf{E}h(\widehat{S}_n) - \mathbf{E}h(G_n)|$ by a Taylor expansion upon successively replacing the $X_{n,k}$ by $Y_{n,k}$. This indeed is the outline of Lindeberg's proof, whose core is the following lemma.

Lemma 3.1.5. *For $h : \mathbb{R} \mapsto \mathbb{R}$ of continuous and uniformly bounded second and third derivatives, G_n having the $\mathcal{N}(0, v_n)$ law, every n and $\varepsilon > 0$, we have that*

$$|\mathbf{E}h(\widehat{S}_n) - \mathbf{E}h(G_n)| \le \Big(\frac{\varepsilon}{6} + \frac{r_n}{2}\Big)\, v_n\|h'''\|_\infty + g_n(\varepsilon)\|h''\|_\infty\,,$$

with $\|f\|_\infty = \sup_{x\in\mathbb{R}}|f(x)|$ denoting the supremum norm.

Remark. Recall that $G_n \overset{\mathcal{D}}{=} \sigma_n G$ for $\sigma_n = \sqrt{v_n}$. So, assuming $v_n \to 1$ and Lindeberg's condition which implies that $r_n \to 0$ for $n \to \infty$, it follows from the lemma that $|\mathbf{E}h(\widehat{S}_n) - \mathbf{E}h(\sigma_n G)| \to 0$ as $n \to \infty$. Further, $|h(\sigma_n x) - h(x)| \le |\sigma_n - 1||x|\|h'\|_\infty$, so taking the expectation with respect to the standard normal law we see that $|\mathbf{E}h(\sigma_n G) - \mathbf{E}h(G)| \to 0$ if the first derivative of h is also uniformly bounded. Hence,

$$\lim_{n\to\infty}\mathbf{E}h(\widehat{S}_n) = \mathbf{E}h(G)\,, \tag{3.1.5}$$

for any continuous function $h(\cdot)$ of continuous and uniformly bounded first three derivatives. This is actually all we need from Lemma 3.1.5 in order to prove Lindeberg's CLT. Further, as we show in Section 3.2, convergence in distribution as in (3.1.3) is *equivalent* to (3.1.5) holding for all continuous, bounded functions $h(\cdot)$.

PROOF OF LEMMA 3.1.5. Let $G_n = \sum_{k=1}^n Y_{n,k}$ for mutually independent $Y_{n,k}$, distributed according to $\mathcal{N}(0, v_{n,k})$, that are independent of $\{X_{n,k}\}$. Fixing n and h, we simplify the notations by eliminating n, that is, we write Y_k for $Y_{n,k}$, and X_k for $X_{n,k}$. To facilitate the proof define the mixed sums

$$U_l = \sum_{k=1}^{l-1} X_k + \sum_{k=l+1}^{n} Y_k \,, \qquad l = 1, \ldots, n$$

Note the following identities

$$G_n = U_1 + Y_1, \quad U_l + X_l = U_{l+1} + Y_{l+1}, \;\; l = 1, \ldots, n-1, \quad U_n + X_n = \widehat{S}_n \,,$$

which imply that,

$$|\mathbf{E}h(G_n) - \mathbf{E}h(\widehat{S}_n)| = |\mathbf{E}h(U_1 + Y_1) - \mathbf{E}h(U_n + X_n)| \le \sum_{l=1}^n \Delta_l \,, \tag{3.1.6}$$

where $\Delta_l = |\mathbf{E}[h(U_l + Y_l) - h(U_l + X_l)]|$, for $l = 1, \ldots, n$. For any l and $\xi \in \mathbb{R}$, consider the remainder term

$$R_l(\xi) = h(U_l + \xi) - h(U_l) - \xi h'(U_l) - \frac{\xi^2}{2} h''(U_l)$$

in second order Taylor's expansion of $h(\cdot)$ at U_l. By Taylor's theorem, we have that

$$|R_l(\xi)| \le \|h'''\|_\infty \frac{|\xi|^3}{6}, \qquad \text{(from third order expansion)}$$
$$|R_l(\xi)| \le \|h''\|_\infty |\xi|^2, \qquad \text{(from second order expansion)}$$

whence,

$$|R_l(\xi)| \le \min\left\{ \|h'''\|_\infty \frac{|\xi|^3}{6}, \|h''\|_\infty |\xi|^2 \right\} . \tag{3.1.7}$$

Considering the expectation of the difference between the two identities,

$$h(U_l + X_l) = h(U_l) + X_l h'(U_l) + \frac{X_l^2}{2} h''(U_l) + R_l(X_l) \,,$$
$$h(U_l + Y_l) = h(U_l) + Y_l h'(U_l) + \frac{Y_l^2}{2} h''(U_l) + R_l(Y_l) \,,$$

we get that

$$\Delta_l \le \Big|\mathbf{E}[(X_l - Y_l)h'(U_l)]\Big| + \Big|\mathbf{E}\big[(\frac{X_l^2}{2} - \frac{Y_l^2}{2})h''(U_l)\big]\Big| + |\mathbf{E}[R_l(X_l) - R_l(Y_l)]| \,.$$

Recall that X_l and Y_l are independent of U_l and chosen such that $\mathbf{E}X_l = \mathbf{E}Y_l$ and $\mathbf{E}X_l^2 = \mathbf{E}Y_l^2$. As the first two terms in the bound on Δ_l vanish we have that

$$\Delta_l \le \mathbf{E}|R_l(X_l)| + \mathbf{E}|R_l(Y_l)| \,. \tag{3.1.8}$$

Further, utilizing (3.1.7),

$$\mathbf{E}|R_l(X_l)| \le \|h'''\|_\infty \mathbf{E}\big[\frac{|X_l|^3}{6}; |X_l| \le \varepsilon\big] + \|h''\|_\infty \mathbf{E}[|X_l|^2; |X_l| \ge \varepsilon]$$

$$\le \frac{\varepsilon}{6}\|h'''\|_\infty \mathbf{E}[|X_l|^2] + \|h''\|_\infty \mathbf{E}[X_l^2; |X_l| \ge \varepsilon] .$$

Summing these bounds over $l = 1, \ldots, n$, by our assumption that $\sum_{l=1}^n \mathbf{E}X_l^2 = v_n$ and the definition of $g_n(\varepsilon)$, we get that

$$\sum_{l=1}^n \mathbf{E}|R_l(X_l)| \le \frac{\varepsilon}{6} v_n \|h'''\|_\infty + g_n(\varepsilon)\|h''\|_\infty . \tag{3.1.9}$$

Recall that $Y_l/\sqrt{v_{n,l}}$ is a standard normal random variable, whose fourth moment is 3 (see (1.3.18)). By monotonicity in q of the L^q-norms (c.f. Lemma 1.3.16), it follows that $\mathbf{E}[|Y_l/\sqrt{v_{n,l}}|^3] \le 3$, hence $\mathbf{E}|Y_l|^3 \le 3v_{n,l}^{3/2} \le 3r_n v_{n,l}$. Utilizing once more (3.1.7) and the fact that $v_n = \sum_{l=1}^n v_{n,l}$, we arrive at

$$\sum_{l=1}^n \mathbf{E}|R_l(Y_l)| \le \frac{\|h'''\|_\infty}{6} \sum_{l=1}^n \mathbf{E}|Y_l|^3 \le \frac{r_n}{2} v_n \|h'''\|_\infty . \tag{3.1.10}$$

Plugging (3.1.8)–(3.1.10) into (3.1.6) completes the proof of the lemma. $\square$

In view of (3.1.5), Lindeberg's CLT builds on the following elementary lemma, whereby we approximate the indicator function on $(-\infty, b]$ by continuous, bounded functions $h_k : \mathbb{R} \mapsto \mathbb{R}$ for each of which Lemma 3.1.5 applies.

Lemma 3.1.6. *There exist $h_k^\pm(x)$ of continuous and uniformly bounded first three derivatives, such that $0 \le h_k^-(x) \uparrow I_{(-\infty,b)}(x)$ and $1 \ge h_k^+(x) \downarrow I_{(-\infty,b]}(x)$ as $k \to \infty$.*

Proof. There are many ways to prove this. Here is one which is from first principles, hence requires no analysis knowledge. The function $\psi : [0,1] \mapsto [0,1]$ given by $\psi(x) = 140 \int_x^1 u^3(1-u)^3 du$ is monotone decreasing, with continuous derivatives of all order, such that $\psi(0) = 1$, $\psi(1) = 0$ and whose first three derivatives at 0 and at 1 are all zero. Its extension $\phi(x) = \psi(\min(x,1)_+)$ to a function on $\mathbb{R}$ that is one for $x \le 0$ and zero for $x \ge 1$ is thus non-increasing, with continuous and uniformly bounded first three derivatives. It is easy to check that the translated and scaled functions $h_k^+(x) = \phi(k(x-b))$ and $h_k^-(x) = \phi(k(x-b)+1)$ have all the claimed properties. $\square$

Proof of Theorem 3.1.3. Applying (3.1.5) for $h_k^-(\cdot)$, then taking $k \to \infty$ we have by monotone convergence that

$$\liminf_{n\to\infty} \mathbf{P}(\widehat{S}_n < b) \ge \lim_{n\to\infty} \mathbf{E}[h_k^-(\widehat{S}_n)] = \mathbf{E}[h_k^-(G)] \uparrow F_G(b^-) .$$

Similarly, considering $h_k^+(\cdot)$, then taking $k \to \infty$ we have by bounded convergence that

$$\limsup_{n\to\infty} \mathbf{P}(\widehat{S}_n \le b) \le \lim_{n\to\infty} \mathbf{E}[h_k^+(\widehat{S}_n)] = \mathbf{E}[h_k^+(G)] \downarrow F_G(b) .$$

Since $F_G(\cdot)$ is a continuous function we conclude that $\mathbf{P}(\widehat{S}_n \le b)$ converges to $F_G(b) = F_G(b^-)$, as $n \to \infty$. This holds for every $b \in \mathbb{R}$ as claimed. $\square$

3.1.2. Applications of the CLT. We start with the simpler, i.i.d. case. In doing so, we use the notation $Z_n \xrightarrow{\mathcal{D}} G$ when the analog of (3.1.3) holds for the sequence $\{Z_n\}$, that is $\mathbf{P}(Z_n \le b) \to \mathbf{P}(G \le b)$ as $n \to \infty$ for all $b \in \mathbb{R}$ (where G is a standard normal variable).

Example 3.1.7 (Normal approximation of the Binomial). *Consider i.i.d. random variables $\{B_i\}$, each of whom is Bernoulli of parameter $0 < p < 1$ (i.e. $P(B_1 = 1) = 1 - P(B_1 = 0) = p$). The sum $S_n = B_1 + \cdots + B_n$ has the* Binomial distribution of parameters *(n, p), that is,*

$$\mathbf{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, \ldots, n.$$

For example, if B_i indicates that the ith independent toss of the same coin lands on a Head then S_n counts the total numbers of Heads in the first n tosses of the coin. Recall that $\mathbf{E}B = p$ and $\mathsf{Var}(B) = p(1-p)$ (see Example 1.3.69), so the CLT states that $(S_n - np)/\sqrt{np(1-p)} \xrightarrow{\mathcal{D}} G$. It allows us to approximate, for all large enough n, the typically non-computable weighted sums of binomial terms by integrals with respect to the standard normal density.

Here is another example that is similar and almost as widely used.

Example 3.1.8 (Normal approximation of the Poisson distribution). *It is not hard to verify that the sum of two independent Poisson random variables has the Poisson distribution, with a parameter which is the sum of the parameters of the summands. Thus, by induction, if $\{X_i\}$ are i.i.d. each of Poisson distribution of parameter 1, then $N_n = X_1 + \ldots + X_n$ has a Poisson distribution of parameter n. Since $\mathbf{E}(N_1) = \mathsf{Var}(N_1) = 1$ (see Example 1.3.69), the CLT applies for $(N_n - n)/n^{1/2}$. This provides an approximation for the distribution function of the Poisson variable N_λ of parameter λ that is a large integer. To deal with non-integer values $\lambda = n + \eta$ for some $\eta \in (0, 1)$, consider the mutually independent Poisson variables N_n, N_η and $N_{1-\eta}$. Since $N_\lambda \stackrel{\mathcal{D}}{=} N_n + N_\eta$ and $N_{n+1} \stackrel{\mathcal{D}}{=} N_n + N_\eta + N_{1-\eta}$, this provides a* monotone coupling *, that is, a construction of the random variables N_n, N_λ and N_{n+1} on the same probability space, such that $N_n \le N_\lambda \le N_{n+1}$. Because of this monotonicity, for any $\varepsilon > 0$ and all $n \ge n_0(b, \varepsilon)$ the event $\{(N_\lambda - \lambda)/\sqrt{\lambda} \le b\}$ is between $\{(N_{n+1} - (n+1))/\sqrt{n+1} \le b - \varepsilon\}$ and $\{(N_n - n)/\sqrt{n} \le b + \varepsilon\}$. Considering the limit as $n \to \infty$ followed by $\varepsilon \to 0$, it thus follows that the convergence $(N_n - n)/n^{1/2} \xrightarrow{\mathcal{D}} G$ implies also that $(N_\lambda - \lambda)/\lambda^{1/2} \xrightarrow{\mathcal{D}} G$ as $\lambda \to \infty$. In words, the normal distribution is a good approximation of a Poisson with large parameter.*

In Theorem 2.3.3 we established the strong law of large numbers when the summands X_i are only *pairwise independent*. Unfortunately, as the next example shows, pairwise independence is not good enough for the CLT.

Example 3.1.9. *Consider i.i.d. $\{\xi_i\}$ such that $\mathbf{P}(\xi_i = 1) = \mathbf{P}(\xi_i = -1) = 1/2$ for all i. Set $X_1 = \xi_1$ and successively let $X_{2^k+j} = X_j \xi_{k+2}$ for $j = 1, \ldots, 2^k$ and $k = 0, 1, \ldots$. Note that each X_l is a $\{-1, 1\}$-valued variable, specifically, a product of a different finite subset of ξ_i-s that corresponds to the positions of ones in the binary representation of $2l - 1$ (with ξ_1 for its least significant digit, ξ_2 for the next digit, etc.). Consequently, each X_l is of zero mean and if $l \ne r$ then in $\mathbf{E}X_l X_r$ at least one of the ξ_i-s will appear exactly once, resulting with $\mathbf{E}X_l X_r = 0$, hence with $\{X_l\}$ being uncorrelated variables. Recall part (b) of Exercise 1.4.42, that such* *variables are pairwise independent. Further,* $\mathbf{E}X_l = 0$ *and* $X_l \in \{-1, 1\}$ *mean that* $\mathbf{P}(X_l = -1) = \mathbf{P}(X_l = 1) = 1/2$ *are identically distributed. As for the zero mean variables* $S_n = \sum_{j=1}^n X_j$*, we have arranged things such that* $S_1 = \xi_1$ *and for any* $k \geq 0$

$$S_{2^{k+1}} = \sum_{j=1}^{2^k} (X_j + X_{2^k+j}) = \sum_{j=1}^{2^k} X_j(1 + \xi_{k+2}) = S_{2^k}(1 + \xi_{k+2}),$$

hence $S_{2^k} = \xi_1 \prod_{i=2}^{k+1}(1 + \xi_i)$ *for all* $k \geq 1$*. In particular,* $S_{2^k} = 0$ *unless* $\xi_2 = \xi_3 = \ldots = \xi_{k+1} = 1$*, an event of probability* 2^{-k}*. Thus,* $\mathbf{P}(S_{2^k} \neq 0) = 2^{-k}$ *and certainly the* CLT *result (3.1.3) does not hold along the subsequence* $n = 2^k$*.*

We turn next to applications of Lindeberg's triangular array CLT, starting with the asymptotic of the count of record events till time $n \gg 1$.

Exercise 3.1.10. *Consider the count* R_n *of record events during the first* n *instances of i.i.d. R.V. with a continuous distribution function, as in Example 2.2.27. Recall that* $R_n = B_1 + \cdots + B_n$ *for mutually independent Bernoulli random variables* $\{B_k\}$ *such that* $\mathbf{P}(B_k = 1) = 1 - \mathbf{P}(B_k = 0) = k^{-1}$*.*

(a) *Check that* $b_n / \log n \to 1$ *where* $b_n = \mathsf{Var}(R_n)$*.*
(b) *Show that Lindeberg's* CLT *applies for* $X_{n,k} = (\log n)^{-1/2}(B_k - k^{-1})$*.*
(c) *Recall that* $|\mathbf{E}R_n - \log n| \leq 1$*, and conclude that* $(R_n - \log n)/\sqrt{\log n} \xrightarrow{\mathcal{D}} G$*.*

Remark. Let $\mathcal{S}_n$ denote the symmetric group of permutations on $\{1, \ldots, n\}$. For $s \in \mathcal{S}_n$ and $i \in \{1, \ldots, n\}$, denoting by $L_i(s)$ the smallest $j \leq n$ such that $s^j(i) = i$, we call $\{s^j(i) : 1 \leq j \leq L_i(s)\}$ the cycle of s containing i. If each $s \in \mathcal{S}_n$ is equally likely, then the law of the number $T_n(s)$ of different cycles in s is the same as that of R_n of Example 2.2.27 (for a proof see [**Dur10**, Example 2.2.4]). Consequently, Exercise 3.1.10 also shows that in this setting $(T_n - \log n)/\sqrt{\log n} \xrightarrow{\mathcal{D}} G$.

Part (a) of the following exercise is a special case of Lindeberg's CLT, known also as *Lyapunov's theorem.*

Exercise 3.1.11 (Lyapunov's theorem)**.** *Let* $S_n = \sum_{k=1}^n X_k$ *for* $\{X_k\}$ *mutually independent such that* $v_n = \mathsf{Var}(S_n) < \infty$*.*

(a) *Show that if there exists* $q > 2$ *such that*

$$\lim_{n \to \infty} v_n^{-q/2} \sum_{k=1}^n \mathbf{E}(|X_k - \mathbf{E}X_k|^q) = 0,$$

then $v_n^{-1/2}(S_n - \mathbf{E}S_n) \xrightarrow{\mathcal{D}} G$*.*

(b) *Show that part (a) applies in case* $v_n \to \infty$ *and* $\mathbf{E}(|X_k - \mathbf{E}X_k|^q) \leq C(\mathsf{Var}\, X_k)^r$ *for* $r = 1$*, some* $q > 2$*,* $C < \infty$ *and* $k = 1, 2, \ldots$*.*
(c) *Provide an example where the conditions of part (b) hold with* $r = q/2$ *but* $v_n^{-1/2}(S_n - \mathbf{E}S_n)$ *does not converge in distribution.*

The next application of Lindeberg's CLT involves the use of truncation (which we have already introduced in the context of the weak law of large numbers), to derive the CLT for normalized sums of certain i.i.d. random variables of *infinite variance.*

Proposition 3.1.12. *Suppose $\{X_k\}$ are i.i.d of symmetric distribution, that is $X_1 \stackrel{\mathcal{D}}{=} -X_1$ (or $\mathbf{P}(X_1 > x) = \mathbf{P}(X_1 < -x)$ for all x) such that $\mathbf{P}(|X_1| > x) = x^{-2}$ for $x \geq 1$. Then, $\frac{1}{\sqrt{n \log n}} \sum_{k=1}^n X_k \xrightarrow{\mathcal{D}} G$ as $n \to \infty$.*

Remark 3.1.13. Note that $\mathsf{Var}(X_1) = \mathbf{E}X_1^2 = \int_0^\infty 2x\mathbf{P}(|X_1| > x)dx = \infty$ (c.f. part (a) of Lemma 1.4.31), so the usual CLT of Proposition 3.1.2 does not apply here. Indeed, the infinite variance of the summands results in a different normalization of the sums $S_n = \sum_{k=1}^n X_k$ that is tailored to the specific tail behavior of $x \mapsto \mathbf{P}(|X_1| > x)$.

Caution should be exercised here, since when $\mathbf{P}(|X_1| > x) = x^{-\alpha}$ for $x > 1$ and some $0 < \alpha < 2$, there is no way to approximate the distribution of $(S_n - a_n)/b_n$ by the standard normal distribution. Indeed, in this case $b_n = n^{1/\alpha}$ and the approximation is by an α-stable law (c.f. Definition 3.3.31 and Exercise 3.3.33).

Proof. We plan to apply Lindeberg's CLT for the truncated random variables $X_{n,k} = b_n^{-1} X_k I_{|X_k| \leq c_n}$ where $b_n = \sqrt{n \log n}$ and $c_n \geq 1$ are such that both $c_n/b_n \to 0$ and $c_n/\sqrt{n} \to \infty$. Indeed, for each n the variables $X_{n,k}$, $k = 1, \ldots, n$, are i.i.d. of bounded and symmetric distribution (since both the distribution of X_k and the truncation function are symmetric). Consequently, $\mathbf{E}X_{n,k} = 0$ for all n and k. Further, we have chosen b_n such that

$$
\begin{aligned}
v_n = n\mathbf{E}X_{n,1}^2 = \frac{n}{b_n^2}\mathbf{E}X_1^2 I_{|X_1| \leq c_n} &= \frac{n}{b_n^2}\int_0^{c_n} 2x[\mathbf{P}(|X_1| > x) - \mathbf{P}(|X_1| > c_n)]dx \\
&= \frac{n}{b_n^2}\Big[\int_0^1 2x dx + \int_1^{c_n} \frac{2}{x}dx - \int_0^{c_n} \frac{2x}{c_n^2}dx\Big] = \frac{2n \log c_n}{b_n^2} \to 1
\end{aligned}
$$

as $n \to \infty$. Finally, note that $|X_{n,k}| \leq c_n/b_n \to 0$ as $n \to \infty$, implying that $g_n(\varepsilon) = 0$ for any $\varepsilon > 0$ and all n large enough, hence Lindeberg's condition trivially holds. We thus deduce from Lindeberg's CLT that $\frac{1}{\sqrt{n \log n}}\overline{S}_n \xrightarrow{\mathcal{D}} G$ as $n \to \infty$, where $\overline{S}_n = \sum_{k=1}^n X_k I_{|X_k| \leq c_n}$ is the sum of the truncated variables. We have chosen the truncation level c_n large enough to assure that

$$
\mathbf{P}(S_n \neq \overline{S}_n) \leq \sum_{k=1}^n \mathbf{P}(|X_k| > c_n) = n\mathbf{P}(|X_1| > c_n) = nc_n^{-2} \to 0
$$

for $n \to \infty$, hence we may now conclude that $\frac{1}{\sqrt{n \log n}} S_n \xrightarrow{\mathcal{D}} G$ as claimed. $\square$

We conclude this section with Kolmogorov's three series theorem, the most definitive result on the convergence of random series.

Theorem 3.1.14 (Kolmogorov's three series theorem)**.** *Suppose $\{X_k\}$ are independent random variables. For non-random $c > 0$ let $X_n^{(c)} = X_n I_{|X_n| \leq c}$ be the corresponding truncated variables and consider the three series*

$$
\sum_n \mathbf{P}(|X_n| > c), \qquad \sum_n \mathbf{E}X_n^{(c)}, \qquad \sum_n \mathsf{Var}(X_n^{(c)}). \tag{3.1.11}
$$

Then, the random series $\sum_n X_n$ converges a.s. if and only if for some $c > 0$ all three series of (3.1.11) converge.

Remark. By convergence of a series we mean the existence of a finite limit to the sum of its first m entries when $m \to \infty$. Note that the theorem implies that if all three series of (3.1.11) converge for some $c > 0$, then they necessarily converge for every $c > 0$.

Proof. We prove the sufficiency first, that is, assume that for some $c > 0$ all three series of (3.1.11) converge. By Theorem 2.3.17 and the finiteness of $\sum_n \mathsf{Var}(X_n^{(c)})$ it follows that the random series $\sum_n (X_n^{(c)} - \mathbf{E}X_n^{(c)})$ converges a.s. Then, by our assumption that $\sum_n \mathbf{E}X_n^{(c)}$ converges, also $\sum_n X_n^{(c)}$ converges a.s. Further, by assumption the sequence of probabilities $\mathbf{P}(X_n \neq X_n^{(c)}) = \mathbf{P}(|X_n| > c)$ is summable, hence by Borel-Cantelli I, we have that a.s. $X_n \neq X_n^{(c)}$ for at most finitely many n's. The convergence a.s. of $\sum_n X_n^{(c)}$ thus results with the convergence a.s. of $\sum_n X_n$, as claimed.

We turn to prove the necessity of convergence of the three series in (3.1.11) to the convergence of $\sum_n X_n$, which is where we use the CLT. To this end, assume the random series $\sum_n X_n$ converges a.s. (to a finite limit) and fix an arbitrary constant $c > 0$. The convergence of $\sum_n X_n$ implies that $|X_n| \to 0$, hence a.s. $|X_n| > c$ for only finitely many n's. In view of the independence of these events and Borel-Cantelli II, necessarily the sequence $\mathbf{P}(|X_n| > c)$ is summable, that is, the series $\sum_n \mathbf{P}(|X_n| > c)$ converges. Further, the convergence a.s. of $\sum_n X_n$ then results with the a.s. convergence of $\sum_n X_n^{(c)}$.

Suppose now that the non-decreasing sequence $v_n = \sum_{k=1}^n \mathsf{Var}(X_k^{(c)})$ is unbounded, in which case the latter convergence implies that a.s. $T_n = v_n^{-1/2} \sum_{k=1}^n X_k^{(c)} \to 0$ when $n \to \infty$. We further claim that in this case Lindeberg's CLT applies for $\widehat{S}_n = \sum_{k=1}^n X_{n,k}$, where

$$X_{n,k} = v_n^{-1/2}(X_k^{(c)} - m_k^{(c)}), \quad \text{and} \quad m_k^{(c)} = \mathbf{E}X_k^{(c)}.$$

Indeed, per fixed n the variables $X_{n,k}$ are mutually independent of zero mean and such that $\sum_{k=1}^n \mathbf{E}X_{n,k}^2 = 1$. Further, since $|X_k^{(c)}| \leq c$ and we assumed that $v_n \uparrow \infty$ it follows that $|X_{n,k}| \leq 2c/\sqrt{v_n} \to 0$ as $n \to \infty$, resulting with Lindeberg's condition holding (as $g_n(\varepsilon) = 0$ when $\varepsilon > 2c/\sqrt{v_n}$, i.e. for all n large enough). Combining Lindeberg's CLT conclusion that $\widehat{S}_n \xrightarrow{\mathcal{D}} G$ and $T_n \overset{a.s.}{\to} 0$, we deduce that $(\widehat{S}_n - T_n) \xrightarrow{\mathcal{D}} G$ (c.f. Exercise 3.2.8). However, since $\widehat{S}_n - T_n = -v_n^{-1/2} \sum_{k=1}^n m_k^{(c)}$ are *non-random*, the sequence $\mathbf{P}(\widehat{S}_n - T_n \leq 0)$ is composed of zeros and ones, hence cannot converge to $\mathbf{P}(G \leq 0) = 1/2$. We arrive at a contradiction to our assumption that $v_n \uparrow \infty$, and so conclude that the sequence $\mathsf{Var}(X_n^{(c)})$ is summable, that is, the series $\sum_n \mathsf{Var}(X_n^{(c)})$ converges.

By Theorem 2.3.17, the summability of $\mathsf{Var}(X_n^{(c)})$ implies that the series $\sum_n (X_n^{(c)} - m_n^{(c)})$ converges a.s. We have already seen that $\sum_n X_n^{(c)}$ converges a.s. so it follows that their difference $\sum_n m_n^{(c)}$, which is the middle term of (3.1.11), converges as well. $\square$

3.2. Weak convergence

Focusing here on the theory of *weak convergence*, we first consider in Subsection 3.2.1 the *convergence in distribution* in a more general setting than that of the CLT. This is followed by the study in Subsection 3.2.2 of weak convergence of probability measures and the theory associated with it. Most notably its relation to other modes of convergence, such as convergence in *total variation* or point-wise convergence of probability density functions. We conclude by introducing in Subsection 3.2.3 the key concept of *uniform tightness* which is instrumental to the derivation of weak convergence statements, as demonstrated in later sections of this chapter.

3.2.1. Convergence in distribution. Motivated by the CLT, we explore here the convergence in distribution, its relation to convergence in probability, some additional properties and examples in which the limiting law is not the normal law. To start off, here is the definition of convergence in distribution.

Definition 3.2.1. *We say that R.V.-s X_n* converge in distribution *to a R.V. X_∞, denoted by $X_n \xrightarrow{\mathcal{D}} X_\infty$, if $F_{X_n}(\alpha) \to F_{X_\infty}(\alpha)$ as $n \to \infty$ for each fixed α which is a continuity point of F_{X_∞}.*
Similarly, we say that distribution functions F_n converge weakly *to F_∞, denoted by $F_n \stackrel{w}{\to} F_\infty$, if $F_n(\alpha) \to F_\infty(\alpha)$ as $n \to \infty$ for each fixed α which is a continuity point of F_∞.*

Remark. If the limit R.V. X_∞ has a probability density function, or more generally whenever F_{X_∞} is a continuous function, the convergence in distribution of X_n to X_∞ is equivalent to the point-wise convergence of the corresponding distribution functions. Such is the case of the CLT, since the normal R.V. G has a density. Further,

Exercise 3.2.2. *Show that if $F_n \stackrel{w}{\to} F_\infty$ and $F_\infty(\cdot)$ is a continuous function then also* $\sup_x |F_n(x) - F_\infty(x)| \to 0$.

The CLT is not the only example of convergence in distribution we have already met. Recall the Glivenko-Cantelli theorem (see Theorem 2.3.6), whereby a.s. the empirical distribution functions F_n of an i.i.d. sequence of variables $\{X_i\}$ converge uniformly, hence point-wise to the true distribution function F_X.

Here is an explicit necessary and sufficient condition for the convergence in distribution of integer valued random variables

Exercise 3.2.3. *Let $X_n, 1 \le n \le \infty$ be integer valued R.V.-s. Show that $X_n \xrightarrow{\mathcal{D}} X_\infty$ if and only if $\mathbf{P}(X_n = k) \to_{n\to\infty} \mathbf{P}(X_\infty = k)$ for each $k \in \mathbf{Z}$.*

In contrast with all of the preceding examples, we demonstrate next why the convergence $X_n \xrightarrow{\mathcal{D}} X_\infty$ has been chosen to be strictly weaker than the point-wise convergence of the corresponding distribution functions. We also see that $\mathbf{E}h(X_n) \to \mathbf{E}h(X_\infty)$ or not, depending upon the choice of $h(\cdot)$, and even within the collection of continuous functions with image in $[-1, 1]$, the rate of this convergence is not uniform in h.

Example 3.2.4. *The random variables $X_n = 1/n$ converge in distribution to $X_\infty = 0$. Indeed, it is easy to check that $F_{X_n}(\alpha) = I_{[1/n,\infty)}(\alpha)$ converge to $F_{X_\infty}(\alpha) = I_{[0,\infty)}(\alpha)$ at each $\alpha \neq 0$. However, there is no convergence at the discontinuity point $\alpha = 0$ of F_{X_∞} as $F_{X_\infty}(0) = 1$ while $F_{X_n}(0) = 0$ for all n.*
Further, $\mathbf{E}h(X_n) = h(\frac{1}{n}) \to h(0) = \mathbf{E}h(X_\infty)$ if and only if $h(x)$ is continuous at $x = 0$, and the rate of convergence varies with the modulus of continuity of $h(x)$ at $x = 0$.
More generally, if $X_n = X + 1/n$ then $F_{X_n}(\alpha) = F_X(\alpha - 1/n) \to F_X(\alpha^-)$ as $n \to \infty$. So, in order for $X + 1/n$ to converge in distribution to X as $n \to \infty$, we

have to restrict such convergence to the continuity points of the limiting distribution function F_X, as done in Definition 3.2.1.

We have seen in Examples 3.1.7 and 3.1.8 that the normal distribution is a good approximation for the Binomial and the Poisson distributions (when the corresponding parameter is large). Our next example is of the same type, now with the approximation of the Geometric distribution by the Exponential one.

Example 3.2.5 (Exponential approximation of the Geometric). *Let Z_p be a random variable with a Geometric distribution of parameter $p \in (0,1)$, that is, $\mathbf{P}(Z_p \geq k) = (1-p)^{k-1}$ for any positive integer k. As $p \to 0$, we see that*

$$\mathbf{P}(pZ_p > t) = (1-p)^{\lfloor t/p \rfloor} \to e^{-t} \qquad \text{for all} \quad t \geq 0$$

That is, $pZ_p \xrightarrow{\mathcal{D}} T$, with T having a standard exponential distribution. *As Z_p corresponds to the number of independent trials till the first occurrence of a specific event whose probability is p, this approximation corresponds to waiting for the occurrence of rare events.*

At this point, you are to check that convergence in probability implies the convergence in distribution, which is hence weaker than all notions of convergence explored in Section 1.3.3 (and is perhaps a reason for naming it weak convergence). The converse cannot hold, for example because convergence in distribution does not require X_n and X_∞ to be even defined on the same probability space. However, convergence in distribution is equivalent to convergence in probability when the limiting random variable is a non-random constant.

Exercise 3.2.6. *Show that if $X_n \xrightarrow{p} X_\infty$, then $X_n \xrightarrow{\mathcal{D}} X_\infty$. Conversely, if $X_n \xrightarrow{\mathcal{D}} X_\infty$ and X_∞ is almost surely a non-random constant, then $X_n \xrightarrow{p} X_\infty$.*

Further, as the next theorem shows, given $F_n \xrightarrow{w} F_\infty$, it is possible to construct random variables Y_n, $n \leq \infty$ such that $F_{Y_n} = F_n$ and $Y_n \xrightarrow{a.s.} Y_\infty$. The catch of course is to construct the appropriate *coupling*, that is, to specify the relation between the different Y_n's.

Theorem 3.2.7. *Let F_n be a sequence of distribution functions that converges weakly to F_∞. Then there exist random variables Y_n, $1 \leq n \leq \infty$ on the probability space $((0,1], \mathcal{B}_{(0,1]}, U)$ such that $F_{Y_n} = F_n$ for $1 \leq n \leq \infty$ and $Y_n \xrightarrow{a.s.} Y_\infty$.*

Proof. We use Skorokhod's representation as in the proof of Theorem 1.2.37. That is, for $\omega \in (0,1]$ and $1 \leq n \leq \infty$ let $Y_n^+(\omega) \geq Y_n^-(\omega)$ be

$$Y_n^+(\omega) = \sup\{y : F_n(y) \leq \omega\}, \qquad Y_n^-(\omega) = \sup\{y : F_n(y) < \omega\}\,.$$

While proving Theorem 1.2.37 we saw that $F_{Y_n^-} = F_n$ for any $n \leq \infty$, and as remarked there $Y_n^-(\omega) = Y_n^+(\omega)$ for all but at most countably many values of ω, hence $\mathbf{P}(Y_n^- = Y_n^+) = 1$. It thus suffices to show that for all $\omega \in (0,1)$,

$$\begin{aligned} Y_\infty^+(\omega) &\geq \limsup_{n\to\infty} Y_n^+(\omega) \geq \limsup_{n\to\infty} Y_n^-(\omega) \\ &\geq \liminf_{n\to\infty} Y_n^-(\omega) \geq Y_\infty^-(\omega)\,. \end{aligned} \tag{3.2.1}$$

Indeed, then $Y_n^-(\omega) \to Y_\infty^-(\omega)$ for any $\omega \in A = \{\omega : Y_\infty^+(\omega) = Y_\infty^-(\omega)\}$ where $\mathbf{P}(A) = 1$. Hence, setting $Y_n = Y_n^+$ for $1 \leq n \leq \infty$ would complete the proof of the theorem.

Turning to prove (3.2.1) note that the two middle inequalities are trivial. Fixing $\omega \in (0,1)$ we proceed to show that

$$Y_\infty^+(\omega) \geq \limsup_{n\to\infty} Y_n^+(\omega) \ . \tag{3.2.2}$$

Since the continuity points of F_∞ form a dense subset of $\mathbb{R}$ (see Exercise 1.2.39), it suffices for (3.2.2) to show that if $z > Y_\infty^+(\omega)$ is a continuity point of F_∞, then necessarily $z \geq Y_n^+(\omega)$ for all n large enough. To this end, note that $z > Y_\infty^+(\omega)$ implies by definition that $F_\infty(z) > \omega$. Since z is a continuity point of F_∞ and $F_n \xrightarrow{w} F_\infty$ we know that $F_n(z) \to F_\infty(z)$. Hence, $F_n(z) > \omega$ for all sufficiently large n. By definition of Y_n^+ and monotonicity of F_n, this implies that $z \geq Y_n^+(\omega)$, as needed. The proof of

$$\liminf_{n\to\infty} Y_n^-(\omega) \geq Y_\infty^-(\omega) \,, \tag{3.2.3}$$

is analogous. For $y < Y_\infty^-(\omega)$ we know by monotonicity of F_∞ that $F_\infty(y) < \omega$. Assuming further that y is a continuity point of F_∞, this implies that $F_n(y) < \omega$ for all sufficiently large n, which in turn results with $y \leq Y_n^-(\omega)$. Taking continuity points y_k of F_∞ such that $y_k \uparrow Y_\infty^-(\omega)$ will yield (3.2.3) and complete the proof. $\square$

The next exercise provides useful ways to get convergence in distribution for one sequence out of that of another sequence. Its result is also called *the converging together lemma* or *Slutsky's lemma*.

Exercise 3.2.8. *Suppose that $X_n \xrightarrow{\mathcal{D}} X_\infty$ and $Y_n \xrightarrow{\mathcal{D}} Y_\infty$, where Y_∞ is non-random and for each n the variables X_n and Y_n are defined on the same probability space.*

(a) *Show that then $X_n + Y_n \xrightarrow{\mathcal{D}} X_\infty + Y_\infty$.*
Hint: *Recall that the collection of continuity points of F_{X_∞} is dense.*
(b) *Deduce that if $Z_n - X_n \xrightarrow{\mathcal{D}} 0$ then $X_n \xrightarrow{\mathcal{D}} X$ if and only if $Z_n \xrightarrow{\mathcal{D}} X$.*
(c) *Show that $Y_n X_n \xrightarrow{\mathcal{D}} Y_\infty X_\infty$.*

For example, here is an application of Exercise 3.2.8 en-route to a CLT connected to *renewal theory*.

Exercise 3.2.9.

(a) *Suppose $\{N_m\}$ are non-negative integer-valued random variables and $b_m \to \infty$ are non-random integers such that $N_m/b_m \xrightarrow{p} 1$. Show that if $S_n = \sum_{k=1}^n X_k$ for i.i.d. random variables $\{X_k\}$ with $v = \mathsf{Var}(X_1) \in (0,\infty)$ and $\mathbf{E}(X_1) = 0$, then $S_{N_m}/\sqrt{vb_m} \xrightarrow{\mathcal{D}} G$ as $m \to \infty$.*
Hint: *Use Kolmogorov's inequality to show that $S_{N_m}/\sqrt{vb_m} - S_{b_m}/\sqrt{vb_m} \xrightarrow{p} 0$.*
(b) *Let $N_t = \sup\{n : S_n \leq t\}$ for $S_n = \sum_{k=1}^n Y_k$ and i.i.d. random variables $Y_k > 0$ such that $v = \mathsf{Var}(Y_1) \in (0,\infty)$ and $\mathbf{E}(Y_1) = 1$. Show that $(N_t - t)/\sqrt{vt} \xrightarrow{\mathcal{D}} G$ as $t \to \infty$.*

Theorem 3.2.7 is key to solving the following:

Exercise 3.2.10. *Suppose that $Z_n \xrightarrow{\mathcal{D}} Z_\infty$. Show that then $b_n(f(c + Z_n/b_n) - f(c))/f'(c) \xrightarrow{\mathcal{D}} Z_\infty$ for every positive constants $b_n \to \infty$ and every Borel function*

$f : \mathbb{R} \to \mathbb{R}$ (not necessarily continuous) that is differentiable at $c \in \mathbb{R}$, with a derivative $f'(c) \neq 0$.

Consider the following exercise as a cautionary note about your interpretation of Theorem 3.2.7.

Exercise 3.2.11. *Let $M_n = \sum_{k=1}^n \prod_{i=1}^k U_i$ and $W_n = \sum_{k=1}^n \prod_{i=k}^n U_i$, where $\{U_i\}$ are i.i.d. uniformly on $[0, c]$ and $c > 0$.*

(a) *Show that $M_n \xrightarrow{a.s.} M_\infty$ as $n \to \infty$, with M_∞ taking values in $[0, \infty]$.*
(b) *Prove that M_∞ is a.s. finite if and only if $c < e$ (but $\mathbf{E}M_\infty$ is finite only for $c < 2$).*
(c) *In case $c < e$ prove that $W_n \xrightarrow{\mathcal{D}} M_\infty$ as $n \to \infty$ while W_n can not have an almost sure limit. Explain why this does not contradict Theorem 3.2.7.*

The next exercise relates the decay (in n) of $\sup_s |F_{X_\infty}(s) - F_{X_n}(s)|$ to that of $\sup |\mathbf{E}h(X_n) - \mathbf{E}h(X_\infty)|$ over all functions $h : \mathbb{R} \mapsto [-M, M]$ with $\sup_x |h'(x)| \le L$.

Exercise 3.2.12. *Let $\Delta_n = \sup_s |F_{X_\infty}(s) - F_{X_n}(s)|$.*

(a) *Show that if $\sup_x |h(x)| \le M$ and $\sup_x |h'(x)| \le L$, then for any $b > a$, $C = 4M + L(b - a)$ and all n*
$$|\mathbf{E}h(X_n) - \mathbf{E}h(X_\infty)| \le C\Delta_n + 4M\mathbf{P}(X_\infty \notin [a, b]) .$$
(b) *Show that if $X_\infty \in [a, b]$ and $f_{X_\infty}(x) \ge \eta > 0$ for all $x \in [a, b]$, then $|Q_n(\alpha) - Q_\infty(\alpha)| \le \eta^{-1}\Delta_n$ for any $\alpha \in (\Delta_n, 1 - \Delta_n)$, where $Q_n(\alpha) = \sup\{x : F_{X_n}(x) < \alpha\}$ denotes α-quantile for the law of X_n. Using this, construct $Y_n \overset{\mathcal{D}}{=} X_n$ such that $\mathbf{P}(|Y_n - Y_\infty| > \eta^{-1}\Delta_n) \le 2\Delta_n$ and deduce the bound of part (a), albeit the larger value $4M + L/\eta$ of C.*

Here is another example of convergence in distribution, this time in the context of extreme value theory.

Exercise 3.2.13. *Let $M_n = \max_{1 \le i \le n}\{T_i\}$, where T_i, $i = 1, 2, \ldots$ are i.i.d. random variables of distribution function $F_T(t)$. Noting that $F_{M_n}(x) = F_T(x)^n$, show that $b_n^{-1}(M_n - a_n) \xrightarrow{\mathcal{D}} M_\infty$ when:*

(a) *$F_T(t) = 1 - e^{-t}$ for $t \ge 0$ (i.e. T_i are Exponential of parameter one). Here, $a_n = \log n$, $b_n = 1$ and $F_{M_\infty}(y) = \exp(-e^{-y})$ for $y \in \mathbb{R}$.*
(b) *$F_T(t) = 1 - t^{-\alpha}$ for $t \ge 1$ and $\alpha > 0$. Here, $a_n = 0$, $b_n = n^{1/\alpha}$ and $F_{M_\infty}(y) = \exp(-y^{-\alpha})$ for $y > 0$.*
(c) *$F_T(t) = 1 - |t|^\alpha$ for $-1 \le t \le 0$ and $\alpha > 0$. Here, $a_n = 0$, $b_n = n^{-1/\alpha}$ and $F_{M_\infty}(y) = \exp(-|y|^\alpha)$ for $y \le 0$.*

Remark. Up to the linear transformation $y \mapsto (y - \mu)/\sigma$, the three distributions of M_∞ provided in Exercise 3.2.13 are the only possible limits of maxima of i.i.d. random variables. They are thus called the *extreme value distributions* of Type 1 (or Gumbel-type), in case (a), Type 2 (or Fréchet-type), in case (b), and Type 3 (or Weibull-type), in case (c). Indeed,

Exercise 3.2.14.

(a) *Building upon part* (a) *of Exercise 2.2.24, show that if G has the standard normal distribution, then for any $y \in \mathbb{R}$*
$$\lim_{t\to\infty} \frac{1 - F_G(t + y/t)}{1 - F_G(t)} = e^{-y} .$$

(b) *Let* $M_n = \max_{1\le i\le n}\{G_i\}$ *for i.i.d. standard normal random variables* G_i. *Show that* $b_n(M_n - b_n) \xrightarrow{\mathcal{D}} M_\infty$ *where* $F_{M_\infty}(y) = \exp(-e^{-y})$ *and* b_n *is such that* $1 - F_G(b_n) = n^{-1}$.

(c) *Show that* $b_n/\sqrt{2\log n} \to 1$ *as* $n \to \infty$ *and deduce that* $M_n/\sqrt{2\log n} \xrightarrow{p} 1$.

(d) *More generally, suppose* $T_t = \inf\{x \ge 0 : M_x \ge t\}$, *where* $x \mapsto M_x$ *is some monotone non-decreasing family of random variables such that* $M_0 = 0$. *Show that if* $e^{-t}T_t \xrightarrow{\mathcal{D}} T_\infty$ *as* $t \to \infty$ *with* T_∞ *having the standard* exponential distribution *then* $(M_x - \log x) \xrightarrow{\mathcal{D}} M_\infty$ *as* $x \to \infty$, *where* $F_{M_\infty}(y) = \exp(-e^{-y})$.

Our next example is of a more combinatorial flavor.

Exercise 3.2.15 (THE BIRTHDAY PROBLEM)**.** *Suppose* $\{X_i\}$ *are i.i.d. with each* X_i *uniformly distributed on* $\{1,\dots,n\}$. *Let* $T_n = \min\{k : X_k = X_l, \text{ for some } l < k\}$ *mark the first coincidence among the entries of the sequence* $X_1, X_2, \dots$, *so*

$$\mathbf{P}(T_n > r) = \prod_{k=2}^{r}(1 - \frac{k-1}{n}),$$

is the probability that among r *items chosen uniformly and independently from a set of* n *different objects, no two are the same (the name "birthday problem" corresponds to* $n = 365$ *with the items interpreted as the birthdays for a group of size* r*). Show that* $\mathbf{P}(n^{-1/2}T_n > s) \to \exp(-s^2/2)$ *as* $n \to \infty$, *for any fixed* $s \ge 0$.
Hint: *Recall that* $-x - x^2 \le \log(1-x) \le -x$ *for* $x \in [0, 1/2]$.

The *symmetric, simple random walk* on the integers is the sequence of random variables $S_n = \sum_{k=1}^n \xi_k$ where ξ_k are i.i.d. such that $\mathbf{P}(\xi_k = 1) = \mathbf{P}(\xi_k = -1) = \frac{1}{2}$. From the CLT we already know that $n^{-1/2}S_n \xrightarrow{\mathcal{D}} G$. The next exercise provides the asymptotics of the first and last visits to zero by this random sequence, namely $R = \inf\{\ell \ge 1 : S_\ell = 0\}$ and $L_n = \sup\{\ell \le n : S_\ell = 0\}$. Much more is known about this random sequence (c.f. [**Dur10**, Section 4.3] or [**Fel68**, Chapter 3]).

Exercise 3.2.16. *Let* $q_{n,r} = \mathbf{P}(S_1 > 0, \dots, S_{n-1} > 0, S_n = r)$ *and*

$$p_{n,r} = \mathbf{P}(S_n = r) = 2^{-n}\binom{n}{k} \qquad k = (n+r)/2\,.$$

(a) *Counting paths of the walk, prove the* discrete reflection principle *that* $\mathbf{P}_x(R < n, S_n = y) = \mathbf{P}_{-x}(S_n = y) = p_{n,x+y}$ *for any positive integers* x, y, *where* $\mathbf{P}_x(\cdot)$ *denote probabilities for the walk starting at* $S_0 = x$.

(b) *Verify that* $q_{n,r} = \frac{1}{2}(p_{n-1,r-1} - p_{n-1,r+1})$ *for any* $n, r \ge 1$.
Hint: *Paths of the walk contributing to* $q_{n,r}$ *must have* $S_1 = 1$. *Hence, use part (a) with* $x = 1$ *and* $y = r$.

(c) *Deduce that* $\mathbf{P}(R > n) = p_{n-1,0} + p_{n-1,1}$ *and that* $\mathbf{P}(L_{2n} = 2k) = p_{2k,0}\,p_{2n-2k,0}$ *for* $k = 0, 1, \dots, n$.

(d) *Using* Stirling's formula *(that* $\sqrt{2\pi n}(n/e)^n/n! \to 1$ *as* $n \to \infty$*), show that* $\sqrt{\pi n}\,\mathbf{P}(R > 2n) \to 1$ *and that* $(2n)^{-1}L_{2n} \xrightarrow{\mathcal{D}} X$, *where* X *has the* arc-sine *probability density function* $f_X(x) = \frac{1}{\pi\sqrt{x(1-x)}}$ *on* $[0, 1]$.

(e) *Let* H_{2n} *count the number of* $1 \le k \le 2n$ *such that* $S_k \ge 0$ *and* $S_{k-1} \ge 0$. *Show that* $H_{2n} \overset{\mathcal{D}}{=} L_{2n}$, *hence* $(2n)^{-1}H_{2n} \xrightarrow{\mathcal{D}} X$.

3.2.2. Weak convergence of probability measures. We first extend the definition of weak convergence from distribution functions to measures on Borel σ-algebras.

Definition 3.2.17. *For a topological space* $\mathbb{S}$*, let* $C_b(\mathbb{S})$ *denote the collection of all* continuous bounded functions *on* $\mathbb{S}$*. We say that a sequence of probability measures* ν_n *on a topological space* $\mathbb{S}$ *equipped with its Borel* σ*-algebra (see Example 1.1.15),* converges weakly *to a probability measure* ν_∞*, denoted* $\nu_n \stackrel{w}{\Rightarrow} \nu_\infty$*, if* $\nu_n(h) \to \nu_\infty(h)$ *for each* $h \in C_b(\mathbb{S})$*.*

As we show next, Definition 3.2.17 is an alternative definition of convergence in distribution, which, in contrast to Definition 3.2.1, applies to more general R.V. (for example to the $\mathbb{R}^d$-valued random variables we consider in Section 3.5).

Proposition 3.2.18. *The weak convergence of distribution functions is equivalent to the weak convergence of the corresponding laws as probability measures on* $(\mathbb{R}, \mathcal{B})$*. Consequently,* $X_n \xrightarrow{\mathcal{D}} X_\infty$ *if and only if for each* $h \in C_b(\mathbb{R})$*, we have* $\mathbf{E}h(X_n) \to \mathbf{E}h(X_\infty)$ *as* $n \to \infty$*.*

Proof. Suppose first that $F_n \stackrel{w}{\to} F_\infty$ and let Y_n, $1 \le n \le \infty$ be the random variables given by Theorem 3.2.7 such that $Y_n \stackrel{a.s.}{\to} Y_\infty$. For $h \in C_b(\mathbb{R})$ we have by continuity of h that $h(Y_n) \stackrel{a.s.}{\to} h(Y_\infty)$, and by bounded convergence also

$$\mathcal{P}_n(h) = \mathbf{E}(h(Y_n)) \to \mathbf{E}(h(Y_\infty)) = \mathcal{P}_\infty(h)\,.$$

Conversely, suppose that $\mathcal{P}_n \stackrel{w}{\Rightarrow} \mathcal{P}_\infty$ per Definition 3.2.17. Fixing $\alpha \in \mathbb{R}$, let the non-negative $h_k^\pm \in C_b(\mathbb{R})$ be such that $h_k^-(x) \uparrow I_{(-\infty,\alpha)}(x)$ and $h_k^+(x) \downarrow I_{(-\infty,\alpha]}(x)$ as $k \to \infty$ (c.f. Lemma 3.1.6 for a construction of such functions). We have by the weak convergence of the laws when $n \to \infty$, followed by monotone convergence as $k \to \infty$, that

$$\liminf_{n\to\infty} \mathcal{P}_n((-\infty,\alpha)) \ge \lim_{n\to\infty} \mathcal{P}_n(h_k^-) = \mathcal{P}_\infty(h_k^-) \uparrow \mathcal{P}_\infty((-\infty,\alpha)) = F_\infty(\alpha^-)\,.$$

Similarly, considering $h_k^+(\cdot)$ and then $k \to \infty$, we have by bounded convergence that

$$\limsup_{n\to\infty} \mathcal{P}_n((-\infty,\alpha]) \le \lim_{n\to\infty} \mathcal{P}_n(h_k^+) = \mathcal{P}_\infty(h_k^+) \downarrow \mathcal{P}_\infty((-\infty,\alpha]) = F_\infty(\alpha)\,.$$

For any continuity point α of F_∞ we conclude that $F_n(\alpha) = \mathcal{P}_n((-\infty,\alpha])$ converges as $n \to \infty$ to $F_\infty(\alpha) = F_\infty(\alpha^-)$, thus completing the proof. $\square$

By yet another application of Theorem 3.2.7 we find that convergence in distribution is preserved under a.s. continuous mappings (see Corollary 2.2.13 for the analogous statement for convergence in probability).

Proposition 3.2.19 (Continuous mapping)**.** *For a Borel function* g *let* $\mathbf{D}_g$ *denote its set of points of discontinuity. If* $X_n \xrightarrow{\mathcal{D}} X_\infty$ *and* $\mathbf{P}(X_\infty \in \mathbf{D}_g) = 0$*, then* $g(X_n) \xrightarrow{\mathcal{D}} g(X_\infty)$*. If in addition* g *is bounded then* $\mathbf{E}g(X_n) \to \mathbf{E}g(X_\infty)$*.*

Proof. Given $X_n \xrightarrow{\mathcal{D}} X_\infty$, by Theorem 3.2.7 there exists $Y_n \stackrel{\mathcal{D}}{=} X_n$, such that $Y_n \xrightarrow{a.s.} Y_\infty$. Fixing $h \in C_b(\mathbb{R})$, clearly $\mathbf{D}_{h\circ g} \subseteq \mathbf{D}_g$, so

$$\mathbf{P}(Y_\infty \in \mathbf{D}_{h\circ g}) \le \mathbf{P}(Y_\infty \in \mathbf{D}_g) = 0.$$

Therefore, by Exercise 2.2.12, it follows that $h(g(Y_n)) \xrightarrow{a.s.} h(g(Y_\infty))$. Since $h \circ g$ is bounded and $Y_n \overset{\mathcal{D}}{=} X_n$ for all n, it follows by bounded convergence that

$$\mathbf{E}h(g(X_n)) = \mathbf{E}h(g(Y_n)) \to \mathbf{E}(h(g(Y_\infty)) = \mathbf{E}h(g(X_\infty)) .$$

This holds for any $h \in C_b(\mathbb{R})$, so by Proposition 3.2.18, we conclude that $g(X_n) \xrightarrow{\mathcal{D}} g(X_\infty)$. □

Our next theorem collects several equivalent characterizations of weak convergence of probability measures on $(\mathbb{R}, \mathcal{B})$. To this end we need the following definition.

Definition 3.2.20. *For a subset A of a topological space $\mathbb{S}$, we denote by ∂A the* boundary *of A, that is $\partial A = \overline{A} \setminus A^o$ is the closed set of points in the closure of A but not in the interior of A. For a measure μ on $(\mathbb{S}, \mathcal{B}_\mathbb{S})$ we say that $A \in \mathcal{B}_\mathbb{S}$ is a* μ-continuity set *if $\mu(\partial A) = 0$.*

Theorem 3.2.21 (PORTMANTEAU THEOREM)**.** *The following four statements are equivalent for any probability measures ν_n, $1 \le n \le \infty$ on $(\mathbb{R}, \mathcal{B})$.*

(a) $\nu_n \overset{w}{\Rightarrow} \nu_\infty$
(b) *For every closed set F, one has* $\limsup_{n\to\infty} \nu_n(F) \le \nu_\infty(F)$
(c) *For every open set G, one has* $\liminf_{n\to\infty} \nu_n(G) \ge \nu_\infty(G)$
(d) *For every ν_∞-continuity set A, one has* $\lim_{n\to\infty} \nu_n(A) = \nu_\infty(A)$

Remark. As shown in Subsection 3.5.1, this theorem holds with $(\mathbb{R}, \mathcal{B})$ replaced by any metric space $\mathbb{S}$ and its Borel σ-algebra $\mathcal{B}_\mathbb{S}$.

For $\nu_n = \mathcal{P}_{X_n}$ we get the formulation of the Portmanteau theorem for random variables X_n, $1 \le n \le \infty$, where the following four statements are then equivalent to $X_n \xrightarrow{\mathcal{D}} X_\infty$:

(a) $\mathbf{E}h(X_n) \to \mathbf{E}h(X_\infty)$ for each bounded continuous h
(b) For every closed set F one has $\limsup_{n\to\infty} \mathbf{P}(X_n \in F) \le \mathbf{P}(X_\infty \in F)$
(c) For every open set G one has $\liminf_{n\to\infty} \mathbf{P}(X_n \in G) \ge \mathbf{P}(X_\infty \in G)$
(d) For every Borel set A such that $\mathbf{P}(X_\infty \in \partial A) = 0$, one has $\lim_{n\to\infty} \mathbf{P}(X_n \in A) = \mathbf{P}(X_\infty \in A)$

PROOF. It suffices to show that $(a) \Rightarrow (b) \Rightarrow (c) \Rightarrow (d) \Rightarrow (a)$, which we shall establish in that order. To this end, with $F_n(x) = \nu_n((-\infty, x])$ denoting the corresponding distribution functions, we replace $\nu_n \overset{w}{\Rightarrow} \nu_\infty$ of (a) by the equivalent condition $F_n \overset{w}{\to} F_\infty$ (see Proposition 3.2.18).
$(a) \Rightarrow (b)$. Assuming $F_n \overset{w}{\to} F_\infty$, we have the random variables Y_n, $1 \le n \le \infty$ of Theorem 3.2.7, such that $\mathcal{P}_{Y_n} = \nu_n$ and $Y_n \overset{a.s.}{\to} Y_\infty$. Since F is closed, the function I_F is upper semi-continuous bounded by one, so it follows that a.s.

$$\limsup_{n\to\infty} I_F(Y_n) \le I_F(Y_\infty) ,$$

and by Fatou's lemma,

$$\limsup_{n\to\infty} \nu_n(F) = \limsup_{n\to\infty} \mathbf{E}I_F(Y_n) \le \mathbf{E} \limsup_{n\to\infty} I_F(Y_n) \le \mathbf{E}I_F(Y_\infty) = \nu_\infty(F) ,$$

as stated in (b).

$(b) \Rightarrow (c)$. The complement $F = G^c$ of an open set G is a closed set, so by (b) we have that

$$1 - \liminf_{n\to\infty} \nu_n(G) = \limsup_{n\to\infty} \nu_n(G^c) \le \nu_\infty(G^c) = 1 - \nu_\infty(G)\,,$$

implying that (c) holds. In an analogous manner we can show that $(c) \Rightarrow (b)$, so (b) and (c) are equivalent.

$(c) \Rightarrow (d)$. Since (b) and (c) are equivalent, we assume now that both (b) and (c) hold. Then, applying (c) for the open set $G = A^o$ and (b) for the closed set $F = \overline{A}$ we have that

$$\begin{aligned} \nu_\infty(\overline{A}) &\ge \limsup_{n\to\infty} \nu_n(\overline{A}) \ge \limsup_{n\to\infty} \nu_n(A) \\ &\ge \liminf_{n\to\infty} \nu_n(A) \ge \liminf_{n\to\infty} \nu_n(A^o) \ge \nu_\infty(A^o)\,. \end{aligned} \tag{3.2.4}$$

Further, $\overline{A} = A^o \cup \partial A$ so $\nu_\infty(\partial A) = 0$ implies that $\nu_\infty(\overline{A}) = \nu_\infty(A^o) = \nu_\infty(A)$ (with the last equality due to the fact that $A^o \subseteq A \subseteq \overline{A}$). Consequently, for such a set A all the inequalities in (3.2.4) are equalities, yielding (d).

$(d) \Rightarrow (a)$. Consider the set $A = (-\infty, \alpha]$ where α is a continuity point of F_∞. Then, $\partial A = \{\alpha\}$ and $\nu_\infty(\{\alpha\}) = F_\infty(\alpha) - F_\infty(\alpha^-) = 0$. Applying (d) for this choice of A, we have that

$$\lim_{n\to\infty} F_n(\alpha) = \lim_{n\to\infty} \nu_n((-\infty,\alpha]) = \nu_\infty((-\infty,\alpha]) = F_\infty(\alpha)\,,$$

which is our version of (a). $\square$

We turn to relate the weak convergence to the convergence point-wise of probability density functions. To this end, we first define a new concept of convergence of measures, the *convergence in total-variation*.

Definition 3.2.22. *The total variation norm of a finite signed measure ν on the measurable space $(\mathbb{S}, \mathcal{F})$ is*

$$\|\nu\|_{tv} = \sup\{\nu(h) : h \in m\mathcal{F}, \sup_{s\in\mathbb{S}} |h(s)| \le 1\}.$$

We say that a sequence of probability measures ν_n converges in total variation to a probability measure ν_∞, denoted $\nu_n \xrightarrow{t.v.} \nu_\infty$, if $\|\nu_n - \nu_\infty\|_{tv} \to 0$.

Remark. Note that $\|\nu\|_{tv} = 1$ for any probability measure ν (since $\nu(h) \le \nu(|h|) \le \|h\|_\infty \nu(1) \le 1$ for the functions h considered, with equality for $h = 1$). By a similar reasoning, $\|\nu - \nu'\|_{tv} \le 2$ for any two probability measures ν, ν' on $(\mathbb{S}, \mathcal{F})$.

Convergence in total-variation obviously implies weak convergence of the same probability measures, but the converse fails, as demonstrated for example by $\nu_n = \delta_{1/n}$, the probability measure on $(\mathbb{R}, \mathcal{B})$ assigning probability one to the point $1/n$, which converge weakly to $\nu_\infty = \delta_0$ (see Example 3.2.4), whereas $\|\nu_n - \nu_\infty\| = 2$ for all n. The difference of course has to do with the non-uniformity of the weak convergence with respect to the continuous function h.

To gain a better understanding of the convergence in total-variation, we consider an important special case.

Proposition 3.2.23. *Suppose $\mathbf{P} = f\mu$ and $\mathbf{Q} = g\mu$ for some measure μ on $(\mathbb{S}, \mathcal{F})$ and $f, g \in m\mathcal{F}_+$ such that $\mu(f) = \mu(g) = 1$. Then,*

$$\|\mathbf{P} - \mathbf{Q}\|_{tv} = \int_{\mathbb{S}} |f(s) - g(s)| d\mu(s)\,. \tag{3.2.5}$$

Further, suppose $\nu_n = f_n\mu$ with $f_n \in m\mathcal{F}_+$ such that $\mu(f_n) = 1$ for all $n \le \infty$. Then, $\nu_n \xrightarrow{t.v.} \nu_\infty$ if $f_n(s) \to f_\infty(s)$ for μ-almost-every $s \in \mathbb{S}$.

Proof. For any measurable function $h : \mathbb{S} \mapsto [-1, 1]$ we have that

$$(f\mu)(h) - (g\mu)(h) = \mu(fh) - \mu(gh) = \mu((f - g)h) \le \mu(|f - g|)\,,$$

with equality when $h(s) = \text{sgn}((f(s)-g(s))$ (see Proposition 1.3.56 for the left-most identity and note that fh and gh are in $L^1(\mathbb{S}, \mathcal{F}, \mu)$). Consequently, $\|\mathbf{P} - \mathbf{Q}\|_{tv} = \sup\{(f\mu)(h) - (g\mu)(h) : h \text{ as above }\} = \mu(|f - g|)$, as claimed.

For $\nu_n = f_n\mu$, we thus have that $\|\nu_n - \nu_\infty\|_{tv} = \mu(|f_n - f_\infty|)$, so the convergence in total-variation is equivalent to $f_n \to f_\infty$ in $L^1(\mathbb{S}, \mathcal{F}, \mu)$. Since $f_n \ge 0$ and $\mu(f_n) = 1$ for any $n \le \infty$, it follows from Scheffé's lemma (see Lemma 1.3.35) that the latter convergence is a consequence of $f_n(s) \to f_\infty(s)$ for μ a.e. $s \in \mathbb{S}$. ☐

Two specific instances of Proposition 3.2.23 are of particular value in applications.

Example 3.2.24. *Let $\nu_n = \mathcal{P}_{X_n}$ denote the laws of random variables X_n that have probability density functions f_n, $n = 1, 2, \ldots, \infty$. Recall Exercise 1.3.66 that then $\nu_n = f_n\lambda$ for Lebesgue's measure λ on $(\mathbb{R}, \mathcal{B})$. Hence, by the preceding proposition, the convergence point-wise of $f_n(x)$ to $f_\infty(x)$ implies the convergence in total-variation of $\mathcal{P}_{X_n}$ to $\mathcal{P}_{X_\infty}$, and in particular implies that $X_n \xrightarrow{\mathcal{D}} X_\infty$.*

Example 3.2.25. *Similarly, if X_n are integer valued for $n = 1, 2 \ldots$, then $\nu_n = f_n\widetilde{\lambda}$ for $f_n(k) = \mathbf{P}(X_n = k)$ and the* counting measure *$\widetilde{\lambda}$ on $(\mathbf{Z}, 2^{\mathbf{Z}})$ such that $\widetilde{\lambda}(\{k\}) = 1$ for each $k \in \mathbf{Z}$. So, by the preceding proposition, the point-wise convergence of Exercise 3.2.3 is not only necessary and sufficient for weak convergence but also for convergence in total-variation of the laws of X_n to that of X_∞.*

In the next exercise, you are to rephrase Example 3.2.25 in terms of the topological space of all probability measures on $\mathbf{Z}$.

Exercise 3.2.26. *Show that $d(\mu, \nu) = \|\mu - \nu\|_{tv}$ is a metric on the collection of all probability measures on $\mathbf{Z}$, and that in this space the convergence in total variation is equivalent to the weak convergence which in turn is equivalent to the point-wise convergence at each $x \in \mathbf{Z}$.*

Hence, under the framework of Example 3.2.25, the Glivenko-Cantelli theorem tells us that the empirical measures of integer valued i.i.d. R.V.-s $\{X_i\}$ converge in total-variation to the true law of X_1.

Here is an example from statistics that corresponds to the framework of Example 3.2.24.

Exercise 3.2.27. *Let V_{n+1} denote the central value on a list of $2n+1$ values (that is, the $(n + 1)$th largest value on the list). Suppose the list consists of mutually independent R.V., each chosen uniformly in $[0, 1)$.*

(a) *Show that V_{n+1} has probability density function $(2n + 1)\binom{2n}{n}v^n(1 - v)^n$ at each $v \in [0, 1)$.*
(b) *Verify that the density $f_n(v)$ of $\widehat{V}_n = \sqrt{2n}(2V_{n+1} - 1)$ is of the form $f_n(v) = c_n(1 - v^2/(2n))^n$ for some normalization constant c_n that is independent of $|v| \le \sqrt{2n}$.*
(c) *Deduce that for $n \to \infty$ the densities $f_n(v)$ converge point-wise to the standard normal density, and conclude that $\widehat{V}_n \xrightarrow{\mathcal{D}} G$.*

Here is an interesting interpretation of the CLT in terms of weak convergence of probability measures.

Exercise 3.2.28. *Let $\mathcal{M}$ denote the set of probability measures ν on $(\mathbb{R},\mathcal{B})$ for which $\int x^2 d\nu(x) = 1$ and $\int x d\nu(x) = 0$, and $\gamma \in \mathcal{M}$ denote the standard normal distribution. Consider the mapping $T : \mathcal{M} \mapsto \mathcal{M}$ where $T\nu$ is the law of $(X_1 + X_2)/\sqrt{2}$ for X_1 and X_2 i.i.d. of law ν each. Explain why the* CLT *implies that $T^m\nu \overset{w}{\Rightarrow} \gamma$ as $m \to \infty$, for any $\nu \in \mathcal{M}$. Show that $T\gamma = \gamma$ (see Lemma 3.1.1), and explain why γ is the unique, globally attracting fixed point of T in $\mathcal{M}$.*

Your next exercise is the basis behind the celebrated *method of moments* for weak convergence.

Exercise 3.2.29. *Suppose that X and Y are $[0,1]$-valued random variables such that $\mathbf{E}(X^n) = \mathbf{E}(Y^n)$ for $n = 0,1,2,\dots$.*

(a) *Show that $\mathbf{E}p(X) = \mathbf{E}p(Y)$ for any polynomial $p(\cdot)$.*
(b) *Show that $\mathbf{E}h(X) = \mathbf{E}h(Y)$ for any continuous function $h : [0,1] \mapsto \mathbb{R}$ and deduce that $X \overset{\mathcal{D}}{=} Y$.*

Hint: *Recall Weierstrass approximation theorem, that if h is continuous on $[0,1]$ then there exist polynomials p_n such that $\sup_{x\in[0,1]} |h(x) - p_n(x)| \to 0$ as $n \to \infty$.*

We conclude with the following example about weak convergence of measures in the space of infinite binary sequences.

Exercise 3.2.30. *Consider the topology of coordinate wise convergence on $\mathbb{S} = \{0,1\}^{\mathbb{N}}$ and the Borel probability measures $\{\nu_n\}$ on $\mathbb{S}$, where ν_n is the uniform measure over the $\binom{2n}{n}$ binary sequences of precisely n ones among the first $2n$ coordinates, followed by zeros from position $2n+1$ onwards. Show that $\nu_n \overset{w}{\Rightarrow} \nu_\infty$ where ν_∞ denotes the law of i.i.d. Bernoulli random variables of parameter $p = 1/2$.*
Hint: *Any open subset of $\mathbb{S}$ is a countable union of disjoint sets of the form $A_{\theta,k} = \{\omega \in \mathbb{S} : \omega_i = \theta_i, i = 1,\dots,k\}$ for some $\theta = (\theta_1,\dots,\theta_k) \in \{0,1\}^k$ and $k \in \mathbb{N}$.*

3.2.3. Uniform tightness and vague convergence. So far we have studied the properties of weak convergence. We turn to deal with general ways to establish such convergence, a subject to which we return in Subsection 3.3.2. To this end, the most important concept is that of *uniform tightness*, which we now define.

Definition 3.2.31. *We say that a probability measure μ on $(\mathbb{S},\mathcal{B}_{\mathbb{S}})$ is* tight *if for each $\varepsilon > 0$ there exists a compact set $K_\varepsilon \subseteq \mathbb{S}$ such that $\mu(K_\varepsilon^c) < \varepsilon$. A collection $\{\mu_\beta\}$ of probability measures on $(\mathbb{S},\mathcal{B}_{\mathbb{S}})$ is called* uniformly tight *if for each $\varepsilon > 0$ there exists* one *compact set K_ε such that $\mu_\beta(K_\varepsilon^c) < \varepsilon$ for all β.*

Since bounded closed intervals are compact and $[-M,M]^c \downarrow \emptyset$ as $M \uparrow \infty$, by continuity from above we deduce that each probability measure μ on $(\mathbb{R},\mathcal{B})$ is tight. The same argument applies for a finite collection of probability measures on $(\mathbb{R},\mathcal{B})$ (just choose the maximal value among the finitely many values of $M = M_\varepsilon$ that are needed for the different measures). Further, in the case of $\mathbb{S} = \mathbb{R}$ which we study here one can take without loss of generality the compact K_ε as a symmetric bounded interval $[-M_\varepsilon, M_\varepsilon]$, or even consider instead $(-M_\varepsilon, M_\varepsilon]$ (whose closure is compact) in order to simplify notations. Thus, expressing uniform tightness in terms of the corresponding distribution functions leads in this setting to the following alternative definition.

Definition 3.2.32. *A sequence of distribution functions F_n is called* uniformly tight*, if for every $\varepsilon > 0$ there exists $M = M_\varepsilon$ such that*

$$\limsup_{n\to\infty} [1 - F_n(M) + F_n(-M)] < \varepsilon \,.$$

Remark. As most texts use in the context of Definition 3.2.32 "tight" (or "tight sequence") instead of uniformly tight, we shall adopt the same convention here.

Uniform tightness of distribution functions has some structural resemblance to the U.I. condition (1.3.11). As such we have the following simple sufficient condition for uniform tightness (which is the analog of Exercise 1.3.54).

Exercise 3.2.33. *A sequence of probability measures ν_n on $(\mathbb{R}, \mathcal{B})$ is uniformly tight if $\sup_n \nu_n(f(|x|))$ is finite for some non-negative Borel function such that $f(r) \to \infty$ as $r \to \infty$. Alternatively, if $\sup_n Ef(|X_n|) < \infty$ then the distribution functions F_{X_n} form a tight sequence.*

The importance of uniform tightness is that it guarantees the existence of limit points for weak convergence.

Theorem 3.2.34 (Prohorov theorem). *A collection Γ of probability measures on a complete, separable metric space $\mathbb{S}$ equipped with its Borel σ-algebra $\mathcal{B}_{\mathbb{S}}$, is uniformly tight if and only if for any sequence $\nu_m \in \Gamma$ there exists a subsequence ν_{m_k} that converges weakly to some probability measure ν_∞ on $(\mathbb{S}, \mathcal{B}_{\mathbb{S}})$ (where ν_∞ is not necessarily in Γ and may depend on the subsequence m_k).*

Remark. For a proof of Prohorov's theorem, which is beyond the scope of these notes, see [**Dud89**, Theorem 11.5.4].

Instead of Prohorov's theorem, we prove here a bare-hands substitute for the special case $\mathbb{S} = \mathbb{R}$. When doing so, it is convenient to have the following notion of convergence of distribution functions.

Definition 3.2.35. *When a sequence F_n of distribution functions converges to a right continuous, non-decreasing function F_∞ at all continuous points of F_∞, we say that F_n* converges vaguely *to F_∞, denoted $F_n \xrightarrow{v} F_\infty$.*

In contrast with weak convergence, the vague convergence allows for the limit $F_\infty(x) = \nu_\infty((-\infty, x])$ to correspond to a measure ν_∞ such that $\nu_\infty(\mathbb{R}) < 1$.

Example 3.2.36. *Suppose $F_n = pI_{[n,\infty)} + qI_{[-n,\infty)} + (1-p-q)F$ for some $p, q \geq 0$ such that $p+q \leq 1$ and a distribution function F that is independent of n. It is easy to check that $F_n \xrightarrow{v} F_\infty$ as $n \to \infty$, where $F_\infty = q + (1-p-q)F$ is the distribution function of an $\overline{\mathbb{R}}$-valued random variable, with probability mass p at $+\infty$ and mass q at $-\infty$. If $p + q > 0$ then F_∞ is not a distribution function of any measure on $\mathbb{R}$ and F_n does not converge weakly.*

The preceding example is generic, that is, the space $\overline{\mathbb{R}}$ is compact, so the only loss of mass when dealing with weak convergence on $\mathbb{R}$ has to do with its escape to $\pm\infty$. It is thus not surprising that *every* sequence of distribution functions have vague limit points, as stated by the following theorem.

Theorem 3.2.37 (Helly's selection theorem). *For every sequence F_n of distribution functions, there is a subsequence F_{n_k} and a non-decreasing right continuous function F_∞ such that $F_{n_k}(y) \to F_\infty(y)$ as $k \to \infty$ at all continuity points y of F_∞, that is $F_{n_k} \xrightarrow{v} F_\infty$.*

Deferring the proof of Helly's theorem to the end of this section, uniform tightness is exactly what prevents probability mass from escaping to $\pm\infty$, thus assuring the existence of limit points for weak convergence.

Lemma 3.2.38. *The sequence of distribution functions $\{F_n\}$ is uniformly tight if and only if each vague limit point of this sequence is a distribution function. That is, if and only if when $F_{n_k} \stackrel{v}{\to} F$, necessarily $1 - F(x) + F(-x) \to 0$ as $x \to \infty$.*

Proof. Suppose first that $\{F_n\}$ is uniformly tight and $F_{n_k} \stackrel{v}{\to} F$. Fixing $\varepsilon > 0$, there exist $r_1 < -M_\varepsilon$ and $r_2 > M_\varepsilon$ that are both continuity points of F. Then, by the definition of vague convergence and the monotonicity of F_n,

$$\begin{aligned} 1 - F(r_2) + F(r_1) &= \lim_{k\to\infty} (1 - F_{n_k}(r_2) + F_{n_k}(r_1)) \\ &\le \limsup_{n\to\infty} (1 - F_n(M_\varepsilon) + F_n(-M_\varepsilon)) < \varepsilon\ . \end{aligned}$$

It follows that $\limsup_{x\to\infty}(1 - F(x) + F(-x)) \le \varepsilon$ and since $\varepsilon > 0$ is arbitrarily small, F must be a distribution function of some probability measure on $(\mathbb{R}, \mathcal{B})$.

Conversely, suppose $\{F_n\}$ is not uniformly tight, in which case by Definition 3.2.32, for some $\varepsilon > 0$ and $n_k \uparrow \infty$

$$1 - F_{n_k}(k) + F_{n_k}(-k) \ge \varepsilon \qquad \text{for all } k. \tag{3.2.6}$$

By Helly's theorem, there exists a vague limit point F to F_{n_k} as $k \to \infty$. That is, for some $k_l \uparrow \infty$ as $l \to \infty$ we have that $F_{n_{k_l}} \stackrel{v}{\to} F$. For any two continuity points $r_1 < 0 < r_2$ of F, we thus have by the definition of vague convergence, the monotonicity of $F_{n_{k_l}}$, and (3.2.6), that

$$\begin{aligned} 1 - F(r_2) + F(r_1) &= \lim_{l\to\infty} (1 - F_{n_{k_l}}(r_2) + F_{n_{k_l}}(r_1)) \\ &\ge \liminf_{l\to\infty} (1 - F_{n_{k_l}}(k_l) + F_{n_{k_l}}(-k_l)) \ge \varepsilon. \end{aligned}$$

Considering now $r = \min(-r_1, r_2) \to \infty$, this shows that $\inf_r(1-F(r)+F(-r)) \ge \varepsilon$, hence the vague limit point F cannot be a distribution function of a probability measure on $(\mathbb{R}, \mathcal{B})$. ∎

Remark. Comparing Definitions 3.2.31 and 3.2.32 we see that if a collection Γ of probability measures on $(\mathbb{R}, \mathcal{B})$ is uniformly tight, then for any sequence $\nu_m \in \Gamma$ the corresponding sequence F_m of distribution functions is uniformly tight. In view of Lemma 3.2.38 and Helly's theorem, this implies the existence of a subsequence m_k and a distribution function F_∞ such that $F_{m_k} \stackrel{w}{\to} F_\infty$. By Proposition 3.2.18 we deduce that $\nu_{m_k} \stackrel{w}{\Rightarrow} \nu_\infty$, a probability measure on $(\mathbb{R}, \mathcal{B})$, thus proving the only direction of Prohorov's theorem that we ever use.

Proof of Theorem 3.2.37. Fix a sequence of distribution function F_n. The key to the proof is to observe that there exists a sub-sequence n_k and a non-decreasing function $H : \mathbb{Q} \mapsto [0, 1]$ such that $F_{n_k}(q) \to H(q)$ for any $q \in \mathbb{Q}$.

This is done by a standard analysis argument called the principle of 'diagonal selection'. That is, let $q_1, q_2, \ldots$, be an enumeration of the set $\mathbb{Q}$ of all rational numbers. There exists then a limit point $H(q_1)$ to the sequence $F_n(q_1) \in [0, 1]$, that is a sub-sequence $n_k^{(1)}$ such that $F_{n_k^{(1)}}(q_1) \to H(q_1)$. Since $F_{n_k^{(1)}}(q_2) \in [0, 1]$, there exists a further sub-sequences $n_k^{(2)}$ of $n_k^{(1)}$ such that

$$F_{n_k^{(i)}}(q_i) \to H(q_i) \quad \text{for } i = 1, 2.$$

In the same manner we get a collection of nested sub-sequences $n_k^{(i)} \subseteq n_k^{(i-1)}$ such that

$$F_{n_k^{(i)}}(q_j) \to H(q_j), \quad \text{for all } j \leq i.$$

The diagonal $n_k^{(k)}$ then has the property that

$$F_{n_k^{(k)}}(q_j) \to H(q_j), \quad \text{for all } j,$$

so $n_k = n_k^{(k)}$ is our desired sub-sequence, and since each F_n is non-decreasing, the limit function H must also be non-decreasing on $\mathbb{Q}$.

Let $F_\infty(x) := \inf\{H(q) : q \in \mathbb{Q}, q > x\}$, noting that $F_\infty \in [0,1]$ is non-decreasing. Further, F_∞ is right continuous, since

$$\begin{aligned}\lim_{x_n \downarrow x} F_\infty(x_n) &= \inf\{H(q) : q \in \mathbb{Q}, q > x_n \text{ for some } n\} \\ &= \inf\{H(q) : q \in \mathbb{Q}, q > x\} = F_\infty(x).\end{aligned}$$

Suppose that x is a continuity point of the non-decreasing function F_∞. Then, for any $\varepsilon > 0$ there exists $y < x$ such that $F_\infty(x) - \varepsilon < F_\infty(y)$ and rational numbers $y < r_1 < x < r_2$ such that $H(r_2) < F_\infty(x) + \varepsilon$. It follows that

$$F_\infty(x) - \varepsilon < F_\infty(y) \leq H(r_1) \leq H(r_2) < F_\infty(x) + \varepsilon\,. \tag{3.2.7}$$

Recall that $F_{n_k}(x) \in [F_{n_k}(r_1), F_{n_k}(r_2)]$ and $F_{n_k}(r_i) \to H(r_i)$ as $k \to \infty$, for $i = 1, 2$. Thus, by (3.2.7) for all k large enough

$$F_\infty(x) - \varepsilon < F_{n_k}(r_1) \leq F_{n_k}(x) \leq F_{n_k}(r_2) < F_\infty(x) + \varepsilon,$$

which since $\varepsilon > 0$ is arbitrary implies $F_{n_k}(x) \to F_\infty(x)$ as $k \to \infty$. □

Exercise 3.2.39. *Suppose that the sequence of distribution functions $\{F_{X_k}\}$ is uniformly tight and $\mathbf{E}X_k^2 < \infty$ are such that $\mathbf{E}X_n^2 \to \infty$ as $n \to \infty$. Show that then also $\mathsf{Var}(X_n) \to \infty$ as $n \to \infty$.*

Hint: *If $|\mathbf{E}X_{n_l}|^2 \to \infty$ then $\sup_l \mathsf{Var}(X_{n_l}) < \infty$ yields $X_{n_l}/\mathbf{E}X_{n_l} \xrightarrow{L^2} 1$, whereas the uniform tightness of $\{F_{X_{n_l}}\}$ implies that $X_{n_l}/\mathbf{E}X_{n_l} \xrightarrow{p} 0$.*

Using Lemma 3.2.38 and Helly's theorem, you next explore the possibility of establishing weak convergence for non-negative random variables out of the convergence of the corresponding *Laplace transforms*.

Exercise 3.2.40.

(a) *Based on Exercise 3.2.29 show that if $Z \geq 0$ and $W \geq 0$ are such that $\mathbf{E}(e^{-sZ}) = \mathbf{E}(e^{-sW})$ for each $s > 0$, then $Z \overset{\mathcal{D}}{=} W$.*

(b) *Further, show that for any $Z \geq 0$, the function $L_Z(s) = \mathbf{E}(e^{-sZ})$ is infinitely differentiable at all $s > 0$ and for any positive integer k,*

$$\mathbf{E}[Z^k] = (-1)^k \lim_{s \downarrow 0} \frac{d^k}{ds^k} L_Z(s)\,,$$

even when (both sides are) infinite.

(c) *Suppose that $X_n \geq 0$ are such that $L(s) = \lim_n \mathbf{E}(e^{-sX_n})$ exists for all $s > 0$ and $L(s) \to 1$ for $s \downarrow 0$. Show that then the sequence of distribution functions $\{F_{X_n}\}$ is uniformly tight and that there exists a random variable $X_\infty \geq 0$ such that $X_n \xrightarrow{\mathcal{D}} X_\infty$ and $L(s) = \mathbf{E}(e^{-sX_\infty})$ for all $s > 0$.*

Hint: *To show that* $X_n \xrightarrow{\mathcal{D}} X_\infty$ *try reading and adapting the proof of Theorem 3.3.17.*

(d) *Let* $X_n = n^{-1}\sum_{k=1}^n kI_k$ *for* $I_k \in \{0,1\}$ *independent random variables, with* $\mathbf{P}(I_k = 1) = k^{-1}$. *Show that there exists* $X_\infty \geq 0$ *such that* $X_n \xrightarrow{\mathcal{D}} X_\infty$ *and* $\mathbf{E}(e^{-sX_\infty}) = \exp(\int_0^1 t^{-1}(e^{-st}-1)dt)$ *for all* $s > 0$.

Remark. The idea of using transforms to establish weak convergence shall be further developed in Section 3.3, with the *Fourier transform* instead of the Laplace transform.

3.3. Characteristic functions

This section is about the fundamental concept of characteristic function, its relevance for the theory of weak convergence, and in particular for the CLT.

In Subsection 3.3.1 we define the characteristic function, providing illustrating examples and certain general properties such as the relation between finite moments of a random variable and the degree of smoothness of its characteristic function. In Subsection 3.3.2 we recover the distribution of a random variable from its characteristic function, and building upon it, relate tightness and weak convergence with the point-wise convergence of the associated characteristic functions. We conclude with Subsection 3.3.3 in which we re-prove the CLT of Section 3.1 as an application of the theory of characteristic functions we have thus developed. The same approach will serve us well in other settings which we consider in the sequel (c.f. Sections 3.4 and 3.5).

3.3.1. Definition, examples, moments and derivatives.

We start off with the definition of the characteristic function of a random variable. To this end, recall that a $\mathbb{C}$-valued random variable is a function $Z : \Omega \mapsto \mathbb{C}$ such that the real and imaginary parts of Z are measurable, and for $Z = X + iY$ with $X, Y \in \mathbb{R}$ integrable random variables (and $i = \sqrt{-1}$), let $\mathbf{E}(Z) = \mathbf{E}(X) + i\mathbf{E}(Y) \in \mathbb{C}$.

Definition 3.3.1. *The* characteristic function Φ_X *of a random variable* X *is the map* $\mathbb{R} \mapsto \mathbb{C}$ *given by*

$$\Phi_X(\theta) = \mathbf{E}[e^{i\theta X}] = \mathbf{E}[\cos(\theta X)] + i\mathbf{E}[\sin(\theta X)]$$

where $\theta \in \mathbb{R}$ *and obviously both* $\cos(\theta X)$ *and* $\sin(\theta X)$ *are integrable R.V.-s.*

We also denote by $\Phi_\mu(\theta)$ *the characteristic function associated with a probability measure* μ *on* $(\mathbb{R}, \mathcal{B})$. *That is,* $\Phi_\mu(\theta) = \mu(e^{i\theta x})$ *is the characteristic function of a R.V.* X *whose law* $\mathcal{P}_X$ *is* μ.

Here are some of the properties of characteristic functions, where the complex conjugate $x - iy$ of $z = x + iy \in \mathbb{C}$ is denoted throughout by $\overline{z}$ and the *modulus* of $z = x + iy$ is $|z| = \sqrt{x^2 + y^2}$.

Proposition 3.3.2. *Let* X *be a R.V. and* Φ_X *its characteristic function, then*

(a) $\Phi_X(0) = 1$
(b) $\Phi_X(-\theta) = \overline{\Phi_X(\theta)}$
(c) $|\Phi_X(\theta)| \leq 1$
(d) $\theta \mapsto \Phi_X(\theta)$ *is a uniformly continuous function on* $\mathbb{R}$
(e) $\Phi_{aX+b}(\theta) = e^{ib\theta}\Phi_X(a\theta)$

Proof. For (a), $\Phi_X(0) = \mathbf{E}[e^{i0X}] = \mathbf{E}[1] = 1$. For (b), note that

$$\begin{aligned}\Phi_X(-\theta) &= \mathbf{E}\cos(-\theta X) + i\mathbf{E}\sin(-\theta X)\\ &= \mathbf{E}\cos(\theta X) - i\mathbf{E}\sin(\theta X) = \overline{\Phi_X(\theta)}\,.\end{aligned}$$

For (c), note that the function $|z| = \sqrt{x^2+y^2} : \mathbb{R}^2 \mapsto \mathbb{R}$ is convex, hence by Jensen's inequality (c.f. Exercise 1.3.20),

$$|\Phi_X(\theta)| = |\mathbf{E}e^{i\theta X}| \le \mathbf{E}|e^{i\theta X}| = 1$$

(since the modulus $|e^{i\theta x}| = 1$ for any real x and θ).

For (d), since $\Phi_X(\theta+h)-\Phi_X(\theta) = \mathbf{E}e^{i\theta X}(e^{ihX}-1)$, it follows by Jensen's inequality for the modulus function that

$$|\Phi_X(\theta+h) - \Phi_X(\theta)| \le \mathbf{E}[|e^{i\theta X}||e^{ihX}-1|] = \mathbf{E}|e^{ihX}-1| = \delta(h)$$

(using the fact that $|zv| = |z||v|$). Since $2 \ge |e^{ihX}-1| \to 0$ as $h \to 0$, by bounded convergence $\delta(h) \to 0$. As the bound $\delta(h)$ on the modulus of continuity of $\Phi_X(\theta)$ is independent of θ, we have uniform continuity of $\Phi_X(\cdot)$ on $\mathbb{R}$.

For (e) simply note that $\Phi_{aX+b}(\theta) = \mathbf{E}e^{i\theta(aX+b)} = e^{i\theta b}\mathbf{E}e^{i(a\theta)X} = e^{i\theta b}\Phi_X(a\theta)$. □

We also have the following relation between finite moments of the random variable and the derivatives of its characteristic function.

Lemma 3.3.3. *If $\mathbf{E}|X|^n < \infty$, then the characteristic function $\Phi_X(\theta)$ of X has continuous derivatives up to the n-th order, given by*

$$\frac{d^k}{d\theta^k}\Phi_X(\theta) = \mathbf{E}[(iX)^k e^{i\theta X}]\,, \qquad \textit{for} \quad k = 1,\dots,n \tag{3.3.1}$$

Proof. Note that for any $x, h \in \mathbb{R}$

$$e^{ihx} - 1 = ix\int_0^h e^{iux}du\,.$$

Consequently, for any $h \neq 0$, $\theta \in \mathbb{R}$ and positive integer k we have the identity

$$\begin{aligned}\Delta_{k,h}(x) &= h^{-1}\big((ix)^{k-1}e^{i(\theta+h)x} - (ix)^{k-1}e^{i\theta x}\big) - (ix)^k e^{i\theta x}\\ &= (ix)^k e^{i\theta x} h^{-1}\int_0^h (e^{iux}-1)du\,,\end{aligned} \tag{3.3.2}$$

from which we deduce that $|\Delta_{k,h}(x)| \le 2|x|^k$ for all θ and $h \neq 0$, and further that $|\Delta_{k,h}(x)| \to 0$ as $h \to 0$. Thus, for $k = 1,\dots,n$ we have by dominated convergence (and Jensen's inequality for the modulus function) that

$$|\mathbf{E}\Delta_{k,h}(X)| \le \mathbf{E}|\Delta_{k,h}(X)| \to 0 \quad \text{ for } \quad h \to 0.$$

Taking $k = 1$, we have from (3.3.2) that

$$\mathbf{E}\Delta_{1,h}(X) = h^{-1}(\Phi_X(\theta+h) - \Phi_X(\theta)) - \mathbf{E}[iXe^{i\theta X}]\,,$$

so its convergence to zero as $h \to 0$ amounts to the identity (3.3.1) holding for $k = 1$. In view of this, considering now (3.3.2) for $k = 2$, we have that

$$\mathbf{E}\Delta_{2,h}(X) = h^{-1}(\Phi'_X(\theta+h) - \Phi'_X(\theta)) - \mathbf{E}[(iX)^2 e^{i\theta X}]\,,$$

and its convergence to zero as $h \to 0$ amounts to (3.3.1) holding for $k = 2$. We continue in this manner for $k = 3,\dots,n$ to complete the proof of (3.3.1). The continuity of the derivatives follows by dominated convergence from the convergence to zero of $|(ix)^k e^{i(\theta+h)x} - (ix)^k e^{i\theta x}| \le 2|x|^k$ as $h \to 0$ (with $k = 1,\dots,n$). □

The converse of Lemma 3.3.3 does not hold. That is, there exist random variables with $\mathbf{E}|X| = \infty$ for which $\Phi_X(\theta)$ is differentiable at $\theta = 0$ (c.f. Exercise 3.3.23).

However, as we see next, the existence of a finite second derivative of $\Phi_X(\theta)$ at $\theta = 0$ implies that $\mathbf{E}X^2 < \infty$.

Lemma 3.3.4. *If* $\liminf_{\theta\to 0} \theta^{-2}(2\Phi_X(0) - \Phi_X(\theta) - \Phi_X(-\theta)) < \infty$*, then* $\mathbf{E}X^2 < \infty$.

Proof. Note that $\theta^{-2}(2\Phi_X(0) - \Phi_X(\theta) - \Phi_X(-\theta)) = \mathbf{E}g_\theta(X)$, where

$$g_\theta(x) = \theta^{-2}(2 - e^{i\theta x} - e^{-i\theta x}) = 2\theta^{-2}[1 - \cos(\theta x)] \to x^2 \quad \text{for} \quad \theta \to 0\,.$$

Since $g_\theta(x) \geq 0$ for all θ and x, it follows by Fatou's lemma that

$$\liminf_{\theta\to 0} \mathbf{E}g_\theta(X) \geq \mathbf{E}[\liminf_{\theta\to 0} g_\theta(X)] = \mathbf{E}X^2\,,$$

thus completing the proof of the lemma. $\square$

We continue with a few explicit computations of the characteristic function.

Example 3.3.5. *Consider a* Bernoulli *random variable B of parameter p, that is, $\mathbf{P}(B = 1) = p$ and $\mathbf{P}(B = 0) = 1 - p$. Its characteristic function is by definition*

$$\Phi_B(\theta) = \mathbf{E}[e^{i\theta B}] = pe^{i\theta} + (1-p)e^{i0\theta} = pe^{i\theta} + 1 - p\,.$$

The same type of explicit formula applies to any discrete valued R.V. For example, if N has the Poisson *distribution of parameter λ then*

$$\Phi_N(\theta) = \mathbf{E}[e^{i\theta N}] = \sum_{k=0}^{\infty} \frac{(\lambda e^{i\theta})^k}{k!} e^{-\lambda} = \exp(\lambda(e^{i\theta} - 1))\,. \tag{3.3.3}$$

The characteristic function has an explicit form also when the R.V. X has a probability density function f_X as in Definition 1.2.40. Indeed, then by Corollary 1.3.62 we have that

$$\Phi_X(\theta) = \int_{\mathbb{R}} e^{i\theta x} f_X(x) dx\,, \tag{3.3.4}$$

which is merely the Fourier transform of the density f_X (and is well defined since $\cos(\theta x) f_X(x)$ and $\sin(\theta x) f_X(x)$ are both integrable with respect to Lebesgue's measure).

Example 3.3.6. *If G has the $\mathcal{N}(\mu, v)$ distribution, namely, the probability density function $f_G(y)$ is given by (3.1.1), then its characteristic function is*

$$\Phi_G(\theta) = e^{i\mu\theta - v\theta^2/2}\,.$$

Indeed, recall Example 1.3.68 that $G = \sigma X + \mu$ for $\sigma = \sqrt{v}$ and X of a standard normal distribution $\mathcal{N}(0, 1)$. Hence, considering part (e) of Proposition 3.3.2 for $a = \sqrt{v}$ and $b = \mu$, it suffices to show that $\Phi_X(\theta) = e^{-\theta^2/2}$. To this end, as X is integrable, we have from Lemma 3.3.3 that

$$\Phi'_X(\theta) = \mathbf{E}(iXe^{i\theta X}) = \int_{\mathbb{R}} -x \sin(\theta x) f_X(x) dx$$

(since $x \cos(\theta x) f_X(x)$ is an integrable odd function, whose integral is thus zero). The standard normal density is such that $f'_X(x) = -x f_X(x)$, hence integrating by parts we find that

$$\Phi'_X(\theta) = \int_{\mathbb{R}} \sin(\theta x) f'_X(x) dx = -\int_{\mathbb{R}} \theta \cos(\theta x) f_X(x) dx = -\theta \Phi_X(\theta)$$

(since $\sin(\theta x) f_X(x)$ is an integrable odd function). We know that $\Phi_X(0) = 1$ and since $\varphi(\theta) = e^{-\theta^2/2}$ is the unique solution of the ordinary differential equation $\varphi'(\theta) = -\theta\varphi(\theta)$ with $\varphi(0) = 1$, it follows that $\Phi_X(\theta) = \varphi(\theta)$.

Example 3.3.7. *In another example, applying the formula (3.3.4) we see that the random variable $U = U(a, b)$ whose probability density function is $f_U(x) = (b-a)^{-1}\mathbf{1}_{a<x<b}$, has the characteristic function*

$$\Phi_U(\theta) = \frac{e^{i\theta b} - e^{i\theta a}}{i\theta(b-a)}$$

(recall that $\int_a^b e^{zx} dx = (e^{zb} - e^{za})/z$ for any $z \in \mathbb{C}$). For $a = -b$ the characteristic function simplifies to $\sin(b\theta)/(b\theta)$. Or, in case $b = 1$ and $a = 0$ we have $\Phi_U(\theta) = (e^{i\theta} - 1)/(i\theta)$ for the random variable U of Example 1.1.26.

For $a = 0$ and $z = -\lambda + i\theta$, $\lambda > 0$, the same integration identity applies also when $b \to \infty$ (since the real part of z is negative). Consequently, by (3.3.4), the exponential distribution *of parameter $\lambda > 0$ whose density is $f_T(t) = \lambda e^{-\lambda t}\mathbf{1}_{t>0}$ (see Example 1.3.68), has the characteristic function $\Phi_T(\theta) = \lambda/(\lambda - i\theta)$.*

Finally, for the density $f_S(s) = 0.5e^{-|s|}$ it is not hard to check that $\Phi_S(\theta) = 0.5/(1 - i\theta) + 0.5/(1 + i\theta) = 1/(1 + \theta^2)$ (just break the integration over $s \in \mathbb{R}$ in (3.3.4) according to the sign of s).

We next express the characteristic function of the sum of independent random variables in terms of the characteristic functions of the summands. This relation makes the characteristic function a useful tool for proving weak convergence statements involving sums of independent variables.

Lemma 3.3.8. *If X and Y are two independent random variables, then*

$$\Phi_{X+Y}(\theta) = \Phi_X(\theta)\Phi_Y(\theta)$$

PROOF. By the definition of the characteristic function

$$\Phi_{X+Y}(\theta) = \mathbf{E}e^{i\theta(X+Y)} = \mathbf{E}[e^{i\theta X}e^{i\theta Y}] = \mathbf{E}[e^{i\theta X}]\mathbf{E}[e^{i\theta Y}]\,,$$

where the right-most equality is obtained by the independence of X and Y (i.e. applying (1.4.12) for the integrable $f(x) = g(x) = e^{i\theta x}$). Observing that the right-most expression is $\Phi_X(\theta)\Phi_Y(\theta)$ completes the proof. $\square$

Here are three simple applications of this lemma.

Example 3.3.9. *If X and Y are independent and uniform on $(-1/2, 1/2)$ then by Corollary 1.4.33 the random variable $\Delta = X + Y$ has the* triangular density, *$f_\Delta(x) = (1 - |x|)\mathbf{1}_{|x|\le 1}$. Thus, by Example 3.3.7, Lemma 3.3.8, and the trigonometric identity $\cos\theta = 1 - 2\sin^2(\theta/2)$ we have that its characteristic function is*

$$\Phi_\Delta(\theta) = [\Phi_X(\theta)]^2 = \Big(\frac{2\sin(\theta/2)}{\theta}\Big)^2 = \frac{2(1-\cos\theta)}{\theta^2}\,.$$

Exercise 3.3.10. *Let X, $\widetilde{X}$ be i.i.d. random variables.*

(a) *Show that the characteristic function of $Z = X - \widetilde{X}$ is a non-negative, real-valued function.*
(b) *Prove that there do not exist $a < b$ and i.i.d. random variables X, $\widetilde{X}$ such that $X - \widetilde{X}$ is the uniform random variable on (a, b).*

In the next exercise you construct a random variable X whose law has no atoms while its characteristic function does not converge to zero for $\theta \to \infty$.

Exercise 3.3.11. *Let $X = 2\sum_{k=1}^{\infty} 3^{-k} B_k$ for $\{B_k\}$ i.i.d. Bernoulli random variables such that $\mathbf{P}(B_k = 1) = \mathbf{P}(B_k = 0) = 1/2$.*

(a) *Show that $\Phi_X(3^k\pi) = \Phi_X(\pi) \neq 0$ for $k = 1, 2, \ldots$.*
(b) *Recall that X has the uniform distribution on the Cantor set C, as specified in Example 1.2.43. Verify that $x \mapsto F_X(x)$ is everywhere continuous, hence the law $\mathcal{P}_X$ has no atoms (i.e. points of positive probability).*

3.3.2. Inversion, continuity and convergence. Is it possible to recover the distribution function from the characteristic function? Then answer is essentially yes.

Theorem 3.3.12 (Lévy's inversion theorem)**.** *Suppose Φ_X is the characteristic function of random variable X whose distribution function is F_X. For any real numbers $a < b$ and θ, let*

$$\psi_{a,b}(\theta) = \frac{1}{2\pi}\int_a^b e^{-i\theta u} du = \frac{e^{-i\theta a} - e^{-i\theta b}}{i2\pi\theta}\,. \tag{3.3.5}$$

Then,

$$\lim_{T\uparrow\infty}\int_{-T}^{T}\psi_{a,b}(\theta)\Phi_X(\theta)d\theta = \frac{1}{2}[F_X(b) + F_X(b^-)] - \frac{1}{2}[F_X(a) + F_X(a^-)]\,. \tag{3.3.6}$$

Furthermore, if $\int_{\mathbb{R}} |\Phi_X(\theta)| d\theta < \infty$, then X has the bounded continuous probability density function

$$f_X(x) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-i\theta x}\Phi_X(\theta)d\theta\,. \tag{3.3.7}$$

Remark. The identity (3.3.7) is a special case of the Fourier transform inversion formula, and as such is in 'duality' with $\Phi_X(\theta) = \int_{\mathbb{R}} e^{i\theta x} f_X(x) dx$ of (3.3.4). The formula (3.3.6) should be considered its integrated version, which thereby holds even in the absence of a density for X.

Here is a simple application of the 'duality' between (3.3.7) and (3.3.4).

Example 3.3.13. *The* Cauchy density *is $f_X(x) = 1/[\pi(1 + x^2)]$. Recall Example 3.3.7 that the density $f_S(s) = 0.5e^{-|s|}$ has the positive, integrable characteristic function $1/(1 + \theta^2)$. Thus, by (3.3.7),*

$$0.5e^{-|s|} = \frac{1}{2\pi}\int_{\mathbb{R}} \frac{1}{1+t^2} e^{-its} dt\,.$$

Multiplying both sides by two, then changing t to x and s to $-\theta$, we get (3.3.4) for the Cauchy density, resulting with its characteristic function $\Phi_X(\theta) = e^{-|\theta|}$.

When using characteristic functions for proving limit theorems we do not need the explicit formulas of Lévy's inversion theorem, but rather only the fact that the characteristic function determines the law, that is:

Corollary 3.3.14. *If the characteristic functions of two random variables X and Y are the same, that is $\Phi_X(\theta) = \Phi_Y(\theta)$ for all θ, then $X \stackrel{\mathcal{D}}{=} Y$.*

Remark. While the real-valued *moment generating function* $M_X(s) = \mathbf{E}[e^{sX}]$ is perhaps a simpler object than the characteristic function, it has a somewhat limited scope of applicability. For example, the law of a random variable X is uniquely determined by $M_X(\cdot)$ provided $M_X(s)$ is finite for all $s \in [-\delta, \delta]$, some $\delta > 0$ (c.f. [**Bil95**, Theorem 30.1]). More generally, assuming all moments of X are finite, the *Hamburger moment problem* is about uniquely determining the law of X from a given sequence of moments $\mathbf{E}X^k$. You saw in Exercise 3.2.29 that this is always possible when X has bounded support, but unfortunately, this is not always the case when X has unbounded support. For more on this issue, see [**Dur10**, Subsection 3.3.5].

Proof of Corollary 3.3.14. Since $\Phi_X = \Phi_Y$, comparing the right side of (3.3.6) for X and Y shows that

$$[F_X(b) + F_X(b^-)] - [F_X(a) + F_X(a^-)] = [F_Y(b) + F_Y(b^-)] - [F_Y(a) + F_Y(a^-)]\,.$$

As F_X is a distribution function, both $F_X(a) \to 0$ and $F_X(a^-) \to 0$ when $a \downarrow -\infty$. For this reason also $F_Y(a) \to 0$ and $F_Y(a^-) \to 0$. Consequently,

$$F_X(b) + F_X(b^-) = F_Y(b) + F_Y(b^-) \quad \text{for all} \quad b \in \mathbb{R}\,.$$

In particular, this implies that $F_X = F_Y$ on the collection $\mathcal{C}$ of continuity points of both F_X and F_Y. Recall that F_X and F_Y have each at most a countable set of points of discontinuity (see Exercise 1.2.39), so the complement of $\mathcal{C}$ is countable, and consequently $\mathcal{C}$ is a dense subset of $\mathbb{R}$. Thus, as distribution functions are non-decreasing and right-continuous we know that $F_X(b) = \inf\{F_X(x) : x > b, x \in \mathcal{C}\}$ and $F_Y(b) = \inf\{F_Y(x) : x > b, x \in \mathcal{C}\}$. Since $F_X(x) = F_Y(x)$ for all $x \in \mathcal{C}$, this identity extends to all $b \in \mathbb{R}$, resulting with $X \stackrel{\mathcal{D}}{=} Y$. ☐

Remark. In Lemma 3.1.1, it was shown directly that the sum of independent random variables of normal distributions $\mathcal{N}(\mu_k, v_k)$ has the normal distribution $\mathcal{N}(\mu, v)$ where $\mu = \sum_k \mu_k$ and $v = \sum_k v_k$. The proof easily reduces to dealing with two independent random variables, X of distribution $\mathcal{N}(\mu_1, v_1)$ and Y of distribution $\mathcal{N}(\mu_2, v_2)$ and showing that $X+Y$ has the normal distribution $\mathcal{N}(\mu_1 + \mu_2, v_1 + v_2)$. Here is an easy proof of this result via characteristic functions. First by the independence of X and Y (see Lemma 3.3.8), and their normality (see Example 3.3.6),

$$\begin{aligned}\Phi_{X+Y}(\theta) &= \Phi_X(\theta)\Phi_Y(\theta) = \exp(i\mu_1\theta - v_1\theta^2/2)\exp(i\mu_2\theta - v_2\theta^2/2) \\ &= \exp(i(\mu_1 + \mu_2)\theta - \frac{1}{2}(v_1 + v_2)\theta^2)\end{aligned}$$

We recognize this expression as the characteristic function corresponding to the $\mathcal{N}(\mu_1 + \mu_2, v_1 + v_2)$ distribution, which by Corollary 3.3.14 must indeed be the distribution of $X + Y$.

Proof of Lévy's inversion theorem. Consider the product μ of the law $\mathcal{P}_X$ of X which is a probability measure on $\mathbb{R}$ and Lebesgue's measure of $\theta \in [-T, T]$, noting that μ is a finite measure on $\mathbb{R} \times [-T, T]$ of total mass $2T$.

Fixing $a < b \in \mathbb{R}$ let $h_{a,b}(x, \theta) = \psi_{a,b}(\theta)e^{i\theta x}$, where by (3.3.5) and Jensen's inequality for the modulus function (and the uniform measure on $[a, b]$),

$$|h_{a,b}(x, \theta)| = |\psi_{a,b}(\theta)| \le \frac{1}{2\pi}\int_a^b |e^{-i\theta u}|du = \frac{b-a}{2\pi}\,.$$

Consequently, $\int |h_{a,b}| d\mu < \infty$, and applying Fubini's theorem, we conclude that

$$J_T(a,b) := \int_{-T}^{T} \psi_{a,b}(\theta)\Phi_X(\theta) d\theta = \int_{-T}^{T} \psi_{a,b}(\theta) \Big[\int_{\mathbb{R}} e^{i\theta x} d\mathcal{P}_X(x) \Big] d\theta$$
$$= \int_{-T}^{T} \Big[\int_{\mathbb{R}} h_{a,b}(x,\theta) d\mathcal{P}_X(x) \Big] d\theta = \int_{\mathbb{R}} \Big[\int_{-T}^{T} h_{a,b}(x,\theta) d\theta \Big] d\mathcal{P}_X(x) \,.$$

Since $h_{a,b}(x,\theta)$ is the difference between the function $e^{i\theta u}/(i2\pi\theta)$ at $u = x-a$ and the same function at $u = x-b$, it follows that

$$\int_{-T}^{T} h_{a,b}(x,\theta) d\theta = R(x-a,T) - R(x-b,T) \,.$$

Further, as the cosine function is even and the sine function is odd,

$$R(u,T) = \int_{-T}^{T} \frac{e^{i\theta u}}{i2\pi\theta} d\theta = \int_0^T \frac{\sin(\theta u)}{\pi\theta} d\theta = \frac{\mathrm{sgn}(u)}{\pi} S(|u|T) \,,$$

with $S(r) = \int_0^r x^{-1} \sin x \, dx$ for $r > 0$.

Even though the Lebesgue integral $\int_0^\infty x^{-1} \sin x \, dx$ does not exist, because both the integral of the positive part and the integral of the negative part are infinite, we still have that $S(r)$ is uniformly bounded on $(0,\infty)$ and

$$\lim_{r\uparrow\infty} S(r) = \frac{\pi}{2}$$

(c.f. Exercise 3.3.15). Consequently,

$$\lim_{T\uparrow\infty} [R(x-a,T) - R(x-b,T)] = g_{a,b}(x) = \begin{cases} 0 & \text{if } \ x < a \ \text{ or } \ x > b \\ \frac{1}{2} & \text{if } \ x = a \ \text{ or } \ x = b \\ 1 & \text{if } \ a < x < b \end{cases} .$$

Since $S(\cdot)$ is uniformly bounded, so is $|R(x-a,T) - R(x-b,T)|$ and by bounded convergence,

$$\lim_{T\uparrow\infty} J_T(a,b) = \lim_{T\uparrow\infty} \int_{\mathbb{R}} [R(x-a,T) - R(x-b,T)] d\mathcal{P}_X(x) = \int_{\mathbb{R}} g_{a,b}(x) d\mathcal{P}_X(x)$$
$$= \frac{1}{2}\mathcal{P}_X(\{a\}) + \mathcal{P}_X((a,b)) + \frac{1}{2}\mathcal{P}_X(\{b\}) \,.$$

With $\mathcal{P}_X(\{a\}) = F_X(a) - F_X(a^-)$, $\mathcal{P}_X((a,b)) = F_X(b^-) - F_X(a)$ and $\mathcal{P}_X(\{b\}) = F_X(b) - F_X(b^-)$, we arrive at the assertion (3.3.6).

Suppose now that $\int_{\mathbb{R}} |\Phi_X(\theta)| d\theta = C < \infty$. This implies that both the real and the imaginary parts of $e^{i\theta x}\Phi_X(\theta)$ are integrable with respect to Lebesgue's measure on $\mathbb{R}$, hence $f_X(x)$ of (3.3.7) is well defined. Further, $|f_X(x)| \le C$ is uniformly bounded and by dominated convergence with respect to Lebesgue's measure on $\mathbb{R}$,

$$\lim_{h\to 0} |f_X(x+h) - f_X(x)| \le \lim_{h\to 0} \frac{1}{2\pi} \int_{\mathbb{R}} |e^{-i\theta x}| |\Phi_X(\theta)| |e^{-i\theta h} - 1| d\theta = 0,$$

implying that $f_X(\cdot)$ is also continuous. Turning to prove that $f_X(\cdot)$ is the density of X, note that

$$|\psi_{a,b}(\theta)\Phi_X(\theta)| \le \frac{b-a}{2\pi} |\Phi_X(\theta)| \,,$$

so by dominated convergence we have that

$$\lim_{T\uparrow\infty} J_T(a,b) = J_\infty(a,b) = \int_{\mathbb{R}} \psi_{a,b}(\theta)\Phi_X(\theta)d\theta\,. \tag{3.3.8}$$

Further, in view of (3.3.5), upon applying Fubini's theorem for the integrable function $e^{-i\theta u}I_{[a,b]}(u)\Phi_X(\theta)$ with respect to Lebesgue's measure on $\mathbb{R}^2$, we see that

$$J_\infty(a,b) = \frac{1}{2\pi}\int_{\mathbb{R}} \Big[\int_a^b e^{-i\theta u}du\Big]\Phi_X(\theta)d\theta = \int_a^b f_X(u)du\,,$$

for the bounded continuous function $f_X(\cdot)$ of (3.3.7). In particular, $J_\infty(a,b)$ must be continuous in both a and b. Comparing (3.3.8) with (3.3.6) we see that

$$J_\infty(a,b) = \frac{1}{2}[F_X(b)+F_X(b^-)] - \frac{1}{2}[F_X(a)+F_X(a^-)]\,,$$

so the continuity of $J_\infty(\cdot,\cdot)$ implies that $F_X(\cdot)$ must also be continuous everywhere, with

$$F_X(b) - F_X(a) = J_\infty(a,b) = \int_a^b f_X(u)du\,,$$

for all $a < b$. This shows that necessarily $f_X(x)$ is a non-negative real-valued function, which is the density of X. □

Exercise 3.3.15. *Integrating $\int z^{-1}e^{iz}dz$ around the contour formed by the "upper" semi-circles of radii ε and r and the intervals $[-r,-\varepsilon]$ and $[r,\varepsilon]$, deduce that $S(r) = \int_0^r x^{-1}\sin x dx$ is uniformly bounded on $(0,\infty)$ with $S(r)\to\pi/2$ as $r\to\infty$.*

Our strategy for handling the CLT and similar limit results is to establish the convergence of characteristic functions and deduce from it the corresponding convergence in distribution. One ingredient for this is of course the fact that the characteristic function uniquely determines the corresponding law. Our next result provides an important second ingredient, that is, an explicit sufficient condition for uniform tightness in terms of the limit of the characteristic functions.

Lemma 3.3.16. *Suppose $\{\nu_n\}$ are probability measures on $(\mathbb{R},\mathcal{B})$ and $\Phi_{\nu_n}(\theta) = \nu_n(e^{i\theta x})$ the corresponding characteristic functions. If $\Phi_{\nu_n}(\theta)\to\Phi(\theta)$ as $n\to\infty$, for each $\theta\in\mathbb{R}$ and further $\Phi(\theta)$ is continuous at $\theta=0$, then the sequence $\{\nu_n\}$ is uniformly tight.*

Remark. To see why continuity of the limit $\Phi(\cdot)$ at 0 is required, consider the sequence ν_n of normal distributions $\mathcal{N}(0,n^2)$. From Example 3.3.6 we see that the point-wise limit $\Phi(\theta) = I_{\theta=0}$ of $\Phi_{\nu_n}(\theta) = \exp(-n^2\theta^2/2)$ exists but is discontinuous at $\theta=0$. However, for any $M<\infty$ we know that $\nu_n([-M,M]) = \nu_1([-M/n,M/n])\to 0$ as $n\to\infty$, so clearly the sequence $\{\nu_n\}$ is not uniformly tight. Indeed, the corresponding distribution functions $F_n(x) = F_1(x/n)$ converge vaguely to $F_\infty(x) = F_1(0) = 1/2$ which is not a distribution function (reflecting escape of all the probability mass to $\pm\infty$).

Proof. We start the proof by deriving the key inequality

$$\frac{1}{r}\int_{-r}^{r}(1-\Phi_\mu(\theta))d\theta \ge \mu([-2/r,2/r]^c)\,, \tag{3.3.9}$$

which holds for every probability measure μ on $(\mathbb{R}, \mathcal{B})$ and any $r > 0$, relating the smoothness of the characteristic function at 0 with the tail decay of the corresponding probability measure at $\pm\infty$. To this end, fixing $r > 0$, note that

$$J(x) := \int_{-r}^{r} (1 - e^{i\theta x}) d\theta = 2r - \int_{-r}^{r} (\cos\theta x + i \sin\theta x) d\theta = 2r - \frac{2 \sin rx}{x} .$$

So $J(x)$ is non-negative (since $|\sin u| \le |u|$ for all u), and bounded below by $2r - 2/|x|$ (since $|\sin u| \le 1$). Consequently,

$$J(x) \ge \max(2r - \frac{2}{|x|}, 0) \ge r I_{\{|x|>2/r\}} . \tag{3.3.10}$$

Now, applying Fubini's theorem for the function $1 - e^{i\theta x}$ whose modulus is bounded by 2 and the product of the probability measure μ and Lebesgue's measure on $[-r, r]$, which is a finite measure of total mass $2r$, we get the identity

$$\int_{-r}^{r} (1 - \Phi_\mu(\theta)) d\theta = \int_{-r}^{r} \Big[\int_{\mathbb{R}} (1 - e^{i\theta x}) d\mu(x) \Big] d\theta = \int_{\mathbb{R}} J(x) d\mu(x) .$$

Thus, the lower bound (3.3.10) and monotonicity of the integral imply that

$$\frac{1}{r} \int_{-r}^{r} (1 - \Phi_\mu(\theta)) d\theta = \frac{1}{r} \int_{\mathbb{R}} J(x) d\mu(x) \ge \int_{\mathbb{R}} I_{\{|x|>2/r\}} d\mu(x) = \mu([-2/r, 2/r]^c) ,$$

hence establishing (3.3.9).

We turn to the application of this inequality for proving the uniform tightness. Since $\Phi_{\nu_n}(0) = 1$ for all n and $\Phi_{\nu_n}(0) \to \Phi(0)$, it follows that $\Phi(0) = 1$. Further, $\Phi(\theta)$ is continuous at $\theta = 0$, so for any $\varepsilon > 0$, there exists $r = r(\varepsilon) > 0$ such that

$$\frac{\varepsilon}{4} \ge |1 - \Phi(\theta)| \quad \text{for all} \quad \theta \in [-r, r],$$

and hence also

$$\frac{\varepsilon}{2} \ge \frac{1}{r} \int_{-r}^{r} |1 - \Phi(\theta)| d\theta .$$

The point-wise convergence of Φ_{ν_n} to Φ implies that $|1 - \Phi_{\nu_n}(\theta)| \to |1 - \Phi(\theta)|$. By bounded convergence with respect to Uniform measure of θ on $[-r, r]$, it follows that for some finite $n_0 = n_0(\varepsilon)$ and all $n \ge n_0$,

$$\varepsilon \ge \frac{1}{r} \int_{-r}^{r} |1 - \Phi_{\nu_n}(\theta)| d\theta ,$$

which in view of (3.3.9) results with

$$\varepsilon \ge \frac{1}{r} \int_{-r}^{r} [1 - \Phi_{\nu_n}(\theta)] d\theta \ge \nu_n([-2/r, 2/r]^c) .$$

Since $\varepsilon > 0$ is arbitrary and $M = 2/r$ is independent of n, by Definition 3.2.32 this amounts to the uniform tightness of the sequence $\{\nu_n\}$. □

Building upon Corollary 3.3.14 and Lemma 3.3.16 we can finally relate the point-wise convergence of characteristic functions to the weak convergence of the corresponding measures.

Theorem 3.3.17 (Lévy's continuity theorem)**.** *Let ν_n, $1 \le n \le \infty$ be probability measures on $(\mathbb{R}, \mathcal{B})$.*

(a) *If $\nu_n \stackrel{w}{\Rightarrow} \nu_\infty$, then $\Phi_{\nu_n}(\theta) \to \Phi_{\nu_\infty}(\theta)$ for each $\theta \in \mathbb{R}$.*

(b) *Conversely, if $\Phi_{\nu_n}(\theta)$ converges point-wise to a limit $\Phi(\theta)$ that is continuous at $\theta = 0$, then $\{\nu_n\}$ is a uniformly tight sequence and $\nu_n \stackrel{w}{\Rightarrow} \nu$ such that $\Phi_\nu = \Phi$.*

Proof. For part (a), since both $x \mapsto \cos(\theta x)$ and $x \mapsto \sin(\theta x)$ are bounded continuous functions, the assumed weak convergence of ν_n to ν_∞ implies that $\Phi_{\nu_n}(\theta) = \nu_n(e^{i\theta x}) \to \nu_\infty(e^{i\theta x}) = \Phi_{\nu_\infty}(\theta)$ (c.f. Definition 3.2.17).
Turning to deal with part (b), recall that by Lemma 3.3.16 we know that the collection $\Gamma = \{\nu_n\}$ is uniformly tight. Hence, by Prohorov's theorem (see the remark preceding the proof of Lemma 3.2.38), for every subsequence $\nu_{n(m)}$ there is a further sub-subsequence $\nu_{n(m_k)}$ that converges weakly to some probability measure ν_∞. Though in general ν_∞ might depend on the specific choice of $n(m)$, we deduce from part (a) of the theorem that necessarily $\Phi_{\nu_\infty} = \Phi$. Since the characteristic function uniquely determines the law (see Corollary 3.3.14), here the same limit $\nu = \nu_\infty$ applies for *all choices* of $n(m)$. In particular, fixing $h \in C_b(\mathbb{R})$, the sequence $y_n = \nu_n(h)$ is such that every subsequence $y_{n(m)}$ has a further sub-subsequence $y_{n(m_k)}$ that converges to $y = \nu(h)$. Consequently, $y_n = \nu_n(h) \to y = \nu(h)$ (see Lemma 2.2.11), and since this applies for all $h \in C_b(\mathbb{R})$, we conclude that $\nu_n \stackrel{w}{\Rightarrow} \nu$ such that $\Phi_\nu = \Phi$. □

Here is a direct consequence of Lévy's continuity theorem.

Exercise 3.3.18. *Show that if $X_n \xrightarrow{\mathcal{D}} X_\infty$, $Y_n \xrightarrow{\mathcal{D}} Y_\infty$ and Y_n is independent of X_n for $1 \le n \le \infty$, then $X_n + Y_n \xrightarrow{\mathcal{D}} X_\infty + Y_\infty$.*

Combining Exercise 3.3.18 with the Portmanteau theorem and the CLT, you can now show that a finite second moment is necessary for the convergence in distribution of $n^{-1/2}\sum_{k=1}^n X_k$ for i.i.d. $\{X_k\}$.

Exercise 3.3.19. *Suppose $\{X_k, \widetilde{X}_k\}$ are i.i.d. and $n^{-1/2}\sum_{k=1}^n X_k \xrightarrow{\mathcal{D}} Z$ (with the limit $Z \in \mathbb{R}$).*

(a) *Set $Y_k = X_k - \widetilde{X}_k$ and show that $n^{-1/2}\sum_{k=1}^n Y_k \xrightarrow{\mathcal{D}} Z - \widetilde{Z}$, with Z and $\widetilde{Z}$ i.i.d.*

(b) *Let $U_k = Y_k I_{|Y_k| \le b}$ and $V_k = Y_k I_{|Y_k|>b}$. Show that for any $u < \infty$ and all n,*

$$\mathbf{P}(\sum_{k=1}^n Y_k \ge u\sqrt{n}) \ge \mathbf{P}(\sum_{k=1}^n U_k \ge u\sqrt{n}, \sum_{k=1}^n V_k \ge 0) \ge \frac{1}{2}\mathbf{P}(\sum_{k=1}^n U_k \ge u\sqrt{n}) .$$

(c) *Apply the Portmanteau theorem and the CLT for the bounded i.i.d. $\{U_k\}$ to get that for any $u, b < \infty$,*

$$\mathbf{P}(Z - \widetilde{Z} \ge u) \ge \frac{1}{2}\mathbf{P}(G \ge u/\sqrt{\mathbf{E}U_1^2}) .$$

Considering the limit $b \to \infty$ followed by $u \to \infty$ deduce that $\mathbf{E}Y_1^2 < \infty$.

(d) *Conclude that if $n^{-1/2}\sum_{k=1}^n X_k \xrightarrow{\mathcal{D}} Z$, then necessarily $\mathbf{E}X_1^2 < \infty$.*

Remark. The trick of replacing X_k by the variables $Y_k = X_k - \widetilde{X}_k$ whose law is symmetric (i.e. $Y_k \stackrel{\mathcal{D}}{=} -Y_k$), is very useful in many problems. It is often called the *symmetrization trick*.

Exercise 3.3.20. *Provide an example of a random variable X with a bounded probability density function but for which $\int_{\mathbb{R}} |\Phi_X(\theta)|d\theta = \infty$, and another example of a random variable X whose characteristic function $\Phi_X(\theta)$ is not differentiable at $\theta = 0$.*

As you find out next, Lévy's inversion theorem can help when computing densities.

Exercise 3.3.21. *Suppose the random variables U_k are i.i.d. where the law of each U_k is the* uniform probability measure *on $(-1,1)$. Considering Example 3.3.7, show that for each $n \ge 2$, the probability density function of $S_n = \sum_{k=1}^n U_k$ is*

$$f_{S_n}(s) = \frac{1}{\pi} \int_0^\infty \cos(\theta s)(\sin\theta/\theta)^n d\theta \,,$$

and deduce that $\int_0^\infty \cos(\theta s)(\sin\theta/\theta)^n d\theta = 0$ for all $s > n \ge 2$.

Exercise 3.3.22. *Deduce from Example 3.3.13 that if $\{X_k\}$ are i.i.d. each having the Cauchy density, then $n^{-1}\sum_{k=1}^n X_k$ has the same distribution as X_1, for any value of n.*

We next relate differentiability of $\Phi_X(\cdot)$ with the weak law of large numbers and show that it does not imply that $\mathbf{E}|X|$ is finite.

Exercise 3.3.23. *Let $S_n = \sum_{k=1}^n X_k$ where the i.i.d. random variables $\{X_k\}$ have each the characteristic function $\Phi_X(\cdot)$.*

(a) *Show that if $\frac{d\Phi_X}{d\theta}(0) = z \in \mathbb{C}$, then $z = ia$ for some $a \in \mathbb{R}$ and $n^{-1}S_n \xrightarrow{p} a$ as $n \to \infty$.*
(b) *Show that if $n^{-1}S_n \xrightarrow{p} a$, then $\Phi_X(\pm h_k)^{n_k} \to e^{\pm ia\theta}$ for any $h_k \downarrow 0$, $\theta > 0$ and $n_k = [\theta/h_k]$, and deduce that $\frac{d\Phi_X}{d\theta}(0) = ia$.*
(c) *Conclude that the weak law of large numbers holds (i.e. $n^{-1}S_n \xrightarrow{p} a$ for some non-random a), if and only if $\Phi_X(\cdot)$ is differentiable at $\theta = 0$ (this result is due to E.J.G. Pitman, see* [**Pit56**]*).*
(d) *Use Exercise 2.1.13 to provide a random variable X for which $\Phi_X(\cdot)$ is differentiable at $\theta = 0$ but $\mathbf{E}|X| = \infty$.*

As you show next, $X_n \xrightarrow{\mathcal{D}} X_\infty$ yields convergence of $\Phi_{X_n}(\cdot)$ to $\Phi_{X_\infty}(\cdot)$, uniformly over compact subsets of $\mathbb{R}$.

Exercise 3.3.24. *Show that if $X_n \xrightarrow{\mathcal{D}} X_\infty$ then for any r finite,*

$$\lim_{n\to\infty} \sup_{|\theta| \le r} |\Phi_{X_n}(\theta) - \Phi_{X_\infty}(\theta)| = 0 \,.$$

Hint: *By Theorem 3.2.7 you may further assume that $X_n \xrightarrow{a.s.} X_\infty$.*

Characteristic functions of modulus one correspond to lattice or degenerate laws, as you show in the following refinement of part (c) of Proposition 3.3.2.

Exercise 3.3.25. *Suppose $|\Phi_Y(\theta)| = 1$ for some $\theta \neq 0$.*

(a) *Show that Y is a* $(2\pi/\theta)$-lattice random variable*, namely, that $Y \bmod (2\pi/\theta)$ is $\mathbf{P}$-degenerate.*
Hint: *Check conditions for equality when applying Jensen's inequality for $(\cos\theta Y, \sin\theta Y)$ and the convex function $g(x,y) = \sqrt{x^2+y^2}$.*
(b) *Deduce that if in addition $|\Phi_Y(\lambda\theta)| = 1$ for some $\lambda \notin \mathcal{Q}$ then Y must be $\mathbf{P}$-degenerate, in which case $\Phi_Y(\theta) = \exp(i\theta c)$ for some $c \in \mathbb{R}$.*

Building on the preceding two exercises, you are to prove next the following *convergence of types* result.

Exercise 3.3.26. *Suppose* $Z_n \xrightarrow{\mathcal{D}} Y$ *and* $\beta_n Z_n + \gamma_n \xrightarrow{\mathcal{D}} \widehat{Y}$ *for some* $\widehat{Y}$*, non-**P**-degenerate* Y*, and non-random* $\beta_n \geq 0$*,* γ_n*.*

(a) *Show that* $\beta_n \to \beta \geq 0$ *finite.*
Hint: *Start with the finiteness of limit points of* $\{\beta_n\}$*.*

(b) *Deduce that* $\gamma_n \to \gamma$ *finite.*

(c) *Conclude that* $\widehat{Y} \stackrel{\mathcal{D}}{=} \beta Y + \gamma$*.*
Hint: *Recall Slutsky's lemma.*

Remark. This convergence of types fails for **P**-degenerate Y. For example, if $Z_n \stackrel{\mathcal{D}}{=} \mathcal{N}(0, n^{-3})$, then both $Z_n \xrightarrow{\mathcal{D}} 0$ and $nZ_n \xrightarrow{\mathcal{D}} 0$. Similarly, if $Z_n \stackrel{\mathcal{D}}{=} \mathcal{N}(0,1)$ then $\beta_n Z_n \stackrel{\mathcal{D}}{=} \mathcal{N}(0,1)$ for the non-converging sequence $\beta_n = (-1)^n$ (of alternating signs).

Mimicking the proof of Lévy's inversion theorem, for random variables of bounded support you get the following alternative inversion formula, based on the theory of Fourier series.

Exercise 3.3.27. *Suppose R.V. X supported on* $(0,t)$ *has the characteristic function* Φ_X *and the distribution function* F_X*. Let* $\theta_0 = 2\pi/t$ *and* $\psi_{a,b}(\cdot)$ *be as in (3.3.5), with* $\psi_{a,b}(0) = \frac{b-a}{2\pi}$*.*

(a) *Show that for any* $0 < a < b < t$

$$\lim_{T\uparrow\infty} \sum_{k=-T}^{T} \theta_0 (1 - \frac{|k|}{T}) \psi_{a,b}(k\theta_0) \Phi_X(k\theta_0) = \frac{1}{2}[F_X(b) + F_X(b^-)] - \frac{1}{2}[F_X(a) + F_X(a^-)] \,.$$

Hint: *Recall that* $S_T(r) = \sum_{k=1}^{T} (1 - k/T) \frac{\sin kr}{k}$ *is uniformly bounded for* $r \in (0, 2\pi)$ *and integer* $T \geq 1$*, and* $S_T(r) \to \frac{\pi - r}{2}$ *as* $T \to \infty$*.*

(b) *Show that if* $\sum_k |\Phi_X(k\theta_0)| < \infty$ *then X has the bounded continuous probability density function, given for* $x \in (0,t)$ *by*

$$f_X(x) = \frac{\theta_0}{2\pi} \sum_{k\in\mathbb{Z}} e^{-ik\theta_0 x} \Phi_X(k\theta_0) \,.$$

(c) *Deduce that if R.V.s X and Y supported on* $(0,t)$ *are such that* $\Phi_X(k\theta_0) = \Phi_Y(k\theta_0)$ *for all* $k \in \mathbb{Z}$*, then* $X \stackrel{\mathcal{D}}{=} Y$*.*

Here is an application of the preceding exercise for the *random walk* on the circle S^1 of radius one (c.f. Definition 5.1.6 for the random walk on $\mathbb{R}$).

Exercise 3.3.28. *Let* $t = 2\pi$ *and* Ω *denote the unit circle* S^1 *parametrized by the angular coordinate to yield the identification* $\Omega = [0,t]$ *where both end-points are considered the same point. We equip* Ω *with the topology induced by* $[0,t]$ *and the surface measure* λ_Ω *similarly induced by Lebesgue's measure (as in Exercise 1.4.37). In particular, R.V.-s on* $(\Omega, \mathcal{B}_\Omega)$ *correspond to Borel periodic functions on* $\mathbb{R}$*, of period t. In this context we call U of law* $t^{-1}\lambda_\Omega$ *a uniform R.V. and call* $S_n = (\sum_{k=1}^{n} \xi_k) \text{mod } t$*, with i.i.d* $\xi, \xi_k \in \Omega$*, a random walk.*

(a) *Verify that Exercise 3.3.27 applies for* $\theta_0 = 1$ *and R.V.-s on* Ω*.*

(b) *Show that if probability measures* ν_n *on* $(\Omega, \mathcal{B}_\Omega)$ *are such that* $\Phi_{\nu_n}(k) \to \varphi(k)$ *for* $n \to \infty$ *and fixed* $k \in \mathbb{Z}$, *then* $\nu_n \overset{w}{\Rightarrow} \nu_\infty$ *and* $\varphi(k) = \Phi_{\nu_\infty}(k)$ *for all* $k \in \mathbb{Z}$.
Hint: *Since* Ω *is compact the sequence* $\{\nu_n\}$ *is uniformly tight.*

(c) *Show that* $\Phi_U(k) = \mathbf{1}_{k=0}$ *and* $\Phi_{S_n}(k) = \Phi_\xi(k)^n$. *Deduce from these facts that if* ξ *has a density with respect to* λ_Ω *then* $S_n \xrightarrow{\mathcal{D}} U$ *as* $n \to \infty$.
Hint: *Recall part (a) of Exercise 3.3.25.*

(d) *Check that if* $\xi = \alpha$ *is non-random for some* $\alpha/t \notin \mathbb{Q}$, *then* S_n *does not converge in distribution, but* $S_{K_n} \xrightarrow{\mathcal{D}} U$ *for* K_n *which are uniformly chosen in* $\{1, 2, \dots, n\}$, *independently of the sequence* $\{\xi_k\}$.

3.3.3. Revisiting the CLT. Applying the theory of Subsection 3.3.2 we provide an alternative proof of the CLT, based on characteristic functions. One can prove many other weak convergence results for sums of random variables by properly adapting this approach, which is exactly what we will do when demonstrating the convergence to stable laws (see Exercise 3.3.33), and in proving the Poisson approximation theorem (in Subsection 3.4.1), and the multivariate CLT (in Section 3.5).

To this end, we start by deriving the analog of the bound (3.1.7) for the characteristic function.

Lemma 3.3.29. *If a random variable* X *has* $\mathbf{E}(X) = 0$ *and* $\mathbf{E}(X^2) = v < \infty$, *then for all* $\theta \in \mathbb{R}$,

$$\Big|\Phi_X(\theta) - \big(1 - \frac{1}{2}v\theta^2\big)\Big| \le \theta^2 \mathbf{E}\min(|X|^2, |\theta||X|^3/6).$$

Proof. Let $R_2(x) = e^{ix} - 1 - ix - (ix)^2/2$. Then, rearranging terms, recalling $\mathbf{E}(X) = 0$ and using Jensen's inequality for the modulus function, we see that

$$\Big|\Phi_X(\theta) - \Big(1 - \frac{1}{2}v\theta^2\Big)\Big| = \Big|\mathbf{E}\big[e^{i\theta X} - 1 - i\theta X - \frac{i^2}{2}\theta^2 X^2\big]\Big| = \Big|\mathbf{E}R_2(\theta X)\Big| \le \mathbf{E}|R_2(\theta X)|.$$

Since $|R_2(x)| \le \min(|x|^2, |x|^3/6)$ for any $x \in \mathbb{R}$ (see also Exercise 3.3.34), by monotonicity of the expectation we get that $\mathbf{E}|R_2(\theta X)| \le \mathbf{E}\min(|\theta X|^2, |\theta X|^3/6)$, completing the proof of the lemma. $\square$

The following simple complex analysis estimate is needed for relating the approximation of the characteristic function of summands to that of their sum.

Lemma 3.3.30. *Suppose* $z_{n,k} \in \mathbb{C}$ *are such that* $z_n = \sum_{k=1}^n z_{n,k} \to z_\infty$ *and* $\eta_n = \sum_{k=1}^n |z_{n,k}|^2 \to 0$ *when* $n \to \infty$. *Then,*

$$\varphi_n := \prod_{k=1}^n (1 + z_{n,k}) \to \exp(z_\infty) \quad \textit{for} \quad n \to \infty.$$

Proof. Recall that the power series expansion

$$\log(1+z) = \sum_{k=1}^\infty \frac{(-1)^{k-1} z^k}{k}$$

converges for $|z| < 1$. In particular, for $|z| \le 1/2$ it follows that

$$|\log(1+z) - z| \le \sum_{k=2}^{\infty} \frac{|z|^k}{k} \le |z|^2 \sum_{k=2}^{\infty} \frac{2^{-(k-2)}}{k} \le |z|^2 \sum_{k=2}^{\infty} 2^{-(k-1)} = |z|^2 \,.$$

Let $\delta_n = \max\{|z_{n,k}| : k = 1, \dots, n\}$. Note that $\delta_n^2 \le \eta_n$, so our assumption that $\eta_n \to 0$ implies that $\delta_n \le 1/2$ for all n sufficiently large, in which case

$$|\log \varphi_n - z_n| = |\log \prod_{k=1}^{n} (1 + z_{n,k}) - \sum_{k=1}^{n} z_{n,k}| \le \sum_{k=1}^{n} |\log(1 + z_{n,k}) - z_{n,k}| \le \eta_n \,.$$

With $z_n \to z_\infty$ and $\eta_n \to 0$, it follows that $\log \varphi_n \to z_\infty$. Consequently, $\varphi_n \to \exp(z_\infty)$ as claimed. $\square$

We will give now an alternative proof of the CLT of Theorem 3.1.2.

Proof of Theorem 3.1.2. From Example 3.3.6 we know that $\Phi_G(\theta) = e^{-\frac{\theta^2}{2}}$ is the characteristic function of the standard normal distribution. So, by Lévy's continuity theorem it suffices to show that $\Phi_{\widehat{S}_n}(\theta) \to \exp(-\theta^2/2)$ as $n \to \infty$, for each $\theta \in \mathbb{R}$. Recall that $\widehat{S}_n = \sum_{k=1}^n X_{n,k}$, with $X_{n,k} = (X_k - \mu)/\sqrt{vn}$ i.i.d. random variables, so by independence (see Lemma 3.3.8) and scaling (see part (e) of Proposition 3.3.2), we have that

$$\varphi_n := \Phi_{\widehat{S}_n}(\theta) = \prod_{k=1}^{n} \Phi_{X_{n,k}}(\theta) = \Phi_Y(n^{-1/2}\theta)^n = (1 + z_n/n)^n,$$

where $Y = (X_1 - \mu)/\sqrt{v}$ and $z_n = z_n(\theta) := n[\Phi_Y(n^{-1/2}\theta) - 1]$. Applying Lemma 3.3.30 for $z_{n,k} = z_n/n$ it remains only to show that $z_n \to -\theta^2/2$ (for then $\eta_n = |z_n|^2/n \to 0$). Indeed, since $\mathbf{E}(Y) = 0$ and $\mathbf{E}(Y^2) = 1$, we have from Lemma 3.3.29 that

$$|z_n + \theta^2/2| = \big|n[\Phi_Y(n^{-1/2}\theta) - 1] + \theta^2/2\big| \le \mathbf{E}V_n \,,$$

for $V_n = \min(|\theta Y|^2, n^{-1/2}|\theta Y|^3/6)$. With $V_n \xrightarrow{a.s.} 0$ as $n \to \infty$ and $V_n \le |\theta|^2|Y|^2$ which is integrable, it follows by dominated convergence that $\mathbf{E}V_n \to 0$ as $n \to \infty$, hence $z_n \to -\theta^2/2$ completing the proof of Theorem 3.1.2. $\square$

We proceed with a brief introduction of stable laws, their domain of attraction and the corresponding limit theorems (which are a natural generalization of the CLT).

Definition 3.3.31. *Random variable Y has a* stable law *if it is non-degenerate and for any $m \ge 1$ there exist constants $d_m > 0$ and c_m, such that $Y_1 + \ldots + Y_m \overset{\mathcal{D}}{=} d_m Y + c_m$, where $\{Y_i\}$ are i.i.d. copies of Y. Such variable has a* symmetric stable law *if in addition $Y \overset{\mathcal{D}}{=} -Y$. We further say that random variable X is in the* domain of attraction *of non-degenerate Y if there exist constants $b_n > 0$ and a_n such that $Z_n(X) = (S_n - a_n)/b_n \xrightarrow{\mathcal{D}} Y$ for $S_n = \sum_{k=1}^n X_k$ and i.i.d. copies X_k of X.*

By definition, the collection of stable laws is closed under the affine map $Y \mapsto \pm\sqrt{v}Y + \mu$ for $\mu \in \mathbb{R}$ and $v > 0$ (which correspond to the centering and scale of the law, but not necessarily its mean and variance). Clearly, each stable law is in its own domain of attraction and as we see next, only stable laws have a non-empty domain of attraction.

Proposition 3.3.32. *If X is in the domain of attraction of some non-degenerate variable Y, then Y must have a stable law.*

Proof. Fix $m \geq 1$, and setting $n = km$ let $\beta_n = b_n/b_k > 0$ and $\gamma_n = (a_n - ma_k)/b_k$. We then have the representation

$$\beta_n Z_n(X) + \gamma_n = \sum_{i=1}^{m} Z_k^{(i)} ,$$

where $Z_k^{(i)} = (X_{(i-1)k+1} + \ldots + X_{ik} - a_k)/b_k$ are i.i.d. copies of $Z_k(X)$. From our assumption that $Z_k(X) \xrightarrow{\mathcal{D}} Y$ we thus deduce (by at most $m-1$ applications of Exercise 3.3.18), that $\beta_n Z_n(X)+\gamma_n \xrightarrow{\mathcal{D}} \widehat{Y}$, where $\widehat{Y} = Y_1+\ldots+Y_m$ for i.i.d. copies $\{Y_i\}$ of Y. Moreover, by assumption $Z_n(X) \xrightarrow{\mathcal{D}} Y$, hence by the convergence of types $\widehat{Y} \overset{\mathcal{D}}{=} d_m Y + c_m$ for some finite non-random $d_m \geq 0$ and c_m (c.f. Exercise 3.3.26). Recall Lemma 3.3.8 that $\Phi_{\widehat{Y}}(\theta) = [\Phi_Y(\theta)]^m$. So, with Y assumed non-degenerate the same applies to $\widehat{Y}$ (see Exercise 3.3.25), and in particular $d_m > 0$. Since this holds for any $m \geq 1$, by definition Y has a stable law. $\square$

We have already seen two examples of symmetric stable laws, namely those associated with the zero-mean normal density and with the *Cauchy* density of Example 3.3.13. Indeed, as you show next, for each $\alpha \in (0,2)$ there corresponds the symmetric α-stable variable Y_α whose characteristic function is $\Phi_{Y_\alpha}(\theta) = \exp(-|\theta|^\alpha)$ (so the Cauchy distribution corresponds to the symmetric stable of *index* $\alpha = 1$ and the normal distribution corresponds to index $\alpha = 2$).

Exercise 3.3.33. *Fixing $\alpha \in (0,2)$, suppose $X \overset{\mathcal{D}}{=} -X$ and $\mathbf{P}(|X| > x) = x^{-\alpha}$ for all $x \geq 1$.*

(a) *Check that $\Phi_X(\theta) = 1 - \gamma(|\theta|)|\theta|^\alpha$ where $\gamma(r) = \alpha \int_r^\infty (1-\cos u) u^{-(\alpha+1)} du$ converges as $r \downarrow 0$ to $\gamma(0)$ finite and positive.*

(b) *Setting $\varphi_{\alpha,0}(\theta) = \exp(-|\theta|^\alpha)$, $b_n = (\gamma(0)n)^{1/\alpha}$ and $\widehat{S}_n = b_n^{-1} \sum_{k=1}^n X_k$ for i.i.d. copies X_k of X, deduce that $\Phi_{\widehat{S}_n}(\theta) \to \varphi_{\alpha,0}(\theta)$ as $n \to \infty$, for any fixed $\theta \in \mathbb{R}$.*

(c) *Conclude that X is in the domain of attraction of a symmetric stable variable Y_α, whose characteristic function is $\varphi_{\alpha,0}(\cdot)$.*

(d) *Fix $\alpha = 1$ and show that with probability one $\limsup_{n\to\infty} \widehat{S}_n = \infty$ and $\liminf_{n\to\infty} \widehat{S}_n = -\infty$.*
Hint: *Recall Kolmogorov's 0-1 law. The same proof applies for any $\alpha > 0$ once we verify that Y_α has unbounded support.*

(e) *Show that if $\alpha = 1$ then $\frac{1}{n \log n} \sum_{k=1}^n |X_k| \to 1$ in probability but not almost surely (in contrast, X is integrable when $\alpha > 1$, in which case the strong law of large numbers applies).*

Remark. While outside the scope of these notes, one can show that (up to scaling) *any* symmetric stable variable must be of the form Y_α for some $\alpha \in (0,2]$. Further, for any $\alpha \in (0,2)$ the necessary and sufficient condition for $X \overset{\mathcal{D}}{=} -X$ to be in the domain of attraction of Y_α is that the function $L(x) = x^\alpha \mathbf{P}(|X| > x)$ is *slowly varying* at ∞ (that is, $L(ux)/L(x) \to 1$ for $x \to \infty$ and fixed $u > 0$). Indeed, as shown for example in [**Bre92**, Theorem 9.32], up to the mapping $Y \mapsto \sqrt{v}Y + \mu$, the collection of all stable laws forms a two parameter family $Y_{\alpha,\kappa}$, parametrized by the index $\alpha \in (0,2]$ and *skewness* $\kappa \in [-1,1]$. The corresponding characteristic functions are

$$\varphi_{\alpha,\kappa}(\theta) = \exp(-|\theta|^\alpha(1 + i\kappa\text{sgn}(\theta)g_\alpha(\theta)))\,, \tag{3.3.11}$$

where $g_1(r) = (2/\pi)\log|r|$ and $g_\alpha = \tan(\pi\alpha/2)$ is constant for all $\alpha \neq 1$ (in particular, $g_2 = 0$ so the parameter κ is irrelevant when $\alpha = 2$). Further, in case $\alpha < 2$ the domain of attraction of $Y_{\alpha,\kappa}$ consists precisely of the random variables X for which $L(x) = x^\alpha \mathbf{P}(|X| > x)$ is slowly varying at ∞ and $(\mathbf{P}(X > x) - \mathbf{P}(X < -x))/\mathbf{P}(|X| > x) \to \kappa$ as $x \to \infty$ (for example, see [**Bre92**, Theorem 9.34]). To complete this picture, we recall [**Fel71**, Theorem XVII.5.1], that X is in the domain of attraction of the normal variable Y_2 if and only if $L(x) = \mathbf{E}[X^2 I_{|X|\le x}]$ is slowly varying (as is of course the case whenever $\mathbf{E}X^2$ is finite).

As shown in the following exercise, controlling the modulus of the remainder term for the n-th order Taylor approximation of e^{ix} one can generalize the bound on $\Phi_X(\theta)$ beyond the case $n = 2$ of Lemma 3.3.29.

Exercise 3.3.34. *For any $x \in \mathbb{R}$ and non-negative integer n, let*

$$R_n(x) = e^{ix} - \sum_{k=0}^{n} \frac{(ix)^k}{k!}\,.$$

(a) *Show that $R_n(x) = \int_0^x iR_{n-1}(y)dy$ for all $n \ge 1$ and deduce by induction on n that*

$$|R_n(x)| \le \min\Big(\frac{2|x|^n}{n!}, \frac{|x|^{n+1}}{(n+1)!}\Big) \quad \text{for all} \quad x \in \mathbb{R}, n = 0,1,2,\ldots.$$

(b) *Conclude that if $\mathbf{E}|X|^n < \infty$ then*

$$\big|\Phi_X(\theta) - \sum_{k=0}^{n} \frac{(i\theta)^k \mathbf{E}X^k}{k!}\big| \le |\theta|^n \mathbf{E}\Big[\min\big(\frac{2|X|^n}{n!}, \frac{|\theta||X|^{n+1}}{(n+1)!}\big)\Big]\,.$$

By solving the next exercise you generalize the proof of Theorem 3.1.2 via characteristic functions to the setting of Lindeberg's CLT.

Exercise 3.3.35. *Consider $\widehat{S}_n = \sum_{k=1}^n X_{n,k}$ for mutually independent random variables $X_{n,k}$, $k = 1,\ldots,n$, of zero mean and variance $v_{n,k}$, such that $v_n = \sum_{k=1}^n v_{n,k} \to 1$ as $n \to \infty$.*

(a) *Fixing $\theta \in \mathbb{R}$ show that*

$$\varphi_n = \Phi_{\widehat{S}_n}(\theta) = \prod_{k=1}^{n}(1 + z_{n,k})\,,$$

where $z_{n,k} = \Phi_{X_{n,k}}(\theta) - 1$.

(b) *With $z_\infty = -\theta^2/2$, use Lemma 3.3.29 to verify that $|z_{n,k}| \le 2\theta^2 v_{n,k}$ and further, for any $\varepsilon > 0$,*

$$|z_n - v_n z_\infty| \le \sum_{k=1}^{n} |z_{n,k} - v_{n,k} z_\infty| \le \theta^2 g_n(\varepsilon) + \frac{|\theta|^3}{6}\varepsilon v_n\,,$$

where $z_n = \sum_{k=1}^n z_{n,k}$ and $g_n(\varepsilon)$ is given by (3.1.4).

(c) *Recall that Lindeberg's condition $g_n(\varepsilon) \to 0$ implies that $r_n^2 = \max_k v_{n,k} \to 0$ as $n \to \infty$. Deduce that then $z_n \to z_\infty$ and $\eta_n = \sum_{k=1}^n |z_{n,k}|^2 \to 0$ when $n \to \infty$.*

(d) *Applying Lemma 3.3.30, conclude that* $\widehat{S}_n \xrightarrow{\mathcal{D}} G$.

We conclude this section with an exercise that reviews various techniques one may use for establishing convergence in distribution for sums of independent random variables.

Exercise 3.3.36. *Throughout this problem* $S_n = \sum_{k=1}^n X_k$ *for mutually independent random variables* $\{X_k\}$.

(a) *Suppose that* $\mathbf{P}(X_k = k^\alpha) = \mathbf{P}(X_k = -k^\alpha) = 1/(2k^\beta)$ *and* $\mathbf{P}(X_k = 0) = 1 - k^{-\beta}$. *Show that for any fixed* $\alpha \in \mathbb{R}$ *and* $\beta > 1$, *the series* $S_n(\omega)$ *converges almost surely as* $n \to \infty$.
(b) *Consider the setting of part (a) when* $0 \le \beta < 1$ *and* $\gamma = 2\alpha - \beta + 1$ *is positive. Find non-random* b_n *such that* $b_n^{-1} S_n \xrightarrow{\mathcal{D}} Z$ *and* $0 < F_Z(z) < 1$ *for some* $z \in \mathbb{R}$. *Provide also the characteristic function* $\Phi_Z(\theta)$ *of* Z.
(c) *Repeat part (b) in case* $\beta = 1$ *and* $\alpha > 0$ *(see Exercise 3.1.11 for* $\alpha = 0$*).*
(d) *Suppose now that* $\mathbf{P}(X_k = 2k) = \mathbf{P}(X_k = -2k) = 1/(2k^2)$ *and* $\mathbf{P}(X_k = 1) = \mathbf{P}(X_k = -1) = 0.5(1 - k^{-2})$. *Show that* $S_n/\sqrt{n} \xrightarrow{\mathcal{D}} G$.

3.4. Poisson approximation and the Poisson process

Subsection 3.4.1 deals with the Poisson approximation theorem and few of its applications. It leads naturally to the introduction of the Poisson process in Subsection 3.4.2, where we also explore its relation to sums of i.i.d. Exponential variables and to order statistics of i.i.d. uniform random variables.

3.4.1. Poisson approximation.

The Poisson approximation theorem is about the law of the sum S_n of a large number ($= n$) of independent random variables. In contrast to the CLT that also deals with such objects, here all variables are non-negative integer valued and the variance of S_n remains bounded, allowing for the approximation in law of S_n by an integer valued variable. The Poisson distribution results when the number of terms in the sum grows while the probability that each of them is non-zero decays. As such, the Poisson approximation is about counting the number of occurrences among many independent rare events.

Theorem 3.4.1 (Poisson approximation). *Let* $S_n = \sum_{k=1}^{n} Z_{n,k}$, *where for each* n *the random variables* $Z_{n,k}$ *for* $1 \le k \le n$, *are mutually independent, each taking value in the set of non-negative integers. Suppose that* $p_{n,k} = \mathbf{P}(Z_{n,k} = 1)$ *and* $\varepsilon_{n,k} = \mathbf{P}(Z_{n,k} \ge 2)$ *are such that as* $n \to \infty$,

(a) $\displaystyle\sum_{k=1}^{n} p_{n,k} \to \lambda < \infty$,
(b) $\displaystyle\max_{k=1,\cdots,n} \{p_{n,k}\} \to 0$,
(c) $\displaystyle\sum_{k=1}^{n} \varepsilon_{n,k} \to 0$.

Then, $S_n \xrightarrow{\mathcal{D}} N_\lambda$ *of a Poisson distribution with parameter* λ, *as* $n \to \infty$.

Proof. The first step of the proof is to apply truncation by comparing S_n with

$$\overline{S}_n = \sum_{k=1}^{n} \overline{Z}_{n,k} \,,$$

where $\overline{Z}_{n,k} = Z_{n,k} I_{Z_{n,k} \le 1}$ for $k = 1, \dots, n$. Indeed, observe that,

$$\begin{aligned} \mathbf{P}(\overline{S}_n \neq S_n) &\le \sum_{k=1}^{n} \mathbf{P}(\overline{Z}_{n,k} \neq Z_{n,k}) = \sum_{k=1}^{n} \mathbf{P}(Z_{n,k} \ge 2) \\ &= \sum_{k=1}^{n} \varepsilon_{n,k} \to 0 \quad \text{for } n \to \infty, \qquad \text{by assumption} \quad (c)\,. \end{aligned}$$

Hence, $(\overline{S}_n - S_n) \xrightarrow{p} 0$. Consequently, the convergence $\overline{S}_n \xrightarrow{\mathcal{D}} N_\lambda$ of the sums of truncated variables imply that also $S_n \xrightarrow{\mathcal{D}} N_\lambda$ (c.f. Exercise 3.2.8).

As seen in the context of the CLT, characteristic functions are a powerful tool for the convergence in distribution of sums of independent random variables (see Subsection 3.3.3). This is also evident in our proof of the Poisson approximation theorem. That is, to prove that $\overline{S}_n \xrightarrow{\mathcal{D}} N_\lambda$, if suffices by Levy's continuity theorem to show the convergence of the characteristic functions $\Phi_{\overline{S}_n}(\theta) \to \Phi_{N_\lambda}(\theta)$ for each $\theta \in \mathbb{R}$.

To this end, recall that $\overline{Z}_{n,k}$ are independent Bernoulli variables of parameters $p_{n,k}$, $k = 1, \dots, n$. Hence, by Lemma 3.3.8 and Example 3.3.5 we have that for $z_{n,k} = p_{n,k}(e^{i\theta} - 1)$,

$$\Phi_{\overline{S}_n}(\theta) = \prod_{k=1}^{n} \Phi_{\overline{Z}_{n,k}}(\theta) = \prod_{k=1}^{n} (1 - p_{n,k} + p_{n,k} e^{i\theta}) = \prod_{k=1}^{n} (1 + z_{n,k})\,.$$

Our assumption (a) implies that for $n \to \infty$

$$z_n := \sum_{k=1}^{n} z_{n,k} = (\sum_{k=1}^{n} p_{n,k})(e^{i\theta} - 1) \to \lambda(e^{i\theta} - 1) := z_\infty\,.$$

Further, since $|z_{n,k}| \le 2p_{n,k}$, our assumptions (a) and (b) imply that for $n \to \infty$,

$$\eta_n = \sum_{k=1}^{n} |z_{n,k}|^2 \le 4 \sum_{k=1}^{n} p_{n,k}^2 \le 4(\max_{k=1,\dots,n}\{p_{n,k}\})(\sum_{k=1}^{n} p_{n,k}) \to 0\,.$$

Applying Lemma 3.3.30 we conclude that when $n \to \infty$,

$$\Phi_{\overline{S}_n}(\theta) \to \exp(z_\infty) = \exp(\lambda(e^{i\theta} - 1)) = \Phi_{N_\lambda}(\theta)$$

(see (3.3.3) for the last identity), thus completing the proof. □

Remark. Recall Example 3.2.25 that the weak convergence of the laws of the integer valued S_n to that of N_λ also implies their convergence in total variation. In the setting of the Poisson approximation theorem, taking $\lambda_n = \sum_{k=1}^{n} p_{n,k}$, the more quantitative result

$$||\mathcal{P}_{\overline{S}_n} - \mathcal{P}_{N_{\lambda_n}}||_{tv} = \sum_{k=0}^{\infty} |\mathbf{P}(\overline{S}_n = k) - \mathbf{P}(N_{\lambda_n} = k)| \le 2\min(\lambda_n^{-1}, 1) \sum_{k=1}^{n} p_{n,k}^2$$

due to Stein (1987) also holds (see also [**Dur10**, (3.6.1)] for a simpler argument, due to Hodges and Le Cam (1960), which is just missing the factor $\min(\lambda_n^{-1}, 1)$).

For the remainder of this subsection we list applications of the Poisson approximation theorem, starting with

Example 3.4.2 (Poisson approximation for the Binomial). *Take independent variables $Z_{n,k} \in \{0,1\}$, so $\varepsilon_{n,k} = 0$, with $p_{n,k} = p_n$ that does not depend on k. Then, the variable $S_n = \overline{S}_n$ has the Binomial distribution of parameters (n, p_n). By Stein's result, the Binomial distribution of parameters (n, p_n) is approximated well by the Poisson distribution of parameter $\lambda_n = np_n$, provided $p_n \to 0$. In case $\lambda_n = np_n \to \lambda < \infty$, Theorem 3.4.1 yields that the Binomial (n, p_n) laws converge weakly as $n \to \infty$ to the Poisson distribution of parameter λ. This is in agreement with Example 3.1.7 where we approximate the Binomial distribution of parameters (n, p) by the normal distribution, for in Example 3.1.8 we saw that, upon the same scaling, N_{λ_n} is also approximated well by the normal distribution when $\lambda_n \to \infty$.*

Recall the occupancy problem where we distribute at random r distinct balls among n distinct boxes and each of the possible n^r assignments of balls to boxes is equally likely. In Example 2.1.10 we considered the asymptotic fraction of empty boxes when $r/n \to \alpha$ and $n \to \infty$. Noting that the number of balls $M_{n,k}$ in the k-th box follows the Binomial distribution of parameters (r, n^{-1}), we deduce from Example 3.4.2 that $M_{n,k} \xrightarrow{\mathcal{D}} N_\alpha$. Thus, $\mathbf{P}(M_{n,k} = 0) \to \mathbf{P}(N_\alpha = 0) = e^{-\alpha}$. That is, for large n each box is empty with probability about $e^{-\alpha}$, which may explain (though not prove) the result of Example 2.1.10. Here we use the Poisson approximation theorem to tackle a different regime, in which $r = r_n$ is of order $n \log n$, and consequently, there are fewer empty boxes.

Proposition 3.4.3. *Let S_n denote the number of empty boxes. Assuming $r = r_n$ is such that $ne^{-r/n} \to \lambda \in [0, \infty)$, we have that $S_n \xrightarrow{\mathcal{D}} N_\lambda$ as $n \to \infty$.*

Proof. Let $Z_{n,k} = I_{M_{n,k}=0}$ for $k = 1, \ldots, n$, that is $Z_{n,k} = 1$ if the k-th box is empty and $Z_{n,k} = 0$ otherwise. Note that $S_n = \sum_{k=1}^n Z_{n,k}$, with each $Z_{n,k}$ having the Bernoulli distribution of parameter $p_n = (1 - n^{-1})^r$. Our assumption about r_n guarantees that $np_n \to \lambda$. If the occupancy $Z_{n,k}$ of the various boxes were mutually independent, then the stated convergence of S_n to N_λ would have followed from Theorem 3.4.1. Unfortunately, this is not the case, so we present a bare-hands approach showing that the dependence is weak enough to retain the same conclusion. To this end, first observe that for any $l = 1, 2, \ldots, n$, the probability that given boxes $k_1 < k_2 < \ldots < k_l$ are all empty is,

$$\mathbf{P}(Z_{n,k_1} = Z_{n,k_2} = \cdots = Z_{n,k_l} = 1) = (1 - \frac{l}{n})^r .$$

Let $p_l = p_l(r, n) = \mathbf{P}(S_n = l)$ denote the probability that exactly l boxes are empty out of the n boxes into which the r balls are placed at random. Then, considering all possible choices of the locations of these $l \geq 1$ empty boxes we get the identities $p_l(r, n) = b_l(r, n) p_0(r, n - l)$ for

$$b_l(r, n) = \binom{n}{l} \big(1 - \frac{l}{n}\big)^r . \tag{3.4.1}$$

Further, $p_0(r, n) = 1 - \mathbf{P}($ at least one empty box$)$, so that by the inclusion-exclusion formula,

$$p_0(r, n) = \sum_{l=0}^{n} (-1)^l b_l(r, n) . \tag{3.4.2}$$

According to part (b) of Exercise 3.4.4, $p_0(r,n) \to e^{-\lambda}$. Further, for fixed l we have that $(n-l)e^{-r/(n-l)} \to \lambda$, so as before we conclude that $p_0(r,n-l) \to e^{-\lambda}$. By part (a) of Exercise 3.4.4 we know that $b_l(r,n) \to \lambda^l/l!$ for fixed l, hence $p_l(r,n) \to e^{-\lambda}\lambda^l/l!$. As $p_l = \mathbf{P}(S_n = l)$, the proof of the proposition is thus complete. □

The following exercise provides the estimates one needs during the proof of Proposition 3.4.3 (for more details, see [**Dur10**, Theorem 3.6.5]).

Exercise 3.4.4. *Assuming $ne^{-r/n} \to \lambda$, show that*

(a) *$b_l(r,n)$ of (3.4.1) converges to $\lambda^l/l!$ for each fixed l.*
(b) *$p_0(r,n)$ of (3.4.2) converges to $e^{-\lambda}$.*

Finally, here is an application of Proposition 3.4.3 to the coupon collector's problem of Example 2.1.8, where T_n denotes the number of independent trials, it takes to have at least one representative of each of the n possible values (and each trial produces a value U_i that is distributed uniformly on the set of n possible values).

Example 3.4.5 (Revisiting the coupon collector's problem)**.** *For any $x \in \mathbb{R}$, we have that*

$$\lim_{n\to\infty} \mathbf{P}(T_n - n\log n \le nx) = \exp(-e^{-x}), \tag{3.4.3}$$

which is an improvement over our weak law result that $T_n/n\log n \to 1$. Indeed, to derive (3.4.3) view the first r trials of the coupon collector as the random placement of r balls into n distinct boxes that correspond to the n possible values. From this point of view, the event $\{T_n \le r\}$ corresponds to filling all n boxes with the r balls, that is, having none empty. Taking $r = r_n = [n\log n + nx]$ we have that $ne^{-r/n} \to \lambda = e^{-x}$, and so it follows from Proposition 3.4.3 that $\mathbf{P}(T_n \le r_n) \to \mathbf{P}(N_\lambda = 0) = e^{-\lambda}$, as stated in (3.4.3).

Note that though $T_n = \sum_{k=1}^n X_{n,k}$ with $X_{n,k}$ independent, the convergence in distribution of T_n, given by (3.4.3), is to a non-normal limit. This should not surprise you, for the terms $X_{n,k}$ with k near n are large and do not satisfy Lindeberg's condition.

Exercise 3.4.6. *Recall that τ_ℓ^n denotes the first time one has ℓ distinct values when collecting coupons that are uniformly distributed on $\{1,2,\dots,n\}$. Using the Poisson approximation theorem show that if $n \to \infty$ and $\ell = \ell(n)$ is such that $n^{-1/2}\ell \to \lambda \in [0,\infty)$, then $\tau_\ell^n - \ell \xrightarrow{\mathcal{D}} N$ with N a Poisson random variable of parameter $\lambda^2/2$.*

3.4.2. Poisson Process.

The Poisson process is a continuous time stochastic process $\omega \mapsto N_t(\omega)$, $t \ge 0$ which belongs to the following class of counting processes.

Definition 3.4.7. *A* counting process *is a mapping $\omega \longmapsto N_t(\omega)$, where $N_t(\omega)$ is a piecewise constant, non-decreasing, right continuous function of $t \ge 0$, with $N_0(\omega) = 0$ and (countably) infinitely many jump discontinuities, each of whom is of size one.*

Associated with each sample path *$N_t(\omega)$ of such a process are the jump times $0 = T_0 < T_1 < \cdots < T_n < \cdots$ such that $T_k = \inf\{t \ge 0 : N_t \ge k\}$ for each k, or equivalently*

$$N_t = \sup\{k \ge 0 : T_k \le t\}.$$

In applications we find such N_t as counting the number of discrete events occurring in the interval $[0,t]$ for each $t \geq 0$, with T_k denoting the arrival or occurrence time of the k-th such event.

Remark. It is possible to extend the notion of counting processes to discrete events indexed on $\mathbb{R}^d$, $d \geq 2$. This is done by assigning random integer counts N_A to Borel subsets A of $\mathbb{R}^d$ in an additive manner, that is, $N_{A \cup B} = N_A + N_B$ whenever A and B are disjoint. Such processes are called *point processes*. See also Exercise 7.1.13 for more about *Poisson point process* and inhomogeneous Poisson processes of non-constant rate.

Among all counting processes we characterize the Poisson process by the joint distribution of its jump (arrival) times $\{T_k\}$.

Definition 3.4.8. *The* Poisson process *of* rate $\lambda > 0$ *is the unique counting process with the gaps between jump times $\tau_k = T_k - T_{k-1}$, $k = 1, 2, \ldots$ being i.i.d. random variables, each having the* exponential distribution *of parameter λ.*

Thus, from Exercise 1.4.46 we deduce that the k-th arrival time T_k of the Poisson process of rate λ has the *gamma density* of parameters $\alpha = k$ and λ,

$$f_{T_k}(u) = \frac{\lambda^k u^{k-1}}{(k-1)!} e^{-\lambda u} \mathbf{1}_{u>0} \,.$$

As we have seen in Example 2.3.7, counting processes appear in the context of renewal theory. In particular, as shown in Exercise 2.3.8, the Poisson process of rate λ satisfies the strong law of large numbers $t^{-1} N_t \overset{a.s.}{\to} \lambda$.

Recall that a random variable N has the Poisson(μ) law if

$$\mathbf{P}(N = n) = \frac{\mu^n}{n!} e^{-\mu}, \quad n = 0, 1, 2, \ldots \,.$$

Our next proposition, which is often used as an alternative definition of the Poisson process, also explains its name.

Proposition 3.4.9. *For any ℓ and any $0 = t_0 < t_1 < \cdots < t_\ell$, the increments N_{t_1}, $N_{t_2} - N_{t_1}$, $\ldots$, $N_{t_\ell} - N_{t_{\ell-1}}$, are independent random variables and for some $\lambda > 0$ and all $t > s \geq 0$, the increment $N_t - N_s$ has the* Poisson($\lambda(t-s)$) *law.*

Thus, the Poisson process has independent increments, each having a Poisson law, where the parameter of the count $N_t - N_s$ is proportional to the length of the corresponding interval $[s, t]$.

The proof of Proposition 3.4.9 relies on the *lack of memory* of the exponential distribution. That is, if the law of a random variable T is exponential (of some parameter $\lambda > 0$), then for all $t, s \geq 0$,

$$\mathbf{P}(T > t + s | T > t) = \frac{\mathbf{P}(T > t + s)}{\mathbf{P}(T > t)} = \frac{e^{-\lambda(t+s)}}{e^{-\lambda t}} = e^{-\lambda s} = \mathbf{P}(T > s)\,. \tag{3.4.4}$$

Indeed, the key to the proof of Proposition 3.4.9 is the following lemma.

Lemma 3.4.10. *Fixing $t > 0$, the variables $\{\tau'_j\}$ with $\tau'_1 = T_{N_t+1} - t$, and $\tau'_j = T_{N_t+j} - T_{N_t+j-1}$, $j \geq 2$ are i.i.d. each having the exponential distribution of parameter λ. Further, the collection $\{\tau'_j\}$ is independent of N_t which has the Poisson distribution of parameter λt.*

Remark. Note that in particular, $E_t = T_{N_t+1} - t$ which counts the time till next arrival occurs, hence called *the excess life time* at t, follows the exponential distribution of parameter λ.

PROOF. Fixing $t > 0$ and $n \geq 1$ let $H_n(x) = \mathbf{P}(t \geq T_n > t - x)$. With $H_n(x) = \int_0^x f_{T_n}(t-y)dy$ and T_n independent of τ_{n+1}, we get by Fubini's theorem (for $I_{t \geq T_n > t - \tau_{n+1}}$), and the integration by parts of Lemma 1.4.30 that

$$\begin{aligned}\mathbf{P}(N_t = n) &= \mathbf{P}(t \geq T_n > t - \tau_{n+1}) = \mathbf{E}[H_n(\tau_{n+1})] \\ &= \int_0^t f_{T_n}(t-y)\mathbf{P}(\tau_{n+1} > y)dy \\ &= \int_0^t \frac{\lambda^n (t-y)^{n-1}}{(n-1)!} e^{-\lambda(t-y)} e^{-\lambda y} dy = e^{-\lambda t}\frac{(\lambda t)^n}{n!}\,. \end{aligned} \tag{3.4.5}$$

As this applies for any $n \geq 1$, it follows that N_t has the Poisson distribution of parameter λt. Similarly, observe that for any $s_1 \geq 0$ and $n \geq 1$,

$$\begin{aligned}\mathbf{P}(N_t = n, \tau_1' > s_1) &= \mathbf{P}(t \geq T_n > t - \tau_{n+1} + s_1) \\ &= \int_0^t f_{T_n}(t-y)\mathbf{P}(\tau_{n+1} > s_1 + y)dy \\ &= e^{-\lambda s_1}\mathbf{P}(N_t = n) = \mathbf{P}(\tau_1 > s_1)\mathbf{P}(N_t = n)\,. \end{aligned}$$

Since $T_0 = 0$, $\mathbf{P}(N_t = 0) = e^{-\lambda t}$ and $T_1 = \tau_1$, in view of (3.4.4) this conclusion extends to $n = 0$, proving that τ_1' is independent of N_t and has the same exponential law as τ_1.

Next, fix arbitrary integer $k \geq 2$ and non-negative $s_j \geq 0$ for $j = 1, \ldots, k$. Then, for any $n \geq 0$, since $\{\tau_{n+j}, j \geq 2\}$ are i.i.d. and independent of (T_n, τ_{n+1}),

$$\begin{aligned}&\mathbf{P}(N_t = n, \tau_j' > s_j, j = 1, \ldots, k) \\ &= \mathbf{P}(t \geq T_n > t - \tau_{n+1} + s_1, T_{n+j} - T_{n+j-1} > s_j, j = 2, \ldots, k) \\ &= \mathbf{P}(t \geq T_n > t - \tau_{n+1} + s_1)\prod_{j=2}^k \mathbf{P}(\tau_{n+j} > s_j) = \mathbf{P}(N_t = n)\prod_{j=1}^k \mathbf{P}(\tau_j > s_j). \end{aligned}$$

Since $s_j \geq 0$ and $n \geq 0$ are arbitrary, this shows that the random variables N_t and τ_j', $j = 1, \ldots, k$ are mutually independent (c.f. Corollary 1.4.12), with each τ_j' having an exponential distribution of parameter λ. As k is arbitrary, the independence of N_t and the countable collection $\{\tau_j'\}$ follows by Definition 1.4.3. $\square$

PROOF OF PROPOSITION 3.4.9. Fix $t, s_j \geq 0$, $j = 1, \ldots, k$, and non-negative integers n and m_j, $1 \leq j \leq k$. The event $\{N_{s_j} = m_j, 1 \leq j \leq k\}$ is of the form $\{(\tau_1, \ldots, \tau_r) \in H\}$ for $r = m_k + 1$ and

$$H = \bigcap_{j=1}^k \{\underline{x} \in [0,\infty)^r : x_1 + \cdots + x_{m_j} \leq s_j < x_1 + \cdots + x_{m_j+1}\}\,.$$

Since the event $\{(\tau_1', \ldots, \tau_r') \in H\}$ is merely $\{N_{t+s_j} - N_t = m_j, 1 \leq j \leq k\}$, it follows form Lemma 3.4.10 that

$$\begin{aligned}&\mathbf{P}(N_t = n, N_{t+s_j} - N_t = m_j, 1 \leq j \leq k) = \mathbf{P}(N_t = n, (\tau_1', \ldots, \tau_r') \in H) \\ &= \mathbf{P}(N_t = n)\mathbf{P}((\tau_1, \ldots, \tau_r) \in H) = \mathbf{P}(N_t = n)\mathbf{P}(N_{s_j} = m_j, 1 \leq j \leq k)\,. \end{aligned}$$

By induction on ℓ this identity implies that if $0 = t_0 < t_1 < t_2 < \cdots < t_\ell$, then

$$\mathbf{P}(N_{t_i} - N_{t_{i-1}} = n_i, 1 \le i \le \ell) = \prod_{i=1}^{\ell} \mathbf{P}(N_{t_i - t_{i-1}} = n_i) \tag{3.4.6}$$

(the case $\ell = 1$ is trivial, and to advance the induction to $\ell + 1$ set $k = \ell$, $t = t_1$, $n = n_1$ and $s_j = t_{j+1} - t_1$, $m_j = \sum_{i=2}^{j+1} n_i$).

Considering (3.4.6) for $\ell = 2$, $t_2 = t > s = t_1$, and summing over the values of n_1 we see that $\mathbf{P}(N_t - N_s = n_2) = \mathbf{P}(N_{t-s} = n_2)$, hence by (3.4.5) we conclude that $N_t - N_s$ has the Poisson distribution of parameter $\lambda(t-s)$, as claimed. $\square$

The Poisson process is also related to the *order statistics* $\{V_{n,k}\}$ for the uniform measure, as stated in the next two exercises.

Exercise 3.4.11. *Let $U_1, U_2, \ldots, U_n$ be i.i.d. with each U_i having the uniform measure on $(0, 1]$. Denote by $V_{n,k}$ the k-th smallest number in $\{U_1, \ldots, U_n\}$.*

(a) *Show that $(V_{n,1}, \ldots, V_{n,n})$ has the same law as $(T_1/T_{n+1}, \ldots, T_n/T_{n+1})$, where $\{T_k\}$ are the jump (arrival) times for a Poisson process of rate λ (see Subsection 1.4.2 for the definition of the* law $\mathcal{P}_{\underline{X}}$ *of a random vector $\underline{X}$).*

(b) *Taking $\lambda = 1$, deduce that $nV_{n,k} \xrightarrow{\mathcal{D}} T_k$ as $n \to \infty$ while k is fixed, where T_k has the* gamma density *of parameters $\alpha = k$ and $s = 1$.*

Exercise 3.4.12. *Fixing any positive integer n and $0 \le t_1 \le t_2 \le \cdots \le t_n \le t$, show that*

$$\mathbf{P}(T_k \le t_k,\ k = 1, \ldots, n | N_t = n) = \frac{n!}{t^n} \int_0^{t_1} \int_{x_1}^{t_2} \cdots \left(\int_{x_{n-1}}^{t_n} dx_n \right) dx_{n-1} \cdots dx_1 \,.$$

That is, conditional on the event $N_t = n$, the first n jump times $\{T_k : k = 1, \ldots, n\}$ have the same law as the order statistics *$\{V_{n,k} : k = 1, \ldots, n\}$ of a sample of n i.i.d random variables $U_1, \ldots, U_n$, each of which is uniformly distributed in $[0, t]$.*

Here is an application of Exercise 3.4.12.

Exercise 3.4.13. *Consider a Poisson process N_t of rate λ and jump times $\{T_k\}$.*

(a) *Compute the values of $g(n) = \mathbf{E}(I_{N_t = n} \sum_{k=1}^{n} T_k)$.*

(b) *Compute the value of $v = \mathbf{E}(\sum_{k=1}^{N_t} (t - T_k))$.*

(c) *Suppose that T_k is the arrival time to the train station of the k-th passenger on a train that departs the station at time t. What is the meaning of N_t and of v in this case?*

The representation of the *order statistics* $\{V_{n,k}\}$ in terms of the jump times of a Poisson process is very useful when studying the large n asymptotics of their spacings $\{R_{n,k}\}$. For example,

Exercise 3.4.14. *Let $R_{n,k} = V_{n,k} - V_{n,k-1}$, $k = 1, \ldots, n$, denote the spacings between $V_{n,k}$ of Exercise 3.4.11 (with $V_{n,0} = 0$). Show that as $n \to \infty$,*

$$\frac{n}{\log n} \max_{k=1,\ldots,n} R_{n,k} \xrightarrow{p} 1 \,, \tag{3.4.7}$$

and further for each fixed $x \geq 0$,

$$G_n(x) := n^{-1} \sum_{k=1}^{n} I_{\{R_{n,k} > x/n\}} \xrightarrow{p} e^{-x}, \tag{3.4.8}$$

$$B_n(x) := \mathbf{P}(\min_{k=1,\ldots,n} R_{n,k} > x/n^2) \to e^{-x}. \tag{3.4.9}$$

As we show next, the Poisson approximation theorem provides a characterization of the Poisson process that is very attractive for modeling real-world phenomena.

Corollary 3.4.15. *If N_t is a Poisson process of rate $\lambda > 0$, then for any fixed k, $0 < t_1 < t_2 < \cdots < t_k$ and non-negative integers $n_1, n_2, \cdots, n_k$,*

$$\mathbf{P}(N_{t_k+h} - N_{t_k} = 1 | N_{t_j} = n_j,\ j \leq k) = \lambda h + o(h),$$
$$\mathbf{P}(N_{t_k+h} - N_{t_k} \geq 2 | N_{t_j} = n_j,\ j \leq k) = o(h),$$

where $o(h)$ denotes a function $f(h)$ such that $h^{-1} f(h) \to 0$ as $h \downarrow 0$.

Proof. Fixing k, the t_j and the n_j, denote by A the event $\{N_{t_j} = n_j,\ j \leq k\}$. For a Poisson process of rate λ the random variable $N_{t_k+h} - N_{t_k}$ is independent of A with $\mathbf{P}(N_{t_k+h} - N_{t_k} = 1) = e^{-\lambda h} \lambda h$ and $\mathbf{P}(N_{t_k+h} - N_{t_k} \geq 2) = 1 - e^{-\lambda h}(1 + \lambda h)$. Since $e^{-\lambda h} = 1 - \lambda h + o(h)$ we see that the Poisson process satisfies this corollary. □

Our next exercise explores the phenomenon of *thinning*, that is, the partitioning of Poisson variables as sums of mutually independent Poisson variables of smaller parameter.

Exercise 3.4.16. *Suppose $\{X_i\}$ are i.i.d. with $\mathbf{P}(X_i = j) = p_j$ for $j = 0, 1, \ldots, k$ and N a Poisson random variable of parameter λ that is independent of $\{X_k\}$. Let*

$$N_j = \sum_{i=1}^{N} I_{X_i = j} \quad j = 0, \ldots, k.$$

(a) *Show that the variables N_j, $j = 0, 1, \ldots, k$ are mutually independent with N_j having a Poisson distribution of parameter λp_j.*
(b) *Show that the sub-sequence of jump times $\{\widetilde{T}_k\}$ obtained by independently keeping with probability p each of the jump times $\{T_k\}$ of a Poisson process N_t of rate λ, yields in turn a Poisson process $\widetilde{N}_t$ of rate λp.*

We conclude this section noting the *superposition* property, namely that the sum of two independent Poisson processes is yet another Poisson process.

Exercise 3.4.17. *Suppose $N_t = N_t^{(1)} + N_t^{(2)}$ where $N_t^{(1)}$ and $N_t^{(2)}$ are two independent Poisson processes of rates $\lambda_1 > 0$ and $\lambda_2 > 0$, respectively. Show that N_t is a Poisson process of rate $\lambda_1 + \lambda_2$.*

3.5. Random vectors and the multivariate CLT

The goal of this section is to extend the CLT to random vectors, that is, $\mathbb{R}^d$-valued random variables. Towards this end, we revisit in Subsection 3.5.1 the theory of weak convergence, this time in the more general setting of $\mathbb{R}^d$-valued random variables. Subsection 3.5.2 is devoted to the extension of characteristic functions and Lévy's theorems to the multivariate setting, culminating with the Cramér-wold reduction of convergence in distribution of random vectors to that of their one dimensional linear projections. Finally, in Subsection 3.5.3 we introduce the important concept of Gaussian random vectors and prove the multivariate CLT.

3.5.1. Weak convergence revisited. Recall Definition 3.2.17 of weak convergence for a sequence of probability measures on a topological space $\mathbb{S}$, which suggests the following definition for convergence in distribution of $\mathbb{S}$-valued random variables.

Definition 3.5.1. *We say that $(\mathbb{S}, \mathcal{B}_{\mathbb{S}})$-valued random variables X_n* converge in distribution *to a $(\mathbb{S}, \mathcal{B}_{\mathbb{S}})$-valued random variable X_∞, denoted by $X_n \xrightarrow{\mathcal{D}} X_\infty$, if $\mathcal{P}_{X_n} \overset{w}{\Rightarrow} \mathcal{P}_{X_\infty}$.*

As already remarked, the *Portmanteau theorem* about equivalent characterizations of the weak convergence holds also when the probability measures ν_n are on a Borel measurable space $(\mathbb{S}, \mathcal{B}_{\mathbb{S}})$ with $(\mathbb{S}, \rho)$ any metric space (and in particular for $\mathbb{S} = \mathbb{R}^d$).

Theorem 3.5.2 (PORTMANTEAU THEOREM)**.** *The following five statements are equivalent for any probability measures ν_n, $1 \le n \le \infty$ on $(\mathbb{S}, \mathcal{B}_{\mathbb{S}})$, with $(\mathbb{S}, \rho)$ any metric space.*

(a) $\nu_n \overset{w}{\Rightarrow} \nu_\infty$
(b) *For every closed set F, one has* $\limsup_{n\to\infty} \nu_n(F) \le \nu_\infty(F)$
(c) *For every open set G, one has* $\liminf_{n\to\infty} \nu_n(G) \ge \nu_\infty(G)$
(d) *For every ν_∞-continuity set A, one has* $\lim_{n\to\infty} \nu_n(A) = \nu_\infty(A)$
(e) *If the Borel function $g : \mathbb{S} \mapsto \mathbb{R}$ is such that $\nu_\infty(\mathbf{D}_g) = 0$, then $\nu_n \circ g^{-1} \overset{w}{\Rightarrow} \nu_\infty \circ g^{-1}$ and if in addition g is bounded then $\nu_n(g) \to \nu_\infty(g)$.*

Remark. For $\mathbb{S} = \mathbb{R}$, the equivalence of (a)–(d) is the content of Theorem 3.2.21 while Proposition 3.2.19 derives (e) out of (a) (in the context of convergence in distribution, that is, $X_n \xrightarrow{\mathcal{D}} X_\infty$ and $\mathbf{P}(X_\infty \in \mathbf{D}_g) = 0$ implying that $g(X_n) \xrightarrow{\mathcal{D}} g(X_\infty)$). In addition to proving the converse of the continuous mapping property, we extend the validity of this equivalence to any metric space $(\mathbb{S}, \rho)$, for we shall apply it again in Subsection 9.2, considering there $\mathbb{S} = C([0,\infty))$, the metric space of all continuous functions on $[0, \infty)$.

PROOF. The derivation of $(b) \Rightarrow (c) \Rightarrow (d)$ in Theorem 3.2.21 applies for any topological space. The direction $(e) \Rightarrow (a)$ is also obvious since $h \in C_b(\mathbb{S})$ has $\mathbf{D}_h = \emptyset$ and $C_b(\mathbb{S})$ is a subset of the bounded Borel functions on the same space (c.f. Exercise 1.2.20). So taking $g \in C_b(\mathbb{S})$ in (e) results with (a). It thus remains only to show that $(a) \Rightarrow (b)$ and that $(d) \Rightarrow (e)$, which we proceed to show next.
$(a) \Rightarrow (b)$. Fixing $A \in \mathcal{B}_{\mathbb{S}}$ let $\rho_A(x) = \inf_{y \in A} \rho(x, y) : \mathbb{S} \mapsto [0, \infty)$. Since $|\rho_A(x) - \rho_A(x')| \le \rho(x, x')$ for any x, x', it follows that $x \mapsto \rho_A(x)$ is a continuous function on $(\mathbb{S}, \rho)$. Consequently, $h_r(x) = (1 - r\rho_A(x))_+ \in C_b(\mathbb{S})$ for all $r \ge 0$. Further, $\rho_A(x) = 0$ for all $x \in A$, implying that $h_r \ge I_A$ for all r. Thus, applying part (a) of the Portmanteau theorem for h_r we have that

$$\limsup_{n\to\infty} \nu_n(A) \le \lim_{n\to\infty} \nu_n(h_r) = \nu_\infty(h_r) .$$

As $\rho_A(x) = 0$ if and only if $x \in \overline{A}$ it follows that $h_r \downarrow I_{\overline{A}}$ as $r \to \infty$, resulting with

$$\limsup_{n\to\infty} \nu_n(A) \le \nu_\infty(\overline{A}) .$$

Taking $A = \overline{A} = F$ a closed set, we arrive at part (b) of the theorem.
$(d) \Rightarrow (e)$. Fix a Borel function $g : \mathbb{S} \mapsto \mathbb{R}$ with $K = \sup_x |g(x)| < \infty$ such that $\nu_\infty(\mathbf{D}_g) = 0$. Clearly, $\{\alpha \in \mathbb{R} : \nu_\infty \circ g^{-1}(\{\alpha\}) > 0\}$ is a countable set. Thus, fixing $\varepsilon > 0$ we can pick $\ell < \infty$ and $\alpha_0 < \alpha_1 < \cdots < \alpha_\ell$ such that $\nu_\infty \circ g^{-1}(\{\alpha_i\}) = 0$ for $0 \le i \le \ell$, $\alpha_0 < -K < K < \alpha_\ell$ and $\alpha_i - \alpha_{i-1} < \varepsilon$ for $1 \le i \le \ell$. Let $A_i = \{x : \alpha_{i-1} < g(x) \le \alpha_i\}$ for $i = 1, \ldots, \ell$, noting that $\partial A_i \subset \{x : g(x) = \alpha_{i-1},$ or $g(x) = \alpha_i\} \cup \mathbf{D}_g$. Consequently, by our assumptions about $g(\cdot)$ and $\{\alpha_i\}$ we have that $\nu_\infty(\partial A_i) = 0$ for each $i = 1, \ldots, \ell$. It thus follows from part (d) of the Portmanteau theorem that

$$\sum_{i=1}^{\ell} \alpha_i \nu_n(A_i) \to \sum_{i=1}^{\ell} \alpha_i \nu_\infty(A_i)$$

as $n \to \infty$. Our choice of α_i and A_i is such that $g \le \sum_{i=1}^{\ell} \alpha_i I_{A_i} \le g + \varepsilon$, resulting with

$$\nu_n(g) \le \sum_{i=1}^{\ell} \alpha_i \nu_n(A_i) \le \nu_n(g) + \varepsilon$$

for $n = 1, 2, \ldots, \infty$. Considering first $n \to \infty$ followed by $\varepsilon \downarrow 0$, we establish that $\nu_n(g) \to \nu_\infty(g)$. More generally, recall that $\mathbf{D}_{h \circ g} \subseteq \mathbf{D}_g$ for any $g : \mathbb{S} \mapsto \mathbb{R}$ and $h \in C_b(\mathbb{R})$. Thus, by the preceding proof $\nu_n(h \circ g) \to \nu_\infty(h \circ g)$ as soon as $\nu_\infty(\mathbf{D}_g) = 0$. This applies for every $h \in C_b(\mathbb{R})$, so in this case $\nu_n \circ g^{-1} \stackrel{w}{\Rightarrow} \nu_\infty \circ g^{-1}$. ∎

We next show that the relation of Exercise 3.2.6 between convergences in probability and in distribution also extends to any metric space $(\mathbb{S}, \rho)$, a fact we will later use in Subsection 9.2, when considering the metric space of all continuous functions on $[0, \infty)$.

Corollary 3.5.3. *If random variables X_n, $1 \le n \le \infty$ on the same probability space and taking value in a metric space $(\mathbb{S}, \rho)$ are such that $\rho(X_n, X_\infty) \xrightarrow{p} 0$, then $X_n \xrightarrow{\mathcal{D}} X_\infty$.*

Proof. Fixing $h \in C_b(\mathbb{S})$ and $\varepsilon > 0$, we have by continuity of $h(\cdot)$ that $G_r \uparrow \mathbb{S}$, where

$$G_r = \{y \in \mathbb{S} : |h(x) - h(y)| \le \varepsilon \text{ whenever } \rho(x, y) \le r^{-1}\}.$$

By definition, if $X_\infty \in G_r$ and $\rho(X_n, X_\infty) \le r^{-1}$ then $|h(X_n) - h(X_\infty)| \le \varepsilon$. Hence, for any $n, r \ge 1$,

$$\mathbf{E}[|h(X_n) - h(X_\infty)|] \le \varepsilon + 2\|h\|_\infty (\mathbf{P}(X_\infty \notin G_r) + \mathbf{P}(\rho(X_n, X_\infty) > r^{-1})),$$

where $\|h\|_\infty = \sup_{x \in \mathbb{S}} |h(x)|$ is finite (by the boundedness of h). Considering $n \to \infty$ followed by $r \to \infty$ we deduce from the convergence in probability of $\rho(X_n, X_\infty)$ to zero, that

$$\limsup_{n \to \infty} \mathbf{E}[|h(X_n) - h(X_\infty)|] \le \varepsilon + 2\|h\|_\infty \lim_{r \to \infty} \mathbf{P}(X_\infty \notin G_r) = \varepsilon.$$

Since this applies for any $\varepsilon > 0$, it follows by the triangle inequality that $\mathbf{E}h(X_n) \to \mathbf{E}h(X_\infty)$ for all $h \in C_b(\mathbb{S})$, i.e. $X_n \xrightarrow{\mathcal{D}} X_\infty$. ∎

Remark. The notion of *distribution function* for an $\mathbb{R}^d$-valued random vector $\underline{X} = (X_1, \ldots, X_d)$ is

$$F_{\underline{X}}(\underline{x}) = \mathbf{P}(X_1 \le x_1, \ldots, X_d \le x_d).$$

Inducing a partial order on $\mathbb{R}^d$ by $\underline{x} \le \underline{y}$ if and only if $\underline{x} - \underline{y}$ has only non-negative coordinates, each distribution function $F_{\underline{X}}(\underline{x})$ has the three properties listed in Theorem 1.2.37. Unfortunately, these three properties are not sufficient for a given function $F : \mathbb{R}^d \mapsto [0,1]$ to be a distribution function. For example, since the measure of each rectangle $A = \prod_{i=1}^d (a_i, b_i]$ should be positive, the additional constraint of the form $\Delta_A F = \sum_{j=1}^{2^d} \pm F(\underline{x}_j) \ge 0$ should hold if $F(\cdot)$ is to be a distribution function. Here $\underline{x}_j$ enumerates the 2^d corners of the rectangle A and each corner is taken with a positive sign if and only if it has an even number of coordinates from the collection $\{a_1, \dots, a_d\}$. Adding the fourth property that $\Delta_A F \ge 0$ for each rectangle $A \subset \mathbb{R}^d$, we get the necessary and sufficient conditions for $F(\cdot)$ to be a distribution function of some $\mathbb{R}^d$-valued random variable (c.f. [**Bil95**, Theorem 12.5] for a detailed proof).

Recall Definition 3.2.31 of *uniform tightness*, where for $\mathbb{S} = \mathbb{R}^d$ we can take $K_\varepsilon = [-M_\varepsilon, M_\varepsilon]^d$ with no loss of generality. Though Prohorov's theorem about uniform tightness (i.e. Theorem 3.2.34) is beyond the scope of these notes, we shall only need in the sequel the fact that a uniformly tight sequence of probability measures has at least one limit point. This can be proved for $\mathbb{S} = \mathbb{R}^d$ in a manner similar to what we have done in Theorem 3.2.37 and Lemma 3.2.38 for $\mathbb{S} = \mathbb{R}^1$, using the corresponding concept of distribution function $F_{\underline{X}}(\cdot)$ (see [**Dur10**, Theorem 3.9.2] for more details).

3.5.2. Characteristic function.

We start by extending the useful notion of characteristic function to the context of $\mathbb{R}^d$-valued random variables (which we also call hereafter random vectors).

Definition 3.5.4. *Adopting the notation* $(\underline{x}, \underline{y}) = \sum_{i=1}^d x_i y_i$ *for* $\underline{x}, \underline{y} \in \mathbb{R}^d$, *a* random vector $\underline{X} = (X_1, X_2, \cdots, X_d)$ *with values in* $\mathbb{R}^d$ *has the* characteristic function

$$\Phi_{\underline{X}}(\underline{\theta}) = \mathbf{E}[e^{i(\underline{\theta}, \underline{X})}] ,$$

where $\underline{\theta} = (\theta_1, \theta_2, \cdots, \theta_d) \in \mathbb{R}^d$ *and* $i = \sqrt{-1}$.

Remark. The characteristic function $\Phi_{\underline{X}} : \mathbb{R}^d \mapsto \mathbb{C}$ exists for any $\underline{X}$ since

$$e^{i(\underline{\theta}, \underline{X})} = \cos(\underline{\theta}, \underline{X}) + i \sin(\underline{\theta}, \underline{X}) , \tag{3.5.1}$$

with both real and imaginary parts being bounded (hence integrable) random variables. Actually, it is easy to check that all five properties of Proposition 3.3.2 hold, where part (e) is modified to $\Phi_{\mathbf{A}^t \underline{X} + \underline{b}}(\underline{\theta}) = \exp(i(\underline{b}, \underline{\theta}))\Phi_{\underline{X}}(\mathbf{A}\underline{\theta})$, for any non-random $d \times d$-dimensional matrix $\mathbf{A}$ and $\underline{b} \in \mathbb{R}^d$ (with $\mathbf{A}^t$ denoting the transpose of the matrix $\mathbf{A}$).

Here is the extension of the notion of probability density function (as in Definition 1.2.40) to a random vector.

Definition 3.5.5. *Suppose* $f_{\underline{X}}$ *is a non-negative Borel measurable function with* $\int_{\mathbb{R}^d} f_{\underline{X}}(\underline{x}) d\underline{x} = 1$. *We say that a random vector* $\underline{X} = (X_1, \dots, X_d)$ *has a* probability density function $f_{\underline{X}}(\cdot)$ *if for every* $\underline{b} = (b_1, \dots, b_d)$,

$$F_{\underline{X}}(\underline{b}) = \int_{-\infty}^{b_1} \cdots \int_{-\infty}^{b_d} f_{\underline{X}}(x_1, \dots, x_d) dx_d \cdots dx_1$$

(such $f_{\underline{X}}$ is sometimes called the joint density of $X_1, \ldots, X_d$). This is the same as saying that the law of $\underline{X}$ is of the form $f_{\underline{X}} \lambda^d$ with λ^d the d-fold product Lebesgue measure on $\mathbb{R}^d$ (i.e. the $d > 1$ extension of Example 1.3.60).

Example 3.5.6. *We have the following extension of the Fourier transform formula (3.3.4) to random vectors $\underline{X}$ with density,*

$$\Phi_{\underline{X}}(\underline{\theta}) = \int_{\mathbb{R}^d} e^{i(\underline{\theta}, \underline{x})} f_{\underline{X}}(\underline{x}) d\underline{x}$$

(this is merely a special case of the extension of Corollary 1.3.62 to $h : \mathbb{R}^d \mapsto \mathbb{R}$).

We next state and prove the corresponding extension of Lévy's inversion theorem.

Theorem 3.5.7 (Lévy's inversion theorem)**.** *Suppose $\Phi_{\underline{X}}(\underline{\theta})$ is the characteristic function of random vector $\underline{X} = (X_1, \ldots, X_d)$ whose law is $\mathcal{P}_{\underline{X}}$, a probability measure on $(\mathbb{R}^d, \mathcal{B}_{\mathbb{R}^d})$. If $A = [a_1, b_1] \times \cdots \times [a_d, b_d]$ with $\mathcal{P}_{\underline{X}}(\partial A) = 0$, then*

$$\mathcal{P}_{\underline{X}}(A) = \lim_{T \to \infty} \int_{[-T,T]^d} \prod_{j=1}^d \psi_{a_j, b_j}(\theta_j) \Phi_{\underline{X}}(\underline{\theta}) d\underline{\theta} \tag{3.5.2}$$

for $\psi_{a,b}(\cdot)$ of (3.3.5). Further, the characteristic function determines the law *of a random vector. That is, if $\Phi_{\underline{X}}(\underline{\theta}) = \Phi_{\underline{Y}}(\underline{\theta})$ for all $\underline{\theta}$ then $\underline{X}$ has the same law as $\underline{Y}$.*

Proof. We derive (3.5.2) by adapting the proof of Theorem 3.3.12. First apply Fubini's theorem with respect to the product of Lebesgue's measure on $[-T, T]^d$ and the law of $\underline{X}$ (both of which are finite measures on $\mathbb{R}^d$) to get the identity

$$J_T(\underline{a}, \underline{b}) := \int_{[-T,T]^d} \prod_{j=1}^d \psi_{a_j, b_j}(\theta_j) \Phi_{\underline{X}}(\underline{\theta}) d\underline{\theta} = \int_{\mathbb{R}^d} \Big[\prod_{j=1}^d \int_{-T}^{T} h_{a_j, b_j}(x_j, \theta_j) d\theta_j \Big] d\mathcal{P}_{\underline{X}}(\underline{x})$$

(where $h_{a,b}(x, \theta) = \psi_{a,b}(\theta) e^{i\theta x}$). In the course of proving Theorem 3.3.12 we have seen that for $j = 1, \ldots, d$ the integral over θ_j is uniformly bounded in T and that it converges to $g_{a_j, b_j}(x_j)$ as $T \uparrow \infty$. Thus, by bounded convergence it follows that

$$\lim_{T \uparrow \infty} J_T(\underline{a}, \underline{b}) = \int_{\mathbb{R}^d} g_{\underline{a}, \underline{b}}(\underline{x}) d\mathcal{P}_{\underline{X}}(\underline{x}) \,,$$

where

$$g_{\underline{a}, \underline{b}}(\underline{x}) = \prod_{j=1}^d g_{a_j, b_j}(x_j) \,,$$

is zero on A^c and one on A^o (see the explicit formula for $g_{a,b}(x)$ provided there). So, our assumption that $\mathcal{P}_{\underline{X}}(\partial A) = 0$ implies that the limit of $J_T(\underline{a}, \underline{b})$ as $T \uparrow \infty$ is merely $\mathcal{P}_{\underline{X}}(A)$, thus establishing (3.5.2).

Suppose now that $\Phi_{\underline{X}}(\underline{\theta}) = \Phi_{\underline{Y}}(\underline{\theta})$ for all $\underline{\theta}$. Adapting the proof of Corollary 3.3.14 to the current setting, let $\mathcal{J} = \{\alpha \in \mathbb{R} : \mathbf{P}(X_j = \alpha) > 0 \text{ or } \mathbf{P}(Y_j = \alpha) > 0 \text{ for some } j = 1, \ldots, d\}$ noting that if all the coordinates $\{a_j, b_j, j = 1, \ldots, d\}$ of a rectangle A are from the complement of $\mathcal{J}$ then both $\mathcal{P}_{\underline{X}}(\partial A) = 0$ and $\mathcal{P}_{\underline{Y}}(\partial A) = 0$. Thus, by (3.5.2) we have that $\mathcal{P}_{\underline{X}}(A) = \mathcal{P}_{\underline{Y}}(A)$ for any A in the collection $\mathcal{C}$ of rectangles with coordinates in the complement of $\mathcal{J}$. Recall that $\mathcal{J}$ is countable, so for any rectangle A there exists $A_n \in \mathcal{C}$ such that $A_n \downarrow A$, and by continuity from above of both $\mathcal{P}_{\underline{X}}$ and $\mathcal{P}_{\underline{Y}}$ it follows that $\mathcal{P}_{\underline{X}}(A) = \mathcal{P}_{\underline{Y}}(A)$ for *every rectangle* A. In view of Proposition 1.1.39 and Exercise 1.1.21 this implies that the probability measures $\mathcal{P}_{\underline{X}}$ and $\mathcal{P}_{\underline{Y}}$ agree on all Borel subsets of $\mathbb{R}^d$. ∎

We next provide the ingredients needed when using characteristic functions en-route to the derivation of a convergence in distribution result for random vectors. To this end, we start with the following analog of Lemma 3.3.16.

Lemma 3.5.8. *Suppose the random vectors $\underline{X}_n$, $1 \le n \le \infty$ on $\mathbb{R}^d$ are such that $\Phi_{\underline{X}_n}(\underline{\theta}) \to \Phi_{\underline{X}_\infty}(\underline{\theta})$ as $n \to \infty$ for each $\underline{\theta} \in \mathbb{R}^d$. Then, the corresponding sequence of laws $\{\mathcal{P}_{\underline{X}_n}\}$ is uniformly tight.*

PROOF. Fixing $\underline{\theta} \in \mathbb{R}^d$ consider the sequence of random variables $Y_n = (\underline{\theta}, \underline{X}_n)$. Since $\Phi_{Y_n}(\alpha) = \Phi_{\underline{X}_n}(\alpha\underline{\theta})$ for $1 \le n \le \infty$, we have that $\Phi_{Y_n}(\alpha) \to \Phi_{Y_\infty}(\alpha)$ for all $\alpha \in \mathbb{R}$. The uniform tightness of the laws of Y_n then follows by Lemma 3.3.16. Considering $\underline{\theta}_1, \dots, \underline{\theta}_d$ which are the unit vectors in the d different coordinates, we have the uniform tightness of the laws of $X_{n,j}$ for the sequence of random vectors $\underline{X}_n = (X_{n,1}, X_{n,2}, \dots, X_{n,d})$ and each fixed coordinate $j = 1, \dots, d$. For the compact sets $K_\varepsilon = [-M_\varepsilon, M_\varepsilon]^d$ and all n,

$$\mathbf{P}(\underline{X}_n \notin K_\varepsilon) \le \sum_{j=1}^{d} \mathbf{P}(|X_{n,j}| > M_\varepsilon)\,.$$

As d is finite, this leads from the uniform tightness of the laws of $X_{n,j}$ for each $j = 1, \dots, d$ to the uniform tightness of the laws of $\underline{X}_n$. □

Equipped with Lemma 3.5.8 we are ready to state and prove Lévy's continuity theorem.

Theorem 3.5.9 (LÉVY'S CONTINUITY THEOREM)**.** *Let $\underline{X}_n$, $1 \le n \le \infty$ be random vectors with characteristic functions $\Phi_{\underline{X}_n}(\underline{\theta})$. Then, $\underline{X}_n \xrightarrow{\mathcal{D}} \underline{X}_\infty$ if and only if $\Phi_{\underline{X}_n}(\underline{\theta}) \to \Phi_{\underline{X}_\infty}(\underline{\theta})$ as $n \to \infty$ for each fixed $\underline{\theta} \in \mathbb{R}^d$.*

PROOF. This is a re-run of the proof of Theorem 3.3.17, adapted to $\mathbb{R}^d$-valued random variables. First, both $\underline{x} \mapsto \cos((\underline{\theta}, \underline{x}))$ and $\underline{x} \mapsto \sin((\underline{\theta}, \underline{x}))$ are bounded continuous functions, so if $\underline{X}_n \xrightarrow{\mathcal{D}} \underline{X}_\infty$, then clearly as $n \to \infty$,

$$\Phi_{\underline{X}_n}(\underline{\theta}) = \mathbf{E}\Big[e^{i(\underline{\theta}, \underline{X}_n)}\Big] \to \mathbf{E}\Big[e^{i(\underline{\theta}, \underline{X}_\infty)}\Big] = \Phi_{\underline{X}_\infty}(\underline{\theta}).$$

For the converse direction, assuming that $\Phi_{\underline{X}_n} \to \Phi_{\underline{X}_\infty}$ point-wise, we know from Lemma 3.5.8 that the collection $\{\mathcal{P}_{\underline{X}_n}\}$ is uniformly tight. Hence, by Prohorov's theorem, for every subsequence $n(m)$ there is a further sub-subsequence $n(m_k)$ such that $\mathcal{P}_{\underline{X}_{n(m_k)}}$ converges weakly to some probability measure $\mathcal{P}_{\underline{Y}}$, possibly dependent upon the choice of $n(m)$. As $\underline{X}_{n(m_k)} \xrightarrow{\mathcal{D}} \underline{Y}$, we have by the preceding part of the proof that $\Phi_{\underline{X}_{n(m_k)}} \to \Phi_{\underline{Y}}$, and necessarily $\Phi_{\underline{Y}} = \Phi_{\underline{X}_\infty}$. The characteristic function determines the law (see Theorem 3.5.7), so $\underline{Y} \stackrel{\mathcal{D}}{=} \underline{X}_\infty$ is independent of the choice of $n(m)$. Thus, fixing $h \in C_b(\mathbb{R}^d)$, the sequence $y_n = \mathbf{E}h(\underline{X}_n)$ is such that every subsequence $y_{n(m)}$ has a further sub-subsequence $y_{n(m_k)}$ that converges to y_∞. Consequently, $y_n \to y_\infty$ (see Lemma 2.2.11). This applies for all $h \in C_b(\mathbb{R}^d)$, so we conclude that $\underline{X}_n \xrightarrow{\mathcal{D}} \underline{X}_\infty$, as stated. □

Remark. As in the case of Theorem 3.3.17, it is not hard to show that if $\Phi_{\underline{X}_n}(\underline{\theta}) \to \Phi(\underline{\theta})$ as $n \to \infty$ and $\Phi(\underline{\theta})$ is continuous at $\underline{\theta} = \underline{0}$ then Φ is necessarily the characteristic function of some random vector $\underline{X}_\infty$ and consequently $\underline{X}_n \xrightarrow{\mathcal{D}} \underline{X}_\infty$.

The proof of the multivariate CLT is just one of the results that rely on the following immediate corollary of Lévy's continuity theorem.

Corollary 3.5.10 (Cramér-Wold device). *A sufficient condition for $\underline{X}_n \xrightarrow{\mathcal{D}} \underline{X}_\infty$ is that $(\underline{\theta}, \underline{X}_n) \xrightarrow{\mathcal{D}} (\underline{\theta}, \underline{X}_\infty)$ for each $\underline{\theta} \in \mathbb{R}^d$.*

Proof. Since $(\underline{\theta}, \underline{X}_n) \xrightarrow{\mathcal{D}} (\underline{\theta}, \underline{X}_\infty)$ it follows by Lévy's continuity theorem (for $d = 1$, that is, Theorem 3.3.17), that

$$\lim_{n\to\infty} \mathbf{E}\left[e^{i(\underline{\theta}, \underline{X}_n)}\right] = \mathbf{E}\left[e^{i(\underline{\theta}, \underline{X}_\infty)}\right].$$

As this applies for any $\underline{\theta} \in \mathbb{R}^d$, we get that $\underline{X}_n \xrightarrow{\mathcal{D}} \underline{X}_\infty$ by applying Lévy's continuity theorem in $\mathbb{R}^d$ (i.e., Theorem 3.5.9), now in the converse direction. □

Remark. Beware that it is not enough to consider only finitely many values of $\underline{\theta}$ in the Cramér-Wold device. For example, consider the random vectors $\underline{X}_n = (X_n, Y_n)$ with $\{X_n, Y_{2n}\}$ i.i.d. and $Y_{2n+1} = X_{2n+1}$. Convince yourself that in this case $X_n \xrightarrow{\mathcal{D}} X_1$ and $Y_n \xrightarrow{\mathcal{D}} Y_1$ but the random vectors $\underline{X}_n$ do not converge in distribution (to any limit).

The computation of the characteristic function is much simplified in the presence of independence.

Exercise 3.5.11. *Show that if $\underline{Y} = (Y_1, \ldots, Y_d)$ with Y_k mutually independent R.V., then for all $\underline{\theta} = (\theta_1, \ldots, \theta_d) \in \mathbb{R}^d$,*

$$\Phi_{\underline{Y}}(\underline{\theta}) = \prod_{k=1}^{d} \Phi_{Y_k}(\theta_k) \tag{3.5.3}$$

Conversely, show that if (3.5.3) holds for all $\underline{\theta} \in \mathbb{R}^d$, the random variables Y_k, $k = 1, \ldots, d$ are mutually independent of each other.

3.5.3. Gaussian random vectors and the multivariate CLT.

Recall the following linear algebra concept.

Definition 3.5.12. *An $d \times d$ matrix $\mathbf{A}$ with entries A_{jk} is called* non-negative definite *(or positive semidefinite) if $A_{jk} = A_{kj}$ for all j, k, and for any $\underline{\theta} \in \mathbb{R}^d$*

$$(\underline{\theta}, \mathbf{A}\underline{\theta}) = \sum_{j=1}^{d} \sum_{k=1}^{d} \theta_j A_{jk} \theta_k \geq 0.$$

We are ready to define the class of multivariate normal distributions via the corresponding characteristic functions.

Definition 3.5.13. *We say that a random vector $\underline{X} = (X_1, X_2, \cdots, X_d)$ is* Gaussian*, or alternatively that it has a* multivariate normal *distribution if*

$$\Phi_{\underline{X}}(\underline{\theta}) = e^{-\frac{1}{2}(\underline{\theta}, \mathbf{V}\underline{\theta})} e^{i(\underline{\theta}, \underline{\mu})}, \tag{3.5.4}$$

for some non-negative definite $d \times d$ matrix $\mathbf{V}$, some $\underline{\mu} = (\mu_1, \ldots, \mu_d) \in \mathbb{R}^d$ and all $\underline{\theta} = (\theta_1, \ldots, \theta_d) \in \mathbb{R}^d$. We denote such a law by $\mathcal{N}(\underline{\mu}, \mathbf{V})$.

Remark. For $d = 1$ this definition coincides with Example 3.3.6.

Our next proposition proves that the multivariate $\mathcal{N}(\underline{\mu}, \mathbf{V})$ distribution is well defined and further links the vector $\underline{\mu}$ and the matrix $\mathbf{V}$ to the first two moments of this distribution.

Proposition 3.5.14. *The formula (3.5.4) corresponds to the characteristic function of a probability measure on $\mathbb{R}^d$. Further, the parameters $\underline{\mu}$ and $\mathbf{V}$ of the Gaussian random vector $\underline{X}$ are merely $\mu_j = \mathbf{E}X_j$ and $V_{jk} = \mathsf{Cov}(X_j, X_k)$, $j, k = 1, \ldots, d$.*

Proof. Any non-negative definite matrix $\mathbf{V}$ can be written as $\mathbf{V} = \mathbf{U}^t\mathbf{D}^2\mathbf{U}$ for some orthogonal matrix $\mathbf{U}$ (i.e., such that $\mathbf{U}^t\mathbf{U} = \mathbf{I}$, the $d \times d$-dimensional identity matrix), and some diagonal matrix $\mathbf{D}$. Consequently,

$$(\underline{\theta}, \mathbf{V}\underline{\theta}) = (\mathbf{A}\underline{\theta}, \mathbf{A}\underline{\theta})$$

for $\mathbf{A} = \mathbf{DU}$ and all $\underline{\theta} \in \mathbb{R}^d$. We claim that (3.5.4) is the characteristic function of the random vector $\underline{X} = \mathbf{A}^t\underline{Y} + \underline{\mu}$, where $\underline{Y} = (Y_1, \ldots, Y_d)$ has i.i.d. coordinates Y_k, each of which has the standard normal distribution. Indeed, by Exercise 3.5.11 $\Phi_{\underline{Y}}(\underline{\theta}) = \exp(-\frac{1}{2}(\underline{\theta}, \underline{\theta}))$ is the product of the characteristic functions $\exp(-\theta_k^2/2)$ of the standard normal distribution (see Example 3.3.6), and by part (e) of Proposition 3.3.2, $\Phi_{\underline{X}}(\underline{\theta}) = \exp(i(\underline{\theta}, \underline{\mu}))\Phi_{\underline{Y}}(\mathbf{A}\underline{\theta})$, yielding the formula (3.5.4).

We have just shown that $\underline{X}$ has the $\mathcal{N}(\underline{\mu}, \mathbf{V})$ distribution if $\underline{X} = \mathbf{A}^t\underline{Y} + \underline{\mu}$ for a Gaussian random vector $\underline{Y}$ (whose distribution is $\mathcal{N}(\underline{0}, \mathbf{I})$), such that $\mathbf{E}Y_j = 0$ and $\mathsf{Cov}(Y_j, Y_k) = \mathbf{1}_{j=k}$ for $j, k = 1, \ldots, d$. It thus follows by linearity of the expectation and the bi-linearity of the covariance that $\mathbf{E}X_j = \mu_j$ and $\mathsf{Cov}(X_j, X_k) = [\mathbf{E}\mathbf{A}^t\underline{Y}(\mathbf{A}^t\underline{Y})^t]_{jk} = (\mathbf{A}^t\mathbf{I}\mathbf{A})_{jk} = V_{jk}$, as claimed. $\square$

Definition 3.5.13 allows for $\mathbf{V}$ that is non-invertible, so for example the constant random vector $\underline{X} = \underline{\mu}$ is considered a Gaussian random vector though it obviously does not have a density. The reason we make this choice is to have the collection of multivariate normal distributions closed with respect to L^2-convergence, as we prove below to be the case.

Proposition 3.5.15. *Suppose Gaussian random vectors $\underline{X}_n$ converge in L^2 to a random vector $\underline{X}_\infty$, that is, $\mathbf{E}[\|\underline{X}_n - \underline{X}_\infty\|^2] \to 0$ as $n \to \infty$. Then, $\underline{X}_\infty$ is a Gaussian random vector, whose parameters are the limits of the corresponding parameters of $\underline{X}_n$.*

Proof. Recall that the convergence in L^2 of $\underline{X}_n$ to $\underline{X}_\infty$ implies that $\underline{\mu}_n = \mathbf{E}\underline{X}_n$ converge to $\underline{\mu}_\infty = \mathbf{E}\underline{X}_\infty$ and the element-wise convergence of the covariance matrices $\mathbf{V}_n$ to the corresponding covariance matrix $\mathbf{V}_\infty$. Further, the L^2-convergence implies the corresponding convergence in probability and hence, by bounded convergence $\Phi_{\underline{X}_n}(\underline{\theta}) \to \Phi_{\underline{X}_\infty}(\underline{\theta})$ for each $\underline{\theta} \in \mathbb{R}^d$. Since $\Phi_{\underline{X}_n}(\underline{\theta}) = e^{-\frac{1}{2}(\underline{\theta}, \mathbf{V}_n\underline{\theta})}e^{i(\underline{\theta}, \underline{\mu}_n)}$, for any $n < \infty$, it follows that the same applies for $n = \infty$. It is a well known fact of linear algebra that the element-wise limit $\mathbf{V}_\infty$ of non-negative definite matrices $\mathbf{V}_n$ is necessarily also non-negative definite. In view of Definition 3.5.13, we see that the limit $\underline{X}_\infty$ is a Gaussian random vector, whose parameters are the limits of the corresponding parameters of $\underline{X}_n$. $\square$

One of the main reasons for the importance of the multivariate normal distribution is the following CLT (which is the multivariate extension of Proposition 3.1.2).

Theorem 3.5.16 (Multivariate clt). *Let* $\widehat{\underline{S}}_n = n^{-\frac{1}{2}} \sum_{k=1}^{n} (\underline{X}_k - \underline{\mu})$*, where* $\{\underline{X}_k\}$ *are i.i.d. random vectors with finite second moments and such that* $\underline{\mu} = \mathbf{E}\underline{X}_1$*. Then,* $\widehat{\underline{S}}_n \xrightarrow{\mathcal{D}} \underline{G}$*, with* $\underline{G}$ *having the* $\mathcal{N}(\underline{0}, \mathbf{V})$ *distribution and where* $\mathbf{V}$ *is the* $d \times d$*-dimensional covariance matrix of* $\underline{X}_1$*.*

Proof. Consider the i.i.d. random vectors $\underline{Y}_k = \underline{X}_k - \underline{\mu}$ each having also the covariance matrix $\mathbf{V}$. Fixing an arbitrary vector $\underline{\theta} \in \mathbb{R}^d$ we proceed to show that $(\underline{\theta}, \widehat{\underline{S}}_n) \xrightarrow{\mathcal{D}} (\underline{\theta}, \underline{G})$, which in view of the Cramér-Wold device completes the proof of the theorem. Indeed, note that $(\underline{\theta}, \widehat{\underline{S}}_n) = n^{-\frac{1}{2}} \sum_{k=1}^{n} Z_k$, where $Z_k = (\underline{\theta}, \underline{Y}_k)$ are i.i.d. $\mathbb{R}$-valued random variables, having zero mean and variance

$$v_{\underline{\theta}} = \mathsf{Var}(Z_1) = \mathbf{E}[(\underline{\theta}, \underline{Y}_1)^2] = (\underline{\theta}, \mathbf{E}[\underline{Y}_1 \underline{Y}_1^t]\, \underline{\theta}) = (\underline{\theta}, \mathbf{V}\underline{\theta})\,.$$

Observing that the CLT of Proposition 3.1.2 thus applies to $(\underline{\theta}, \widehat{\underline{S}}_n)$, it remains only to verify that the resulting limit distribution $\mathcal{N}(0, v_{\underline{\theta}})$ is indeed the law of $(\underline{\theta}, \underline{G})$. To this end note that by Definitions 3.5.4 and 3.5.13, for any $s \in \mathbb{R}$,

$$\Phi_{(\underline{\theta},\underline{G})}(s) = \Phi_{\underline{G}}(s\underline{\theta}) = e^{-\frac{1}{2}s^2(\underline{\theta},\mathbf{V}\underline{\theta})} = e^{-v_{\underline{\theta}}s^2/2}\,,$$

which is the characteristic function of the $\mathcal{N}(0, v_{\underline{\theta}})$ distribution (see Example 3.3.6). Since the characteristic function uniquely determines the law (see Corollary 3.3.14), we are done. □

Here is an explicit example for which the multivariate CLT applies.

Example 3.5.17. *The* simple random walk *on* $\mathbf{Z}^d$ *is* $\underline{S}_n = \sum_{k=1}^{n} \underline{X}_k$ *where* $\underline{X}$*,* $\underline{X}_k$ *are i.i.d. random vectors such that*

$$\mathbf{P}(\underline{X} = +e_i) = \mathbf{P}(\underline{X} = -e_i) = \frac{1}{2d} \qquad i = 1, \ldots, d,$$

and e_i *is the unit vector in the* i*-th direction,* $i = 1, \ldots, d$*. In this case* $\mathbf{E}\underline{X} = \underline{0}$ *and if* $i \neq j$ *then* $\mathbf{E}X_iX_j = 0$*, resulting with the covariance matrix* $\mathbf{V} = (1/d)\mathbf{I}$ *for the multivariate normal limit in distribution of* $n^{-1/2}\underline{S}_n$*.*

Building on Lindeberg's CLT for weighted sums of i.i.d. random variables, the following multivariate normal limit is the basis for the convergence of random walks to *Brownian motion*, to which Section 9.2 is devoted.

Exercise 3.5.18. *Suppose* $\{\xi_k\}$ *are i.i.d. with* $\mathbf{E}\xi_1 = 0$ *and* $\mathbf{E}\xi_1^2 = 1$*. Consider the random functions* $\widehat{S}_n(t) = n^{-1/2}S(nt)$ *where* $S(t) = \sum_{k=1}^{[t]} \xi_k + (t - [t])\xi_{[t]+1}$ *and* $[t]$ *denotes the integer part of* t*.*

(a) *Verify that Lindeberg's* CLT *applies for* $\widehat{S}_n = \sum_{k=1}^{n} a_{n,k}\xi_k$ *whenever the non-random* $\{a_{n,k}\}$ *are such that* $r_n = \max\{|a_{n,k}| : k = 1, \ldots, n\} \to 0$ *and* $v_n = \sum_{k=1}^{n} a_{n,k}^2 \to 1$*.*

(b) *Let* $c(s,t) = \min(s,t)$ *and fixing* $0 = t_0 \leq t_1 < \cdots < t_d$*, denote by* $\mathbf{C}$ *the* $d \times d$ *matrix of entries* $C_{jk} = c(t_j, t_k)$*. Show that for any* $\underline{\theta} \in \mathbb{R}^d$*,*

$$\sum_{r=1}^{d} (t_r - t_{r-1}) (\sum_{j=1}^{r} \theta_j)^2 = (\underline{\theta}, \mathbf{C}\underline{\theta})\,,$$

(c) *Using the Cramér-Wold device deduce that* $(\widehat{S}_n(t_1), \ldots, \widehat{S}_n(t_d)) \xrightarrow{\mathcal{D}} \underline{G}$ *with* $\underline{G}$ *having the* $\mathcal{N}(\underline{0}, \mathbf{C})$ *distribution.*

As we see in the next exercise, there is more to a Gaussian random vector than each coordinate having a normal distribution.

Exercise 3.5.19. *Suppose X_1 has a standard normal distribution and S is independent of X_1 and such that $\mathbf{P}(S = 1) = \mathbf{P}(S = -1) = 1/2$.*

(a) *Check that $X_2 = SX_1$ also has a standard normal distribution.*
(b) *Check that X_1 and X_2 are uncorrelated random variables, each having the standard normal distribution, while $\underline{X} = (X_1, X_2)$ is not a Gaussian random vector and where X_1 and X_2 are not independent variables.*

Motivated by the proof of Proposition 3.5.14 here is an important property of Gaussian random vectors which may also be considered to be an alternative to Definition 3.5.13.

Exercise 3.5.20. *A random vector $\underline{X}$ has the multivariate normal distribution if and only if $(\sum_{i=1}^d a_{ji} X_i, j = 1, \dots, m)$ is a Gaussian random vector for any non-random coefficients $a_{11}, a_{12}, \dots, a_{md} \in \mathbb{R}$.*

The classical definition of the multivariate normal density applies for a strict subset of the distributions we consider in Definition 3.5.13.

Definition 3.5.21. *We say that $\underline{X}$ has a* non-degenerate *multivariate normal distribution if the matrix $\mathbf{V}$ is invertible, or alternatively, when $\mathbf{V}$ is (strictly) positive definite matrix, that is $(\underline{\theta}, \mathbf{V}\underline{\theta}) > 0$ whenever $\underline{\theta} \neq \underline{0}$.*

We next relate the density of a random vector with its characteristic function, and provide the density for the non-degenerate multivariate normal distribution.

Exercise 3.5.22.

(a) *Show that if $\int_{\mathbb{R}^d} |\Phi_{\underline{X}}(\underline{\theta})| d\underline{\theta} < \infty$, then $\underline{X}$ has the bounded continuous probability density function*

$$f_{\underline{X}}(\underline{x}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{-i(\underline{\theta}, \underline{x})} \Phi_{\underline{X}}(\underline{\theta}) d\underline{\theta} \,. \tag{3.5.5}$$

(b) *Show that a random vector $\underline{X}$ with a non-degenerate multivariate normal distribution $\mathcal{N}(\underline{\mu}, \mathbf{V})$ has the probability density function*

$$f_{\underline{X}}(\underline{x}) = (2\pi)^{-d/2} (\det \mathbf{V})^{-1/2} \exp\Big(-\frac{1}{2} (\underline{x} - \underline{\mu}, \mathbf{V}^{-1}(\underline{x} - \underline{\mu})) \Big) \,.$$

Here is an application to the uniform distribution over the sphere in $\mathbb{R}^n$, as $n \to \infty$.

Exercise 3.5.23. *Suppose $\{Y_k\}$ are i.i.d. random variables with $\mathbf{E}Y_1^2 = 1$ and $\mathbf{E}Y_1 = 0$. Let $W_n = n^{-1} \sum_{k=1}^n Y_k^2$ and $X_{n,k} = Y_k / \sqrt{W_n}$ for $k = 1, \dots, n$.*

(a) *Noting that $W_n \stackrel{a.s.}{\to} 1$ deduce that $X_{n,1} \xrightarrow{\mathcal{D}} Y_1$.*
(b) *Show that $n^{-1/2} \sum_{k=1}^n X_{n,k} \xrightarrow{\mathcal{D}} G$ whose distribution is $\mathcal{N}(0,1)$.*
(c) *Show that if $\{Y_k\}$ are standard normal random variables, then the random vector $\underline{X}_n = (X_{n,1}, \dots, X_{n,n})$ has the uniform distribution over the surface of the sphere of radius $\sqrt{n}$ in $\mathbb{R}^n$ (i.e., the unique measure supported on this sphere and invariant under orthogonal transformations), and interpret the preceding results for this special case.*

We conclude the section with the following exercise, which is a *multivariate, Lindeberg's type* CLT.

Exercise 3.5.24. *Let $\underline{y}^t$ denotes the transpose of the vector $\underline{y} \in \mathbb{R}^d$ and $\|\underline{y}\|$ its Euclidean norm. The independent random vectors $\{\underline{Y}_k\}$ on $\mathbb{R}^d$ are such that $\underline{Y}_k \stackrel{\mathcal{D}}{=} -\underline{Y}_k$,*

$$\lim_{n\to\infty} \sum_{k=1}^{n} \mathbf{P}(\|\underline{Y}_k\| > \sqrt{n}) = 0,$$

and for some symmetric, (strictly) positive definite matrix $\mathbf{V}$ and any fixed $\varepsilon \in (0,1]$,

$$\lim_{n\to\infty} n^{-1} \sum_{k=1}^{n} \mathbf{E}(\underline{Y}_k \underline{Y}_k^t I_{\|\underline{Y}_k\| \le \varepsilon\sqrt{n}}) = \mathbf{V}.$$

(a) *Let $\underline{T}_n = \sum_{k=1}^n \underline{X}_{n,k}$ for $\underline{X}_{n,k} = n^{-1/2}\underline{Y}_k I_{\|\underline{Y}_k\| \le \sqrt{n}}$. Show that $\underline{T}_n \xrightarrow{\mathcal{D}} \underline{G}$, with $\underline{G}$ having the $\mathcal{N}(\underline{0}, \mathbf{V})$ multivariate normal distribution.*
(b) *Let $\underline{\widehat{S}}_n = n^{-1/2}\sum_{k=1}^n \underline{Y}_k$ and show that $\underline{\widehat{S}}_n \xrightarrow{\mathcal{D}} \underline{G}$.*
(c) *Show that $(\underline{\widehat{S}}_n)^t \mathbf{V}^{-1} \underline{\widehat{S}}_n \xrightarrow{\mathcal{D}} Z$ and identify the law of Z.*

CHAPTER 4

Conditional expectations and probabilities

The most important concept in probability theory is the conditional expectation to which this chapter is devoted. In contrast with the elementary definition often used for a finite or countable sample space, the conditional expectation, as defined in Section 4.1, is itself a random variable. Section 4.2 details the important properties of the conditional expectation. Section 4.3 provides a representation of the conditional expectation as an orthogonal projection in Hilbert space. Finally, in Section 4.4 we represent the conditional expectation also as the expectation with respect to the *random* regular conditional probability distribution.

4.1. Conditional expectation: existence and uniqueness

In Subsection 4.1.1 we review the elementary definition of the conditional expectation $\mathbf{E}(X|Y)$ in case of discrete valued R.V.-s X and Y. This motivates our formal definition of the conditional expectation for any pair of R.V.s. such that X is integrable. The existence and uniqueness of the conditional expectation is shown there based on the Radon-Nikodym theorem, the proof of which we provide in Subsection 4.1.2.

4.1.1. Conditional expectation: motivation and definition. Suppose the R.V.s X and Z on a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ are both simple functions. More precisely, let X take the distinct values $x_1, \ldots, x_m \in \mathbb{R}$ and Z take the distinct values $z_1, \ldots, z_n \in \mathbb{R}$, where without loss of generality we assume that $\mathbf{P}(Z = z_i) > 0$ for $i = 1, \ldots, n$. Then, from elementary probability theory, we know that for any $i = 1, \ldots, n$, $j = 1, \ldots, m$,

$$\mathbf{P}(X = x_j | Z = z_i) = \frac{\mathbf{P}(X = x_j,\ Z = z_i)}{\mathbf{P}(Z = z_i)},$$

and we can compute the corresponding conditional expectation

$$\mathbf{E}[X|Z = z_i] = \sum_{j=1}^{m} x_j \mathbf{P}(X = x_j | Z = z_i).$$

Noting that this conditional expectation is a function of $\omega \in \Omega$ (via the value of $Z(\omega)$), we define the R.V. $Y = \mathbf{E}[X|Z]$ on the same probability space such that $Y(\omega) = \mathbf{E}[X|Z = z_i]$ whenever ω is such that $Z(\omega) = z_i$.

Example 4.1.1. *Suppose that $X = \omega_1$ and $Z = \omega_2$ on the probability space $\mathcal{F} = 2^{\Omega}$, $\Omega = \{1, 2\}^2$ with*

$$\mathbf{P}(1, 1) = .5, \ \ \mathbf{P}(1, 2) = .1, \ \ \mathbf{P}(2, 1) = .1, \ \ \mathbf{P}(2, 2) = .3.$$

Then,

$$\mathbf{P}(X=1|Z=1)=\frac{\mathbf{P}(X=1,Z=1)}{\mathbf{P}(Z=1)}=\frac{5}{6},$$

implying that $\mathbf{P}(X=2|Z=1)=\frac{1}{6}$ *and*

$$\mathbf{E}[X|Z=1]=1\cdot\frac{5}{6}+2\cdot\frac{1}{6}=\frac{7}{6}.$$

Likewise, check that $\mathbf{E}[X|Z=2]=\frac{7}{4}$, *hence* $\mathbf{E}[X|Z]=\frac{7}{6}I_{Z=1}+\frac{7}{4}I_{Z=2}$.

Partitioning Ω into the discrete collection of Z-atoms, namely the sets $G_i=\{\omega: Z(\omega)=z_i\}$ for $i=1,\ldots,n$, observe that $Y(\omega)$ is constant on each of these sets. The σ-algebra $\mathcal{G}=\mathcal{F}^Z=\sigma(Z)=\{Z^{-1}(B),\ B\in\mathcal{B}\}$ is in this setting merely the collection of all 2^n possible unions of various Z-atoms. Hence, $\mathcal{G}$ is finitely generated and since $Y(\omega)$ is constant on each generator G_i of $\mathcal{G}$, we see that $Y(\omega)$ is measurable on $(\Omega,\mathcal{G})$. Further, since any $G\in\mathcal{G}$ is of the form $G=\bigcup_{i\in\mathcal{I}}G_i$ for the disjoint sets G_i and some $\mathcal{I}\subseteq\{1,\ldots,n\}$, we find that

$$\begin{aligned}\mathbf{E}[YI_G]&=\sum_{i\in\mathcal{I}}\mathbf{E}[YI_{G_i}]=\sum_{i\in\mathcal{I}}\mathbf{E}[X|Z=z_i]\mathbf{P}(Z=z_i)\\&=\sum_{i\in\mathcal{I}}\sum_{j=1}^{m}x_j\mathbf{P}(X=x_j,\ Z=z_i)=\mathbf{E}[XI_G]\,.\end{aligned}$$

To summarize, in case X and Z are simple functions and $\mathcal{G}=\sigma(Z)$, we have $Y=\mathbf{E}[X|Z]$ as a R.V. on $(\Omega,\mathcal{G})$ such that $\mathbf{E}[YI_G]=\mathbf{E}[XI_G]$ for all $G\in\mathcal{G}$. Since both properties make sense for any σ-algebra $\mathcal{G}$ and any integrable R.V. X this suggests the definition of the conditional expectation as given by the following theorem.

Theorem 4.1.2. *Given* $X\in L^1(\Omega,\mathcal{F},\mathbf{P})$ *and* $\mathcal{G}\subseteq\mathcal{F}$ *a* σ*-algebra there exists a R.V.* Y *called the* conditional expectation *(C.E.) of* X *given* $\mathcal{G}$, *denoted by* $\mathbf{E}[X|\mathcal{G}]$, *such that* $Y\in L^1(\Omega,\mathcal{G},\mathbf{P})$ *and for any* $G\in\mathcal{G}$,

$$\mathbf{E}\ [(X-Y)\ I_G]=0\,. \tag{4.1.1}$$

Moreover, if (4.1.1) holds for any $G\in\mathcal{G}$ *and R.V.s* Y *and* $\widetilde{Y}$, *both of which are in* $L^1(\Omega,\mathcal{G},\mathbf{P})$, *then* $\mathbf{P}(\widetilde{Y}=Y)=1$. *In other words, the C.E. is uniquely defined for* $\mathbf{P}$*-almost every* ω.

Remark. We call $Y\in L^1(\Omega,\mathcal{G},\mathbf{P})$ that satisfies (4.1.1) for all $G\in\mathcal{G}$ a *version* of the C.E. $\mathbf{E}[X|\mathcal{G}]$. In view of the preceding theorem, unless stated otherwise we consider all versions of the C.E. as being the same R.V.

Given our motivation for Theorem 4.1.2, we let $\mathbf{E}[X|Z]$ stand for $\mathbf{E}[X|\mathcal{F}^Z]$ and likewise $\mathbf{E}[X|Z_1,Z_2,\ldots]$ stand for $\mathbf{E}[X|\mathcal{F}^{\mathbf{Z}}]$, where $\mathcal{F}^{\mathbf{Z}}=\sigma(Z_1,Z_2,\ldots)$.

To check whether a R.V. is a C.E. with respect to a given σ-algebra $\mathcal{G}$, it suffices to verify (4.1.1) for some π-system that contains Ω and generates $\mathcal{G}$, as you show in the following exercise. This useful general observation is often key to determining an explicit formula for the C.E.

Exercise 4.1.3. *Suppose that* $\mathcal{P}$ *is a* π*-system of subsets of* Ω *such that* $\Omega\in\mathcal{P}$ *and* $\mathcal{G}=\sigma(\mathcal{P})\subseteq\mathcal{F}$. *Show that if* $X\in L^1(\Omega,\mathcal{F},\mathbf{P})$ *and* $Y\in L^1(\Omega,\mathcal{G},\mathbf{P})$ *are such that* $\mathbf{E}[XI_G]=\mathbf{E}[YI_G]$ *for every* $G\in\mathcal{P}$ *then* $Y=\mathbf{E}[X|\mathcal{G}]$.

To prove the existence of the C.E. we need the following definition of absolute continuity of measures.

Definition 4.1.4. *Let ν and μ be two measures on measurable space $(\mathbb{S}, \mathcal{F})$. We say that ν is* absolutely continuous *with respect to μ, denoted by $\nu \ll \mu$, if*

$$\mu(A) = 0 \qquad \Longrightarrow \qquad \nu(A) = 0$$

for any set $A \in \mathcal{F}$.

Recall Proposition 1.3.56 that given a measure μ on $(\mathbb{S}, \mathcal{F})$, any $f \in m\mathcal{F}_+$ induces a new measure $f\mu$ on $(\mathbb{S}, \mathcal{F})$. The next theorem, whose proof is deferred to Subsection 4.1.2, shows that all absolutely continuous σ-finite measures with respect to a σ-finite measure μ are of this form.

Theorem 4.1.5 (Radon-Nikodym theorem)**.** *If ν and μ are two σ-finite measures on $(\mathbb{S}, \mathcal{F})$ such that $\nu \ll \mu$, then there exists $f \in m\mathcal{F}_+$ finite valued such that $\nu = f\mu$. Further, if $f\mu = g\mu$ then $\mu(\{s : f(s) \neq g(s)\}) = 0$.*

Remark. The assumption in Radon-Nikodym theorem that μ is a σ-finite measure can be somewhat relaxed, but not completely dispensed off.

Definition 4.1.6. *The function f such that $\nu = f\mu$ is called the* Radon-Nikodym derivative *(or density) of ν with respect to μ and denoted $f = \frac{d\nu}{d\mu}$.*

We note in passing that a real-valued R.V. has a probability density function f if and only if its law is absolutely continuous with respect to the completion $\overline{\lambda}$ of Lebesgue measure on $(\mathbb{R}, \overline{\mathcal{B}})$, with f being the corresponding Radon-Nikodym derivative (c.f. Example 1.3.60).

Proof of Theorem 4.1.2. Given two versions Y and $\widetilde{Y}$ of $\mathbf{E}[X|\mathcal{G}]$ we apply (4.1.1) for the set $G_\delta = \{\omega : Y(\omega) - \widetilde{Y}(\omega) > \delta\}$ to see that (by linearity of the expectation),

$$0 = \mathbf{E}\,[XI_{G_\delta}] - \mathbf{E}\,[XI_{G_\delta}] = \mathbf{E}\,[(Y - \widetilde{Y})\; I_{G_\delta}] \geq \delta \mathbf{P}(G_\delta)\,.$$

Hence, $\mathbf{P}(G_\delta) = 0$. Since this applies for any $\delta > 0$ and $G_\delta \uparrow G_0$ as $\delta \downarrow 0$, we deduce that $\mathbf{P}(Y - \widetilde{Y} > 0) = 0$. The same argument applies with the roles of Y and $\widetilde{Y}$ reversed, so $\mathbf{P}(Y - \widetilde{Y} = 0) = 1$ as claimed.

We turn to the existence of the C.E. assuming first that $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ is also non-negative. Let μ denote the probability measure obtained by restricting $\mathbf{P}$ to the measurable space $(\Omega, \mathcal{G})$ and ν denote the measure obtained by restricting $X\mathbf{P}$ of Proposition 1.3.56 to this measurable space, noting that ν is a finite measure (since $\nu(\Omega) = (X\mathbf{P})(\Omega) = \mathbf{E}[X] < \infty$). If $G \in \mathcal{G}$ is such that $\mu(G) = \mathbf{P}(G) = 0$, then by definition also $\nu(G) = (X\mathbf{P})(G) = 0$. Therefore, ν is absolutely continuous with respect to μ, and by the Radon-Nikodym theorem there exists $Y \in m\mathcal{G}_+$ such that $\nu = Y\mu$. This implies that for any $G \in \mathcal{G}$,

$$\mathbf{E}[XI_G] = \mathbf{P}(XI_G) = \nu(G) = (Y\mu)(G) = \mu(YI_G) = \mathbf{E}[YI_G]$$

(and in particular, that $\mathbf{E}[Y] = \nu(\Omega) < \infty$), proving the existence of the C.E. for non-negative R.V.s.

Turning to deal with the case of a general integrable R.V. X we use the representation $X = X_+ - X_-$ with $X_+ \geq 0$ and $X_- \geq 0$ such that both $\mathbf{E}[X_+]$ and $\mathbf{E}[X_-]$ are finite. Set $Y = Y^+ - Y^-$ where the integrable, non-negative R.V.s $Y^\pm = \mathbf{E}[X_\pm|\mathcal{G}]$

exist by the preceding argument. Then, $Y \in m\mathcal{G}$ is integrable, and by definition of $Y^{\pm}$ we have that for any $G \in \mathcal{G}$

$$\mathbf{E}[YI_G] = \mathbf{E}[Y^+ I_G] - \mathbf{E}[Y^- I_G] = \mathbf{E}[X_+ I_G] - \mathbf{E}[X_- I_G] = \mathbf{E}[XI_G] .$$

This establishes (4.1.1) and completes the proof of the theorem. □

Remark. Beware that for $Y = \mathbf{E}[X|\mathcal{G}]$ often $Y_\pm \neq \mathbf{E}[X_\pm|\mathcal{G}]$ (for example, take the trivial $\mathcal{G} = \{\emptyset, \Omega\}$ and $\mathbf{P}(X = 1) = \mathbf{P}(X = -1) = 1/2$ for which $Y = 0$ while $\mathbf{E}[X_\pm|\mathcal{G}] = 1$).

Exercise 4.1.7. *Suppose either $\mathbf{E}(Y_k)_+$ is finite or $\mathbf{E}(Y_k)_-$ is finite for random variables Y_k, $k = 1, 2$ on $(\Omega, \mathcal{F}, \mathbf{P})$ such that $\mathbf{E}[Y_1 I_A] \leq \mathbf{E}[Y_2 I_A]$ for any $A \in \mathcal{F}$. Show that then $\mathbf{P}(Y_1 \leq Y_2) = 1$.*

In the next exercise you show that the Radon-Nikodym density preserves the product structure.

Exercise 4.1.8. *Suppose that $\nu_k \ll \mu_k$ are pairs of σ-finite measures on $(\mathbb{S}_k, \mathcal{F}_k)$ for $k = 1, \ldots, n$ with the corresponding Radon-Nikodym derivatives $f_k = d\nu_k/d\mu_k$.*

(a) *Show that the σ-finite product measure $\nu = \nu_1 \times \cdots \times \nu_n$ on the product space $(\mathbb{S}, \mathcal{F})$ is absolutely continuous with respect to the σ-finite measure $\mu = \mu_1 \times \cdots \times \mu_n$ on $(\mathbb{S}, \mathcal{F})$, with $d\nu/d\mu(\mathbf{s}) = \prod_{k=1}^n f_k(s_k)$ for $\mathbf{s} = (s_1, \ldots, s_n)$.*

(b) *Suppose μ and ν are probability measures on $\mathbb{S} = \{(s_1, \ldots, s_n) : s_k \in \mathbb{S}_k, k = 1, \ldots, n\}$. Show that $f_k(s_k)$, $k = 1, \ldots, n$, are both mutually μ-independent and mutually ν-independent.*

4.1.2. Proof of the Radon-Nikodym theorem. This section is devoted to proving the Radon-Nikodym theorem, which we have already used for establishing the existence of C.E. This is done by proving the more general Lebesgue decomposition, based on the following definition.

Definition 4.1.9. *Two measures μ_1 and μ_2 on the same measurable space $(\mathbb{S}, \mathcal{F})$ are* mutually singular *if there is a set $A \in \mathcal{F}$ such that $\mu_1(A) = 0$ and $\mu_2(A^c) = 0$. This is denoted by $\mu_1 \perp \mu_2$, and we sometimes state that μ_1 is singular with respect to μ_2, instead of μ_1 and μ_2 mutually singular.*

Equipped with the concept of mutually singular measures, we next state the Lebesgue decomposition and show that the Radon-Nikodym theorem is a direct consequence of this decomposition.

Theorem 4.1.10 (Lebesgue decomposition). *Suppose μ and ν are measures on the same measurable space $(\mathbb{S}, \mathcal{F})$ such that $\mu(\mathbb{S})$ and $\nu(\mathbb{S})$ are finite. Then, $\nu = \nu_{\mathrm{ac}} + \nu_s$ where the measure ν_s is singular with respect to μ and $\nu_{\mathrm{ac}} = f\mu$ for some $f \in m\mathcal{F}_+$. Further, such a decomposition of ν is unique (per given μ).*

Remark. To build your intuition, note that Lebesgue decomposition is quite explicit for σ-finite measures on a countable space $\mathbb{S}$ (with $\mathcal{F} = 2^{\mathbb{S}}$). Indeed, then ν_{ac} and ν_s are the restrictions of ν to the *support* $S_\mu = \{s \in \mathbb{S} : \mu(\{s\}) > 0\}$ of μ and its complement, respectively, with $f(s) = \nu(\{s\})/\mu(\{s\})$ for $s \in S_\mu$ the Radon-Nikodym derivative of ν_{ac} with respect to μ (see Exercise 1.2.48 for more on the support of a measure).

Proof of the Radon-Nikodym theorem. Assume first that $\nu(\mathbb{S})$ and $\mu(\mathbb{S})$ are finite. Let $\nu = \nu_{\rm ac} + \nu_s$ be the unique Lebesgue decomposition induced by μ. Then, by definition there exists a set $A \in \mathcal{F}$ such that $\nu_s(A^c) = \mu(A) = 0$. Further, our assumption that $\nu \ll \mu$ implies that $\nu_s(A) \le \nu(A) = 0$ as well, hence $\nu_s(\mathbb{S}) = 0$, i.e. $\nu = \nu_{\rm ac} = f\mu$ for some $f \in m\mathcal{F}_+$.

Next, in case ν and μ are σ-finite measures the sample space $\mathbb{S}$ is a countable union of disjoint sets $A_n \in \mathcal{F}$ such that both $\nu(A_n)$ and $\mu(A_n)$ are finite. Considering the measures $\nu_n = I_{A_n}\nu$ and $\mu_n = I_{A_n}\mu$ such that $\nu_n(\mathbb{S}) = \nu(A_n)$ and $\mu_n(\mathbb{S}) = \mu(A_n)$ are finite, our assumption that $\nu \ll \mu$ implies that $\nu_n \ll \mu_n$. Hence, by the preceding argument for each n there exists $f_n \in m\mathcal{F}_+$ such that $\nu_n = f_n\mu_n$. With $\nu = \sum_n \nu_n$ and $\nu_n = (f_n I_{A_n})\mu$ (by the composition relation of Proposition 1.3.56), it follows that $\nu = f\mu$ for $f = \sum_n f_n I_{A_n} \in m\mathcal{F}_+$ finite valued.

As for the uniqueness of the Radon-Nikodym derivative f, suppose that $f\mu = g\mu$ for some $g \in m\mathcal{F}_+$ and a σ-finite measure μ. Consider $E_n = D_n \bigcap \{s : g(s) - f(s) \ge 1/n,\, g(s) \le n\}$ and measurable $D_n \uparrow \mathbb{S}$ such that $\mu(D_n) < \infty$. Then, necessarily both $\mu(f I_{E_n})$ and $\mu(g I_{E_n})$ are finite with

$$n^{-1}\mu(E_n) \le \mu((g-f)I_{E_n}) = (g\mu)(E_n) - (f\mu)(E_n) = 0\,,$$

implying that $\mu(E_n) = 0$. Considering the union over $n = 1, 2, \ldots$ we deduce that $\mu(\{s : g(s) > f(s)\}) = 0$, and upon reversing the roles of f and g, also $\mu(\{s : g(s) < f(s)\}) = 0$. □

Remark. Following the same argument as in the preceding proof of the Radon-Nikodym theorem, one easily concludes that Lebesgue decomposition applies also for any two σ-finite measures ν and μ.

Our next lemma is the key to the proof of Lebesgue decomposition.

Lemma 4.1.11. *If the finite measures μ and ν on $(\mathbb{S}, \mathcal{F})$ are not mutually singular, then there exists $B \in \mathcal{F}$ and $\epsilon > 0$ such that $\mu(B) > 0$ and $\nu(A) \ge \epsilon\mu(A)$ for all $A \in \mathcal{F}_B$.*

The proof of this lemma is based on the Hahn-Jordan decomposition of a finite *signed measure* to its positive and negative parts (for a definition of a finite signed measure see the remark after Definition 1.1.2).

Theorem 4.1.12 (Hahn decomposition)**.** *For any finite signed measure $\nu : \mathcal{F} \mapsto \mathbb{R}$ there exists $D \in \mathcal{F}$ such that $\nu_+ = I_D\nu$ and $\nu_- = -I_{D^c}\nu$ are measures on $(\mathbb{S}, \mathcal{F})$.*

See [**Bil95**, Theorem 32.1] for a proof of the Hahn decomposition as stated here, or [**Dud89**, Theorem 5.6.1] for the same conclusion in case of a general, that is $[-\infty, \infty]$-valued signed measure, where uniqueness of the Hahn-Jordan decomposition of a signed measure as the difference between the mutually singular measures $\nu_\pm$ is also shown (see also [**Dur10**, Theorems A.4.3 and A.4.4]).

Remark. If $I_B\nu$ is a measure we call $B \in \mathcal{F}$ a *positive set* for the signed measure ν and if $-I_B\nu$ is a measure we say that $B \in \mathcal{F}$ is a *negative set* for ν. So, the Hahn decomposition provides a partition of $\mathbb{S}$ into a positive set (for ν) and a negative set (for ν).

Proof of Lemma 4.1.11. Let $A = \bigcup_n D_n$ where D_n, $n = 1, 2 \ldots$, is a positive set for the Hahn decomposition of the finite signed measure $\nu - n^{-1}\mu$. Since A^c is contained in the negative set D_n^c for $\nu - n^{-1}\mu$, it follows that $\nu(A^c) \le n^{-1}\mu(A^c)$.

Taking $n \to \infty$ we deduce that $\nu(A^c) = 0$. If $\mu(D_n) = 0$ for all n then $\mu(A) = 0$ and necessarily ν is singular with respect to μ, contradicting the assumptions of the lemma. Therefore, $\mu(D_n) > 0$ for some finite n. Taking $\epsilon = n^{-1}$ and $B = D_n$ results with the thesis of the lemma. □

Proof of Lebesgue decomposition. Our goal is to construct $f \in m\mathcal{F}_+$ such that the measure $\nu_s = \nu - f\mu$ is singular with respect to μ. Since necessarily $\nu_s(A) \geq 0$ for any $A \in \mathcal{F}$, such a function f must belong to

$$\mathcal{H} = \{h \in m\mathcal{F}_+ : \nu(A) \geq (h\mu)(A), \quad \text{for all } A \in \mathcal{F}\}.$$

Indeed, we take f to be an element of $\mathcal{H}$ for which $(f\mu)(\mathbb{S})$ is maximal. To show that such f exists note first that $\mathcal{H}$ is closed under non-decreasing passages to the limit (by monotone convergence). Further, if h and $\widetilde{h}$ are both in $\mathcal{H}$ then also $\max\{h, \widetilde{h}\} \in \mathcal{H}$ since with $\Gamma = \{s : h(s) > \widetilde{h}(s)\}$ we have that for any $A \in \mathcal{F}$,

$$\nu(A) = \nu(A \cap \Gamma) + \nu(A \cap \Gamma^c) \geq \mu(hI_{A\cap\Gamma}) + \mu(\widetilde{h}I_{A\cap\Gamma^c}) = \mu(\max\{h, \widetilde{h}\}I_A)\,.$$

That is, $\mathcal{H}$ is also closed under the formation of finite maxima and in particular, the function $\lim_n \max(h_1, \ldots, h_n)$ is in $\mathcal{H}$ for any $h_n \in \mathcal{H}$. Now let $\kappa = \sup\{(h\mu)(\mathbb{S}) : h \in \mathcal{H}\}$ noting that $\kappa \leq \nu(\mathbb{S})$ is finite. Choosing $h_n \in \mathcal{H}$ such that $(h_n\mu)(\mathbb{S}) \geq \kappa - n^{-1}$ results with $f = \lim_n \max(h_1, \ldots, h_n)$ in $\mathcal{H}$ such that $(f\mu)(\mathbb{S}) \geq \lim_n (h_n\mu)(\mathbb{S}) = \kappa$. Since f is an element of $\mathcal{H}$ both $\nu_{\text{ac}} = f\mu$ and $\nu_s = \nu - f\mu$ are finite measures.

If ν_s fails to be singular with respect to μ then by Lemma 4.1.11 there exists $B \in \mathcal{F}$ and $\epsilon > 0$ such that $\mu(B) > 0$ and $\nu_s(A) \geq (\epsilon I_B\mu)(A)$ for all $A \in \mathcal{F}$. Since $\nu = \nu_s + f\mu$, this implies that $f + \epsilon I_B \in \mathcal{H}$. However, $((f + \epsilon I_B)\mu)(\mathbb{S}) \geq \kappa + \epsilon\mu(B) > \kappa$ contradicting the fact that κ is the finite maximal value of $(h\mu)(\mathbb{S})$ over $h \in \mathcal{H}$. Consequently, this construction of f has $\nu = f\mu + \nu_s$ with a finite measure ν_s that is singular with respect to μ.

Finally, to prove the uniqueness of the Lebesgue decomposition, suppose there exist $f_1, f_2 \in m\mathcal{F}_+$, such that both $\nu - f_1\mu$ and $\nu - f_2\mu$ are singular with respect to μ. That is, there exist $A_1, A_2 \in \mathcal{F}$ such that $\mu(A_i) = 0$ and $(\nu - f_i\mu)(A_i^c) = 0$ for $i = 1, 2$. Considering $A = A_1 \cup A_2$ it follows that $\mu(A) = 0$ and $(\nu - f_i\mu)(A^c) = 0$ for $i = 1, 2$. Consequently, for any $E \in \mathcal{F}$ we have that $(\nu - f_1\mu)(E) = \nu(E \cap A) = (\nu - f_2\mu)(E)$, proving the uniqueness of ν_s, and hence of the decomposition of ν as $\nu_{\text{ac}} + \nu_s$. □

We conclude with a simple application of Radon-Nikodym theorem in conjunction with Lemma 1.3.8.

Exercise 4.1.13. *Suppose ν and μ are two σ-finite measures on the same measurable space $(\mathbb{S}, \mathcal{F})$ such that $\nu(A) \leq \mu(A)$ for all $A \in \mathcal{F}$. Show that if $\nu(g) = \mu(g)$ is finite for some $g \in m\mathcal{F}$ such that $\mu(\{s : g(s) \leq 0\}) = 0$ then $\nu(\cdot) = \mu(\cdot)$.*

4.2. Properties of the conditional expectation

In some generic settings the C.E. is rather explicit. One such example is when X is measurable on the conditioning σ-algebra $\mathcal{G}$.

Example 4.2.1. *If $X \in L^1(\Omega, \mathcal{G}, \mathbf{P})$ then $Y = X \in m\mathcal{G}$ satisfies (4.1.1) so $\mathbf{E}[X|\mathcal{G}] = X$. In particular, if $X = c$ is a constant R.V. then $\mathbf{E}[X|\mathcal{G}] = c$ for any σ-algebra $\mathcal{G}$.*

Here is an extension of this example.

Exercise 4.2.2. *Suppose that* $(\mathbb{Y},\mathcal{Y})$*-valued random variable* Y *is measurable on* $\mathcal{G}$ *and* $(\mathbb{X},\mathfrak{X})$*-valued random variable* Z *is* $\mathbf{P}$*-independent of* $\mathcal{G}$*. Show that if* φ *is measurable on the product space* $(\mathbb{X}\times\mathbb{Y},\mathfrak{X}\times\mathcal{Y})$ *and* $\varphi(Z,Y)$ *is integrable, then* $\mathbf{E}[\varphi(Z,Y)|\mathcal{G}] = g(Y)$ *where* $g(y) = \mathbf{E}[\varphi(Z,y)]$*.*

Since only constant random variables are measurable on $\mathcal{F}_0 = \{\emptyset,\Omega\}$, by definition of the C.E. clearly $\mathbf{E}[X|\mathcal{F}_0] = \mathbf{E}X$. We show next that $\mathbf{E}[X|\mathcal{H}] = \mathbf{E}X$ also whenever the conditioning σ-algebra $\mathcal{H}$ is independent of $\sigma(X)$ (and in particular, when $\mathcal{H}$ is $\mathbf{P}$-trivial).

Proposition 4.2.3. *If* $X \in L^1(\Omega,\mathcal{F},\mathbf{P})$ *and the* σ*-algebra* $\mathcal{H}$ *is independent of* $\sigma(\sigma(X),\mathcal{G})$*, then*

$$\mathbf{E}[X|\sigma(\mathcal{H},\mathcal{G})] = \mathbf{E}[X\,|\,\mathcal{G}]\,.$$

For $\mathcal{G} = \{\emptyset,\Omega\}$ *this implies that*

$$\mathcal{H} \text{ independent of } \sigma(X) \qquad \Longrightarrow \qquad \mathbf{E}[X|\mathcal{H}] = \mathbf{E}X\,.$$

Remark. Recall that a $\mathbf{P}$-trivial σ-algebra $\mathcal{H} \subseteq \mathcal{F}$ is independent of $\sigma(X)$ for any $X \in m\mathcal{F}$. Hence, by Proposition 4.2.3 in this case $\mathbf{E}[X|\mathcal{H}] = \mathbf{E}X$ for all $X \in L^1(\Omega,\mathcal{F},P)$.

Proof. Let $Y = \mathbf{E}[X|\mathcal{G}] \in m\mathcal{G}$. Because $\mathcal{H}$ is independent of $\sigma(\mathcal{G},\sigma(X))$, it follows that for any $G \in \mathcal{G}$ and $H \in \mathcal{H}$ the random variable I_H is independent of both XI_G and YI_G. Consequently,

$$\mathbf{E}[XI_{G\cap H}] = \mathbf{E}[XI_GI_H] = \mathbf{E}[XI_G]\mathbf{E}I_H$$
$$\mathbf{E}[YI_{G\cap H}] = \mathbf{E}[YI_GI_H] = \mathbf{E}[YI_G]\mathbf{E}I_H$$

Further, $\mathbf{E}[XI_G] = \mathbf{E}[YI_G]$ by the definition of Y, hence $\mathbf{E}[XI_A] = \mathbf{E}[YI_A]$ for any $A \in \mathcal{A} = \{G\cap H : G \in \mathcal{G}, H \in \mathcal{H}\}$. Applying Exercise 4.1.3 with $Y \in L^1(\Omega,\mathcal{G},\mathbf{P}) \subseteq L^1(\Omega,\sigma(\mathcal{H},\mathcal{G}),\mathbf{P})$ and $\mathcal{A}$ a π-system of subsets containing Ω and generating $\sigma(\mathcal{G},\mathcal{H})$, we thus conclude that

$$\mathbf{E}[X|\sigma(\mathcal{G},\mathcal{H})] = Y = \mathbf{E}[X|\mathcal{G}]$$

as claimed. $\square$

We turn to derive various properties of the C.E. operation, starting with its positivity and linearity (per fixed conditioning σ-algebra).

Proposition 4.2.4. *Let* $X \in L^1(\Omega,\mathcal{F},\mathbf{P})$ *and set* $Y = \mathbf{E}[X|\mathcal{G}]$ *for some* σ*-algebra* $\mathcal{G} \subseteq \mathcal{F}$*. Then,*

(a) $\mathbf{E}X = \mathbf{E}Y$

(b) *(*Positivity*)* $X \geq 0 \implies Y \geq 0$ *a.s. and* $X > 0 \implies Y > 0$ *a.s.*

Proof. Considering $G = \Omega \in \mathcal{G}$ in the definition of the C.E. we find that $\mathbf{E}X = \mathbf{E}[XI_G] = \mathbf{E}[YI_G] = \mathbf{E}Y$.
Turning to the positivity of the C.E. note that if $X \geq 0$ a.s. then $0 \leq \mathbf{E}[XI_G] = \mathbf{E}[YI_G] \leq 0$ for $G = \{\omega : Y(\omega) \leq 0\} \in \mathcal{G}$. Hence, in this case $\mathbf{E}[YI_{Y\leq 0}] = 0$. That is, almost surely $Y \geq 0$. Further, $\delta\mathbf{P}(X > \delta, Y \leq 0) \leq \mathbf{E}[XI_{X>\delta}I_{Y\leq 0}] \leq \mathbf{E}[XI_{Y\leq 0}] = 0$ for any $\delta > 0$, so $\mathbf{P}(X > 0, Y = 0) = 0$ as well. $\square$

We next show that the C.E. operator is linear.

Proposition 4.2.5. *(*LINEARITY*) Let $X, Y \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ and $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra. Then, for any $\alpha, \beta \in \mathbb{R}$,*

$$\mathbf{E}[\alpha X + \beta Y|\mathcal{G}] = \alpha \mathbf{E}[X|\mathcal{G}] + \beta \mathbf{E}[Y|\mathcal{G}]\,.$$

PROOF. Let $Z = \mathbf{E}[X|\mathcal{G}]$ and $V = \mathbf{E}[Y|\mathcal{G}]$. Since $Z, V \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ the same applies for $\alpha Z + \beta V$. Further, for any $G \in \mathcal{G}$, by linearity of the expectation operator and the definition of the C.E.

$$\mathbf{E}[(\alpha Z + \beta V)I_G] = \alpha \mathbf{E}[Z I_G] + \beta \mathbf{E}[V I_G] = \alpha \mathbf{E}[X I_G] + \beta \mathbf{E}[Y I_G] = \mathbf{E}[(\alpha X + \beta Y)I_G],$$

as claimed. □

From its positivity and linearity we immediately get the monotonicity of the C.E.

Corollary 4.2.6 (MONOTONICITY)**.** *If $X, Y \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ are such that $X \le Y$, then $\mathbf{E}[X|\mathcal{G}] \le \mathbf{E}[Y|\mathcal{G}]$ for any σ-algebra $\mathcal{G} \subseteq \mathcal{F}$.*

In the following exercise you are to combine the linearity and positivity of the C.E. with Fubini's theorem.

Exercise 4.2.7. *Show that if $X, Y \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ are such that $\mathbf{E}[X|Y] = Y$ and $\mathbf{E}[Y|X] = X$ then almost surely $X = Y$.*
Hint: *First show that $\mathbf{E}[(X - Y)I_{\{X>c\ge Y\}}] = 0$ for any non-random c.*

We next deal with the relationship between the C.E.s of the same R.V. for nested conditioning σ-algebras.

Proposition 4.2.8 (TOWER PROPERTY)**.** *Suppose $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ and the σ-algebras $\mathcal{H}$ and $\mathcal{G}$ are such that $\mathcal{H} \subseteq \mathcal{G} \subseteq \mathcal{F}$. Then, $\mathbf{E}[X|\mathcal{H}] = \mathbf{E}[\mathbf{E}(X|\mathcal{G})|\mathcal{H}]$.*

PROOF. Recall that $Y = \mathbf{E}[X|\mathcal{G}]$ is integrable, hence $Z = \mathbf{E}[Y|\mathcal{H}]$ is integrable. Fixing $A \in \mathcal{H}$ we have that $\mathbf{E}[Y I_A] = \mathbf{E}[Z I_A]$ by the definition of the C.E. Z. Since $\mathcal{H} \subseteq \mathcal{G}$, also $A \in \mathcal{G}$ hence $\mathbf{E}[X I_A] = \mathbf{E}[Y I_A]$ by the definition of the C.E. Y. We deduce that $\mathbf{E}[X I_A] = \mathbf{E}[Z I_A]$ for all $A \in \mathcal{H}$. It then follows from the definition of the C.E. that Z is a version of $\mathbf{E}[X|\mathcal{H}]$. □

Remark. The tower property is also called *the law of iterated expectations*.

Any σ-algebra $\mathcal{G}$ contains the *trivial σ-algebra* $\mathcal{F}_0 = \{\emptyset, \Omega\}$. Applying the tower property with $\mathcal{H} = \mathcal{F}_0$ and using the fact that $\mathbf{E}[Y|\mathcal{F}_0] = \mathbf{E}Y$ for any integrable random variable Y, it follows that for any σ-algebra $\mathcal{G}$

$$\mathbf{E}X = \mathbf{E}[X|\mathcal{F}_0] = \mathbf{E}[\mathbf{E}(X|\mathcal{G})|\mathcal{F}_0] = \mathbf{E}[\mathbf{E}(X|\mathcal{G})]\,. \tag{4.2.1}$$

Here is an application of the tower property, leading to stronger conclusion than what one has from Proposition 4.2.3.

Lemma 4.2.9. *If integrable R.V. X and σ-algebra $\mathcal{G}$ are such that $\mathbf{E}[X|\mathcal{G}]$ is independent of X, then $\mathbf{E}[X|\mathcal{G}] = \mathbf{E}[X]$.*

PROOF. Let $Z = \mathbf{E}[X|\mathcal{G}]$. Applying the tower property for $\mathcal{H} = \sigma(Z) \subseteq \mathcal{G}$ we have that $\mathbf{E}[X|\mathcal{H}] = \mathbf{E}[Z|\mathcal{H}]$. Clearly, $\mathbf{E}[Z|\mathcal{H}] = Z$ (see Example 4.2.1), whereas our assumption that X is independent of Z implies that $\mathbf{E}[X|\mathcal{H}] = \mathbf{E}[X]$ (see Proposition 4.2.3). Consequently, $Z = \mathbf{E}[X]$, as claimed. □

As shown next, we can *take out what is known* when computing the C.E.

Proposition 4.2.10. *Suppose $Y \in m\mathcal{G}$ and $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ are such that $XY \in L^1(\Omega, \mathcal{F}, \mathbf{P})$. Then, $\mathbf{E}[XY|\mathcal{G}] = Y\mathbf{E}[X|\mathcal{G}]$.*

Proof. Let $Z = \mathbf{E}[X|\mathcal{G}]$ which is well defined due to our assumption that $\mathbf{E}|X| < \infty$. With $YZ \in m\mathcal{G}$ and $\mathbf{E}|XY| < \infty$, it suffices to check that

$$\mathbf{E}[XYI_A] = \mathbf{E}[ZYI_A] \tag{4.2.2}$$

for all $A \in \mathcal{G}$. Indeed, if $Y = I_B$ for $B \in \mathcal{G}$ then $YI_A = I_G$ for $G = B \cap A \in \mathcal{G}$ so (4.2.2) follows from the definition of the C.E. Z. By linearity of the expectation, this extends to Y which is a simple function on $(\Omega, \mathcal{G})$. Recall that for $X \geq 0$ by positivity of the C.E. also $Z \geq 0$, so by monotone convergence (4.2.2) then applies for all $Y \in m\mathcal{G}_+$. In general, let $X = X_+ - X_-$ and $Y = Y_+ - Y_-$ for $Y_\pm \in m\mathcal{G}_+$ and the integrable $X_\pm \geq 0$. Since $|XY| = (X_+ + X_-)(Y_+ + Y_-)$ is integrable, so are the products $X_\pm Y_\pm$ and (4.2.2) holds for each of the four possible choices of the pair $(X_\pm, Y_\pm)$, with $Z^\pm = \mathbf{E}[X_\pm|\mathcal{G}]$ instead of Z. Upon noting that $Z = Z^+ - Z^-$ (by linearity of the C.E.), and $XY = X_+Y_+ - X_+Y_- - X_-Y_+ + X_-Y_-$, it readily follows that (4.2.2) applies also for X and Y. $\square$

Adopting hereafter the notation $\mathbf{P}(A|\mathcal{G})$ for $\mathbf{E}[I_A|\mathcal{G}]$, the following exercises illustrate some of the many applications of Propositions 4.2.8 and 4.2.10.

Exercise 4.2.11. *For any σ-algebras $\mathcal{G}_i \subseteq \mathcal{F}$, $i = 1, 2, 3$, let $\mathcal{G}_{ij} = \sigma(\mathcal{G}_i, \mathcal{G}_j)$ and prove that the following conditions are equivalent:*

(a) $\mathbf{P}[A_3|\mathcal{G}_{12}] = \mathbf{P}[A_3|\mathcal{G}_2]$ *for all* $A_3 \in \mathcal{G}_3$.
(b) $\mathbf{P}[A_1 \cap A_3|\mathcal{G}_2] = \mathbf{P}[A_1|\mathcal{G}_2]\mathbf{P}[A_3|\mathcal{G}_2]$ *for all* $A_1 \in \mathcal{G}_1$ *and* $A_3 \in \mathcal{G}_3$.
(c) $\mathbf{P}[A_1|\mathcal{G}_{23}] = \mathbf{P}[A_1|\mathcal{G}_2]$ *for all* $A_1 \in \mathcal{G}_1$.

Remark. Taking $\mathcal{G}_1 = \sigma(X_k, k < n)$, $\mathcal{G}_2 = \sigma(X_n)$ and $\mathcal{G}_3 = \sigma(X_k, k > n)$, condition (a) of the preceding exercise states that the sequence of random variables $\{X_k\}$ has the *Markov property*. That is, the conditional probability of a future event A_3 given the past and present information $\mathcal{G}_{12}$ is the same as its conditional probability given the present $\mathcal{G}_2$ alone. Condition (c) makes the same statement, but with time reversed, while condition (b) says that past and future events A_1 and A_3 are conditionally independent given the present information, that is, $\mathcal{G}_2$.

Exercise 4.2.12. *Let $Z = (X, Y)$ be a uniformly chosen point in $(0, 1)^2$. That is, X and Y are independent random variables, each having the $U(0, 1)$ measure of Example 1.1.26. Set $T = 2I_A(Z) + 10I_B(Z) + 4I_C(Z)$ where $A = \{(x, y) : 0 < x < 1/4, 3/4 < y < 1\}$, $B = \{(x, y) : 1/4 < x < 3/4, 0 < y < 1/2\}$ and $C = \{(x, y) : 3/4 < x < 1, 1/4 < y < 1\}$.*

(a) *Find an explicit formula for the conditional expectation $W = \mathbf{E}(T|X)$ and use it to determine the conditional expectation $U = \mathbf{E}(TX|X)$.*
(b) *Find the value of $\mathbf{E}[(T - W)\sin(e^X)]$.*

Exercise 4.2.13. *Fixing a positive integer k, compute $\mathbf{E}(X|Y)$ in case $Y = kX - [kX]$ for X having the $U(0, 1)$ measure of Example 1.1.26 (and where $[x]$ denotes the integer part of x).*

Exercise 4.2.14. *Fixing $t \in \mathbb{R}$ and X integrable random variable, let $Y = \max(X, t)$ and $Z = \min(X, t)$. Setting $a_t = \mathbf{E}[X|X \leq t]$ and $b_t = \mathbf{E}[X|X \geq t]$, show that $\mathbf{E}[X|Y] = YI_{Y>t} + a_tI_{Y=t}$ and $\mathbf{E}[X|Z] = ZI_{Z<t} + b_tI_{Z=t}$.*

Exercise 4.2.15. *Let X, Y be i.i.d. random variables. Suppose θ is independent of (X, Y), with $\mathbf{P}(\theta = 1) = p$, $\mathbf{P}(\theta = 0) = 1 - p$. Let $Z = (Z_1, Z_2)$ where $Z_1 = \theta X + (1 - \theta)Y$ and $Z_2 = \theta Y + (1 - \theta)X$.*

(a) *Prove that Z and θ are independent.*
(b) *Obtain an explicit expression for $\mathbf{E}[g(X,Y)|Z]$, in terms of Z_1 and Z_2, where $g : \mathbb{R}^2 \mapsto \mathbb{R}$ is a bounded Borel function.*

Exercise 4.2.16. *Suppose $\mathbf{E}X^2 < \infty$ and define $\mathsf{Var}(X|\mathcal{G}) = \mathbf{E}[(X - \mathbf{E}(X|\mathcal{G}))^2|\mathcal{G}]$.*
(a) *Show that, $\mathbf{E}[\mathsf{Var}(X|\mathcal{G}_2)] \le \mathbf{E}[\mathsf{Var}(X|\mathcal{G}_1)]$ for any two σ-algebras $\mathcal{G}_1 \subseteq \mathcal{G}_2$ (that is, the dispersion of X about its conditional mean decreases as the $\sigma-$algebra grows).*
(b) *Show that for any σ-algebra $\mathcal{G}$,*

$$\mathsf{Var}[X] = \mathbf{E}[\mathsf{Var}(X|\mathcal{G})] + \mathsf{Var}[\mathbf{E}(X|\mathcal{G})] \,.$$

Exercise 4.2.17. *Suppose N is a non-negative, integer valued R.V. which is independent of the independent, integrable R.V.-s ξ_i on the same probability space, and that $\sum_i \mathbf{P}(N \ge i)\mathbf{E}|\xi_i|$ is finite.*
(a) *Check that*

$$X(\omega) = \sum_{i=1}^{N(\omega)} \xi_i(\omega) \,,$$

is integrable and deduce that $\mathbf{E}X = \sum_i \mathbf{P}(N \ge i)\mathbf{E}\xi_i$.
(b) *Suppose in addition that ξ_i are identically distributed, in which case this is merely* Wald's identity $\mathbf{E}X = \mathbf{E}N\mathbf{E}\xi_1$. *Show that if both ξ_1 and N are square-integrable, then so is X and*

$$\mathsf{Var}(X) = \mathsf{Var}(\xi_1)\mathbf{E}N + \mathsf{Var}(N)(\mathbf{E}\xi_1)^2 \,.$$

Suppose XY, X and Y are integrable. Combining Proposition 4.2.10 and (4.2.1) convince yourself that if $\mathbf{E}[X|Y] = \mathbf{E}X$ then $\mathbf{E}[XY] = \mathbf{E}X\mathbf{E}Y$. Recall that if X and Y are independent and integrable then $\mathbf{E}[X|Y] = \mathbf{E}X$ (c.f. Proposition 4.2.3). As you show next, the converse implications are false and further, one cannot dispense of the nesting relationship between the two σ-algebras in the tower property.

Exercise 4.2.18. *Provide examples of $X, Y \in \{-1, 0, 1\}$ such that*
(a) $\mathbf{E}[XY] = \mathbf{E}X\mathbf{E}Y$ *but* $\mathbf{E}[X|Y] \ne \mathbf{E}X$.
(b) $\mathbf{E}[X|Y] = \mathbf{E}X$ *but X is not independent of Y.*
(c) *For $\Omega = \{1, 2, 3\}$ and $\mathcal{F}_i = \sigma(\{i\})$, $i = 1, 2, 3$,*

$$\mathbf{E}[\mathbf{E}(X|\mathcal{F}_1)|\mathcal{F}_2] \ne \mathbf{E}[\mathbf{E}(X|\mathcal{F}_2)|\mathcal{F}_1].$$

As shown in the sequel, per fixed conditioning σ-algebra we can interpret the C.E. as an expectation in a different (conditional) probability space. Indeed, every property of the expectation has a corresponding extension to the C.E. For example, the extension of Jensen's inequality is

Proposition 4.2.19 (Jensen's inequality). *Suppose $g(\cdot)$ is a convex function on an open interval G of $\mathbb{R}$, that is,*

$$\lambda g(x) + (1-\lambda)g(y) \ge g(\lambda x + (1-\lambda)y) \qquad \forall\, x, y \in G, \qquad 0 \le \lambda \le 1.$$

If X is an integrable R.V. with $\mathbf{P}(X \in G) = 1$ and $g(X)$ is also integrable, then almost surely $\mathbf{E}[g(X)|\mathcal{H}] \ge g(\mathbf{E}[X|\mathcal{H}])$ for any σ-algebra $\mathcal{H}$.

Proof. Recall our derivation of (1.3.3) showing that

$$g(x) \ge g(c) + (D_-g)(c)(x - c) \qquad \forall c, x \in G$$

Further, with $(D_- g)(\cdot)$ a finite, non-decreasing function on G where $g(\cdot)$ is continuous, it follows that

$$g(x) = \sup_{c \in G \cap \mathbb{Q}} \{g(c) + (D_- g)(c)(x-c)\} = \sup_n \{a_n x + b_n\}$$

for some sequences $\{a_n\}$ and $\{b_n\}$ in $\mathbb{R}$ and all $x \in G$.

Since $\mathbf{P}(X \in G) = 1$, almost surely $g(X) \geq a_n X + b_n$ and by monotonicity of the C.E. also $\mathbf{E}[g(X)|\mathcal{H}] \geq a_n Y + b_n$ for $Y = \mathbf{E}[X|\mathcal{H}]$. Further, $\mathbf{P}(Y \in G) = 1$ due to the linearity and positivity of the C.E., so almost surely $\mathbf{E}[g(X)|\mathcal{H}] \geq \sup_n\{a_n Y + b_n\} = g(Y)$, as claimed. $\square$

Example 4.2.20. *Fixing $q \geq 1$ and applying (the conditional) Jensen's inequality for the convex function $g(x) = |x|^q$, we have that $\mathbf{E}[|X|^q|\mathcal{H}] \geq |\mathbf{E}[X|\mathcal{H}]|^q$ for any $X \in L^q(\Omega, \mathcal{F}, \mathbf{P})$. So, by the tower property and the monotonicity of the expectation,*

$$\begin{aligned}\|X\|_q^q &= \mathbf{E}|X|^q = \mathbf{E}[\mathbf{E}(|X|^q|\mathcal{H})] \\ &\geq \mathbf{E}[|\mathbf{E}(X|\mathcal{H})|^q] = \|\mathbf{E}(X|\mathcal{H})\|_q^q .\end{aligned}$$

In conclusion, $\|X\|_q \geq \|\mathbf{E}(X|\mathcal{H})\|_q$ for all $q \geq 1$.

Exercise 4.2.21. *Let $Z = \mathbf{E}[X|\mathcal{G}]$ for an integrable random variable X and a σ-algebra $\mathcal{G}$.*

(a) *Show that if $\mathbf{E}Z^2 = \mathbf{E}X^2 < \infty$ then $Z = X$ a.s.*

(b) *Suppose that $Z = \mathbf{E}[X|\mathcal{G}]$ has the same law as X. Show that then $Z = X$ a.s. even if $\mathbf{E}X^2 = \infty$.*

Hint: *Show that $\mathbf{E}[(|X| - X)I_A] = 0$ for $A = \{Z \geq 0\} \in \mathcal{G}$, so $X \geq 0$ for almost every $\omega \in A$. Applying this for $X - c$ with c non-random deduce that $\mathbf{P}(X < c \leq Z) = 0$ and conclude that $X \geq Z$ a.s.*

In the following exercises you are to derive the conditional versions of Markov's and Hölder's inequalities.

Exercise 4.2.22. *Suppose $p > 0$ is non-random and X is a random variable in $(\Omega, \mathcal{F}, \mathbf{P})$ with $\mathbf{E}|X|^p$ finite.*

(a) *Prove that for every σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, with probability one*

$$\mathbf{E}[|X|^p\,|\mathcal{G}] = \int_0^\infty p x^{p-1} \mathbf{P}(|X| > x\,|\mathcal{G}) dx .$$

(b) *Deduce the conditional version of Markov's inequality, that for any $a > 0$*

$$\mathbf{P}(|X| \geq a\,|\mathcal{G}) \leq a^{-p} \mathbf{E}[|X|^p\,|\mathcal{G}]$$

(compare with Lemma 1.4.31 and Example 1.3.14).

Exercise 4.2.23. *Suppose $\mathbf{E}|X|^p < \infty$ and $\mathbf{E}|Y|^q < \infty$ for some $p, q > 1$ such that $\frac{1}{p} + \frac{1}{q} = 1$. Prove the conditional Hölder's inequality*

$$\mathbf{E}[|XY|\,|\mathcal{G}] \leq (\mathbf{E}[|X|^p\,|\mathcal{G}])^{1/p} (\mathbf{E}[|Y|^q\,|\mathcal{G}])^{1/q}$$

(compare with Proposition 1.3.17).

Here are the corresponding extensions of some of the convergence theorems of Section 1.3.3.

Theorem 4.2.24 (Monotone convergence for C.E.). *If $0 \leq X_m \uparrow X_\infty$ a.s. and $\mathbf{E}X_\infty < \infty$, then $\mathbf{E}[X_m|\mathcal{G}] \uparrow \mathbf{E}[X_\infty|\mathcal{G}]$.*

PROOF. Let $Y_m = \mathbf{E}[X_m|\mathcal{G}] \in m\mathcal{G}_+$. By monotonicity of the C.E. we have that the sequence Y_m is a.s. non-decreasing, hence it has a limit $Y_\infty \in m\mathcal{G}_+$ (possibly infinite). We complete the proof by showing that $Y_\infty = \mathbf{E}[X_\infty|\mathcal{G}]$. Indeed, for any $G \in \mathcal{G}$,

$$\mathbf{E}[Y_\infty I_G] = \lim_{m\to\infty} \mathbf{E}[Y_m I_G] = \lim_{m\to\infty} \mathbf{E}[X_m I_G] = \mathbf{E}[X_\infty I_G],$$

where since $Y_m \uparrow Y_\infty$ and $X_m \uparrow X_\infty$ the first and third equalities follow by the monotone convergence theorem (the unconditional version), and the second equality from the definition of the C.E. Y_m. Considering $G = \Omega$ we see that Y_∞ is integrable. In conclusion, $\mathbf{E}[X_m|\mathcal{G}] = Y_m \uparrow Y_\infty = \mathbf{E}[X_\infty|\mathcal{G}]$, as claimed. □

Lemma 4.2.25 (FATOU'S LEMMA FOR C.E.). *If the non-negative, integrable X_n on same measurable space $(\Omega, \mathcal{F})$ are such that $\liminf_{n\to\infty} X_n$ is integrable, then for any σ-algebra $\mathcal{G} \subseteq \mathcal{F}$,*

$$\mathbf{E}\left(\liminf_{n\to\infty} X_n \middle|\, \mathcal{G}\right) \le \liminf_{n\to\infty} \mathbf{E}[X_n|\mathcal{G}] \quad \textit{a.s.}$$

PROOF. Applying the monotone convergence theorem for the C.E. of the non-decreasing sequence of non-negative R.V.s $Z_n = \inf\{X_k : k \ge n\}$ (whose limit is the integrable $\liminf_{n\to\infty} X_n$), results with

$$\mathbf{E}\left(\liminf_{n\to\infty} X_n|\mathcal{G}\right) = \mathbf{E}(\lim_{n\to\infty} Z_n|\mathcal{G}) = \lim_{n\to\infty} \mathbf{E}[Z_n|\mathcal{G}] \quad \text{a.s.} \tag{4.2.3}$$

Since $Z_n \le X_n$ it follows that $\mathbf{E}[Z_n|\mathcal{G}] \le \mathbf{E}[X_n|\mathcal{G}]$ for all n and

$$\lim_{n\to\infty} \mathbf{E}[Z_n|\mathcal{G}] = \liminf_{n\to\infty} \mathbf{E}[Z_n|\mathcal{G}] \le \liminf_{n\to\infty} \mathbf{E}[X_n|\mathcal{G}] \quad \text{a.s.} \tag{4.2.4}$$

Upon combining (4.2.3) and (4.2.4) we obtain the thesis of the lemma. □

Fatou's lemma leads to the C.E. version of the dominated convergence theorem.

Theorem 4.2.26 (DOMINATED CONVERGENCE FOR C.E.). *If $\sup_m |X_m|$ is integrable and $X_m \stackrel{a.s}{\to} X_\infty$, then $\mathbf{E}[X_m|\mathcal{G}] \stackrel{a.s.}{\to} \mathbf{E}[X_\infty|\mathcal{G}]$.*

PROOF. Let $Y = \sup_m |X_m|$ and $Z_m = Y - X_m \ge 0$. Applying Fatou's lemma for the C.E. of the non-negative, integrable R.V.s $Z_m \le 2Y$, we see that

$$\mathbf{E}\left(\liminf_{m\to\infty} Z_m \middle|\, \mathcal{G}\right) \le \liminf_{m\to\infty} \mathbf{E}[Z_m|\mathcal{G}] \quad \text{a.s.}$$

Since X_m converges, by the linearity of the C.E. and integrability of Y this is equivalent to

$$\mathbf{E}\left(\lim_{m\to\infty} X_m \middle|\, \mathcal{G}\right) \ge \limsup_{m\to\infty} \mathbf{E}[X_m|\mathcal{G}] \quad \text{a.s.}$$

Applying the same argument for the non-negative, integrable R.V.s $W_m = Y + X_m$ results with

$$\mathbf{E}\left(\lim_{m\to\infty} X_m \middle|\, \mathcal{G}\right) \le \liminf_{m\to\infty} \mathbf{E}[X_m|\mathcal{G}] \quad \text{a.s.}.$$

We thus conclude that a.s. the lim inf and lim sup of the sequence $\mathbf{E}[X_m|\mathcal{G}]$ coincide and are equal to $\mathbf{E}[X_\infty|\mathcal{G}]$, as stated. □

Exercise 4.2.27. *Let X_1, X_2 be random variables defined on same probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra. Prove that (a), (b) and (c) below are equivalent.*

(a) *For any Borel sets B_1 and B_2,*

$$\mathbf{P}(X_1 \in B_1, X_2 \in B_2|\mathcal{G}) = \mathbf{P}(X_1 \in B_1|\mathcal{G})\mathbf{P}(X_2 \in B_2|\mathcal{G}).$$

(b) *For any bounded Borel functions* h_1 *and* h_2,

$$\mathbf{E}[h_1(X_1)h_2(X_2)|\mathcal{G}] = \mathbf{E}[h_1(X_1)|\mathcal{G}]\mathbf{E}[h_2(X_2)|\mathcal{G}]\,.$$

(c) *For any bounded Borel function* h,

$$\mathbf{E}[h(X_1)|\sigma(\mathcal{G}, \sigma(X_2))] = \mathbf{E}[h(X_1)|\mathcal{G}]\,.$$

Definition 4.2.28. *If one of the equivalent conditions of Exercise 4.2.27 holds we say that* X_1 *and* X_2 *are* conditionally independent *given* $\mathcal{G}$.

Exercise 4.2.29. *Suppose that* X *and* Y *are conditionally independent given* $\sigma(Z)$ *and that* X *and* Z *are conditionally independent given* $\mathcal{F}$, *where* $\mathcal{F} \subseteq \sigma(Z)$. *Prove that then* X *and* Y *are conditionally independent given* $\mathcal{F}$.

Our next result shows that the C.E. operation is continuous with respect to L^q convergence.

Theorem 4.2.30. *Suppose* $X_n \xrightarrow{L^q} X_\infty$. *That is,* $X_n, X_\infty \in L^q(\Omega, \mathcal{F}, \mathbf{P})$ *are such that* $\mathbf{E}(|X_n - X_\infty|^q) \to 0$. *Then,* $\mathbf{E}[X_n|\mathcal{G}] \xrightarrow{L^q} \mathbf{E}[X_\infty|\mathcal{G}]$ *for any* σ*-algebra* $\mathcal{G} \subseteq \mathcal{F}$.

Proof. We saw already in Example 4.2.20 that $\mathbf{E}[X_n|\mathcal{G}]$ are in $L^q(\Omega, \mathcal{G}, \mathbf{P})$ for $n \le \infty$. Further, by the linearity of C.E., Jensen's Inequality for the convex function $|x|^q$ as in this example, and the tower property of (4.2.1),

$$\begin{aligned}\mathbf{E}[|\mathbf{E}(X_n|\mathcal{G}) - \mathbf{E}(X_\infty|\mathcal{G})|^q] &= \mathbf{E}[|\mathbf{E}(X_n - X_\infty|\mathcal{G})|^q] \\ &\le \mathbf{E}[\mathbf{E}(|X_n - X_\infty|^q|\mathcal{G})] = \mathbf{E}[|X_n - X_\infty|^q] \to 0,\end{aligned}$$

by our hypothesis, yielding the thesis of the theorem. □

As you will show, the C.E. operation is also continuous with respect to the following topology of weak L^q convergence.

Definition 4.2.31. *Let* $L^\infty(\Omega, \mathcal{F}, \mathbf{P})$ *denote the collection of all random variables on* $(\Omega, \mathcal{F})$ *which are* $\mathbf{P}$*-a.s. bounded, with* $\|Y\|_\infty$ *denoting the smallest non-random* K *such that* $\mathbf{P}(|Y| \le K) = 1$. *Setting* $p(q) : [1, \infty] \to [1, \infty]$ *via* $p(q) = q/(q-1)$, *we say that* X_n converges weakly in L^q *to* X_∞, *denoted* $X_n \xrightarrow{wL^q} X_\infty$, *if* $X_n, X_\infty \in L^q$ *and* $\mathbf{E}[(X_n - X_\infty)Y] \to 0$ *for each fixed* Y *such that* $\|Y\|_{p(q)}$ *is finite (compare with Definition 1.3.26).*

Exercise 4.2.32. *Show that* $\mathbf{E}[Y\mathbf{E}(X|\mathcal{G})] = \mathbf{E}[X\mathbf{E}(Y|\mathcal{G})]$ *for any* σ*-algebra* $\mathcal{G} \subseteq \mathcal{F}$, *provided that for some* $q \ge 1$ *and* $p = q/(q-1)$ *both* $\|X\|_q$ *and* $\|Y\|_p$ *are finite. Deduce that if* $X_n \xrightarrow{wL^q} X_\infty$ *then* $\mathbf{E}[X_n|\mathcal{G}] \xrightarrow{wL^q} \mathbf{E}[X_\infty|\mathcal{G}]$ *for any* σ*-algebra* $\mathcal{G} \subseteq \mathcal{F}$.

In view of Example 4.2.20 we already know that for each integrable random variable X the collection $\{\mathbf{E}[X|\mathcal{G}] : \mathcal{G} \subseteq \mathcal{F}$ is a σ-algebra$\}$ is a bounded in $L^1(\Omega, \mathcal{F}, \mathbf{P})$. As we show next, this collection is even *uniformly integrable* (U.I.), a key fact in our study of uniformly integrable martingales (see Subsection 5.3.1).

Proposition 4.2.33. *For any* $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$, *the collection* $\{\mathbf{E}[X|\mathcal{H}] : \mathcal{H} \subseteq \mathcal{F}$ *is a* σ*-algebra*$\}$ *is U.I.*

Proof. Fixing $\varepsilon > 0$, let $\delta = \delta(X, \varepsilon) > 0$ be as in part (b) of Exercise 1.3.43 and set the finite constant $M = \delta^{-1}\mathbf{E}|X|$. By Markov's inequality and Example 4.2.20 we get that $M\mathbf{P}(A) \le \mathbf{E}|Y| \le \mathbf{E}|X|$ for $A = \{|Y| \ge M\} \in \mathcal{H}$ and $Y = \mathbf{E}[X|\mathcal{H}]$. Hence, $\mathbf{P}(A) \le \delta$ by our choice of M, whereby our choice of δ results with $\mathbf{E}[|X|I_A] \leq \varepsilon$ (c.f. part (b) of Exercise 1.3.43). Further, by (the conditional) Jensen's inequality $|Y| \leq \mathbf{E}[|X|\,|\mathcal{H}]$ (see Example 4.2.20). Therefore, by definition of the C.E. $\mathbf{E}[|X|\,|\mathcal{H}]$,

$$\mathbf{E}[|Y|I_{|Y|>M}] \leq \mathbf{E}[|Y|I_A] \leq \mathbf{E}[\mathbf{E}[|X||\mathcal{H}]I_A] = \mathbf{E}[|X|I_A] \leq \varepsilon\,.$$

Since this applies for any σ-algebra $\mathcal{H} \subseteq \mathcal{F}$ and the value of $M = M(X,\varepsilon)$ does not depend on Y, we conclude that the collection of such $Y = \mathbf{E}[X|\mathcal{H}]$ is U.I. $\square$

To check your understanding of the preceding derivation, prove the following natural extension of Proposition 4.2.33.

Exercise 4.2.34. *Let $\mathcal{C}$ be a uniformly integrable collection of random variables on $(\Omega, \mathcal{F}, \mathbf{P})$. Show that the collection $\mathcal{D}$ of all R.V. Y such that $Y \overset{a.s.}{=} \mathbf{E}[X|\mathcal{H}]$ for some $X \in \mathcal{C}$ and σ-algebra $\mathcal{H} \subseteq \mathcal{F}$, is U.I.*

Here is a somewhat counter intuitive fact about the conditional expectation.

Exercise 4.2.35. *Suppose $Y_n \overset{a.s}{\to} Y_\infty$ in $(\Omega, \mathcal{F}, \mathbf{P})$ when $n \to \infty$ and $\{Y_n\}$ are uniformly integrable.*

(a) *Show that $\mathbf{E}[Y_n|\mathcal{G}] \overset{L^1}{\to} \mathbf{E}[Y_\infty|\mathcal{G}]$ for any σ-algebra $\mathcal{G} \subseteq \mathcal{F}$.*
(b) *Provide an example of such sequence $\{Y_n\}$ and a σ-algebra $\mathcal{G} \subset \mathcal{F}$ such that $\mathbf{E}[Y_n|\mathcal{G}]$ does not converge almost surely to $\mathbf{E}[Y_\infty|\mathcal{G}]$.*

4.3. The conditional expectation as an orthogonal projection

It readily follows from our next proposition that for $X \in L^2(\Omega, \mathcal{F}, \mathbf{P})$ and σ-algebras $\mathcal{G} \subseteq \mathcal{F}$ the C.E. $Y = \mathbf{E}[X|\mathcal{G}]$ is the unique $Y \in L^2(\Omega, \mathcal{G}, \mathbf{P})$ such that

$$\|X - Y\|_2 = \inf\{\|X - W\|_2 : W \in L^2(\Omega, \mathcal{G}, \mathbf{P})\}. \tag{4.3.1}$$

Proposition 4.3.1. *For any $X \in L^2(\Omega, \mathcal{F}, \mathbf{P})$ and σ-algebras $\mathcal{G} \subseteq \mathcal{F}$, a R.V. $Y \in L^2(\Omega, \mathcal{G}, \mathbf{P})$ is optimal in the sense of (4.3.1) if and only if it satisfies the orthogonality relations*

$$\mathbf{E}[(X - Y)Z] = 0 \qquad \textit{for all} \quad Z \in L^2(\Omega, \mathcal{G}, \mathbf{P})\,. \tag{4.3.2}$$

Further, any such R.V. Y is a version of $\mathbf{E}[X|\mathcal{G}]$.

Proof. If $Y \in L^2(\Omega, \mathcal{G}, \mathbf{P})$ satisfies (4.3.1) then considering $W = Y + \alpha Z$ it follows that for any $Z \in L^2(\Omega, \mathcal{G}, \mathbf{P})$ and $\alpha \in \mathbb{R}$,

$$0 \leq \|X - Y - \alpha Z\|_2^2 - \|X - Y\|_2^2 = \alpha^2 \mathbf{E} Z^2 - 2\alpha \mathbf{E}[(X - Y)Z]\,.$$

By elementary calculus, this inequality holds for all $\alpha \in \mathbb{R}$ if and only if $\mathbf{E}[(X - Y)Z] = 0$. Conversely, suppose $Y \in L^2(\Omega, \mathcal{G}, \mathbf{P})$ satisfies (4.3.2) and fix $W \in L^2(\Omega, \mathcal{G}, \mathbf{P})$. Then, considering (4.3.2) for $Z = W - Y$ we see that

$$\|X - W\|_2^2 = \|X - Y\|_2^2 - 2\mathbf{E}[(X - Y)(W - Y)] + \|W - Y\|_2^2 \geq \|X - Y\|_2^2\,,$$

so necessarily Y satisfies (4.3.1). Finally, since $I_G \in L^2(\Omega, \mathcal{G}, \mathbf{P})$ for any $G \in \mathcal{G}$, if Y satisfies (4.3.2) then it also satisfies the identity (4.1.1) which characterizes the C.E. $\mathbf{E}[X|\mathcal{G}]$. $\square$

Example 4.3.2. *If $\mathcal{G} = \sigma(A_1, \ldots, A_n)$ for finite n and disjoint sets A_i such that $\mathbf{P}(A_i) > 0$ for $i = 1, \ldots, n$, then $L^2(\Omega, \mathcal{G}, \mathbf{P})$ consists of all variables of the form $W = \sum_{i=1}^n v_i I_{A_i}$, $v_i \in \mathbb{R}$. A R.V. Y of this form satisfies (4.3.1) if and only if the corresponding $\{v_i\}$ minimizes*

$$\mathbf{E}[(X - \sum_{i=1}^n v_i I_{A_i})^2] - \mathbf{E}X^2 = \Big\{ \sum_{i=1}^n \mathbf{P}(A_i) v_i^2 - 2 \sum_{i=1}^n v_i \mathbf{E}[X I_{A_i}] \Big\} ,$$

which amounts to $v_i = \mathbf{E}[X I_{A_i}]/\mathbf{P}(A_i)$. In particular, if $Z = \sum_{i=1}^n z_i I_{A_i}$ for distinct z_i-s, then $\sigma(Z) = \mathcal{G}$ and we thus recover our first definition of the C.E.

$$\mathbf{E}[X|Z] = \sum_{i=1}^n \frac{\mathbf{E}[X I_{Z=z_i}]}{\mathbf{P}(Z = z_i)} I_{Z=z_i} .$$

As shown in the sequel, using (4.3.1) as an alternative characterization of the C.E. of $X \in L^2(\Omega, \mathcal{F}, \mathbf{P})$ we can prove the existence of the C.E. without invoking the Radon-Nikodym theorem. We start by defining the relevant concepts from the theory of Hilbert spaces on which this approach is based.

Definition 4.3.3. *A* linear vector space *is a set $\mathbb{H}$ that is closed under operations of addition and multiplication by (real-valued) scalars. That is, if $h_1, h_2 \in \mathbb{H}$ then $h_1 + h_2 \in \mathbb{H}$ and $\alpha h \in \mathbb{H}$ for all $\alpha \in \mathbb{R}$, where $\alpha(h_1 + h_2) = \alpha h_1 + \alpha h_2$, $(\alpha + \beta)h = \alpha h + \beta h$, $\alpha(\beta h) = (\alpha\beta)h$ and $1h = h$. A* normed vector space *is a linear vector space $\mathbb{H}$ equipped with a* norm $\|\cdot\|$. *That is, a non-negative function on $\mathbb{H}$ such that $\|\alpha h\| = |\alpha|\|h\|$ for all $\alpha \in \mathbb{R}$ and $d(h_1, h_2) = \|h_1 - h_2\|$ is a metric on $\mathbb{H}$.*

Definition 4.3.4. *A sequence $\{h_n\}$ in a normed vector space is called a* Cauchy sequence *if $\sup_{k,m\geq n} \|h_k - h_m\| \to 0$ as $n \to \infty$ and we say that $\{h_n\}$ converges to $h \in \mathbb{H}$ if $\|h_n - h\| \to 0$ as $n \to \infty$. A* Banach space *is a normed vector space in which every Cauchy sequence converges.*

Building on the preceding, we define the concept of inner product and the corresponding Hilbert spaces and sub-spaces.

Definition 4.3.5. *A* Hilbert space *is a Banach space $\mathbb{H}$ whose norm is of the form $(h,h)^{1/2}$ for a bi-linear, symmetric function $(h_1, h_2) : \mathbb{H} \times \mathbb{H} \mapsto \mathbb{R}$ such that $(h,h) \geq 0$ and we call such (h_1, h_2) an* inner product *for $\mathbb{H}$. A subset $\mathbb{K}$ of a Hilbert space which is closed under addition and under multiplication by a scalar is called a* Hilbert sub-space *if every Cauchy sequence $\{h_n\} \subseteq \mathbb{K}$ has a limit in $\mathbb{K}$.*

Here are two elementary properties of inner products we use in the sequel.

Exercise 4.3.6. *Let $\|h\| = (h,h)^{\frac{1}{2}}$ with (h_1, h_2) an inner product for a linear vector space $\mathbb{H}$. Show that* Schwarz inequality

$$(u, v)^2 \leq \|u\|^2 \|v\|^2 ,$$

and the parallelogram law $\|u+v\|^2 + \|u-v\|^2 = 2\|u\|^2 + 2\|v\|^2$ *hold for any $u, v \in \mathbb{H}$.*

Our next proposition shows that for each finite $q \geq 1$ the space $L^q(\Omega, \mathcal{F}, \mathbf{P})$ is a Banach space for the norm $\|\cdot\|_q$, the usual addition of R.V.s and the multiplication of a R.V. $X(\omega)$ by a non-random (scalar) constant. Further, $L^2(\Omega, \mathcal{G}, \mathbf{P})$ is a Hilbert sub-space of $L^2(\Omega, \mathcal{F}, \mathbf{P})$ for any σ-algebras $\mathcal{G} \subset \mathcal{F}$.

Proposition 4.3.7. *Upon identifying $\overline{\mathbb{R}}$-valued R.V. which are equal with probability one as being in the same equivalence class, for each $q \geq 1$ and a σ-algebra $\mathcal{F}$, the space $L^q(\Omega, \mathcal{F}, \mathbf{P})$ is a Banach space for the norm $\|\cdot\|_q$. Further, $L^2(\Omega, \mathcal{G}, \mathbf{P})$ is then a Hilbert sub-space of $L^2(\Omega, \mathcal{F}, \mathbf{P})$ for the inner product $(X, Y) = \mathbf{E}XY$ and any σ-algebras $\mathcal{G} \subseteq \mathcal{F}$.*

Proof. Fixing $q \geq 1$, we identify X and Y such that $\mathbf{P}(X \neq Y) = 0$ as being the same element of $L^q(\Omega, \mathcal{F}, \mathbf{P})$. The resulting set of equivalence classes is a normed vector space. Indeed, both $\|\cdot\|_q$, the addition of R.V. and the multiplication by a non-random scalar are compatible with this equivalence relation. Further, if $X, Y \in L^q(\Omega, \mathcal{F}, \mathbf{P})$ then $\|\alpha X\|_q = |\alpha|\|X\|_q < \infty$ for all $\alpha \in \mathbb{R}$ and by Minkowski's inequality $\|X + Y\|_q \leq \|X\|_q + \|Y\|_q < \infty$. Consequently, $L^q(\Omega, \mathcal{F}, \mathbf{P})$ is closed under the operations of addition and multiplication by a non-random scalar, with $\|\cdot\|_q$ a norm on this collection of equivalence classes.

Suppose next that $\{X_n\} \subseteq L^q$ is a Cauchy sequence for $\|\cdot\|_q$. Then, by definition, there exist $k_n \uparrow \infty$ such that $\|X_r - X_s\|_q^q < 2^{-n(q+1)}$ for all $r, s \geq k_n$. Observe that by Markov's inequality

$$\mathbf{P}(|X_{k_{n+1}} - X_{k_n}| \geq 2^{-n}) \leq 2^{nq}\|X_{k_{n+1}} - X_{k_n}\|_q^q < 2^{-n} \,,$$

and consequently the sequence $\mathbf{P}(|X_{k_{n+1}} - X_{k_n}| \geq 2^{-n})$ is summable. By Borel-Cantelli I it follows that $\sum_n |X_{k_{n+1}}(\omega) - X_{k_n}(\omega)|$ is finite with probability one, in which case clearly

$$X_{k_n} = X_{k_1} + \sum_{i=1}^{n-1}(X_{k_{i+1}} - X_{k_i})$$

converges to a finite limit $X(\omega)$. Next let, $X = \limsup_{n\to\infty} X_{k_n}$ (which per Theorem 1.2.22 is an $\overline{\mathbb{R}}$-valued R.V.). Then, fixing n and $r \geq k_n$, for any $t \geq n$,

$$\mathbf{E}\Big[|X_r - X_{k_t}|^q\Big] = \|X_r - X_{k_t}\|_q^q \leq 2^{-nq} \,,$$

so that by the a.s. convergence of X_{k_t} to X and Fatou's lemma

$$\mathbf{E}|X_r - X|^q = \mathbf{E}\Big[\lim_{t\to\infty} |X_r - X_{k_t}|^q\Big] \leq \liminf_{t\to\infty} \mathbf{E}|X_r - X_{k_t}|^q \leq 2^{-nq} \,.$$

This inequality implies that $X_r - X \in L^q$ and hence also $X \in L^q$. As $r \to \infty$ so does n and we can further deduce from the preceding inequality that $X_r \xrightarrow{L^q} X$.

Recall that $|\mathbf{E}XY| \leq \sqrt{\mathbf{E}X^2\mathbf{E}Y^2}$ by the Cauchy-Schwarz inequality. Thus, the bilinear, symmetric function $(X, Y) = \mathbf{E}XY$ on $L^2 \times L^2$ is real-valued and compatible with our equivalence relation. As $\|X\|_2^2 = (X, X)$, the Banach space $L^2(\Omega, \mathcal{F}, \mathbf{P})$ is a Hilbert space with respect to this inner product.

Finally, observe that for any σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ the subset $L^2(\Omega, \mathcal{G}, \mathbf{P})$ of the Hilbert space $L^2(\Omega, \mathcal{F}, \mathbf{P})$ is closed under addition of R.V.s and multiplication by a non-random constant. Further, as shown before, the L^2 limit of a Cauchy sequence $\{X_n\} \subseteq L^2(\Omega, \mathcal{G}, \mathbf{P})$ is $\limsup_n X_{k_n}$ which also belongs to $L^2(\Omega, \mathcal{G}, \mathbf{P})$. Hence, the latter is a Hilbert subspace of $L^2(\Omega, \mathcal{F}, \mathbf{P})$. □

Remark. With minor notational modifications, this proof shows that for any measure μ on $(\mathbb{S}, \mathcal{F})$ and $q \geq 1$ finite, the set $L^q(\mathbb{S}, \mathcal{F}, \mu)$ of μ-a.e. equivalence classes of $\overline{\mathbb{R}}$-valued, measurable functions f such that $\mu(|f|^q) < \infty$, is a Banach space. This is merely a special case of a general extension of this property, corresponding to $\mathbb{Y} = \overline{\mathbb{R}}$ in your next exercise.

Exercise 4.3.8. *For $q \geq 1$ finite and a given Banach space $(\mathbb{Y}, \|\cdot\|)$, consider the space $L^q(\mathbb{S}, \mathcal{F}, \mu; \mathbb{Y})$ of all μ-a.e. equivalence classes of functions $f : \mathbb{S} \mapsto \mathbb{Y}$, measurable with respect to the Borel σ-algebra induced on $\mathbb{Y}$ by $\|\cdot\|$ and such that $\mu(\|f(\cdot)\|^q) < \infty$.*

(a) *Show that $\|f\|_q = \mu(\|f(\cdot)\|^q)^{1/q}$ makes $L^q(\mathbb{S}, \mathcal{F}, \mu; \mathbb{Y})$ into a Banach space.*
(b) *For future applications of the preceding, verify that the space $\mathbb{Y} = C_b(\mathbb{T})$ of bounded, continuous real-valued functions on a topological space $\mathbb{T}$ is a Banach space for the supremum norm $\|f\| = \sup\{|f(t)| : t \in \mathbb{T}\}$.*

Your next exercise extends Proposition 4.3.7 to the collection $L^\infty(\Omega, \mathcal{F}, \mathbf{P})$ of all $\overline{\mathbb{R}}$-valued R.V. which are in equivalence classes of bounded random variables.

Exercise 4.3.9. *Fixing a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ prove the following facts:*

(a) *$L^\infty(\Omega, \mathcal{F}, \mathbf{P})$ is a Banach space for $\|X\|_\infty = \inf\{M : \mathbf{P}(|X| \leq M) = 1\}$.*
(b) *$\|X\|_q \uparrow \|X\|_\infty$ as $q \uparrow \infty$, for any $X \in L^\infty(\Omega, \mathcal{F}, \mathbf{P})$.*
(c) *If $\mathbf{E}[|X|^q] < \infty$ for some $q > 0$ then $\mathbf{E}|X|^q \to \mathbf{P}(|X| > 0)$ as $q \to 0$.*
(d) *The collection* SF *of simple functions is dense in $L^q(\Omega, \mathcal{F}, \mathbf{P})$ for any $1 \leq q \leq \infty$.*
(e) *The collection $C_b(\mathbb{R})$ of bounded, continuous real-valued functions, is dense in $L^q(\mathbb{R}, \mathcal{B}, \lambda)$ for any $q \geq 1$ finite.*
Hint: *The (bounded) monotone class theorem might be handy.*

In view of Proposition 4.3.7, the existence of the C.E. of $X \in L^2$ which satisfies (4.3.1), or the equivalent condition (4.3.2), is a special instance of the following fundamental geometric property of Hilbert spaces.

Theorem 4.3.10 (Orthogonal projection)**.** *Given $h \in \mathbb{H}$ and a Hilbert sub-space $\mathbb{G}$ of $\mathbb{H}$, let $d = \inf\{\|h - g\| : g \in \mathbb{G}\}$. Then, there exists a unique $\widehat{h} \in \mathbb{G}$, called the orthogonal projection of h on $\mathbb{G}$, such that $d = \|h - \widehat{h}\|$. This is also the unique $\widehat{h} \in \mathbb{G}$ such that $(h - \widehat{h}, f) = 0$ for all $f \in \mathbb{G}$.*

Proof. We start with the existence of $\widehat{h} \in \mathbb{G}$ such that $d = \|h - \widehat{h}\|$. To this end, let $g_n \in \mathbb{G}$ be such that $\|h - g_n\| \to d$. Applying the parallelogram law for $u = h - \frac{1}{2}(g_m + g_k)$ and $v = \frac{1}{2}(g_m - g_k)$ we find that

$$\|h-g_k\|^2+\|h-g_m\|^2 = 2\|h-\frac{1}{2}(g_m+g_k)\|^2+2\|\frac{1}{2}(g_m-g_k)\|^2 \geq 2d^2+\frac{1}{2}\|g_m-g_k\|^2$$

since $\frac{1}{2}(g_m + g_k) \in \mathbb{G}$. Taking $k, m \to \infty$, both $\|h - g_k\|^2$ and $\|h - g_m\|^2$ approach d^2 and hence by the preceding inequality $\|g_m - g_k\| \to 0$. In conclusion, $\{g_n\}$ is a Cauchy sequence in the Hilbert sub-space $\mathbb{G}$, which thus converges to some $\widehat{h} \in \mathbb{G}$. Recall that $\|h - \widehat{h}\| \geq d$ by the definition of d. Since for $n \to \infty$ both $\|h - g_n\| \to d$ and $\|g_n - \widehat{h}\| \to 0$, the converse inequality is a consequence of the triangle inequality $\|h - \widehat{h}\| \leq \|h - g_n\| + \|g_n - \widehat{h}\|$.

Next, suppose there exist $g_1, g_2 \in \mathbb{G}$ such that $(h - g_i, f) = 0$ for $i = 1, 2$ and all $f \in \mathbb{G}$. Then, by linearity of the inner product $(g_1 - g_2, f) = 0$ for all $f \in \mathbb{G}$. Considering $f = g_1 - g_2 \in \mathbb{G}$ we see that $(g_1 - g_2, g_1 - g_2) = \|g_1 - g_2\|^2 = 0$ so necessarily $g_1 = g_2$.

We complete the proof by showing that $\widehat{h} \in \mathbb{G}$ is such that $\|h - \widehat{h}\|^2 \leq \|h - g\|^2$ for all $g \in \mathbb{G}$ if and only if $(h - \widehat{h}, f) = 0$ for all $f \in \mathbb{G}$. This is done exactly as in the proof of Proposition 4.3.1. That is, by symmetry and bi-linearity of the inner product, for all $f \in \mathbb{G}$ and $\alpha \in \mathbb{R}$,

$$\|h - \widehat{h} - \alpha f\|^2 - \|h - \widehat{h}\|^2 = \alpha^2 \|f\|^2 - 2\alpha(h - \widehat{h}, f)$$

We arrive at the stated conclusion upon noting that fixing f, this function is non-negative for all α if and only if $(h - \widehat{h}, f) = 0$. $\square$

Applying Theorem 4.3.10 for the Hilbert subspace $\mathbb{G} = L^2(\Omega, \mathcal{G}, \mathbf{P})$ of $L^2(\Omega, \mathcal{F}, \mathbf{P})$ (see Proposition 4.3.7), you have the existence of a unique $Y \in \mathbb{G}$ satisfying (4.3.2) for each non-negative $X \in L^2$.

Exercise 4.3.11. *Show that for any non-negative integrable X, not necessarily in L^2, the sequence $Y_n \in \mathbb{G}$ corresponding to $X_n = \min(X, n)$ is non-decreasing and that its limit Y satisfies (4.1.1). Verify that this allows you to prove Theorem 4.1.2 without ever invoking the Radon-Nikodym theorem.*

Exercise 4.3.12. *Suppose $\mathcal{G} \subseteq \mathcal{F}$ is a σ-algebra.*

(a) *Show that for any $X \in L^1(\Omega, \mathcal{F}, \mathbf{P})$ there exists some $G \in \mathcal{G}$ such that*

$$\mathbf{E}[XI_G] = \sup_{A \in \mathcal{G}} \mathbf{E}[XI_A] \, .$$

Any G with this property is called $\mathcal{G}$-optimal for X.

(b) *Show that $Y = \mathbf{E}[X|\mathcal{G}]$ almost surely, if and only if for any $r \in \mathbb{R}$, the event $\{\omega : Y(\omega) > r\}$ is $\mathcal{G}$-optimal for the random variable $(X - r)$.*

Here is an alternative proof of the existence of $\mathbf{E}[X|\mathcal{G}]$ for non-negative $X \in L^2$ which avoids the orthogonal projection, as well as the Radon-Nikodym theorem (the general case then follows as in Exercise 4.3.11).

Exercise 4.3.13. *Suppose $X \in L^2(\Omega, \mathcal{F}, \mathbf{P})$ is non-negative. Assume first that the σ-algebra $\mathcal{G} \subseteq \mathcal{F}$ is countably generated. That is $\mathcal{G} = \sigma(B_1, B_2, \ldots)$ for some $B_k \in \mathcal{F}$.*

(a) *Let $Y_n = \mathbf{E}[X|\mathcal{G}_n]$ for the finitely generated $\mathcal{G}_n = \sigma(B_k, k \le n)$ (for its existence, see Example 4.3.2). Show that Y_n is a Cauchy sequence in $L^2(\Omega, \mathcal{G}, \mathbf{P})$, hence it has a limit $Y \in L^2(\Omega, \mathcal{G}, \mathbf{P})$.*

(b) *Show that $Y = \mathbf{E}[X|\mathcal{G}]$.*

Hint: *Theorem 2.2.10 might be of some help.*

Assume now that $\mathcal{G}$ is not countably generated.

(c) *Let $\mathcal{H}_1 \subseteq \mathcal{H}_2$ be finite σ-algebras. Show that*

$$\mathbf{E}[\mathbf{E}(X|\mathcal{H}_1)^2] \le \mathbf{E}[\mathbf{E}(X|\mathcal{H}_2)^2] \, .$$

(d) *Let $\alpha = \sup \mathbf{E}[\mathbf{E}(X|\mathcal{H})^2]$, where the supremum is over all finite σ-algebras $\mathcal{H} \subseteq \mathcal{G}$. Show that α is finite, and that there exists an increasing sequence of finite σ-algebras $\mathcal{H}_n$ such that $\mathbf{E}[\mathbf{E}(X|\mathcal{H}_n)^2] \uparrow \alpha$ as $n \to \infty$.*

(e) *Let $\mathcal{H}_\infty = \sigma(\cup_n \mathcal{H}_n)$ and $Y_n = \mathbf{E}[X|\mathcal{H}_n]$ for the $\mathcal{H}_n$ in part (d). Explain why your proof of part (b) implies the L^2 convergence of Y_n to a R.V. Y such that $\mathbf{E}[YI_A] = \mathbf{E}[XI_A]$ for any $A \in \mathcal{H}_\infty$.*

(f) *Fixing $A \in \mathcal{G}$ such that $A \notin \mathcal{H}_\infty$, let $\mathcal{H}_{n,A} = \sigma(A, \mathcal{H}_n)$ and $Z_n = \mathbf{E}[X|\mathcal{H}_{n,A}]$. Explain why some sub-sequence of $\{Z_n\}$ has an a.s. and L^2 limit, denoted Z. Show that $\mathbf{E}Z^2 = \mathbf{E}Y^2 = \alpha$ and deduce that $\mathbf{E}[(Y - Z)^2] = 0$, hence $Z = Y$ a.s.*

(g) *Show that Y is a version of the C.E. $\mathbf{E}[X|\mathcal{G}]$.*

4.4. Regular conditional probability distributions

We first show that if the random vector $(X, Z) \in \mathbb{R}^2$ has a probability density function $f_{X,Z}(x, z)$ (per Definition 3.5.5), then the C.E. $\mathbf{E}[X|Z]$ can be computed out of the corresponding conditional probability density (as done in a typical elementary probability course). To this end, let $f_Z(z) = \int_{\mathbb{R}} f_{X,Z}(x, z)dx$ and $f_X(x) = \int_{\mathbb{R}} f_{X,Z}(x, z)dz$ denote the probability density functions of Z and X. That is, $f_Z(z) = \lambda(f_{X,Z}(\cdot, z))$ and $f_X(x) = \lambda(f_{X,Z}(x, \cdot))$ for Lebesgue measure λ and the Borel function $f_{X,Z}$ on $\mathbb{R}^2$. Recall that $f_Z(\cdot)$ and $f_X(\cdot)$ are non-negative Borel functions (for example, consider our proof of Fubini's theorem in case of Lebesgue measure on $(\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2})$ and the non-negative integrable Borel function $h = f_{X,Z}$). So, defining the *conditional probability density function* of X given Z as

$$f_{X|Z}(x|z) = \begin{cases} \frac{f_{X,Z}(x,z)}{f_Z(z)} & \text{if } f_Z(z) > 0\,, \\ f_X(x) & \text{otherwise}\,, \end{cases}$$

guarantees that $f_{X|Z} : \mathbb{R}^2 \mapsto \mathbb{R}_+$ is Borel measurable and $\int_{\mathbb{R}} f_{X|Z}(x|z)dx = 1$ for all $z \in \mathbb{R}$.

Proposition 4.4.1. *Suppose the random vector (X, Z) has a probability density function $f_{X,Z}(x, z)$ and $g(\cdot)$ is a Borel function on $\mathbb{R}$ such that $\mathbf{E}|g(X)| < \infty$. Then, $\widehat{g}(Z)$ is a version of $\mathbf{E}[g(X)|Z]$ for the Borel function*

$$\widehat{g}(z) = \int_{\mathbb{R}} g(x) f_{X|Z}(x|z)dx\,, \tag{4.4.1}$$

in case $\int_{\mathbb{R}} |g(x)| f_{X,Z}(x, z)dx$ is finite (taking otherwise $\widehat{g}(z) = 0$).

Proof. Since the Borel function $h(x, z) = g(x)f_{X,Z}(x, z)$ is integrable with respect to Lebesgue measure on $(\mathbb{R}^2, \mathcal{B}_{\mathbb{R}^2})$, it follows that $\widehat{g}(\cdot)$ is also a Borel function (c.f. our proof of Fubini's theorem). Further, by Fubini's theorem the integrability of $g(X)$ implies that $\lambda(\mathbb{R} \setminus A) = 0$ for $A = \{z : \int |g(x)| f_{X,Z}(x, z)dx < \infty\}$, and with $\mathcal{P}_Z = f_Z\lambda$ this implies that $\mathbf{P}(Z \in A) = 1$. By Jensen's inequality,

$$|\widehat{g}(z)| \le \int |g(x)| f_{X|Z}(x|z)dx\,, \qquad \forall z \in A\,.$$

Thus, by Fubini's theorem and the definition of $f_{X|Z}$ we have that

$$\begin{aligned} \infty > \mathbf{E}|g(X)| &= \int |g(x)| f_X(x)dx \ge \int |g(x)| \Big[\int_A f_{X|Z}(x|z) f_Z(z)dz\Big]dx \\ &= \int_A \Big[\int |g(x)| f_{X|Z}(x|z)dx\Big] f_Z(z)dz \ge \int_A |\widehat{g}(z)| f_Z(z)dz = \mathbf{E}|\widehat{g}(Z)|\,. \end{aligned}$$

So, $\widehat{g}(Z)$ is integrable. With (4.4.1) holding for all $z \in A$ and $\mathbf{P}(Z \in A) = 1$, by Fubini's theorem and the definition of $f_{X|Z}$ we have that for any Borel set B,

$$\begin{aligned} \mathbf{E}[\widehat{g}(Z) I_B(Z)] &= \int_{B\cap A} \widehat{g}(z) f_Z(z)dz = \int \Big[\int g(x) f_{X|Z}(x|z)dx\Big] I_{B\cap A}(z) f_Z(z)dz \\ &= \int_{\mathbb{R}^2} g(x) I_{B\cap A}(z) f_{X,Z}(x, z)dxdz = \mathbf{E}[g(X) I_B(Z)]\,. \end{aligned}$$

This amounts to $\mathbf{E}[\widehat{g}(Z) I_G] = \mathbf{E}[g(X) I_G]$ for any $G \in \sigma(Z) = \{Z^{-1}(B) : B \in \mathcal{B}\}$ so indeed $\widehat{g}(Z)$ is a version of $\mathbf{E}[g(X)|Z]$. $\square$

To each conditional probability density $f_{X|Z}(\cdot|\cdot)$ corresponds the collection of conditional probability measures $\widehat{\mathbf{P}}_{X|Z}(B,\omega) = \int_B f_{X|Z}(x|Z(\omega))dx$. The remainder of this section deals with the following generalization of the latter object.

Definition 4.4.2. *Let $Y : \Omega \mapsto \mathbb{S}$ be an $(\mathbb{S}, \mathcal{S})$-valued R.V. in the probability space $(\Omega, \mathcal{F}, \mathbf{P})$, per Definition 1.2.1, and $\mathcal{G} \subseteq \mathcal{F}$ a σ-algebra. The collection $\widehat{\mathbf{P}}_{Y|\mathcal{G}}(\cdot,\cdot) : \mathcal{S} \times \Omega \mapsto [0,1]$ is called the* regular conditional probability distribution *(R.C.P.D.) of Y given $\mathcal{G}$ if:*

- (a) *$\widehat{\mathbf{P}}_{Y|\mathcal{G}}(A,\cdot)$ is a version of the C.E. $\mathbf{E}[I_{Y\in A}|\mathcal{G}]$ for each fixed $A \in \mathcal{S}$.*
- (b) *For any fixed $\omega \in \Omega$, the set function $\widehat{\mathbf{P}}_{Y|\mathcal{G}}(\cdot,\omega)$ is a probability measure on $(\mathbb{S}, \mathcal{S})$.*

In case $\mathbb{S} = \Omega$, $\mathcal{S} = \mathcal{F}$ and $Y(\omega) = \omega$, we call this collection the regular conditional probability *(R.C.P.) on $\mathcal{F}$ given $\mathcal{G}$, denoted also by $\widehat{\mathbf{P}}(A|\mathcal{G})(\omega)$.*

If the R.C.P. exists, then we can define all conditional expectations through the R.C.P. Unfortunately, the R.C.P. might not exist (see [**Bil95**, Exercise 33.11] for an example in which there exists no R.C.P. on $\mathcal{F}$ given $\mathcal{G}$).

Recall that each C.E. is uniquely determined only a.e. Hence, for any countable collection of disjoint sets $A_n \in \mathcal{F}$ there is possibly a set of $\omega \in \Omega$ of probability zero for which a given collection of C.E. is such that

$$\mathbf{P}(\bigcup_n A_n|\mathcal{G})(\omega) \neq \sum_n \mathbf{P}(A_n|\mathcal{G})(\omega) .$$

In case we need to examine an uncountable number of such collections in order to see whether $\mathbf{P}(\cdot|\mathcal{G})$ is a measure on $(\Omega, \mathcal{F})$, the corresponding exceptional sets of ω can pile up to a non-negligible set, hence the reason why a R.C.P. might not exist.

Nevertheless, as our next proposition shows, the R.C.P.D. exists for any conditioning σ-algebra $\mathcal{G}$ and any real-valued random variable X. In this setting, the R.C.P.D. is the analog of the law of X as in Definition 1.2.34, but now given the information contained in $\mathcal{G}$.

Proposition 4.4.3. *For any real-valued random variable X and any σ-algebra $\mathcal{G} \subseteq \mathcal{F}$, there exists a R.C.P.D. $\widehat{\mathbf{P}}_{X|\mathcal{G}}(\cdot,\cdot)$.*

Proof. Consider the random variables $H(q,\omega) = \mathbf{E}[I_{\{X\leq q\}}|\mathcal{G}](\omega)$, indexed by $q \in \mathbb{Q}$. By monotonicity of the C.E. we know that if $q \leq r$ then $H(q,\omega) \leq H(r,\omega)$ for all $\omega \notin A_{qr}$ where $A_{qr} \in \mathcal{G}$ is such that $\mathbf{P}(A_{qr}) = 0$. Further, by linearity and dominated convergence of C.E.s $H(q + n^{-1},\omega) \to H(q,\omega)$ as $n \to \infty$ for all $\omega \notin B_q$, where $B_q \in \mathcal{G}$ is such that $\mathbf{P}(B_q) = 0$. For the same reason, $H(q,\omega) \to 0$ as $q \to -\infty$ and $H(q,\omega) \to 1$ as $q \to \infty$ for all $\omega \notin C$, where $C \in \mathcal{G}$ is such that $\mathbf{P}(C) = 0$. Since $\mathbb{Q}$ is countable, the set $D = C \bigcup_{r,q} A_{rq} \bigcup_q B_q$ is also in $\mathcal{G}$ with $\mathbf{P}(D) = 0$. Next, for a fixed non-random distribution function $G(\cdot)$, let $F(x,\omega) = \inf\{G(r,\omega) : r \in \mathbb{Q}, r > x\}$, where $G(r,\omega) = H(r,\omega)$ if $\omega \notin D$ and $G(r,\omega) = G(r)$ otherwise. Clearly, for all $\omega \in \Omega$ the non-decreasing function $x \mapsto F(x,\omega)$ converges to zero when $x \to -\infty$ and to one when $x \to \infty$, as $C \subseteq D$. Furthermore, $x \mapsto F(x,\omega)$ is right continuous, hence a distribution function, since

$$\begin{aligned} \lim_{x_n \downarrow x} F(x_n,\omega) &= \inf\{G(r,\omega) : r \in \mathbb{Q}, r > x_n \,\text{for some}\, n\} \\ &= \inf\{G(r,\omega) : r \in \mathbb{Q}, r > x\} = F(x,\omega) . \end{aligned}$$

Thus, to each $\omega \in \Omega$ corresponds a unique probability measure $\widehat{\mathbf{P}}(\cdot, \omega)$ on $(\mathbb{R}, \mathcal{B})$ such that $\widehat{\mathbf{P}}((-\infty, x], \omega) = F(x, \omega)$ for all $x \in \mathbb{R}$ (recall Theorem 1.2.37 for its existence and Proposition 1.2.45 for its uniqueness).

Note that $G(q, \cdot) \in m\mathcal{G}$ for all $q \in \mathbb{Q}$, hence so is $F(x, \cdot)$ for each $x \in \mathbb{R}$ (see Theorem 1.2.22). It follows that $\{B \in \mathcal{B} : \widehat{\mathbf{P}}(B, \cdot) \in m\mathcal{G}\}$ is a λ-system (see Corollary 1.2.19 and Theorem 1.2.22), containing the π-system $\mathcal{P} = \{\mathbb{R}, (-\infty, q] : q \in \mathbb{Q}\}$, hence by Dynkin's $\pi - \lambda$ theorem $\widehat{\mathbf{P}}(B, \cdot) \in m\mathcal{G}$ for all $B \in \mathcal{B}$. Further, for $\omega \notin D$ and $q \in \mathbb{Q}$,

$$H(q, \omega) = G(q, \omega) \leq F(q, \omega) \leq G(q + n^{-1}, \omega) = H(q + n^{-1}, \omega) \to H(q, \omega)$$

as $n \to \infty$ (specifically, the left-most inequality holds for $\omega \notin \cup_r A_{r,q}$ and the right-most limit holds for $\omega \notin B_q$). Hence, $\widehat{\mathbf{P}}(B, \omega) = \mathbf{E}[I_{\{X \in B\}} | \mathcal{G}](\omega)$ for any $B \in \mathcal{P}$ and $\omega \notin D$. Since $\mathbf{P}(D) = 0$ it follows from the definition of the C.E. that for any $G \in \mathcal{G}$ and $B \in \mathcal{P}$,

$$\int_G \widehat{\mathbf{P}}(B, \omega) d\mathbf{P}(\omega) = \mathbf{E}[I_{\{X \in B\}} \cap I_G] \, .$$

Fixing $G \in \mathcal{G}$, by monotone convergence and linearity of the expectation, the set $\mathcal{L}$ of $B \in \mathcal{B}$ for which this equation holds is a λ-system. Consequently, $\mathcal{L} = \sigma(\mathcal{P}) = \mathcal{B}$. Since this applies for all $G \in \mathcal{G}$, we conclude that $\widehat{\mathbf{P}}(B, \cdot)$ is a version of $\mathbf{E}[I_{X \in B} | \mathcal{G}]$ for each $B \in \mathcal{B}$. That is, $\widehat{\mathbf{P}}(B, \omega)$ is per Definition 4.4.2 the R.C.P.D. of X given $\mathcal{G}$. $\square$

Remark. The reason behind Proposition 4.4.3 is that $\sigma(X)$ inherits the structure of the Borel σ-algebra $\mathcal{B}$ which in turn is "not too big" due to the fact the rational numbers are dense in $\mathbb{R}$. Indeed, as you are to deduce in the next exercise, there exists a R.C.P.D. for any $(\mathbb{S}, \mathcal{S})$-valued R.V. X with a $\mathcal{B}$-isomorphic $(\mathbb{S}, \mathcal{S})$.

Exercise 4.4.4. *Suppose $(\mathbb{S}, \mathcal{S})$ is $\mathcal{B}$-isomorphic, that is, there exists a Borel set $\mathbb{T}$ (equipped with the induced Borel σ-algebra $\mathcal{T} = \{B \cap \mathbb{T} : B \in \mathcal{B}\}$) and a one to one and onto mapping $g : \mathbb{S} \mapsto \mathbb{T}$ such that both g and g^{-1} are measurable. For any σ-algebra $\mathcal{G}$ and $(\mathbb{S}, \mathcal{S})$-valued R.V. X let $\widehat{\mathbf{P}}_{Y|\mathcal{G}}(\cdot, \cdot)$ denote the R.C.P.D. of the real-valued random variable $Y = g(X)$.*

(a) *Explain why without loss of generality $\widehat{\mathbf{P}}_{Y|\mathcal{G}}(\mathbb{T}, \omega) = 1$ for all $\omega \in \Omega$.*
(b) *Verify that for any $A \in \mathcal{S}$ both $\{\omega : X(\omega) \in A\} = \{\omega : Y(\omega) \in g(A)\}$ and $g(A) \in \mathcal{B}$.*
(c) *Deduce that $\widehat{\mathbf{Q}}(A, \omega) = \widehat{\mathbf{P}}_{Y|\mathcal{G}}(g(A), \omega)$ is the R.C.P.D. of X given $\mathcal{G}$.*

Our next exercise provides a generalization of Proposition 4.4.3 which is key to the canonical construction of Markov chains in Section 6.1. We note in passing that to conform with the notation for Markov chains, we reverse the order of the arguments in the transition probabilities $\widehat{\mathbf{P}}_{X|Y}(y, A)$ with respect to that of the R.C.P.D. $\widehat{\mathbf{P}}_{X|\sigma(Y)}(A, \omega)$.

Exercise 4.4.5. *Suppose $(\mathbb{S}, \mathcal{S})$ is $\mathcal{B}$-isomorphic and X and Y are $(\mathbb{S}, \mathcal{S})$-valued R.V. in the same probability space $(\Omega, \mathcal{F}, \mathbf{P})$. Prove that there exists (regular)* transition probability $\widehat{\mathbf{P}}_{X|Y}(\cdot, \cdot) : \mathbb{S} \times \mathcal{S} \mapsto [0, 1]$ *such that*

(a) *For each $A \in \mathcal{S}$ fixed, $y \mapsto \widehat{\mathbf{P}}_{X|Y}(y, A)$ is a measurable function and $\widehat{\mathbf{P}}_{X|Y}(Y(\omega), A)$ is a version of the C.E. $\mathbf{E}[I_{X \in A} | \sigma(Y)](\omega)$.*

(b) *For any fixed* $\omega \in \Omega$, *the set function* $\widehat{\mathbf{P}}_{X|Y}(Y(\omega), \cdot)$ *is a probability measure on* $(\mathbb{S}, \mathcal{S})$.

Hint: *With* $g : \mathbb{S} \mapsto \mathbb{T}$ *as before, show that* $\sigma(Y) = \sigma(g(Y))$ *and deduce from Theorem 1.2.26 that* $\widehat{\mathbf{P}}_{X|\sigma(g(Y))}(A, \omega) = f(A, g(Y(\omega))$ *for each* $A \in \mathcal{S}$, *where* $z \mapsto f(A, z)$ *is a Borel function.*

Here is the extension of the change of variables formula (1.3.14) to the setting of conditional distributions.

Exercise 4.4.6. *Suppose* $X \in m\mathcal{F}$ *and* $Y \in m\mathcal{G}$ *for some* σ*-algebras* $\mathcal{G} \subseteq \mathcal{F}$ *are real-valued. Prove that, for any Borel function* $h : \mathbb{R}^2 \mapsto \mathbb{R}$ *such that* $\mathbf{E}|h(X, Y)| < \infty$, *almost surely,*

$$\mathbf{E}[h(X, Y)|\mathcal{G}] = \int_{\mathbb{R}} h(x, Y(\omega)) d\widehat{\mathbf{P}}_{X|\mathcal{G}}(x, \omega) .$$

For an integrable R.V. X (and a non-random constant $Y = c$), this exercise provides the representation

$$\mathbf{E}[X|\mathcal{G}] = \int_{\mathbb{R}} x d\widehat{\mathbf{P}}_{X|\mathcal{G}}(x, \omega) ,$$

of the C.E. in terms of the corresponding R.C.P.D. (with the right side denoting the *Lebesgue's integral* of Definition 1.3.1 for the probability space $(\mathbb{R}, \mathcal{B}, \widehat{\mathbf{P}}_{X|\mathcal{G}}(\cdot, \omega))$.

Solving the next exercise should improve your understanding of the relation between the R.C.P.D. and the conditional probability density function.

Exercise 4.4.7. *Suppose that the random vector* (X, Y, Z) *has a probability density function* $f_{X,Y,Z}$ *per Definition 3.5.5.*

(a) *Express the R.C.P.D.* $\widehat{\mathbf{P}}_{Y|\sigma(X,Z)}$ *in terms of* $f_{X,Y,Z}$.

(b) *Using this expression show that if* X *is independent of* $\sigma(Y, Z)$, *then*

$$\mathbf{E}[Y|X, Z] = \mathbf{E}[Y|Z] .$$

(c) *Provide an example of random variables* X, Y, Z, *such that* X *is independent of* Y *and*

$$\mathbf{E}[Y|X, Z] \neq \mathbf{E}[Y|Z] .$$

Exercise 4.4.8. *Let* $S_n = \sum_{k=1}^n \xi_k$ *for i.i.d. integrable random variables* ξ_k.

(a) *Show that* $\mathbf{E}[\xi_1|S_n] = n^{-1} S_n$.
Hint: *Consider* $\mathbf{E}[\xi_{\pi(1)} I_{S_n \in B}]$ *for* $B \in \mathcal{B}$ *and* π *a uniformly chosen random permutation of* $\{1, \ldots, n\}$ *which is independent of* $\{\xi_k\}$.

(b) *Find* $\mathbf{P}(\xi_1 \leq b|S_2)$ *in case the i.i.d.* ξ_k *are Exponential of parameter* λ.
Hint: *See the representation of Exercise 3.4.11.*

Exercise 4.4.9. *Let* $\mathbf{E}[X|X < Y] = \mathbf{E}[X I_{X<Y}]/\mathbf{P}(X < Y)$ *for integrable* X *and* Y *such that* $\mathbf{P}(X < Y) > 0$. *For each of the following statements, either show that it implies* $\mathbf{E}[X|X < Y] \leq \mathbf{E}X$ *or provide a counter example.*

(a) X *and* Y *are independent.*

(b) *The random vector* (X, Y) *has the same joint law as the random vector* (Y, X) *and* $\mathbf{P}(X = Y) = 0$.

(c) $\mathbf{E}X^2 < \infty$, $\mathbf{E}Y^2 < \infty$ *and* $\mathbf{E}[XY] \leq \mathbf{E}X\mathbf{E}Y$.

Exercise 4.4.10. *Suppose* (X, Y) *are distributed according to a multivariate normal distribution, with* $\mathbf{E}X = \mathbf{E}Y = 0$ *and* $\mathbf{E}Y^2 > 0$*. Show that* $\mathbf{E}[X|Y] = \rho Y$ *with* $\rho = \mathbf{E}[XY]/\mathbf{E}Y^2$.

CHAPTER 5

Discrete time martingales and stopping times

In this chapter we study a collection of stochastic processes called martingales. To simplify our presentation we focus on discrete time martingales and filtrations, also called discrete parameter martingales and filtrations, with definitions and examples provided in Section 5.1 (indeed, a discrete time stochastic process is merely a sequence of random variables defined on the same probability space). As we shall see in Section 5.4, martingales play a key role in computations involving stopping times. Martingales share many other useful properties, chiefly among which are tail bounds and convergence theorems. Section 5.2 deals with martingale representations and tail inequalities, some of which are applied in Section 5.3 to prove various convergence theorems. Section 5.5 further demonstrates the usefulness of martingales in the study of branching processes, likelihood ratios, and exchangeable processes.

5.1. Definitions and closure properties

Subsection 5.1.1 introduces the concepts of filtration, martingale and stopping time and provides a few illustrating examples and interpretations. Subsection 5.1.2 introduces the related super-martingales and sub-martingales, as well as the powerful martingale transform and other closure properties of this collection of stochastic processes.

5.1.1. Martingales, filtrations and stopping times: definitions and examples. Intuitively, a filtration represents any procedure of collecting more and more information as times goes on. Our starting point is the following rigorous mathematical definition of a (discrete time) filtration.

Definition 5.1.1. *A* filtration *is a non-decreasing family of sub-σ-algebras $\{\mathcal{F}_n\}$ of our measurable space $(\Omega, \mathcal{F})$. That is, $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}_n \cdots \subseteq \mathcal{F}$ and $\mathcal{F}_n$ is a σ-algebra for each n. We denote by $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ a filtration $\{\mathcal{F}_n\}$ and the associated σ-algebra $\mathcal{F}_\infty = \sigma(\bigcup_k \mathcal{F}_k)$ such that the relation $\mathcal{F}_k \subseteq \mathcal{F}_\ell$ applies for all $0 \leq k \leq \ell \leq \infty$.*

Given a filtration, we are interested in *stochastic processes* (S.Ps) such that for each n the information gathered by that time suffices for evaluating the value of the n-th element of the process. That is,

Definition 5.1.2. *A S.P. $\{X_n, n = 0, 1, \ldots\}$ is* adapted *to a filtration $\{\mathcal{F}_n\}$, also denoted $\mathcal{F}_n$-*adapted*, if $\sigma(X_n) \subseteq \mathcal{F}_n$ for each n (that is, $X_n \in m\mathcal{F}_n$ for each n).*

At this point you should convince yourself that $\{X_n\}$ is adapted to the filtration $\{\mathcal{F}_n\}$ if and only if $\sigma(X_0, X_1, \ldots, X_n) \subseteq \mathcal{F}_n$ for all n. That is,

Definition 5.1.3. *The filtration $\{\mathcal{F}_n^{\mathbf{X}}\}$ with $\mathcal{F}_n^{\mathbf{X}} = \sigma(X_0, X_1, \ldots, X_n)$ is the minimal filtration with respect to which $\{X_n\}$ is adapted. We therefore call it the* canonical filtration *for the S.P. $\{X_n\}$.*

Whenever clear from the context what it means, we shall use the notation X_n both for the whole S.P. $\{X_n\}$ and for the n-th R.V. of this process, and likewise we may sometimes use $\mathcal{F}_n$ to denote the whole filtration $\{\mathcal{F}_n\}$.

A martingale consists of a filtration and an adapted S.P. which can represent the outcome of a "fair gamble". That is, the expected future reward given current information is exactly the current value of the process, or as a rigorous definition:

Definition 5.1.4. *A* martingale *(denoted MG) is a pair $(X_n, \mathcal{F}_n)$, where $\{\mathcal{F}_n\}$ is a filtration and $\{X_n\}$ is an* integrable S.P.*, that is, $\mathbf{E}|X_n| < \infty$ for all n, adapted to this filtration, such that*

$$\mathbf{E}[X_{n+1}|\mathcal{F}_n] = X_n \qquad \forall n, \quad a.s. \tag{5.1.1}$$

Remark. The "slower" a filtration $n \mapsto \mathcal{F}_n$ grows, the easier it is for an adapted S.P. to be a martingale. That is, if $\mathcal{H}_n \subseteq \mathcal{F}_n$ for all n and S.P. $\{X_n\}$ adapted to filtration $\{\mathcal{H}_n\}$ is such that $(X_n, \mathcal{F}_n)$ is a martingale, then by the tower property $(X_n, \mathcal{H}_n)$ is also a martingale. In particular, if $(X_n, \mathcal{F}_n)$ is a martingale then $\{X_n\}$ is also a martingale with respect to its canonical filtration. For this reason, hereafter the statement $\{X_n\}$ is a MG (without explicitly specifying the filtration), means that $\{X_n\}$ is a MG with respect to its canonical filtration $\mathcal{F}_n^{\mathbf{X}} = \sigma(X_k, k \le n)$.

We next provide an alternative characterization of the martingale property.

Proposition 5.1.5. *If $X_n = \sum_{k=0}^n D_k$ then the canonical filtration for $\{X_n\}$ is the same as the canonical filtration for $\{D_n\}$. Further, $(X_n, \mathcal{F}_n)$ is a martingale if and only if $\{D_n\}$ is an integrable S.P., adapted to $\{\mathcal{F}_n\}$, such that $\mathbf{E}[D_{n+1}|\mathcal{F}_n] = 0$ a.s. for all n.*

Remark. The *martingale differences* associated with $\{X_n\}$ are $D_n = X_n - X_{n-1}$, $n \ge 1$ and $D_0 = X_0$.

Proof. With both the transformation from $(X_0, \ldots, X_n)$ to $(D_0, \ldots, D_n)$ and its inverse being continuous (hence Borel), it follows that $\mathcal{F}_n^{\mathbf{X}} = \mathcal{F}_n^{\mathbf{D}}$ for each n (c.f. Exercise 1.2.33). Therefore, $\{X_n\}$ is adapted to a given filtration $\{\mathcal{F}_n\}$ if and only if $\{D_n\}$ is adapted to this filtration (see Definition 5.1.3). It is easy to show by induction on n that $\mathbf{E}|X_k| < \infty$ for $k = 0, \ldots, n$ if and only if $\mathbf{E}|D_k| < \infty$ for $k = 0, \ldots, n$. Hence, $\{X_n\}$ is an integrable S.P. if and only if $\{D_n\}$ is. Finally, with $X_n \in m\mathcal{F}_n$ it follows from the linearity of the C.E. that

$$\mathbf{E}[X_{n+1}|\mathcal{F}_n] - X_n = \mathbf{E}[X_{n+1} - X_n|\mathcal{F}_n] = \mathbf{E}[D_{n+1}|\mathcal{F}_n]\,,$$

and the alternative expression for the martingale property follows from (5.1.1). □

Our first example of a martingale, is the random walk, perhaps the most fundamental stochastic process.

Definition 5.1.6. *The* random walk *is the stochastic process $S_n = S_0 + \sum_{k=1}^n \xi_k$ with real-valued, independent, identically distributed $\{\xi_k\}$ which are also independent of S_0. Unless explicitly stated otherwise, we always set S_0 to be zero. We say that the random walk is* symmetric *if the law of ξ_k is the same as that of $-\xi_k$. We* *call it a* simple random walk *(on $\mathbb{Z}$), in short* SRW*, if $\xi_k \in \{-1,1\}$. The* SRW *is completely characterized by the parameter $p = \mathbf{P}(\xi_k = 1)$ which is always assumed to be in $(0,1)$ (or alternatively, by $q = 1-p = \mathbf{P}(\xi_k = -1)$). Thus, the* symmetric SRW *corresponds to $p = 1/2 = q$ (and the* asymmetric SRW *corresponds to $p \neq 1/2$).*

The random walk is a MG (with respect to its canonical filtration), whenever $\mathbf{E}|\xi_1| < \infty$ and $\mathbf{E}\xi_1 = 0$.

Remark. More generally, such partial sums $\{S_n\}$ form a MG even when the independent and integrable R.V. ξ_k of zero mean have non-identical distributions, and the canonical filtration of $\{S_n\}$ is merely $\{\mathcal{F}_n^{\boldsymbol{\xi}}\}$, where $\mathcal{F}_n^{\boldsymbol{\xi}} = \sigma(\xi_1,\ldots,\xi_n)$. Indeed, this is an application of Proposition 5.1.5 for independent, integrable $D_k = S_k - S_{k-1} = \xi_k$, $k \geq 1$ (with $D_0 = 0$), where $\mathbf{E}[D_{n+1}|D_0,D_1,\ldots,D_n] = \mathbf{E}D_{n+1} = 0$ for all $n \geq 0$ by our assumption that $\mathbf{E}\xi_k = 0$ for all k.

Definition 5.1.7. *We say that a stochastic process $\{X_n\}$ is* square-integrable *if $\mathbf{E}X_n^2 < \infty$ for all n. Similarly, we call a martingale $(X_n, \mathcal{F}_n)$ such that $\mathbf{E}X_n^2 < \infty$ for all n, an L^2-MG (or a* square-integrable MG*).*

Square-integrable martingales have zero-mean, uncorrelated differences and admit an elegant decomposition of conditional second moments.

Exercise 5.1.8. *Suppose $(X_n, \mathcal{F}_n)$ and $(Y_n, \mathcal{F}_n)$ are square-integrable martingales.*

(a) *Show that the corresponding martingale differences D_n are uncorrelated and that each D_n, $n \geq 1$, has zero mean.*

(b) *Show that for any $\ell \geq n \geq 0$,*

$$\begin{aligned}\mathbf{E}[X_\ell Y_\ell|\mathcal{F}_n] - X_n Y_n &= \mathbf{E}[(X_\ell - X_n)(Y_\ell - Y_n)|\mathcal{F}_n] \\ &= \sum_{k=n+1}^{\ell} \mathbf{E}[(X_k - X_{k-1})(Y_k - Y_{k-1})|\mathcal{F}_n]\,.\end{aligned}$$

(c) *Deduce that if $\sup_k |X_k| \leq C$ non-random then for any $\ell \geq 1$,*

$$\mathbf{E}\Big[\big(\sum_{k=1}^{\ell} D_k^2\big)^2\Big] \leq 6C^4\,.$$

Remark. A square-integrable stochastic process with zero-mean mutually independent differences is necessarily a martingale (consider Proposition 5.1.5). So, in view of part (a) of Exercise 5.1.8, the MG property is between the more restrictive requirement of having zero-mean, independent differences, and the not as useful property of just having zero-mean, uncorrelated differences. While in general these three conditions are not the same, as you show next they do coincide in case of Gaussian stochastic processes.

Exercise 5.1.9. *A stochastic process $\{X_n\}$ is* Gaussian *if for each n the random vector $(X_1,\ldots,X_n)$ has the* multivariate normal *distribution (c.f. Definition 3.5.13). Show that having independent or uncorrelated differences are equivalent properties for such processes, which together with each of these differences having a zero mean is then also equivalent to the MG property.*

Products of R.V. is another classical source for martingales.

Example 5.1.10. *Consider the stochastic process $M_n = \prod_{k=1}^n Y_k$ for independent, integrable random variables $Y_k \geq 0$. Its canonical filtration coincides with $\mathcal{F}_n^{\mathbf{Y}}$ (see Exercise 1.2.33), and taking out what is known we get by independence that*

$$\mathbf{E}[M_{n+1}|\mathcal{F}_n^{\mathbf{Y}}] = \mathbf{E}[Y_{n+1}M_n|\mathcal{F}_n^{\mathbf{Y}}] = M_n\mathbf{E}[Y_{n+1}|\mathcal{F}_n^{\mathbf{Y}}] = M_n\mathbf{E}[Y_{n+1}]\,,$$

so $\{M_n\}$ is a MG, which we then call the product martingale*, if and only if $\mathbf{E}Y_k = 1$ for all $k \geq 1$ (for general sequence $\{Y_n\}$ we need instead that a.s. $\mathbf{E}[Y_{n+1}|Y_1, \ldots, Y_n] = 1$ for all n).*

Remark. In investment applications, the MG condition $\mathbf{E}Y_k = 1$ corresponds to a neutral return rate, and is not the same as the condition $\mathbf{E}[\log Y_k] = 0$ under which the associated partial sums $S_n = \log M_n$ form a MG.

We proceed to define the important concept of stopping time (in the simpler context of a discrete parameter filtration).

Definition 5.1.11. *A random variable τ taking values in $\{0, 1, \ldots, n, \ldots, \infty\}$ is a* stopping time *for the filtration $\{\mathcal{F}_n\}$ (also denoted $\mathcal{F}_n$-*stopping time*), if the event $\{\omega : \tau(\omega) \leq n\}$ is in $\mathcal{F}_n$ for each finite $n \geq 0$.*

Remark. Intuitively, a stopping time corresponds to a situation where the decision whether to stop or not at any given (non-random) time step is based on the information available by that time step. As we shall amply see in the sequel, one of the advantages of MGs is in providing a handle on explicit computations associated with various stopping times.

The next two exercises provide examples of stopping times. Practice your understanding of this concept by solving them.

Exercise 5.1.12. *Suppose that θ and τ are stopping times for the same filtration $\{\mathcal{F}_n\}$. Show that then $\theta \wedge \tau$, $\theta \vee \tau$ and $\theta + \tau$ are also stopping times for this filtration.*

Exercise 5.1.13. *Show that the* first hitting time *$\tau(\omega) = \min\{k \geq 0 : X_k(\omega) \in B\}$ of a Borel set $B \subseteq \mathbb{R}$ by a sequence $\{X_k\}$, is a stopping time for the canonical filtration $\{\mathcal{F}_n^{\mathbf{X}}\}$. Provide an example where the* last hitting time *$\theta = \sup\{k \geq 0 : X_k \in B\}$ of a set B by the sequence, is not a stopping time (not surprising, since we need to know the whole sequence $\{X_k\}$ in order to verify that there are no visits to B after a given time n).*

Here is an elementary application of first hitting times.

Exercise 5.1.14 (Reflection principle)**.** *Suppose $\{S_n\}$ is a symmetric random walk starting at $S_0 = 0$ (see Definition 5.1.6).*

(a) *Show that $\mathbf{P}(S_n - S_k \geq 0) \geq 1/2$ for $k = 1, 2, \ldots, n$.*
(b) *Fixing $x > 0$, let $\tau = \inf\{k \geq 0 : S_k > x\}$ and show that*

$$\mathbf{P}(S_n > x) \geq \sum_{k=1}^n \mathbf{P}(\tau = k, S_n - S_k \geq 0) \geq \frac{1}{2}\sum_{k=1}^n \mathbf{P}(\tau = k)\,.$$

(c) *Deduce that for any n and $x > 0$,*

$$\mathbf{P}(\max_{k=1}^n S_k > x) \leq 2\mathbf{P}(S_n > x)\,.$$

(d) *Considering now the* symmetric SRW, *show that for any positive integers* n, x,

$$\mathbf{P}(\max_{k=1}^{n} S_k \geq x) = 2\mathbf{P}(S_n \geq x) - \mathbf{P}(S_n = x)$$

and that $Z_{2n+1} \stackrel{\mathcal{D}}{=} (|S_{2n+1}|-1)/2$, *where* Z_n *denotes the number of (strict) sign changes within* $\{S_0 = 0, S_1, \ldots, S_n\}$.

We conclude this subsection with a useful sufficient condition for the integrability of a stopping time.

Exercise 5.1.15. *Suppose the* $\mathcal{F}_n$*-stopping time* τ *is such that a.s.*

$$\mathbf{P}[\tau \leq n + r | \mathcal{F}_n] \geq \varepsilon$$

for some positive integer r, *some* $\varepsilon > 0$ *and all* n.

(a) *Show that* $\mathbf{P}(\tau > kr) \leq (1-\varepsilon)^k$ *for any positive integer* k.
Hint: *Use induction on* k.
(b) *Deduce that in this case* $\mathbf{E}\tau < \infty$.

5.1.2. Sub-martingales, super-martingales and stopped martingales. Often when operating on a MG, we naturally end up with a sub-martingale or a super-martingale, as defined next. Moreover, these processes share many of the properties of martingales, so it is useful to develop a unified theory for them.

Definition 5.1.16. *A* sub-martingale *(denoted sub-MG) is an integrable S.P.* $\{X_n\}$, *adapted to the filtration* $\{\mathcal{F}_n\}$, *such that*

$$\mathbf{E}[X_{n+1}|\mathcal{F}_n] \geq X_n \quad \forall n, \quad a.s.$$

A super-martingale *(denoted sup-MG) is an integrable S.P.* $\{X_n\}$, *adapted to the filtration* $\{\mathcal{F}_n\}$ *such that*

$$\mathbf{E}[X_{n+1}|\mathcal{F}_n] \leq X_n \quad \forall n, \quad a.s.$$

(A typical S.P. $\{X_n\}$ *is neither a sub-MG nor a sup-MG, as the sign of the R.V.* $\mathbf{E}[X_{n+1}|\mathcal{F}_n] - X_n$ *may well be random, or possibly dependent upon* n*).*

Remark 5.1.17. Note that $\{X_n\}$ is a sub-MG if and only if $\{-X_n\}$ is a sup-MG. By this identity, all results about sub-MGs have dual statements for sup-MGs and vice verse. We often state only one out of each such pair of statements. Further, $\{X_n\}$ is a MG if and only if $\{X_n\}$ is both a sub-MG and a sup-MG. As a result, every statement holding for either sub-MGs or sup-MGs, also hold for MGs.

Example 5.1.18. *Expanding on Example 5.1.10, if the non-negative, integrable random variables* Y_k *are such that* $\mathbf{E}[Y_n|Y_1, \ldots, Y_{n-1}] \geq 1$ *a.s. for all* n *then* $M_n = \prod_{k=1}^n Y_k$ *is a sub-MG, and if* $\mathbf{E}[Y_n|Y_1, \ldots, Y_{n-1}] \leq 1$ *a.s. for all* n *then* $\{M_n\}$ *is a sup-MG. Such martingales appear for example in mathematical finance, where* Y_k *denotes the random proportional change in the value of a risky asset at the* k*-th trading round. So, positive conditional mean return rate yields a sub-MG while negative conditional mean return rate gives a sup-MG.*

The sub-martingale (and super-martingale) property is closed with respect to the addition of S.P.

Exercise 5.1.19. *Show that if $\{X_n\}$ and $\{Y_n\}$ are sub-MGs with respect to a filtration $\{\mathcal{F}_n\}$, then so is $\{X_n + Y_n\}$. In contrast, show that for any sub-MG $\{Y_n\}$ there exists integrable $\{X_n\}$ adapted to $\{\mathcal{F}_n^{\mathbf{Y}}\}$ such that $\{X_n + Y_n\}$ is not a sub-MG with respect to any filtration.*

Here are some of the properties of sub-MGs (and of sup-MGs).

Proposition 5.1.20. *If $(X_n, \mathcal{F}_n)$ is a sub-MG, then a.s. $\mathbf{E}[X_\ell|\mathcal{F}_m] \geq X_m$ for any $\ell > m$. Consequently, for s sub-MG necessarily $n \mapsto \mathbf{E}X_n$ is non-decreasing. Similarly, for a sup-MG a.s. $\mathbf{E}[X_\ell|\mathcal{F}_m] \leq X_m$ (with $n \mapsto \mathbf{E}X_n$ non-increasing), and for a martingale a.s. $\mathbf{E}[X_\ell|\mathcal{F}_m] = X_m$ for all $\ell > m$ (with $\mathbf{E}[X_n]$ independent of n).*

Proof. Suppose $\{X_n\}$ is a sub-MG and $\ell = m + k$ for $k \geq 1$. Then,

$$\mathbf{E}[X_{m+k}|\mathcal{F}_m] = \mathbf{E}[\mathbf{E}(X_{m+k}|\mathcal{F}_{m+k-1})|\mathcal{F}_m] \geq \mathbf{E}[X_{m+k-1}|\mathcal{F}_m]$$

with the equality due to the tower property and the inequality by the definition of a sub-MG and monotonicity of the C.E. Iterating this inequality for decreasing values of k we deduce that $\mathbf{E}[X_{m+k}|\mathcal{F}_m] \geq \mathbf{E}[X_m|\mathcal{F}_m] = X_m$ for all non-negative integers k, m, as claimed. Next taking the expectation of this inequality, we have by monotonicity of the expectation and (4.2.1) that $\mathbf{E}[X_{m+k}] \geq \mathbf{E}[X_m]$ for all $k, m \geq 0$, or equivalently, that $n \mapsto \mathbf{E}X_n$ is non-decreasing.

To get the corresponding results for a super-martingale $\{X_n\}$ note that then $\{-X_n\}$ is a sub-martingale, see Remark 5.1.17. As already mentioned there, if $\{X_n\}$ is a MG then it is both a super-martingale and a sub-martingale, hence both $\mathbf{E}[X_\ell|\mathcal{F}_m] \geq X_m$ and $\mathbf{E}[X_\ell|\mathcal{F}_m] \leq X_m$, resulting with $\mathbf{E}[X_\ell|\mathcal{F}_m] = X_m$, as stated. □

Exercise 5.1.21. *Show that a sub-martingale $(X_n, \mathcal{F}_n)$ is a martingale if and only if $\mathbf{E}X_n = \mathbf{E}X_0$ for all n.*

We next detail a few examples in which sub-MGs or sup-MGs naturally appear, starting with an immediate consequence of Jensen's inequality

Proposition 5.1.22. *Suppose $\Phi : \mathbb{R} \mapsto \mathbb{R}$ is convex and $\mathbf{E}[|\Phi(X_n)|] < \infty$.*

(a) *If $(X_n, \mathcal{F}_n)$ is a martingale then $(\Phi(X_n), \mathcal{F}_n)$ is a sub-martingale.*
(b) *If $x \mapsto \Phi(x)$ is also non-decreasing, $(\Phi(X_n), \mathcal{F}_n)$ is a sub-martingale even when $(X_n, \mathcal{F}_n)$ is only a sub-martingale.*

Proof. With $\Phi(X_n)$ integrable and adapted, it suffices to check that a.s. $\mathbf{E}[\Phi(X_{n+1})|\mathcal{F}_n] \geq \Phi(X_n)$ for all n. To this end, since $\Phi(\cdot)$ is convex and X_n is integrable, by the conditional Jensen's inequality,

$$\mathbf{E}[\Phi(X_{n+1})|\mathcal{F}_n] \geq \Phi(\mathbf{E}[X_{n+1}|\mathcal{F}_n]),$$

so it remains only to verify that $\Phi(\mathbf{E}[X_{n+1}|\mathcal{F}_n]) \geq \Phi(X_n)$. This clearly applies when $(X_n, \mathcal{F}_n)$ is a MG, and even for a sub-MG $(X_n, \mathcal{F}_n)$, provided that $\Phi(\cdot)$ is monotone non-decreasing. □

Example 5.1.23. *Typical convex functions for which the preceding proposition is often applied are $\Phi(x) = |x|^p$, $p \geq 1$, $\Phi(x) = (x-c)_+$, $\Phi(x) = \max(x, c)$ (for $c \in \mathbb{R}$), $\Phi(x) = e^x$ and $\Phi(x) = x \log x$ (the latter only for non-negative S.P.). Considering instead $\Phi(\cdot)$ concave leads to a sup-MG, as for example when $\Phi(x) = \min(x, c)$ or*

$\Phi(x) = x^p$ for some $p \in (0,1)$ or $\Phi(x) = \log x$ (latter two cases restricted to non-negative S.P.). For example, if $\{X_n\}$ is a sub-martingale then $(X_n - c)_+$ is also a sub-martingale (since $(x-c)_+$ is a convex, non-decreasing function). Similarly, if $\{X_n\}$ is a super-martingale, then $\min(X_n, c)$ is also a super-martingale (since $-X_n$ is a sub-martingale and the function $-\min(-x, c) = \max(x, -c)$ is convex and non-decreasing).

Here is a concrete application of Proposition 5.1.22.

Exercise 5.1.24. *Suppose $\{\xi_i\}$ are mutually independent, $\mathbf{E}\xi_i = 0$ and $\mathbf{E}\xi_i^2 = \sigma_i^2$.*

(a) *Let $S_n = \sum_{i=1}^n \xi_i$ and $s_n^2 = \sum_{i=1}^n \sigma_i^2$. Show that $\{S_n^2\}$ is a sub-martingale and $\{S_n^2 - s_n^2\}$ is a martingale.*

(b) *Show that if in addition $m_n = \prod_{i=1}^n \mathbf{E}e^{\xi_i}$ are finite, then $\{e^{S_n}\}$ is a sub-martingale and $M_n = e^{S_n}/m_n$ is a martingale.*

Remark. A special case of Exercise 5.1.24 is the random walk S_n of Definition 5.1.6, with $S_n^2 - n\mathbf{E}\xi_1^2$ being a MG when ξ_1 is square-integrable and of zero mean. Likewise, e^{S_n} is a sub-MG whenever $\mathbf{E}\xi_1 = 0$ and $\mathbf{E}e^{\xi_1}$ is finite. Though e^{S_n} is in general not a MG, the normalized $M_n = e^{S_n}/[\mathbf{E}e^{\xi_1}]^n$ is merely the product MG of Example 5.1.10 for the i.i.d. variables $Y_i = e^{\xi_i}/\mathbf{E}(e^{\xi_1})$.

Here is another family of super-martingales, this time related to super-harmonic functions.

Definition 5.1.25. *A lower semi-continuous function $f : \mathbb{R}^d \mapsto \mathbb{R}$ is* super-harmonic *if for any x and $r > 0$,*

$$f(x) \geq \frac{1}{|B(0,r)|} \int_{B(x,r)} f(y)dy$$

where $B(x,r) = \{y : |x-y| \leq r\}$ is the ball of radius r centered at x and $|B(x,r)|$ denotes its volume.

Exercise 5.1.26. *Suppose $S_n = x + \sum_{k=1}^n \xi_k$ for i.i.d. ξ_k that are chosen uniformly on the ball $B(0,1)$ in $\mathbb{R}^d$ (i.e. using Lebesgue's measure on this ball, scaled by its volume). Show that if $f(\cdot)$ is super-harmonic on $\mathbb{R}^d$ then $f(S_n)$ is a super-martingale.*
Hint: *When checking the integrability of $f(S_n)$ recall that a lower semi-continuous function is bounded below on any compact set.*

We next define the important concept of a martingale transform, and show that it is a powerful and flexible method for generating martingales.

Definition 5.1.27. *We call a sequence $\{V_n\}$* predictable *(or pre-visible) for the filtration $\{\mathcal{F}_n\}$, also denoted $\mathcal{F}_n$-*predictable*, if V_n is measurable on $\mathcal{F}_{n-1}$ for all $n \geq 1$. The sequence of random variables*

$$Y_n = \sum_{k=1}^n V_k(X_k - X_{k-1}) \,, n \geq 1, \quad Y_0 = 0$$

is called the martingale transform *of the $\mathcal{F}_n$-predictable $\{V_n\}$ with respect to a sub or super martingale $(X_n, \mathcal{F}_n)$.*

Theorem 5.1.28. *Suppose $\{Y_n\}$ is the martingale transform of $\mathcal{F}_n$-predictable $\{V_n\}$ with respect to a sub or super martingale $(X_n, \mathcal{F}_n)$.*

(a) *If Y_n is integrable and $(X_n, \mathcal{F}_n)$ is a martingale, then $(Y_n, \mathcal{F}_n)$ is also a martingale.*
(b) *If Y_n is integrable, $V_n \geq 0$ and $(X_n, \mathcal{F}_n)$ is a sub-martingale (or super-martingale) then $(Y_n, \mathcal{F}_n)$ is also a sub-martingale (super-martingale, respectively).*
(c) *For the integrability of Y_n it suffices in both cases to have $|V_n| \leq c_n$ for some non-random finite constants c_n, or alternatively to have $V_n \in L^q$ and $X_n \in L^p$ for all n and some $p, q > 1$ such that $\frac{1}{q} + \frac{1}{p} = 1$.*

Proof. With $\{V_n\}$ and $\{X_n\}$ adapted to the filtration $\mathcal{F}_n$, it follows that $V_k X_l \in m\mathcal{F}_k \subseteq m\mathcal{F}_n$ for all $l \leq k \leq n$. By inspection $Y_n \in m\mathcal{F}_n$ as well (see Corollary 1.2.19), i.e. $\{Y_n\}$ is adapted to $\{\mathcal{F}_n\}$.

Turning to prove part (c) of the theorem, note that for each n the variable Y_n is a finite sum of terms of the form $\pm V_k X_l$. If $V_k \in L^q$ and $X_l \in L^p$ for some $p, q > 1$ such that $\frac{1}{q} + \frac{1}{p} = 1$, then by Hölder's inequality $V_k X_l$ is integrable. Alternatively, since a super-martingale X_l is in particular integrable, $V_k X_l$ is integrable as soon as $|V_k|$ is bounded by a non-random finite constant. In conclusion, if either of these conditions applies for all k, l then obviously $\{Y_n\}$ is an integrable S.P.

Recall that $Y_{n+1} - Y_n = V_{n+1}(X_{n+1} - X_n)$ and $V_{n+1} \in m\mathcal{F}_n$ (since $\{V_n\}$ is $\mathcal{F}_n$-predictable). Therefore, taking out V_{n+1} which is measurable on $\mathcal{F}_n$ we find that

$$\mathbf{E}[Y_{n+1} - Y_n|\mathcal{F}_n] = \mathbf{E}[V_{n+1}(X_{n+1} - X_n)|\mathcal{F}_n] = V_{n+1}\mathbf{E}[X_{n+1} - X_n|\mathcal{F}_n] \,.$$

This expression is zero when $(X_n, \mathcal{F}_n)$ is a MG and non-negative when $V_{n+1} \geq 0$ and $(X_n, \mathcal{F}_n)$ is a sub-MG. Since the preceding applies for all n, we consequently have that $(Y_n, \mathcal{F}_n)$ is a MG in the former case and a sub-MG in the latter. Finally, to complete the proof also in case of a sup-MG $(X_n, \mathcal{F}_n)$, note that then $-Y_n$ is the MG transform of $\{V_n\}$ with respect to the sub-MG $(-X_n, \mathcal{F}_n)$. □

Here are two concrete examples of a martingale transform.

Example 5.1.29. *The S.P. $Y_n = \sum_{k=1}^{n} X_{k-1}(X_k - X_{k-1})$ is a MG whenever $X_n \in L^2(\Omega, \mathcal{F}, \mathbf{P})$ is a MG (indeed, $V_n = X_{n-1}$ is predictable for the canonical filtration of $\{X_n\}$ and consider $p = q = 2$ in part (c) of Theorem 5.1.28).*

Example 5.1.30. *Given an integrable process $\{V_n\}$ suppose that for each $k \geq 1$ the bounded ξ_k has zero mean and is independent of $\mathcal{F}_{k-1} = \sigma(\xi_1, \ldots, \xi_{k-1}, V_1, \ldots, V_k)$. Then, $Y_n = \sum_{k=1}^{n} V_k \xi_k$ is a martingale for the filtration $\{\mathcal{F}_n\}$. Indeed, by assumption, the differences ξ_n of $X_n = \sum_{k=1}^{n} \xi_k$ are such that $\mathbf{E}[\xi_k|\mathcal{F}_{k-1}] = 0$ for all $k \geq 1$. Hence, $(X_n, \mathcal{F}_n)$ is a martingale (c.f. Proposition 5.1.5), and $\{Y_n\}$ is the martingale transform of the $\mathcal{F}_n$-predictable $\{V_n\}$ with respect to the martingale $(X_n, \mathcal{F}_n)$ (where the integrability of Y_n is a consequence of the boundedness of each ξ_k and integrability of each V_k). In discrete mathematics applications one often uses a special case of this construction, with an auxiliary sequence of random i.i.d. signs $\xi_k \in \{-1, 1\}$ such that $\mathbf{P}(\xi_1 = 1) = \frac{1}{2}$ and $\{\xi_n\}$ is independent of the given integrable S.P. $\{V_n\}$.*

We next define the important concept of a stopped stochastic process and then use the martingale transform to show that stopped sub and super martingales are also sub-MGs (sup-MGs, respectively).

Definition 5.1.31. *Given a stochastic process* $\{X_n\}$ *and a a random variable* τ *taking values in* $\{0, 1, \dots, n, \dots, \infty\}$*, the* stopped at τ stochastic process*, denoted* $\{X_{n\wedge\tau}\}$*, is given by*

$$X_{n\wedge\tau}(\omega) = \begin{cases} X_n(\omega), & n \le \tau(\omega) \\ X_{\tau(\omega)}(\omega), & n > \tau(\omega) \end{cases}$$

Theorem 5.1.32. *If* $(X_n, \mathcal{F}_n)$ *is a sub-MG (or a sup-MG or a MG) and* $\theta \le \tau$ *are stopping times for* $\{\mathcal{F}_n\}$*, then* $(X_{n\wedge\tau} - X_{n\wedge\theta}, \mathcal{F}_n)$ *is also a sub-MG (or sup-MG or MG, respectively). In particular, taking* $\theta = 0$ *we have that* $(X_{n\wedge\tau}, \mathcal{F}_n)$ *is then a sub-MG (or sup-MG or MG, respectively).*

Proof. We may and shall assume that $(X_n, \mathcal{F}_n)$ is a sub-MG (just consider $-X_n$ in case X_n is a sup-MG and both when X_n is a MG). Let $V_k(\omega) = I_{\{\theta(\omega)<k\le\tau(\omega)\}}$. Since $\theta \le \tau$ are two $\mathcal{F}_n$-stopping times, it follows that $V_k(\omega) = I_{\{\theta(\omega)\le(k-1)\}} - I_{\{\tau(\omega)\le(k-1)\}}$ is measurable on $\mathcal{F}_{k-1}$ for all $k \ge 1$. Thus, $\{V_n\}$ is a bounded, non-negative $\mathcal{F}_n$-predictable sequence. Further, since

$$X_{n\wedge\tau}(\omega) - X_{n\wedge\theta}(\omega) = \sum_{k=1}^{n} I_{\{\theta(\omega)<k\le\tau(\omega)\}}(X_k(\omega) - X_{k-1}(\omega))$$

is the martingale transform of $\{V_n\}$ with respect to sub-MG $(X_n, \mathcal{F}_n)$, we know from Theorem 5.1.28 that $(X_{n\wedge\tau} - X_{n\wedge\theta}, \mathcal{F}_n)$ is also a sub-MG. Finally, considering the latter sub-MG for $\theta = 0$ and adding to it the sub-MG $(X_0, \mathcal{F}_n)$, we conclude that $(X_{n\wedge\tau}, \mathcal{F}_n)$ is a sub-MG (c.f. Exercise 5.1.19 and note that $X_{n\wedge 0} = X_0$). $\square$

Theorem 5.1.32 thus implies the following key ingredient in the proof of Doob's optional stopping theorem (to which we return in Section 5.4).

Corollary 5.1.33. *If* $(X_n, \mathcal{F}_n)$ *is a sub-MG and* $\tau \ge \theta$ *are* $\mathcal{F}_n$*-stopping times, then* $\mathbf{E}X_{n\wedge\tau} \ge \mathbf{E}X_{n\wedge\theta}$ *for all* n*. The reverse inequality holds in case* $(X_n, \mathcal{F}_n)$ *is a sup-MG, with* $\mathbf{E}X_{n\wedge\theta} = \mathbf{E}X_{n\wedge\tau}$ *for all* n *in case* $(X_n, \mathcal{F}_n)$ *is a MG.*

Proof. Suffices to consider X_n which is a sub-MG for the filtration $\mathcal{F}_n$. In this case we have from Theorem 5.1.32 that $Y_n = X_{n\wedge\tau} - X_{n\wedge\theta}$ is also a sub-MG for this filtration. Noting that $Y_0 = 0$ we thus get from Proposition 5.1.20 that $\mathbf{E}Y_n \ge 0$. Theorem 5.1.32 also implies the integrability of $X_{n\wedge\theta}$ so by linearity of the expectation we conclude that $\mathbf{E}X_{n\wedge\tau} \ge \mathbf{E}X_{n\wedge\theta}$. $\square$

An important concept associated with each stopping time is the *stopped* σ*-algebra* defined next.

Definition 5.1.34. *The* stopped σ-algebra $\mathcal{F}_\tau$ *associated with the stopping time* τ *for a filtration* $\{\mathcal{F}_n\}$ *is the collection of events* $A \in \mathcal{F}_\infty$ *such that* $A \cap \{\omega : \tau(\omega) \le n\} \in \mathcal{F}_n$ *for all* n*.*

With $\mathcal{F}_n$ representing the information known at time n, think of $\mathcal{F}_\tau$ as quantifying the information known upon stopping at τ. Some of the properties of these stopped σ-algebras are detailed in the next exercise.

Exercise 5.1.35. *Let* θ *and* τ *be* $\mathcal{F}_n$*-stopping times.*

(a) *Verify that* $\mathcal{F}_\tau$ *is a* σ*-algebra and if* $\tau(\omega) = n$ *is non-random then* $\mathcal{F}_\tau = \mathcal{F}_n$*.*

(b) *Suppose* $X_n \in m\mathcal{F}_n$ *for all* n *(including* $n = \infty$ *unless* τ *is finite for all* ω*). Show that then* $X_\tau \in m\mathcal{F}_\tau$*. Deduce that* $\sigma(\tau) \subseteq \mathcal{F}_\tau$ *and* $X_k I_{\{\tau=k\}} \in m\mathcal{F}_\tau$ *for any* k *non-random.*

(c) *Show that for any integrable* $\{Y_n\}$ *and non-random* k*,*

$$\mathbf{E}[Y_\tau I_{\{\tau=k\}} | \mathcal{F}_\tau] = \mathbf{E}[Y_k | \mathcal{F}_k] I_{\{\tau=k\}} .$$

(d) *Show that if* $\theta \leq \tau$ *then* $\mathcal{F}_\theta \subseteq \mathcal{F}_\tau$*.*

Our next exercise shows that the martingale property is equivalent to the "strong martingale property" whereby conditioning at stopped σ-algebras $\mathcal{F}_\theta$ replaces the one at $\mathcal{F}_n$ for non-random n.

Exercise 5.1.36. *Given an integrable stochastic process* $\{X_n\}$ *adapted to a filtration* $\{\mathcal{F}_n\}$*, show that* $(X_n, \mathcal{F}_n)$ *is a martingale if and only if* $\mathbf{E}[X_n | \mathcal{F}_\theta] = X_\theta$ *for any non-random, finite* n *and all* $\mathcal{F}_n$*-stopping times* $\theta \leq n$*.*

For non-integrable stochastic processes we generalize the concept of a martingale into that of a local martingale.

Exercise 5.1.37. *The pair* $(X_n, \mathcal{F}_n)$ *is called a* local martingale *if* $\{X_n\}$ *is adapted to the filtration* $\{\mathcal{F}_n\}$ *and there exist* $\mathcal{F}_n$*-stopping times* τ_k *such that* $\tau_k \uparrow \infty$ *with probability one and* $(X_{n\wedge\tau_k}, \mathcal{F}_n)$ *is a martingale for each* k*. Show that any martingale is a local martingale and any integrable, local martingale is a martingale.*

We conclude with the renewal property of stopping times with respect to the canonical filtration of an i.i.d. sequence.

Exercise 5.1.38. *Suppose* τ *is an a.s. finite stopping time with respect to the canonical filtration* $\{\mathcal{F}_n^{\mathbf{Z}}\}$ *of a sequence* $\{Z_k\}$ *of i.i.d. R.V-s.*

(a) *Show that* $\mathcal{T}_\tau^{\mathbf{Z}} = \sigma(Z_{\tau+k}, k \geq 1)$ *is independent of the stopped* σ*-algebra* $\mathcal{F}_\tau^{\mathbf{Z}}$*.*

(b) *Provide an example of a finite* $\mathcal{F}_n^{\mathbf{Z}}$*-stopping time* τ *and independent* $\{Z_k\}$ *for which* $\mathcal{T}_\tau^{\mathbf{Z}}$ *is not independent of* $\mathcal{F}_\tau^{\mathbf{Z}}$*.*

5.2. Martingale representations and inequalities

In Subsection 5.2.1 we show that martingales are at the core of all adapted processes. We further explore there the structure of certain sub-martingales, introducing the increasing process associated with square-integrable martingales. This is augmented in Subsection 5.2.2 by the study of maximal inequalities for sub-martingales (and martingales). Such inequalities are an important technical tool in many applications of probability theory. In particular, they are the key to the convergence results of Section 5.3.

5.2.1. Martingale decompositions.

To demonstrate the relevance of martingales to the study of general stochastic processes, we start with a representation of any adapted, integrable, discrete-time S.P. as the sum of a martingale and a predictable process.

Theorem 5.2.1 (Doob's decomposition)**.** *Given an integrable stochastic process* $\{X_n\}$*, adapted to a discrete parameter filtration* $\{\mathcal{F}_n\}$*,* $n \geq 0$*, there exists a decomposition* $X_n = Y_n + A_n$ *such that* $(Y_n, \mathcal{F}_n)$ *is a MG and* $\{A_n\}$ *is an* $\mathcal{F}_n$*-predictable sequence. This decomposition is unique up to the value of* $Y_0 \in m\mathcal{F}_0$*.*

PROOF. Let $A_0 = 0$ and for $n \geq 1$ set

$$A_n = A_{n-1} + \mathbf{E}[X_n - X_{n-1}|\mathcal{F}_{n-1}].$$

By definition of the conditional expectation we see that $A_k - A_{k-1}$ is measurable on $\mathcal{F}_{k-1}$ for any $k \geq 1$. Since $\mathcal{F}_{k-1} \subseteq \mathcal{F}_{n-1}$ for all $k \leq n$ and $A_n = A_0 + \sum_{k=1}^{n}(A_k - A_{k-1})$, it follows that $\{A_n\}$ is $\mathcal{F}_n$-predictable. We next check that $Y_n = X_n - A_n$ is a MG. To this end, recall that since $\{X_n\}$ is integrable so is $\{X_n - X_{n-1}\}$, whereas the C.E. only reduces the L^1 norm (see Example 4.2.20). Therefore, $\mathbf{E}|A_n - A_{n-1}| \leq \mathbf{E}|X_n - X_{n-1}| < \infty$. Hence, A_n is integrable, as is X_n, implying by Minkowski's inequality that Y_n is integrable as well. With $\{X_n\}$ adapted and $\{A_n\}$ predictable, hence adapted, to $\{\mathcal{F}_n\}$, we see that $\{Y_n\}$ is also $\mathcal{F}_n$-adapted. It remains to check the martingale condition, that almost surely $\mathbf{E}[Y_n - Y_{n-1}|\mathcal{F}_{n-1}] = 0$ for all $n \geq 1$. Indeed, by linearity of the C.E. and the construction of the $\mathcal{F}_n$-predictable sequence $\{A_n\}$, for any $n \geq 1$,

$$\begin{aligned}\mathbf{E}[Y_n - Y_{n-1}|\mathcal{F}_{n-1}] &= \mathbf{E}[X_n - X_{n-1} - (A_n - A_{n-1})|\mathcal{F}_{n-1}] \\ &= \mathbf{E}[X_n - X_{n-1}|\mathcal{F}_{n-1}] - (A_n - A_{n-1}) = 0\,.\end{aligned}$$

We finish the proof by checking that such a decomposition is unique up to the choice of Y_0. To this end, suppose that $X_n = Y_n + A_n = \widetilde{Y}_n + \widetilde{A}_n$ are two such decompositions of a given stochastic process $\{X_n\}$. Then, $\widetilde{Y}_n - Y_n = A_n - \widetilde{A}_n$. Since $\{A_n\}$ and $\{\widetilde{A}_n\}$ are both $\mathcal{F}_n$-predictable sequences while $(Y_n, \mathcal{F}_n)$ and $(\widetilde{Y_n}, \mathcal{F}_n)$ are martingales, we find that

$$\begin{aligned}A_n - \widetilde{A}_n &= \mathbf{E}[A_n - \widetilde{A}_n|\mathcal{F}_{n-1}] = \mathbf{E}[\widetilde{Y}_n - Y_n|\mathcal{F}_{n-1}] \\ &= \widetilde{Y}_{n-1} - Y_{n-1} = A_{n-1} - \widetilde{A}_{n-1}\,.\end{aligned}$$

Thus, $A_n - \widetilde{A}_n$ is independent of n and if in addition $Y_0 = \widetilde{Y}_0$ then $A_n - \widetilde{A}_n = A_0 - \widetilde{A}_0 = \widetilde{Y}_0 - Y_0 = 0$ for all n. In conclusion, both sequences $\{A_n\}$ and $\{Y_n\}$ are uniquely determined as soon as we determine Y_0, a R.V. measurable on $\mathcal{F}_0$. $\square$

Doob's decomposition has more structure when $(X_n, \mathcal{F}_n)$ is a sub-MG.

Exercise 5.2.2. *Check that the predictable part of Doob's decomposition of a sub-martingale $(X_n, \mathcal{F}_n)$ is a non-decreasing sequence, that is, $A_n \leq A_{n+1}$ for all n.*

Remark. As shown in Subsection 5.3.2, Doob's decomposition is particularly useful in connection with square-integrable martingales $\{X_n\}$, where one can relate the limit of X_n as $n \to \infty$ with that of the non-decreasing sequence $\{A_n\}$ in the decomposition of $\{X_n^2\}$.

We next evaluate Doob's decomposition for two classical sub-MGs.

Example 5.2.3. *Consider the sub-MG $\{S_n^2\}$ for the random walk $S_n = \sum_{k=1}^{n} \xi_k$, where ξ_k are i.i.d. random variables with $\mathbf{E}\xi_1 = 0$ and $\mathbf{E}\xi_1^2 = 1$. Since $Y_n = S_n^2 - n$ is a martingale (see Exercise 5.1.24), and Doob's decomposition $S_n^2 = Y_n + A_n$ is unique, it follows that the non-decreasing predictable part in the decomposition of S_n^2 is $A_n = n$.*

In contrast with the preceding example, the non-decreasing predictable part in Doob's decomposition is for most sub-MGs a non-degenerate random sequence, as is the case in our next example.

Example 5.2.4. *Consider the sub-MG* $(M_n, \mathcal{F}_n^{\mathbf{Z}})$ *where* $M_n = \prod_{i=1}^n Z_i$ *for i.i.d. integrable* $Z_i \geq 0$ *such that* $\mathbf{E}Z_1 > 1$ *(see Example 5.1.10). The non-decreasing predictable part of its Doob's decomposition is such that for* $n \geq 1$

$$\begin{aligned} A_{n+1} - A_n &= \mathbf{E}[M_{n+1} - M_n | \mathcal{F}_n^{\mathbf{Z}}] = \mathbf{E}[Z_{n+1}M_n - M_n | \mathcal{F}_n^{\mathbf{Z}}] \\ &= M_n \mathbf{E}[Z_{n+1} - 1 | \mathcal{F}_n^{\mathbf{Z}}] = M_n(\mathbf{E}Z_1 - 1) \end{aligned}$$

(since Z_{n+1} *is independent of* $\mathcal{F}_n^{\mathbf{Z}}$*). In this case* $A_n = (\mathbf{E}Z_1 - 1)\sum_{k=1}^{n-1} M_k + A_1$*, where we are free to choose for* A_1 *any non-random constant. We see that* $\{A_n\}$ *is a non-degenerate random sequence (assuming the R.V.* Z_i *are not a.s. constant).*

We conclude with the representation of any L^1-bounded martingale as the difference of two non-negative martingales (resembling the representation $X = X_+ - X_-$ for an integrable R.V. X and non-negative $X_\pm$).

Exercise 5.2.5. *Let* $(X_n, \mathcal{F}_n)$ *be a martingale with* $\sup_n \mathbf{E}|X_n| < \infty$*. Show that there is a representation* $X_n = Y_n - Z_n$ *with* $(Y_n, \mathcal{F}_n)$ *and* $(Z_n, \mathcal{F}_n)$ *non-negative martingales such that* $\sup_n \mathbf{E}|Y_n| < \infty$ *and* $\sup_n \mathbf{E}|Z_n| < \infty$*.*

5.2.2. Maximal and up-crossing inequalities.

Martingales are rather tame stochastic processes. In particular, as we see next, the tail of $\max_{k \leq n} X_k$ is bounded by moments of X_n. This is a major improvement over Markov's inequality, relating the typically much smaller tail of the R.V. X_n to its moments (see part (b) of Example 1.3.14).

Theorem 5.2.6 (Doob's inequality)**.** *For any sub-martingale* $\{X_n\}$ *and* $x > 0$ *let* $\tau_x = \min\{k \geq 0 : X_k \geq x\}$*. Then, for any finite* $n \geq 0$*,*

$$\mathbf{P}(\max_{k=0}^{n} X_k \geq x) \leq x^{-1}\mathbf{E}[X_n I_{\{\tau_x \leq n\}}] \leq x^{-1}\mathbf{E}[(X_n)_+]\,. \tag{5.2.1}$$

Proof. Since $X_{\tau_x} \geq x$ whenever τ_x is finite, setting

$$A_n = \{\omega : \tau_x(\omega) \leq n\} = \{\omega : \max_{k=0}^{n} X_k(\omega) \geq x\}\,,$$

it follows that

$$\mathbf{E}[X_{n\wedge\tau_x}] = \mathbf{E}[X_{\tau_x} I_{\tau_x \leq n}] + \mathbf{E}[X_n I_{\tau_x > n}] \geq x\mathbf{P}(A_n) + \mathbf{E}[X_n I_{A_n^c}].$$

With $\{X_n\}$ a sub-MG and $\tau_x \leq \infty$ a pair of $\mathcal{F}_n^{\mathbf{X}}$-stopping times, it follows from Corollary 5.1.33 that $\mathbf{E}[X_{n\wedge\tau_x}] \leq \mathbf{E}[X_n]$. Therefore, $\mathbf{E}[X_n] - \mathbf{E}[X_n I_{A_n^c}] \geq x\mathbf{P}(A_n)$ which is exactly the left inequality in (5.2.1). The right inequality there holds by monotonicity of the expectation and the trivial fact $XI_A \leq (X)_+$ for any R.V. X and any measurable set A. □

Remark. Doob's inequality generalizes Kolmogorov's maximal inequality. Indeed, consider $X_k = Z_k^2$ for the L^2-martingale $Z_k = Y_1 + \cdots + Y_k$, where $\{Y_l\}$ are mutually independent with $\mathbf{E}Y_l = 0$ and $\mathbf{E}Y_l^2 < \infty$. By Proposition 5.1.22 $\{X_k\}$ is a sub-MG, so by Doob's inequality we obtain that for any $z > 0$,

$$\mathbf{P}(\max_{1\leq k\leq n} |Z_k| \geq z) = \mathbf{P}(\max_{1\leq k\leq n} X_k \geq z^2) \leq z^{-2}\mathbf{E}[(X_n)_+] = z^{-2}\,\mathsf{Var}(Z_n)$$

which is exactly Kolmogorov's maximal inequality of Proposition 2.3.16.

Combining Doob's inequality with Doob's decomposition of non-negative sub-martingales, we arrive at the following bounds, due to Lenglart.

Lemma 5.2.7. *Let $V_n = \max_{k=0}^n Z_k$ and A_n denote the $\mathcal{F}_n$-predictable sequence in Doob's decomposition of a non-negative submartingale $(Z_n, \mathcal{F}_n)$ with $Z_0 = 0$. Then, for any $\mathcal{F}_n$-stopping time τ and all $x, y > 0$,*

$$\mathbf{P}(V_\tau \geq x, A_\tau \leq y) \leq x^{-1}\mathbf{E}(A_\tau \wedge y)\,. \tag{5.2.2}$$

Further, in this case $\mathbf{E}[V_\tau^p] \leq c_p \mathbf{E}[A_\tau^p]$ for $c_p = 1 + 1/(1-p)$ and any $p \in (0,1)$.

Proof. Since $M_n = Z_n - A_n$ is a MG with respect to the filtration $\{\mathcal{F}_n\}$ (starting at $M_0 = 0$), by Theorem 5.1.32 the same applies for the stopped stochastic process $M_{n\wedge\theta}$, with θ any $\mathcal{F}_n$-stopping time. By the same reasoning $Z_{n\wedge\theta} = M_{n\wedge\theta} + A_{n\wedge\theta}$ is a sub-MG with respect to $\{\mathcal{F}_n\}$. Applying Doob's inequality (5.2.1) for this non-negative sub-MG we deduce that for any n and $x > 0$,

$$\mathbf{P}(V_{n\wedge\theta} \geq x) = \mathbf{P}(\max_{k=0}^n Z_{k\wedge\theta} \geq x) \leq x^{-1}\mathbf{E}[\,Z_{n\wedge\theta}\,] = x^{-1}\mathbf{E}[\,A_{n\wedge\theta}\,]\,.$$

Both $V_{n\wedge\theta}$ and $A_{n\wedge\theta}$ are non-negative and non-decreasing in n (see Exercise 5.2.2), so by monotone convergence we have that $\mathbf{P}(V_\theta \geq x) \leq x^{-1}\mathbf{E}A_\theta$. In particular, fixing $y > 0$, since $\{A_n\}$ is $\mathcal{F}_n$-predictable, $\theta = \tau \wedge \min\{n \geq 0 : A_{n+1} > y\}$ is an $\mathcal{F}_n$-stopping time. Further, with A_n non-decreasing, $\theta < \tau$ if and only if $A_\tau > y$ in which case $A_\theta \leq y$ (by the definition of θ). Consequently, $A_\theta \leq A_\tau \wedge y$ and as $\{V_\tau \geq x, A_\tau \leq y\} \subseteq \{V_\theta \geq x\}$ we arrive at the inequality (5.2.2).

Next, considering (5.2.2) for $x = y$ we see that for $Y = A_\tau$ and any $y > 0$,

$$\mathbf{P}(V_\tau \geq y) \leq \mathbf{P}(Y \geq y) + \mathbf{E}[\min(Y/y, 1)]\,.$$

Multiplying both sides of this inequality by py^{p-1} and integrating over $y \in (0,\infty)$, upon taking $r = 1 > p$ in part (a) of Lemma 1.4.31 we conclude that

$$\mathbf{E}V_\tau^p \leq \mathbf{E}Y^p + (1-p)^{-1}\mathbf{E}Y^p\,,$$

as claimed. □

To practice your understanding, adapt the proof of Doob's inequality en-route to the following dual inequality (which is often called *Doob's second sub-MG inequality*).

Exercise 5.2.8. *Show that for any sub-MG $\{X_n\}$, finite $n \geq 0$ and $x > 0$,*

$$\mathbf{P}(\min_{k=0}^n X_k \leq -x) \leq x^{-1}(\mathbf{E}[(X_n)_+] - \mathbf{E}[X_0])\,. \tag{5.2.3}$$

Here is a typical example of an application of Doob's inequality.

Exercise 5.2.9. *Fixing $s > 0$, the independent variables Z_n are such that $\mathbf{P}(Z_n = -1) = \mathbf{P}(Z_n = 1) = n^{-s}/2$ and $\mathbf{P}(Z_n = 0) = 1 - n^{-s}$. Starting at $Y_0 = 0$, for $n \geq 1$ let*

$$Y_n = n^s Y_{n-1}|Z_n| + Z_n I_{\{Y_{n-1}=0\}}\,.$$

(a) *Show that $\{Y_n\}$ is a martingale and that for any $x > 0$ and $n \geq 1$,*

$$\mathbf{P}(\max_{k=1}^n Y_k \geq x) \leq \frac{1}{2x}[1 + \sum_{k=1}^{n-1}(k+1)^{-s}(1-k^{-s})]\,.$$

(b) *Show that $Y_n \xrightarrow{p} 0$ as $n \to \infty$ and further $Y_n \xrightarrow{a.s.} 0$ if and only if $s > 1$, but there is no value of s for which $Y_n \xrightarrow{L^1} 0$.*

Martingales also provide bounds on the probability that the sum of bounded independent variables is too close to its mean (in lieu of the CLT).

Exercise 5.2.10. *Let $S_n = \sum_{k=1}^n \xi_k$ where $\{\xi_k\}$ are independent and $\mathbf{E}\xi_k = 0$, $|\xi_k| \le K$ for all k. Let $s_n^2 = \sum_{k=1}^n \mathbf{E}\xi_k^2$. Using Corollary 5.1.33 for the martingale $S_n^2 - s_n^2$ and a suitable stopping time show that*

$$\mathbf{P}(\max_{k=1}^{n} |S_k| \le x) \le (x+K)^2/s_n^2\,.$$

If the positive part of the sub-MG has finite p-th moment you can improve the rate of decay in x in Doob's inequality by an application of Proposition 5.1.22 for the convex non-decreasing $\Phi(y) = \max(y,0)^p$, denoted hereafter by $(y)_+^p$. Further, in case of a MG the same argument yields comparable bounds on tail probabilities for the maximum of $|Y_k|$.

Exercise 5.2.11.

(a) *Show that for any sub-MG $\{Y_n\}$, $p \ge 1$, finite $n \ge 0$ and $y > 0$,*

$$\mathbf{P}(\max_{k=0}^{n} Y_k \ge y) \le y^{-p}\mathbf{E}\Big[\max(Y_n,0)^p\Big]\,.$$

(b) *Show that in case $\{Y_n\}$ is a martingale, also*

$$\mathbf{P}(\max_{k=1}^{n} |Y_k| \ge y) \le y^{-p}\mathbf{E}\Big[|Y_n|^p\Big]\,.$$

(c) *Suppose the martingale $\{Y_n\}$ is such that $Y_0 = 0$. Using the fact that $(Y_n + c)^2$ is a sub-martingale and optimizing over c, show that for $y > 0$,*

$$\mathbf{P}(\max_{k=0}^{n} Y_k \ge y) \le \frac{\mathbf{E}Y_n^2}{\mathbf{E}Y_n^2 + y^2}$$

Here is the version of Doob's inequality for non-negative sup-MGs and its application for the random walk.

Exercise 5.2.12.

(a) *Show that if τ is a stopping time for the canonical filtration of a non-negative super-martingale $\{X_n\}$ then $\mathbf{E}X_0 \ge \mathbf{E}X_{n\wedge\tau} \ge \mathbf{E}[X_\tau I_{\tau\le n}]$ for any finite n.*

(b) *Deduce that if $\{X_n\}$ is a non-negative super-martingale then for any $x > 0$*

$$\mathbf{P}(\sup_k X_k \ge x) \le x^{-1}\mathbf{E}X_0\,.$$

(c) *Suppose S_n is a random walk with $\mathbf{E}\xi_1 = -\mu < 0$ and $\mathsf{Var}(\xi_1) = \sigma^2 > 0$. Let $\alpha = \mu/(\sigma^2+\mu^2)$ and $f(x) = 1/(1+\alpha(z-x)_+)$. Show that $f(S_n)$ is a super-martingale and use this to conclude that for any $z > 0$,*

$$\mathbf{P}(\sup_k S_k \ge z) \le \frac{1}{1+\alpha z}\,.$$

Hint: *Taking $v(x) = \alpha f(x)^2 \mathbf{1}_{x<z}$ show that $g_x(y) = f(x) + v(x)[(y-x) + \alpha(y-x)^2] \ge f(y)$ for all x and y. Then show that $f(S_n) = \mathbf{E}[g_{S_n}(S_{n+1})|S_k, k \le n]$.*

Integrating Doob's inequality we next get bounds on the moments of the maximum of a sub-MG.

Corollary 5.2.13 (L^p MAXIMAL INEQUALITIES). *If $\{X_n\}$ is a sub-MG then for any n and $p > 1$,*

$$\mathbf{E}\big[(\max_{k\le n} X_k)_+^p\big] \le q^p \mathbf{E}\big[(X_n)_+^p\big], \tag{5.2.4}$$

where $q = p/(p-1)$ is a finite universal constant. Consequently, if $\{Y_n\}$ is a MG then for any n and $p > 1$,

$$\mathbf{E}\Big[\big(\max_{k\le n} |Y_k|\big)^p\Big] \le q^p \mathbf{E}\big[|Y_n|^p\big] . \tag{5.2.5}$$

PROOF. The bound (5.2.4) is obtained by applying part (b) of Lemma 1.4.31 for the non-negative variables $X = (X_n)_+$ and $Y = (\max_{k\le n} X_k)_+$. Indeed, the hypothesis $\mathbf{P}(Y \ge y) \le y^{-1}\mathbf{E}[X I_{Y\ge y}]$ of this lemma is provided by the left inequality in (5.2.1) and its conclusion that $\mathbf{E}Y^p \le q^p \mathbf{E}X^p$ is precisely (5.2.4). In case $\{Y_n\}$ is a martingale, we get (5.2.5) by applying (5.2.4) for the non-negative sub-MG $X_n = |Y_n|$. □

Remark. A bound such as (5.2.5) can not hold for all sub-MGs. For example, the non-random sequence $Y_k = (k-n) \wedge 0$ is a sub-MG with $|Y_0| = n$ but $Y_n = 0$.

The following two exercises show that while L^p maximal inequalities as in Corollary 5.2.13 can not hold for $p = 1$, such an inequality does hold provided we replace $\mathbf{E}(X_n)_+$ in the bound by $\mathbf{E}[(X_n)_+ \log \min(X_n, 1)]$.

Exercise 5.2.14. *Consider the martingale $M_n = \prod_{k=1}^n Y_k$ for i.i.d. non-negative random variables $\{Y_k\}$ with $\mathbf{E}Y_1 = 1$ and $\mathbf{P}(Y_1 = 1) < 1$.*

(a) *Explain why $\mathbf{E}(\log Y_1)_+$ is finite and why the strong law of large numbers implies that $n^{-1}\log M_n \stackrel{a.s.}{\to} \mu < 0$ when $n \to \infty$.*
(b) *Deduce that $M_n \stackrel{a.s.}{\to} 0$ as $n \to \infty$ and that consequently $\{M_n\}$ is not uniformly integrable.*
(c) *Show that if (5.2.4) applies for $p = 1$ and some $q < \infty$, then any non-negative martingale would have been uniformly integrable.*

Exercise 5.2.15. *Show that if $\{X_n\}$ is a non-negative sub-MG then*

$$\mathbf{E}\big[\max_{k\le n} X_k\big] \le (1-e^{-1})^{-1}\{1 + \mathbf{E}[X_n(\log X_n)_+]\} .$$

Hint: *Apply part (c) of Lemma 1.4.31 and recall that $x(\log y)_+ \le e^{-1}y + x(\log x)_+$ for any $x, y \ge 0$.*

We just saw that in general L^1-bounded martingales might not be U.I. Nevertheless, as you show next, for sums of independent zero-mean random variables these two properties are equivalent.

Exercise 5.2.16. *Suppose $S_n = \sum_{k=1}^n \xi_k$ with ξ_k independent.*

(a) *Prove* Ottaviani's inequality. *Namely, show that for any n and $t, s \ge 0$,*

$$\mathbf{P}(\max_{k=1}^n |S_k| \ge t+s) \le \mathbf{P}(|S_n| \ge t) + \mathbf{P}(\max_{k=1}^n |S_k| \ge t+s) \max_{k=1}^n \mathbf{P}(|S_n - S_k| > s) .$$

(b) *Suppose further that $\{\xi_k\}$ is integrable and $\sup_n \mathbf{E}|S_n| < \infty$. Show that then $\mathbf{E}[\sup_k |S_k|]$ is finite.*

In the spirit of Doob's inequality bounding the tail probability of the maximum of a sub-MG $\{X_k, k = 0, 1, \ldots, n\}$ in terms of the value of X_n, we will bound the oscillations of $\{X_k, k = 0, 1, \ldots, n\}$ over an interval $[a, b]$ in terms of X_0 and X_n. To this end, we require the following definition of up-crossings.

Definition 5.2.17. *The number of* up-crossings *of the interval* $[a, b]$ *by* $\{X_k(\omega), k = 0, 1, \ldots, n\}$*, denoted* $U_n[a, b](\omega)$*, is the largest* $\ell \in \mathbb{Z}_+$ *such that* $X_{s_i}(\omega) < a$ *and* $X_{t_i}(\omega) > b$ *for* $1 \le i \le \ell$ *and some* $0 \le s_1 < t_1 < \cdots < s_\ell < t_\ell \le n$*.*

For example, Fig. 1 depicts two up-crossings of $[a, b]$.

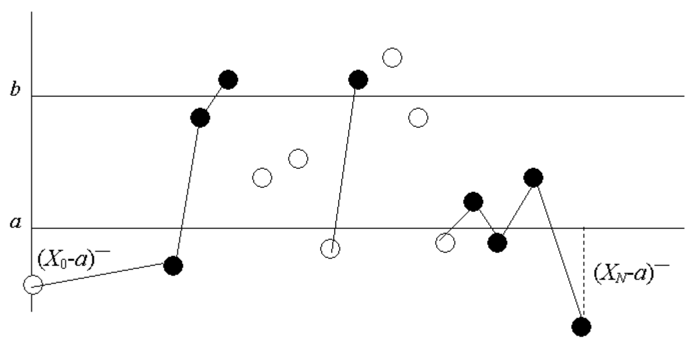

FIGURE 1. Illustration of up-crossings of $[a, b]$ by $X_k(\omega)$

Our next result, Doob's up-crossing inequality, is the key to the a.s. convergence of sup-MGs (and sub-MGs) on which Section 5.3 is based.

Lemma 5.2.18 (DOOB'S UP-CROSSING INEQUALITY)**.** *If* $\{X_n\}$ *is a sup-MG then*

$$(b - a)\mathbf{E}(U_n[a, b]) \le \mathbf{E}[(X_n - a)_-] - \mathbf{E}[(X_0 - a)_-] \qquad \forall a < b\ . \tag{5.2.6}$$

PROOF. Fixing $a < b$, let $V_1 = I_{\{X_0 < a\}}$ and for $n = 2, 3, \ldots$, define recursively $V_n = I_{\{V_{n-1}=1, X_{n-1} \le b\}} + I_{\{V_{n-1}=0, X_{n-1} < a\}}$. Informally, the sequence V_k is zero while waiting for the process $\{X_n\}$ to enter $(-\infty, a)$ after which time it reverts to one and stays so while waiting for this process to enter (b, ∞). See Figure 1 for an illustration in which black circles depict indices k such that $V_k = 1$ and open circles indicate those values of k with $V_k = 0$. Clearly, the sequence $\{V_n\}$ is predictable for the canonical filtration of $\{X_n\}$. Let $\{Y_n\}$ denote the martingale transform of $\{V_n\}$ with respect to $\{X_n\}$ (per Definition 5.1.27). By the choice of $V_\cdot$ every up-crossing of the interval $[a, b]$ by $\{X_k, k = 0, 1, \ldots, n\}$ contributes to Y_n the difference between the value of $X_\cdot$ at the end of the up-crossing (i.e. the last in the corresponding run of black circles), which is at least b and its value at the start of the up-crossing (i.e. the last in the preceding run of open circles), which is at most a. Thus, each up-crossing increases Y_n by at least $(b - a)$ and if $X_0 < a$ then the first up-crossing must have contributed at least $(b - X_0) = (b - a) + (X_0 - a)_-$ to Y_n. The only other contribution to Y_n is by the up-crossing of the interval $[a, b]$ that is in progress at time n (if there is such), and since it started at value at most a, its contribution to Y_n is at least $-(X_n - a)_-$. We thus conclude that

$$Y_n \geq (b - a)U_n[a, b] + (X_0 - a)_- - (X_n - a)_-$$

for all $\omega \in \Omega$. With $\{V_n\}$ predictable, bounded and non-negative it follows that $\{Y_n\}$ is a super-martingale (see parts (b) and (c) of Theorem 5.1.28). Thus, considering the expectation of the preceding inequality yields the up-crossing inequality (5.2.6) since $0 = \mathbf{E}Y_0 \geq \mathbf{E}Y_n$ for the sup-MG $\{Y_n\}$. $\square$

Doob's up-crossing inequality implies that the total number of up-crossings of $[a, b]$ by a non-negative sup-MG has a finite expectation. In this context, Dubins' up-crossing inequality, which you are to derive next, provides universal (i.e. depending only on a/b), exponential bounds on tail probabilities of this random variable.

Exercise 5.2.19. *Suppose $(X_n^1, \mathcal{F}_n)$ and $(X_n^2, \mathcal{F}_n)$ are both sup-MGs and τ is an $\mathcal{F}_n$-stopping time such that $X_\tau^1 \geq X_\tau^2$.*

(a) *Show that $W_n = X_n^1 I_{\tau > n} + X_n^2 I_{\tau \leq n}$ is a sup-MG with respect to $\mathcal{F}_n$ and deduce that so is $Y_n = X_n^1 I_{\tau \geq n} + X_n^2 I_{\tau < n}$ (this is sometimes called* the switching principle*).*

(b) *For a sup-MG $X_n \geq 0$ and constants $b > a > 0$ define the $\mathcal{F}_n^{\mathbf{X}}$-stopping times $\tau_0 = -1$, $\theta_\ell = \inf\{k > \tau_\ell : X_k \leq a\}$ and $\tau_{\ell+1} = \inf\{k > \theta_\ell : X_k \geq b\}$, $\ell = 0, 1, \ldots$. That is, the ℓ-th up-crossing of (a, b) by $\{X_n\}$ starts at $\theta_{\ell-1}$ and ends at τ_ℓ. For $\ell = 0, 1, \ldots$ let $Z_n = a^{-\ell} b^\ell$ when $n \in [\tau_\ell, \theta_\ell)$ and $Z_n = a^{-\ell-1} b^\ell X_n$ for $n \in [\theta_\ell, \tau_{\ell+1})$. Show that $(Z_n, \mathcal{F}_n^{\mathbf{X}})$ is a sup-MG.*

(c) *For $b > a > 0$ let $U_\infty[a, b]$ denote the total number of up-crossings of the interval $[a, b]$ by a non-negative super-martingale $\{X_n\}$. Deduce from the preceding that for any positive integer ℓ,*

$$\mathbf{P}(U_\infty[a, b] \geq \ell) \leq \left(\frac{a}{b}\right)^\ell \mathbf{E}[\min(X_0/a, 1)]$$

(this is Dubins' up-crossing inequality*).*

5.3. The convergence of Martingales

As we shall see in this section, a sub-MG (or a sup-MG), has an integrable limit under relatively mild integrability assumptions. For example, in this context L^1-boundedness (i.e. the finiteness of $\sup_n \mathbf{E}|X_n|$), yields a.s. convergence (see Doob's convergence theorem), while the L^1-convergence of $\{X_n\}$ is equivalent to the stronger hypothesis of uniform integrability of this process (see Theorem 5.3.12). Finally, the even stronger L^p-convergence applies for the smaller sub-class of L^p-bounded martingales (see Doob's L^p martingale convergence).

Indeed, these convergence results are closely related to the fact that the maximum and up-crossings counts of a sub-MG do not grow too rapidly (and same applies for sup-MGs and martingales). To further explore this direction, we next link the finiteness of the total number of up-crossings $U_\infty[a,b]$ of intervals $[a,b]$, $b > a$, by a process $\{X_n\}$ to its a.s. convergence.

Lemma 5.3.1. *If for each $b > a$ almost surely $U_\infty[a,b] < \infty$, then $X_n \stackrel{a.s}{\to} X_\infty$ where X_∞ is an $\overline{\mathbb{R}}$-valued random variable.*

PROOF. Note that the event that X_n has an almost sure ($\overline{\mathbb{R}}$-valued) limit as $n \to \infty$ is the complement of

$$\Gamma = \bigcup_{\substack{a,b\in\mathcal{Q}\\ a<b}} \Gamma_{a,b}\,,$$

where for each $b > a$,

$$\Gamma_{a,b} = \{\omega : \liminf_{n\to\infty} X_n(\omega) < a < b < \limsup_{n\to\infty} X_n(\omega)\}\,.$$

Since Γ is a countable union of these events, it thus suffices to show that $\mathbf{P}(\Gamma_{a,b}) = 0$ for any $a, b \in \mathcal{Q}$, $a < b$. To this end note that if $\omega \in \Gamma_{a,b}$ then $\limsup_n X_n(\omega) > b$ and $\liminf_n X_n(\omega) < a$ are both limit points of the sequence $\{X_n(\omega)\}$, hence the total number of up-crossings of the interval $[a,b]$ by this sequence is infinite. That is, $\Gamma_{a,b} \subseteq \{\omega : U_\infty[a,b](\omega) = \infty\}$. So, from our hypothesis that $U_\infty[a,b]$ is finite almost surely it follows that $\mathbf{P}(\Gamma_{a,b}) = 0$ for each $a < b$, resulting with the stated conclusion. □

Combining Doob's up-crossing inequality of Lemma 5.2.18 with Lemma 5.3.1 we now prove Doob's a.s. convergence theorem for sup-MGs (and sub-MGs).

Theorem 5.3.2 (DOOB'S CONVERGENCE THEOREM)**.** *Suppose sup-MG $(X_n, \mathcal{F}_n)$ is such that $\sup_n\{\mathbf{E}[(X_n)_-]\} < \infty$. Then, $X_n \stackrel{a.s.}{\to} X_\infty$ and $\mathbf{E}|X_\infty| \le \liminf_n \mathbf{E}|X_n|$ is finite.*

PROOF. Fixing $b > a$, recall that $0 \le U_n[a,b] \uparrow U_\infty[a,b]$ as $n \uparrow \infty$, where $U_\infty[a,b]$ denotes the total number of up-crossings of $[a,b]$ by the sequence $\{X_n\}$. Hence, by monotone convergence $\mathbf{E}(U_\infty[a,b]) = \sup_n \mathbf{E}(U_n[a,b])$. Further, with $(x-a)_- \le |a|+x_-$, we get from Doob's up-crossing inequality and the monotonicity of the expectation that

$$\mathbf{E}(U_n[a,b]) \le \frac{1}{(b-a)}\mathbf{E}(X_n-a)_- \le \frac{1}{(b-a)}\Big(|a| + \sup_n \mathbf{E}[(X_n)_-]\Big)\,.$$

Thus, our hypothesis that $\sup_n \mathbf{E}[(X_n)_-] < \infty$ implies that $\mathbf{E}(U_\infty[a,b])$ is finite, hence in particular $U_\infty[a,b]$ is finite almost surely.

Since this applies for any $b > a$, we have from Lemma 5.3.1 that $X_n \stackrel{a.s}{\to} X_\infty$. Further, with X_n a sup-MG, we have that $\mathbf{E}|X_n| = \mathbf{E}X_n + 2\mathbf{E}(X_n)_- \le \mathbf{E}X_0 + 2\mathbf{E}(X_n)_-$ for all n. Using this observation in conjunction with Fatou's lemma for $0 \le |X_n| \stackrel{a.s.}{\to} |X_\infty|$ and our hypothesis, we find that

$$\mathbf{E}|X_\infty| \le \liminf_{n\to\infty} \mathbf{E}|X_n| \le \mathbf{E}X_0 + 2\sup_n\{\mathbf{E}[(X_n)_-]\} < \infty\,,$$

as stated. □

Remark. In particular, Doob's convergence theorem implies that if $(X_n, \mathcal{F}_n)$ is a non-negative sup-MG then $X_n \stackrel{a.s.}{\to} X_\infty$ for some integrable X_∞ (and in this case $\mathbf{E}X_\infty \le \mathbf{E}X_0$). The same convergence applies for a non-positive sub-MG and more generally, for any sub-MG with $\sup_n\{\mathbf{E}(X_n)_+\} < \infty$. Further, the following exercise provides alternative equivalent conditions for the applicability of Doob's convergence theorem.

Exercise 5.3.3. *Show that the following five conditions are equivalent for any sub-MG $\{X_n\}$ (and if $\{X_n\}$ is a sup-MG, just replace $(X_n)_+$ by $(X_n)_-$).*

(a) $\lim_n \mathbf{E}|X_n|$ *exists and is finite.*
(b) $\sup_n \mathbf{E}|X_n| < \infty$.
(c) $\liminf_n \mathbf{E}|X_n| < \infty$.
(d) $\lim_n \mathbf{E}(X_n)_+$ *exists and is finite.*
(e) $\sup_n \mathbf{E}(X_n)_+ < \infty$.

Our first application of Doob's convergence theorem extends Doob's inequality (5.2.1) to the following bound on the maximal value of a U.I. sub-MG.

Corollary 5.3.4. *For any U.I. sub-MG $\{X_n\}$ and $x > 0$,*

$$\mathbf{P}(X_k \ge x \text{ for some } k < \infty) \le x^{-1}\mathbf{E}[X_\infty I_{\tau_x<\infty}] \le x^{-1}\mathbf{E}[(X_\infty)_+]\,, \tag{5.3.1}$$

where $\tau_x = \min\{k \ge 0 : X_k \ge x\}$.

Proof. Let $A_n = \{\tau_x \le n\} = \{\max_{k\le n} X_k \ge x\}$ and $A_\infty = \{\tau_x < \infty\} = \{X_k \ge x$ for some $k < \infty\}$. Then, $A_n \uparrow A_\infty$ and as the U.I. sub-MG $\{X_n\}$ is L^1-bounded, we have from Doob's convergence theorem that $X_n \stackrel{a.s.}{\to} X_\infty$. Consequently, $X_n I_{A_n}$ and $(X_n)_+$ converge almost surely to $X_\infty I_{A_\infty}$ and $(X_\infty)_+$, respectively. Since these two sequences are U.I. we further have that $\mathbf{E}[X_n I_{A_n}] \to \mathbf{E}[X_\infty I_{A_\infty}]$ and $\mathbf{E}[(X_n)_+] \to \mathbf{E}[(X_\infty)_+]$. Recall Doob's inequality (5.2.1) that

$$\mathbf{P}(A_n) \le x^{-1}\mathbf{E}[X_n I_{A_n}] \le x^{-1}\mathbf{E}[(X_n)_+] \tag{5.3.2}$$

for any n finite. Taking $n \to \infty$ we conclude that

$$\mathbf{P}(A_\infty) \le x^{-1}\mathbf{E}[X_\infty I_{A_\infty}] \le x^{-1}\mathbf{E}[(X_\infty)_+]$$

which is precisely our stated inequality (5.3.1). $\square$

Applying Doob's convergence theorem we also find that martingales of bounded differences either converge to a finite limit or oscillate between $-\infty$ and $+\infty$.

Proposition 5.3.5. *Suppose $\{X_n\}$ is a martingale of uniformly bounded differences. That is, almost surely $\sup_n |X_n - X_{n-1}| \le c$ for some finite non-random constant c. Then, $\mathbf{P}(A \cup B) = 1$ for the events*

$$A = \{\omega : \lim_{n\to\infty} X_n(\omega) \text{ exists and is finite}\},$$

$$B = \{\omega : \limsup_{n\to\infty} X_n(\omega) = \infty \text{ and } \liminf_{n\to\infty} X_n(\omega) = -\infty\}.$$

Proof. We may and shall assume without loss of generality that $X_0 = 0$ (otherwise, apply the proposition for the MG $Y_n = X_n - X_0$). Fixing a positive integer k, consider the stopping time $\tau_k(\omega) = \inf\{n \ge 0 : X_n(\omega) \le -k\}$ for the canonical filtration of $\{X_n\}$ and the associated stopped sup-MG $Y_n = X_{n\wedge\tau_k}$ (per Theorem 5.1.32). By definition of τ_k and our hypothesis of X_n having uniformly bounded differences, it follows that $Y_n(\omega) \ge -k - c$ for all n. Consequently,

$\sup_n \mathbf{E}(Y_n)_- \le k + c$ and by Doob's convergence theorem $Y_n(\omega) \to Y_\infty(\omega) \in \mathbb{R}$ for all $\omega \notin \Gamma_k$ and some measurable Γ_k such that $\mathbf{P}(\Gamma_k) = 0$. In particular, if $\tau_k(\omega) = \infty$ and $\omega \notin \Gamma_k$ then $X_n(\omega) = Y_n(\omega)$ has a finite limit, so $\omega \in A$. This shows that $A^c \subseteq \{\tau_k < \infty\} \cup \Gamma_k$ for all k, and hence $A^c \subseteq B_- \cup_k \Gamma_k$ where $B_- = \cap_k \{\tau_k < \infty\} = \{\omega : \liminf_n X_n(\omega) = -\infty\}$. With $\mathbf{P}(\Gamma_k) = 0$ for all k, we thus deduce that $\mathbf{P}(A \cup B_-) = 1$. Applying the preceding argument for the sup-MG $\{-X_n\}$ we find that $\mathbf{P}(A \cup B_+) = 1$ for $B_+ = \{\omega : \limsup_n X_n(\omega) = \infty\}$. Combining these two results we conclude that $\mathbf{P}(A \cup (B_- \cap B_+)) = 1$ as stated. $\square$

Remark. Consider a random walk $S_n = \sum_{k=1}^n \xi_k$ with zero-mean, bounded increments $\{\xi_k\}$ (i.e. $|\xi_k| \le c$ with c a finite non-random constant). Then, $v = \mathbf{E}\xi_k^2$ is finite and the event A where $S_n(\omega) \to S_\infty(\omega)$ as $n \to \infty$ for some $S_\infty(\omega)$ finite, implies that $\widehat{S}_n = (nv)^{-1/2} S_n \to 0$. Combining the CLT $\widehat{S}_n \xrightarrow{\mathcal{D}} G$ with Fatou's lemma and part (d) of the Portmanteau theorem we find that for any $\varepsilon > 0$,

$$\mathbf{P}(A) \le \mathbf{E}[\liminf_{n\to\infty} I_{|\widehat{S}_n| \le \varepsilon}] \le \liminf_{n\to\infty} \mathbf{P}(|\widehat{S}_n| \le \varepsilon) = \mathbf{P}(|G| \le \varepsilon)\,.$$

Taking $\varepsilon \downarrow 0$ we deduce that $\mathbf{P}(A) = 0$. Hence, by Proposition 5.3.5 such random walk is an example of a non-converging MG for which a.s.

$$\limsup_{n\to\infty} S_n = \infty = -\liminf_{n\to\infty} S_n\,.$$

Here is another application of the preceding proposition.

Exercise 5.3.6. *Consider the $\mathcal{F}_n$-adapted $W_n \ge 0$, such that $\sup_n |W_{n+1} - W_n| \le K$ for some finite non-random constant K and $W_0 = 0$. Suppose there exist non-random, positive constants a and b such that for all $n \ge 0$,*

$$\mathbf{E}[W_{n+1} - W_n + a|\mathcal{F}_n] I_{\{W_n \ge b\}} \le 0\,.$$

With $N_n = \sum_{k=1}^n I_{\{W_k < b\}}$, show that $\mathbf{P}(N_\infty$ is finite$) = 0$.
Hint: *Check that $X_n = W_n + an - (K+a)N_{n-1}$ is a sup-MG of uniformly bounded differences.*

As we show next, Doob's convergence theorem leads to the integrability of X_θ for any L^1 bounded sub-MG X_n and any stopping time θ.

Lemma 5.3.7. *If $(X_n, \mathcal{F}_n)$ is a sub-MG and $\sup_n \mathbf{E}[(X_n)_+] < \infty$ then $\mathbf{E}|X_\theta| < \infty$ for any $\mathcal{F}_n$-stopping time θ.*

Proof. Since $((X_n)_+, \mathcal{F}_n)$ is a sub-MG (see Proposition 5.1.22), it follows that $\mathbf{E}[(X_{n\wedge\theta})_+] \le \mathbf{E}[(X_n)_+]$ for all n (consider Theorem 5.1.32 for the sub-MG $(X_n)_+$ and $\tau = \infty$). Thus, our hypothesis that $\sup_n \mathbf{E}[(X_n)_+]$ is finite results with $\sup_n \mathbf{E}[(Y_n)_+]$ finite, where $Y_n = X_{n\wedge\theta}$. Applying Doob's convergence theorem for the sub-MG $(Y_n, \mathcal{F}_n)$ we have that $Y_n \stackrel{a.s}{\to} Y_\infty$ with $Y_\infty = X_\theta$ integrable. $\square$

We further get the following relation, which is key to establishing Doob's optional stopping for certain sup-MGs (and sub-MGs).

Proposition 5.3.8. *Suppose $(X_n, \mathcal{F}_n)$ is a non-negative sup-MG and $\tau \ge \theta$ are stopping times for the filtration $\{\mathcal{F}_n\}$. Then, $\mathbf{E}X_\theta \ge \mathbf{E}X_\tau$ are finite valued.*

PROOF. From Theorem 5.1.32 we know that $Z_n = X_{n\wedge\tau} - X_{n\wedge\theta}$ is a sup-MG (as are $X_{n\wedge\tau}$ and $X_{n\wedge\theta}$), with $Z_0 = 0$. Thus, $\mathbf{E}[X_{n\wedge\theta}] \geq \mathbf{E}[X_{n\wedge\tau}]$ are finite and since $\tau \geq \theta$, subtracting from both sides the finite $\mathbf{E}[X_n I_{\theta\geq n}]$ we find that

$$\mathbf{E}[X_\theta I_{\theta<n}] \geq \mathbf{E}[X_\tau I_{\tau<n}] + \mathbf{E}[X_n I_{\tau\geq n} I_{\theta<n}]\,.$$

The sup-MG $\{X_n\}$ is non-negative, so by Doob's convergence theorem $X_n \stackrel{a.s.}{\to} X_\infty$ and in view of Fatou's lemma

$$\liminf_{n\to\infty} \mathbf{E}[X_n I_{\tau\geq n} I_{\theta<n}] \geq \mathbf{E}[X_\infty I_{\tau=\infty} I_{\theta<\infty}]\,.$$

Further, by monotone convergence $\mathbf{E}[X_\tau I_{\tau<n}] \uparrow \mathbf{E}[X_\tau I_{\tau<\infty}]$ and $\mathbf{E}[X_\theta I_{\theta<n}] \uparrow \mathbf{E}[X_\theta I_{\theta<\infty}]$. Hence, taking $n \to \infty$ results with

$$\mathbf{E}[X_\theta I_{\theta<\infty}] \geq \mathbf{E}[X_\tau I_{\tau<\infty}] + \mathbf{E}[X_\tau I_{\tau=\infty} I_{\theta<\infty}]\,.$$

Adding the identity $\mathbf{E}[X_\theta I_{\theta=\infty}] = \mathbf{E}[X_\tau I_{\theta=\infty}]$, which holds for $\tau \geq \theta$, yields the stated inequality $\mathbf{E}[X_\theta] \geq \mathbf{E}[X_\tau]$. Considering $0 \leq \theta$ we further see that $\mathbf{E}[X_0] \geq \mathbf{E}[X_\theta] \geq \mathbf{E}[X_\tau] \geq 0$ are finite, as claimed. □

Solving the next exercise should improve your intuition about the domain of validity of Proposition 5.1.22 and of Doob's convergence theorem.

Exercise 5.3.9.

(a) *Provide an example of a sub-martingale $\{X_n\}$ for which $\{X_n^2\}$ is a super-martingale and explain why it does not contradict Proposition 5.1.22.*

(b) *Provide an example of a martingale which converges a.s. to $-\infty$ and explain why it does not contradict Theorem 5.3.2.*
Hint: *Try $S_n = \sum_{i=1}^n \xi_i$, with zero-mean, independent but not identically distributed ξ_i.*

We conclude this sub-section with few additional applications of Doob's convergence theorem.

Exercise 5.3.10. *Suppose $\{X_n\}$ and $\{Y_n\}$ are non-negative, integrable processes adapted to the filtration $\mathcal{F}_n$ such that $\sum_{n\geq 1} Y_n < \infty$ a.s. and $\mathbf{E}[X_{n+1}|\mathcal{F}_n] \leq (1+Y_n)X_n + Y_n$ for all n. Show that X_n converges a.s. to a finite limit as $n \to \infty$.*
Hint: *Find a non-negative super-martingale $(W_n, \mathcal{F}_n)$ whose convergence implies that of X_n.*

Exercise 5.3.11. *Let $\{X_k\}$ be mutually independent but not necessarily integrable random variables, such that*

(a) *Fixing $c < \infty$ non-random, let $Y_n^{(c)} = \sum_{k=1}^n |S_{k-1}| I_{|S_{k-1}|\leq c} X_k I_{|X_k|\leq c}$. Show that $Y_n^{(c)}$ is a martingale with respect to the filtration $\{\mathcal{F}_n^{\mathbf{X}}\}$ and that $\sup_n \|Y_n^{(c)}\|_2 < \infty$.*
Hint: *Kolmogorov's three series theorem may help in proving that $\{Y_n^{(c)}\}$ is L^2-bounded.*

(b) *Show that $Y_n = \sum_{k=1}^n |S_{k-1}| X_k$ converges a.s.*

5.3.1. Uniformly integrable martingales. The main result of this subsection is the following L^1 convergence theorem for *uniformly integrable* (U.I.) sub-MGs (and sup-MGs).

Theorem 5.3.12. *If $(X_n, \mathcal{F}_n)$ is a sub-MG, then $\{X_n\}$ is U.I. (c.f. Definition 1.3.47), if and only if $X_n \stackrel{L^1}{\to} X_\infty$, in which case also $X_n \stackrel{a.s.}{\to} X_\infty$ and $X_n \le \mathbf{E}[X_\infty|\mathcal{F}_n]$ for all n.*

Remark. If $\{X_n\}$ is uniformly integrable then $\sup_n \mathbf{E}|X_n|$ is finite (see Lemma 1.3.48). Thus, the assumption of Theorem 5.3.12 is stronger than that of Theorem 5.3.2, as is its conclusion.

Proof. If $\{X_n\}$ is U.I. then $\sup_n \mathbf{E}|X_n| < \infty$. For $\{X_n\}$ sub-MG it thus follows by Doob's convergence theorem that $X_n \stackrel{a.s.}{\to} X_\infty$ with X_∞ integrable. Obviously, this implies that $X_n \stackrel{p}{\to} X_\infty$. Similarly, if we start instead by assuming that $X_n \stackrel{L^1}{\to} X_\infty$ then also $X_n \stackrel{p}{\to} X_\infty$. Either way, Vitali's convergence theorem (i.e. Theorem 1.3.49), tells us that uniform integrability is equivalent to L^1 convergence when $X_n \stackrel{p}{\to} X_\infty$. We thus deduce that for sub-MGs the U.I. property is equivalent to L^1 convergence and either one of these yields also the corresponding a.s. convergence.

Turning to show that $X_n \le \mathbf{E}[X_\infty|\mathcal{F}_n]$ for all n, recall that $X_m \le \mathbf{E}[X_\ell|\mathcal{F}_m]$ for all $\ell > m$ and any sub-MG (see Proposition 5.1.20). Further, since $X_\ell \stackrel{L^1}{\to} X_\infty$ it follows that $\mathbf{E}[X_\ell|\mathcal{F}_m] \stackrel{L^1}{\to} \mathbf{E}[X_\infty|\mathcal{F}_m]$ as $\ell \to \infty$, per fixed m (see Theorem 4.2.30). The latter implies the convergence a.s. of these conditional expectations along some sub-sequence ℓ_k (c.f. Theorem 2.2.10). Hence, we conclude that for any m, a.s.

$$X_m \le \liminf_{\ell\to\infty} \mathbf{E}[X_\ell|\mathcal{F}_m] \le \mathbf{E}[X_\infty|\mathcal{F}_m]\,,$$

i.e., $X_n \le \mathbf{E}[X_\infty|\mathcal{F}_n]$ for all n. $\square$

The preceding theorem identifies the collection of U.I. martingales as merely the set of all Doob's martingales, a concept we now define.

Definition 5.3.13. *The sequence $X_n = \mathbf{E}[X|\mathcal{F}_n]$ with X an integrable R.V. and $\{\mathcal{F}_n\}$ a filtration, is called* Doob's martingale *of X with respect to $\{\mathcal{F}_n\}$.*

Corollary 5.3.14. *A martingale $(X_n, \mathcal{F}_n)$ is U.I. if and only if $X_n = \mathbf{E}[X_\infty|\mathcal{F}_n]$ is a Doob's martingale with respect to $\{\mathcal{F}_n\}$, or equivalently if and only if $X_n \stackrel{L^1}{\to} X_\infty$.*

Proof. Theorem 5.3.12 states that a sub-MG (hence also a MG) is U.I. if and only if it converges in L^1 and in this case $X_n \le \mathbf{E}[X_\infty|\mathcal{F}_n]$. Applying this theorem also for $-X_n$ we deduce that a U.I. martingale is necessarily a Doob's martingale of the form $X_n = \mathbf{E}[X_\infty|\mathcal{F}_n]$. Conversely, the sequence $X_n = \mathbf{E}[X|\mathcal{F}_n]$ for some integrable X and a filtration $\{\mathcal{F}_n\}$ is U.I. (see Proposition 4.2.33). $\square$

We next generalize Theorem 4.2.26 about dominated convergence of C.E.

Theorem 5.3.15 (Lévy's upward theorem)**.** *Suppose $\sup_m |X_m|$ is integrable, $X_n \stackrel{a.s.}{\to} X_\infty$ and $\mathcal{F}_n \uparrow \mathcal{F}_\infty$. Then $\mathbf{E}[X_n|\mathcal{F}_n] \to \mathbf{E}[X_\infty|\mathcal{F}_\infty]$ both a.s. and in L^1.*

Remark. Lévy's upward theorem is trivial if $\{X_n\}$ is adapted to $\{\mathcal{F}_n\}$ (which is obviously not part of its assumptions). On the other hand, recall that in view of part (b) of Exercise 4.2.35, having $\{X_n\}$ U.I. and $X_n \stackrel{a.s.}{\to} X_\infty$ is in general not enough even for the a.s. convergence of $\mathbf{E}[X_n|\mathcal{G}]$ to $\mathbf{E}[X_\infty|\mathcal{G}]$.

PROOF. Consider first the special case where $X_n = X$ does not depend on n. Then, $Y_n = \mathbf{E}[X|\mathcal{F}_n]$ is a U.I. martingale. Therefore, $\mathbf{E}[Y_\infty|\mathcal{F}_n] = \mathbf{E}[X|\mathcal{F}_n]$ for all n, where Y_∞ denotes the a.s. and L^1 limit of Y_n (see Corollary 5.3.14). As $Y_n \in m\mathcal{F}_n \subseteq m\mathcal{F}_\infty$ clearly $Y_\infty = \lim_n Y_n \in m\mathcal{F}_\infty$. Further, by definition of the C.E. $\mathbf{E}[XI_A] = \mathbf{E}[Y_\infty I_A]$ for all A in the π-system $\mathcal{P} = \bigcup_n \mathcal{F}_n$ hence with $\mathcal{F}_\infty = \sigma(\mathcal{P})$ it follows that $Y_\infty = \mathbf{E}[X|\mathcal{F}_\infty]$ (see Exercise 4.1.3).

Turning to the general case, with $Z = \sup_m |X_m|$ integrable and $X_m \stackrel{a.s.}{\to} X_\infty$, we deduce that X_∞ and $W_k = \sup\{|X_n - X_\infty| : n \geq k\} \leq 2Z$ are both integrable. So, the conditional Jensen's inequality and the monotonicity of the C.E. imply that for all $n \geq k$,

$$|\mathbf{E}[X_n|\mathcal{F}_n] - \mathbf{E}[X_\infty|\mathcal{F}_n]| \leq \mathbf{E}[|X_n - X_\infty|\,|\mathcal{F}_n] \leq \mathbf{E}[W_k|\mathcal{F}_n]\,.$$

Consequently, considering $n \to \infty$ we find by the special case of the theorem where X_n is replaced by W_k independent of n (which we already proved), that

$$\limsup_{n\to\infty} |\mathbf{E}[X_n|\mathcal{F}_n] - \mathbf{E}[X_\infty|\mathcal{F}_n]| \leq \lim_{n\to\infty} \mathbf{E}[W_k|\mathcal{F}_n] = \mathbf{E}[W_k|\mathcal{F}_\infty]\,.$$

Similarly, we know that $\mathbf{E}[X_\infty|\mathcal{F}_n] \stackrel{a.s}{\to} \mathbf{E}[X_\infty|\mathcal{F}_\infty]$. Further, by definition $W_k \downarrow 0$ a.s. when $k \to \infty$, so also $\mathbf{E}[W_k|\mathcal{F}_\infty] \downarrow 0$ by the usual dominated convergence of C.E. (see Theorem 4.2.26). Combining these two a.s. convergence results and the preceding inequality, we deduce that $\mathbf{E}[X_n|\mathcal{F}_n] \stackrel{a.s}{\to} \mathbf{E}[X_\infty|\mathcal{F}_\infty]$ as stated. Finally, since $|\mathbf{E}[X_n|\mathcal{F}_n]| \leq \mathbf{E}[Z|\mathcal{F}_n]$ for all n, it follows that $\{\mathbf{E}[X_n|\mathcal{F}_n]\}$ is U.I. and hence the a.s. convergence of this sequence to $\mathbf{E}[X_\infty|\mathcal{F}_\infty]$ yields its convergence in L^1 as well (c.f. Theorem 1.3.49). $\square$

Considering Lévy's upward theorem for $X_n = X_\infty = I_A$ and $A \in \mathcal{F}_\infty$ yields the following corollary.

Corollary 5.3.16 (LÉVY'S 0-1 LAW)**.** *If $\mathcal{F}_n \uparrow \mathcal{F}_\infty$, $A \in \mathcal{F}_\infty$, then $\mathbf{E}[I_A|\mathcal{F}_n] \stackrel{a.s.}{\to} I_A$.*

As shown in the sequel, Kolmogorov's 0-1 law about $\mathbf{P}$-triviality of the tail σ-algebra $\mathcal{T}^{\mathbf{X}} = \cap_n \mathcal{T}_n^{\mathbf{X}}$ of independent random variables is a special case of Lévy's 0-1 law.

PROOF OF COROLLARY 1.4.10. Let $\mathcal{F}^{\mathbf{X}} = \sigma(\bigcup_n \mathcal{F}_n^{\mathbf{X}})$. Recall Definition 1.4.9 that $\mathcal{T}^{\mathbf{X}} \subseteq \mathcal{T}_n^{\mathbf{X}} \subseteq \mathcal{F}^{\mathbf{X}}$ for all n. Thus, by Lévy's 0-1 law $\mathbf{E}[I_A|\mathcal{F}_n^{\mathbf{X}}] \stackrel{a.s.}{\to} I_A$ for any $A \in \mathcal{T}^{\mathbf{X}}$. By assumption $\{X_k\}$ are $\mathbf{P}$-mutually independent, hence for any $A \in \mathcal{T}^{\mathbf{X}}$ the R.V. $I_A \in m\mathcal{T}_n^{\mathbf{X}}$ is independent of the σ-algebra $\mathcal{F}_n^{\mathbf{X}}$. Consequently, $\mathbf{E}[I_A|\mathcal{F}_n^{\mathbf{X}}] \stackrel{a.s.}{=} \mathbf{P}(A)$ for all n. We deduce that $\mathbf{P}(A) \stackrel{a.s.}{=} I_A$, implying that $\mathbf{P}(A) \in \{0,1\}$ for all $A \in \mathcal{T}^{\mathbf{X}}$, as stated. $\square$

The generalization of Theorem 4.2.30 which you derive next also relaxes the assumptions of Lévy's upward theorem in case only L^1 convergence is of interest.

Exercise 5.3.17. *Show that if $X_n \stackrel{L^1}{\to} X_\infty$ and $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ then $\mathbf{E}[X_n|\mathcal{F}_n] \stackrel{L^1}{\to} \mathbf{E}[X_\infty|\mathcal{F}_\infty]$.*

Here is an example of the importance of uniform integrability when dealing with convergence.

Exercise 5.3.18. *Suppose $X_n \stackrel{a.s.}{\to} 0$ are $[0,1]$-valued random variables and $\{M_n\}$ is a non-negative MG.*

(a) *Provide an example where $\mathbf{E}[X_n M_n] = 1$ for all n finite.*

(b) *Show that if* $\{M_n\}$ *is U.I. then* $\mathbf{E}[X_n M_n] \to 0$.

Definition 5.3.19. *A continuous function* $x : [0,1) \mapsto \mathbb{R}$ *is* absolutely continuous *if for every* $\varepsilon > 0$ *there exists* $\delta > 0$ *such that for all* $k < \infty$, $s_1 < t_1 \leq s_2 < t_2 \leq \cdots \leq s_k < t_k \in [0,1)$

$$\sum_{\ell=1}^{k} |t_\ell - s_\ell| \leq \delta \qquad \Longrightarrow \qquad \sum_{\ell=1}^{k} |x(t_\ell) - x(s_\ell)| \leq \varepsilon.$$

The next exercise uses convergence properties of MGs to prove a classical result in real analysis, namely, that an absolutely continuous function is differentiable for Lebesgue a.e. $t \in [0,1)$.

Exercise 5.3.20. *On the probability space* $([0,1), \mathcal{B}, U)$ *consider the events*

$$A_{i,n} = [(i-1)2^{-n}, i2^{-n}) \quad \textit{for} \quad i = 1, \ldots, 2^n, \quad n = 0, 1, \ldots,$$

and the associated σ*-algebras* $\mathcal{F}_n = \sigma(A_{i,n}, i = 1, \ldots, 2^n)$.

(a) *Write an explicit formula for* $\mathbf{E}[h|\mathcal{F}_n]$ *and* $h \in L^1([0,1), \mathcal{B}, U)$.

(b) *For* $h_{i,n} = 2^n(x(i2^{-n}) - x((i-1)2^{-n}))$, *show that* $X_n(t) = \sum_{i=1}^{2^n} h_{i,n} I_{A_{i,n}}(t)$ *is a martingale with respect to* $\{\mathcal{F}_n\}$.

(c) *Show that for absolutely continuous* $x(\cdot)$ *the martingale* $\{X_n\}$ *is U.I.*
Hint: *Show that* $\mathbf{P}(|X_n| > \rho) \leq c/\rho$ *for some constant* $c < \infty$ *and all* $n, \rho > 0$.

(d) *Show that then there exists* $h \in L^1([0,1), \mathcal{B}, U)$ *such that*

$$x(t) - x(s) = \int_s^t h(u)du \qquad \textit{for all} \qquad 1 > t \geq s \geq 0.$$

(e) *Recall Lebesgue's theorem, that* $\Delta^{-1} \int_s^{s+\Delta} |h(s) - h(u)| du \stackrel{a.s}{\to} 0$ *as* $\Delta \to 0$, *for a.e.* $s \in [0,1)$. *Using it, conclude that* $\frac{dx}{dt} = h$ *for almost every* $t \in [0,1)$.

Recall that uniformly bounded p-th moment for some $p > 1$ implies U.I. (see Exercise 1.3.54). Strengthening the L^1 convergence of Theorem 5.3.12, the next proposition shows that an L^p-bounded martingale converges to its a.s. limit also in L^p (provided $p > 1$). In contrast to the preceding convergence results, this one *does not hold* for sub-MGs (or sup-MGs) which are not MGs (for example, let $\tau = \inf\{k \geq 1 : \xi_k = 0\}$ for independent $\{\xi_k\}$ such that $\mathbf{P}(\xi_k \neq 0) = k^2/(k+1)^2$, so $\mathbf{P}(\tau \geq n) = n^{-2}$ and verify that $X_n = nI_{\{n \leq \tau\}} \stackrel{a.s.}{\to} 0$ but $\mathbf{E}X_n^2 = 1$, so this L^2-bounded sup-MG does not converge to zero in L^2).

Proposition 5.3.21 (Doob's L^p martingale convergence)**.** *If the MG* $\{X_n\}$ *is such that* $\sup_n \mathbf{E}|X_n|^p < \infty$ *for some* $p > 1$, *then there exists a R.V.* X_∞ *such that* $X_n \to X_\infty$ *almost surely and in* L^p *(so* $\|X_n\|_p \to \|X_\infty\|_p$*).*

Proof. Being L^p bounded, the MG $\{X_n\}$ is L^1 bounded and Doob's martingale convergence theorem applies here, so $X_n \stackrel{a.s.}{\to} X_\infty$ for some integrable R.V. X_∞. Further, applying Fatou's lemma for $|X_n|^p \geq 0$ we have that,

$$\liminf_{n\to\infty} \mathbf{E}(|X_n|^p) \geq \mathbf{E}(\liminf_{n\to\infty} |X_n|^p) = \mathbf{E}|X_\infty|^p$$

as claimed, with $X_\infty \in L^p$. It thus suffices to verify that $\mathbf{E}|X_n - X_\infty|^p \to 0$ as $n \to \infty$ (as in Exercise 1.3.28, this implies that $\|X_n\|_p \to \|X_\infty\|_p$). To this end, with

$c = \sup_n \mathbf{E}|X_n|^p$ finite we have by the L^p maximal inequality of (5.2.5) that $\mathbf{E}Z_n \leq q^p c$ for $Z_n = \max_{k\leq n} |X_k|^p$ and any finite n. Since $0 \leq Z_n \uparrow Z = \sup_{k<\infty} |X_k|^p$, we have by monotone convergence that $\mathbf{E}Z \leq q^p c$ is finite. As X_∞ is the a.s. limit of X_n it follows that $|X_\infty|^p \leq Z$ as well. Hence, $Y_n = |X_n - X_\infty|^p \leq (|X_n|+|X_\infty|)^p \leq 2^p Z$. With $Y_n \overset{a.s.}{\to} 0$ and $Y_n \leq 2^p Z$ for integrable Z, we deduce by dominated convergence that $\mathbf{E}Y_n \to 0$ as $n \to \infty$, thus completing the proof of the proposition. $\square$

Remark. Proposition 5.3.21 does not have an L^1 analog. Indeed, as we have seen already in Exercise 5.2.14, there exists a non-negative MG $\{M_n\}$ such that $\mathbf{E}M_n = 1$ for all n and $M_n \to M_\infty = 0$ almost surely, so obviously, M_n does not converge to M_∞ in L^1.

Example 5.3.22. *Consider the martingale $S_n = \sum_{k=1}^n \xi_k$ for independent, square-integrable, zero-mean random variables ξ_k such that $\sum_k \mathbf{E}\xi_k^2 < \infty$. Since $\mathbf{E}S_n^2 = \sum_{k=1}^n \mathbf{E}\xi_k^2$, it follows from Proposition 5.3.21 that the random series $S_n(\omega) \to S_\infty(\omega)$ almost surely and in L^2 (see also Theorem 2.3.17 for a direct proof of this result, based on Kolmogorov's maximal inequality).*

Exercise 5.3.23. *Suppose $Z_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n \xi_k$ for i.i.d. $\xi_k \in L^2(\Omega, \mathcal{F}, \mathbf{P})$ of zero-mean and unit variance. Let $\mathcal{F}_n = \sigma(\xi_k, k \leq n)$ and $\mathcal{F}_\infty = \sigma(\xi_k, k < \infty)$.*

(a) *Prove that $\mathbf{E}WZ_n \to 0$ for any fixed $W \in L^2(\Omega, \mathcal{F}_\infty, \mathbf{P})$.*

(b) *Deduce that the same applies for any $W \in L^2(\Omega, \mathcal{F}, \mathbf{P})$ and conclude that Z_n does not converge in L^2.*

(c) *Show that though $Z_n \xrightarrow{\mathcal{D}} G$, a standard normal variable, there exists no $Z_\infty \in m\mathcal{F}$ such that $Z_n \overset{p}{\to} Z_\infty$.*

We conclude this sub-section with the application of martingales to the study of Pólya's urn scheme.

Example 5.3.24 (Pólya's urn)**.** *Consider an urn that initially contains r red and b blue marbles. At the k-th step a marble is drawn at random from the urn, with all possible choices being equally likely, and it and c_k more marbles of the same color are then returned to the urn. With $N_n = r+b+\sum_{k=1}^n c_k$ counting the number of marbles in the urn after n iterations of this procedure, let R_n denote the number of red marbles at that time and $M_n = R_n/N_n$ the corresponding fraction of red marbles. Since $R_{n+1} \in \{R_n, R_n+c_n\}$ with $\mathbf{P}(R_{n+1} = R_n+c_n|\mathcal{F}_n^{\mathbf{M}}) = R_n/N_n = M_n$ it follows that $\mathbf{E}[R_{n+1}|\mathcal{F}_n^{\mathbf{M}}] = R_n + c_n M_n = N_{n+1}M_n$. Consequently, $\mathbf{E}[M_{n+1}|\mathcal{F}_n^{\mathbf{M}}] = M_n$ for all n with $\{M_n\}$ a uniformly bounded martingale.*

For the study of Pólya's urn scheme we need the following definition.

Definition 5.3.25. *The* beta density *with parameters $\alpha > 0$ and $\beta > 0$ is*

$$f_\beta(u) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} u^{\alpha-1}(1-u)^{\beta-1}\mathbf{1}_{u\in[0,1]}\,,$$

where $\Gamma(\alpha) = \int_0^\infty s^{\alpha-1}e^{-s}ds$ is finite and positive (compare with Definition 1.4.45). In particular, $\alpha = \beta = 1$ corresponds to the density $f_U(u)$ of the uniform measure on $(0, 1]$, as in Example 1.2.41.

Exercise 5.3.26. *Let $\{M_n\}$ be the martingale of Example 5.3.24.*

(a) *Show that $M_n \to M_\infty$ a.s. and in L^p for any $p > 1$.*

(b) *Assuming further that $c_k = c$ for all $k \geq 1$, show that for $\ell = 0, \ldots, n$,*

$$\mathbf{P}(R_n = r + \ell c) = \binom{n}{\ell} \frac{\prod_{i=0}^{\ell-1}(r+ic)\prod_{j=0}^{n-\ell-1}(b+jc)}{\prod_{k=0}^{n-1}(r+b+kc)},$$

and deduce that M_∞ has the beta density with parameters $\alpha = b/c$ and $\beta = r/c$ (in particular, M_∞ has the law of $U(0,1]$ when $r = b = c_k > 0$).

(c) *For $r = b = c_k > 0$ show that $\mathbf{P}(\sup_{k\geq 1} M_k > 3/4) \leq 2/3$.*

Exercise 5.3.27 (Bernard Friedman's urn)**.** *Consider the following variant of Pólya's urn scheme, where after the k-th step one returns to the urn in addition to the marble drawn and c_k marbles of its color, also $d_k \geq 1$ marbles of the opposite color. Show that if c_k, d_k are uniformly bounded and $r + b > 0$, then $M_n \stackrel{a.s.}{\to} 1/2$.*
Hint: *With $X_n = (M_n - 1/2)^2$ check that $\mathbf{E}[X_n|\mathcal{F}^{\mathbf{M}}_{n-1}] = (1-a_n)X_{n-1} + u_n$, where the non-negative constants a_n and u_n are such that $\sum_k u_k < \infty$ and $\sum_k a_k = \infty$.*

Exercise 5.3.28. *Fixing $b_n \in [\delta, 1]$ for some $\delta > 0$, suppose $\{X_n\}$ are $[0,1]$-valued, $\mathcal{F}_n$-adapted such that $X_{n+1} = (1-b_n)X_n + b_n B_n$, $n \geq 0$, and $\mathbf{P}(B_n = 1|\mathcal{F}_n) = 1 - \mathbf{P}(B_n = 0|\mathcal{F}_n) = X_n$. Show that $X_n \stackrel{a.s.}{\to} X_\infty \in \{0,1\}$ and $\mathbf{P}(X_\infty = 1|\mathcal{F}_0) = X_0$.*

5.3.2. Square-integrable martingales. If $(X_n, \mathcal{F}_n)$ is a square-integrable martingale then $(X_n^2, \mathcal{F}_n)$ is a sub-MG, so by Doob's decomposition $X_n^2 = M_n + A_n$ for a non-decreasing $\mathcal{F}_n$-predictable sequence $\{A_n\}$ and a MG $(M_n, \mathcal{F}_n)$ with $M_0 = 0$. In the course of proving Doob's decomposition we saw that $A_n - A_{n-1} = \mathbf{E}[X_n^2 - X_{n-1}^2|\mathcal{F}_{n-1}]$ and part (b) of Exercise 5.1.8 provides an alternative expression $A_n - A_{n-1} = \mathbf{E}[(X_n - X_{n-1})^2|\mathcal{F}_{n-1}]$, motivating the following definition.

Definition 5.3.29. *The sequence $A_n = X_0^2 + \sum_{k=1}^n \mathbf{E}[(X_k - X_{k-1})^2|\mathcal{F}_{k-1}]$ is called the* predictable compensator *of an L^2-martingale $(X_n, \mathcal{F}_n)$ and denoted by* angle-brackets*, that is, $A_n = \langle X\rangle_n$.*

With $\mathbf{E}M_n = \mathbf{E}M_0 = 0$ it follows that $\mathbf{E}X_n^2 = \mathbf{E}\langle X\rangle_n$, so $\{X_n\}$ is L^2-bounded if and only if $\sup_n \mathbf{E}\langle X\rangle_n$ is finite. Further, $\langle X\rangle_n$ is non-decreasing, so it converges to a limit, denoted hereafter $\langle X\rangle_\infty$. With $\langle X\rangle_n \geq \langle X\rangle_0 = X_0^2$ integrable it further follows by monotone convergence that $\mathbf{E}\langle X\rangle_n \uparrow \mathbf{E}\langle X\rangle_\infty$, so $\{X_n\}$ is L^2 bounded if and only if $\langle X\rangle_\infty$ is integrable, in which case X_n converges a.s. and in L^2 (see Doob's L^2 convergence theorem). As we show in the sequel much more can be said about the relation between convergence of $X_n(\omega)$ and the random variable $\langle X\rangle_\infty$. To simplify the notations assume hereafter that $X_0 = 0 = \langle X\rangle_0$ so $\langle X\rangle_n \geq 0$ (the transformation of our results to the general case is trivial).

We start with the following explicit bounds on $\mathbf{E}[\sup_n |X_n|^p]$ for $p \leq 2$, from which we deduce that $\{X_n\}$ is U.I. (hence converges a.s. and in L^1), whenever $\langle X\rangle_\infty^{1/2}$ is integrable.

Proposition 5.3.30. *There exist finite constants c_q, $q \in (0,1]$, such that if $(X_n, \mathcal{F}_n)$ is an L^2-MG with $X_0 = 0$, then*

$$\mathbf{E}[\sup_k |X_k|^{2q}] \leq c_q \mathbf{E}[\,\langle X\rangle_\infty^q\,].$$

Remark. Our proof gives $c_q = (2-q)/(1-q)$ for $q < 1$ and $c_1 = 4$.

Proof. Let $V_n = \max_{k=0}^n |X_k|^2$, noting that $V_n \uparrow V_\infty = \sup_k |X_k|^2$ when $n \to \infty$. As already explained $\mathbf{E}X_n^2 \uparrow \mathbf{E}\langle X\rangle_\infty$ for $n \to \infty$. Thus, applying the bound (5.2.5) of Corollary 5.2.13 for $p = 2$ we find that

$$\mathbf{E}[V_n] \leq 4\mathbf{E}X_n^2 \leq 4\mathbf{E}\langle X\rangle_\infty \,,$$

and considering $n \to \infty$ we get our thesis for $q = 1$ (by monotone convergence).

Turning to the case of $0 < q < 1$, note that $(V_\infty)^q = \sup_k |X_k|^{2q}$. Further, the $\mathcal{F}_n$-predictable part in Doob's decomposition of the non-negative sub-martingale $Z_n = X_n^2$ is $A_n = \langle X\rangle_n$. Hence, applying Lemma 5.2.7 with $p = q$ and $\tau = \infty$ yields the stated bound. $\square$

Here is an application of Proposition 5.3.30 to the study of a certain class of random walks.

Exercise 5.3.31. *Let $S_n = \sum_{k=1}^n \xi_k$ for i.i.d. $\{\xi_k\}$ of zero mean and finite second moment. Suppose τ is an $\mathcal{F}_n^{\boldsymbol{\xi}}$-stopping time such that $\mathbf{E}[\sqrt{\tau}]$ is finite.*

(a) *Compute the predictable compensator of the L^2-martingale $(S_{n\wedge\tau}, \mathcal{F}_n^{\boldsymbol{\xi}})$.*
(b) *Deduce that $\{S_{n\wedge\tau}\}$ is U.I. and that $\mathbf{E}S_\tau = 0$.*

We deduce from Proposition 5.3.30 that $X_n(\omega)$ converges a.s. to a finite limit when $\langle X\rangle_\infty^{1/2}$ is integrable. A considerable refinement of this conclusion is offered by our next result, relating such convergence to $\langle X\rangle_\infty$ being finite *at ω!*

Theorem 5.3.32. *Suppose $(X_n, \mathcal{F}_n)$ is an L^2 martingale with $X_0 = 0$.*

(a) *$X_n(\omega)$ converges to a finite limit for a.e. ω for which $\langle X\rangle_\infty(\omega)$ is finite.*
(b) *$X_n(\omega)/\langle X\rangle_n(\omega) \to 0$ for a.e. ω for which $\langle X\rangle_\infty(\omega)$ is infinite.*
(c) *If the martingale differences $X_n - X_{n-1}$ are uniformly bounded then the converse to part (a) holds. That is, $\langle X\rangle_\infty(\omega)$ is finite for a.e. ω for which $X_n(\omega)$ converges to a finite limit.*

Proof. (a) Recall that for any n and $\mathcal{F}_n$-stopping time τ we have the identity $X_{n\wedge\tau}^2 = M_{n\wedge\tau} + \langle X\rangle_{n\wedge\tau}$ with $\mathbf{E}M_{n\wedge\tau} = 0$, resulting with $\sup_n \mathbf{E}[X_{n\wedge\tau}^2] = \mathbf{E}\langle X\rangle_\tau$. For example, see the proof of Proposition 5.3.30, where we also noted that $\theta_v = \min\{n \geq 0 : \langle X\rangle_{n+1} > v\}$ are $\mathcal{F}_n$-stopping times such that $\langle X\rangle_{\theta_v} \leq v$. Thus, setting $Y_n = X_{n\wedge\theta_k}$ for a positive integer k, the martingale $(Y_n, \mathcal{F}_n)$ is L^2-bounded and as such it almost surely has a finite limit. Further, we saw there that if $\langle X\rangle_\infty(\omega)$ is finite then $\theta_k(\omega) = \infty$ for some random positive integer $k = k(\omega)$, in which case $X_{n\wedge\theta_k} = X_n$ for all n. Since we consider only countably many values of k, this yields the thesis of part (a) of the theorem.

(b). Since $V_n = (1 + \langle X\rangle_n)^{-1}$ is an $\mathcal{F}_n$-predictable sequence of bounded variables, its martingale transform $Y_n = \sum_{k=1}^n V_k(X_k - X_{k-1})$ with respect to the square-integrable martingale $\{X_n\}$ is also a square-integrable martingale for the filtration $\{\mathcal{F}_n\}$ (c.f. Theorem 5.1.28). Further, since $V_k \in m\mathcal{F}_{k-1}$ it follows that for all $k \geq 1$,

$$\begin{aligned}\langle Y\rangle_k - \langle Y\rangle_{k-1} &= \mathbf{E}[(Y_k - Y_{k-1})^2|\mathcal{F}_{k-1}] = V_k^2\mathbf{E}[(X_k - X_{k-1})^2|\mathcal{F}_{k-1}] \\ &= \frac{\langle X\rangle_k - \langle X\rangle_{k-1}}{(1 + \langle X\rangle_k)^2} \leq \frac{1}{1 + \langle X\rangle_{k-1}} - \frac{1}{1 + \langle X\rangle_k}\end{aligned}$$

(as $(x - y)/(1 + x)^2 \leq (1 + y)^{-1} - (1 + x)^{-1}$ for all $x \geq y \geq 0$ and $\langle X\rangle_k \geq 0$ is non-decreasing in k). With $\langle Y\rangle_0 = \langle X\rangle_0 = 0$, adding the preceding inequalities over $k = 1, \ldots, n$, we deduce that $\langle Y\rangle_n \leq 1 - 1/(1 + \langle X\rangle_n) \leq 1$ for all n. Thus, by part (a) of the theorem, for almost every ω, $Y_n(\omega)$ has a finite limit. That is, for a.e. ω the series $\sum_n x_n/b_n$ converges, where $b_n = 1 + \langle X\rangle_n(\omega)$ is a positive, non-decreasing sequence and $X_n(\omega) = \sum_{k=1}^n x_k$ for all n. If in addition to the convergence of this series also $\langle X\rangle_\infty(\omega) = \infty$ then $b_n \uparrow \infty$ and by Kronecker's lemma $X_n(\omega)/b_n \to 0$. In this case $b_n/(b_n-1) \to 1$ so we conclude that then also $X_n/(b_n-1) \to 0$, which is exactly the thesis of part (b) of the theorem.

(c). Suppose that $\mathbf{P}\big(\langle X\rangle_\infty = \infty,\ \sup_n |X_n| < \infty\big) > 0$. Then, there exists some r such that $\mathbf{P}\big(\langle X\rangle_\infty = \infty,\ \tau_r = \infty\big) > 0$ for the $\mathcal{F}_n$-stopping time $\tau_r = \inf\{m \ge 0 : |X_m| > r\}$. Since $\sup_m |X_m - X_{m-1}| \le c$ for some non-random finite constant c, we have that $|X_{n\wedge\tau_r}| \le r+c$, from which we deduce that $\mathbf{E}\langle X\rangle_{n\wedge\tau_r} = \mathbf{E}X^2_{n\wedge\tau_r} \le (r+c)^2$ for all n. With $0 \le \langle X\rangle_{n\wedge\tau_r} \uparrow \langle X\rangle_{\tau_r}$, by monotone convergence also

$$\mathbf{E}\big[\langle X\rangle_\infty I_{\tau_r=\infty}\big] \le \mathbf{E}\big[\langle X\rangle_{\tau_r}\big] \le (r+c)^2 .$$

This contradicts our assumption that $\mathbf{P}\big(\langle X\rangle_\infty = \infty,\ \tau_r = \infty\big) > 0$. In conclusion, necessarily, $\mathbf{P}\big(\langle X\rangle_\infty = \infty,\ \sup_n |X_n| < \infty\big) = 0$. Consequently, with $\sup_n |X_n|$ finite on the set of ω values for which $X_n(\omega)$ converges to a finite limit, it follows that $\langle X\rangle_\infty(\omega)$ is finite for a.e. such ω. $\square$

We next prove Lévy's extension of both Borel-Cantelli lemmas (which is a neat application of the preceding theorem).

Proposition 5.3.33 (Borel-Cantelli III). *Consider events $A_n \in \mathcal{F}_n$ for some filtration $\{\mathcal{F}_n\}$. Let $S_n = \sum_{k=1}^n I_{A_k}$ count the number of events occurring among the first n, with $S_\infty = \sum_k I_{A_k}$ the corresponding total number of occurrences. Similarly, let $Z_n = \sum_{k=1}^n \xi_k$ denote the sum of the first n conditional probabilities $\xi_k = \mathbf{P}(A_k|\mathcal{F}_{k-1})$ and $Z_\infty = \sum_k \xi_k$. Then, for almost every ω,*

(a) *If $Z_\infty(\omega)$ is finite, then so is $S_\infty(\omega)$.*

(b) *If $Z_\infty(\omega)$ is infinite, then $S_n(\omega)/Z_n(\omega) \to 1$.*

Remark. Given any sequence of events, by the tower property $\mathbf{E}\xi_k = \mathbf{P}(A_k)$ for all k and setting $\mathcal{F}_n = \sigma(A_k, k \le n)$ guarantees that $A_k \in \mathcal{F}_k$ for all k. Hence,
(a) If $\mathbf{E}Z_\infty = \sum_k \mathbf{P}(A_k)$ is finite, then from part (a) of Proposition 5.3.33 we deduce that $\sum_k I_{A_k}$ is finite a.s., thus recovering the first Borel-Cantelli lemma.
(b) For $\mathcal{F}_n = \sigma(A_k, k \le n)$ and mutually independent events $\{A_k\}$ we have that $\xi_k = \mathbf{P}(A_k)$ and $Z_n = \mathbf{E}S_n$ for all n. Thus, in this case, part (b) of Proposition 5.3.33 is merely the statement that $S_n/\mathbf{E}S_n \stackrel{a.s.}{\to} 1$ when $\sum_k \mathbf{P}(A_k) = \infty$, which is your extension of the second Borel-Cantelli via Exercise 2.2.26.

Proof. Clearly, $M_n = S_n - Z_n$ is square-integrable and $\mathcal{F}_n$-adapted. Further, as $M_n - M_{n-1} = I_{A_n} - \mathbf{E}[I_{A_n}|\mathcal{F}_{n-1}]$ and $\mathsf{Var}(I_{A_n}|\mathcal{F}_{n-1}) = \xi_n(1-\xi_n)$, it follows that the predictable compensator of the L^2 martingale $(M_n, \mathcal{F}_n)$ is $\langle M\rangle_n = \sum_{k=1}^n \xi_k(1-\xi_k)$. Hence, $\langle M\rangle_n \le Z_n$ for all n, and if $Z_\infty(\omega)$ is finite, then so is $\langle M\rangle_\infty(\omega)$. By part (a) of Theorem 5.3.32, for a.e. such ω the finite limit $M_\infty(\omega)$ of $M_n(\omega)$ exists, implying that $S_\infty = M_\infty + Z_\infty$ is finite as well.

With $S_n = M_n + Z_n$, it suffices for part (b) of the proposition to show that $M_n/Z_n \to 0$ for a.e. ω for which $Z_\infty(\omega) = \infty$. To this end, note first that by the preceding argument, the finite limit $M_\infty(\omega)$ exists also for a.e. ω for which $Z_\infty(\omega) = \infty$ while $\langle M\rangle_\infty(\omega)$ is finite. For such ω we have that $M_n/Z_n \to 0$ (since $M_n(\omega)$ is a bounded sequence while $Z_n(\omega)$ is unbounded). Finally, from part (b) of Theorem 5.3.32 we know that $M_n/\langle M\rangle_n \ge M_n/Z_n$ converges to zero for a.e. ω for which $\langle M\rangle_\infty(\omega)$ is infinite. $\square$

Here is a direct appliction of Theorem 5.3.32.

Exercise 5.3.34. *Given a martingale $(M_n, \mathcal{F}_n)$ and positive, non-random $b_n \uparrow \infty$, show that $b_n^{-1} M_n \to 0$ for a.e. ω such that $\sum_{k \geq 1} b_k^{-2} \mathbf{E}[(M_k - M_{k-1})^2 | \mathcal{F}_{k-1}]$ is finite.*
Hint: *Consider $X_n = \sum_{k=1}^n b_k^{-1}(M_k - M_{k-1})$ and recall Kronecker's lemma.*

The following extension of Kolmogorov's three series theorem uses both Theorem 5.3.32 and Lévy's extension of the Borel-Cantelli lemmas.

Exercise 5.3.35. *Suppose $\{X_n\}$ is adapted to filtration $\{\mathcal{F}_n\}$ and for any n, the R.C.P.D. of X_n given $\mathcal{F}_{n-1}$ equals the R.C.P.D. of $-X_n$ given $\mathcal{F}_{n-1}$. For non-random $c > 0$ let $X_n^{(c)} = X_n I_{|X_n| \leq c}$ be the corresponding truncated variables.*

(a) *Verify that $(Z_n, \mathcal{F}_n)$ is a MG, where $Z_n = \sum_{k=1}^n X_k^{(c)}$.*
(b) *Considering the series*

$$\sum_n \mathbf{P}(|X_n| > c \,|\mathcal{F}_{n-1}), \qquad \text{and} \qquad \sum_n \mathsf{Var}(X_n^{(c)} \,|\mathcal{F}_{n-1}), \tag{5.3.3}$$

show that for a.e. ω the series $\sum_n X_n(\omega)$ has a finite limit if and only if both series in (5.3.3) converge.
(c) *Provide an example where the convergence in part (b) occurs with probability $0 < p < 1$.*

We now consider sufficient conditions for convergence almost surely of the martingale transform.

Exercise 5.3.36. *Suppose $Y_n = \sum_{k=1}^n V_k(Z_k - Z_{k-1})$ is the martingale transform of the $\mathcal{F}_n$-predictable $\{V_n\}$ with respect to the martingale $(Z_n, \mathcal{F}_n)$, per Definition 5.1.27.*

(a) *Show that if $\{Z_n\}$ is L^2-bounded and $\{V_n\}$ is uniformly bounded then $Y_n \stackrel{a.s.}{\to} Y_\infty$ finite.*
(b) *Deduce that for L^2-bounded MG $\{Z_n\}$ the sequence $Y_n(\omega)$ converges to a finite limit for a.e. ω for which $\sup_{k \geq 1} |V_k(\omega)|$ is finite.*
(c) *Suppose now that $\{V_k\}$ is predictable for the canonical filtration $\{\mathcal{F}_n\}$ of the i.i.d. $\{\xi_k\}$. Show that if $\xi_k \stackrel{\mathcal{D}}{=} -\xi_k$ and $u \mapsto u\mathbf{P}(|\xi_1| \geq u)$ is bounded above, then the series $\sum_n V_n \xi_n$ has a finite limit for a.e. ω for which $\sum_{k \geq 1} |V_k(\omega)|$ is finite.*
Hint: *Consider Exercise 5.3.35 for the adapted sequence $X_k = V_k \xi_k$.*

Here is another application of Lévy's extension of the Borel-Cantelli lemmas.

Exercise 5.3.37. *Suppose $X_n = 1 + \sum_{k=1}^n D_k$, $n \geq 0$, where the $\{-1, 1\}$-valued D_k is $\mathcal{F}_k$-adapted and such that $\mathbf{E}[D_k | \mathcal{F}_{k-1}] \geq \epsilon$ for some non-random $1 > \epsilon > 0$ and all $k \geq 1$.*

(a) *Show that $(X_n, \mathcal{F}_n)$ is a sub-martingale and provide its Doob decomposition.*
(b) *Using this decomposition and Lévy's extension of the Borel-Cantelli lemmas, show that $X_n \to \infty$ almost surely.*
(c) *Let $Z_n = \phi^{X_n}$ for $\phi = (1 - \epsilon)/(1 + \epsilon)$. Show that $(Z_n, \mathcal{F}_n)$ is a super-martingale and deduce that $\mathbf{P}(\inf_n X_n \leq 0) \leq \phi$.*

As we show next, the predictable compensator controls the exponential tails for martingales of bounded differences.

Exercise 5.3.38. *Fix $\lambda > 0$ non-random and an L^2 martingale $(M_n, \mathcal{F}_n)$ with $M_0 = 0$ and bounded differences $\sup_k |M_k - M_{k-1}| \le 1$.*

(a) *Show that $N_n = \exp(\lambda M_n - (e^\lambda - \lambda - 1)\langle M\rangle_n)$ is a sup-MG for $\{\mathcal{F}_n\}$.*
Hint: *Recall part (a) of Exercise 1.4.40.*

(b) *Show that for any a.s. finite $\mathcal{F}_n$-stopping time τ and constants $u, r > 0$,*

$$\mathbf{P}(M_\tau \ge u,\ \langle M\rangle_\tau \le r) \le \exp(-\lambda u + r(e^\lambda - \lambda - 1))\,.$$

(c) *Show that if the martingale $\{S_n\}$ of Example 5.3.22 has uniformly bounded differences $|\xi_k| \le 1$, then $\mathbf{E}\exp(\lambda S_\infty)$ is finite for $S_\infty = \sum_k \xi_k$ and any $\lambda \in \mathbb{R}$.*

Applying part (c) of the preceding exercise, you are next to derive the following tail estimate, due to Dvoretsky, in the context of Lévy's extension of the Borel-Cantelli lemmas.

Exercise 5.3.39. *Suppose $A_k \in \mathcal{F}_k$ for some filtration $\{\mathcal{F}_k\}$. Let $S_n = \sum_{k=1}^n I_{A_k}$ and $Z_n = \sum_{k=1}^n \mathbf{P}(A_k|\mathcal{F}_{k-1})$. Show that $\mathbf{P}(S_n \ge r+u, Z_n \le r) \le e^u (r/(r+u))^{r+u}$ for all n and $u, r > 0$, then deduce that for any $0 < r < 1$,*

$$\mathbf{P}\Big(\bigcup_{k=1}^n A_k\Big) \le er + \mathbf{P}\Big(\sum_{k=1}^n \mathbf{P}(A_k|\mathcal{F}_{k-1}) > r\Big)\,.$$

Hint: *Recall the proof of Borel-Cantelli III that the L^2-martingale $M_n = S_n - Z_n$ has differences bounded by one and $\langle M\rangle_n \le Z_n$.*

We conclude this section with a refinement of the well known *Azuma-Hoeffding* concentration inequality for martingales of bounded differences, from which we deduce the *strong law of large numbers* for martingales of bounded differences.

Exercise 5.3.40. *Suppose $(M_n, \mathcal{F}_n)$ is a martingale with $M_0 = 0$ and differences $D_k = M_k - M_{k-1}$, $k \ge 1$ such that for some finite γ_k,*

$$\mathbf{E}[D_k^2 e^{D_k}|\mathcal{F}_{k-1}] \le \gamma_k^2\, \mathbf{E}[e^{D_k}|\mathcal{F}_{k-1}] < \infty\,.$$

(a) *Show that $N_n = \exp(\lambda M_n - \lambda^2 r_n/2)$ is a sup-MG for $\mathcal{F}_n$ provided $\lambda \in [0,1]$ and $r_n = \sum_{k=1}^n \gamma_k^2$.*
Hint: *Recall part (b) of Exercise 1.4.40.*

(b) *Deduce that for $I(x) = (x \wedge 1)(2x - x \wedge 1)$ and any $u \ge 0$,*

$$\mathbf{P}(M_n \ge u) \le \exp(-r_n I(u/r_n)/2)\,.$$

(c) *Conclude that $b_n^{-1} M_n \overset{a.s.}{\to} 0$ for any martingale $\{M_n\}$ of uniformly bounded differences and non-random $\{b_n\}$ such that $b_n/\sqrt{n \log n} \to \infty$.*

5.4. The optional stopping theorem

This section is about the use of martingales in computations involving stopping times. The key tool for doing so is the following theorem.

Theorem 5.4.1 (Doob's optional stopping). *Suppose $\theta \le \tau$ are $\mathcal{F}_n$-stopping times and $X_n = Y_n + V_n$ for sub-MGs $(V_n, \mathcal{F}_n)$, $(Y_n, \mathcal{F}_n)$ such that V_n is non-positive and $\{Y_{n\wedge\tau}\}$ is uniformly integrable. Then, the R.V. X_θ and X_τ are integrable and $\mathbf{E}X_\tau \ge \mathbf{E}X_\theta \ge \mathbf{E}X_0$ (where $X_\tau(\omega)$ and $X_\theta(\omega)$ are set as $\limsup_n X_n(\omega)$ in case the corresponding stopping time is infinite).*

Remark 5.4.2. Doob's optional stopping theorem holds for any sub-MG $(X_n, \mathcal{F}_n)$ such that $\{X_{n\wedge\tau}\}$ is uniformly integrable (just set $V_n = 0$). Alternatively, it holds also whenever $\mathbf{E}[X_\infty|\mathcal{F}_n] \geq X_n$ for some integrable X_∞ and all n (for then the martingale $Y_n = \mathbf{E}[X_\infty|\mathcal{F}_n]$ is U.I. by Corollary 5.3.14, hence $\{Y_{n\wedge\tau}\}$ also U.I. by Proposition 5.4.4, and the sub-MG $V_n = X_n - Y_n$ is by assumption non-positive).
By far the most common application has $(X_n, \mathcal{F}_n)$ a martingale, in which case it yields that $\mathbf{E}X_0 = \mathbf{E}X_\tau$ for any $\mathcal{F}_n$-stopping time τ such that $\{X_{n\wedge\tau}\}$ is U.I. (for example, whenever τ is bounded, or under the more general conditions of Proposition 5.4.4).

Proof. By linearity of the expectation, it suffices to prove the claim separately for $Y_n = 0$ and for V_n. Dealing first with $Y_n = 0$, i.e. with a non-positive sub-MG $(V_n, \mathcal{F}_n)$, note that $(-V_n, \mathcal{F}_n)$ is then a non-negative sup-MG. Thus, the inequality $\mathbf{E}[V_\tau] \geq \mathbf{E}[V_\theta] \geq \mathbf{E}[V_0]$ and the integrability of V_θ and V_τ are immediate consequences of Proposition 5.3.8.
Considering hereafter the sub-MG $(Y_n, \mathcal{F}_n)$ such that $\{Y_{n\wedge\tau}\}$ is U.I., since $\theta \leq \tau$ are $\mathcal{F}_n$-stopping times it follows by Theorem 5.1.32 that $U_n = Y_{n\wedge\tau}$, $Z_n = Y_{n\wedge\theta}$ and $U_n - Z_n$ are all sub-MGs with respect to $\mathcal{F}_n$. In particular, $\mathbf{E}U_n \geq \mathbf{E}Z_n \geq \mathbf{E}Z_0$ for all n. Our assumption that the sub-MG $(U_n, \mathcal{F}_n)$ is U.I. results with $U_n \to U_\infty$ a.s. and in L^1 (see Theorem 5.3.12). Further, as we show in part (c) of Proposition 5.4.4, in this case $U_{n\wedge\theta} = Z_n$ is U.I. so by the same reasoning, $Z_n \to Z_\infty$ a.s. and in L^1. We thus deduce that $\mathbf{E}U_\infty \geq \mathbf{E}Z_\infty \geq \mathbf{E}Z_0$. By definition, $U_\infty = \lim_n Y_{n\wedge\tau} = Y_\tau$ and $Z_\infty = \lim_n Y_{n\wedge\theta} = Y_\theta$. Consequently, $\mathbf{E}Y_\tau \geq \mathbf{E}Y_\theta \geq \mathbf{E}Y_0$, as claimed. ☐

We complement Theorem 5.4.1 by first strengthening its conclusion and then providing explicit sufficient conditions for the uniform integrability of $\{Y_{n\wedge\tau}\}$.

Lemma 5.4.3. *Suppose $\{X_n\}$ is adapted to filtration $\{\mathcal{F}_n\}$ and the $\mathcal{F}_n$-stopping time τ is such that for any $\mathcal{F}_n$-stopping time $\tau \geq \theta$ the R.V. X_θ is integrable and $\mathbf{E}[X_\tau] \geq \mathbf{E}[X_\theta]$. Then, also $\mathbf{E}[X_\tau|\mathcal{F}_\theta] \geq X_\theta$ a.s.*

Proof. Fixing $A \in \mathcal{F}_\theta$ set $\eta = \theta I_A + \tau I_{A^c}$. Note that $\eta \leq \tau$ is also an $\mathcal{F}_n$-stopping time since for any n,

$$\begin{aligned}\{\eta \leq n\} &= (A \cap \{\theta \leq n\}) \bigcup (A^c \cap \{\tau \leq n\}) \\ &= (A \cap \{\theta \leq n\}) \bigcup ((A^c \cap \{\theta \leq n\}) \cap \{\tau \leq n\}) \in \mathcal{F}_n\end{aligned}$$

because both A and A^c are in $\mathcal{F}_\theta$ and $\{\tau \leq n\} \in \mathcal{F}_n$ (c.f. Definition 5.1.34 of the σ-algebra $\mathcal{F}_\theta$). By assumption, X_η, X_θ, X_τ are integrable and $\mathbf{E}X_\tau \geq \mathbf{E}X_\eta$. Since $X_\eta = X_\theta I_A + X_\tau I_{A^c}$ subtracting the finite $\mathbf{E}[X_\tau I_{A^c}]$ from both sides of this inequality results with $\mathbf{E}[X_\tau I_A] \geq \mathbf{E}[X_\theta I_A]$. This holds for all $A \in \mathcal{F}_\theta$ and with $\mathbf{E}[X_\tau I_A] = \mathbf{E}[Z I_A]$ for $Z = \mathbf{E}[X_\tau|\mathcal{F}_\theta]$ (by definition of the conditional expectation), we see that $\mathbf{E}[(Z - X_\theta)I_A] \geq 0$ for all $A \in \mathcal{F}_\theta$. Since both Z and X_θ are measurable on $\mathcal{F}_\theta$ (see part (b) of Exercise 5.1.35), it thus follows that a.s. $Z \geq X_\theta$, as claimed. ☐

Proposition 5.4.4. *Suppose $\{Y_n\}$ is integrable and τ is a stopping time for a filtration $\{\mathcal{F}_n\}$. Then, $\{Y_{n\wedge\tau}\}$ is uniformly integrable if any one of the following conditions hold.*

(a) $\mathbf{E}\tau < \infty$ *and a.s.* $\mathbf{E}[|Y_n - Y_{n-1}||\mathcal{F}_{n-1}] \leq c$ *for some finite, non-random* c.

(b) $\{Y_n I_{\tau>n}\}$ *is uniformly integrable and* $Y_\tau I_{\tau<\infty}$ *is integrable.*
(c) $(Y_n, \mathcal{F}_n)$ *is a uniformly integrable sub-MG (or sup-MG).*

PROOF. (a) Clearly, $|Y_{n\wedge\tau}| \le Z_n$, where

$$Z_n = |Y_0| + \sum_{k=1}^{n\wedge\tau} |Y_k - Y_{k-1}| = |Y_0| + \sum_{k=1}^{n} |Y_k - Y_{k-1}| I_{\tau\ge k}\,,$$

is non-decreasing in n. Hence, $\sup_n |Y_{n\wedge\tau}| \le Z_\infty$, implying that $\{Y_{n\wedge\tau}\}$ is U.I. whenever $\mathbf{E}Z_\infty$ is finite (c.f. Lemma 1.3.48). Proceeding to show that this is the case under condition (a), recall that $I_{\tau\ge k} \in m\mathcal{F}_{k-1}$ for all k (since τ is an $\mathcal{F}_n$-stopping time). Thus, taking out what is known, by the tower property we find that under condition (a),

$$\mathbf{E}[|Y_k - Y_{k-1}| I_{\tau\ge k}] = \mathbf{E}[\mathbf{E}(|Y_k - Y_{k-1}|\,|\mathcal{F}_{k-1}) I_{\tau\ge k}] \le c\,\mathbf{P}(\tau \ge k)$$

for all $k \ge 1$. Summing this bound over $k = 1, 2, \ldots$ results with

$$\mathbf{E}Z_\infty \le \mathbf{E}|Y_0| + c\sum_{k=1}^{\infty} \mathbf{P}(\tau \ge k) = \mathbf{E}|Y_0| + c\,\mathbf{E}\tau\,,$$

with the integrability of Z_∞ being a consequence of the hypothesis in condition (a) that τ is integrable.

(b) Next note that $|X_{n\wedge\tau}| \le |X_\tau| I_{\tau<\infty} + |X_n| I_{\tau>n}$ for every n, any sequence of random variables $\{X_n\}$ and any $\tau \in \{0, 1, 2, \ldots, \infty\}$. Condition (b) states that the sequence $\{|Y_n| I_{\tau>n}\}$ is U.I. and that the variable $|Y_\tau| I_{\tau<\infty}$ is integrable. Thus, taking the expectation of the preceding inequality in case $X_n = Y_n I_{|Y_n|>M}$, we find that when condition (b) holds,

$$\sup_n \mathbf{E}[|Y_{n\wedge\tau}| I_{|Y_{n\wedge\tau}|>M}] \le \mathbf{E}[|Y_\tau| I_{|Y_\tau|>M} I_{\tau<\infty}] + \sup_n \mathbf{E}[|Y_n| I_{|Y_n|>M} I_{\tau>n}]\,,$$

converges to zero as $M \uparrow \infty$. That is, $\{|Y_{n\wedge\tau}|\}$ is then a U.I. sequence.

(c) The hypothesis of (c) that $\{Y_n\}$ is U.I. implies that $\{Y_n I_{\tau>n}\}$ is also U.I. and that $\sup_n \mathbf{E}[(Y_n)_+]$ is finite. With τ an $\mathcal{F}_n$-stopping time and $(Y_n, \mathcal{F}_n)$ a sub-MG, it further follows by Lemma 5.3.7 that $Y_\tau I_{\tau<\infty}$ is integrable. Having arrived at the hypothesis of part (b), we are done. $\square$

Since $\{Y_{n\wedge\tau}\}$ is U.I. whenever τ is bounded, we have the following immediate consequences of Doob's optional stopping theorem, Remark 5.4.2 and Lemma 5.4.3.

Corollary 5.4.5. *For any sub-MG* $(X_n, \mathcal{F}_n)$ *and any non-decreasing sequence* $\{\tau_k\}$ *of* $\mathcal{F}_n$*-stopping times,* $(X_{\tau_k}, \mathcal{F}_{\tau_k}, k \ge 0)$ *is a sub-MG when either* $\sup_k \tau_k \le \ell$ *a non-random finite integer, or a.s.* $X_n \le \mathbf{E}[X_\infty|\mathcal{F}_n]$ *for an integrable* X_∞ *and all* $n \ge 0$.

Check that by part (b) of Exercise 5.2.16 and part (c) of Proposition 5.4.4 it follows from Doob's optional stopping theorem that $\mathbf{E}S_\tau = 0$ for any stopping time τ with respect to the canonical filtration of $S_n = \sum_{k=1}^n \xi_k$ provided the independent ξ_k are integrable with $\mathbf{E}\xi_k = 0$ and $\sup_n \mathbf{E}|S_n| < \infty$.

Sometimes Doob's optional stopping theorem is applied en-route to a useful contradiction. For example,

Exercise 5.4.6. *Show that if* $\{X_n\}$ *is a sub-martingale such that* $\mathbf{E}X_0 \ge 0$ *and* $\inf_n X_n < 0$ *a.s. then necessarily* $\mathbf{E}[\sup_n X_n] = \infty$.

Hint: *Assuming first that* $\sup_n |X_n|$ *is integrable, apply Doob's optional stopping theorem to arrive at a contradiction. Then consider the same argument for the sub-MG* $Z_n = \max\{X_n, -1\}$.

Exercise 5.4.7. *Fixing* $b > 0$, *let* $\tau_b = \min\{n \geq 0 : S_n \geq b\}$ *for the random walk* $\{S_n\}$ *of Definition 5.1.6 and suppose* $\xi_n = S_n - S_{n-1}$ *are uniformly bounded, of zero mean and positive variance.*

(a) *Show that* τ_b *is almost surely finite.*
Hint: *See Proposition 5.3.5.*

(b) *Show that* $\mathbf{E}[\min\{S_n : n \leq \tau_b\}] = -\infty$.

Martingales often provide much information about specific stopping times. We detail below one such example, pertaining to the SRW of Definition 5.1.6.

Corollary 5.4.8 (GAMBLER'S RUIN). *Fixing positive integers* a *and* b *the probability that a* SRW $\{S_n\}$, *starting at* $S_0 = 0$, *hits* $-a$ *before first hitting* $+b$ *is* $r = (e^{\lambda b} - 1)/(e^{\lambda b} - e^{-\lambda a})$ *for* $\lambda = \log[(1-p)/p] \neq 0$. *For the symmetric* SRW, *i.e. when* $p = 1/2$, *this probability is* $r = b/(a+b)$.

Remark. The probability r is often called the gambler's ruin, or *ruin probability* for a gambler with initial capital of $+a$, betting on the outcome of independent rounds of the same game, a unit amount per round, gaining or losing an amount equal to his bet in each round and stopping when either all his capital is lost (the ruin event), or his accumulated gains reach the amount $+b$.

PROOF. Consider the stopping time $\tau_{a,b} = \inf\{n \geq 0 : S_n \geq b, \text{ or } S_n \leq -a\}$ for the canonical filtration of the SRW. That is, $\tau_{a,b}$ is the first time that the SRW exits the interval $(-a, b)$. Since $(S_k + k)/2$ has the Binomial(k, p) distribution it is not hard to check that $\sup_\ell \mathbf{P}(S_k = \ell) \to 0$ hence $\mathbf{P}(\tau_{a,b} > k) \leq \mathbf{P}(-a < S_k < b) \to 0$ as $k \to \infty$. Consequently, $\tau_{a,b}$ is finite a.s. Further, starting at $S_0 \in (-a, b)$ and using only increments $\xi_k \in \{-1, 1\}$, necessarily $S_{\tau_{a,b}} \in \{-a, b\}$ with probability one. Our goal is thus to compute the ruin probability $r = \mathbf{P}(S_{\tau_{a,b}} = -a)$. To this end, note that $\mathbf{E}e^{\lambda \xi_k} = pe^\lambda + (1-p)e^{-\lambda} = 1$ for $\lambda = \log[(1-p)/p]$. Thus, $M_n = \exp(\lambda S_n) = \prod_{k=1}^n e^{\lambda \xi_k}$ is, for such λ, a non-negative MG with $M_0 = 1$ (c.f. Example 5.1.10). Clearly, $M_{n \wedge \tau_{a,b}} = \exp(\lambda S_{n \wedge \tau_{a,b}}) \leq \exp(|\lambda| \max(a, b))$ is uniformly bounded (in n), hence uniformly integrable. So, applying Doob's optional stopping theorem for this MG and stopping time, we have that

$$1 = \mathbf{E}M_0 = \mathbf{E}[M_{\tau_{a,b}}] = \mathbf{E}[e^{\lambda S_{\tau_{a,b}}}] = re^{-\lambda a} + (1-r)e^{\lambda b},$$

which easily yields the stated explicit formula for r in case $\lambda \neq 0$ (i.e. $p \neq 1/2$). Finally, recall that $\{S_n\}$ is a martingale for the symmetric SRW, with $S_{n \wedge \tau_{a,b}}$ uniformly bounded, hence uniformly integrable. So, applying Doob's optional stopping theorem for this MG, we find that in the symmetric case

$$0 = \mathbf{E}S_0 = \mathbf{E}[S_{\tau_{a,b}}] = -ar + b(1-r),$$

that is, $r = b/(a+b)$ when $p = 1/2$. □

Here is an interesting consequence of the Gambler's ruin formula.

Example 5.4.9. *Initially, at step* $k = 0$ *zero is the only occupied site in* $\mathbb{Z}$. *Then, at each step a new particle starts at zero and follows a symmetric* SRW, *independently of the previous particles, till it lands on an unoccupied site, whereby it stops and* *thereafter occupies this site. The set of occupied sites after k steps is thus an interval of length $k+1$ and we let $R_k \in \{1, \ldots, k+1\}$ count the number of non-negative integers occupied after k steps (starting at $R_0 = 1$).*

Clearly, $R_{k+1} \in \{R_k, R_k + 1\}$ and $\mathbf{P}(R_{k+1} = R_k | \mathcal{F}_k^{\mathbf{M}}) = R_k/(k+2)$ by the preceding Gambler's ruin formula. Thus, $\{R_k\}$ follows the evolution of Bernard Friedman's urn *with parameters $d_k = r = b = 1$ and $c_k = 0$. Consequently, by Exercise 5.3.27 we have that $(n+1)^{-1}R_n \stackrel{a.s.}{\to} 1/2$.*

You are now to derive Wald's identities about stopping times for the random walk, and use them to gain further information about the stopping times $\tau_{a,b}$ of the preceding corollary.

Exercise 5.4.10. *Let τ be an integrable stopping time for the canonical filtration of the random walk $\{S_n\}$.*

(a) *Show that if ξ_1 is integrable, then* Wald's identity $\mathbf{E}S_\tau = \mathbf{E}\xi_1 \mathbf{E}\tau$ *holds.*
Hint: *Use the representation $S_\tau = \sum_{k=1}^{\infty} \xi_k I_{k \le \tau}$ and independence.*
(b) *Show that if in addition ξ_1 is square-integrable, then* Wald's second identity $\mathbf{E}[(S_\tau - \tau\mathbf{E}\xi_1)^2] = \mathsf{Var}(\xi_1)\mathbf{E}\tau$ *holds as well.*
Hint: *Explain why you may assume that $\mathbf{E}\xi_1 = 0$, prove the identity with $n \wedge \tau$ instead of τ and use Doob's L^2 convergence theorem.*
(c) *Show that if $\xi_1 \ge 0$ then Wald's identity applies also when $\mathbf{E}\tau = \infty$ (under the convention that $0 \times \infty = 0$).*

Exercise 5.4.11. *For the* SRW *S_n and positive integers a, b consider the stopping time $\tau_{a,b} = \min\{n \ge 0 : S_n \notin (-a, b)\}$ as in proof of Corollary 5.4.8.*

(a) *Check that $\mathbf{E}[\tau_{a,b}] < \infty$.*
Hint: *See Exercise 5.1.15.*
(b) *Combining Corollary 5.4.8 with Wald's identities, compute the value of $\mathbf{E}[\tau_{a,b}]$.*
(c) *Show that $\tau_{a,b} \uparrow \tau_b = \min\{n \ge 0 : S_n = b\}$ for $a \uparrow \infty$ (where the minimum over the empty set is ∞), and deduce that $\mathbf{E}\tau_b = b/(2p-1)$ when $p \ge 1/2$.*
(d) *Show that τ_b is almost surely finite when $p \ge 1/2$.*
(e) *Find constants c_1 and c_2 such that $Y_n = S_n^4 - 6nS_n^2 + c_1 n^2 + c_2 n$ is a martingale for the symmetric* SRW, *and use it to evaluate $\mathbf{E}[(\tau_{b,b})^2]$ in this case.*

We next provide a few applications of Doob's optional stopping theorem, starting with information on the law of τ_b for SRW (and certain other random walks).

Exercise 5.4.12. *Consider the stopping time $\tau_b = \inf\{n \ge 0 : S_n = b\}$ and the martingale $M_n = \exp(\lambda S_n) M(\lambda)^{-n}$ for a* SRW *$\{S_n\}$, with b a positive integer and $M(\lambda) = \mathbf{E}[e^{\lambda \xi_1}]$.*

(a) *Show that if $p = 1 - q \in [1/2, 1)$ then $e^{\lambda b}\mathbf{E}[M(\lambda)^{-\tau_b}] = 1$ for every $\lambda > 0$.*
(b) *Deduce that for $p \in [1/2, 1)$ and every $0 < s < 1$,*

$$\mathbf{E}[s^{\tau_1}] = \frac{1}{2qs}\left[1 - \sqrt{1 - 4pqs^2}\right],$$

and $\mathbf{E}[s^{\tau_b}] = (\mathbf{E}[s^{\tau_1}])^b$.
(c) *Show that if $0 < p < 1/2$ then $\mathbf{P}(\tau_b < \infty) = \exp(-\lambda_\star b)$ for $\lambda_\star = \log[(1-p)/p] > 0$.*

(d) *Deduce that for* $p \in (0, 1/2)$ *the variable* $Z = 1 + \max_{k \geq 0} S_k$ *has a Geometric distribution of success probability* $1 - e^{-\lambda_\star}$.

Exercise 5.4.13. *Consider* $\tau_b = \min\{n \geq 0 : S_n \geq b\}$ *for* $b > 0$, *in case the i.i.d. increments* $\xi_n = S_n - S_{n-1}$ *of the random walk* $\{S_n\}$ *are such that* $\mathbf{P}(\xi_1 > 0) > 0$ *and* $\{\xi_1 | \xi_1 > 0\}$ *has the Exponential law of parameter* α.

(a) *Show that for any* n *finite, conditional on* $\{\tau_b = n\}$ *the law of* $S_{\tau_b} - b$ *is also Exponential of parameter* α.
Hint: *Recall the memory-less property of the exponential distribution.*

(b) *With* $M(\lambda) = \mathbf{E}[e^{\lambda \xi_1}]$ *and* $\lambda_\star \geq 0$ *denoting the maximal solution of* $M(\lambda) = 1$, *verify the existence of a monotone decreasing, continuous function* $u : (0, 1] \mapsto [\lambda_\star, \alpha)$ *such that* $M(u(s)) = 1/s$.

(c) *Evaluate* $\mathbf{E}[s^{\tau_b} I_{\tau_b < \infty}]$, $0 < s < 1$, *and* $\mathbf{P}(\tau_b < \infty)$ *in terms of* $u(s)$ *and* $\lambda_\star$.

Exercise 5.4.14. *A monkey types a random sequence of capital letters* $\{\xi_k\}$ *that are chosen independently of each other, with each* ξ_k *chosen uniformly from amongst the 26 possible values* $\{A, B, \ldots, Z\}$.

(a) *Suppose that just before* each *time step* $n = 1, 2, \cdots$, *a new gambler arrives on the scene and bets* \$1 *that* $\xi_n = P$. *If he loses, he leaves, whereas if he wins, he receives* \$26, *all of which he bets on the event* $\xi_{n+1} = R$. *If he now loses, he leaves, whereas if he wins, he bets his current fortune of* $\$26^2$ *on the event that* $\xi_{n+2} = O$, *and so on, through the word* $PROBABILITY$. *Show that the amount of money* M_n *that the gamblers have collectively earned by time* n *is a martingale with respect to* $\{\mathcal{F}_n^{\boldsymbol{\xi}}\}$.

(b) *Let* L_n *denote the number of occurrences of the word* $PROBABILITY$ *in the first* n *letters typed by the monkey and* $\widehat{\tau} = \inf\{n \geq 11 : L_n = 1\}$ *the first time by which it produced this word. Using Doob's optional stopping theorem show that* $\mathbf{E}\widehat{\tau} = a$ *for* $a = 26^{11}$. *Does the same apply for the first time* τ *by which the monkey produces the word* $ABRACADABRA$ *and if not, what is* $\mathbf{E}\tau$*?*

(c) *Show that* $n^{-1} L_n \stackrel{a.s.}{\to} \frac{1}{a}$ *and further that* $(L_n - n/a)/\sqrt{vn} \xrightarrow{\mathcal{D}} G$ *for some finite, positive constant* v.
Hint: *Fixing* $\eta < 1/2$, *partition* $\{11, \ldots, n\}$ *into* $m = \lfloor n^\eta \rfloor$ *consecutive blocks* K_i, *each of length either* $\ell = \lfloor n/m \rfloor - 10$ *or* $\ell + 1$, *with gaps of length* 10 *between them. With* W_i *denoting the number of occurrences of* $PROBABILITY$ *within* K_i, *apply Lindeberg's* CLT *to* $(L_n^* - n'/a)/\sqrt{v_n}$, *where* $L_n^* = \sum_{i=1}^m W_i$, $n' = n - 10m$ *and* $v_n = \sum_{i=1}^m \mathsf{Var}(W_i)$, *then show that* $n^{-1} v_n$ *converges.*

Exercise 5.4.15. *Consider a fair game consisting of successive turns whose outcome are the i.i.d. signs* $\xi_k \in \{-1, 1\}$ *such that* $\mathbf{P}(\xi_1 = 1) = \frac{1}{2}$, *and where upon betting the wagers* $\{V_k\}$ *in each turn, your gain (or loss) after* n *turns is* $Y_n = \sum_{k=1}^n \xi_k V_k$. *Here is a betting system* $\{V_k\}$, *predictable with respect to the canonical filtration* $\{\mathcal{F}_n^{\boldsymbol{\xi}}\}$, *as in Example 5.1.30, that* surely makes a profit *in this fair game!*

Choose a finite sequence $x_1, x_2, \ldots, x_\ell$ *of non-random positive numbers. For each* $k \geq 1$, *wager an amount* V_k *that equals the sum of the first and last terms in your sequence prior to your* k*-s turn. Then, to update your sequence, if you just won* *your bet delete those two numbers while if you lost it, append their sum as an extra term $x_{\ell+1} = x_1 + x_\ell$ at the right-hand end of the sequence. You play iteratively according to this rule till your sequence is empty (and if your sequence ever consists of one term only, you wager that amount, so upon wining you delete this term, while upon losing you append it to the sequence to obtain two terms).*

(a) *Let $v = \sum_{i=1}^{\ell} x_i$. Show that the sum of terms in your sequence after n turns is a martingale $S_n = v + Y_n$ with respect to $\{\mathcal{F}_n^{\boldsymbol{\xi}}\}$. Deduce that with probability one you terminate playing with a profit v at the finite $\mathcal{F}_n^{\boldsymbol{\xi}}$-stopping time $\tau = \inf\{n \geq 0 : S_n = 0\}$.*

(b) *Show that $\mathbf{E}\tau$ is finite.*
Hint: *Consider the number of terms N_n in your sequence after n turns.*

(c) *Show that the expected value of your aggregate maximal loss till termination, namely $\mathbf{E}L$ for $L = -\min_{k\leq\tau} Y_k$, is infinite (which is why you are not to attempt this gambling scheme).*

In the next exercise you derive a time-reversed version of the L^2 maximal inequality (5.2.4) by an application of Corollary 5.4.5.

Exercise 5.4.16. *Associate to any given martingale $(Y_n, \mathcal{H}_n)$ the* record times *$\theta_{k+1} = \min\{j \geq 0 : Y_j > Y_{\theta_k}\}$, $k = 0, 1, \ldots$ starting at $\theta_0 = 0$.*

(a) *Fixing m finite, set $\tau_k = \theta_k \wedge m$ and explain why $(Y_{\tau_k}, \mathcal{H}_{\tau_k})$ is a MG.*

(b) *Deduce that if $\mathbf{E}Y_m^2$ is finite then*

$$\sum_{k=1}^{m} \mathbf{E}[(Y_{\tau_k} - Y_{\tau_{k-1}})^2] = \mathbf{E}Y_m^2 - \mathbf{E}Y_0^2 .$$

Hint: *Apply Exercise 5.1.8.*

(c) *Conclude that for any martingale $\{Y_n\}$ and all m*

$$\mathbf{E}[(\max_{\ell\leq m} Y_\ell - Y_m)^2] \leq \mathbf{E}Y_m^2 .$$

5.5. Reversed MGs, likelihood ratios and branching processes

With martingales applied throughout probability theory, we present here just a few selected applications. Our first example, Sub-section 5.5.1, deals with the analysis of extinction probabilities for branching processes. We then study in Sub-section 5.5.2 the likelihood ratios for independent experiments with the help of Kakutani's theorem about product martingales. Finally, in Sub-section 5.5.3 we develop the theory of reversed martingales and applying it, provide zero-one law and representation results for exchangeable processes.

5.5.1. Branching processes: extinction probabilities.

We use martingales to study the extinction probabilities of branching processes, the object we define next.

Definition 5.5.1 (Branching process). *The* branching process *is a discrete time stochastic process $\{Z_n\}$ taking non-negative integer values, such that $Z_0 = 1$ and for any $n \geq 1$,*

$$Z_n = \sum_{j=1}^{Z_{n-1}} N_j^{(n)},$$

where N and $N_j^{(n)}$ for $j = 1, 2, \ldots$ are i.i.d. non-negative integer valued R.V.s with finite mean $m_N = \mathbf{E}N < \infty$, and where we use the convention that if $Z_{n-1} = 0$ then also $Z_n = 0$. We call a branching process sub-critical *when $m_N < 1$,* critical *when $m_N = 1$ and* super-critical *when $m_N > 1$.*

Remark. The S.P. $\{Z_n\}$ is interpreted as counting the size of an evolving population, with $N_j^{(n)}$ being the number of offspring of j^{th} individual of generation $(n-1)$ and Z_n being the size of the n-th generation. Associated with the branching process is the family tree with the root denoting the 0-th generation and having $N_j^{(n)}$ edges from vertex j at distance n from the root to vertices of distance $(n+1)$ from the root. Random trees generated in such a fashion are called *Galton-Watson trees* and are the subject of much research. We focus here on the simpler S.P. $\{Z_n\}$ and shall use throughout the filtration $\mathcal{F}_n = \sigma(\{N_j^{(k)}, k \le n, j = 1, 2, \ldots\})$. We note in passing that in general $\mathcal{F}_n^{\mathbf{Z}}$ is a strict subset of $\mathcal{F}_n$ (since in general one can not recover the number of offspring of each individual knowing only the total population sizes at the different generations). Though not dealt with here, more sophisticated related models have also been successfully studied by probabilists. For example, *branching process with immigration*, where one adds to Z_n an external random variable I_n that count the number of individuals immigrating into the population at the n^{th} generation; *Age-dependent branching process* where individuals have random lifetimes during which they produce offspring according to age-dependent probability generating function; *Multi-type branching process* where each individual is assigned a label (type), possibly depending on the type of its parent and with a different law for the number of offspring in each type, and *branching process in random environment* where the law of the number of offspring per individual is itself a random variable (part of the a-apriori given random environment).

Our goal here is to find the probability p_{ex} of population extinction, formally defined as follows.

Definition 5.5.2. *The* extinction probability *of a branching process is*

$$p_{\text{ex}} := \mathbf{P}(\{\omega : Z_n(\omega) = 0 \text{ for all } n \text{ large enough}\}).$$

Obviously, $p_{\text{ex}} = 0$ whenever $\mathbf{P}(N = 0) = 0$ and $p_{\text{ex}} = 1$ whenever $\mathbf{P}(N = 0) = 1$. Hereafter we exclude these degenerate cases by assuming that $1 > \mathbf{P}(N = 0) > 0$.

To this end, we first deduce that with probability one, conditional upon non-extinction the branching process grows unboundedly.

Lemma 5.5.3. *If $\mathbf{P}(N = 0) > 0$ then with probability one either $Z_n \to \infty$ or $Z_n = 0$ for all n large enough.*

Proof. We start by proving that for any filtration $\mathcal{F}_n \uparrow \mathcal{F}_\infty$ and any S.P. $Z_n \ge 0$ if for $A \in \mathcal{F}_\infty$, some non-random $\eta_k > 0$ and all large positive integers k, n

$$\mathbf{P}(A|\mathcal{F}_n) I_{[0,k]}(Z_n) \ge \eta_k I_{[0,k]}(Z_n), \tag{5.5.1}$$

then $\mathbf{P}(A \cup B) = 1$ for $B = \{\lim_n Z_n = \infty\}$. Indeed, $C_k = \{Z_n \le k, \text{ i.o. in } n\}$ are by (5.5.1) such that $C_k \subseteq \{\mathbf{P}(A|\mathcal{F}_n) \ge \eta_k, \text{ i.o. in } n\}$. By Lévy's 0-1 law $\mathbf{P}(A|\mathcal{F}_n) \to I_A$ except on a set D such that $\mathbf{P}(D) = 0$, hence also $C_k \subseteq D \cup \{I_A \ge \eta_k\} = D \cup A$ for all k. With $C_k \uparrow B^c$ it follows that $B^c \subseteq D \cup A$ yielding our claim that $\mathbf{P}(A \cup B) = 1$.

Turning now to the branching process Z_n, let $A = \{\omega : Z_n(\omega) = 0 \text{ for all } n \text{ large enough}\}$ which is in $\mathcal{F}_\infty$, noting that if $Z_n \le k$ and $N_j^{(n+1)} = 0$, $j = 1, \ldots, k$, then

$Z_{n+1}=0$ hence $\omega \in A$. Consequently, by the independence of $\{N_j^{(n+1)}, j=1,\ldots\}$ and $\mathcal{F}_n$ it follows that

$$\mathbf{E}[I_A|\mathcal{F}_n]I_{\{Z_n\le k\}} \ge \mathbf{E}[I_{\{Z_{n+1}=0\}}|\mathcal{F}_n]I_{\{Z_n\le k\}} \ge \mathbf{P}(N=0)^k I_{\{Z_n\le k\}}$$

for all n and k. That is, (5.5.1) holds in this case for $\eta_k=\mathbf{P}(N=0)^k>0$. As shown already, this implies that with probability one either $Z_n\to\infty$ or $Z_n=0$ for all n large enough. $\square$

The generating function

$$L(s)=\mathbf{E}[s^N]=\mathbf{P}(N=0)+\sum_{k=1}^{\infty}\mathbf{P}(N=k)s^k \tag{5.5.2}$$

plays a key role in analyzing the branching process. In this task, we employ the following martingales associated with branching process.

Lemma 5.5.4. *Suppose $1>\mathbf{P}(N=0)>0$. Then, $(X_n,\mathcal{F}_n)$ is a martingale where $X_n=m_N^{-n}Z_n$. In the super-critical case we also have the martingale $(M_n,\mathcal{F}_n)$ for $M_n=\rho^{Z_n}$ and $\rho\in(0,1)$ the unique solution of $s=L(s)$. The same applies in the sub-critical case if there exists a solution $\rho\in(1,\infty)$ of $s=L(s)$.*

Proof. Since the value of Z_n is a non-random function of $\{N_j^{(k)}, k\le n, j=1,2,\ldots\}$, it follows that both X_n and M_n are $\mathcal{F}_n$-adapted. We proceed to show by induction on n that the non-negative processes Z_n and s^{Z_n} for each $s>0$ such that $L(s)\le\max(s,1)$ are integrable with

$$\mathbf{E}[Z_{n+1}|\mathcal{F}_n]=m_N Z_n\,, \qquad \mathbf{E}[s^{Z_{n+1}}|\mathcal{F}_n]=L(s)^{Z_n}\,. \tag{5.5.3}$$

Indeed, recall that the i.i.d. random variables $N_j^{(n+1)}$ of finite mean m_N are independent of $\mathcal{F}_n$ on which Z_n is measurable. Hence, by linearity of the expectation it follows that for any $A\in\mathcal{F}_n$,

$$\mathbf{E}[Z_{n+1}I_A]=\sum_{j=1}^{\infty}\mathbf{E}[N_j^{(n+1)}I_{\{Z_n\ge j\}}I_A]=\sum_{j=1}^{\infty}\mathbf{E}[N_j^{(n+1)}]\mathbf{E}[I_{\{Z_n\ge j\}}I_A]=m_N\mathbf{E}[Z_nI_A]\,.$$

This verifies the integrability of $Z_n\ge 0$ as well as the identity $\mathbf{E}[Z_{n+1}|\mathcal{F}_n]=m_NZ_n$ of (5.5.3), which amounts to the martingale condition $\mathbf{E}[X_{n+1}|\mathcal{F}_n]=X_n$ for $X_n=m_N^{-n}Z_n$. Similarly, fixing $s>0$,

$$s^{Z_{n+1}}=\sum_{\ell=0}^{\infty}I_{\{Z_n=\ell\}}\prod_{j=1}^{\ell}s^{N_j^{(n+1)}}\,.$$

Hence, by linearity of the expectation and independence of $s^{N_j^{(n+1)}}$ and $\mathcal{F}_n$,

$$\begin{aligned}\mathbf{E}[s^{Z_{n+1}}I_A]&=\sum_{\ell=0}^{\infty}\mathbf{E}[I_{\{Z_n=\ell\}}I_A\prod_{j=1}^{\ell}s^{N_j^{(n+1)}}]\\&=\sum_{\ell=0}^{\infty}\mathbf{E}[I_{\{Z_n=\ell\}}I_A]\prod_{j=1}^{\ell}\mathbf{E}[s^{N_j^{(n+1)}}]=\sum_{\ell=0}^{\infty}\mathbf{E}[I_{\{Z_n=\ell\}}I_A]L(s)^{\ell}=\mathbf{E}[L(s)^{Z_n}I_A]\,.\end{aligned}$$

Since $Z_n\ge 0$ and $L(s)\le\max(s,1)$ this implies that $\mathbf{E}s^{Z_{n+1}}\le 1+\mathbf{E}s^{Z_n}$ and the integrability of s^{Z_n} follows by induction on n. Given that s^{Z_n} is integrable and the preceding identity holds for all $A \in \mathcal{F}_n$, we have thus verified the right identity in (5.5.3), which in case $s = L(s)$ is precisely the martingale condition for $M_n = s^{Z_n}$.

Finally, to prove that $s = L(s)$ has a unique solution in $(0,1)$ when $m_N = \mathbf{E}N > 1$, note that the function $s \mapsto L(s)$ of (5.5.2) is continuous and bounded on $[0,1]$. Further, since $L(1) = 1$ and $L'(1) = \mathbf{E}N > 1$, it follows that $L(s) < s$ for some $0 < s < 1$. With $L(0) = \mathbf{P}(N = 0) > 0$ we have by continuity that $s = L(s)$ for some $s \in (0,1)$. To show the uniqueness of such solution note that $\mathbf{E}N > 1$ implies that $\mathbf{P}(N = k) > 0$ for some $k > 1$, so $L''(s) = \sum_{k=2}^{\infty} k(k-1)\mathbf{P}(N = k)s^{k-2}$ is positive and finite on $(0,1)$. Consequently, $L(\cdot)$ is strictly convex there. Hence, if $\rho \in (0,1)$ is such that $\rho = L(\rho)$, then $L(s) < s$ for $s \in (\rho, 1)$, so such a solution $\rho \in (0,1)$ is unique. □

Remark. Since $X_n = m_N^{-n} Z_n$ is a martingale with $X_0 = 1$, it follows that $\mathbf{E}Z_n = m_N^n$ for all $n \geq 0$. Thus, a sub-critical branching process, i.e. when $m_N < 1$, has mean total population size

$$\mathbf{E}[\sum_{n=0}^{\infty} Z_n] = \sum_{n=0}^{\infty} m_N^n = \frac{1}{1 - m_N} < \infty \,,$$

which is finite.

We now determine the extinction probabilities for branching processes.

Proposition 5.5.5. *Suppose* $1 > \mathbf{P}(N = 0) > 0$. *If* $m_N \leq 1$ *then* $p_{\text{ex}} = 1$. *In contrast, if* $m_N > 1$ *then* $p_{\text{ex}} = \rho$, *with* $m_N^{-n} Z_n \overset{a.s.}{\to} X_\infty$ *and* $Z_n \overset{a.s.}{\to} Z_\infty \in \{0, \infty\}$.

Remark. In words, we find that for sub-critical and non-degenerate critical branching processes the population eventually dies off, whereas non-degenerate super-critical branching processes survive forever with positive probability and conditional upon such survival their population size grows unboundedly in time.

Proof. Applying Doob's martingale convergence theorem to the non-negative MG X_n of Lemma 5.5.4 we have that $X_n \overset{a.s.}{\to} X_\infty$ with X_∞ almost surely finite. In case $m_N \leq 1$ this implies that $Z_n = m_N^n X_n$ is almost surely bounded (in n), hence by Lemma 5.5.3 necessarily $Z_n = 0$ for all large n, i.e. $p_{\text{ex}} = 1$. In case $m_N > 1$ we have by Doob's martingale convergence theorem that $M_n \overset{a.s.}{\to} M_\infty$ for the non-negative MG $M_n = \rho^{Z_n}$ of Lemma 5.5.4. Since $\rho \in (0,1)$ and $Z_n \geq 0$, it follows that this MG is bounded by one, hence U.I. and with $Z_0 = 1$ it follows that $\mathbf{E}M_\infty = \mathbf{E}M_0 = \rho$ (see Theorem 5.3.12). Recall Lemma 5.5.3 that $Z_n \overset{a.s.}{\to} Z_\infty \in \{0, \infty\}$, so $M_\infty = \rho^{Z_\infty} \in \{0, 1\}$ with

$$p_{\text{ex}} = \mathbf{P}(Z_\infty = 0) = \mathbf{P}(M_\infty = 1) = \mathbf{E}M_\infty = \rho$$

as stated. □

Remark. For a non-degenerate critical branching process (i.e. when $m_N = 1$ and $\mathbf{P}(N = 0) > 0$), we have seen that the martingale $\{Z_n\}$ converges to 0 with probability one, while $\mathbf{E}Z_n = \mathbf{E}Z_0 = 1$. Consequently, this MG is L^1-bounded but not U.I. (for another example, see Exercise 5.2.14). Further, as either $Z_n = 0$ or $Z_n \geq 1$, it follows that in this case $1 = \mathbf{E}(Z_n | Z_n \geq 1)(1 - q_n)$ for $q_n = \mathbf{P}(Z_n = 0)$. Further, here $q_n \uparrow p_{\text{ex}} = 1$ so we deduce that conditional upon non-extinction, the mean population size $\mathbf{E}(Z_n | Z_n \geq 1) = 1/(1 - q_n)$ grows to infinity as $n \to \infty$.

As you show next, if super-critical branching process has a square-integrable offspring distribution then $m_N^{-n} Z_n$ converges in law to a non-degenerate random variable. The Kesten-Stigum $L \log L$-theorem, (which we do not prove here), states that the latter property holds if and only if $\mathbf{E}[N \log N]$ is finite.

Exercise 5.5.6. *Consider a super-critical branching process $\{Z_n\}$ where the number of offspring is of mean $m_N = \mathbf{E}[N] > 1$ and variance $v_N = \mathsf{Var}(N) < \infty$.*

(a) *Compute $\mathbf{E}[X_n^2]$ for $X_n = m_N^{-n} Z_n$.*
(b) *Show that $\mathbf{P}(X_\infty > 0) > 0$ for the a.s. limit X_∞ of the martingale X_n.*
(c) *Show that $\mathbf{P}(X_\infty = 0) = \rho$ and deduce that for a.e. ω, if the branching process survives forever, that is $Z_n(\omega) > 0$ for all n, then $X_\infty(\omega) > 0$.*

The generating function $L(s) = \mathbf{E}[s^N]$ yields information about the laws of Z_n and that of X_∞ of Proposition 5.5.5.

Proposition 5.5.7. *Consider the generating functions $L_n(s) = \mathbf{E}[s^{Z_n}]$ for $s \in [0,1]$ and a branching process $\{Z_n\}$ starting with $Z_0 = 1$. Then, $L_0(s) = s$ and $L_n(s) = L[L_{n-1}(s)]$ for $n \geq 1$ and $L(\cdot)$ of (5.5.2). Consequently, the generating function $\widehat{L}_\infty(s) = \mathbf{E}[s^{X_\infty}]$ of X_∞ is a solution of $\widehat{L}_\infty(s) = L[\widehat{L}_\infty(s^{1/m_N})]$ which converges to one as $s \uparrow 1$.*

Remark. In particular, the probability $q_n = \mathbf{P}(Z_n = 0) = L_n(0)$ that the branching process is extinct after n generations is given by the recursion $q_n = L(q_{n-1})$ for $n \geq 1$, starting at $q_0 = 0$. Since the continuous function $L(s)$ is above s on the interval from zero to the smallest positive solution of $s = L(s)$ it follows that q_n is a monotone non-decreasing sequence that converges to this solution, which is thus the value of p_{ex}. This alternative evaluation of p_{ex} does not use martingales. Though implicit here, it instead relies on the Markov property of the branching process (c.f. Example 6.1.10).

Proof. Recall that $Z_1 = N_1^{(1)}$ and if $Z_1 = k$ then the branching process Z_n for $n \geq 2$ has the same law as the sum of k i.i.d. variables, each having the same law as Z_{n-1} (with the j^{th} such variable counting the number of individuals in the n^{th} generation who are descendants of the j^{th} individual of the first generation). Consequently, $\mathbf{E}[s^{Z_n}|Z_1 = k] = \mathbf{E}[s^{Z_{n-1}}]^k$ for all $n \geq 2$ and $k \geq 0$. Summing over the disjoint events $\{Z_1 = k\}$ we have by the tower property that for $n \geq 2$,

$$L_n(s) = \mathbf{E}[\mathbf{E}(s^{Z_n}|Z_1)] = \sum_{k=0}^{\infty} \mathbf{P}(N = k) L_{n-1}(s)^k = L[L_{n-1}(s)]$$

for $L(\cdot)$ of (5.5.2), as claimed. Obviously, $L_0(s) = s$ and $L_1(s) = \mathbf{E}[s^N] = L(s)$. From this identity we conclude that $\widehat{L}_n(s) = L[\widehat{L}_{n-1}(s^{1/m_N})]$ for $\widehat{L}_n(s) = \mathbf{E}[s^{X_n}]$ and $X_n = m_N^{-n} Z_n$. With $X_n \overset{a.s.}{\to} X_\infty$ we have by bounded convergence that $\widehat{L}_n(s) \to \widehat{L}_\infty(s) = \mathbf{E}[s^{X_\infty}]$, which by the continuity of $r \mapsto L(r)$ on $[0,1]$ is thus a solution of the identity $\widehat{L}_\infty(s) = L[\widehat{L}_\infty(s^{1/m_N})]$. Further, by monotone convergence $\widehat{L}_\infty(s) \uparrow \widehat{L}_\infty(1) = 1$ as $s \uparrow 1$. $\square$

Remark. Of course, $q_n = \mathbf{P}(T \leq n)$ provides the distribution function of the time of extinction $T = \min\{k \geq 0 : Z_k = 0\}$. For example, if N has the Bernoulli(p) distribution for some $0 < p < 1$ then T is merely a Geometric($1 - p$) random variable, but in general the law of T is more involved.

The generating function $\widehat{L}_\infty(\cdot)$ determines the law of $X_\infty \geq 0$ (see Exercise 3.2.40). For example, as you show next, in the special case where N has the Geometric distribution, conditioned on non-extinction X_∞ is an exponential random variable.

Exercise 5.5.8. *Suppose Z_n is a branching process with $Z_0 = 1$ and $N+1$ having a Geometric(p) distribution for some $0 < p < 1$ (that is, $\mathbf{P}(N = k) = p(1-p)^k$ for $k = 0, \ldots$). Here $m = m_N = (1-p)/p$ so the branching process is sub-critical if $p > 1/2$, critical if $p = 1/2$ and super-critical if $p < 1/2$.*

(a) *Check that $L(s) = p/(1-(1-p)s)$ and $\rho = 1/m$. Then verify that $L_n(s) = (pm^n(1-s) + (1-p)s - p)/((1-p)(1-s)m^n + (1-p)s - p)$ except in the critical case for which $L_n(s) = (n-(n-1)s)/((n+1)-ns)$.*
(b) *Show that in the super-critical case $\widehat{L}_\infty(e^{-\lambda}) = \rho + (1-\rho)^2/(\lambda + (1-\rho))$ for all $\lambda \geq 0$ and deduce that conditioned on non-extinction X_∞ has the exponential distribution of parameter $(1-\rho)$.*
(c) *Show that in the sub-critical case $\mathbf{E}[s^{Z_n}|Z_n \neq 0] \to (1-m)s/[1-ms]$ and deduce that then the law of Z_n conditioned upon non-extinction converges weakly to a Geometric($1-m$) distribution.*
(d) *Show that in the critical case $\mathbf{E}[e^{-\lambda Z_n/n}|Z_n \neq 0] \to 1/(1+\lambda)$ for all $\lambda \geq 0$ and deduce that then the law of $n^{-1}Z_n$ conditioned upon non-extinction converges weakly to an exponential distribution (of parameter one).*

The following exercise demonstrates that martingales are also useful in the study of Galton-Watson trees.

Exercise 5.5.9. *Consider a super-critical branching process Z_n such that $1 \leq N \leq \ell$ for some non-random finite ℓ. A vertex of the corresponding* Galton-Watson tree *T_∞ is called a* branch point *if it has more than one offspring. For each vertex $v \in T_\infty$ let $C(v)$ count the number of branch points one encounters when traversing along a path from the root of the tree to v (possibly counting the root, but not counting v among these branch points).*

(a) *Let ∂T_n denote the set of vertices in T_∞ of distance n from the root. Show that for each $\lambda > 0$,*

$$X_n := M(\lambda)^{-n} \sum_{v \in \partial T_n} e^{-\lambda C(v)}$$

is a martingale when $M(\lambda) = m_N e^{-\lambda} + \mathbf{P}(N = 1)(1 - e^{-\lambda})$.
(b) *Let $B_n = \min\{C(v) : v \in \partial T_n\}$. Show that a.s. $\liminf_{n\to\infty} n^{-1}B_n \geq \delta$ where $\delta > 0$ is non-random (and possibly depends on the offspring distribution).*

5.5.2. Product martingales and Radon-Nikodym derivatives. We start with an explicit characterization of uniform integrability for the product martingale of Example 5.1.10.

Theorem 5.5.10 (Kakutani's Theorem)**.** *Let M_∞ denote the a.s. limit of the product martingale $M_n = \prod_{k=1}^n Y_k$, with $M_0 = 1$ and independent, integrable $Y_k \geq 0$ such that $\mathbf{E}Y_k = 1$ for all $k \geq 1$. By Jensen's inequality, $a_k = \mathbf{E}[\sqrt{Y_k}]$ is in $(0, 1]$*

for all $k \geq 1$. The following five statements are then equivalent:

$$\text{(a) } \{M_n\} \text{ is } U.I., \quad \text{(b) } M_n \xrightarrow{L^1} M_\infty; \quad \text{(c) } \mathbf{E}M_\infty = 1;$$
$$\text{(d) } \prod_k a_k > 0\,; \quad \text{(e) } \sum_k (1 - a_k) < \infty\,,$$

and if any (every) one of them fails, then $M_\infty = 0$ a.s.

Proof. Statement (a) implies statement (b) because any U.I. martingale converges in L^1 (see Theorem 5.3.12). Further, the L^1 convergence per statement (b) implies that $\mathbf{E}M_n \to \mathbf{E}M_\infty$ and since $\mathbf{E}M_n = \mathbf{E}M_0 = 1$ for all n, this results with $\mathbf{E}M_\infty = 1$ as well, which is statement (c).

Considering the non-negative martingale $N_n = \prod_{k=1}^n (\sqrt{Y_k}/a_k)$ we next show that (c) implies (d) by proving the contra-positive. Indeed, by Doob's convergence theorem $N_n \xrightarrow{a.s.} N_\infty$ with N_∞ finite a.s. Hence, if statement (d) fails to hold (that is, $\prod_{k=1}^n a_k \to 0$), then $M_n = N_n^2 (\prod_{k=1}^n a_k)^2 \xrightarrow{a.s.} 0$. So in this case $M_\infty = 0$ a.s. and statement (c) also fails to hold.

In contrast, if statement (d) holds then $\{N_n\}$ is L^2-bounded since for all n,

$$\mathbf{E}N_n^2 = (\prod_{k=1}^n a_k)^{-2}\mathbf{E}M_n \leq \big(\prod_k a_k\big)^{-2} = c < \infty\,.$$

Thus, with $M_k \leq N_k^2$ it follows by the L^2-maximal inequality that for all n,

$$\mathbf{E}\big[\max_{k=0}^n M_k\big] \leq \mathbf{E}\big[\max_{k=0}^n N_k^2\big] \leq 4\mathbf{E}[N_n^2] \leq 4c\,.$$

Hence, $M_k \geq 0$ are such that $\sup_k M_k$ is integrable and in particular, $\{M_n\}$ is U.I. (that is, (a) holds).

Finally, to see why the statements (d) and (e) are equivalent note that upon applying the Borel Cantelli lemmas for independent events A_n with $\mathbf{P}(A_n) = 1 - a_n$ the divergence of the series $\sum_k (1 - a_k)$ is equivalent to $\mathbf{P}(A_n^c \text{ eventually}) = 0$, which for strictly positive a_k is equivalent to $\prod_k a_k = 0$. $\square$

We next consider another martingale that is key to the study of likelihood ratios in sequential statistics. To this end, let $\mathbf{P}$ and $\mathbf{Q}$ be two probability measures on the same measurable space $(\Omega, \mathcal{F}_\infty)$ with $\mathbf{P}_n = \mathbf{P}\big|_{\mathcal{F}_n}$ and $\mathbf{Q}_n = \mathbf{Q}\big|_{\mathcal{F}_n}$ denoting the restrictions of $\mathbf{P}$ and $\mathbf{Q}$ to a filtration $\mathcal{F}_n \uparrow \mathcal{F}_\infty$.

Theorem 5.5.11. *Suppose $\mathbf{Q}_n \ll \mathbf{P}_n$ for all n, with $M_n = d\mathbf{Q}_n/d\mathbf{P}_n$ denoting the corresponding Radon-Nikodym derivatives on $(\Omega, \mathcal{F}_n)$. Then,*

(a) *$(M_n, \mathcal{F}_n)$ is a martingale on the probability space $(\Omega, \mathcal{F}_\infty, \mathbf{P})$ where $M_n \xrightarrow{a.s.} M_\infty$ as $n \to \infty$ and M_∞ is $\mathbf{P}$-a.s. finite.*
(b) *If $\{M_n\}$ is uniformly $\mathbf{P}$-integrable then $\mathbf{Q} \ll \mathbf{P}$ and $d\mathbf{Q}/d\mathbf{P} = M_\infty$.*
(c) *More generally, the Lebesgue decomposition of $\mathbf{Q}$ to its absolutely continuous and singular parts with respect to $\mathbf{P}$ is*

$$\mathbf{Q} = \mathbf{Q}_{\mathrm{ac}} + \mathbf{Q}_s = M_\infty \mathbf{P} + I_{\{M_\infty = \infty\}}\mathbf{Q}\,. \tag{5.5.4}$$

Remark. From the decomposition of (5.5.4) it follows that if $\mathbf{Q} \ll \mathbf{P}$ then both $\mathbf{Q}(M_\infty < \infty) = 1$ and $\mathbf{P}(M_\infty) = 1$ while if $\mathbf{Q} \perp \mathbf{P}$ then both $\mathbf{Q}(M_\infty = \infty) = 1$ and $\mathbf{P}(M_\infty = 0) = 1$.

Example 5.5.12. *Suppose $\mathcal{F}_n = \sigma(\Pi_n)$ and the countable partitions $\Pi_n = \{A_{i,n}\} \subset \mathcal{F}$ of Ω are nested (that is, for each n the partition Π_{n+1} is a refinement of Π_n). It is not hard to check directly that*

$$M_n = \sum_{\{i:\mathbf{P}(A_{i,n})>0\}} \frac{\mathbf{Q}(A_{i,n})}{\mathbf{P}(A_{i,n})} I_{A_{i,n}} ,$$

is an $\mathcal{F}_n$-sup-MG for $(\Omega, \mathcal{F}, \mathbf{P})$ and is further an $\mathcal{F}_n$-martingale if $\mathbf{Q}(A_{i,n}) = 0$ whenever $\mathbf{P}(A_{i,n}) = 0$ (which is precisely the assumption made in Theorem 5.5.11). We have seen this construction in Exercise 5.3.20, where Π_n are the dyadic partitions of $\Omega = [0,1)$, $\mathbf{P}$ is taken to be Lebesgue's measure on $[0,1)$ and $\mathbf{Q}([s,t)) = x(t) - x(s)$ is the signed measure associated with the function $x(\cdot)$.

Proof. (a). By the Radon-Nikodym theorem, $M_n \in m\mathcal{F}_n$ is non-negative and $\mathbf{P}$-integrable (since $\mathbf{P}_n(M_n) = \mathbf{Q}_n(\Omega) = 1$). Further, $\mathbf{Q}(A) = \mathbf{Q}_n(A) = M_n\mathbf{P}_n(A) = M_n\mathbf{P}(A)$ for all $A \in \mathcal{F}_n$. In particular, if $k \le n$ and $A \in \mathcal{F}_k$ then (since $\mathcal{F}_k \subseteq \mathcal{F}_n$),

$$\mathbf{P}(M_n I_A) = \mathbf{Q}(A) = \mathbf{P}(M_k I_A) ,$$

so in $(\Omega, \mathcal{F}_\infty, \mathbf{P})$ we have $M_k = \mathbf{E}[M_n|\mathcal{F}_k]$ by definition of the conditional expectation. Finally, by Doob's convergence theorem the non-negative MG M_n converges $\mathbf{P}$-a.s. to M_∞ which is $\mathbf{P}$-a.s. finite.
(b). We have seen already that if $A \in \mathcal{F}_k$ then $\mathbf{Q}(A) = \mathbf{P}(M_n I_A)$ for all $n \ge k$. Hence, if $\{M_n\}$ is further uniformly $\mathbf{P}$-integrable then also $\mathbf{P}(M_n I_A) \to \mathbf{P}(M_\infty I_A)$, so taking $n \to \infty$ we deduce that in this case $\mathbf{Q}(A) = \mathbf{P}(M_\infty I_A)$ for any $A \in \cup_k \mathcal{F}_k$ (and in particular for $A = \Omega$). Since the probability measures $\mathbf{Q}$ and $M_\infty\mathbf{P}$ then coincide on the π-system $\cup_k \mathcal{F}_k$ they agree also on the σ-algebra $\mathcal{F}_\infty$ generated by this π-system (recall Proposition 1.1.39).
(c). To deal with the general case, where M_n is not necessarily uniformly $\mathbf{P}$-integrable, consider the probability measure $\mathbf{S} = (\mathbf{P}+\mathbf{Q})/2$ and its restrictions $\mathbf{S}_n = (\mathbf{P}_n + \mathbf{Q}_n)/2$ to $\mathcal{F}_n$. As $\mathbf{P}_n \ll \mathbf{S}_n$ and $\mathbf{Q}_n \ll \mathbf{S}_n$ there exist $V_n = d\mathbf{P}_n/d\mathbf{S}_n \ge 0$ and $W_n = d\mathbf{Q}_n/d\mathbf{S}_n \ge 0$ such that $V_n + W_n = 2$. Per part (a) the bounded $(V_n, \mathcal{F}_n)$ and $(W_n, \mathcal{F}_n)$ are martingales on $(\Omega, \mathcal{F}_\infty, \mathbf{S})$, having the $\mathbf{S}$-a.s. finite limits V_∞ and W_∞, respectively. Further, as shown in part (b), $V_\infty = d\mathbf{P}/d\mathbf{S}$ and $W_\infty = d\mathbf{Q}/d\mathbf{S}$. Recall that $W_n\mathbf{S}_n = \mathbf{Q}_n = M_n\mathbf{P}_n = M_n V_n \mathbf{S}_n$, so $\mathbf{S}$-a.s. $M_n V_n = W_n = 2 - V_n$ for all n. Consequently, $\mathbf{S}$-a.s. $V_n > 0$ and $M_n = (2 - V_n)/V_n$. Considering $n \to \infty$ we deduce that $\mathbf{S}$-a.s. $M_n \to M_\infty = (2 - V_\infty)/V_\infty = W_\infty/V_\infty$, possibly infinite, and $I_{\{M_\infty < \infty\}} = I_{\{V_\infty > 0\}}$. Thus,

$$\begin{aligned} \mathbf{Q} = W_\infty \mathbf{S} &= I_{\{V_\infty > 0\}} M_\infty V_\infty \mathbf{S} + I_{\{V_\infty = 0\}} W_\infty \mathbf{S} \\ &= I_{\{M_\infty < \infty\}} M_\infty \mathbf{P} + I_{\{M_\infty = \infty\}} \mathbf{Q} , \end{aligned}$$

and since M_∞ is finite $\mathbf{P}$-a.s. this is precisely the stated Lebesgue decomposition of $\mathbf{Q}$ with respect to $\mathbf{P}$. □

Combining Theorem 5.5.11 and Kakutani's theorem we next deduce that if the marginals of one infinite product measure are absolutely continuous with respect to those of another, then either the former product measure is absolutely continuous with respect to the latter, or these two measures are mutually singular. This dichotomy is a key result in the treatment by theoretical statistics of the problem of hypothesis testing (with independent observables under both the null hypothesis and the alternative hypothesis).

Proposition 5.5.13. *Suppose that* $\mathbf{P}$ *and* $\mathbf{Q}$ *are product measures on* $(\mathbb{R}^{\mathbf{N}}, \mathcal{B}_c)$ *which make the coordinates* $X_n(\omega) = \omega_n$ *independent with the respective laws* $\mathbf{Q} \circ X_k^{-1} \ll \mathbf{P} \circ X_k^{-1}$ *for each* $k \in \mathbf{N}$. *Let* $Y_k(\omega) = d(\mathbf{Q} \circ X_k^{-1})/d(\mathbf{P} \circ X_k^{-1})(X_k(\omega))$ *then denote the* likelihood ratios *of the marginals. Then,* $M_\infty = \prod_k Y_k$ *exists a.s. under both* $\mathbf{P}$ *and* $\mathbf{Q}$. *If* $\alpha = \prod_k \mathbf{P}(\sqrt{Y_k})$ *is positive then* $\mathbf{Q}$ *is absolutely continuous with respect to* $\mathbf{P}$ *with* $d\mathbf{Q}/d\mathbf{P} = M_\infty$, *whereas if* $\alpha = 0$ *then* $\mathbf{Q}$ *is singular with respect to* $\mathbf{P}$ *such that* $\mathbf{Q}$*-a.s.* $M_\infty = \infty$ *while* $\mathbf{P}$*-a.s.* $M_\infty = 0$.

Remark 5.5.14. Note that the preceding Y_k are identically distributed when both $\mathbf{P}$ and $\mathbf{Q}$ are products of i.i.d. random variables. Hence in this case $\alpha > 0$ if and only if $\mathbf{P}(\sqrt{Y_1}) = 1$, which with $\mathbf{P}(Y_1) = 1$ is equivalent to $\mathbf{P}[(\sqrt{Y_1} - 1)^2] = 0$, i.e. to having $\mathbf{P}$-a.s. $Y_1 = 1$. The latter condition implies that $\mathbf{P}$-a.s. $M_\infty = 1$, so $\mathbf{Q} = \mathbf{P}$. We thus deduce that any $\mathbf{Q} \neq \mathbf{P}$ that are both products of i.i.d. random variables, are mutually singular, and for n large enough the likelihood test of comparing M_n to a fixed threshold decides correctly between the two hypothesis regarding the law of $\{X_k\}$, since $\mathbf{P}$-a.s. $M_n \to 0$ while $\mathbf{Q}$-a.s. $M_n \to \infty$.

PROOF. We are in the setting of Theorem 5.5.11 for $\Omega = \mathbb{R}^{\mathbf{N}}$ and the filtration

$$\mathcal{F}_n^{\mathbf{X}} = \sigma(X_k : 1 \leq k \leq n) \uparrow \mathcal{F}^{\mathbf{X}} = \sigma(X_k, k < \infty) = \mathcal{B}_c$$

(c.f. Exercise 1.2.14 and the definition of $\mathcal{B}_c$ preceding Kolmogorov's extension theorem). Here $M_n = d\mathbf{Q}_n/d\mathbf{P}_n = \prod_{k=1}^n Y_k$ and the mutual independence of $\{X_k\}$ imply that $\{Y_k\}$ are both $\mathbf{P}$-independent and $\mathbf{Q}$-independent (c.f. part (b) of Exercise 4.1.8). In the course of proving part (c) of Theorem 5.5.11 we have shown that $M_n \to M_\infty$ both $\mathbf{P}$-a.s. and $\mathbf{Q}$-a.s. Further, recall part (a) of Theorem 5.5.11 that M_n is a martingale on $(\Omega, \mathcal{F}^{\mathbf{X}}, \mathbf{P})$. From Kakutani's theorem we know that the product martingale $\{M_n\}$ is uniformly $\mathbf{P}$-integrable when $\alpha > 0$ (see (d) implying (a) there), whereas if $\alpha = 0$ then $\mathbf{P}$-a.s. $M_\infty = 0$. By part (b) of Theorem 5.5.11 the uniform $\mathbf{P}$-integrability of M_n results with $\mathbf{Q} = M_\infty \mathbf{P} \ll \mathbf{P}$. In contrast, when $\mathbf{P}$-a.s. $M_\infty = 0$ we get from the decomposition of part (c) of Theorem 5.5.11 that $\mathbf{Q}_{\rm ac} = 0$ and $\mathbf{Q} = I_{\{M_\infty = \infty\}} \mathbf{Q}$ so in this case $\mathbf{Q}$-a.s. $M_\infty = \infty$ and $\mathbf{Q} \perp \mathbf{P}$. □

Here is a concrete application of the preceding proposition.

Exercise 5.5.15. *Suppose* $\mathbf{P}$ *and* $\mathbf{Q}$ *are two product probability measures on the set* $\Omega_\infty = \{0,1\}^{\mathbf{N}}$ *of infinite binary sequences equipped with the product* σ*-algebra generated by its cylinder sets, with* $p_k = \mathbf{P}(\{\omega : \omega_k = 1\})$ *strictly between zero and one and* $q_k = \mathbf{Q}(\{\omega : \omega_k = 1\}) \in [0,1]$.

(a) *Deduce from Proposition 5.5.13 that* $\mathbf{Q}$ *is absolutely continuous with respect to* $\mathbf{P}$ *if and only if* $\sum_k (1 - \sqrt{p_k q_k} - \sqrt{(1-p_k)(1-q_k)})$ *is finite.*
(b) *Show that if* $\sum_k |p_k - q_k|$ *is finite then* $\mathbf{Q}$ *is absolutely continuous with respect to* $\mathbf{P}$.
(c) *Show that if* $p_k, q_k \in [\varepsilon, 1-\varepsilon]$ *for some* $\varepsilon > 0$ *and all* k, *then* $\mathbf{Q} \ll \mathbf{P}$ *if and only if* $\sum_k (p_k - q_k)^2 < \infty$.
(d) *Show that if* $\sum_k q_k < \infty$ *and* $\sum_k p_k = \infty$ *then* $\mathbf{Q} \perp \mathbf{P}$ *so in general the condition* $\sum_k (p_k - q_k)^2 < \infty$ *is not sufficient for absolute continuity of* $\mathbf{Q}$ *with respect to* $\mathbf{P}$.

In the spirit of Theorem 5.5.11, as you show next, a positive martingale $(Z_n, \mathcal{F}_n)$ induces a collection of probability measures $\mathbf{Q}_n$ that are *equivalent* to $\mathbf{P}_n = \mathbf{P}\big|_{\mathcal{F}_n}$ (i.e. both $\mathbf{Q}_n \ll \mathbf{P}_n$ and $\mathbf{P}_n \ll \mathbf{Q}_n$), and satisfy a certain *martingale Bayes rule*.

In particular, the following discrete time analog of *Girsanov's theorem*, shows that such construction can significantly simplify certain computations upon moving from $\mathbf{P}_n$ to $\mathbf{Q}_n$.

Exercise 5.5.16. *Suppose $(Z_n, \mathcal{F}_n)$ is a (strictly) positive MG on $(\Omega, \mathcal{F}, \mathbf{P})$, normalized so that $\mathbf{E}Z_0 = 1$. Let $\mathbf{P}_n = \mathbf{P}\big|_{\mathcal{F}_n}$ and consider the equivalent probability measure $\mathbf{Q}_n$ on $(\Omega, \mathcal{F}_n)$ of Radon-Nikodym derivative $d\mathbf{Q}_n/d\mathbf{P}_n = Z_n$.*

(a) *Show that $\mathbf{Q}_k = \mathbf{Q}_n\big|_{\mathcal{F}_k}$ for any $0 \le k \le n$.*

(b) *Fixing $0 \le k \le m \le n$ and $Y \in L^1(\Omega, \mathcal{F}_m, \mathbf{P})$ show that $\mathbf{Q}_n$-a.s. (hence also $\mathbf{P}$-a.s.), $\mathbf{E}_{\mathbf{Q}_n}[Y|\mathcal{F}_k] = \mathbf{E}[YZ_m|\mathcal{F}_k]/Z_k$.*

(c) *For $\mathcal{F}_n = \mathcal{F}_n^{\boldsymbol{\xi}}$, the canonical filtration of i.i.d. standard normal variables $\{\xi_k\}$ and any bounded, $\mathcal{F}_n^{\boldsymbol{\xi}}$-predictable V_n, consider the measures $\mathbf{Q}_n$ induced by the exponential martingale $Z_n = \exp(Y_n - \frac{1}{2}\sum_{k=1}^n V_k^2)$, where $Y_n = \sum_{k=1}^n \xi_k V_k$. Show that $\underline{X}$ of coordinates $X_m = \sum_{k=1}^m (\xi_k - V_k)$, $1 \le m \le n$, is under $\mathbf{Q}_n$ a Gaussian random vector whose law is the same as that of $\{\sum_{k=1}^m \xi_k : 1 \le m \le n\}$ under $\mathbf{P}$.*
H*int: Use characteristic functions.*

5.5.3. Reversed martingales and 0-1 laws. Reversed martingales which we next define, though less common than martingales, are key tools in the proof of many asymptotics (e.g. 0-1 laws).

Definition 5.5.17. *A* reversed martingale *(in short* RMG*), is a martingale indexed by non-negative integers. That is, integrable X_n, $n \le 0$, adapted to a filtration $\mathcal{F}_n$, $n \le 0$, such that $\mathbf{E}[X_{n+1}|\mathcal{F}_n] = X_n$ for all $n \le -1$. We denote by $\mathcal{F}_n \downarrow \mathcal{F}_{-\infty}$ a filtration $\{\mathcal{F}_n\}_{n\le 0}$ and the associated σ-algebra $\mathcal{F}_{-\infty} = \bigcap_{n\le 0}\mathcal{F}_n$ such that the relation $\mathcal{F}_k \subseteq \mathcal{F}_\ell$ applies for any $-\infty \le k \le \ell \le 0$.*

Remark. One similarly defines reversed subMG-s (and supMG-s), by replacing $\mathbf{E}[X_{n+1}|\mathcal{F}_n] = X_n$ for all $n \le -1$ with the condition $\mathbf{E}[X_{n+1}|\mathcal{F}_n] \ge X_n$, for all $n \le -1$ (or the condition $\mathbf{E}[X_{n+1}|\mathcal{F}_n] \le X_n$, for all $n \le -1$, respectively). Since $(X_{n+k}, \mathcal{F}_{n+k})$, $k = 0, \ldots, -n$, is then a MG (or sub-MG, or sup-MG), any result about subMG-s, sup-MG-s and MG-s that does not involve the limit as $n \to \infty$ (such as, Doob's decomposition, maximal and up-crossing inequalities), shall apply also for reversed subMG-s, reversed supMG-s and RMG-s.

As we see next, RMG-s are the dual of Doob's martingales (with time moving backwards), hence U.I. and as such each RMG converges both a.s. and in L^1 as $n \to -\infty$.

Theorem 5.5.18 (Lévy's downward theorem)**.** *With X_0 integrable, $(X_n, \mathcal{F}_n)$, $n \le 0$ is a RMG if and only if $X_n = \mathbf{E}[X_0|\mathcal{F}_n]$ for all $n \le 0$. Further, $\mathbf{E}[X_0|\mathcal{F}_n] \to \mathbf{E}[X_0|\mathcal{F}_{-\infty}]$ almost surely and in L^1 when $n \to -\infty$.*

Remark. Actually, $(X_n, \mathcal{F}_n)$ is a RMG for $X_n = \mathbf{E}[Y|\mathcal{F}_n]$, $n \ge 0$ and *any* integrable Y (possibly $Y \notin m\mathcal{F}_0$). Further, $\mathbf{E}[Y|\mathcal{F}_n] \to \mathbf{E}[Y|\mathcal{F}_{-\infty}]$ almost surely and in L^1. This is merely a restatement of Lévy's downward theorem, since for $X_0 = \mathbf{E}[Y|\mathcal{F}_0]$ we have by the tower property that $\mathbf{E}[Y|\mathcal{F}_n] = \mathbf{E}[X_0|\mathcal{F}_n]$ for any $-\infty \le n \le 0$.

Proof. Suppose $(X_n, \mathcal{F}_n)$ is a RMG. Then, fixing $n < 0$ and applying Proposition 5.1.20 for the MG $(X_{n+k}, \mathcal{F}_{n+k})$, $k = 0, \ldots, -n$ (taking there $\ell = -n > m =$

0), we deduce that $\mathbf{E}[X_0|\mathcal{F}_n] = X_n$. Conversely, suppose $X_n = \mathbf{E}[X_0|\mathcal{F}_n]$ for X_0 integrable and all $n \le 0$. Then, $X_n \in L^1(\Omega, \mathcal{F}_n, \mathbf{P})$ by the definition of C.E. and further, with $\mathcal{F}_n \subseteq \mathcal{F}_{n+1}$, we have by the tower property that

$$X_n = \mathbf{E}[X_0|\mathcal{F}_n] = \mathbf{E}[\mathbf{E}(X_0|\mathcal{F}_{n+1})|\mathcal{F}_n] = \mathbf{E}[X_{n+1}|\mathcal{F}_n],$$

so any such $(X_n, \mathcal{F}_n)$ is a RMG.

Setting hereafter $X_n = \mathbf{E}[X_0|\mathcal{F}_n]$, note that for each $n \le 0$ and $a < b$, by Doob's up-crossing inequality for the MG $(X_{n+k}, \mathcal{F}_{n+k})$, $k = 0, \ldots, -n$, we have that $\mathbf{E}(U_n[a,b]) \le (b-a)^{-1}\mathbf{E}[(X_0-a)_-]$ (where $U_n[a,b]$ denotes the number of up-crossings of the interval $[a,b]$ by $\{X_k(\omega), k = n, \ldots, 0\}$). By monotone convergence this implies that $\mathbf{E}(U_{-\infty}[a,b]) \le (b-a)^{-1}\mathbf{E}[(X_0-a)_-]$ is finite (for any $a < b$). Repeating the proof of Lemma 5.3.1, now for $n \to -\infty$, we thus deduce that $X_n \stackrel{a.s.}{\to} X_{-\infty}$ as $n \to -\infty$. Recall Proposition 4.2.33 that $\{\mathbf{E}[X_0|\mathcal{F}_n]\}$ is U.I. hence by Vitali's convergence theorem also $X_n \stackrel{L^1}{\to} X_{-\infty}$ when $n \to -\infty$ (and in particular the random variable $X_{-\infty}$ is integrable).

We now complete the proof by showing that $X_{-\infty} = \mathbf{E}[X_0|\mathcal{F}_{-\infty}]$. Indeed, fixing $k \le 0$, since $X_n \in m\mathcal{F}_k$ for all $n \le k$ it follows that $X_{-\infty} = \limsup_{n\to-\infty} X_n$ is also in $m\mathcal{F}_k$. This applies for all $k \le 0$, hence $X_{-\infty} \in m[\bigcap_{k\le 0}\mathcal{F}_k] = m\mathcal{F}_{-\infty}$. Further, $\mathbf{E}[X_n I_A] \to \mathbf{E}[X_{-\infty} I_A]$ for any $A \in \mathcal{F}_{-\infty}$ (by the L^1 convergence of X_n to $X_{-\infty}$), and as $A \in \mathcal{F}_{-\infty} \subseteq \mathcal{F}_n$ also $\mathbf{E}[X_0 I_A] = \mathbf{E}[X_n I_A]$ for all $n \le 0$. Thus, $\mathbf{E}[X_{-\infty} I_A] = \mathbf{E}[X_0 I_A]$ for all $A \in \mathcal{F}_{-\infty}$, so by the definition of conditional expectation, $X_{-\infty} = \mathbf{E}[X_0|\mathcal{F}_{-\infty}]$. □

Similarly to Lévy's upward theorem, as you show next, Lévy's downward theorem can be extended to accommodate a dominated sequences of random variables and if $X_0 \in L^p$ for some $p > 1$, then $X_n \stackrel{L^p}{\to} X_{-\infty}$ as $n \to -\infty$ (which is the analog of Doob's L^p martingale convergence).

Exercise 5.5.19. *Suppose $\mathcal{F}_n \downarrow \mathcal{F}_{-\infty}$ and $Y_n \stackrel{a.s.}{\to} Y_{-\infty}$ as $n \to -\infty$. Show that if $\sup_n |Y_n|$ is integrable, then $\mathbf{E}[Y_n|\mathcal{F}_n] \stackrel{a.s.}{\to} \mathbf{E}[Y_{-\infty}|\mathcal{F}_{-\infty}]$ when $n \to -\infty$.*

Exercise 5.5.20. *Suppose $(X_n, \mathcal{F}_n)$ is a RMG. Show that if $\mathbf{E}|X_0|^p$ is finite and $p > 1$, then $X_n \stackrel{L^p}{\to} X_{-\infty}$ when $n \to -\infty$*

Not all reversed sub-MGs are U.I. but here is an explicit characterization of those that are.

Exercise 5.5.21. *Show that a reversed sub-MG $\{X_n\}$ is U.I. if and only if $\inf_n \mathbf{E}X_n$ is finite.*

Our first application of RMG-s is to provide an alternative proof of the strong law of large numbers of Theorem 2.3.3, with the added bonus of L^1 convergence.

Theorem 5.5.22 (Strong law of large numbers). *Suppose $S_n = \sum_{k=1}^n \xi_k$ for i.i.d. integrable $\{\xi_k\}$. Then, $n^{-1}S_n \to \mathbf{E}\xi_1$ a.s. and in L^1 when $n \to \infty$.*

Proof. Let $X_{-m} = (m+1)^{-1}S_{m+1}$ for $m \ge 0$, and define the corresponding filtration $\mathcal{F}_{-m} = \sigma(X_{-k}, k \ge m)$. Recall part (a) of Exercise 4.4.8, that $X_n = \mathbf{E}[\xi_1|X_n]$ for each $n \le 0$. Further, clearly $\mathcal{F}_n = \sigma(\mathcal{G}_n, \mathcal{T}_{-n})$ for $\mathcal{G}_n = \sigma(X_n)$ and $\mathcal{T}_\ell = \sigma(\xi_r, r > \ell)$. With $\mathcal{T}^{\mathbf{X}}_{-n}$ independent of $\sigma(\sigma(\xi_1), \mathcal{G}_n)$, we thus have that $X_n = \mathbf{E}[\xi_1|\mathcal{F}_n]$ for each $n \le 0$ (see Proposition 4.2.3). Consequently, $(X_n, \mathcal{F}_n)$ is a RMG which by Lévy's downward theorem converges for $n \to -\infty$ both a.s. and in L^1 to the finite valued random variable $X_{-\infty} = \mathbf{E}[\xi_1|\mathcal{F}_{-\infty}]$. Combining this and the tower property leads to $\mathbf{E}X_{-\infty} = \mathbf{E}\xi_1$ so it only remains to show that $\mathbf{P}(X_{-\infty} \neq c) = 0$ for some non-random constant c. To this end, note that for any ℓ finite,

$$X_{-\infty} = \limsup_{m\to\infty} \frac{1}{m}\sum_{k=1}^{m} \xi_k = \limsup_{m\to\infty} \frac{1}{m}\sum_{k=\ell+1}^{m} \xi_k .$$

Clearly, $X_{-\infty} \in m\mathcal{T}_\ell^{\mathbf{X}}$ for any ℓ so $X_{-\infty}$ is also measurable on the tail σ-algebra $\mathcal{T} = \bigcap_\ell \mathcal{T}_\ell$ of the sequence $\{\xi_k\}$. We complete the proof upon noting that the σ-algebra $\mathcal{T}$ is $\mathbf{P}$-trivial (by Kolmogorov's 0-1 law and the independence of ξ_k), so in particular, a.s. $X_{-\infty}$ equals a non-random constant (see Proposition 1.2.47). $\square$

In this conext, you find next that while any RMG X_{-m} is U.I., it is not necessarily dominated by an integrable variable, and its a.s. convergence may not translate to conditional expectations $\mathbf{E}[X_{-m}|\mathcal{H}]$.

Exercise 5.5.23. *Consider integrable i.i.d. copies of ξ_1, having distribution function $F_{\xi_1}(x) = 1 - x^{-1}(\log x)^{-2}$ for $x \geq e$ and $\mathbf{P}(\xi_1 = -e/(e-1)) = 1 - e^{-1}$, so $\mathbf{E}\xi_1 = 0$. Let $\mathcal{H} = \sigma(A_n, n \geq 3)$ for $A_n = \{\xi_n \geq en/(\log n)\}$ and recall Theorem 5.5.22 that for $m \to \infty$ the U.I. RMG $X_{-m} = (m+1)^{-1}S_{m+1}$ converges a.s. to zero.*

(a) *Verify that $m^{-1}\mathbf{E}[\xi_m|\mathcal{H}] \geq I_{A_m}$ for all $m \geq 3$ and deduce that a.s. $\limsup_{m\to\infty} m^{-1}\mathbf{E}[\xi_m|\mathcal{H}] \geq 1$.*
(b) *Conclude that $\mathbf{E}[X_{-m}|\mathcal{H}]$ does not converge to zero a.s. and $\sup_m |X_{-m}|$ is not integrable.*

In preparation for the Hewitt-Savage 0-1 law and de-Finetti's theorem we now define the exchangeable σ-algebra and random variables.

Definition 5.5.24 (EXCHANGEABLE σ-ALGEBRA AND RANDOM VARIABLES)**.** *Consider the product measurable space $(\mathbb{R}^{\mathbf{N}}, \mathcal{B}_c)$ as in Kolmogorov's extension theorem. Let $\mathcal{E}_m \subseteq \mathcal{B}_c$ denote the σ-algebra of events that are invariant under permutations of the first m coordinates; that is, $A \in \mathcal{E}_m$ if $(\omega_{\pi(1)}, \ldots, \omega_{\pi(m)}, \omega_{m+1}, \ldots) \in A$ for any permutation π of $\{1, \ldots, m\}$ and all $(\omega_1, \omega_2, \ldots) \in A$. The* exchangeable σ-algebra *$\mathcal{E} = \bigcap_m \mathcal{E}_m$ consists of all events that are invariant under all finite permutations of coordinates. Similarly, we call an infinite sequence of R.V.s $\{\xi_k\}_{k\geq 1}$ on the same measurable space* exchangeable *if $(\xi_1, \ldots, \xi_m) \stackrel{\mathcal{D}}{=} (\xi_{\pi(1)}, \ldots, \xi_{\pi(m)})$ for any m and any permutation π of $\{1, \ldots, m\}$; that is, the joint law of an exchangeable sequence is invariant under finite permutations of coordinates.*

Our next lemma summarizes the use of RMG-s in this context.

Lemma 5.5.25. *Suppose $\xi_k(\omega) = \omega_k$ is an exchangeable sequence of random variables on $(\mathbb{R}^{\mathbf{N}}, \mathcal{B}_c)$. For any bounded Borel function $\varphi : \mathbb{R}^\ell \mapsto \mathbb{R}$ and $m \geq \ell$ let $\widehat{S}_m(\varphi) = \frac{1}{(m)_\ell}\sum_{\underline{i}} \varphi(\xi_{i_1}, \ldots, \xi_{i_\ell})$, where $\underline{i} = (i_1, \ldots, i_\ell)$ is an ℓ-tuple of distinct integers from $\{1, \ldots, m\}$ and $(m)_\ell = \frac{m!}{(m-\ell)!}$ is the number of such ℓ-tuples. Then,*

$$\widehat{S}_m(\varphi) \to \mathbf{E}[\varphi(\xi_1, \ldots, \xi_\ell)|\mathcal{E}] \tag{5.5.5}$$

a.s. and in L^1 when $m \to \infty$.

PROOF. Fixing $m \geq \ell$ since the value of $\widehat{S}_m(\varphi)$ is invariant under any permutation of the first m coordinates of ω we have that $\widehat{S}_m(\varphi)$ is measurable on $\mathcal{E}_m$. Further, this bounded R.V. is obviously integrable, so

$$\widehat{S}_m(\varphi) = \mathbf{E}[\widehat{S}_m(\varphi)|\mathcal{E}_m] = \frac{1}{(m)_\ell}\sum_{\underline{i}} \mathbf{E}[\varphi(\xi_{i_1},\dots,\xi_{i_\ell})|\mathcal{E}_m]\,. \tag{5.5.6}$$

Fixing any ℓ-tuple of distinct integers $i_1,\dots,i_\ell$ from $\{1,\dots,m\}$, by our exchangeability assumption, the probability measure on $(\mathbb{R}^{\mathbf{N}},\mathcal{B}_c)$ is invariant under any permutation π of the first m coordinates of ω such that $\pi(i_k)=k$ for $k=1,\dots,\ell$. Consequently, $\mathbf{E}[\varphi(\xi_{i_1},\dots,\xi_{i_\ell})I_A] = \mathbf{E}[\varphi(\xi_1,\dots,\xi_\ell)I_A]$ for any $A\in\mathcal{E}_m$, implying that $\mathbf{E}[\varphi(\xi_{i_1},\dots,\xi_{i_\ell})|\mathcal{E}_m] = \mathbf{E}[\varphi(\xi_1,\dots,\xi_\ell)|\mathcal{E}_m]$. Since this applies for any ℓ-tuple of distinct integers from $\{1,\dots,m\}$ it follows by (5.5.6) that $\widehat{S}_m(\varphi) = \mathbf{E}[\varphi(\xi_1,\dots,\xi_\ell)|\mathcal{E}_m]$ for all $m\geq\ell$. In conclusion, considering the filtration $\mathcal{F}_n=\mathcal{E}_{\ell-n}$, $n\leq 0$ for which $\mathcal{F}_{-\infty}=\mathcal{E}$, we have in view of the remark following Lévy's downward theorem that $(\widehat{S}_{\ell-n}(\varphi),\mathcal{E}_{\ell-n})$, $n\leq 0$ is a RMG and the convergence in (5.5.5) holds a.s. and in L^1. □

Remark. Noting that any sequence of i.i.d. random variables is also exchangeable, our first application of Lemma 5.5.25 is the following zero-one law.

Theorem 5.5.26 (HEWITT-SAVAGE 0-1 LAW). *The exchangeable σ-algebra $\mathcal{E}$ of a sequence of i.i.d. random variables $\xi_k(\omega)=\omega_k$ is $\mathbf{P}$-trivial (that is, $\mathbf{P}(A)\in\{0,1\}$ for any $A\in\mathcal{E}$).*

Remark. Given the Hewitt-Savage 0-1 law, we can simplify the proof of Theorem 5.5.22 upon noting that for each m the σ-algebra $\mathcal{F}_{-m}$ is contained in $\mathcal{E}_{m+1}$, hence $\mathcal{F}_{-\infty}\subseteq\mathcal{E}$ must also be $\mathbf{P}$-trivial.

PROOF. As the i.i.d. $\xi_k(\omega)=\omega_k$ are exchangeable, from Lemma 5.5.25 we have that for any bounded Borel $\varphi:\mathbb{R}^\ell\to\mathbb{R}$, almost surely $\widehat{S}_m(\varphi)\to\widehat{S}_\infty(\varphi) = \mathbf{E}[\varphi(\xi_1,\dots,\xi_\ell)|\mathcal{E}]$.

We proceed to show that $\widehat{S}_\infty(\varphi)=\mathbf{E}[\varphi(\xi_1,\dots,\xi_\ell)]$. To this end, fixing a finite integer $r\leq m$ let

$$\widehat{S}_{m,r}(\varphi) = \frac{1}{(m)_\ell}\sum_{\{\underline{i}:\, i_1>r,\dots,i_\ell>r\}} \varphi(\xi_{i_1},\dots,\xi_{i_\ell})$$

denote the contribution of the ℓ-tuples $\underline{i}$ that do not intersect $\{1,\dots,r\}$. Since there are exactly $(m-r)_\ell$ such ℓ-tuples and φ is bounded, it follows that

$$|\widehat{S}_m(\varphi)-\widehat{S}_{m,r}(\varphi)| \leq [1-\frac{(m-r)_\ell}{(m)_\ell}]\|\varphi(\cdot)\|_\infty \leq \frac{c}{m}$$

for some $c=c(r,\ell,\varphi)$ finite and all m. Consequently, for any r,

$$\widehat{S}_\infty(\varphi) = \lim_{m\to\infty}\widehat{S}_m(\varphi) = \lim_{m\to\infty}\widehat{S}_{m,r}(\varphi)\,.$$

Further, by the mutual independence of $\{\xi_k\}$ we have that $\widehat{S}_{m,r}(\varphi)$ are independent of $\mathcal{F}_r^{\boldsymbol{\xi}}$, hence the same applies for their limit $\widehat{S}_\infty(\varphi)$. Applying Lemma 4.2.9 for $X=\varphi(\xi_1,\dots,\xi_\ell)$ we deduce that $\mathbf{E}[\varphi(\xi_1,\dots,\xi_\ell)|\mathcal{E}] = \mathbf{E}[\varphi(\xi_1,\dots,\xi_\ell)]$. Recall that I_G is, for each $G\in\mathcal{F}_\ell^{\boldsymbol{\xi}}$, a bounded Borel function of $(\xi_1,\dots,\xi_\ell)$. Hence, $\mathbf{E}[I_G|\mathcal{E}]=\mathbf{E}[I_G]$. Thus, by the tower property and taking out what is known, $\mathbf{P}(A\cap G)=\mathbf{E}[I_A\mathbf{E}[I_G|\mathcal{E}]]=\mathbf{E}[I_A]\mathbf{E}[I_G]$ for any $A\in\mathcal{E}$. That is, $\mathcal{E}$ and $\mathcal{F}_\ell^{\boldsymbol{\xi}}$ are independent for each finite ℓ, so by Lemma 1.4.8 we conclude that $\mathcal{E}$ is a $\mathbf{P}$-trivial σ-algebra, as claimed. □

The proof of de Finetti's theorem requires the following algebraic identity which we leave as an exercise for the reader.

Exercise 5.5.27. *Fixing bounded Borel functions $f : \mathbb{R}^{\ell-1} \mapsto \mathbb{R}$ and $g : \mathbb{R} \mapsto \mathbb{R}$, let $h_j(x_1, \ldots, x_\ell) = f(x_1, \ldots, x_{\ell-1})g(x_j)$ for $j = 1, \ldots, \ell$. Show that for any sequence $\{\xi_k\}$ and any $m \geq \ell$,*

$$\widehat{S}_m(h_\ell) = \frac{m}{m-\ell+1}\widehat{S}_m(f)\widehat{S}_m(g) - \frac{1}{m-\ell+1}\sum_{j=1}^{\ell-1}\widehat{S}_m(h_j)\,.$$

Theorem 5.5.28 (de Finetti's theorem)**.** *If $\xi_k(\omega) = \omega_k$ is an exchangeable sequence then conditional on $\mathcal{E}$ the random variables ξ_k, $k \geq 1$ are mutually independent and identically distributed.*

Remark. For example, if exchangeable $\{\xi_k\}$ are $\{0, 1\}$-valued, then by de Finetti's theorem these are i.i.d. Bernoulli variables of parameter p, conditional on $\mathcal{E}$. The joint (unconditional) law of $\{\xi_k\}$ is thus that of a mixture of i.i.d. Bernoulli(p) sequences with p a $[0, 1]$-valued random variable, measurable on $\mathcal{E}$.

Proof. In view of Exercise 5.5.27, upon applying (5.5.5) of Lemma 5.5.25 for the exchangeable sequence $\{\xi_k\}$ and bounded Borel functions f, g and h_ℓ, we deduce that

$$\mathbf{E}[f(\xi_1, \ldots, \xi_{\ell-1})g(\xi_\ell)|\mathcal{E}] = \mathbf{E}[f(\xi_1, \ldots, \xi_{\ell-1})|\mathcal{E}]\mathbf{E}[g(\xi_\ell)|\mathcal{E}]\,.$$

By induction on ℓ this leads to the identity

$$\mathbf{E}\big[\prod_{k=1}^{\ell} g_k(\xi_k)|\mathcal{E}\big] = \prod_{k=1}^{\ell} E[g_k(\xi_k)|\mathcal{E}]$$

for all ℓ and bounded Borel $g_k : \mathbb{R} \mapsto \mathbb{R}$. Taking $g_k = I_{B_k}$ for $B_k \in \mathcal{B}$ we have

$$\mathbf{P}[(\xi_1, \ldots, \xi_\ell) \in B_1 \times \cdots \times B_\ell|\mathcal{E}] = \prod_{k=1}^{\ell}\mathbf{P}(\xi_k \in B_k|\mathcal{E})$$

which implies that conditional on $\mathcal{E}$ the R.V.-s $\{\xi_k\}$ are mutually independent (see Proposition 1.4.21). Further, $\mathbf{E}[g(\xi_1)I_A] = \mathbf{E}[g(\xi_r)I_A]$ for any $A \in \mathcal{E}$, bounded Borel $g(\cdot)$, positive integer r and exchangeable variables $\xi_k(\omega) = \omega_k$, from which it follows that conditional on $\mathcal{E}$ these R.V.-s are also identically distributed. □

We conclude this section with exercises detailing further applications of RMG-s for the study of a certain U-statistics, for solving the *ballot's problem* and in the context of mixing conditions.

Exercise 5.5.29. *Suppose $\{\xi_k\}$ are i.i.d. random variables and $h : \mathbb{R}^2 \mapsto \mathbb{R}$ a Borel function such that $\mathbf{E}[|h(\xi_1, \xi_2)|] < \infty$. For each $m \geq 2$ let*

$$W_{2-m} = \frac{1}{m(m-1)}\sum_{1\leq i\neq j\leq m} h(\xi_i, \xi_j)\,.$$

For example, note that $W_{2-m} = \frac{1}{m-1}\sum_{k=1}^{m}(\xi_k - m^{-1}\sum_{i=1}^{m}\xi_i)^2$ is of this form, corresponding to $h(x, y) = (x - y)^2/2$.

(a) *Show that* $W_n = \mathbf{E}[h(\xi_1, \xi_2)|\mathcal{F}_n^{\mathbf{W}}]$ *for* $n \le 0$ *hence* $(W_n, \mathcal{F}_n^{\mathbf{W}})$ *is a RMG and determine its almost sure limit as* $n \to -\infty$.

(b) *Assuming in addition that* $v = \mathbf{E}[h(\xi_1, \xi_2)^2]$ *is finite, find the limit of* $\mathbf{E}[W_n^2]$ *as* $n \to -\infty$.

Exercise 5.5.30 (The ballot problem)**.** *Let* $S_k = \sum_{i=1}^k \xi_i$ *for i.i.d. integer valued* $\xi_j \ge 0$ *and for* $n \ge 2$ *consider the event* $\Gamma_n = \{S_j < j \text{ for } 1 \le j \le n\}$.

(a) *Show that* $X_{-k} = k^{-1} S_k$ *is a RMG for the filtration* $\mathcal{F}_{-k} = \sigma(S_j, j \ge k)$ *and that* $\tau = \inf\{\ell \ge -n : X_\ell \ge 1\} \wedge -1$ *is a stopping time for it.*

(b) *Show that* $I_{\Gamma_n} = 1 - X_\tau$ *whenever* $S_n \le n$, *hence* $\mathbf{P}(\Gamma_n|S_n) = (1 - S_n/n)_+$.

The name ballot problem is attached to Exercise 5.5.30 since for $\xi_j \in \{0, 2\}$ we interpret 0's and 2's as n votes for two candidates A and B in a ballot, with $\Gamma_n = \{$A leads B throughout the counting$\}$ and $\mathbf{P}(\Gamma_n | B \text{ gets } r \text{ votes}) = (1 - 2r/n)_+$.

As you find next, the ballot problem yields explicit formulas for the probability distributions of the stopping times $\tau_b = \inf\{n \ge 0 : \ S_n = b\}$ associated with the SRW $\{S_n\}$.

Exercise 5.5.31. *Let* $R = \inf\{\ell \ge 1 : S_\ell = 0\}$ *denote the first visit to zero by the* SRW $\{S_n\}$. *Using a path reversal counting argument followed by the ballot problem, show that for any positive integers* n, b,

$$\mathbf{P}(\tau_b = n | S_n = b) = \mathbf{P}(R > n | S_n = b) = \frac{b}{n}$$

and deduce that for any $k \ge 0$,

$$\mathbf{P}(\tau_b = b + 2k) = b \frac{(b + 2k - 1)!}{k!(k+b)!} p^{b+k} q^k \,.$$

Exercise 5.5.32. *Show that for any* $A \in \mathcal{F}$ *and* σ*-algebra* $\mathcal{G} \subseteq \mathcal{F}$

$$\sup_{B \in \mathcal{G}} |\mathbf{P}(A \cap B) - \mathbf{P}(A)\mathbf{P}(B)| \le \mathbf{E}[|\mathbf{P}(A|\mathcal{G}) - \mathbf{P}(A)|] \,.$$

Next, deduce that if $\mathcal{G}_n \downarrow \mathcal{G}$ *as* $n \downarrow -\infty$ *and* $\mathcal{G}$ *is* $\mathbf{P}$*-trivial, then*

$$\lim_{m \to \infty} \sup_{B \in \mathcal{G}_{-m}} |\mathbf{P}(A \cap B) - \mathbf{P}(A)\mathbf{P}(B)| = 0 \,.$$

CHAPTER 6

Markov chains

The rich theory of Markov processes is the subject of many text books and one can easily teach a full course on this subject alone. Thus, we limit ourselves here to the discrete time Markov chains and to their most fundamental properties. Specifically, in Section 6.1 we provide definitions and examples, and prove the strong Markov property of such chains. Section 6.2 explores the key concepts of recurrence, transience, invariant and reversible measures, as well as the asymptotic (long time) behavior for time homogeneous Markov chains of countable state space. These concepts and results are then generalized in Section 6.3 to the class of Harris Markov chains.

6.1. Canonical construction and the strong Markov property

We start with the definition of a Markov chain.

Definition 6.1.1. *Given a filtration $\{\mathcal{F}_n\}$, an $\mathcal{F}_n$-adapted stochastic process $\{X_n\}$ taking values in a measurable space $(\mathbb{S}, \mathcal{S})$ is called an $\mathcal{F}_n$-*Markov chain *with* state space $(\mathbb{S}, \mathcal{S})$ *if for any $A \in \mathcal{S}$,*

$$\mathbf{P}[X_{n+1} \in A|\mathcal{F}_n] = \mathbf{P}[X_{n+1} \in A|X_n] \qquad \forall n, \quad a.s. \tag{6.1.1}$$

Remark. We call $\{X_n\}$ a *Markov chain* in case $\mathcal{F}_n = \sigma(X_k, k \leq n)$, noting that if $\{X_n\}$ is an $\mathcal{F}_n$-Markov chain then it is also a Markov chain. Indeed, $\mathcal{F}_n^{\mathbf{X}} = \sigma(X_k, k \leq n) \subseteq \mathcal{F}_n$ since $\{X_n\}$ is adapted to $\{\mathcal{F}_n\}$, so by the tower property we have that for any $\mathcal{F}_n$-Markov chain, any $A \in \mathcal{S}$ and all n, almost surely,

$$\begin{aligned}\mathbf{P}[X_{n+1} \in A|\mathcal{F}_n^{\mathbf{X}}] &= \mathbf{E}[\mathbf{E}[I_{X_{n+1}\in A}|\mathcal{F}_n]|\mathcal{F}_n^{\mathbf{X}}] = \mathbf{E}[\mathbf{E}[I_{X_{n+1}\in A}|X_n]|\mathcal{F}_n^{\mathbf{X}}] \\ &= \mathbf{E}[I_{X_{n+1}\in A}|X_n] = \mathbf{P}[X_{n+1} \in A|X_n]\,.\end{aligned}$$

The key object in characterizing an $\mathcal{F}_n$-Markov chain are its transition probabilities, as defined next.

Definition 6.1.2. *A set function $p : \mathbb{S} \times \mathcal{S} \mapsto [0, 1]$ is a* transition probability *if*

(a) *For each $x \in \mathbb{S}$, $A \mapsto p(x, A)$ is a probability measure on $(\mathbb{S}, \mathcal{S})$.*
(b) *For each $A \in \mathcal{S}$, $x \mapsto p(x, A)$ is a measurable function on $(\mathbb{S}, \mathcal{S})$.*

We say that an $\mathcal{F}_n$-Markov chain $\{X_n\}$ has transition probabilities $p_n(x, A)$, if almost surely $\mathbf{P}[X_{n+1} \in A|\mathcal{F}_n] = p_n(X_n, A)$ for every $n \geq 0$ and every $A \in \mathcal{S}$ and call it a homogeneous $\mathcal{F}_n$-Markov chain *if $p_n(x, A) = p(x, A)$ for all n, $x \in \mathbb{S}$ and $A \in \mathcal{S}$.*

With $b\mathcal{S} \subseteq m\mathcal{S}$ denoting the collection of all *bounded* $(\mathbb{R}, \mathcal{B}_{\mathbb{R}})$-valued measurable mappings on $(\mathbb{S}, \mathcal{S})$, we next express $\mathbf{E}[h(X_{k+1})|\mathcal{F}_k]$ for $h \in b\mathcal{S}$ in terms of the transition probabilities of the $\mathcal{F}_n$-Markov chain $\{X_n\}$.

Lemma 6.1.3. *If $\{X_n\}$ is an $\mathcal{F}_n$-Markov chain with state space $(\mathbb{S}, \mathcal{S})$ and transition probabilities $p_n(\cdot\,,\cdot)$, then for any $h \in b\mathcal{S}$ and all $k \geq 0$*

$$\mathbf{E}[h(X_{k+1})|\mathcal{F}_k] = (p_k h)(X_k)\,, \tag{6.1.2}$$

where $h \mapsto (p_k h) : b\mathcal{S} \mapsto b\mathcal{S}$ and $(p_k h)(x) = \int p_k(x, dy)h(y)$ denotes the Lebesgue integral of $h(\cdot)$ under the probability measure $p_k(x, \cdot)$ per fixed $x \in \mathbb{S}$.

Proof. Let $\mathcal{H} \subseteq b\mathcal{S}$ denote the collection of bounded, measurable $\mathbb{R}$-valued functions $h(\cdot)$ for which $(p_k h)(x) \in b\mathcal{S}$ and (6.1.2) holds for all $k \geq 0$. Since $p_k(\cdot, \cdot)$ are transition probabilities of the chain, $I_A \in \mathcal{H}$ for all $A \in \mathcal{S}$ (c.f. Definition 6.1.2). Thus, we complete the proof of the lemma upon checking that $\mathcal{H}$ satisfies the conditions of the (bounded version of the) monotone class theorem (i.e. Theorem 1.2.7). To this end, for a constant h we have that $p_k h$ is also constant and evidently (6.1.2) then holds. Further, with $b\mathcal{S}$ a vector space over $\mathbb{R}$, due to the linearity of both the conditional expectation on the left side of (6.1.2) and the expectation on its right side, so is $\mathcal{H}$. Next, suppose $h_m \in \mathcal{H}$, $h_m \geq 0$ and $h_m \uparrow h \in b\mathcal{S}$. Then, by monotone convergence $(p_k h_m)(x) \uparrow (p_k h)(x)$ for each $x \in \mathbb{S}$ and all $k \geq 0$. In particular, with $p_k h_m \in b\mathcal{S}$ and $p_k h$ bounded by the bound on h, it follows that $p_k h \in b\mathcal{S}$. Further, by the monotone convergence of conditional expectations and the boundedness of $h(X_{k+1})$ also $\mathbf{E}[h_m(X_{k+1})|\mathcal{F}_k] \uparrow \mathbf{E}[h(X_{k+1})|\mathcal{F}_k]$. It thus follows that $h \in \mathcal{H}$ and with all conditions of the monotone class theorem holding for $\mathcal{H}$ and the π-system $\mathcal{S}$, we have that $b\mathcal{S} \subseteq \mathcal{H}$, as stated. □

Our construction of product measures extends to products of transition probabilities. Indeed, you should check at this point that the proof of Theorem 1.4.19 easily adapts to yield the following proposition.

Proposition 6.1.4. *Given a σ-finite measure ν_1 on $(\mathbb{X}, \mathfrak{X})$ and $\nu_2 : \mathbb{X} \times \mathcal{S} \mapsto [0, 1]$ such that $B \mapsto \nu_2(x, B)$ is a probability measure on $(\mathbb{S}, \mathcal{S})$ for each fixed $x \in \mathbb{X}$ and $x \mapsto \nu_2(x, B)$ is measurable on $(\mathbb{X}, \mathfrak{X})$ for each fixed $B \in \mathcal{S}$, there exists a unique σ-finite measure μ on the product space $(\mathbb{X} \times \mathbb{S}, \mathfrak{X} \times \mathcal{S})$, denoted hereafter by $\mu = \nu_1 \otimes \nu_2$, such that*

$$\mu(A \times B) = \int_A \nu_1(dx)\nu_2(x, B), \quad \forall A \in \mathfrak{X}, B \in \mathcal{S}\,.$$

We turn to show how relevant the preceding proposition is for Markov chains.

Proposition 6.1.5. *To any σ-finite measure ν on $(\mathbb{S}, \mathcal{S})$ and any sequence of transition probabilities $p_n(\cdot, \cdot)$ there correspond unique σ-finite measures $\mu_k = \nu \otimes p_0 \cdots \otimes p_{k-1}$ on $(\mathbb{S}^{k+1}, \mathcal{S}^{k+1})$, $k = 1, 2, \ldots$ such that*

$$\mu_k(A_0 \times \cdots \times A_k) = \int_{A_0} \nu(dx_0) \int_{A_1} p_0(x_0, dx_1) \cdots \int_{A_k} p_{k-1}(x_{k-1}, dx_k)$$

for any $A_i \in \mathcal{S}$, $i = 0, \ldots, k$. If ν is a probability measure, then μ_k is a consistent sequence of probability measures (that is, $\mu_{k+1}(A \times \mathbb{S}) = \mu_k(A)$ for any k finite and $A \in \mathbb{S}^{k+1}$).

Further, if $\{X_n\}$ is a Markov chain with state space $(\mathbb{S}, \mathcal{S})$, transition probabilities $p_n(\cdot\,, \cdot)$ and initial distribution *$\nu(A) = \mathbf{P}(X_0 \in A)$ on $(\mathbb{S}, \mathcal{S})$, then for any $h_\ell \in b\mathcal{S}$*

and all $k \geq 0$,

$$\mathbf{E}[\prod_{\ell=0}^{k} h_\ell(X_\ell)] = \int \nu(dx_0)h_0(x_0)\cdots\int p_{k-1}(x_{k-1},dx_k)h_k(x_k)\,, \tag{6.1.3}$$

so in particular, $\{X_n\}$ *has the* finite dimensional distributions *(f.d.d.)*

$$\mathbf{P}(X_0 \in A_0,\ldots,X_n \in A_n) = \nu \otimes p_0 \cdots \otimes p_{n-1}(A_0 \times \ldots \times A_n)\,. \tag{6.1.4}$$

Proof. Starting at a σ-finite measure $\nu_1 = \nu$ on $(\mathbb{S},\mathcal{S})$ and applying Proposition 6.1.4 for $\nu_2(x,B) = p_0(x,B)$ on $\mathbb{S}\times\mathcal{S}$ yields the σ-finite measure $\mu_1 = \nu\otimes p_0$ on $(\mathbb{S}^2,\mathcal{S}^2)$. Applying this proposition once more, now with $\nu_1 = \mu_1$ and $\nu_2((x_0,x_1),B) = p_1(x_1,B)$ for $x = (x_0,x_1) \in \mathbb{S}\times\mathbb{S}$ yields the σ-finite measure $\mu_2 = \nu\otimes p_0\otimes p_1$ on $(\mathbb{S}^3,\mathcal{S}^3)$ and upon repeating this procedure k times we arrive at the σ-finite measure $\mu_k = \nu\otimes p_0\cdots\otimes p_{k-1}$ on $(\mathbb{S}^{k+1},\mathcal{S}^{k+1})$. Since $p_n(x,\mathbb{S}) = 1$ for all n and $x \in \mathbb{S}$, it follows that if ν is a probability measure, so are μ_k which by construction are also consistent.

Suppose next that the Markov chain $\{X_n\}$ has transition probabilities $p_n(\cdot,\cdot)$ and initial distribution ν. Fixing k and $h_\ell \in b\mathcal{S}$ we have by the tower property and (6.1.2) that

$$\mathbf{E}[\prod_{\ell=0}^{k} h_\ell(X_\ell)] = \mathbf{E}[\prod_{\ell=0}^{k-1} h_\ell(X_\ell)\mathbf{E}(h_k(X_k)|\mathcal{F}^{\mathbf{X}}_{k-1})] = \mathbf{E}[\prod_{\ell=0}^{k-1} h_\ell(X_\ell)(p_{k-1}h_k)(X_{k-1})]\,.$$

Further, with $p_{k-1}h_k \in b\mathcal{S}$ (see Lemma 6.1.3), also $h_{k-1}(p_{k-1}h_k) \in b\mathcal{S}$ and we get (6.1.3) by induction on k starting at $\mathbf{E}h_0(X_0) = \int \nu(dx_0)h_0(x_0)$. The formula (6.1.4) for the f.d.d. is merely the special case of (6.1.3) corresponding to indicator functions $h_\ell = I_{A_\ell}$. □

Remark 6.1.6. Using (6.1.1) we deduce from Exercise 4.4.5 that any $\mathcal{F}_n$-Markov chain with a $\mathcal{B}$-isomorphic state space has transition probabilities. We proceed to define the *law* of such a Markov chain and building on Proposition 6.1.5 show that it is uniquely determined by the initial distribution and transition probabilities of the chain.

Definition 6.1.7. *The law of a Markov chain* $\{X_n\}$ *with a* $\mathcal{B}$*-isomorphic state space* $(\mathbb{S},\mathcal{S})$ *and initial distribution* ν *is the unique probability measure* $\mathbf{P}_\nu$ *on* $(\mathbb{S}_\infty,\mathcal{S}_c)$ *with* $\mathbb{S}_\infty = \mathbb{S}^{\mathbb{Z}_+}$*, per Corollary 1.4.25, with the specified f.d.d.*

$$\mathbf{P}_\nu(\{\mathbf{s} : s_i \in A_i, i = 0,\ldots,n\}) = \mathbf{P}(X_0 \in A_0,\ldots,X_n \in A_n)\,,$$

for $A_i \in \mathcal{S}$*. We denote by* $\mathbf{P}_x$ *the law* $\mathbf{P}_\nu$ *in case* $\nu(A) = I_{x\in A}$ *(i.e. when* $X_0 = x$ *is non-random).*

Remark. Definition 6.1.7 provides the (joint) law for any stochastic process $\{X_n\}$ with a $\mathcal{B}$-isomorphic state space (that is, it applies for any sequence of $(\mathbb{S},\mathcal{S})$-valued R.V. on the same probability space).

Here is our *canonical construction* of Markov chains out of their transition probabilities and initial distributions.

Theorem 6.1.8. *If* $(\mathbb{S},\mathcal{S})$ *is* $\mathcal{B}$*-isomorphic, then to any collection of transition probabilities* $p_n : \mathbb{S}\times\mathcal{S} \mapsto [0,1]$ *and any probability measure* ν *on* $(\mathbb{S},\mathcal{S})$ *there* *corresponds a Markov chain* $Y_n(\mathbf{s}) = s_n$ *on the measurable space* $(\mathbb{S}_\infty, \mathcal{S}_c)$ *with state space* $(\mathbb{S}, \mathcal{S})$*, transition probabilities* $p_n(\cdot, \cdot)$*, initial distribution* ν *and f.d.d.*

$$\mathbf{P}_\nu(\{\mathbf{s} : (s_0, \dots, s_k) \in A\}) = \nu \otimes p_0 \cdots \otimes p_{k-1}(A) \;\; \forall A \in \mathcal{S}^{k+1}, \;\; k < \infty. \tag{6.1.5}$$

Remark. In particular, this construction implies that for any probability measure ν on $(\mathbb{S}, \mathcal{S})$ and all $A \in \mathcal{S}_c$

$$\mathbf{P}_\nu(A) = \int \nu(dx) \mathbf{P}_x(A)\,. \tag{6.1.6}$$

We shall use the latter identity as an alternative definition for $\mathbf{P}_\nu$, that is applicable even for a non-finite initial measure (namely, when $\nu(\mathbb{S}) = \infty$), noting that if ν is σ-finite then $\mathbf{P}_\nu$ is also the unique σ-finite measure on $(\mathbb{S}_\infty, \mathcal{S}_c)$ for which (6.1.5) holds (see the remark following Corollary 1.4.25).

Proof. The given transition probabilities $p_n(\cdot, \cdot)$ and probability measure ν on $(\mathbb{S}, \mathcal{S})$ determine the consistent probability measures $\mu_k = \nu \otimes p_0 \cdots \otimes p_{k-1}$ per Proposition 6.1.5 and thereby via Corollary 1.4.25 yield the stochastic process $Y_n(\mathbf{s}) = s_n$ on $(\mathbb{S}_\infty, \mathcal{S}_c)$, of law $\mathbf{P}_\nu$, state space $(\mathbb{S}, \mathcal{S})$ and f.d.d. μ_k. Taking $k = 0$ in (6.1.5) confirms that its initial distribution is indeed ν. Further, fixing $k \geq 0$ finite, let $\mathbf{Y} = (Y_0, \dots, Y_k)$ and note that for any $A \in \mathcal{S}^{k+1}$ and $B \in \mathcal{S}$

$$\mathbf{E}[I_{\{\mathbf{Y} \in A\}} I_{\{Y_{k+1} \in B\}}] = \mu_{k+1}(A \times B) = \int_A \mu_k(dy) p_k(y_k, B) = \mathbf{E}[I_{\{\mathbf{Y} \in A\}} p_k(Y_k, B)]$$

(where the first and last equalities are due to (6.1.5)). Consequently, for any $B \in \mathcal{S}$ and $k \geq 0$ finite, $p_k(Y_k, B)$ is a version of the C.E. $\mathbf{E}[I_{\{Y_{k+1} \in B\}} | \mathcal{F}_k^{\mathbf{Y}}]$ for $\mathcal{F}_k^{\mathbf{Y}} = \sigma(Y_0, \dots, Y_k)$, thus showing that $\{Y_n\}$ is a Markov chain of transition probabilities $p_n(\cdot, \cdot)$. □

Remark. Conversely, given a Markov chain $\{X_n\}$ of state space $(\mathbb{S}, \mathcal{S})$, applying this construction for its transition probabilities and initial distribution yields a Markov chain $\{Y_n\}$ that has the same law as $\{X_n\}$. To see this, recall (6.1.4) that the f.d.d. of a Markov chain are uniquely determined by its transition probabilities and initial distribution, and further for a $\mathcal{B}$-isomorphic state space, the f.d.d. uniquely determine the law $\mathbf{P}_\nu$ of the corresponding stochastic process. For this reason we consider $(\mathbb{S}_\infty, \mathcal{S}_c, \mathbf{P}_\nu)$ to be the *canonical probability space* for Markov chains, with $X_n(\omega) = \omega_n$ given by the coordinate maps.

The evaluation of the f.d.d. of a Markov chain is considerably more explicit when the state space $\mathbb{S}$ is a countable set (in which case $\mathcal{S} = 2^{\mathbb{S}}$), as then

$$p_n(x, A) = \sum_{y \in A} p_n(x, y)\,,$$

for any $A \subseteq \mathbb{S}$, so the transition probabilities are determined by $p_n(x, y) \geq 0$ such that $\sum_{y \in \mathbb{S}} p_n(x, y) = 1$ for all n and $x \in \mathbb{S}$ (and all Lebesgue integrals are in this case merely sums). In particular, if $\mathbb{S}$ is a finite set and the chain is homogeneous, then identifying $\mathbb{S}$ with $\{1, \dots, m\}$ for some $m < \infty$, we view $p(x, y)$ as the (x, y)-th entry of an $m \times m$ dimensional *transition probability matrix*, and express probabilities of interest in terms of powers of the latter matrix.

For homogeneous Markov chains whose state space is $\mathbb{S} = \mathbb{R}^d$ (or a product of closed intervals thereof), equipped with the corresponding Borel σ-algebra, computations are relatively explicit when for each $x \in \mathbb{S}$ the transition probability $p(x, \cdot)$

is absolutely continuous with respect to (the completion of) Lebesgue measure on $\mathbb{S}$. Its non-negative Radon-Nikodym derivative $p(x,y)$ is then called the *transition probability kernel* of the chain. In this case $(ph)(x) = \int h(y)p(x,y)dy$ and the right side of (6.1.4) amounts to iterated integrations of the kernel $p(x,y)$ with respect to Lebesgue measure on $\mathbb{S}$.

Here are few homogeneous Markov chains of considerable interest in probability theory and its applications.

Example 6.1.9 (Random walk). *The random walk $S_n = S_0 + \sum_{k=1}^{n} \xi_k$, where $\{\xi_k\}$ are i.i.d. $\mathbb{R}^d$-valued random variables that are also independent of S_0 is an example of a homogeneous Markov chain. Indeed, $S_{n+1} = S_n + \xi_{n+1}$ with ξ_{n+1} independent of $\mathcal{F}_n^{\mathbf{S}} = \sigma(S_0, \dots, S_n)$. Hence, $\mathbf{P}[S_{n+1} \in A|\mathcal{F}_n^{\mathbf{S}}] = \mathbf{P}[S_n + \xi_{n+1} \in A|S_n]$. With ξ_{n+1} having the same law as ξ_1, we thus get that $\mathbf{P}[S_n + \xi_{n+1} \in A|S_n] = p(S_n, A)$ for the transition probabilities $p(x,A) = \mathbf{P}(\xi_1 \in \{y - x : y \in A\})$ (c.f. Exercise 4.2.2) and the state space $\mathbb{S} = \mathbb{R}^d$ (with its Borel σ-algebra).*

Example 6.1.10 (Branching process). *Another homogeneous Markov chain is the branching process $\{Z_n\}$ of Definition 5.5.1 having the countable state space $\mathbb{S} = \{0, 1, 2, \dots\}$ (and the σ-algebra $\mathcal{S} = 2^{\mathbb{S}}$). The transition probabilities are in this case $p(x,A) = \mathbf{P}(\sum_{j=1}^{x} N_j \in A)$, for integer $x \geq 1$ and $p(0,A) = \mathbf{1}_{0 \in A}$.*

Example 6.1.11 (Renewal Markov chain). *Suppose $q_k \geq 0$ and $\sum_{k=1}^{\infty} q_k = 1$. Taking $\mathbb{S} = \{0, 1, 2, \dots\}$ (and $\mathcal{S} = 2^{\mathbb{S}}$), a homogeneous Markov chain with transition probabilities $p(0,j) = q_{j+1}$ for $j \geq 0$ and $p(i, i-1) = 1$ for $i \geq 1$ is called a* renewal chain.

As you are now to show, in a renewal (Markov) chain $\{X_n\}$ the value of X_n is the amount of time from n to the first of the (integer valued) *renewal times* $\{T_k\}$ in $[n, \infty)$, where $\tau_m = T_m - T_{m-1}$ are i.i.d. and $\mathbf{P}(\tau_1 = j) = q_j$ (compare with Example 2.3.7).

Exercise 6.1.12. *Suppose $\{\tau_k\}$ are i.i.d. positive integer valued random variables with $\mathbf{P}(\tau_1 = j) = q_j$. Let $T_m = T_0 + \sum_{k=1}^{m} \tau_k$ for non-negative integer random variable T_0 which is independent of $\{\tau_k\}$.*

(a) *Show that $N_\ell = \inf\{k \geq 0 : T_k \geq \ell\}$, $\ell = 0, 1, \dots,$ are finite stopping times for the filtration $\mathcal{G}_n = \sigma(T_0, \tau_k, k \leq n)$.*
(b) *Show that for each fixed non-random ℓ, the random variable $\tau_{N_\ell+1}$ is independent of the stopped σ-algebra $\mathcal{G}_{N_\ell}$ and has the same law as τ_1.*
(c) *Let $X_n = \min\{(T_k - n)_+ : T_k \geq n\}$. Show that $X_{n+1} = X_n + \tau_{N_n+1} I_{X_n=0} - 1$ is a homogeneous Markov chain whose transition probabilities are given in Example 6.1.11.*

Example 6.1.13 (Birth and death chain). *A homogeneous Markov chain $\{X_n\}$ whose state space is $\mathbb{S} = \{0, 1, 2, \dots\}$ and for which $X_{n+1} - X_n \in \{-1, 0, 1\}$ is called a* birth and death chain.

Exercise 6.1.14 (Bayesian estimator). *Let θ and $\{U_k\}$ be independent random variables, each of which is uniformly distributed on $(0,1)$. Let $S_n = \sum_{k=1}^{n} X_k$ for $X_k = \operatorname{sgn}(\theta - U_k)$. That is, first pick θ according to the uniform distribution and then generate a* SRW *S_n with each of its increments being $+1$ with probability θ and -1 otherwise.*

(a) *Compute* $\mathbf{P}(X_{n+1}=1|X_1,\ldots,X_n)$.
(b) *Show that* $\{S_n\}$ *is a Markov chain. Is it a homogeneous chain?*

Exercise 6.1.15 (First order auto-regressive process)**.** *The first order auto-regressive process* $\{X_k\}$ *is defined via* $X_n=\alpha X_{n-1}+\xi_n$ *for* $n\geq 1$*, where* α *is a non-random scalar constant and* $\{\xi_k\}$ *are i.i.d.* $\mathbb{R}^d$*-valued random variables that are independent of* X_0.

(a) *With* $\mathcal{F}_n=\sigma(X_0,\xi_k,k\leq n)$ *verify that* $\{X_n\}$ *is a homogeneous* $\mathcal{F}_n$*-Markov chain of state space* $\mathbb{S}=\mathbb{R}^d$ *(equipped with its Borel* σ*-algebra), and provide its transition probabilities.*
(b) *Suppose* $|\alpha|<1$ *and* $X_0=\beta\xi_0$ *for non-random scalar* β*, with each* ξ_k *having the* multivariate normal distribution $\mathcal{N}(\underline{0},\mathbf{V})$ *of zero mean and covariance matrix* $\mathbf{V}$*. Find the values of* β *for which the law of* X_n *is independent of* n.

As we see in the sequel, our next result, the *strong Markov property*, is extremely useful. It applies to any homogeneous Markov chain with a $\mathcal{B}$-isomorphic state space and allows us to handle expectations of random variables shifted by any stopping time τ with respect to the canonical filtration of the chain.

Proposition 6.1.16 (Strong Markov property)**.** *Consider a canonical probability space* $(\mathbb{S}_\infty,\mathcal{S}_c,\mathbf{P}_\nu)$*, a homogeneous Markov chain* $X_n(\omega)=\omega_n$ *constructed on it via Theorem 6.1.8, its canonical filtration* $\mathcal{F}^{\mathbf{X}}_n=\sigma(X_k,k\leq n)$ *and the shift operator* $\theta:\mathbb{S}_\infty\mapsto\mathbb{S}_\infty$ *such that* $(\theta\omega)_k=\omega_{k+1}$ *for all* $k\geq 0$ *(with the corresponding iterates* $(\theta^n\omega)_k=\omega_{k+n}$ *for* $k,n\geq 0$*). Then, for any* $\{h_n\}\subseteq b\mathcal{S}_c$ *with* $\sup_{n,\omega}|h_n(\omega)|$ *finite, and any* $\mathcal{F}^{\mathbf{X}}_n$*-stopping time* τ

$$\mathbf{E}_\nu[h_\tau(\theta^\tau\omega)\,|\,\mathcal{F}^{\mathbf{X}}_\tau]I_{\{\tau<\infty\}}=\mathbf{E}_{X_\tau}\left[h_\tau\right]I_{\{\tau<\infty\}}\,. \tag{6.1.7}$$

Remark. Here $\mathcal{F}^{\mathbf{X}}_\tau$ is the stopped σ-algebra associated with the stopping time τ (c.f. Definition 5.1.34) and $\mathbf{E}_\nu$ (or $\mathbf{E}_x$) indicates expectation taken with respect to $\mathbf{P}_\nu$ ($\mathbf{P}_x$, respectively). Both sides of (6.1.7) are set to zero when $\tau(\omega)=\infty$ and otherwise its right hand side is $g(n,x)=\mathbf{E}_x[h_n]$ evaluated at $n=\tau(\omega)$ and $x=X_{\tau(\omega)}(\omega)$.

The strong Markov property is a significant extension of the *Markov property*:

$$\mathbf{E}_\nu[\,h(\theta^n\omega)\,|\,\mathcal{F}^{\mathbf{X}}_n]=\mathbf{E}_{X_n}\left[h\right], \tag{6.1.8}$$

holding almost surely for any non-negative integer n and fixed $h\in b\mathcal{S}_c$ (that is, the identity (6.1.7) with $\tau=n$ non-random). This in turn generalizes Lemma 6.1.3 where (6.1.8) is proved in the special case of $h(\omega_1)$ and $h\in b\mathcal{S}$.

Proof. We first prove (6.1.8) for $h(\omega)=\prod_{\ell=0}^{k}g_\ell(\omega_\ell)$ with $g_\ell\in b\mathcal{S}$, $\ell=0,\ldots,k$. To this end, fix $B\in\mathcal{S}^{n+1}$ and recall that $\mu_m=\nu\otimes p_0\cdots\otimes p_{m-1}$ are the f.d.d. for

$\mathbf{P}_\nu$. Consequently, by (6.1.3) and the definition of θ^n,

$$\mathbf{E}_\nu[h(\theta^n\omega)I_B(\omega_0,\dots,\omega_n)] = \mu_{n+k}[I_B(x_0,\dots,x_n)\prod_{\ell=0}^{k} g_\ell(x_{\ell+n})]$$
$$= \mu_n\Big[I_B(x_0,\dots,x_n)g_0(x_n)\int p(x_n,dy_1)g_1(y_1)\cdots\int p(y_{k-1},dy_k)g_k(y_k)\Big]$$
$$= \mathbf{E}_\nu[I_B(X_0,\dots,X_n)\mathbf{E}_{X_n}(h)\,]\,.$$

This holds for all $B \in \mathcal{S}^{n+1}$, which by definition of the conditional expectation amounts to (6.1.8).

The collection $\mathcal{H} \subseteq b\mathcal{S}_c$ of bounded, measurable $h : \mathbb{S}_\infty \to \mathbb{R}$ for which (6.1.8) holds, clearly contains the constant functions and is a vector space over $\mathbb{R}$ (by linearity of the expectation and the conditional expectation). Moreover, by the monotone convergence theorem for conditional expectations, if $h_m \in \mathcal{H}$ are non-negative and $h_m \uparrow h$ which is bounded, then also $h \in \mathcal{H}$. Taking in the preceding $g_\ell = I_{B_\ell}$ we see that $I_A \in \mathcal{H}$ for any A in the π-system $\mathcal{P}$ of cylinder sets (i.e. whenever $A = \{\omega : \omega_0 \in B_0, \dots, \omega_k \in B_k\}$ for some k finite and $B_\ell \in \mathcal{S}$). We thus deduce by the (bounded version of the) monotone class theorem that $\mathcal{H} = b\mathcal{S}_c$, the collection of all bounded functions on $\mathbb{S}_\infty$ that are measurable with respect to the σ-algebra $\mathcal{S}_c$ generated by $\mathcal{P}$.

Having established the Markov property (6.1.8), fixing $\{h_n\} \subseteq b\mathcal{S}_c$ and a $\mathcal{F}^{\mathbf{X}}_n$-stopping time τ, we proceed to prove (6.1.7) by decomposing both sides of the latter identity according to the value of τ. Specifically, the bounded random variables $Y_n = h_n(\theta^n\omega)$ are integrable and applying (6.1.8) for $h = h_n$ we have that $\mathbf{E}_\nu[Y_n|\mathcal{F}^{\mathbf{X}}_n] = g(n, X_n)$. Hence, by part (c) of Exercise 5.1.35, for any finite integer $k \geq 0$,

$$\mathbf{E}_\nu[h_\tau(\theta^\tau\omega)I_{\{\tau=k\}}|\mathcal{F}^{\mathbf{X}}_\tau] = g(k, X_k)I_{\{\tau=k\}} = g(\tau, X_\tau)I_{\{\tau=k\}}$$

The identity (6.1.7) is then established by taking out the $\mathcal{F}^{\mathbf{X}}_\tau$-measurable indicator on $\{\tau = k\}$ and summing over $k = 0, 1, \dots$ (where the finiteness of $\sup_{n,\omega}|h_n(\omega)|$ provides the required integrability). $\square$

Exercise 6.1.17. *Modify the last step of the proof of Proposition 6.1.16 to show that (6.1.7) holds as soon as* $\sum_k \mathbf{E}_{X_k}[\,|h_k|\,]I_{\{\tau=k\}}$ *is* $\mathbf{P}_\nu$*-integrable.*

Here are few applications of the Markov and strong Markov properties.

Exercise 6.1.18. *Consider a homogeneous Markov chain* $\{X_n\}$ *with* $\mathcal{B}$*-isomorphic state space* $(\mathbb{S}, \mathcal{S})$*. Fixing* $\{B_l\} \subseteq \mathcal{S}$*, let* $\Gamma_n = \bigcup_{l>n}\{X_l \in B_l\}$ *and* $\Gamma = \{X_l \in B_l$ *i.o.*$\}$*.*

(a) *Using the Markov property and Lévy's upward theorem (Theorem 5.3.15), show that* $\mathbf{P}(\Gamma_n|X_n) \stackrel{a.s.}{\to} I_\Gamma$*.*

(b) *Show that* $\mathbf{P}(\{X_n \in A_n$ *i.o.*$\} \setminus \Gamma) = 0$ *for any* $\{A_n\} \subseteq \mathcal{S}$ *such that for some* $\eta > 0$ *and all* n*, with probability one,*

$$\mathbf{P}(\Gamma_n|X_n) \geq \eta I_{\{X_n \in A_n\}}\,.$$

(c) *Suppose* $A, B \in \mathcal{S}$ *are such that* $\mathbf{P}_x(X_l \in B$ *for some* $l \geq 1) \geq \eta$ *for some* $\eta > 0$ *and all* $x \in A$*. Deduce that*

$$\mathbf{P}(\{X_n \in A \text{ finitely often}\} \cup \{X_n \in B \text{ i.o.}\}) = 1\,.$$

Exercise 6.1.19 (Reflection principle). *Consider a* symmetric *random walk* $S_n = \sum_{k=1}^n \xi_k$, *that is,* $\{\xi_k\}$ *are i.i.d. real-valued and such that* $\xi_1 \stackrel{\mathcal{D}}{=} -\xi_1$. *With* $\omega_n = S_n$, *use the strong Markov property for the stopping time* $\tau = \inf\{k \le n : \omega_k > b\}$ *and* $h_k(\omega) = I_{\{\omega_{n-k} > b\}}$ *to show that for any* $b > 0$,

$$\mathbf{P}(\max_{k\le n} S_k > b) \le 2\mathbf{P}(S_n > b)\,.$$

Derive also the following, more precise result for the symmetric SRW, *where for any integer* $b > 0$,

$$\mathbf{P}(\max_{k\le n} S_k \ge b) = 2\mathbf{P}(S_n > b) + \mathbf{P}(S_n = b)\,.$$

The concept of *invariant measure* for a homogeneous Markov chain, which we now introduce, plays an important role in our study of such chains throughout Sections 6.2 and 6.3.

Definition 6.1.20. *A measure* ν *on* $(\mathbb{S}, \mathcal{S})$ *such that* $\nu(\mathbb{S}) > 0$ *is called a* positive *or* non-zero *measure. An event* $A \in \mathcal{S}_c$ *is called* shift invariant *if* $A = \theta^{-1}A$ *(i.e.* $A = \{\omega : \theta(\omega) \in A\}$*), and a positive measure* ν *on* $(\mathbb{S}_\infty, \mathcal{S}_c)$ *is called* shift invariant *if* $\nu \circ \theta^{-1}(\cdot) = \nu(\cdot)$ *(i.e.* $\nu(A) = \nu(\{\omega : \theta(\omega) \in A\})$ *for all* $A \in \mathcal{S}_c$*). We say that a stochastic process* $\{X_n\}$ *with a* $\mathcal{B}$*-isomorphic state space* $(\mathbb{S}, \mathcal{S})$ *is (strictly)* stationary *if its joint law* ν *is shift invariant. A positive* σ*-finite measure* μ *on a* $\mathcal{B}$*-isomorphic space* $(\mathbb{S}, \mathcal{S})$ *is called an* invariant measure *for a transition probability* $p(\cdot,\cdot)$ *if it defines via (6.1.6) a shift invariant measure* $\mathbf{P}_\mu(\cdot)$*. In particular, starting at* X_0 *chosen according to an* invariant probability measure μ *results with a* stationary *Markov chain* $\{X_n\}$.

Lemma 6.1.21. *Suppose a* σ*-finite measure* ν *and transition probability* $p_0(\cdot,\cdot)$ *on* $(\mathbb{S}, \mathcal{S})$ *are such that* $\nu \otimes p_0(\mathbb{S} \times A) = \nu(A)$ *for any* $A \in \mathcal{S}$*. Then, for all* $k \ge 1$ *and* $A \in \mathcal{S}^{k+1}$,

$$\nu \otimes p_0 \otimes \cdots \otimes p_k(\mathbb{S} \times A) = \nu \otimes p_1 \otimes \cdots \otimes p_k(A)\,.$$

Proof. Our assumption that $\nu((p_0 f)) = \nu(f)$ for $f = I_A$ and any $A \in \mathcal{S}$ extends by the monotone class theorem to all $f \in b\mathcal{S}$. Fixing $A_i \in \mathcal{S}$ and $k \ge 1$ let $f_k(x) = I_{A_0}(x) p_1 \otimes \cdots \otimes p_k(x, A_1 \times \cdots \times A_k)$ (where $p_1 \otimes \cdots \otimes p_k(x,\cdot)$ are the probability measures of Proposition 6.1.5 in case $\nu = \delta_x$ is the probability measure supported on the singleton $\{x\}$ and $p_0(y, \{y\}) = 1$ for all $y \in \mathbb{S}$). Since $(p_j h) \in b\mathcal{S}$ for any $h \in b\mathcal{S}$ and $j \ge 1$ (see Lemma 6.1.3), it follows that $f_k \in b\mathcal{S}$ as well. Further, $\nu(f_k) = \nu \otimes p_1 \otimes \cdots \otimes p_k(A)$ for $A = A_0 \times A_1 \cdots \times A_k$. By the same reasoning also

$$\nu((p_0 f_k)) = \int_{\mathbb{S}} \nu(dy) \int_{A_0} p_0(y, dx) p_1 \otimes \cdots \otimes p_k(x, A_1 \times \cdots \times A_k) = \nu \otimes p_0 \cdots \otimes p_k(\mathbb{S} \times A)\,.$$

Thus, the stated identity holds for the π-system of product sets $A = A_0 \times \cdots \times A_k$ which generates $\mathcal{S}^{k+1}$ and since $\nu \otimes p_1 \otimes \cdots \otimes p_k(B_n \times \mathbb{S}^k) = \nu(B_n) < \infty$ for some $B_n \uparrow \mathbb{S}$, this identity extends to all of $\mathcal{S}^{k+1}$ (see the remark following Proposition 1.1.39). $\square$

Remark 6.1.22. Let $\mu_k = \nu \otimes^k p$ denote the σ-finite measures of Proposition 6.1.5 in case $p_n(\cdot,\cdot) = p(\cdot,\cdot)$ for all n (with $\mu_0 = \nu \otimes^0 p = \nu$). Specializing Lemma 6.1.21 to this setting we see that if $\mu_1(\mathbb{S} \times A) = \mu_0(A)$ for any $A \in \mathcal{S}$ then $\mu_{k+1}(\mathbb{S} \times A) = \mu_k(A)$ for all $k \ge 0$ and $A \in \mathcal{S}^{k+1}$.

Building on the preceding remark we next characterize the invariant measures for a given transition probability.

Proposition 6.1.23. *A positive σ-finite measure $\mu(\cdot)$ on $\mathcal{B}$-isomorphic $(\mathbb{S}, \mathcal{S})$ is an invariant measure for transition probability $p(\cdot, \cdot)$ if and only if $\mu \otimes p(\mathbb{S} \times A) = \mu(A)$ for all $A \in \mathcal{S}$.*

PROOF. With μ a positive σ-finite measure, so are the measures $\mathbf{P}_\mu$ and $\mathbf{P}_\mu \circ \theta^{-1}$ on $(\mathbb{S}_\infty, \mathcal{S}_c)$ which for a $\mathcal{B}$-isomorphic space $(\mathbb{S}, \mathcal{S})$ are uniquely determined by their finite dimensional distributions (see the remark following Corollary 1.4.25). By (6.1.5) the f.d.d. of $\mathbf{P}_\mu$ are the σ-finite measures $\mu_k(A) = \mu \otimes^k p(A)$ for $A \in \mathcal{S}^{k+1}$ and $k = 0, 1, \dots$ (where $\mu_0 = \mu$). By definition of θ the corresponding f.d.d. of $\mathbf{P}_\mu \circ \theta^{-1}$ are $\mu_{k+1}(\mathbb{S} \times A)$. Therefore, a positive σ-finite measure μ is an invariant measure for $p(\cdot, \cdot)$ if and only if $\mu_{k+1}(\mathbb{S} \times A) = \mu_k(A)$ for any non-negative integer k and $A \in \mathcal{S}^{k+1}$, which by Remark 6.1.22 is equivalent to $\mu \otimes p(\mathbb{S} \times A) = \mu(A)$ for all $A \in \mathcal{S}$. $\square$

6.2. Markov chains with countable state space

Throughout this section we restrict our attention to *homogeneous Markov chains* $\{X_n\}$ on a countable (finite or infinite), state space $\mathbb{S}$, setting as usual $\mathcal{S} = 2^{\mathbb{S}}$ and $p(x, y) = \mathbf{P}_x(X_1 = y)$ for the corresponding transition probabilities. Noting that such chains admit the canonical construction of Theorem 6.1.8 since their state space is $\mathcal{B}$-isomorphic (c.f. Proposition 1.4.27 for $M = \mathbb{S}$ equipped with the metric $d(x, y) = \mathbf{1}_{x \neq y}$), we start with a few useful consequences of the Markov and strong Markov properties that apply for any homogeneous Markov chain on a countable state space.

Proposition 6.2.1 (CHAPMAN-KOLMOGOROV)**.** *For any $x, y \in \mathbb{S}$ and non-negative integers $k \leq n$,*

$$\mathbf{P}_x(X_n = y) = \sum_{z \in \mathbb{S}} \mathbf{P}_x(X_k = z) \mathbf{P}_z(X_{n-k} = y) \tag{6.2.1}$$

PROOF. Using the canonical construction of the chain whereby $X_n(\omega) = \omega_n$, we combine the tower property with the Markov property for $h(\omega) = I_{\{\omega_{n-k} = y\}}$ followed by a decomposition according to the value z of X_k to get that

$$\begin{aligned} \mathbf{P}_x(X_n = y) &= \mathbf{E}_x[h(\theta^k \omega)] = \mathbf{E}_x\Big\{\mathbf{E}_x\big[h(\theta^k \omega) \,|\, \mathcal{F}_k^{\mathbf{X}}\big]\Big\} \\ &= \mathbf{E}_x[\mathbf{E}_{X_k}(h)] = \sum_{z \in \mathbb{S}} \mathbf{P}_x(X_k = z) \mathbf{E}_z(h)\,. \end{aligned}$$

This concludes the proof as $\mathbf{E}_z(h) = \mathbf{P}_z(X_{n-k} = y)$. $\square$

Remark. The Chapman-Kolmogorov equations of (6.2.1) are a concrete special case of the more general Chapman-Kolmogorov *semi-group* representation $p^n = p^k p^{n-k}$ of the n-step transition probabilities $p^n(x, y) = \mathbf{P}_x(X_n = y)$. See [**Dyn65**] for more on this representation, which is at the core of the analytic treatment of general Markov chains and processes (and beyond our scope).

We proceed to derive some results about first hitting times of subsets of the state space by the Markov chain, where by convention we use $\tau_A = \inf\{n \geq 0 : X_n \in A\}$ in case the initial state matters and the strictly positive $T_A = \inf\{n \geq 1 : X_n \in A\}$ when it does not, with $\tau_y = \tau_{\{y\}}$ and $T_y = T_{\{y\}}$. To this end, we start with the *first* *entrance decomposition* of $\{X_n = y\}$ according to the value of T_y (which serves as an alternative to the Chapman-Kolmogorov decomposition of the same event via the value in $\mathbb{S}$ of X_k).

Exercise 6.2.2 (First entrance decomposition)**.**
For a homogeneous Markov chain $\{X_n\}$ on $(\mathbb{S}, \mathcal{S})$, let $T_{y,r} = \inf\{n \geq r : X_n = y\}$ (so $T_y = T_{y,1}$ and $\tau_y = T_{y,0}$).

(a) *Show that for any $x, y \in \mathbb{S}$, $B \in \mathcal{S}$ and positive integers $r \leq n$,*

$$\mathbf{P}_x(X_n \in B, T_{y,r} \leq n) = \sum_{k=0}^{n-r} \mathbf{P}_x(T_{y,r} = n-k)\mathbf{P}_y(X_k \in B)\,.$$

(b) *Deduce that in particular,*

$$\mathbf{P}_x(X_n = y) = \sum_{k=r}^{n} \mathbf{P}_x(T_{y,r} = k)\mathbf{P}_y(X_{n-k} = y).$$

(c) *Conclude that for any $y \in \mathbb{S}$ and non-negative integers r, ℓ,*

$$\sum_{j=0}^{\ell} \mathbf{P}_y(X_j = y) \geq \sum_{n=r}^{\ell+r} \mathbf{P}_y(X_n = y).$$

In contrast, here is an application of the *last entrance decomposition.*

Exercise 6.2.3 (Last entrance decomposition)**.** *Show that for a homogeneous Markov chain $\{X_n\}$ on state space $(\mathbb{S}, \mathcal{S})$, all $x, y \in \mathbb{S}$, $B \in \mathcal{S}$ and $n \geq 1$,*

$$\mathbf{P}_x(X_n \in B, T_y \leq n) = \sum_{k=0}^{n-1} \mathbf{P}_x(X_{n-k} = y)\mathbf{P}_y(X_k \in B, T_y > k)\,.$$

Hint: *With $L_n = \max\{1 \leq \ell \leq n : X_\ell = y\}$ denoting the last visit of y by the chain during $\{1, \ldots, n\}$, observe that $\{T_y \leq n\}$ is the union of the disjoint events $\{L_n = n-k\}$, $k = 0, \ldots, n-1$.*

Next, we express certain hitting probabilities for Markov chains in terms of harmonic functions for these chains.

Definition 6.2.4. *Extending Definition 5.1.25 we say that $f : \mathbb{S} \mapsto \mathbb{R}$ which is either bounded below or bounded above is* super-harmonic *for the transition probability $p(x, y)$ at $x \in \mathbb{S}$ when $f(x) \geq \sum_{y \in \mathbb{S}} p(x, y)f(y)$. Likewise, $f(\cdot)$ is* sub-harmonic *at x when this inequality is reversed and* harmonic *at x in case an equality holds. Such a function is called super-harmonic (or sub-harmonic, harmonic, respectively) for $p(\cdot, \cdot)$ (or for the corresponding chain $\{X_n\}$), if it is super-harmonic (or, sub-harmonic, harmonic, respectively), at all $x \in \mathbb{S}$. Equivalently, $f(\cdot)$ which is either bounded below or bounded above is harmonic provided $\{f(X_n)\}$ is a* martingale *whenever the initial distribution of the chain is such that $f(X_0)$ is integrable. Similarly, $f(\cdot)$ bounded below is super-harmonic if $\{f(X_n)\}$ is a* super-martingale *whenever $f(X_0)$ is integrable.*

Exercise 6.2.5. *Suppose $\mathbb{S} \setminus C$ is finite, $\inf_{x \notin C} \mathbf{P}_x(\tau_C < \infty) > 0$ and $A \subset C$, $B = C \setminus A$ are both non-empty.*

(a) *Show that there exist $N < \infty$ and $\epsilon > 0$ such that $\mathbf{P}_y(\tau_C > kN) \leq (1-\epsilon)^k$ for all $k \geq 1$ and $y \in \mathbb{S}$.*

(b) *Show that $g(x) = \mathbf{P}_x(\tau_A < \tau_B)$ is harmonic at every $x \notin C$.*

(c) *Show that if a bounded function* $g(\cdot)$ *is harmonic at every* $x \notin C$ *then* $g(X_{n\wedge\tau_C})$ *is a martingale.*
(d) *Deduce that* $g(x) = \mathbf{P}_x(\tau_A < \tau_B)$ *is the only bounded function harmonic at every* $x \notin C$ *for which* $g(x) = 1$ *when* $x \in A$ *and* $g(x) = 0$ *when* $x \in B$.
(e) *Show that if* $f : \mathbb{S} \mapsto \mathbb{R}_+$ *satisfies* $f(x) = 1 + \sum_{y\in\mathbb{S}} p(x,y)f(y)$ *at every* $x \notin C$ *then* $f(X_{n\wedge\tau_C}) + n \wedge \tau_C$ *is a martingale. Deduce that if in addition* $f(x) = 0$ *for* $x \in C$ *then* $f(x) = \mathbf{E}_x\tau_C$.

The next exercise demonstrates few of the many interesting explicit formulas one may find for finite state Markov chains.

Exercise 6.2.6. *Throughout,* $\{X_n\}$ *is a Markov chain on* $\mathbb{S} = \{0, 1, \dots, N\}$ *of transition probability* $p(x, y)$.
(a) *Use induction to show that in case* $N = 1$, $p(0, 1) = \alpha$ *and* $p(1, 0) = \beta$ *such that* $\alpha + \beta > 0$,

$$\mathbf{P}_\nu(X_n = 0) = \frac{\beta}{\alpha+\beta} + (1-\alpha-\beta)^n\big\{\nu(0) - \frac{\beta}{\alpha+\beta}\big\}.$$

(b) *Fixing* $\nu(0)$ *and* $\theta_1 \neq \theta_0$ *non-random, suppose* $\alpha = \beta$ *and conditional on* $\{X_n\}$ *the variables* B_k *are independent Bernoulli(*θ_{X_k}*). Evaluate the mean and variance of the additive functional* $S_n = \sum_{k=1}^n B_k$.
(c) *Verify that* $\mathbf{E}_x[(X_n - N/2)] = (1-2/N)^n(x - N/2)$ *for the* Ehrenfest *chain whose transition probabilities are* $p(x, x-1) = x/N = 1 - p(x, x+1)$.

6.2.1. Classification of states, recurrence and transience. We start with the partition of a countable state space of a homogeneous Markov chains to its intercommunicating (equivalence) classes, as defined next.

Definition 6.2.7. *Let* $\rho_{xy} = \mathbf{P}_x(T_y < \infty)$ *denote the probability that starting at* x *the chain eventually visits the state* y. *State* y *is said to be* accessible *from state* $x \neq y$ *if* $\rho_{xy} > 0$ *(or alternatively, we then say that* x leads *to* y*). Two states* $x \neq y$*, each accessible to the other, are said to* intercommunicate*, denoted by* $x \leftrightarrow y$*. A non-empty collection of states* $C \subseteq \mathbb{S}$ *is called* irreducible *if each two states in* C *intercommunicate, and* closed *if there is no* $y \notin C$ *and* $x \in C$ *such that* y *is accessible from* x.

Remark. Evidently an irreducible set C may be a non-closed set and vice verse. For example, if $p(x, y) > 0$ for any $x, y \in \mathbb{S}$ then $\mathbb{S}\setminus\{z\}$ is irreducible and non-closed (for any $z \in \mathbb{S}$). More generally, adopting hereafter the convention that $x \leftrightarrow x$, any non-empty proper subset of an irreducible set is irreducible and non-closed. Conversely, when there exists $y \in \mathbb{S}$ such that $p(x, y) = 0$ for all $x \in \mathbb{S}\setminus\{y\}$, then $\mathbb{S}$ is closed and reducible. More generally, a closed set that has a closed proper subset is reducible. Note however the following elementary properties.

Exercise 6.2.8.
(a) *Show that if* $\rho_{xy} > 0$ *and* $\rho_{yz} > 0$ *then also* $\rho_{xz} > 0$.
(b) *Deduce that intercommunication is an equivalence relation (that is,* $x \leftrightarrow x$*, if* $x \leftrightarrow y$ *then also* $y \leftrightarrow x$ *and if both* $x \leftrightarrow y$ *and* $y \leftrightarrow z$ *then also* $x \leftrightarrow z$*).*
(c) *Explain why its equivalence classes partition* $\mathbb{S}$ *into maximal irreducible sets such that the directed graph indicating which one leads to each other is both transitive (i.e. if* C_1 *leads to* C_2 *and* C_2 *leads to* C_3 *then also* C_1

leads to C_3), and acyclic (i.e. if C_1 leads to C_2 then C_2 does not lead to C_1).

For our study of the qualitative behavior of such chains we further classify each state as either a *transient* state, visited only finitely many times by the chain or as a *recurrent* state to which the chain returns with certainty (infinitely many times) once it has been reached by the chain. To this end, we make use of the following formal definition and key proposition.

Definition 6.2.9. *A state $y \in \mathbb{S}$ is called* recurrent *(or* persistent*) if $\rho_{yy} = 1$ and* transient *if $\rho_{yy} < 1$.*

Proposition 6.2.10. *With $T_y^0 = 0$, let $T_y^k = \inf\{n > T_y^{k-1} : X_n = y\}$ for $k \geq 1$ denote the time of the k-th return to state $y \in \mathbb{S}$ (so $T_y^1 = T_y > 0$ regardless of X_0). Then, for any $x, y \in \mathbb{S}$ and $k \geq 1$,*

$$\mathbf{P}_x(T_y^k < \infty) = \rho_{xy}\rho_{yy}^{k-1}\,. \tag{6.2.2}$$

Further, let $N_\infty(y)$ denote the number of visits to state y by the Markov chain at positive times. Then, $\mathbf{E}_x N_\infty(y) = \frac{\rho_{xy}}{1-\rho_{yy}}$ is positive if and only if $\rho_{xy} > 0$, in which case it is finite when y is transient and infinite when y is recurrent.

Proof. The identity (6.2.2) is merely the observation that starting at x, in order to have k visits to y, one has to first reach y and then to have $k-1$ consecutive returns to y. More formally, the event $\{T_y < \infty\} = \bigcup_n \{T_y \leq n\}$ is in $\mathcal{S}_c$ so fixing $k \geq 2$ the strong Markov property applies for the stopping time $\tau = T_y^{k-1}$ and the indicator function $h = I_{\{T_y < \infty\}}$. Further, $\tau < \infty$ implies that $h(\theta^\tau \omega) = I_{\{T_y^k < \infty\}}(\omega)$ and $X_\tau = y$ so $\mathbf{E}_{X_\tau} h = \mathbf{P}_y(T_y < \infty) = \rho_{yy}$. Combining the tower property with the strong Markov property we thus find that

$$\begin{aligned}\mathbf{P}_x(T_y^k < \infty) &= \mathbf{E}_x[h(\theta^\tau\omega) I_{\tau<\infty}] = \mathbf{E}_x[\mathbf{E}_x[h(\theta^\tau\omega) \,|\, \mathcal{F}_\tau^{\mathbf{X}}] I_{\tau<\infty}] \\ &= \mathbf{E}_x[\rho_{yy} I_{\tau<\infty}] = \rho_{yy}\mathbf{P}_x(T_y^{k-1} < \infty)\,,\end{aligned}$$

and (6.2.2) follows by induction on k, starting with the trivial case $k = 1$.

Next note that if the chain makes at least k visits to state y, then the k-th return to y occurs at finite time, and vice verse. That is, $\{T_y^k < \infty\} = \{N_\infty(y) \geq k\}$, and from the identity (6.2.2), we get that

$$\begin{aligned}\mathbf{E}_x N_\infty(y) &= \sum_{k=1}^\infty \mathbf{P}_x(N_\infty(y) \geq k) = \sum_{k=1}^\infty \mathbf{P}_x(T_y^k < \infty) \\ &= \sum_{k=1}^\infty \rho_{xy}\rho_{yy}^{k-1} = \begin{cases} \frac{\rho_{xy}}{1-\rho_{yy}}, & \rho_{xy} > 0 \\ 0, & \rho_{xy} = 0 \end{cases}\end{aligned} \tag{6.2.3}$$

as claimed. □

In the same spirit as the preceding proof you next show that successive returns to the same state by a Markov chain are *renewal times.*

Exercise 6.2.11. *Fix a recurrent state $y \in \mathbb{S}$ of a Markov chain $\{X_n\}$. Let $R_k = T_y^k$ and $r_k = R_k - R_{k-1}$ the number of steps between consecutive returns to y.*

(a) *Deduce from the strong Markov property that under $\mathbf{P}_y$ the random vectors $\underline{Y}_k = (r_k, X_{R_{k-1}}, \ldots, X_{R_k - 1})$ for $k = 1, 2, \ldots$ are independent and identically distributed.*

(b) *Show that for any probability measure* ν*, under* $\mathbf{P}_\nu$ *and conditional on the event* $\{T_y < \infty\}$*, the random vectors* $\underline{Y}_k$ *are independent of each other and further* $\underline{Y}_k \stackrel{\mathcal{D}}{=} \underline{Y}_2$ *for all* $k \geq 2$*, with* $\underline{Y}_2$ *having then the law of* $\underline{Y}_1$ *under* $\mathbf{P}_y$*.*

Here is a direct consequence of Proposition 6.2.10.

Corollary 6.2.12. *Each of the following characterizes a recurrent state* y*:*

(a) $\rho_{yy} = 1$;
(b) $\mathbf{P}_y(T_y^k < \infty) = 1$ *for all* k;
(c) $\mathbf{P}_y(X_n = y, \ i.o.) = 1$;
(d) $\mathbf{P}_y(N_\infty(y) = \infty) = 1$;
(e) $\mathbf{E}_y N_\infty(y) = \infty$.

Proof. Considering (6.2.2) for $x = y$ we have that (a) implies (b). Given (b) we have w.p.1. that $X_{n_k} = y$ for infinitely many $n_k = T_y^k$, $k = 1, 2, \ldots$, which is (c). Clearly, the events in (c) and (d) are identical, and evidently (d) implies (e). To complete the proof simply note that if $\rho_{yy} < 1$ then by (6.2.3) $\mathbf{E}_y N_\infty(y) = \rho_{yy}/(1 - \rho_{yy})$ is finite. □

We are ready for the main result of this section, a decomposition of the recurrent states to disjoint irreducible closed sets.

Theorem 6.2.13 (Decomposition theorem)**.** *A countable state space* $\mathbb{S}$ *of a homogeneous Markov chain can be partitioned uniquely as*

$$\mathbb{S} = \mathbb{T} \cup \mathbf{R}_1 \cup \mathbf{R}_2 \cup \ldots$$

where $\mathbb{T}$ *is the set of transient states and the* $\mathbf{R}_i$ *are disjoint, irreducible closed sets of recurrent states with* $\rho_{xy} = 1$ *whenever* $x, y \in \mathbf{R}_i$*.*

Remark. An alternative statement of the decomposition theorem is that for any pair of recurrent states $\rho_{xy} = \rho_{yx} \in \{0, 1\}$ while $\rho_{xy} = 0$ if x is recurrent and y is transient (so $x \mapsto \{y \in \mathbb{S} : \rho_{xy} > 0\}$ induces a unique partition of the recurrent states to irreducible closed sets).

Proof. Suppose $x \leftrightarrow y$. Then, $\rho_{xy} > 0$ implies that $\mathbf{P}_x(X_K = y) > 0$ for some finite K and $\rho_{yx} > 0$ implies that $\mathbf{P}_y(X_L = x) > 0$ for some finite L. By the Chapman-Kolmogorov equations we have for any integer $n \geq 0$,

$$\begin{aligned} \mathbf{P}_x(X_{K+n+L} = x) &= \sum_{z,v \in \mathbb{S}} \mathbf{P}_x(X_K = z)\mathbf{P}_z(X_n = v)\mathbf{P}_v(X_L = x) \\ &\geq \mathbf{P}_x(X_K = y)\mathbf{P}_y(X_n = y)\mathbf{P}_y(X_L = x) \,. \end{aligned} \tag{6.2.4}$$

As $\mathbf{E}_y N_\infty(y) = \sum_{n=1}^\infty \mathbf{P}_y(X_n = y)$, summing the preceding inequality over $n \geq 1$ we find that $\mathbf{E}_x N_\infty(x) \geq c\mathbf{E}_y N_\infty(y)$ with $c = \mathbf{P}_x(X_K = y)\mathbf{P}_y(X_L = x)$ positive. If x is a transient state then $\mathbf{E}_x N_\infty(x)$ is finite (see Corollary 6.2.12), hence the same applies for y. Reversing the roles of x and y we conclude that any two intercommunicating states x and y are either both transient or both recurrent. More generally, an irreducible set of states C is either *transient* (i.e. every $x \in C$ is transient) or *recurrent* (i.e. every $x \in C$ is recurrent).

We thus consider the unique partition of $\mathbb{S}$ to (disjoint) maximal irreducible equivalence classes of $\leftrightarrow$ (see Exercise 6.2.8), with $\mathbf{R}_i$ denoting those equivalence classes that are recurrent and proceed to show that if x is recurrent and $\rho_{xy} > 0$ for $y \neq x$, then $\rho_{yx} = 1$. The latter implies that any y accessible from $x \in \mathbf{R}_\ell$ must intercommunicate with x, so with $\mathbf{R}_\ell$ a *maximal* irreducible set, necessarily such y is also in $\mathbf{R}_\ell$. We thus conclude that each $\mathbf{R}_\ell$ is closed, with $\rho_{xy} = 1$ whenever $x, y \in \mathbf{R}_\ell$, as claimed.

To complete the proof fix a state $y \neq x$ that is accessible by the chain from a recurrent state x, noting that then $L = \inf\{n \geq 1 : \mathbf{P}_x(X_n = y) > 0\}$ is finite. Further, because L is the minimal such value there exist $y_0 = x$, $y_L = y$ and $y_i \neq x$ for $1 \leq i \leq L$ such that $\prod_{k=1}^{L} p(y_{k-1}, y_k) > 0$. Consequently, if $\mathbf{P}_y(T_x = \infty) = 1 - \rho_{yx} > 0$, then

$$\mathbf{P}_x(T_x = \infty) \geq \prod_{k=1}^{L} p(y_{k-1}, y_k)(1 - \rho_{yx}) > 0\,,$$

in contradiction of the assumption that x is recurrent. □

The decomposition theorem motivates the following definition, as an irreducible chain is either a recurrent chain or a transient chain.

Definition 6.2.14. *A homogeneous Markov chain is called an* irreducible Markov chain *(or in short, irreducible), if $\mathbb{S}$ is irreducible, a* recurrent Markov chain *(or in short, recurrent), if every $x \in \mathbb{S}$ is recurrent and a* transient Markov chain *(or in short, transient), if every $x \in \mathbb{S}$ is transient.*

By definition once the chain enters a closed set, it remains forever in this set. Hence, if $X_0 \in \mathbf{R}_\ell$ we may as well take $\mathbf{R}_\ell$ to be the whole state space. The case of $X_0 \in \mathbb{T}$ is more involved, for then the chain either remains forever in the set of transient states, or it lies eventually in the first irreducible set of recurrent states it entered. As we next show, the first of these possibilities does not occur when $\mathbb{T}$ (or $\mathbb{S}$) is finite (and any irreducible chain of finite state space is recurrent).

Proposition 6.2.15. *If F is a finite set of transient states then for any initial distribution $\mathbf{P}_\nu(X_n \in F \text{ i.o.}) = 0$. Hence, any finite closed set C contains at least one recurrent state, and if C is also irreducible then C is* recurrent.

Proof. Let $N_\infty(F) = \sum_{y \in F} N_\infty(y)$ denote the totality of positive time the chain spends at a set F. If F is a finite set of transient states then by Proposition 6.2.10 and linearity of the expectation $\mathbf{E}_x N_\infty(F)$ is finite, hence $\mathbf{P}_x(N_\infty(F) = \infty) = 0$. With $\mathbb{S}$ countable and x arbitrary, it follows that $\mathbf{P}_\nu(N_\infty(F) = \infty) = 0$ for any initial distribution ν. This is precisely our first claim (as $N_\infty(F)$ is infinite if and only if $X_n \in F$ for infinitely many values of n). If C is a closed set then starting at $x \in C$ the chain stays in C forever. Thus, $\mathbf{P}_x(N_\infty(C) = \infty) = 1$ and to not contradict our first claim, if such C is finite, then it must contain at least one recurrent state, which is our second claim. Finally, while proving the decomposition theorem we showed that if an irreducible set contains a recurrent state then all its states are recurrent, thus yielding our third and last claim. □

We proceed to study the recurrence versus transience of states for some homogeneous Markov chains we have encountered in Section 6.1. To this end, starting with the branching process we make use of the following definition.

Definition 6.2.16. *If a singleton $\{x\}$ is a closed set of a homogeneous Markov chain, then we call x an* absorbing *state for the chain. Indeed, once the chain visits an absorbing state it remains there (so an absorbing state is recurrent).*

Example 6.2.17 (Branching Processes)**.** *By our definition of the branching process $\{Z_n\}$ we have that 0 is an absorbing state (as $p(0,0)=1$, hence $\rho_{0k}=0$ for all $k\ge 1$). If $\mathbf{P}(N=0)>0$ then clearly $\rho_{k0}\ge p(k,0)=\mathbf{P}(N=0)^k>0$ and $\rho_{kk}\le 1-\rho_{k0}<1$ for all $k\ge 1$, so all states other than 0 are transient.*

Exercise 6.2.18. *Suppose a homogeneous Markov chain $\{X_n\}$ with state space $\mathbb{S}=\{0,1,\dots,N\}$ is a martingale for any initial distribution.*

(a) *Show that 0 and N are absorbing states, that is, $p(0,0)=p(N,N)=1$.*
(b) *Show that if also $\mathbf{P}_x(\tau_{\{0,N\}}<\infty)>0$ for all x then all other states are transient and $\rho_{xN}=\mathbf{P}_x(\tau_N<\tau_0)=x/N$.*
(c) *Check that this applies for the symmetric* SRW *on $\mathbb{S}$ (with absorption at 0 and N), in which case also $\mathbf{E}_x\tau_{\{0,N\}}=x(N-x)$.*

Example 6.2.19 (Renewal Markov chain)**.** *The renewal Markov chain of Example 6.1.11 has $p(i,i-1)=1$ for $i\ge 1$ so evidently $\rho_{i0}=1$ for all $i\ge 1$ and hence also $\rho_{00}=1$, namely 0 is a recurrent state. Recall that $p(0,j)=q_{j+1}$, so if $\{k:q_k>0\}$ is unbounded, then $\rho_{0j}>0$ for all j so the only closed set containing 0 is $\mathbb{S}=\mathbb{Z}_+$. Consequently, in this case the renewal chain is recurrent. If on the other hand $K=\sup\{k:q_k>0\}<\infty$ then $\mathbf{R}=\{0,1,\dots,K-1\}$ is an irreducible closed set of recurrent states and all other states are transient. Indeed, starting at any positive integer j this chain enters its recurrent class of states after at most j steps and stays there forever.*

Your next exercise pursues another approach to the classification of states, expressing the return probabilities ρ_{xx} in terms of limiting values of certain generating functions. Applying this approach to the asymmetric SRW on the integers provides us with an example of a transient (irreducible) chain.

Exercise 6.2.20. *Given a homogeneous Markov chain of countable state space $\mathbb{S}$ and $x\in\mathbb{S}$, consider for $-1<s<1$ the generating functions $f(s)=\mathbf{E}_x[s^{T_x}]$ and*

$$u(s)=\sum_{k\ge 0}\mathbf{E}_x[s^{T_x^k}]=\sum_{n\ge 0}\mathbf{P}_x(X_n=x)s^n\,.$$

(a) *Show that $u(s)=u(s)f(s)+1$.*
(b) *Show that $u(s)\uparrow 1+\mathbf{E}_x[N_\infty(x)]$ as $s\uparrow 1$, while $f(s)\uparrow\rho_{xx}$ and deduce that $\mathbf{E}_x[N_\infty(x)]=\rho_{xx}/(1-\rho_{xx})$.*
(c) *Consider the* SRW *on $\mathbb{Z}$ with $p(i,i+1)=p$ and $p(i,i-1)=q=1-p$. Show that in this case $u(s)=(1-4pqs^2)^{-1/2}$ is independent of the initial state x.*
Hint: *Recall that $(1-t)^{-1/2}=\sum_{m=0}^\infty\binom{2m}{m}2^{-2m}t^m$ for any $0\le t<1$.*
(d) *Deduce that the* SRW *on $\mathbb{Z}$ has $\rho_{xx}=2\min(p,q)$ for all x so for $0<p<1$, $p\ne 1/2$ this irreducible chain is transient, whereas for $p=1/2$ it is recurrent.*

Our next proposition explores a powerful method for proving recurrence of an irreducible chain by the construction of super-harmonic functions (per Definition 6.2.4).

Proposition 6.2.21. *Suppose $\mathbb{S}$ is irreducible for a chain $\{X_n\}$ and there exists $h:\mathbb{S}\mapsto[0,\infty)$ of finite level sets $G_r=\{x:h(x)<r\}$ that is super-harmonic at $\mathbb{S}\setminus G_r$ for this chain and some finite r. Then, the chain $\{X_n\}$ is recurrent.*

Proof. If $\mathbb{S}$ is finite then the chain is recurrent by Proposition 6.2.15. Assuming hereafter that $\mathbb{S}$ is infinite, fix r_0 large enough so the finite set $F = G_{r_0}$ is non-empty and $h(\cdot)$ is super-harmonic at $x \notin F$. By Proposition 6.2.15 and part (c) of Exercise 6.1.18 (for $B = F = \mathbb{S} \setminus A$), if $\mathbf{P}_x(\tau_F < \infty) = 1$ for all $x \in \mathbb{S}$ then F contains at least one recurrent state, so by irreducibility of $\mathbb{S}$ the chain is recurrent, as claimed. Proceeding to show that $\mathbf{P}_x(\tau_F < \infty) = 1$ for all $x \in \mathbb{S}$, fix $r > r_0$ and $C = C_r = F \cup (\mathbb{S} \setminus G_r)$. Note that $h(\cdot)$ super-harmonic at $x \notin C$, hence $h(X_{n \wedge \tau_C})$ is a non-negative sup-MG under $\mathbf{P}_x$ for any $x \in \mathbb{S}$. Further, $\mathbb{S} \setminus C$ is a subset of G_r hence a finite set, so it follows by irreducibility of $\mathbb{S}$ that $\mathbf{P}_x(\tau_C < \infty) = 1$ for all $x \in \mathbb{S}$ (see part (a) of Exercise 6.2.5). Consequently, from Proposition 5.3.8 we get that

$$h(x) \geq \mathbf{E}_x h(X_{\tau_C}) \geq r\, \mathbf{P}_x(\tau_C < \tau_F)$$

(since $h(X_{\tau_C}) \geq r$ when $\tau_C < \tau_F$). Thus,

$$\mathbf{P}_x(\tau_F < \infty) \geq \mathbf{P}_x(\tau_F \leq \tau_C) \geq 1 - h(x)/r$$

and taking $r \to \infty$ we deduce that $\mathbf{P}_x(\tau_F < \infty) = 1$ for all $x \in \mathbb{S}$, as claimed. $\square$

Here is a concrete application of Proposition 6.2.21.

Exercise 6.2.22. *Suppose $\{S_n\}$ is an irreducible* random walk *on $\mathbb{Z}$ with zero-mean increments $\{\xi_k\}$ such that $|\xi_k| \leq r$ for some finite integer r. Show that $\{S_n\}$ is a recurrent chain.*

The following exercises complement Proposition 6.2.21.

Exercise 6.2.23. *Suppose that $\mathbb{S}$ is irreducible for some homogeneous Markov chain. Show that this chain is recurrent if and only if the only non-negative super-harmonic functions for it are the constant functions.*

Exercise 6.2.24. *Suppose $\{X_n\}$ is an irreducible* birth and death *chain with $p_i = p(i, i+1)$, $q_i = p(i, i-1)$ and $r_i = 1 - p_i - q_i = p(i, i) \geq 0$, where p_i and q_i are positive for $i > 0$, $q_0 = 0$ and $p_0 > 0$. Let*

$$h(m) = \sum_{k=0}^{m-1} \prod_{j=1}^{k} \frac{q_j}{p_j},$$

for $m \geq 1$ and $h(0) = 0$.

(a) *Check that $h(\cdot)$ is harmonic for the chain at all positive integers.*
(b) *Fixing $a < x < b$ in $\mathbb{S} = \mathbb{Z}_+$ verify that $\mathbf{P}_x(\tau_C < \infty) = 1$ for $C = \{a, b\}$ and that $h(X_{n \wedge \tau_C})$ is a bounded martingale under $\mathbf{P}_x$. Deduce that*

$$\mathbf{P}_x(T_a < T_b) = \frac{h(b) - h(x)}{h(b) - h(a)}$$

(c) *Considering $a = 0$ and $b \to \infty$ show that the chain is transient if and only if $h(\cdot)$ is bounded above.*
(d) *Suppose $i(p_i/q_i - 1) \to c$ as $i \to \infty$. Show that the chain is recurrent if $c < 1$ and transient if $c > 1$, so in particular, when $p_i = p = 1 - q_i$ for all $i > 0$ the chain is recurrent if and only if $p \leq \frac{1}{2}$.*

6.2.2. Invariant, excessive and reversible measures. Recall Proposition 6.1.23 that an *invariant measure* for the transition probability $p(x,y)$ is uniquely determined by a non-zero $\mu : \mathbb{S} \mapsto [0,\infty)$ such that

$$\mu(y) = \sum_{x\in\mathbb{S}} \mu(x)p(x,y)\,, \qquad \forall y \in \mathbb{S}\,. \tag{6.2.5}$$

To simplify our notations we thus regard such a function μ as the corresponding invariant measure. Similarly, we say that $\mu : \mathbb{S} \mapsto [0,\infty)$ is a finite, positive, or probability measure, when $\sum_x \mu(x)$ is finite (positive, or equals one, respectively), and call $\{x : \mu(x) > 0\}$ the *support* of the measure μ.

Definition 6.2.25. *Relaxing the notion of invariance we say that a non-zero* $\mu : \mathbb{S} \mapsto [0,\infty]$ *is an* excessive measure *if*

$$\mu(y) \geq \sum_{x\in\mathbb{S}} \mu(x)p(x,y)\,, \qquad \forall y \in \mathbb{S}\,.$$

Example 6.2.26. *Some chains do not have any invariant measure. For example, in a birth and death chain with $p_i = 1$, $i \geq 0$ the identity (6.2.5) is merely $\mu(0) = 0$ and $\mu(i) = \mu(i-1)$ for $i \geq 1$, whose only solution is the zero function. However, the totally asymmetric* SRW *on $\mathbb{Z}$ with $p(x, x+1) = 1$ at every integer x has an invariant measure $\mu(x) = 1$, although just as in the preceding birth and death chain all its states are transient with the only closed set being the whole state space.*

Nevertheless, as we show next, to every *recurrent* state corresponds an invariant measure.

Proposition 6.2.27. *Let T_z denote the possibly infinite return time to a state z by a homogeneous Markov chain $\{X_n\}$. Then,*

$$\mu_z(y) = \mathbf{E}_z\Big[\sum_{n=0}^{T_z-1} I_{\{X_n=y\}}\Big]\,,$$

is an excessive measure for $\{X_n\}$, the support of which is the closed set of all states accessible from z. If z is recurrent then $\mu_z(\cdot)$ is an invariant measure, whose support is the closed and recurrent $\leftrightarrow$ equivalence class of z.

Remark. We have by the second claim of Proposition 6.2.15 (for the closed set $\mathbb{S}$), that any chain with a finite state space has at least one recurrent state. Further, recall that any invariant measure is σ-finite, which for a finite state space amounts to being a finite measure. Hence, by Proposition 6.2.27 any chain with a finite state space has at least one invariant probability measure.

Example 6.2.28. *For a transient state z the excessive measure $\mu_z(y)$ may be infinite at some $y \in \mathbb{S}$. For example, the transition probability $p(x,0) = 1$ for all $x \in \mathbb{S} = \{0,1\}$ has 0 as an absorbing (recurrent) state and 1 as a transient state, with $T_1 = \infty$ and $\mu_1(1) = 1$ while $\mu_1(0) = \infty$.*

PROOF. Using the canonical construction of the chain, we set

$$h_k(\omega, y) = \sum_{n=0}^{T_z(\omega)-1} I_{\{\omega_{n+k}=y\}}\,,$$

so that $\mu_z(y) = \mathbf{E}_z h_0(\omega, y)$. By the tower property and the Markov property of the chain,

$$\begin{aligned}
\mathbf{E}_z h_1(\omega, y) &= \mathbf{E}_z \Big[\sum_{n=0}^{\infty} I_{\{T_z > n\}} I_{\{X_{n+1}=y\}} \sum_{x \in \mathbb{S}} I_{\{X_n = x\}} \Big] \\
&= \sum_{x \in \mathbb{S}} \sum_{n=0}^{\infty} \mathbf{E}_z \Big[I_{\{T_z > n\}} I_{\{X_n = x\}} \mathbf{P}_z(X_{n+1} = y | \mathcal{F}_n^{\mathbf{X}}) \Big] \\
&= \sum_{x \in \mathbb{S}} \sum_{n=0}^{\infty} \mathbf{E}_z \Big[I_{\{T_z > n\}} I_{\{X_n = x\}} \Big] p(x, y) = \sum_{x \in \mathbb{S}} \mu_z(x) p(x, y) \,.
\end{aligned}$$

The key to the proof is the observation that if $\omega_0 = z$ then $h_0(\omega, y) \geq h_1(\omega, y)$ for any $y \in \mathbb{S}$, with equality when $y \neq z$ or $T_z(\omega) < \infty$ (in which case $\omega_{T_z(\omega)} = \omega_0$). Consequently, for any state y,

$$\mu_z(y) = \mathbf{E}_z h_0(\omega, y) \geq \mathbf{E}_z h_1(\omega, y) = \sum_{x \in \mathbb{S}} \mu_z(x) p(x, y) \,,$$

with equality when $y \neq z$ or z is recurrent (in which case $\mathbf{P}_z(T_z < \infty) = 1$). By definition $\mu_z(z) = 1$, so $\mu_z(\cdot)$ is an excessive measure. Iterating the preceding inequality k times we further deduce that $\mu_z(y) \geq \sum_x \mu_z(x) \mathbf{P}_x(X_k = y)$ for any $k \geq 1$ and $y \in \mathbb{S}$, with equality when z is recurrent. If $\rho_{zy} = 0$ then clearly $\mu_z(y) = 0$, while if $\rho_{zy} > 0$ then $\mathbf{P}_z(X_k = y) > 0$ for some k finite, hence $\mu_z(y) \geq \mu_z(z) \mathbf{P}_z(X_k = y) > 0$. The support of μ_z is thus the closed set of states accessible from z, which for z recurrent is its $\leftrightarrow$ equivalence class. Finally, note that if $x \leftrightarrow z$ then $\mathbf{P}_x(X_k = z) > 0$ for some k finite, so $1 = \mu_z(z) \geq \mu_z(x) \mathbf{P}_x(X_k = z)$ implying that $\mu_z(x) < \infty$. That is, if z is recurrent then μ_z is a σ-finite, positive invariant measure, as claimed. $\square$

What about uniqueness of the invariant measure for a given transition probability? By definition the set of invariant measures for $p(\cdot, \cdot)$ is a *convex cone* (that is, if μ_1 and μ_2 are invariant measures, possibly the same, then for any positive c_1 and c_2 the measure $c_1\mu_1 + c_2\mu_2$ is also invariant). Thus, hereafter we say that the invariant measure is *unique* whenever it is unique up to multiplication by a positive constant.

The first negative result in this direction comes from Proposition 6.2.27. Indeed, the invariant measures μ_z and μ_x are clearly *mutually singular* (and in particular, not constant multiple of each other), whenever the two recurrent states x and z do not intercommunicate. In contrast, your next exercise yields a positive result, that the invariant measure supported within each recurrent equivalence class of states is unique (and given by Proposition 6.2.27).

Exercise 6.2.29. *Suppose $\mu : \mathbb{S} \mapsto (0, \infty)$ is a strictly positive invariant measure for the transition probability $p(\cdot, \cdot)$ of a Markov chain $\{X_n\}$ on the countable set $\mathbb{S}$.*

(a) *Verify that $q(x, y) = \mu(y) p(y, x) / \mu(x)$ is a transition probability on $\mathbb{S}$.*
(b) *Verify that if $\nu : \mathbb{S} \mapsto [0, \infty)$ is an excessive measure for $p(\cdot, \cdot)$ then $h(x) = \nu(x)/\mu(x)$ is* super-harmonic *for $q(\cdot, \cdot)$.*
(c) *Show that if $p(\cdot, \cdot)$ is irreducible and recurrent, then so is $q(\cdot, \cdot)$. Deduce from Exercise 6.2.23 that then $h(x)$ is a constant function, hence $\nu(x) = c\mu(x)$ for some $c > 0$ and all $x \in \mathbb{S}$.*

Proposition 6.2.30. *If $\mathbf{R}$ is a recurrent $\leftrightarrow$ equivalence class of states then the invariant measure whose support is contained in $\mathbf{R}$ is unique (and has $\mathbf{R}$ as its support). In particular, the invariant measure of an irreducible, recurrent chain is unique (up to multiplication by a constant) and strictly positive.*

PROOF. Recall the decomposition theorem that $\mathbf{R}$ is closed, hence the restriction of $p(\cdot,\cdot)$ to $\mathbf{R}$ is also a transition probability and when considering invariant measures supported within $\mathbf{R}$ we may as well take $\mathbb{S} = \mathbf{R}$. That is, hereafter we assume that the chain is recurrent. In this case we have by Proposition 6.2.27 a strictly positive invariant measure $\mu = \mu_z$ on $\mathbb{S} = \mathbf{R}$. To complete the proof recall the conclusion of Exercise 6.2.29 that any σ-finite excessive measure (and in particular any invariant measure), is then a constant multiple of μ. $\square$

Propositions 6.2.27 and 6.2.30 provide a complete picture of the invariant measures supported outside the set $\mathbb{T}$ of transient states, as the convex cone generated by the mutually singular, unique invariant measures $\mu_z(\cdot)$ supported on each closed recurrent $\leftrightarrow$ equivalence class. Complementing it, your next exercise shows that an invariant measure must be zero at all transient states that lead to at least one recurrent state and if it is positive at some $v \in \mathbb{T}$ then it is also positive at any $y \in \mathbb{T}$ accessible from v.

Exercise 6.2.31. *Let $\mu(\cdot)$ be an invariant measure for a Markov chain $\{X_k\}$ on $\mathbb{S}$.*

(a) *Iterating (6.2.5) verify that $\mu(y) = \sum_x \mu(x)\mathbf{P}_x(X_k = y)$ for all $k \geq 1$ and $y \in \mathbb{S}$.*
(b) *Deduce that if $\mu(v) > 0$ for some $v \in \mathbb{S}$ then $\mu(y) > 0$ for any y accessible from v.*
(c) *Show that if $\mathbf{R}$ is a recurrent $\leftrightarrow$ equivalence class then $\mu(x)p(x,y) = 0$ for all $x \notin \mathbf{R}$ and $y \in \mathbf{R}$.*
Hint: *Exercise 6.2.29 may be handy here.*
(d) *Deduce that if such $\mathbf{R}$ is accessible from $v \notin \mathbf{R}$ then $\mu(v) = 0$.*

We complete our discussion of (non)-uniqueness of the invariant measure with an example of a transient chain having two strictly positive invariant measures that are not constant multiple of each other.

Example 6.2.32 (SRW ON $\mathbb{Z}$)**.** *Consider the* SRW*, a homogeneous Markov chain with state space $\mathbb{Z}$ and transition probability $p(x, x+1) = 1 - p(x, x-1) = p$ for some $0 < p < 1$. You can easily verify that both the counting measure $\widetilde{\lambda}(x) \equiv 1$ and $\mu_0(x) = (p/(1-p))^x$ are invariant measures for this chain, with μ_0 a constant multiple of $\widetilde{\lambda}$ only in the symmetric case $p = 1/2$. Recall Exercise 6.2.20 that this chain is transient for $p \neq 1/2$ and recurrent for $p = 1/2$ and observe that neither $\widetilde{\lambda}$ nor μ_0 is a finite measure. Indeed, as we show in the sequel, a finite invariant measure of a Markov chain must be zero at all transient states.*

Remark. Evidently, having a uniform (or counting) invariant measure (i.e. $\mu(x) \equiv c > 0$ for all $x \in \mathbb{S}$), as in the preceding example, is equivalent to the transition probability being *doubly stochastic*, that is, $\sum_{x \in \mathbb{S}} p(x,y) = 1$ for all $y \in \mathbb{S}$.

Example 6.2.32 motivates our next subject, which are the conditions under which a Markov chain is reversible, starting with the relevant definitions.

Definition 6.2.33. *A non-zero $\mu : \mathbb{S} \mapsto [0,\infty)$ is called a* reversible measure *for the transition probability $p(\cdot,\cdot)$ if the* detailed balance *relation $\mu(x)p(x,y) = \mu(y)p(y,x)$ holds for all $x,y \in \mathbb{S}$. We say that a transition probability $p(\cdot,\cdot)$ (or the corresponding Markov chain) is* reversible *if it has a reversible measure.*

Remark. Every reversible measure is an invariant measure, for summing the detailed balance relation over $x \in \mathbb{S}$ yields the identity (6.2.5), but there are non-reversible invariant measures. For example, the uniform invariant measure of a doubly stochastic transition probability $p(\cdot,\cdot)$ is non-reversible as soon as $p(x,y) \neq p(y,x)$ for some $x,y \in \mathbb{S}$. Indeed, for the asymmetric SRW of Example 6.2.32 (i.e., when $p \neq 1/2$), the (constant) counting measure $\widetilde{\lambda}$ is non-reversible while μ_0 is a reversible measure (as you can easily check on your own).

As their name suggest, reversible measures have to do with the time reversed chain (and the corresponding adjoint transition probability), which we now define.

Definition 6.2.34. *If $\mu(\cdot)$ is an invariant measure for transition probability $p(x,y)$, then $q(x,y) = \mu(y)p(y,x)/\mu(x)$ is a transition probability on the support of $\mu(\cdot)$, which we call the* adjoint *(or* dual*) of $p(\cdot,\cdot)$ with respect to μ. The corresponding chain of law $\mathbf{Q}_\mu$ is called the* time reversed *chain (with respect to μ).*

It is not hard, and left to the reader, to check that for any invariant probability measure μ the stationary Markov chains $\{Y_n\}$ of law $\mathbf{Q}_\mu$ and $\{X_n\}$ of law $\mathbf{P}_\mu$ are such that $(Y_k,\dots,Y_\ell) \stackrel{\mathcal{D}}{=} (X_\ell,\dots,X_k)$ for any $k \leq \ell$ finite. Indeed, this is why $\{Y_n\}$ is called the time reversed chain.

Also note that $\mu(\cdot)$ is a reversible measure if and only if $p(\cdot,\cdot)$ is self-adjoint with respect to $\mu(\cdot)$ (that is, $q(x,y) = p(x,y)$ on the support of $\mu(\cdot)$). Alternatively put, $\mu(\cdot)$ is a reversible measure if and only if $\mathbf{P}_\mu = \mathbf{Q}_\mu$, that is, the shift invariant law of the chain induced by μ is the same as that of its time reversed chain.

By Definition 6.2.33 the set of reversible measures for $p(\cdot,\cdot)$ is a convex cone. The following exercise affirms that reversible measures are zero outside the closed $\leftrightarrow$ equivalence classes of the chain and uniquely determined by it within each such class. It thus reduces the problem of characterizing reversible chains (and measures) to doing so for irreducible chains.

Exercise 6.2.35. *Suppose $\mu(x)$ is a reversible measure for the transition probability $p(x,y)$ of a Markov chain $\{X_n\}$ with a countable state space $\mathbb{S}$.*

(a) *Show that $\mu(x)\mathbf{P}_x(X_k = y) = \mu(y)\mathbf{P}_y(X_k = x)$ for any $x,y \in \mathbb{S}$ and all $k \geq 1$.*
(b) *Deduce that if $\mu(x) > 0$ then any y accessible from x must intercommunicate with x.*
(c) *Conclude that the support of $\mu(\cdot)$ is a disjoint union of closed $\leftrightarrow$ equivalence classes, within each of which the measure μ is uniquely determined by $p(\cdot,\cdot)$ up to a non-negative constant multiple.*

We proceed to characterize reversible irreducible Markov chains as random walks on *networks*.

Definition 6.2.36. *A* network *(or* weighted graph*) consists of a countable (finite or infinite) set of vertices $\mathbb{V}$ with a symmetric weight function $w : \mathbb{V} \times \mathbb{V} \mapsto [0,\infty)$ (i.e. $w_{xy} = w_{yx}$ for all $x,y \in \mathbb{V}$). Further requiring that $\mu(x) = \sum_{y\in\mathbb{V}} w_{xy}$ is* *finite and positive for each* $x \in \mathbb{V}$*, a* random walk *on the network is a homogeneous Markov chain of state space* $\mathbb{V}$ *and transition probability* $p(x, y) = w_{xy}/\mu(x)$*. That is, when at state* x *the probability of the chain moving to state* y *is proportional to the weight* w_{xy} *of the pair* $\{x, y\}$*.*

Remark. For example, an undirected graph is merely a network the weights w_{xy} of which are either one (indicating an edge in the graph whose ends are x and y) or zero (no such edge). Assuming such graph has positive and finite degrees, the random walker moves at each time step to a vertex chosen uniformly at random from those adjacent in the graph to its current position.

Exercise 6.2.37. *Check that a random walk on a network has a strictly positive reversible measure* $\mu(x) = \sum_y w_{xy}$ *and that a Markov chain is reversible if and only if there exists an irreducible closed set* $\mathbb{V}$ *on which it is a random walk (with weights* $w_{xy} = \mu(x)p(x, y)$*).*

Example 6.2.38 (BIRTH AND DEATH CHAIN)**.** *We leave for the reader to check that the irreducible* birth and death chain *of Exercise 6.2.24 is a random walk on the network* $\mathbb{Z}_+$ *with weights* $w_{x,x+1} = p_x\mu(x) = q_{x+1}\mu(x + 1)$*,* $w_{xx} = r_x\mu(x)$ *and* $w_{xy} = 0$ *for* $|x - y| > 1$*, and the unique reversible measure* $\mu(x) = \prod_{i=1}^{x} \frac{p_{i-1}}{q_i}$ *(with* $\mu(0) = 1$*).*

Remark. Though irreducibility does not imply uniqueness of the invariant measure (c.f. Example 6.2.32), if μ is an invariant measure of the preceding birth and death chain then $\mu(x + 1)$ is determined by (6.2.5) from $\mu(x)$ and $\mu(x - 1)$, so starting at $\mu(0) = 1$ we conclude that the reversible measure of Example 6.2.38 is also the *unique* invariant measure for this chain.

We conclude our discussion of reversible measures with an explicit condition for reversibility of an irreducible chain, whose proof is left for the reader (for example, see [**Dur10**, Theorem 6.5.1]).

Exercise 6.2.39 (KOLMOGOROV'S CYCLE CONDITION)**.** *Show that an irreducible chain of transition probability* $p(x, y)$ *is reversible if and only if* $p(x, y) > 0$ *whenever* $p(y, x) > 0$ *and*

$$\prod_{i=1}^{k} p(x_{i-1}, x_i) = \prod_{i=1}^{k} p(x_i, x_{i-1}) \,,$$

for any $k \geq 3$ *and any cycle* $x_0, x_1, \ldots, x_k = x_0$*.*

Remark. The renewal Markov chain of Example 6.1.11 is one of the many recurrent chains that fail to satisfy Kolmogorov's condition (and thus are not reversible).

Turning to investigate the existence and support of *finite* invariant measures (or equivalently, that of *invariant probability measures*), we further partition the recurrent states of the chain according to the integrability (or lack thereof) of the corresponding return times.

Definition 6.2.40. *With* T_z *denoting the first return time to state* z*, a* recurrent *state* z *is called* positive recurrent *if* $\mathbf{E}_z(T_z) < \infty$ *and* null recurrent *otherwise.*

Indeed, invariant probability measures require the existence of positive recurrent states, on which they are supported.

Proposition 6.2.41. *If $\pi(\cdot)$ is an invariant probability measure then all states z with $\pi(z) > 0$ are positive recurrent. Further, if the support of $\pi(\cdot)$ is an irreducible set $\mathbf{R}$ of positive recurrent states then $\pi(z) = 1/\mathbf{E}_z(T_z)$ for all $z \in \mathbf{R}$.*

PROOF. Recall Proposition 6.2.10 that for any initial probability measure $\pi(\cdot)$ the number of visits $N_\infty(z) = \sum_{n\geq 1} I_{X_n=z}$ to a state z by the chain is such that

$$\sum_{n=1}^{\infty} \mathbf{P}_\pi(X_n = z) = \mathbf{E}_\pi N_\infty(z) = \sum_{x\in\mathbb{S}} \pi(x)\mathbf{E}_x N_\infty(z) = \sum_{x\in\mathbb{S}} \pi(x)\frac{\rho_{xz}}{1-\rho_{zz}} \leq \frac{1}{1-\rho_{zz}}$$

(since $\rho_{xz} \leq 1$ for all x). Starting at X_0 chosen according to an invariant probability measure $\pi(\cdot)$ results with a stationary Markov chain $\{X_n\}$ and in particular $\mathbf{P}_\pi(X_n = z) = \pi(z)$ for all n. The left side of the preceding inequality is thus infinite for positive $\pi(z)$ and invariant probability measure $\pi(\cdot)$. Consequently, in this case $\rho_{zz} = 1$, or equivalently z must be a recurrent state of the chain. Since this applies for any $z \in \mathbb{S}$ we conclude that $\pi(\cdot)$ is supported outside the set $\mathbb{T}$ of transient states.

Next, recall that for any $z \in \mathbb{S}$,

$$\mu_z(\mathbb{S}) = \sum_{y\in\mathbb{S}} \mu_z(y) = \mathbf{E}_z\Big[\sum_{y\in\mathbb{S}}\sum_{n=0}^{T_z-1} I_{\{X_n=y\}}\Big] = \mathbf{E}_z T_z\,,$$

so μ_z is a finite measure if and only if z is a positive recurrent state of the chain. If the support of $\pi(\cdot)$ is an irreducible $\leftrightarrow$ equivalence class $\mathbf{R}$ then we deduce from Propositions 6.2.27 and 6.2.30 that μ_z is a finite measure and $\pi(z) = \mu_z(z)/\mu_z(\mathbb{S}) = 1/\mathbf{E}_z T_z$ for any $z \in \mathbf{R}$. Consequently, $\mathbf{R}$ must be a positive recurrent equivalence class, that is, all states of $\mathbf{R}$ are positive recurrent.

To complete the proof, note that by the decomposition theorem any invariant probability measure $\pi(\cdot)$ is a mixture of such invariant probability measures, each supported on a different closed recurrent class $\mathbf{R}_i$, which by the preceding argument must all be positive recurrent. □

In the course of proving Proposition 6.2.41 we have shown that positive and null recurrence are $\leftrightarrow$ equivalence *class properties*. That is, an irreducible set of states C is either *positive recurrent* (i.e. every $z \in C$ is positive recurrent), *null recurrent* (i.e. every $z \in C$ is null recurrent), or transient. Further, recall the discussion after Proposition 6.2.27, that any chain with a finite state space has an invariant probability measure, from which we get the following corollary.

Corollary 6.2.42. *For an irreducible Markov chain the existence of an invariant probability measure is equivalent to the existence of a positive recurrent state, in which case every state is positive recurrent. We call such a chain* positive recurrent *and note that any irreducible chain with a finite state space is positive recurrent.*

For the remainder of this section we consider the existence and non-existence of invariant probability measures for some Markov chains of interest.

Example 6.2.43. *Since the invariant measure of a recurrent chain is unique up to a constant multiple (see Proposition 6.2.30) and a transient chain has no invariant probability measure (see Corollary 6.2.42), if an irreducible chain has an invariant measure $\mu(\cdot)$ for which $\sum_x \mu(x) = \infty$ then it has no invariant probability measure.*

For example, since the counting measure $\widetilde{\lambda}$ *is an invariant measure for the (irreducible)* SRW *of Example 6.2.32, this chain does not have an invariant probability measure, regardless of the value of* p. *For the same reason, the symmetric* SRW *on* $\mathbb{Z}$ *(i.e. where* $p = 1/2$*), is a null recurrent chain.*

Similarly, the irreducible birth and death chain *of Exercise 6.2.24 has an invariant probability measure if and only if its reversible measure* $\mu(x) = \prod_{i=1}^{x} \frac{p_{i-1}}{q_i}$ *is finite (c.f. Example 6.2.38). In particular, if* $p_j = 1 - q_j = p$ *for all* $j \geq 1$ *then this chain is positive recurrent with an invariant probability measure when* $p < 1/2$ *but null recurrent for* $p = 1/2$ *(and transient when* $1 > p > 1/2$*).*

Finally, a random walk on a graph *is irreducible if and only if the graph is connected. With* $\mu(v) \geq 1$ *for all* $v \in \mathbb{V}$ *(see Definition 6.2.36), it is positive recurrent only for finite graphs.*

Exercise 6.2.44. *Check that* $\mu(j) = \sum_{k>j} q_k$ *is an invariant measure for the recurrent renewal Markov chain of Example 6.1.11 in case* $\{k : q_k > 0\}$ *is unbounded (see Example 6.2.19). Conclude that this chain is positive recurrent if and only if* $\sum_k kq_k$ *is finite.*

In the next exercise you find how the invariant probability measure is modified by the introduction of *holding times*.

Exercise 6.2.45. *Let* $\pi(\cdot)$ *be the unique invariant probability measure of an irreducible, positive recurrent Markov chain* $\{X_n\}$ *with transition probability* $p(x, y)$ *such that* $p(x, x) = 0$ *for all* $x \in \mathbb{S}$. *Fixing* $r(x) \in (0, 1)$, *consider the Markov chain* $\{Y_n\}$ *whose transition probability is* $q(x, x) = 1 - r(x)$ *and* $q(x, y) = r(x)p(x, y)$ *for all* $y \neq x$. *Show that* $\{Y_n\}$ *is an irreducible, recurrent chain of invariant measure* $\mu(x) = \pi(x)/r(x)$ *and deduce that* $\{Y_n\}$ *is further positive recurrent if and only if* $\sum_x \pi(x)/r(x) < \infty$.

Though we have established the next result in a more general setting, the proof we outline here is elegant, self-contained and instructive.

Exercise 6.2.46. *Suppose* $g(\cdot)$ *is a strictly concave bounded function on* $[0, \infty)$ *and* $\pi(\cdot)$ *is a strictly positive invariant probability measure for irreducible transition probability* $p(x, y)$. *For any* $\nu : \mathbb{S} \mapsto [0, \infty)$ *let* $(\nu p)(y) = \sum_{x \in \mathbb{S}} \nu(x)p(x, y)$ *and*

$$\mathcal{E}(\nu) = \sum_{y \in \mathbb{S}} g\left(\frac{\nu(y)}{\pi(y)}\right) \pi(y) \,.$$

(a) *Show that* $\mathcal{E}(\nu p) \geq \mathcal{E}(\nu)$.
(b) *Assuming* $p(x, y) > 0$ *for all* $x, y \in \mathbb{S}$ *deduce from part (a) that any invariant measure* $\mu(\cdot)$ *for* $p(x, y)$ *is a constant multiple of* $\pi(\cdot)$.
(c) *Extend this conclusion to any irreducible* $p(x, y)$ *by checking that*

$$\widehat{p}(x, y) = \sum_{n=1}^{\infty} 2^{-n} \mathbf{P}_x(X_n = y) > 0 \,, \qquad \forall x, y \in \mathbb{S} \,,$$

and that invariant measures for $p(x, y)$ *are also invariant for* $\widehat{p}(x, y)$.

Here is an introduction to the powerful method of Lyapunov (or energy) functions.

Exercise 6.2.47. *Let* $\tau_z = \inf\{n \geq 0 : Z_n = z\}$ *and* $\mathcal{F}_n^{\mathbf{Z}} = \sigma(Z_k, k \leq n)$, *for Markov chain* $\{Z_n\}$ *of transition probabilities* $p(x, y)$ *on a countable state space* $\mathbb{S}$.

(a) *Show that* $V_n = Z_{n\wedge\tau_z}$ *is a* $\mathcal{F}_n^{\mathbf{Z}}$*-Markov chain and compute its transition probabilities* $q(x,y)$.

(b) *Suppose* $h : \mathbb{S} \mapsto [0,\infty)$ *is such that* $h(z) = 0$*, the function* $(ph)(x) = \sum_y p(x,y)h(y)$ *is finite everywhere and* $h(x) \geq (ph)(x) + \delta$ *for some* $\delta > 0$ *and all* $x \neq z$*. Show that* $(W_n, \mathcal{F}_n^{\mathbf{Z}})$ *is a sup-MG under* $\mathbf{P}_x$ *for* $W_n = h(V_n) + \delta(n \wedge \tau_z)$ *and any* $x \in \mathbb{S}$.

(c) *Deduce that* $\mathbf{E}_x\tau_z \leq h(x)/\delta$ *for any* $x \in \mathbb{S}$ *and conclude that* z *is positive recurrent in the stronger sense that* $\mathbf{E}_xT_z$ *is finite for all* $x \in \mathbb{S}$.

(d) *Fixing* $\delta > 0$ *consider i.i.d. random vectors* $v_k = (\xi_k, \eta_k)$ *such that* $\mathbf{P}(v_1 = (1,0)) = \mathbf{P}(v_1 = (0,1)) = 0.25 - \delta$ *and* $\mathbf{P}(v_1 = (-1,0)) = \mathbf{P}(v_1 = (0,-1)) = 0.25 + \delta$*. The chain* $Z_n = (X_n, Y_n)$ *on* $\mathbb{Z}^2$ *is such that* $X_{n+1} = X_n + \mathrm{sgn}(X_n)\xi_{n+1}$ *and* $Y_{n+1} = Y_n + \mathrm{sgn}(Y_n)\eta_{n+1}$*, where* $\mathrm{sgn}(0) = 0$*. Prove that* $(0,0)$ *is positive recurrent in the sense of part (c).*

Exercise 6.2.48. *Consider the Markov chain* $Z_n = \xi_n + (Z_{n-1} - 1)_+$*,* $n \geq 1$*, on* $\mathbb{S} = \{0,1,2,\ldots\}$*, where* ξ_n *are i.i.d.* $\mathbb{S}$*-valued such that* $\mathbf{P}(\xi_1 > 1) > 0$ *and* $\mathbf{E}\xi_1 = 1 - \delta$ *for some* $\delta > 0$.

(a) *Show that* $\{Z_n\}$ *is positive recurrent.*

(b) *Find its invariant probability measure* $\pi(\cdot)$ *in case* $\mathbf{P}(\xi_1 = k) = p(1-p)^k$*,* $k \in \mathbb{S}$*, for some* $p \in (1/2, 1)$.

(c) *Is this Markov chain reversible?*

6.2.3. Aperiodicity and limit theorems. Building on our classification of states and study of the invariant measures of homogeneous Markov chains with countable state space $\mathbb{S}$, we focus here on the large n asymptotics of the state $X_n(\omega)$ of the chain and its law.

We start with the asymptotic behavior of the *occupation time*

$$N_n(y) = \sum_{\ell=1}^{n} I_{X_\ell = y}\,,$$

of state y by the Markov chain during its first n steps.

Proposition 6.2.49. *For any probability measure* ν *on* $\mathbb{S}$ *and all* $y \in \mathbb{S}$,

$$\lim_{n\to\infty} n^{-1}N_n(y) = \frac{1}{\mathbf{E}_y(T_y)} I_{\{T_y<\infty\}} \qquad \mathbf{P}_\nu\text{-a.s.} \tag{6.2.6}$$

Remark. This special case of the strong law of large numbers for Markov additive functionals (see Exercise 6.2.62 for its generalization), tells us that if a Markov chain visits a positive recurrent state then it asymptotically occupies it for a positive fraction of time, while the fraction of time it occupies each null recurrent or transient state is zero (hence the reason for the name *null recurrent*).

Proof. First note that if y is transient then $\mathbf{E}_xN_\infty(y)$ is finite by (6.2.3) for any $x \in \mathbb{S}$. Hence, $\mathbf{P}_\nu$-a.s. $N_\infty(y)$ is finite and consequently $n^{-1}N_n(y) \to 0$ as $n \to \infty$. Furthermore, since $\mathbf{P}_y(T_y = \infty) = 1 - \rho_{yy} > 0$, in this case $\mathbf{E}_y(T_y) = \infty$ and (6.2.6) follows.

Turning to consider recurrent $y \in \mathbb{S}$, note that if $T_y(\omega) = \infty$ then $N_n(y)(\omega) = 0$ for all n and (6.2.6) trivially holds. Thus, assuming hereafter that $T_y(\omega) < \infty$, we have by recurrence of y that a.s. $T_y^k(\omega) < \infty$ for all k (see Corollary 6.2.12). Recall part (b) of Exercise 6.2.11, that under $\mathbf{P}_\nu$ and conditional on $\{T_y < \infty\}$, the positive, finite random variables $\tau_k = T_y^k - T_y^{k-1}$ are independent of each other, with $\{\tau_k, k \geq 2\}$ further identically distributed and of mean value $\mathbf{E}_y(T_y)$. Since $N_n(y) = \sup\{k \geq 0 : T_y^k \leq n\}$, as you have showed in part (b) of Exercise 2.3.8, it follows from the strong law of large numbers that $n^{-1}N_n(y) \stackrel{a.s.}{\to} 1/\mathbf{E}_y(T_y)$ for $n \to \infty$. This completes the proof, as by assumption $I_{\{T_y<\infty\}} = 1$ in the present case. □

Here is a direct application of Proposition 6.2.49.

Exercise 6.2.50. *Consider the positions $\{X_n\}$ of a particle starting at $X_0 = x \in \mathbb{S}$ and moving in $\mathbb{S} = \{0, \dots, r\}$ according to the following rules. From any position $1 \leq y \leq r-1$ the particle moves to $y-1$ or $y+1$, and each such move is made with probability $1/2$ independently of all other moves, whereas from positions 0 and r the particle moves in one step to position $k \in \mathbb{S}$.*

(a) *Fixing $y \in \mathbb{S}$ and $k \in \{1, \dots, r-1\}$ find the almost sure limit $\pi(k, y)$ of $n^{-1}N_n(y)$ as $n \to \infty$.*

(b) *Find the almost sure limit $\pi(y)$ of $n^{-1}N_n(y)$ in case upon reaching either 0 or r the particle next moves to an independently and uniformly chosen position $K \in \{1, \dots, r-1\}$.*

Your next task is to prove the following *ratio limit theorem* for the occupation times $N_n(y)$ within each irreducible, closed recurrent set of states. In particular, it refines the limited information provided by Proposition 6.2.49 in case y is a null recurrent state.

Exercise 6.2.51. *Suppose $y \in \mathbb{S}$ is a recurrent state for the chain $\{X_n\}$. Let $\mu_y(\cdot)$ denote the invariant measure of the chain per Proposition 6.2.27, whose support is the closed and recurrent $\leftrightarrow$ equivalence class $\mathbf{R}_y$ of y. Decomposing the path $\{X_\ell\}$ at the successive return times T_y^k show that for any $x, w \in \mathbf{R}_y$,*

$$\lim_{n\to\infty} \frac{N_n(w)}{N_n(y)} = \mu_y(w), \qquad \mathbf{P}_x\text{-}a.s.$$

Hint: *Use Exercise 6.2.11 and the monotonicity of $n \mapsto N_n(w)$.*

Proceeding to study the asymptotics of $\mathbf{P}_x(X_n = y)$ we start with the following consequence of Proposition 6.2.49.

Corollary 6.2.52. *For all $x, y \in \mathbb{S}$,*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{\ell=1}^{n} \mathbf{P}_x(X_\ell = y) = \frac{\rho_{xy}}{\mathbf{E}_y(T_y)}. \tag{6.2.7}$$

Further, for any transient state $y \in \mathbb{T}$

$$\lim_{n\to\infty} \mathbf{P}_x(X_n = y) = \frac{\rho_{xy}}{\mathbf{E}_y(T_y)}. \tag{6.2.8}$$

Proof. Since $\sup_n n^{-1}N_n(y) \leq 1$, the convergence in (6.2.7) follows from Proposition 6.2.49 by bounded convergence (i.e. Corollary 1.3.46).

For a transient state y the sequence $\mathbf{P}_x(X_n = y)$ is summable (to the finite value $\mathbf{E}_x N_\infty(y)$, c.f. Proposition 6.2.10), hence $\mathbf{P}_x(X_n = y) \to 0$ as $n \to \infty$. Further, this amounts to (6.2.8) as in this case $\mathbf{E}_y(T_y) = \infty$. □

Corollary 6.2.52 tells us that for every Markov chain the Cesàro averages of $\mathbf{P}_x(X_n = y)$ converge. In contrast, our next example shows that even for an irreducible chain of finite state space the sequence $n \mapsto \mathbf{P}_x(X_n = y)$ may fail to converge *pointwise*.

Example 6.2.53. *Consider the Markov chain $\{X_n\}$ on state space $\mathbb{S} = \{0, 1\}$ with transition probabilities $p(x, y) = \mathbf{1}_{x \neq y}$. Then, $\mathbf{P}_x(X_n = y) = \mathbf{1}_{\{n \text{ even}\}}$ when $x = y$ and $\mathbf{P}_x(X_n = y) = \mathbf{1}_{\{n \text{ odd}\}}$ when $x \neq y$, so the sequence $n \mapsto \mathbf{P}_x(X_n = y)$ alternates between zero and one, having no limit for any fixed $(x, y) \in \mathbb{S}^2$.*

Nevertheless, as we prove in the sequel (more precisely, in Theorem 6.2.59), periodicity of the state y is the only reason for such non-convergence of $\mathbf{P}_x(X_n = y)$.

Definition 6.2.54. *The* period d_x *of a state $x \in \mathbb{S}$ of a Markov chain $\{X_n\}$ is the greatest common divisor (g.c.d.) of the set $\mathcal{I}_x = \{n \geq 1 : \mathbf{P}_x(X_n = x) > 0\}$, with $d_x = 0$ in case $\mathcal{I}_x$ is empty. Similarly, we say that the chain is of period d if $d_x = d$ for all $x \in \mathbb{S}$. A state x is called* aperiodic *if $d_x \leq 1$ and a Markov chain is called* aperiodic *if every $x \in \mathbb{S}$ is aperiodic.*

As the first step in this program, we show that the period is constant on each irreducible set.

Lemma 6.2.55. *The set $\mathcal{I}_x$ contains all large enough integer multiples of d_x and if $x \leftrightarrow y$ then $d_x = d_y$.*

Proof. Considering (6.2.4) for $x = y$ and $L = 0$ we find that $\mathcal{I}_x$ is closed under addition. Hence, this set contains all large enough integer multiples of d_x because *every* non-empty set $\mathcal{I}$ of positive integers which is closed under addition must contain all large enough integer multiples of its g.c.d. d. Indeed, it suffices to prove this fact when $d = 1$ since the general case then follows upon considering the non-empty set $\mathcal{I}' = \{n \geq 1 : nd \in \mathcal{I}\}$ whose g.c.d. is one (and which is also closed under addition). Further, note that any integer $n \geq \ell^2$ is of the form $n = \ell^2 + k\ell + r = r(\ell + 1) + (\ell - r + k)\ell$ for some $k \geq 0$ and $0 \leq r < \ell$. Hence, if two consecutive integers ℓ and $\ell + 1$ are in $\mathcal{I}$ then so are all integers $n \geq \ell^2$. We thus complete the proof by showing that $K = \inf\{m - \ell : m, \ell \in \mathcal{I}, m > \ell > 0\} > 1$ is in contradiction with $\mathcal{I}$ having g.c.d. $d = 1$. Indeed, both m_0 and $m_0 + K$ are in $\mathcal{I}$ for some positive integer m_0 and if $d = 1$ then $\mathcal{I}$ must contain also a positive integer of the form $m_1 = sK + r$ for some $0 < r < K$ and $s \geq 0$. With $\mathcal{I}$ closed under addition, $(s + 1)(m_0 + K) > (s + 1)m_0 + m_1$ must then both be in $\mathcal{I}$ but their difference is $(s+1)K - m_1 = K - r < K$, in contradiction with the definition of K.

If $x \leftrightarrow y$ then in view of the inequality (6.2.4) there exist finite K and L such that $K + n + L \in \mathcal{I}_x$ whenever $n \in \mathcal{I}_y$. Moreover, $K + L \in \mathcal{I}_x$ so every $n \in \mathcal{I}_y$ must also be an integer multiple of d_x. Consequently, d_x is a common divisor of $\mathcal{I}_y$ and therefore d_y, being the greatest common divisor of $\mathcal{I}_y$, is an integer multiple of d_x. Reversing the roles of x and y we likewise have that d_x is an integer multiple of d_y from which we conclude that in this case $d_x = d_y$. □

The key for determining the asymptotics of $\mathbf{P}_x(X_n = y)$ is to handle this question for aperiodic irreducible chains, to which end the next lemma is most useful.

Lemma 6.2.56. *Consider two independent copies $\{X_n\}$ and $\{Y_n\}$ of an aperiodic, irreducible chain on a countable state space $\mathbb{S}$ with transition probabilities $p(\cdot, \cdot)$. The Markov chain $Z_n = (X_n, Y_n)$ on $\mathbb{S}^2$ of transition probabilities $p_2((x', y'), (x, y)) = p(x', x)p(y', y)$ is then also aperiodic and irreducible. If $\{X_n\}$*

has invariant probability measure $\pi(\cdot)$ then $\{Z_n\}$ is further positive recurrent and has the invariant probability measure $\pi_2(x,y) = \pi(x)\pi(y)$.

Remark. Example 6.2.53 shows that for periodic $p(\cdot,\cdot)$ the chain of transition probabilities $p_2(\cdot,\cdot)$ may not be irreducible.

PROOF. Fix states $z' = (x', y') \in \mathbb{S}^2$ and $z = (x,y) \in \mathbb{S}^2$. Since $p(\cdot,\cdot)$ are the transition probabilities of an irreducible chain, there exist K and L finite such that $\mathbf{P}_{x'}(X_K = x) > 0$ and $\mathbf{P}_{y'}(Y_L = y) > 0$. Further, by the aperiodicity of this chain we have from Lemma 6.2.55 that both $\mathbf{P}_x(X_{n+L} = x) > 0$ and $\mathbf{P}_{y'}(Y_{K+n} = y') > 0$ for all n large enough, in which case from (6.2.4) we deduce that $\mathbf{P}_{z'}(Z_{K+n+L} = z) > 0$ as well. As this applies for any $z', z \in \mathbb{S}^2$, the chain $\{Z_n\}$ is irreducible. Further, considering $z' = z$ we see that $\mathcal{I}_z$ contains all large enough integers, hence $\{Z_n\}$ is also aperiodic. Finally, it is easy to verify that if $\pi(\cdot)$ is an invariant probability measure for $p(\cdot,\cdot)$ then $\pi_2(x,y) = \pi(x)\pi(y)$ is an invariant probability measure for $p_2(\cdot,\cdot)$, whose existence implies positive recurrence of the chain $\{Z_n\}$ (see Corollary 6.2.42). $\square$

The following *Markovian coupling* complements Lemma 6.2.56.

Theorem 6.2.57. *Let $\{X_n\}$ and $\{Y_n\}$ be two independent copies of an aperiodic, irreducible Markov chain. Suppose further that the irreducible chain $Z_n = (X_n, Y_n)$ is recurrent. Then, regardless of the initial distribution of (X_0, Y_0), the first* meeting time *$\tau = \min\{\ell \geq 0 : X_\ell = Y_\ell\}$ of the two processes is a.s. finite and for any n,*

$$\|\mathcal{P}_{X_n} - \mathcal{P}_{Y_n}\|_{tv} \leq 2\mathbf{P}(\tau > n)\,, \tag{6.2.9}$$

where $\|\cdot\|_{tv}$ denotes the total variation norm *of Definition 3.2.22.*

PROOF. Recall Lemma 6.2.56 that the Markov chain $Z_n = (X_n, Y_n)$ on $\mathbb{S}^2$ is irreducible. We have further assumed that $\{Z_n\}$ is recurrent, hence $\tau_z = \min\{\ell \geq 0 : Z_\ell = z\}$ is a.s. finite (for any $z \in \mathbb{S} \times \mathbb{S}$), regardless of the initial measure of $Z_0 = (X_0, Y_0)$. Consequently,

$$\tau = \inf\{\tau_z : z = (x,x) \text{ for some } x \in \mathbb{S}\}$$

is also a.s. finite, as claimed.

Turning to prove the inequality (6.2.9), fixing $g \in b\mathcal{S}$ bounded by one, recall that the chains $\{X_n\}$ and $\{Y_n\}$ have the same transition probabilities and further $X_\tau = Y_\tau$. Thus, for any $k \leq n$,

$$I_{\{\tau=k\}}\mathbf{E}_{X_k}[g(X_{n-k})] = I_{\{\tau=k\}}\mathbf{E}_{Y_k}[g(Y_{n-k})]\,.$$

By the Markov property and taking out the known $I_{\{\tau=k\}}$ it thus follows that

$$\begin{aligned}\mathbf{E}[I_{\{\tau=k\}}g(X_n)] &= \mathbf{E}(I_{\{\tau=k\}}\mathbf{E}_{X_k}[g(X_{n-k})]) \\ &= \mathbf{E}(I_{\{\tau=k\}}\mathbf{E}_{Y_k}[g(Y_{n-k})]) = \mathbf{E}[I_{\{\tau=k\}}g(Y_n)]\,.\end{aligned}$$

Summing over $0 \leq k \leq n$ we deduce that $\mathbf{E}[I_{\{\tau\leq n\}}g(X_n)] = \mathbf{E}[I_{\{\tau\leq n\}}g(Y_n)]$ and hence

$$\begin{aligned}\mathbf{E}g(X_n) - \mathbf{E}g(Y_n) &= \mathbf{E}[I_{\{\tau>n\}}g(X_n)] - \mathbf{E}[I_{\{\tau>n\}}g(Y_n)] \\ &= \mathbf{E}[I_{\{\tau>n\}}(g(X_n) - g(Y_n))]\,.\end{aligned}$$

Since $|g(X_n) - g(Y_n)| \leq 2$, we conclude that $|\mathbf{E}g(X_n) - \mathbf{E}g(Y_n)| \leq 2\mathbf{P}(\tau > n)$ for any $g \in b\mathcal{S}$ bounded by one, which is precisely what is claimed in (6.2.9). $\square$

Remark. Another Markovian coupling corresponds to replacing the transition probabilities $p_2((x',y'),(x,y))$ with $p(x',x)\mathbf{1}_{y=x}$ whenever $x'=y'$. Doing so extends the identity $Y_\tau = X_\tau$ to $Y_n = X_n$ for all $n \geq \tau$, thus yielding the bound $\mathbf{P}(X_n \neq Y_n) \leq \mathbf{P}(\tau > n)$ while each coordinate of the coupled chain evolves as before according to the original transition probabilities $p(\cdot,\cdot)$.

The tail behavior of the first meeting time τ controls the rate of convergence of $n \mapsto \mathbf{P}_x(X_n = y)$. As you are to show next, this convergence is exponentially fast when the state space is finite.

Exercise 6.2.58. *Show that if the aperiodic, irreducible Markov chain $\{X_n\}$ has finite state space, then $\mathbf{P}(\tau > n) \leq \exp(-\delta n)$ for the first meeting time τ of Theorem 6.2.57, some $\delta > 0$ and any n large enough.*
Hint: *First assume that $p(x,y) > 0$ for all $x,y \in \mathbb{S}$. Then show that $\mathbf{P}_x(X_r = y) > 0$ for some finite r and all x,y and consider the chain $\{Z_{nr}\}$.*

The following consequence of Theorem 6.2.57 is a major step in our analysis of the asymptotics of $\mathbf{P}_x(X_n = y)$.

Theorem 6.2.59. *The convergence (6.2.8) holds whenever y is an aperiodic state of the Markov chain $\{X_n\}$. In particular, if this Markov chain is irreducible, positive recurrent and aperiodic then for any $x \in \mathbb{S}$,*

$$\lim_{n\to\infty} \|\mathbf{P}_x(X_n \in \cdot) - \pi(\cdot)\|_{tv} = 0\,.$$

PROOF. If $\rho_{xy} = 0$ then $\mathbf{P}_x(X_n = y) = 0$ for all n and (6.2.8) trivially holds. Otherwise,

$$\rho_{xy} = \sum_{k=1}^{\infty} \mathbf{P}_x(T_y = k)\,,$$

is finite. Hence, in view of the first entrance decomposition

$$\mathbf{P}_x(X_n = y) = \sum_{k=1}^{n} \mathbf{P}_x(T_y = k)\mathbf{P}_y(X_{n-k} = y)$$

(see part (b) of Exercise 6.2.2), the asymptotics (6.2.8) follows by bounded convergence (with respect to the law of T_y conditional on $\{T_y < \infty\}$), from

$$\lim_{n\to\infty} \mathbf{P}_y(X_n = y) = \frac{1}{\mathbf{E}_y(T_y)}\,. \tag{6.2.10}$$

Turning to prove (6.2.10), in view of Corollary 6.2.52 we may and shall assume hereafter that y is an aperiodic recurrent state. Further, recall that by Theorem 6.2.13 it then suffices to consider the aperiodic, irreducible, recurrent chain $\{X_n\}$ obtained upon restricting the original Markov chain to the closed $\leftrightarrow$ equivalence class of y, which with some abuse of notation we denote hereafter also by $\mathbb{S}$.

Suppose first that $\{X_n\}$ is *positive recurrent* and so it has the invariant probability measure $\pi(w) = 1/\mathbf{E}_w(T_w)$ (see Proposition 6.2.41). The irreducible chain $Z_n = (X_n, Y_n)$ of Lemma 6.2.56 is then recurrent, so we apply Theorem 6.2.57 for $X_0 = y$ and Y_0 chosen according to the invariant probability measure π. Since Y_n is a *stationary* Markov chain (see Definition 6.1.20), in particular $Y_n \stackrel{\mathcal{D}}{=} Y_0$ has the law π for all n. Moreover, the corresponding first meeting time τ is a.s. finite. Hence, $\mathbf{P}(\tau > n) \downarrow 0$ as $n \to \infty$ and by (6.2.9) the law of X_n converges in total variation to π. This convergence in total variation further implies that $\mathbf{P}_y(X_n = y) \to \pi(y)$ when $n \to \infty$ (c.f. Example 3.2.25), which is precisely the statement of (6.2.10).

Next, consider a *null recurrent* aperiodic, irreducible chain $\{X_n\}$, in which case our thesis is that $\mathbf{P}_y(X_n = y) \to 0$ when $n \to \infty$. This clearly holds if the irreducible chain $\{Z_n\}$ of Lemma 6.2.56 is transient, for setting $z = (y, y)$ we then have upon applying Corollary 6.2.52 for the chain $\{Z_n\}$, that as $n \to \infty$

$$\mathbf{P}_z(Z_n = z) = \mathbf{P}_y(X_n = y)^2 \to 0\,.$$

Proceeding to prove our thesis when the chain $\{Z_n\}$ is recurrent, suppose to the contrary that the sequence $n \mapsto \mathbf{P}_y(X_n = y)$ has a limit point $\nu(y) > 0$. Then, mapping $\mathbb{S}$ in a one to one manner into $\mathbb{Z}$ we deduce from Helly's theorem that along a further sub-sequence n_ℓ the distributions of X_{n_ℓ} under $\mathbf{P}_y$ converge vaguely, hence pointwise (see Exercise 3.2.3), to some finite, positive measure ν on $\mathbb{S}$. We complete the proof of the theorem by showing that ν is an excessive measure for the irreducible, recurrent chain $\{X_n\}$. Indeed, By part (c) of Exercise 6.2.29 this would imply the existence of a finite invariant measure for $\{X_n\}$, in contradiction with our assumption that this chain is null recurrent (see Corollary 6.2.42).

To prove that ν is an excessive measure, note first that considering Theorem 6.2.57 for $Z_0 = (x, y)$ we get from (6.2.9) that $|\mathbf{P}_x(X_n = w) - \mathbf{P}_y(X_n = w)| \to 0$ as $n \to \infty$, for any $x, w \in \mathbb{S}$. Consequently, $\mathbf{P}_x(X_{n_\ell} = w) \to \nu(w)$ as $\ell \to \infty$, for every $x, w \in \mathbb{S}$. Moreover, from the Chapman-Kolmogorov equations we have that for any $w \in \mathbb{S}$, any finite set $F \subset \mathbb{S}$ and all $\ell \geq 1$,

$$\sum_{z \in \mathbb{S}} p(x, z)\mathbf{P}_z(X_{n_\ell} = w) = \mathbf{P}_x(X_{n_\ell + 1} = w) \geq \sum_{z \in F} \mathbf{P}_x(X_{n_\ell} = z)p(z, w)\,.$$

In the limit $\ell \to \infty$ this yields by bounded convergence (with respect to the probability measure $p(x, \cdot)$ on $\mathbb{S}$), that for all $w \in \mathbb{S}$

$$\nu(w) = \sum_{z \in \mathbb{S}} p(x, z)\nu(w) \geq \sum_{z \in F} \nu(z)p(z, w)\,.$$

Taking $F \uparrow \mathbb{S}$ we conclude by monotone convergence that $\nu(\cdot)$ is an excessive measure on $\mathbb{S}$, as we have claimed before. $\square$

Turning to the behavior of $\mathbf{P}_x(X_n = y)$ for periodic state y, we start with the following consequence of Theorem 6.2.59.

Corollary 6.2.60. *The convergence (6.2.8) holds whenever y is a null recurrent state of the Markov chain $\{X_n\}$ and if y is a positive recurrent state of $\{X_n\}$ having period $d = d_y$, then*

$$\lim_{n \to \infty} \mathbf{P}_y(X_{nd} = y) = \frac{d}{\mathbf{E}_y(T_y)}\,. \tag{6.2.11}$$

Proof. If $y \in \mathbb{S}$ has period $d \geq 1$ for the chain $\{X_n\}$ then $\mathbf{P}_y(X_n = y) = 0$ whenever n is not an integer multiple of d. Hence, the expected return time to such state y by the Markov chain $Y_n = X_{nd}$ is precisely $1/d$ of the expected return time $\mathbf{E}_y(T_y)$ for $\{X_n\}$. Therefore, (6.2.11) is merely a reformulation of the limit (6.2.10) for the chain $\{Y_n\}$ at its aperiodic state $y \in \mathbb{S}$.

If y is a null recurrent state of $\{X_n\}$ then $\mathbf{E}_y(T_y) = \infty$ so we have just established that $\mathbf{P}_y(X_n = y) \to 0$ as $n \to \infty$. It thus follows by the first entrance decomposition at T_y that in this case $\mathbf{P}_x(X_n = y) \to 0$ for any $x \in \mathbb{S}$ (as in the opening of the proof of Theorem 6.2.59). $\square$

In the next exercise, you extend (6.2.11) to the asymptotic behavior of $\mathbf{P}_x(X_n = y)$ for any two states x, y in a recurrent chain (which is not necessarily aperiodic).

Exercise 6.2.61. *Suppose $\{X_n\}$ is an irreducible, recurrent chain of period d. For each $x, y \in \mathbb{S}$ let $\mathcal{I}_{x,y} = \{n \geq 1 : \mathbf{P}_x(X_n = y) > 0\}$.*

(a) *Fixing $z \in \mathbb{S}$ show that there exist integers $0 \leq r_y < d$ such that if $n \in \mathcal{I}_{z,y}$ then d divides $n - r_y$.*
(b) *Show that if $n \in \mathcal{I}_{x,y}$ then $n = (r_y - r_x) \mod d$ and deduce that $\mathbb{S}_i = \{y \in \mathbb{S} : r_y = i\}$, $i = 0, \ldots, d-1$ are the irreducible $\leftrightarrow$ equivalence classes of the aperiodic chain $\{X_{nd}\}$ ($\mathbb{S}_i$ are called the* cyclic classes *of $\{X_n\}$).*
(c) *Show that for all $x, y, \in \mathbb{S}$,*

$$\lim_{n\to\infty} \mathbf{P}_x(X_{nd+r_y-r_x} = y) = \frac{d}{\mathbf{E}_y(T_y)} .$$

Remark. It is not always true that if a recurrent state y has period d then $\mathbf{P}_x(X_{nd+r} = y) \to d\rho_{xy}/\mathbf{E}_y(T_y)$ for some $r = r(x, y) \in \{0, \ldots, d-1\}$. Indeed, let $p(x, y)$ be the transition probabilities of the renewal chain with $q_1 = 0$ and $q_k > 0$ for $k \geq 2$ (see Example 6.1.11), *except for* setting $p(1, 2) = 1$ (instead of $p(1, 0) = 1$ in the renewal chain). The corresponding Markov chain has precisely two recurrent states, $y = 1$ and $y = 2$, both of period $d = 2$ and mean return times $\mathbf{E}_1(T_1) = \mathbf{E}_2(T_2) = 2$. Further, $\rho_{02} = 1$ but $\mathbf{P}_0(X_{nd} = 2) \to \eta$ and $\mathbf{P}_0(X_{nd+1} = 2) \to 1 - \eta$, where $\eta = \sum_k q_{2k}$ is strictly between zero and one.

We next consider the large n asymptotic behavior of the *Markov additive functional* $A_n^f = \sum_{\ell=1}^n f(X_\ell)$, where $\{X_\ell\}$ is an irreducible, positive recurrent Markov chain. In the following two exercises you establish first the *strong law of large numbers* (thereby generalizing Proposition 6.2.49), and then the *central limit theorem* for such Markov additive functionals.

Exercise 6.2.62. *Suppose $\{X_n\}$ is an irreducible, positive recurrent chain of initial probability measure ν and invariant probability measure $\pi(\cdot)$. Let $f : \mathbb{S} \mapsto \mathbb{R}$ be such that $\pi(|f|) < \infty$.*

(a) *Fixing $y \in \mathbb{S}$ let $R_k = T_y^k$. Show that the random variables*

$$Z_k^f = \sum_{\ell=R_{k-1}}^{R_k-1} f(X_\ell) , \qquad k \geq 1 ,$$

are mutually independent and moreover Z_k^f, $k \geq 2$ are identically distributed with $\mathbf{E}Z_2^{|f|}$ finite.
Hint: *Consider Exercise 6.2.11.*
(b) *With $S_n^f = \sum_{k=1}^{N_n(y)} Z_{k+1}^f$ show that*

$$\lim_{n\to\infty} n^{-1} S_n^f = \frac{\mathbf{E}Z_2^f}{\mathbf{E}_y(T_y)} = \pi(f) \qquad\qquad \mathbf{P}_\nu\text{-a.s.}$$

(c) *Show that $\mathbf{P}_\nu$-a.s. $\max\{n^{-1} Z_k^{|f|} : k \leq n\} \to 0$ when $n \to \infty$ and deduce that $n^{-1} A_n^f \to \pi(f)$ with $\mathbf{P}_\nu$ probability one.*

Exercise 6.2.63. *For $\{X_n\}$ as in Exercise 6.2.62 suppose that $f : \mathbb{S} \mapsto \mathbb{R}$ is such that $\pi(f) = 0$ and $v_{|f|} = \mathbf{E}_y[(Z_1^{|f|})^2]$ is finite.*

(a) *Show that* $n^{-1/2}S_n^f \xrightarrow{\mathcal{D}} \sqrt{u}G$ *as* $n \to \infty$, *for* $u = v_f/\mathbf{E}_y(T_y)$ *finite and* G *a standard normal variable.*
Hint: *See part (a) of Exercise 3.2.9.*

(b) *Show that* $\max\{n^{-1/2}Z_k^{|f|} : k \le n\} \xrightarrow{p} 0$ *and deduce that* $n^{-1/2}A_n^f \xrightarrow{\mathcal{D}} \sqrt{u}G$.

Building upon their strong law of large number, you are next to show that irreducible, positive recurrent chains have $\mathbf{P}$-trivial tail σ-algebra and the laws of any two such chains are mutually singular (for the analogous results for i.i.d. variables, see Corollary 1.4.10 and Remark 5.5.14, respectively).

Exercise 6.2.64. *Suppose* $\{X_n\}$ *is an irreducible, positive recurrent chain of law* $\mathbf{P}_x$ *on* $(\mathbb{S}_\infty, \mathcal{S}_c)$ *(as in Definition 6.1.7).*

(a) *Show that* $\mathbf{P}_x(A)$ *is independent of* $x \in \mathbb{S}$ *whenever* A *is in the tail* σ*-algebra* $\mathcal{T}^{\mathbf{X}}$ *(of Definition 1.4.9).*

(b) *Deduce that* $\mathcal{T}^{\mathbf{X}}$ *is* $\mathbf{P}$*-trivial.*

Exercise 6.2.65. *Suppose* $\{X_n\}$ *is an irreducible, positive recurrent chain of transition probability* $p(x, y)$, *initial and invariant probability measures* $\nu(\cdot)$ *and* $\pi(\cdot)$, *respectively.*

(a) *Show that* $\{X_n, X_{n+1}\}$ *is an irreducible, positive recurrent chain on* $\mathbb{S}_+^2 = \{(x, y) : x, y \in \mathbb{S}, p(x, y) > 0\}$, *of initial and invariant measures* $\nu(x)p(x, y)$ *and* $\pi(x)p(x, y)$, *respectively.*

(b) *Let* $\mathbf{P}_\nu$ *and* $\mathbf{P}'_\mu$ *denote the laws of two irreducible, positive recurrent chains on the same countable state space* $\mathbb{S}$, *whose transition probabilities* $p(x, y)$ *and* $p'(x, y)$ *are not identical. Show that* $\mathbf{P}_\nu$ *and* $\mathbf{P}'_\mu$ *are mutually singular measures (per Definition 4.1.9).*
Hint: *Consider the conclusion of Exercise 6.2.62 (for* $f(\cdot) = \mathbf{1}_x(\cdot)$, *or, if the invariant measures* π *and* π' *are identical, then for* $f(\cdot) = \mathbf{1}_{(x,y)}(\cdot)$ *and the induced pair-chains of part (a)).*

Exercise 6.2.66. *Fixing* $1 > \alpha > \beta > 0$ *let* $\mathbf{P}_n^{\alpha,\beta}$ *denote the law of* $(X_0, \ldots, X_n)$ *for the Markov chain* $\{X_k\}$ *of state space* $\mathbb{S} = \{-1, 1\}$ *starting from* $X_0 = -1$ *and evolving according to transition probability* $p(-1, -1) = \alpha = 1 - p(-1, 1)$ *and* $p(1, 1) = \beta = 1 - p(1, -1)$. *Fixing an integer* $b > 0$ *consider the stopping time* $\tau_b = \inf\{n \ge 0 : A_n = b\}$ *where* $A_n = \sum_{k=1}^n X_k$.

(a) *Setting* $\lambda_* = \log(\alpha/\beta)$, $h(-1) = 1$ *and* $h(1) = \beta(1 - \beta)/(\alpha(1 - \alpha))$, *show that the Radon-Nikodym derivative* $M_n = d\mathbf{P}_n^{\beta,\alpha}/d\mathbf{P}_n^{\alpha,\beta}$ *is of the form* $M_n = \exp(\lambda_* A_n)h(X_n)$.

(b) *Deduce that* $\mathbf{P}^{\alpha,\beta}(\tau_b < \infty) = \exp(-\lambda_* b)/h(1)$.

Exercise 6.2.67. *Suppose* $\{X_n\}$ *is a Markov chain of transition probability* $p(x, y)$ *and* $g(\cdot) = (ph)(\cdot) - h(\cdot)$ *for some bounded function* $h(\cdot)$ *on* $\mathbb{S}$. *Show that* $h(X_n) - \sum_{\ell=0}^{n-1} g(X_\ell)$ *is then a martingale.*

6.3. General state space: Doeblin and Harris chains

The refined analysis of homogeneous Markov chains with countable state space is possible because such chains hit states with positive probability. This does not happen in many important applications where the state space is uncountable. However, most proofs require only having one point of the state space that the chain hits with probability one. As we shall see, subject to the rather mild irreducibility and recurrence properties of Section 6.3.1, it is possible to create such a point (called a *recurrent atom*), even in an uncountable state space, by splitting the chain transitions. Guided by successive visits of the recurrent atom for the split chain, we establish in Section 6.3.2 the existence and attractiveness of invariant (probability) measures for the split chain (which then yield such results about the original chain).

6.3.1. Minorization, splitting, irreducibility and recurrence. Considering hereafter homogeneous Markov chains, we start by imposing a *minorization* property of the transition probability $p(\cdot,\cdot)$ which yields the *splitting* of these transitions.

Definition 6.3.1. *Consider a $\mathcal{B}$-isomorphic state space $(\mathbb{S},\mathcal{S})$. Suppose there exists a non-zero measurable function $v:\mathbb{S}\mapsto[0,1]$ and a probability measure $q(\cdot)$ on $(\mathbb{S},\mathcal{S})$ such that the transition probability of the chain $\{X_n\}$ is of the form*

$$p(x,\cdot)=(1-v(x))\widehat{p}(x,\cdot)+v(x)q(\cdot)\,, \tag{6.3.1}$$

for some transition probability $\widehat{p}(x,\cdot)$ and $v(x)q(\cdot)\ll\widehat{p}(x,\cdot)$. Amending the state space to $\overline{\mathbb{S}}=\mathbb{S}\cup\{\boldsymbol{\alpha}\}$ with the corresponding σ-algebra $\overline{\mathcal{S}}=\{A,A\cup\{\boldsymbol{\alpha}\}:A\in\mathcal{S}\}$, we then consider the split chain $\{\overline{X}_n\}$ *on $(\overline{\mathbb{S}},\overline{\mathcal{S}})$ with transition probability*

$$\begin{aligned}
\overline{p}(x,A)&=(1-v(x))\widehat{p}(x,A) && x\in\mathbb{S},\quad A\in\mathcal{S}\\
\overline{p}(x,\{\boldsymbol{\alpha}\})&=v(x) && x\in\mathbb{S}\\
\overline{p}(\boldsymbol{\alpha},B)&=\int q(dy)\overline{p}(y,B) && B\in\overline{\mathcal{S}}\,.
\end{aligned}$$

The transitions of $\{X_n\}$ on $\mathbb{S}$ have been split by moving to the pseudo-*atom* $\boldsymbol{\alpha}$ with probability $v(x)$. The random times in which the split chain is at state $\boldsymbol{\alpha}$ are *regeneration times* for $\{X_n\}$. That is, stopping times where future transitions are decoupled from the past. Indeed, the event $\overline{X}_n=\boldsymbol{\alpha}$ corresponds to X_n moving to a second copy of $\mathbb{S}$ where it is distributed according to the so called *regeneration measure* $q(\cdot)$, independently of X_{n-1}.

As the transitions of the split chain outside $\boldsymbol{\alpha}$ occur according to the excess probability $(1-v(x))\widehat{p}(x,\cdot)$, we can further merge the split chain to get back the original. That is,

Definition 6.3.2. *The* merge transition probability $m(\cdot,\cdot)$ *on $(\overline{\mathbb{S}},\overline{\mathcal{S}})$ is such that $m(x,\{x\})=1$ for all $x\in\mathbb{S}$ and $m(\boldsymbol{\alpha},\cdot)=q(\cdot)$. Associated with it is the* split mapping $f\mapsto\overline{f}:b\mathcal{S}\mapsto b\overline{\mathcal{S}}$ *such that $\overline{f}(\cdot)=(mf)(\cdot)=\int m(\cdot,dy)f(y)$.*

We note in passing that $\overline{f}(x)=f(x)$ for all $x\in\mathbb{S}$ and $\overline{f}(\boldsymbol{\alpha})=q(f)$, and further use in the sequel the following elementary fact about the closure of transition probabilities under composition.

Corollary 6.3.3. *Given any transition probabilities $\nu_i:\mathbb{X}\times\mathcal{X}\mapsto[0,1]$, $i=1,2$, the set function $\nu_1\nu_2:\mathbb{X}\times\mathcal{X}\mapsto[0,1]$ such that $\nu_1\nu_2(x,A)=\int\nu_1(x,dy)\nu_2(y,A)$ for all $x\in\mathbb{X}$ and $A\in\mathcal{X}$ is a transition probability.*

Proof. From Proposition 6.1.4 we see that

$$\nu_1\nu_2(x,A)=(\nu_1(x,\cdot)\otimes\nu_2)(\mathbb{X}\times A)=(\nu_1\nu_2(\cdot,A))(x)\,.$$

Now, by the first equality, $A \mapsto \nu_1\nu_2(x, A)$ is a probability measure on $(\mathbb{S}, \mathcal{S})$ for each $x \in \mathbb{S}$, and by the second equality, $x \mapsto \nu_1\nu_2(x, A)$ is a measurable function on $(\mathbb{S}, \mathcal{S})$ for each $A \in \mathcal{S}$, as required in Definition 6.1.2. □

Equipped with these notations we have the following coupling of $\{X_n\}$ and $\{\overline{X}_n\}$.

Proposition 6.3.4. *Consider the setup of Definitions 6.3.1 and 6.3.2.*
(a). $m\overline{p} = \overline{p}$ and the restriction of $\overline{p}m$ to $(\mathbb{S}, \mathcal{S})$ equals to p.
(b). Suppose $\{Z_n\}$ is an inhomogeneous Markov chain on $(\overline{\mathbb{S}}, \overline{\mathcal{S}})$ with transition probability $p_{2k} = m$ and $p_{2k+1} = \overline{p}$. Then, $\overline{X}_n = Z_{2n}$ is a Markov chain of transition probability $\overline{p}$ and $X_n = Z_{2n+1} \in \mathbb{S}$ is a Markov chain of transition probability p. Setting an initial measure $\overline{\nu}$ for $Z_0 = \overline{X}_0$ corresponds to having the initial measure $\nu(A) = \overline{\nu}(A) + \overline{\nu}(\{\boldsymbol{\alpha}\})q(A)$ for $X_0 \in \mathbb{S}$.
(c). $\mathbf{E}_\nu[f(X_n)] = \mathbf{E}_\nu[\overline{f}(\overline{X}_n)]$ for any $f \in b\mathcal{S}$, any initial distribution ν on $(\mathbb{S}, \mathcal{S})$ and all $n \geq 0$.

Proof. (a). Since $m(x, \{x\}) = 1$ it follows that $m\overline{p}(x, B) = \overline{p}(x, B)$ for all $x \in \mathbb{S}$ and $B \in \overline{\mathcal{S}}$. Further, $m(\boldsymbol{\alpha}, \cdot) = q(\cdot)$ so $m\overline{p}(\boldsymbol{\alpha}, B) = \int q(dy)\overline{p}(y, B)$ which by definition of $\overline{p}$ equals $\overline{p}(\boldsymbol{\alpha}, B)$ (see Definition 6.3.1). Similarly, if either $B = A \in \mathcal{S}$ or $B = A \cup \{\boldsymbol{\alpha}\}$, then by definition of the merge m and split $\overline{p}$ transition probabilities we have as claimed that for any $x \in \mathbb{S}$,

$$\overline{p}m(x, B) = \overline{p}(x, A) + \overline{p}(x, \{\boldsymbol{\alpha}\})q(A) = p(x, A)\,.$$

(b). As $m(\overline{x}, \{\boldsymbol{\alpha}\}) = 0$ for all $\overline{x} \in \overline{\mathbb{S}}$, this follows directly from part (a). Indeed, $Z_0 = \overline{X}_0$ of measure $\overline{\nu}$ is mapped by transition m to $X_0 = Z_1 \in \mathbb{S}$ of measure $\nu = \overline{\nu}m$, then by transition $\overline{p}$ to $\overline{X}_1 = Z_2$, followed by transition m to $X_1 = Z_3 \in \mathbb{S}$ and so on. Therefore, the transition probability between $\overline{X}_{n-1}$ and $\overline{X}_n$ is $m\overline{p} = \overline{p}$ and the one between X_{n-1} and X_n is $\overline{p}m$ restricted to $(\mathbb{S}, \mathcal{S})$, namely p.
(c). Constructing X_n and $\overline{X}_n$ as in part (b), if the initial distribution $\overline{\nu}$ of $\overline{X}_0$ assigns zero mass to $\boldsymbol{\alpha}$ then $\overline{\nu} = \nu$ with $X_0 = \overline{X}_0$. Further, by construction $\mathbf{E}_\nu[f(X_n)] = \mathbf{E}_{\overline{\nu}}[(mf)(\overline{X}_n)]$ which by definition of the split mapping is precisely $\mathbf{E}_\nu[\overline{f}(\overline{X}_n)]$, as claimed. □

We plan to study existence and attractiveness of invariant (probability) measures for the split chain $\{\overline{X}_n\}$, then apply Proposition 6.3.4 to transfer such results to the original chain $\{X_n\}$. This however requires the recurrence of the atom $\boldsymbol{\alpha}$. To this end, we must restrict the so called *small function* $v(x)$ of (6.3.1), motivating the next definition.

Definition 6.3.5. *A homogeneous Markov chain $\{X_n\}$ on $(\mathbb{S}, \mathcal{S})$ is called a* strong Doeblin chain *if the minorization condition (6.3.1) holds with a constant small function. That is, when $\inf_x p(x, A) \geq \delta q(A)$ for some probability measure q on $(\mathbb{S}, \mathcal{S})$, a positive constant $\delta > 0$ and all $A \in \mathcal{S}$. We call $\{X_n\}$ a* Doeblin chain *in case $Y_n = X_{rn}$ is a strong Doeblin chain for some finite r, namely when $\mathbf{P}_x(X_r \in A) \geq \delta q(A)$ for all $x \in \mathbb{S}$ and $A \in \mathcal{S}$.*

The Doeblin condition allows us to construct a split chain $\{\overline{Y}_n\}$ that visits its atom $\boldsymbol{\alpha}$ at each time step with probability $\eta \in (0, \delta)$. Considering part (c) of Exercise 6.1.18 (with $A = \mathbb{S}$), it follows that $\mathbf{P}_{\overline{\nu}}(\overline{Y}_n = \boldsymbol{\alpha}$ i.o.$) = 1$ for any initial distribution $\overline{\nu}$. So, in any Doeblin chain the atom $\boldsymbol{\alpha}$ is a recurrent state of the split chain. Further, since $T_{\boldsymbol{\alpha}} = \inf\{n \geq 1 : \overline{Y}_n = \boldsymbol{\alpha}\}$ is such that $\mathbf{P}_{\overline{x}}(T_{\boldsymbol{\alpha}} = 1) = \eta$

for all $\overline{x} \in \overline{\mathbb{S}}$, by the Markov property of $\overline{Y}_n$ (and Exercise 5.1.15), we deduce that $\mathbf{E}_{\overline{\nu}}[T_{\boldsymbol{\alpha}}] \leq 1/\eta$ is finite and uniformly bounded (in terms of the initial distribution $\overline{\nu}$). Consequently, the atom $\boldsymbol{\alpha}$ is a positive recurrent, aperiodic state of the split chain, which is accessible with probability one from each of its states.

As we see in Section 6.3.2, this is more than enough to assure that starting at any initial state, $\mathcal{P}_{Y_n}$ converges in total variation norm to the unique invariant probability measure for $\{Y_n\}$.

You are next going to examine which Markov chains of countable state space are Doeblin chains.

Exercise 6.3.6. *Suppose $\mathcal{S} = 2^{\mathbb{S}}$ with $\mathbb{S}$ a countable set.*

(a) *Show that a Markov chain of state space $(\mathbb{S}, \mathcal{S})$ is a Doeblin chain if and only if there exists $a \in \mathbb{S}$ and r finite such that $\inf_x \mathbf{P}_x(X_r = a) > 0$.*
(b) *Deduce that for any Doeblin chain $\mathbb{S} = \mathbb{T} \cup \mathbf{R}$, where $\mathbf{R} = \{y \in \mathbb{S} : \rho_{ay} > 0\}$ is a non-empty irreducible, closed set of positive recurrent, aperiodic states and $\mathbb{T} = \{y \in \mathbb{S} : \rho_{ay} = 0\}$ consists of transient states, all of which lead to $\mathbf{R}$.*
(c) *Verify that a Markov chain on a* finite *state space is a Doeblin chain if and only if it has an aperiodic state $a \in \mathbb{S}$ that is accessible from any other state.*
(d) *Check that branching processes with $0 < \mathbf{P}(N = 0) < 1$, renewal Markov chains and birth and death chains are never Doeblin chains.*

The preceding exercise shows that the Doeblin (recurrence) condition is too strong for many chains of interest. We thus replace it by the weaker H-irreducibility condition whereby the small function $v(x)$ is only assumed bounded below on a "small", accessible set C. To this end, we start with the definitions of an accessible set and weakly irreducible Markov chain.

Definition 6.3.7. *We say that $A \in \mathcal{S}$ is* accessible *by the Markov chain $\{X_n\}$ if $\mathbf{P}_x(T_A < \infty) > 0$ for all $x \in \mathbb{S}$.*

Given a non-zero σ-finite measure φ on $(\mathbb{S}, \mathcal{S})$, the chain is φ-irreducible *if any set $A \in \mathcal{S}$ with $\varphi(A) > 0$ is accessible by it. Finally, a homogeneous Markov chain on $(\mathbb{S}, \mathcal{S})$ is called* weakly irreducible *if it is φ-irreducible for some non-zero σ-finite measure φ (in particular, any Doeblin chain is weakly irreducible).*

Remark. Modern texts on Markov chains typically refer to the preceding as the standard definition of irreducibility but we use here the term *weak irreducibility* to clearly distinguish it from the elementary definition for a countable $\mathbb{S}$. Indeed, in case $\mathbb{S}$ is a countable set, let $\widetilde{\lambda}$ denote the corresponding counting measure of $\mathbb{S}$. A Markov chain of state space $\mathbb{S}$ is then $\widetilde{\lambda}$-irreducible if and only if $\rho_{xy} > 0$ for all $x, y \in \mathbb{S}$, matching our Definition 6.2.14 of irreducibility, whereas a chain on $\mathbb{S}$ countable is weakly irreducible if and only if $\rho_{xa} > 0$ for some $a \in \mathbb{S}$ and all $x \in \mathbb{S}$. In particular, a weakly irreducible chain of a countable state space $\mathbb{S}$ has exactly one non-empty equivalence class of intercommunicating states (i.e. $\{y \in \mathbb{S} : \rho_{ay} > 0\}$), which is further accessible by the chain.

As we show next, a weakly irreducible chain has a *maximal irreducibility measure* ψ such that $\psi(A) > 0$ if and only if $A \in \mathcal{S}$ is accessible by the chain.

Proposition 6.3.8. *Suppose $\{X_n\}$ is a weakly irreducible Markov chain on $(\mathbb{S}, \mathcal{S})$. Then, there exists a probability measure ψ on $(\mathbb{S}, \mathcal{S})$ such that for any $A \in \mathcal{S}$,*

$$\psi(A) > 0 \quad \Longleftrightarrow \quad \mathbf{P}_x(T_A < \infty) > 0 \quad \forall x \in \mathbb{S} . \tag{6.3.2}$$

We call such ψ a maximal irreducibility measure *for the chain.*

Remark. Clearly, if a chain is φ-irreducible, then any non-zero σ-finite measure absolutely continuous with respect to φ (per Definition 4.1.4), is also an irreducibility measure for this chain. The converse holds in case of a maximal irreducibility measure. That is, unraveling Definition 6.3.7 it follows from (6.3.2) that $\{X_n\}$ is φ-irreducible if and only if the non-zero σ-finite measure φ is absolutely continuous with respect to ψ.

Proof. Let ν be a non-zero σ-finite irreducibility measure of the given weakly irreducible chain $\{X_n\}$. Taking $D \in \mathcal{S}$ such that $\nu(D) \in (0, \infty)$ we see that $\{X_n\}$ is also q-irreducible for the probability measure $q(\cdot) = \nu(\cdot \cap D)/\nu(D)$. We claim that (6.3.2) holds for the probability measure $\psi(A) = \int_{\mathbb{S}} q(dx) k(x, A)$ on $(\mathbb{S}, \mathcal{S})$, where

$$k(x, A) = \sum_{n=1}^{\infty} 2^{-n} \mathbf{P}_x(X_n \in A) . \tag{6.3.3}$$

Indeed, with $\{T_A < \infty\} = \cup_{n \geq 1} \{X_n \in A\}$, clearly $\mathbf{P}_x(T_A < \infty) > 0$ if and only if $k(x, A) > 0$. Consequently, if $\mathbf{P}_x(T_A < \infty)$ is positive for all $x \in \mathbb{S}$ then so is $k(x, A)$ and hence $\psi(A) > 0$. Conversely, if $\psi(A) > 0$ then necessarily $q(C) > 0$ for $C = \{x \in \mathbb{S} : k(x, A) \geq \eta\}$ and some $\eta > 0$ small enough. In particular, fixing $x \in \mathbb{S}$, as $\{X_n\}$ is q-irreducible, also $\mathbf{P}_x(T_C < \infty) > 0$. That is, there exists positive integer $m = m(x)$ such that $\mathbf{P}_x(X_m \in C) > 0$. It now follows by the Markov property at m (for $h(\omega) = \sum_{\ell \geq 1} 2^{-\ell} I_{\omega_\ell \in A}$), that

$$\begin{aligned} k(x, A) &\geq 2^{-m} \sum_{\ell=1}^{\infty} 2^{-\ell} \mathbf{P}_x(X_{m+\ell} \in A) \\ &= 2^{-m} \mathbf{E}_x[k(X_m, A)] \geq 2^{-m} \mathbf{P}_x(X_m \in C) \eta > 0 . \end{aligned}$$

Since this is equivalent to $\mathbf{P}_x(T_A < \infty) > 0$ and applies for all $x \in \mathbb{S}$, we have established the identity (6.3.2). $\square$

We next define the notions of a *small set* and an *H-irreducible* chain.

Definition 6.3.9. *An accessible set $C \in \mathcal{S}$ of a Markov chain $\{X_n\}$ on $(\mathbb{S}, \mathcal{S})$ is called* r-small set *if the transition probability $(x, A) \mapsto \mathbf{P}_x(X_r \in A)$ satisfies the minorization condition (6.3.1) with a small function that is constant and positive on C. That is, when $\mathbf{P}_x(X_r \in \cdot) \geq \delta I_C(x) q(\cdot)$ for some positive constant $\delta > 0$ and probability measure q on $(\mathbb{S}, \mathcal{S})$.*

We further use small set *for 1-small set, and call the chain* H-irreducible *if it has an r-small set for some finite $r \geq 1$ and* strong H-irreducible *in case $r = 1$.*

Clearly, a chain is Doeblin if and only if $\mathbb{S}$ is an r-small set for some $r \geq 1$, and is further strong Doeblin in case $r = 1$. In particular, a Doeblin chain is H-irreducible and a strong Doeblin chain is also strong H-irreducible.

Exercise 6.3.10. *Prove the following properties of H-irreducible chains.*

(a) *Show that an H-irreducible chain is q-irreducible, hence weakly irreducible.*

(b) *Show that if $\{X_n\}$ is strong H-irreducible then the atom $\boldsymbol{\alpha}$ of the split chain $\{\overline{X}_n\}$ is accessible by $\{\overline{X}_n\}$ from all states in $\overline{\mathbb{S}}$.*

(c) *Show that in a countable state space every weakly irreducible chain is strong H-irreducible.*
Hint: *Try $C = \{a\}$ and $q(\cdot) = p(a, \cdot)$ for some $a \in \mathbb{S}$.*

Actually, the converse to part (a) of Exercise 6.3.10 holds as well. That is, weak irreducibility is equivalent to H-irreducibility (for the proof, see [**Num84**, Proposition 2.6]), and weakly irreducible chains can be analyzed via the study of an appropriate split chain. For simplicity we focus hereafter on the somewhat more restricted setting of strong H-irreducible chains. The following example shows that it still applies for many Markov chains of interest.

Example 6.3.11 (Continuous transition densities)**.** *Let $\mathbb{S} = \mathbb{R}^d$ with $\mathcal{S} = \mathcal{B}_{\mathbb{S}}$. Suppose that for each $\underline{x} \in \mathbb{R}^d$ the transition probability has a density $p(\underline{x}, \underline{y})$ with respect to Lebesgue measure $\lambda^d(\cdot)$ on $\mathbb{R}^d$ such that $(\underline{x}, \underline{y}) \mapsto p(\underline{x}, \underline{y})$ is continuous jointly in $\underline{x}$ and $\underline{y}$. Picking $\underline{u}$ and $\underline{v}$ such that $p(\underline{u}, \underline{v}) > 0$, there exists a neighborhood C of $\underline{u}$ and a bounded neighborhood K of $\underline{v}$, such that $\inf\{p(\underline{x}, \underline{y}) : \underline{x} \in C, \underline{y} \in K\} > 0$. Hence, setting $q(\cdot)$ to be the uniform measure on K (i.e. $q(A) = \lambda^d(A \cap K)/\lambda^d(K)$ for any $A \in \mathcal{S}$), such a chain is strong H-irreducible provided C is an accessible set. For example, this occurs whenever $p(\underline{x}, \underline{u}) > 0$ for all $\underline{x} \in \mathbb{R}^d$.*

Remark 6.3.12. Though our study of Markov chains has been mostly concerned with measure theoretic properties of $(\mathbb{S}, \mathcal{S})$ (e.g. being $\mathcal{B}$-isomorphic), quite often $\mathbb{S}$ is actually a topological state space with $\mathcal{S}$ its Borel σ-algebra. As seen in the preceding example, continuity properties of the transition probability are then of much relevance in the study of Markov chains on $\mathbb{S}$. In this context, we say that $p : \mathbb{S} \times \mathcal{B}_{\mathbb{S}} \mapsto [0, 1]$ is a *strong Feller* transition probability, when the linear operator $(ph)(\cdot) = \int p(\cdot, dy)h(y)$ of Lemma 6.1.3 maps every bounded $\mathcal{B}_{\mathbb{S}}$-measurable function h to $ph \in C_b(\mathbb{S})$, a continuous bounded function on $\mathbb{S}$. In case of continuous transition densities, as in Example 6.3.11, the transition probability is strong Feller whenever the collection of probability measures $\{p(x, \cdot), x \in \mathbb{S}\}$ is uniformly tight (per Definition 3.2.31).

In case $\mathcal{S} = \mathcal{B}_{\mathbb{S}}$ we further have the following topological notions of reachability and irreducibility.

Definition 6.3.13. *Suppose $\{X_n\}$ is a Markov chain on a topological space $\mathbb{S}$ equipped with its Borel σ-algebra $\mathcal{S} = \mathcal{B}_{\mathbb{S}}$. We call $x \in \mathbb{S}$ a* reachable state *of $\{X_n\}$ if any neighborhood of x is accessible by this chain and call the chain* O-irreducible *(or open set irreducible), if every $x \in \mathbb{S}$ is reachable, that is, every open set is accessible by $\{X_n\}$.*

Remark. Equipping a countable state space $\mathbb{S}$ with its discrete topology yields the Borel σ-algebra $\mathcal{S} = 2^{\mathbb{S}}$, in which case O-irreducibility is equivalent to our earlier Definition 6.2.14 of irreducibility.

For more general topological state spaces (such as $\mathbb{S} = \mathbb{R}^d$), by their definitions, a weakly irreducible chain is O-irreducible if and only if its maximal irreducibility measure ψ is such that $\psi(O) > 0$ for any open subset O of $\mathbb{S}$. Conversely,

Exercise 6.3.14. *Show that if a strong Feller transition probability $p(\cdot, \cdot)$ has a reachable state $x \in \mathbb{S}$, then it is weakly irreducible.*
Hint: *Try the irreducibility measure $\varphi(\cdot) = p(x, \cdot)$.*

Remark. The minorization (6.3.1) may cause the maximal irreducibility measure for the split chain to be supported on a smaller subset of the state space than the one for the original chain. For example, consider the trivial Doeblin chain of i.i.d. $\{X_n\}$, that is, $p(x, \cdot) = q(\cdot)$. In this case, taking $v(x) = 1$ results with the split chain $\overline{X}_n = \boldsymbol{\alpha}$ for all $n \geq 1$, so the maximal irreducibility measures $\overline{\psi} = \delta_{\boldsymbol{\alpha}}$ and $\psi = q$ of $\{X_n\}$ and $\{\overline{X}_n\}$ are then mutually singular.

This is of course precluded by our additional requirement that $v(x)q(\cdot) \ll \widehat{p}(x, \cdot)$. For a strong H-irreducible chain $\{X_n\}$ it is easily accommodated by, for example, setting $v(x) = \eta I_C(x)$ with $\eta = \delta/2 > 0$, and then the restriction of $\overline{\psi}$ to $\mathcal{S}$ is a maximal irreducibility measure for $\{X_n\}$.

Strong H-irreducible chains with a recurrent atom are called *H-recurrent* chains. That is,

Definition 6.3.15. *A strong H-irreducible chain $\{X_n\}$ is called* H-recurrent *if $\mathbf{P}_{\boldsymbol{\alpha}}(T_{\boldsymbol{\alpha}} < \infty) = 1$. By the strong Markov property of $\overline{X}_n$ at the consecutive visit times $T^k_{\boldsymbol{\alpha}}$ of $\boldsymbol{\alpha}$, H-recurrence further implies that $\mathbf{P}_{\boldsymbol{\alpha}}(T^k_{\boldsymbol{\alpha}}$ finite for all $k) = 1$, or equivalently $\mathbf{P}_{\boldsymbol{\alpha}}(\overline{X}_n = \boldsymbol{\alpha}$ i.o.$) = 1$.*

Here are a few examples and exercises to clarify the concept of H-recurrence.

Example 6.3.16. *Many strong H-irreducible chains are not H-recurrent. For example, combining part (c) of Exercise 6.3.10 with the remark following Definition 6.3.7 we see that such are all irreducible transient chains on a countable state space.*

By the same reasoning, a Markov chain of countable state space $\mathbb{S}$ is H-recurrent if and only if $\mathbb{S} = \mathbb{T} \cup \mathbf{R}$ with $\mathbf{R}$ a non-empty irreducible, closed set of recurrent states and $\mathbb{T}$ a collection of transient states that lead to $\mathbf{R}$ (c.f. part (b) of Exercise 6.3.6 for such a decomposition in case of Doeblin chains). In particular, such chains are not necessarily recurrent *in the sense of Definition 6.2.14. For example, the chain on $\mathbb{S} = \{1, 2, \ldots\}$ with transitions $p(k, 1) = 1 - p(k, k+1) = k^{-s}$ for some constant $s > 0$, is H-recurrent but has only one recurrent state, i.e. $\mathbf{R} = \{1\}$. Further, $\rho_{k1} < 1$ for all $k \neq 1$ when $s > 1$, while $\rho_{k1} = 1$ for all k when $s \leq 1$.*

Remark. Advanced texts on Markov chains refer to what we call H-recurrence as the standard definition of recurrence and call such chains *Harris recurrent* when in addition $\mathbf{P}_x(T_{\boldsymbol{\alpha}} < \infty) = 1$ for all $x \in \mathbb{S}$. As seen in the preceding example, both notions are weaker than the elementary notion of recurrence for countable $\mathbb{S}$, per Definition 6.2.14. For this reason, we adopt here the convention of calling H-recurrence (with H after Harris), what is not the usual definition of Harris recurrence.

Exercise 6.3.17. *Verify that any strong Doeblin chain is also H-recurrent. Conversely show that for any H-recurrent chain $\{X_n\}$ there exists $C \in \mathcal{S}$ and a probability distribution q on $(\mathbb{S}, \mathcal{S})$ such that $\mathbf{P}_q(T^k_C$ finite for all $k) = 1$ and the Markov chain $Z_k = X_{T_C^{k+1}}$ for $k \geq 0$ is then a strong Doeblin chain.*

The next proposition shows that similarly to the elementary notion of recurrence, H-recurrence is transferred from the atom $\boldsymbol{\alpha}$ to all sets that are accessible from it. Building on this proposition, you show in Exercise 6.3.19 that the same applies when starting at any irreducibility probability measure of the split chain and that every set in $\overline{\mathcal{S}}$ is either almost surely visited or almost surely never reached from $\boldsymbol{\alpha}$ by the split chain.

Proposition 6.3.18. *For an H-recurrent chain $\{X_n\}$ consider the probability measure*

$$\overline{\psi}(B) = \sum_{n=1}^{\infty} 2^{-n} \mathbf{P}_{\boldsymbol{\alpha}}(\overline{X}_n \in B)\,. \tag{6.3.4}$$

Then, $\mathbf{P}_{\boldsymbol{\alpha}}(\overline{X}_n \in B \text{ i.o.}) = 1$ whenever $\overline{\psi}(B) > 0$.

Proof. Clearly, $\overline{\psi}(B) > 0$ if and only if $\mathbf{P}_{\boldsymbol{\alpha}}(T_B < \infty) > 0$. Further, if $\eta = \mathbf{P}_{\boldsymbol{\alpha}}(T_B < \infty) > 0$, then considering the split chain starting at $\overline{X}_0 = \boldsymbol{\alpha}$, we have from part (c) of Exercise 6.1.18 that

$$\mathbf{P}_{\boldsymbol{\alpha}}(\{\overline{X}_n = \boldsymbol{\alpha} \text{ finitely often }\} \cup \{\overline{X}_n \in B \text{ i.o.}\}) = 1\ .$$

As $\mathbf{P}_{\boldsymbol{\alpha}}(\overline{X}_n = \boldsymbol{\alpha} \text{ i.o.}) = 1$ by the assumed H-recurrence, our thesis that $\mathbf{P}_{\boldsymbol{\alpha}}(\overline{X}_n \in B \text{ i.o.}) = 1$ follows. □

Exercise 6.3.19. *Suppose $\overline{\psi}$ is the probability measure of (6.3.4) for an H-recurrent chain.*

(a) *Argue that $\{\boldsymbol{\alpha}\}$ is accessible by the split chain $\{\overline{X}_n\}$ and show that $\overline{\psi}$ is a maximal irreducibility measure for it.*
(b) *Show that $\mathbf{P}_{\overline{\psi}}(D) = \mathbf{P}_{\boldsymbol{\alpha}}(D)$ for any shift invariant $D \in \overline{\mathcal{S}}_c$ (i.e. where $D = \theta^{-1}D$).*
(c) *In case $B \in \overline{\mathcal{S}}$ is such that $\overline{\psi}(B) > 0$ explain why $\mathbf{P}_{\overline{x}}(\overline{X}_n \in B \text{ i.o.}) = 1$ for $\overline{\psi}$-a.e. $\overline{x} \in \overline{\mathbb{S}}$ and $\mathbf{P}_{\overline{\nu}}(\overline{X}_n \in B \text{ i.o.}) = 1$ for any probability measure $\overline{\nu} \ll \overline{\psi}$.*
(d) *Show that $\mathbf{P}_{\boldsymbol{\alpha}}(T_B < \infty) \in \{0, 1\}$ for all $B \in \overline{\mathcal{S}}$.*

Given a strong H-irreducible chain $\{X_n\}$ there is no unique way to select the small set C, regeneration measure $q(\cdot)$ and $\delta > 0$ such that $p(x, \cdot) \geq \delta I_C(x) q(\cdot)$. Consequently, there are many different split chains for each chain $\{X_n\}$. Nevertheless, as you show next, H-recurrence is determined by the original chain $\{X_n\}$.

Exercise 6.3.20. *Suppose $\{\overline{X}_n\}$ and $\overline{X}'_n\}$ are two different split chains for the* same *strong H-irreducible chain $\{X_n\}$ with the corresponding atoms $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}'$. Relying on Proposition 6.3.18 prove that $\mathbf{P}_{\boldsymbol{\alpha}}(T_{\boldsymbol{\alpha}} < \infty) = 1$ if and only if $\mathbf{P}_{\boldsymbol{\alpha}'}(T'_{\boldsymbol{\alpha}'} < \infty) = 1$.*

The concept of H-recurrence builds on measure theoretic properties of the chain, namely the minorization associated with strong H-irreducibility. In contrast, for topological state space we have the following topological concept of O-recurrence, built on reachability of states.

Definition 6.3.21. *A state x of a Markov chain $\{X_n\}$ on (topological) state space $(\mathbb{S}, \mathcal{B}_{\mathbb{S}})$ is called* O-recurrent *(or open set recurrent), if $\mathbf{P}_x(X_n \in O \text{ i.o.}) = 1$ for any neighborhood O of x in $\mathbb{S}$. All states $x \in \mathbb{S}$ which are not O-recurrent are called* O-transient. *Such a chain is then called* O-recurrent *if every $x \in \mathbb{S}$ is O-recurrent and* O-transient *if every $x \in \mathbb{S}$ is O-transient.*

Remark. As was the case with O-irreducibility versus irreducibility, for a countable state space $\mathbb{S}$ equipped with its discrete topology, being O-recurrent (or O-transient), is equivalent to being recurrent (or transient, respectively), per Definitions 6.2.9 and 6.2.14.

The concept of O-recurrence is in particular suitable for the study of random walks. Indeed,

Exercise 6.3.22. *Suppose $S_n = S_0 + \sum_{k=1}^n \xi_k$ is a random walk on $\mathbb{R}^d$.*

(a) *Show that if $\{S_n\}$ has one reachable state, then it is O-irreducible.*
(b) *Show that either $\{S_n\}$ is an O-recurrent chain or it is an O-transient chain.*
(c) *Show that if $\{S_n\}$ is O-recurrent, then*

$$\mathbb{S} = \{x \in \mathbb{R}^d : \mathbf{P}_x(\|X_n\| < r \text{ i.o. }) > 0, \text{ for all } r > 0\},$$

is a closed sub-group of $\mathbb{R}^d$ (i.e. $\underline{0} \in \mathbb{S}$ and if $x, y \in \mathbb{S}$ then also $x - y \in \mathbb{S}$), with respect to which $\{S_n\}$ is O-irreducible (i.e. $\mathbf{P}_y(T_{B(x,r)} < \infty) > 0$ for all $r > 0$ and $x, y \in \mathbb{S}$).

In case of one-dimensional random walks, you are to recover next the Chung-Fuchs theorem, stating that if $n^{-1}S_n$ converges to zero in probability, then this Markov chain is O-recurrent.

Exercise 6.3.23 (Chung-Fuchs theorem)**.** *Suppose $\{S_n\}$ is a random walk on $\mathbb{S} \subseteq \mathbb{R}$.*

(a) *Show that such random walk is O-recurrent if and only if for each $r > 0$,*

$$\sum_{n=0}^{\infty} \mathbf{P}_0(|S_n| < r) = \infty .$$

(b) *Show that for any $r > 0$ and $k \in \mathbb{Z}$,*

$$\sum_{n=0}^{\infty} \mathbf{P}_0(S_n \in [kr, (k+1)r)) \le \sum_{m=0}^{\infty} \mathbf{P}_0(|S_m| < r),$$

and deduce that suffices to check divergence of the series in part (a) for large r.
(c) *Conclude that if $n^{-1}S_n \xrightarrow{p} 0$ as $n \to \infty$, then $\{S_n\}$ is O-recurrent.*

6.3.2. Invariant measures, aperiodicity and asymptotic behavior. We consider hereafter an H-recurrent Markov chain $\{X_n\}$ of transition probability $p(\cdot, \cdot)$ on the $\mathcal{B}$-isomorphic state space $(\mathbb{S}, \mathcal{S})$ with its recurrent pseudo-atom $\boldsymbol{\alpha}$ and the corresponding split and merge chains $\overline{p}(\cdot, \cdot)$, $m(\cdot, \cdot)$ on $(\overline{\mathbb{S}}, \overline{\mathcal{S}})$ per Definitions 6.3.1 and 6.3.2.

The following lemma characterizes the invariant measures of the split chain $\overline{p}(\cdot, \cdot)$ and their relation to the invariant measures for $p(\cdot, \cdot)$. To this end, we use hereafter $\nu_1\nu_2$ also for the measure $\nu_1\nu_2(A) = \nu_1(\nu_2(\cdot, A))$ on $(\mathbb{X}, \mathcal{X})$ in case ν_1 is a measure on $(\mathbb{X}, \mathcal{X})$ and ν_2 is a transition probability on this space and let $\overline{p}^n(\overline{x}, B)$ denote the transition probability $\mathbf{P}_{\overline{x}}(\overline{X}_n \in B)$ on $(\overline{\mathbb{S}}, \overline{\mathcal{S}})$.

Lemma 6.3.24. *A measure $\bar{\mu}$ on $(\overline{\mathbb{S}}, \overline{\mathcal{S}})$ is invariant for the split chain $\overline{p}(\cdot, \cdot)$ of a strong H-irreducible chain if and only if $\bar{\mu} = \bar{\mu}\overline{p}$ and $0 < \bar{\mu}(\{\boldsymbol{\alpha}\}) < \infty$. Further, $\bar{\mu}m$ is then an invariant measure for the original chain $p(\cdot, \cdot)$. Conversely, if μ is an invariant measure for $p(\cdot, \cdot)$ then the measure $\mu\overline{p}$ is invariant for the split chain.*

Proof. Recall Proposition 6.1.23 that a measure $\bar{\mu}$ is invariant for the split chain if and only if $\bar{\mu}$ is positive, σ-finite and

$$\bar{\mu}(B) = \bar{\mu} \otimes \overline{p}(\overline{\mathbb{S}} \times B) = \bar{\mu}\overline{p}(B) \qquad \forall B \in \overline{\mathcal{S}} .$$

Likewise, a measure μ is invariant for p if and only if μ is positive, σ-finite and $\mu(A) = \mu p(A)$ for all $A \in \mathcal{S}$.

We first show that if $\bar{\mu}$ is invariant for $\overline{p}$ then $\mu = \bar{\mu} m$ is invariant for p. Indeed, note that from Definition 6.3.2 it follows that

$$\mu(A) = \bar{\mu}(A) + \bar{\mu}(\{\boldsymbol{\alpha}\}) q(A) \qquad \forall A \in \mathcal{S} \tag{6.3.5}$$

and in particular, such μ is a positive, σ-finite measure on $(\mathbb{S}, \mathcal{S})$ for any σ-finite $\bar{\mu}$ on $(\overline{\mathbb{S}}, \overline{\mathcal{S}})$, and any probability measure $q(\cdot)$ on $(\mathbb{S}, \mathcal{S})$. Further, starting the inhomogeneous Markov chain $\{Z_n\}$ of Proposition 6.3.4 with initial measure $\bar{\mu}$ for $Z_0 = \overline{X}_0$ yields the measure μ for $Z_1 = X_0$. By construction, the measure of $Z_2 = \overline{X}_1$ is then $\mu \overline{p}$ and that of $Z_3 = X_1$ is $(\mu \overline{p}) m = \mu(\overline{p} m)$. Next, the invariance of $\bar{\mu}$ for $\overline{p}$ implies that the measure of $\overline{X}_1$ equals that of $\overline{X}_0$. Consequently, the measure of X_1 must equal that of X_0, namely $\mu = \mu(\overline{p} m)$. With $m(\cdot, \{\boldsymbol{\alpha}\}) \equiv 0$ necessarily $\mu(\{\boldsymbol{\alpha}\}) = 0$ and the identity $\mu = \mu(\overline{p} m)$ holds also for the restrictions to $(\mathbb{S}, \mathcal{S})$ of both μ and $\overline{p} m$. Since the latter equals to p (see part (a) of Proposition 6.3.4), we conclude that $\mu = \mu p$, as claimed.

Conversely, let $\bar{\mu} = \mu \overline{p}$ where μ is an invariant measure for p (and we set $\mu(\{\boldsymbol{\alpha}\}) = 0$). Since μ is σ-finite, there exist $A_n \uparrow \mathbb{S}$ such that $\mu(A_n) < \infty$ for all n and necessarily also $q(A_n) > 0$ for all n large enough (by monotonicity from below of the probability measure $q(\cdot)$). Further, the invariance of μ implies that $\bar{\mu} m = (\mu \overline{p}) m = \mu$, i.e. the relation (6.3.5) holds. In particular, $\bar{\mu}(\overline{\mathbb{S}}) = \mu(\mathbb{S})$ so $\bar{\mu}$ inherits the positivity of μ. Moreover, both $\bar{\mu}(\{\boldsymbol{\alpha}\}) = \infty$ and $\bar{\mu}(A_n) = \infty$ contradict the finiteness of $\mu(A_n)$ for all n, so the measure $\bar{\mu}$ is σ-finite on $(\overline{\mathbb{S}}, \overline{\mathcal{S}})$. Next, start the chain $\{Z_n\}$ at $Z_0 = \overline{X}_0 \in \mathbb{S}$ of initial measure μ. It yields the same measure $\mu = \mu m$ for $Z_1 = X_0$, with measure $\bar{\mu} = \mu \overline{p}$ for $Z_2 = \overline{X}_1$ followed by $\bar{\mu} m = \mu$ for $Z_3 = X_1$ and $\bar{\mu} \overline{p}$ for $Z_4 = \overline{X}_2$. As the measure of X_1 equals that of X_0, it follows that the measure $\bar{\mu} \overline{p}$ of $\overline{X}_2$ equals the measure $\bar{\mu}$ of $\overline{X}_1$, i.e. $\bar{\mu}$ is invariant for $\overline{p}$.

Finally, suppose the measure $\bar{\mu}$ satisfies $\bar{\mu} = \bar{\mu} \overline{p}$. Iterating this identity we deduce that $\bar{\mu} = \bar{\mu} \overline{p}^n$ for all $n \geq 1$, hence also $\bar{\mu} = \bar{\mu} k$ for the transition probability

$$k(\overline{x}, B) = \sum_{n=1}^{\infty} 2^{-n} \, \overline{p}^n \, (\overline{x}, B) \, . \tag{6.3.6}$$

Due to its strong H-irreducibility, the atom $\{\boldsymbol{\alpha}\}$ of the split chain is an accessible set for the transition probability $\overline{p}$ (see part (b) of Exercise 6.3.10). So, from (6.3.6) we deduce that $k(\overline{x}, \{\boldsymbol{\alpha}\}) > 0$ for all $\overline{x} \in \overline{\mathbb{S}}$. Consequently, as $n \uparrow \infty$,

$$B_n = \{\overline{x} \in \overline{\mathbb{S}} \, : \, k(\overline{x}, \{\boldsymbol{\alpha}\}) \geq n^{-1}\} \uparrow \overline{\mathbb{S}} \, ,$$

whereas by the identity $\bar{\mu}(\{\boldsymbol{\alpha}\}) = (\bar{\mu} k)(\{\boldsymbol{\alpha}\})$ also $\bar{\mu}(\{\boldsymbol{\alpha}\}) \geq n^{-1} \bar{\mu}(B_n)$. This proves the first claim of the lemma. Indeed, we have just shown that when $\bar{\mu} = \bar{\mu} \overline{p}$ it follows that $\bar{\mu}$ is positive if and only if $\bar{\mu}(\{\boldsymbol{\alpha}\}) > 0$ and σ-finite if and only if $\bar{\mu}(\{\boldsymbol{\alpha}\}) < \infty$. $\square$

Our next result shows that, similarly to Proposition 6.2.27, the recurrent atom $\boldsymbol{\alpha}$ induces an invariant measure for the split chain (and hence also one for the original chain).

Proposition 6.3.25. *If $\{X_n\}$ is H-recurrent of transition probability $p(\cdot, \cdot)$ then*

$$\bar{\nu}_{\boldsymbol{\alpha}}(B) = \mathbf{E}_{\boldsymbol{\alpha}} \Big(\sum_{n=0}^{T_{\boldsymbol{\alpha}}-1} I_{\{\overline{X}_n \in B\}} \Big) \tag{6.3.7}$$

is an invariant measure for $\overline{p}(\cdot,\cdot)$.

PROOF. Let $\bar{\nu}_{\boldsymbol{\alpha},n}(B) = \mathbf{P}_{\boldsymbol{\alpha}}(\overline{X}_n \in B, T_{\boldsymbol{\alpha}} > n)$, noting that

$$\bar{\nu}_{\boldsymbol{\alpha}}(B) = \sum_{n=0}^{\infty} \bar{\nu}_{\boldsymbol{\alpha},n}(B) \qquad \forall B \in \overline{\mathcal{S}} \tag{6.3.8}$$

and $\bar{\nu}_{\boldsymbol{\alpha},n}(g) = \mathbf{E}_{\boldsymbol{\alpha}}[I_{\{T_{\boldsymbol{\alpha}}>n\}} g(\overline{X}_n)]$ for all $g \in b\overline{\mathcal{S}}$. Since $\{T_{\boldsymbol{\alpha}} > n\} \in \mathcal{F}_n^{\overline{\mathbf{X}}} = \sigma(\overline{X}_k, k \le n)$, we have by the tower and Markov properties that, for each $n \ge 0$,

$$\begin{aligned}\mathbf{P}_{\boldsymbol{\alpha}}(\overline{X}_{n+1} \in B, T_{\boldsymbol{\alpha}} > n) &= \mathbf{E}_{\boldsymbol{\alpha}}[I_{\{T_{\boldsymbol{\alpha}}>n\}} \mathbf{P}_{\boldsymbol{\alpha}}(\overline{X}_{n+1} \in B | \mathcal{F}_n^{\overline{\mathbf{X}}})\,] \\ &= \mathbf{E}_{\boldsymbol{\alpha}}[I_{\{T_{\boldsymbol{\alpha}}>n\}} \bar{p}(\overline{X}_n, B)\,] = \bar{\nu}_{\boldsymbol{\alpha},n}(\overline{p}(\cdot, B)) = (\bar{\nu}_{\boldsymbol{\alpha},n}\overline{p})(B)\,.\end{aligned}$$

Hence,

$$\begin{aligned}(\bar{\nu}_{\boldsymbol{\alpha}}\overline{p})(B) &= \sum_{n=0}^{\infty} (\bar{\nu}_{\boldsymbol{\alpha},n}\overline{p})(B) = \sum_{n=0}^{\infty} \mathbf{P}_{\boldsymbol{\alpha}}(\overline{X}_{n+1} \in B, T_{\boldsymbol{\alpha}} > n) \\ &= \mathbf{E}_{\boldsymbol{\alpha}}\Big(\sum_{n=1}^{T_{\boldsymbol{\alpha}}} I_{\{\overline{X}_n \in B\}}\Big) = \bar{\nu}_{\boldsymbol{\alpha}}(B)\end{aligned}$$

since $\mathbf{P}_{\boldsymbol{\alpha}}(T_{\boldsymbol{\alpha}} < \infty, \overline{X}_0 = \overline{X}_{T_{\boldsymbol{\alpha}}}) = 1$. We thus established that $\bar{\nu}_{\boldsymbol{\alpha}}\overline{p} = \bar{\nu}_{\boldsymbol{\alpha}}$ and as $\bar{\nu}_{\boldsymbol{\alpha}}(\{\boldsymbol{\alpha}\}) = 1$ we conclude from Lemma 6.3.24 that it is an invariant measure for the split chain. □

Building on Lemma 6.3.24 and Proposition 6.3.25 we proceed to the uniqueness of the invariant measure for an H-recurrent chain, namely the extension of Proposition 6.2.30 to a typically uncountable state space.

Theorem 6.3.26. *Up to a constant multiple, the unique invariant measure for H-recurrent transition probability* $p(\cdot,\cdot)$ *is the restriction to* $(\mathbb{S}, \mathcal{S})$ *of* $\bar{\nu}_{\boldsymbol{\alpha}} m$*, where* $\bar{\nu}_{\boldsymbol{\alpha}}$ *is per (6.3.7).*

Remark. As $\bar{\nu}_{\boldsymbol{\alpha}}(\overline{\mathbb{S}}) = \mathbf{E}_{\boldsymbol{\alpha}} T_{\boldsymbol{\alpha}}$, it follows from the theorem that an H-recurrent chain has an *invariant probability measure* if and only if $\mathbf{E}_{\boldsymbol{\alpha}}(T_{\boldsymbol{\alpha}}) = \mathbf{E}_q(T_{\boldsymbol{\alpha}}) < \infty$. In accordance with Definition 6.2.40 we call such chains *positive H-recurrent*. While the value of $\mathbf{E}_{\boldsymbol{\alpha}}(T_{\boldsymbol{\alpha}})$ depends on the specific split chain one associates with $\{X_n\}$, it follows from the preceding that positive H-recurrence, i.e. the finiteness of $\mathbf{E}_{\boldsymbol{\alpha}}(T_{\boldsymbol{\alpha}})$, is determined by the original chain. Further, in view of the relation (6.3.5) between $\bar{\nu}_{\boldsymbol{\alpha}} m$ and $\bar{\nu}_{\boldsymbol{\alpha}}$ and the decomposition (6.3.8) of $\bar{\nu}_{\boldsymbol{\alpha}}$, the unique invariant probability measure for $\{X_n\}$ is then

$$\pi(A) = \frac{1}{\mathbf{E}_q(T_{\boldsymbol{\alpha}})} \sum_{n=0}^{\infty} \mathbf{P}_q(\overline{X}_n \in A, T_{\boldsymbol{\alpha}} > n) \qquad \forall A \in \mathcal{S}\,. \tag{6.3.9}$$

PROOF. By Lemma 6.3.24, to any invariant measure μ for p (with $\mu(\{\boldsymbol{\alpha}\}) = 0$), corresponds the invariant measure $\bar{\mu} = \mu\overline{p}$ for the split chain $\overline{p}$. It is also shown there that $0 < \bar{\mu}(\{\boldsymbol{\alpha}\}) < \infty$. Hence, with no loss of generality we assume hereafter that the given invariant measure μ for p has already been divided by this positive, finite constant, and so $\bar{\mu}(\{\boldsymbol{\alpha}\}) = 1$. Recall that while proving Lemma 6.3.24 we further noted that $\mu = \bar{\mu}m$, due to the invariance of μ for p. Consequently, to prove the theorem it suffices to show that $\bar{\mu} = \bar{\nu}_{\boldsymbol{\alpha}}$ (for then $\mu = \bar{\mu}m = \bar{\nu}_{\boldsymbol{\alpha}} m$).

To this end, fix $B \in \overline{\mathcal{S}}$ and recall from the proof of Lemma 6.3.24 that $\bar{\mu}$ is also invariant for $\overline{p}^n$ and any $n \geq 1$. Using the latter invariance property and applying Exercise 6.2.3 for $y = \boldsymbol{\alpha}$ and the split chain $\{\overline{X}_n\}$, we find that

$$\begin{aligned}\bar{\mu}(B) = (\bar{\mu}\overline{p}^n)(B) &= \int_{\overline{\mathbb{S}}} \bar{\mu}(d\overline{x})\mathbf{P}_{\overline{x}}(\overline{X}_n \in B) \geq \int_{\overline{\mathbb{S}}} \bar{\mu}(d\overline{x})\mathbf{P}_{\overline{x}}(\overline{X}_n \in B, T_{\boldsymbol{\alpha}} \leq n) \\ &= \sum_{k=0}^{n-1} (\bar{\mu}\overline{p}^{n-k})(\{\boldsymbol{\alpha}\})\mathbf{P}_{\boldsymbol{\alpha}}(\overline{X}_k \in B, T_{\boldsymbol{\alpha}} > k) = \sum_{k=0}^{n-1} \bar{\nu}_{\boldsymbol{\alpha},k}(B)\,,\end{aligned}$$

with $\bar{\nu}_{\boldsymbol{\alpha},k}(\cdot)$ per the decomposition (6.3.8) of $\bar{\nu}_{\boldsymbol{\alpha}}(\cdot)$. Taking $n \to \infty$, we thus deduce that

$$\bar{\mu}(B) \geq \sum_{k=0}^{\infty} \bar{\nu}_{\boldsymbol{\alpha},k}(B) = \bar{\nu}_{\boldsymbol{\alpha}}(B) \qquad \forall B \in \overline{\mathcal{S}}\,. \tag{6.3.10}$$

We proceed to show that this inequality actually holds with equality, namely, that $\bar{\mu} = \bar{\nu}_{\boldsymbol{\alpha}}$. To this end, recall that while proving Lemma 6.3.24 we showed that invariant measures for $\overline{p}$, such as $\bar{\mu}$ and $\bar{\nu}_{\boldsymbol{\alpha}}$ are also invariant for the transition probability $k(\cdot,\cdot)$ of (6.3.6), and by strong H-irreducibility the measurable function $g(\cdot) = k(\cdot, \{\boldsymbol{\alpha}\})$ is strictly positive on $\overline{\mathbb{S}}$. Therefore,

$$\bar{\mu}(g) = (\bar{\mu}k)(\{\boldsymbol{\alpha}\}) = \bar{\mu}(\{\boldsymbol{\alpha}\}) = 1 = \bar{\nu}_{\boldsymbol{\alpha}}(\{\boldsymbol{\alpha}\}) = (\bar{\nu}_{\boldsymbol{\alpha}}k)(\{\boldsymbol{\alpha}\}) = \bar{\nu}_{\boldsymbol{\alpha}}(g)\,.$$

Recall Exercise 4.1.13 that identity such as $\bar{\mu}(g) = \bar{\nu}_{\boldsymbol{\alpha}}(g) = 1$ for a strictly positive $g \in m\overline{\mathcal{S}}$, strengthens the inequality (6.3.10) between two σ-finite measures $\bar{\mu}$ and $\bar{\nu}_{\boldsymbol{\alpha}}$ on $(\overline{\mathbb{S}}, \overline{\mathcal{S}})$ into the claimed equality $\bar{\mu} = \bar{\nu}_{\boldsymbol{\alpha}}$. $\square$

The next result is a natural extension of Theorem 6.2.57.

Theorem 6.3.27. *Suppose $\{X_n\}$ and $\{Y_n\}$ are independent copies of a strong H-irreducible chain. Then, for any initial distribution of (X_0, Y_0) and all n,*

$$\|\mathcal{P}_{X_n} - \mathcal{P}_{Y_n}\|_{tv} \leq 2\mathbf{P}(\tau > n)\,, \tag{6.3.11}$$

where $\|\cdot\|_{tv}$ denotes the total variation norm *of Definition 3.2.22 and $\tau = \min\{\ell \geq 0 : \overline{X}_\ell = \overline{Y}_\ell = \boldsymbol{\alpha}\}$ is the time of the first joint visit of the atom by the corresponding copies of the split chain under the coupling of Proposition 6.3.4.*

Proof. Fixing $g \in b\mathcal{S}$ bounded by one, recall that the split mapping yields $\overline{g} \in b\overline{\mathcal{S}}$ of the same bound, and by part (c) of Proposition 6.3.4

$$\mathbf{E}\, g(X_n) - \mathbf{E}\, g(Y_n) = \mathbf{E}\, \overline{g}(\overline{X}_n) - \mathbf{E}\, \overline{g}(\overline{Y}_n)$$

for any joint initial distribution of (X_0, Y_0) on $(\mathbb{S}^2, \mathcal{S} \times \mathcal{S})$ and all $n \geq 0$. Further, since $\overline{X}_\tau = \overline{Y}_\tau$ in case $\tau \leq n$, following the proof of Theorem 6.2.57 one finds that $|\mathbf{E}\, \overline{g}(\overline{X}_n) - \mathbf{E}\, \overline{g}(\overline{Y}_n)| \leq 2\mathbf{P}(\tau > n)$. Since this applies for all $g \in b\mathcal{S}$ bounded by one, we are done. $\square$

Our goal is to extend the scope of the convergence result of Theorem 6.2.59 to the setting of positive H-recurrent chains. To this end, we first adapt Definition 6.2.54 of an aperiodic chain.

Definition 6.3.28. *The period of a strongly H-irreducible chain is the g.c.d. $d_{\boldsymbol{\alpha}}$ of the set $\mathcal{I}_{\boldsymbol{\alpha}} = \{n \geq 1 : \mathbf{P}_{\boldsymbol{\alpha}}(\overline{X}_n = \boldsymbol{\alpha}) > 0\}$, of return times to its pseudo-atom and such chain is called* aperiodic *if it has period one. For example, $q(C) > 0$ implies aperiodicity of the chain.*

Remark. Recall that being (strongly) H-irreducible amounts for a countable state space to having exactly one non-empty equivalence class of intercommunicating states (which is accessible from any other state). The preceding definition then coincides with the common period of these intercommunicating states per Definition 6.2.54.

More generally, our definition of the period of the chain seems to depend on which small set and regeneration measure one chooses. However, in analogy with Exercise 6.3.20, after some work it can be shown that any two split chains for the *same* strong H-irreducible chain induce the same period.

Theorem 6.3.29. *Let $\pi(\cdot)$ denote the unique invariant probability measure of an aperiodic positive H-recurrent Markov chain $\{X_n\}$. If $x \in \mathbb{S}$ is such that $\mathbf{P}_x(T_{\boldsymbol{\alpha}} < \infty) = 1$, then*

$$\lim_{n\to\infty} \|\mathbf{P}_x(X_n \in \cdot) - \pi(\cdot)\|_{tv} = 0\,. \tag{6.3.12}$$

Remark. It follows from (6.3.9) that $\pi(\cdot)$ is absolutely continuous with respect to $\overline{\psi}(\cdot)$ of Proposition 6.3.18. Hence, by parts (a) and (c) of Exercise 6.3.19, both

$$\mathbf{P}_\pi(T_{\boldsymbol{\alpha}} < \infty) = 1\,, \tag{6.3.13}$$

and $\mathbf{P}_x(T_{\boldsymbol{\alpha}} < \infty) = 1$ for π-a.e. $x \in \mathbb{S}$. Consequently, the convergence result (6.3.12) holds for π-a.e. $x \in \mathbb{S}$.

Proof. Consider independent copies $\overline{X}_n$ and $\overline{Y}_n$ of the split chain starting at $\overline{X}_0 = x$ and at $\overline{Y}_0$ whose law is the invariant probability measure $\overline{\pi} = \pi\overline{p}$ of the split chain. The corresponding X_n and Y_n per Proposition 6.3.4 have the laws $\mathbf{P}_x(X_n \in \cdot)$ and $\pi(\cdot)$, respectively. Hence, in view of Theorem 6.3.27, to establish (6.3.12) it suffices to show that with probability one $\overline{X}_n = \overline{Y}_n = \boldsymbol{\alpha}$ for some finite, possibly random value of n. Proceeding to prove the latter fact, recall (6.3.13) and the H-recurrence of the chain, in view of which we have with probability one that $\overline{Y}_n = \boldsymbol{\alpha}$ for infinitely many values of n, say at random times $\{R_k\}$. Similarly, our assumption that $\mathbf{P}_x(T_{\boldsymbol{\alpha}} < \infty) = 1$ implies that with probability one $\overline{X}_n = \boldsymbol{\alpha}$ for infinitely many values of n, say at another sequence of random times $\{\widetilde{R}_k\}$ and it remains to show that these two random subsets of $\{1, 2, \ldots\}$ intersect with probability one. To this end, note that upon adapting the argument used in solving Exercise 6.2.11 you find that R_1, $\widetilde{R}_1$, $r_k = R_{k+1} - R_k$ and $\widetilde{r}_k = \widetilde{R}_{k+1} - \widetilde{R}_k$ for $k \geq 1$ are mutually independent, with $\{r_k, \widetilde{r}_k, k \geq 1\}$ identically distributed, each following the law of $T_{\boldsymbol{\alpha}}$ under $\mathbf{P}_{\boldsymbol{\alpha}}$. Let $W_{n+1} = W_n + Z_n$ and $\widetilde{W}_{n+1} = \widetilde{W}_n + \widetilde{Z}_n$, starting at $W_0 = \widetilde{W}_0 = 1$, where the i.i.d. $\{Z, Z_\ell, \widetilde{Z}_\ell\}$ are independent of $\{\overline{X}_n\}$ and $\{\overline{Y}_n\}$ and such that $\mathbf{P}(Z = k) = 2^{-k}$ for $k \geq 1$. It then follows by the strong Markov property of the split chains that $S_n = R_{W_n} - \widetilde{R}_{\widetilde{W}_n}$, $n \geq 0$, is a random walk on $\mathbb{Z}$, whose i.i.d. increments $\{\xi_n\}$ have each the law of the difference between two independent copies of $T^Z_{\boldsymbol{\alpha}}$ under $\mathbf{P}_{\boldsymbol{\alpha}}$. As mentioned already, our thesis follows from $\mathbf{P}(S_n = 0 \text{ i.o.}) = 1$, which in view of Corollary 6.2.12 and Theorem 6.2.13 is in turn an immediate consequence of our claim that $\{S_n\}$ is an irreducible, recurrent Markov chain.

Turning to prove that $\{S_n\}$ is irreducible, note that since Z is independent of $\{T^k_{\boldsymbol{\alpha}}\}$, for any $n \geq 1$

$$\mathbf{P}_{\boldsymbol{\alpha}}(T^Z_{\boldsymbol{\alpha}} = n) = \sum_{k=1}^\infty 2^{-k}\mathbf{P}_{\boldsymbol{\alpha}}(T^k_{\boldsymbol{\alpha}} = n) = \sum_{k=1}^\infty 2^{-k}\mathbf{P}_{\boldsymbol{\alpha}}(N_n(\boldsymbol{\alpha}) = k, \overline{X}_n = \boldsymbol{\alpha})\,.$$

Consequently, $\mathbf{P}_{\boldsymbol{\alpha}}(T^Z_{\boldsymbol{\alpha}} = n) > 0$ if and only if $\mathbf{P}_{\boldsymbol{\alpha}}(\overline{X}_n = \boldsymbol{\alpha}) > 0$. That is, the support of the law of $T^Z_{\boldsymbol{\alpha}}$ is the set $\mathcal{I}_{\boldsymbol{\alpha}}$ of possible return times to $\boldsymbol{\alpha}$. By the assumed aperiodicity of the chain, the g.c.d. of $\mathcal{I}_{\boldsymbol{\alpha}}$ is one (see Definition 6.3.28). Further, by definition this subset of positive integers is closed under addition, hence as we have seen in the course of proving Lemma 6.2.55, the set $\mathcal{I}_{\boldsymbol{\alpha}}$ contains all large enough integers. As ξ_1 is the difference between two independent copies of $T^Z_{\boldsymbol{\alpha}}$, the law of each of which is strictly positive for all large enough positive integers, clearly $\mathbf{P}(\xi_1 = z) > 0$ for all $z \in \mathbb{Z}$, out of which the irreducibility of $\{S_n\}$ follows.

As for the recurrence of $\{S_n\}$, note that by the assumed positive H-recurrence of $\{X_n\}$ and the independence of Z and this chain,

$$\mathbf{E}_{\boldsymbol{\alpha}}(T^Z_{\boldsymbol{\alpha}}) = \sum_{k=1}^{\infty} \mathbf{E}_{\boldsymbol{\alpha}}(T^k_{\boldsymbol{\alpha}})\mathbf{P}(Z=k) = \mathbf{E}_{\boldsymbol{\alpha}}(T_{\boldsymbol{\alpha}}) \sum_{k=1}^{\infty} k\mathbf{P}(Z=k) = \mathbf{E}_{\boldsymbol{\alpha}}(T_{\boldsymbol{\alpha}})\mathbf{E}(Z) < \infty\,.$$

Hence, the increments ξ_n of the irreducible random walk $\{S_n\}$ on $\mathbb{Z}$ are integrable and of zero mean. Consequently, $n^{-1}S_n \xrightarrow{p} 0$ as $n \to \infty$ which by the Chung-Fuchs theorem implies the recurrence of $\{S_n\}$ (see Exercise 6.3.23). $\square$

Exercise 6.3.30. *Suppose $\{X_k\}$ is the first order auto-regressive process $X_n = \beta X_{n-1} + \xi_n$, $n \geq 1$ with $|\beta| < 1$ and where the integrable i.i.d. $\{\xi_n\}$ have a strictly positive, continuous density $f_\xi(\cdot)$ with respect to Lebesgue measure on $\mathbb{R}^d$.*

(a) *Show that $\{X_k\}$ is a strong H-irreducible chain.*
(b) *Show that $V_n = \sum_{k=0}^n \beta^k \xi_k$ converges a.s. to $V_\infty = \sum_{k\geq 0} \beta^k \xi_k$ whose law $\pi(\cdot)$ is an invariant probability measure for $\{X_k\}$.*
(c) *Show that $\{X_k\}$ is positive H-recurrent.*
(d) *Explain why $\{X_k\}$ is aperiodic and deduce that starting at any fixed $x \in \mathbb{R}^d$ the law of X_n converges in total variation to $\pi(\cdot)$.*

Exercise 6.3.31. *Show that if $\{X_n\}$ is an aperiodic, positive H-recurrent chain and $x, y \in \mathbb{S}$ are such that $\mathbf{P}_x(T_{\boldsymbol{\alpha}} < \infty) = \mathbf{P}_y(T_{\boldsymbol{\alpha}} < \infty) = 1$, then for any $A \in \mathcal{S}$,*

$$\lim_{n\to\infty} |\mathbf{P}_x(X_n \in A) - \mathbf{P}_y(X_n \in A)| = 0\,.$$

Exercise 6.3.32. *Suppose $\{\xi_n\}$ are i.i.d. with $\mathbf{P}(\xi_1 = 1) = 1 - \mathbf{P}(\xi_1 = -1) = b$ and $\{U_n\}$ are i.i.d. uniform on $[-5, 5]$ and independent of $\{\xi_n\}$. Consider the Markov chain $\{X_n\}$ with state space $\mathbb{S} = \mathbb{R}$ such that $X_n = X_{n-1} + \xi_n \mathrm{sign}(X_{n-1})$ when $|X_{n-1}| > 5$ and $X_n = X_{n-1} + U_n$ otherwise.*

(a) *Show that this chain is strongly H-irreducible for any $0 \leq b < 1$.*
(b) *Show that it has a unique invariant measure (up to a constant multiple), when $0 \leq b \leq 1/2$.*
(c) *Show that if $0 \leq b < 1/2$ the chain has a unique invariant probability measure $\pi(\cdot)$ and that $\mathbf{P}_x(X_n \in B) \to \pi(B)$ as $n \to \infty$ for any $x \in \mathbb{R}$ and every Borel set B.*

CHAPTER 7

Continuous, Gaussian and stationary processes

A discrete parameter *stochastic process* (S.P.) is merely a sequence of random variables. We have encountered and constructed many such processes when considering martingales and Markov chains in Sections 5.1 and 6.1, respectively. Our focus here is on continuous time processes, each of which consists of an *uncountable* collection of random variables (defined on the same probability space).

We have successfully constructed by an ad-hoc method one such process, namely the Poisson process of Section 3.4. In contrast, Section 7.1 provides a *canonical construction* of S.P., viewed as a collection of R.V.-s $\{X_t(\omega), t \in \mathbb{T}\}$. This construction, based on the specification of finite dimensional distributions, applies for any index set $\mathbb{T}$ and any S.P. taking values in a $\mathcal{B}$-isomorphic measurable space.

However, this approach ignores the *sample function* $t \mapsto X_t(\omega)$ of the process. Consequently, the resulting law of the S.P. provides no information about probabilities such as that of continuity of the sample function, or whether it is ever zero, or the distribution of $\sup_{t\in\mathbb{T}} X_t$. We thus detail in Section 7.2 a way to circumvent this difficulty, whereby we guarantee, under suitable conditions, the continuity of the sample function for almost all outcomes ω, or at the very least, its (Borel) measurability.

We conclude this chapter by studying in Section 7.3 the concept of stationary (of processes and their increments), and the class of Gaussian (stochastic) processes, culminating with the definition and construction of the Brownian motion.

7.1. Definition, canonical construction and law

We start with the definition of a stochastic process.

Definition 7.1.1. *Given* $(\Omega, \mathcal{F}, \mathbf{P})$*, a* stochastic process*, denoted* $\{X_t\}$*, is a collection* $\{X_t : t \in \mathbb{T}\}$ *of R.V.-s. In case the index set* $\mathbb{T}$ *is an interval in* $\mathbb{R}$ *we call it a* continuous time *S.P. The function* $t \mapsto X_t(\omega)$ *is called the* sample function *(or* sample path*,* realization*, or* trajectory*), of the S.P. at* $\omega \in \Omega$*.*

We shall follow the approach we have taken in constructing product measures (in Section 1.4.2) and repeated for dealing with Markov chains (in Section 6.1). To this end, we start with the *finite dimensional distributions* associated with the S.P.

Definition 7.1.2. *By* finite dimensional distributions *(f.d.d.) of a S.P.* $\{X_t, t \in \mathbb{T}\}$ *we refer to the collection of probability measures* $\mu_{t_1,t_2,\cdots,t_n}(\cdot)$ *on* $\mathcal{B}^n$*, indexed by* n *and distinct* $t_k \in \mathbb{T}$*,* $k = 1, \ldots, n$*, where*

$$\mu_{t_1,t_2,\cdots,t_n}(B) = \mathbf{P}((X_{t_1}, X_{t_2}, \cdots, X_{t_n}) \in B)\,,$$

for any Borel subset B *of* $\mathbb{R}^n$*.*

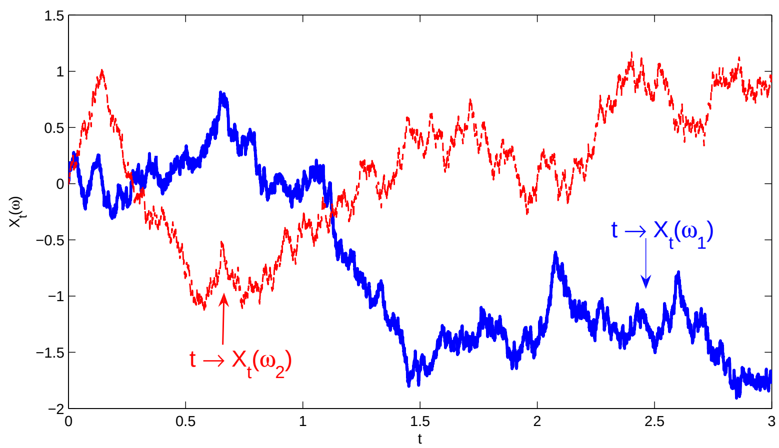

FIGURE 1. Sample functions of a continuous time stochastic process, corresponding to two outcomes ω_1 and ω_2.

Not all f.d.d. are relevant here, for you should convince yourself that the f.d.d. of any S.P. should be consistent, as specified next.

Definition 7.1.3. *We say that a collection of finite dimensional distributions is* consistent *if for any $B_k \in \mathcal{B}$, distinct $t_k \in \mathbb{T}$ and finite n,*

$$\mu_{t_1,\cdots,t_n}(B_1 \times \cdots \times B_n) = \mu_{t_{\pi(1)},\cdots,t_{\pi(n)}}(B_{\pi(1)} \times \cdots \times B_{\pi(n)})\,, \tag{7.1.1}$$

for any permutation π of $\{1, 2, \cdots, n\}$ and

$$\mu_{t_1,\cdots,t_{n-1}}(B_1 \times \cdots \times B_{n-1}) = \mu_{t_1,\cdots,t_{n-1},t_n}(B_1 \times \cdots \times B_{n-1} \times \mathbb{R})\,. \tag{7.1.2}$$

Here is a simpler, equivalent definition of consistent f.d.d. in case $\mathbb{T}$ is linearly (i.e. totally) ordered.

Lemma 7.1.4. *In case $\mathbb{T}$ is a linearly ordered set (for example, $\mathbb{T}$ countable, or $\mathbb{T} \subseteq \mathbb{R}$), it suffices to define as f.d.d. the collection of probability measures $\mu_{s_1,\ldots,s_n}(\cdot)$ running over $s_1 < s_2 < \cdots < s_n$ in $\mathbb{T}$ and finite n, where such collection is* consistent *if and only if for any $A_i \in \mathcal{B}$ and $k = 1, \ldots, n$,*

$$\begin{aligned} &\mu_{s_1,\cdots,s_n}(A_1 \times \cdots \times A_{k-1} \times \mathbb{R} \times A_{k+1} \times \cdots \times A_n) \\ &= \mu_{s_1,\cdots,s_{k-1},s_{k+1},\cdots,s_n}(A_1 \times \cdots \times A_{k-1} \times A_{k+1} \times \cdots \times A_n) \end{aligned} \tag{7.1.3}$$

PROOF. Since the set $\mathbb{T}$ is linearly ordered, for any distinct $t_i \in \mathbb{T}$, $i = 1, \ldots, n$ there exists a unique permutation π on $\{1, \ldots, n\}$ such that $s_i = t_{\pi(i)}$ are in increasing order and taking the random vector $(X_{s_1}, \cdots, X_{s_n})$ of (joint) distribution $\mu_{s_1,\cdots,s_n}(\cdot)$, we set $\mu_{t_1,\cdots,t_n}$ as the distribution of the vector $(X_{t_1}, \cdots, X_{t_n})$ of permuted coordinates. This unambiguously extends the definition of the f.d.d. from the ordered $s_1 < \cdots < s_n$ to all distinct $t_i \in \mathbb{T}$. Proceeding to verify the consistency of these f.d.d. note that by our definition, the identity (7.1.1) holds whenever $\{t_{\pi(i)}\}$ are in increasing order. Permutations of $\{1, \ldots, n\}$ form a group with respect to composition, so (7.1.1) extends to $\{t_{\pi(i)}\}$ of arbitrary order. Next suppose that in the permutation π of $\{1, \ldots, n\}$ such that $s_i = t_{\pi(i)}$ are in increasing order we have $n = \pi(k)$ for some $1 \le k \le n$. Then, setting $B_n = \mathbb{R}$ and $A_i = B_{\pi(i)}$ leads to $A_k = \mathbb{R}$ and from (7.1.1) and (7.1.3) it follows that

$$\mu_{t_1,\cdots,t_n}(B_1\times\cdots\times B_{n-1}\times\mathbb{R}) = \mu_{s_1,\cdots,s_{k-1},s_{k+1},\cdots,s_n}(A_1\times\cdots\times A_{k-1}\times A_{k+1}\times\cdots\times A_n)\,.$$

Further, $(t_1,\dots,t_{n-1})$ is the image of $(s_1,\dots,s_{k-1},s_{k+1},\dots,s_n)$ under the permutation π^{-1} restricted to $\{1,\dots,k-1,k+1,\dots,n\}$ so a second application of (7.1.1) results with the consistency condition (7.1.2). $\square$

Our goal is to establish the existence and uniqueness (in law) of the S.P. associated with any given consistent collection of f.d.d. We shall do so via a *canonical construction*, whereby we set $\Omega = \mathbb{R}^{\mathbb{T}}$ and $\mathcal{F} = \mathcal{B}^{\mathbb{T}}$ as follows.

Definition 7.1.5. *Let $\mathbb{R}^{\mathbb{T}}$ denote the collection of all functions $x(t) : \mathbb{T} \mapsto \mathbb{R}$. A* finite dimensional measurable rectangle *in $\mathbb{R}^{\mathbb{T}}$ is any set of the form $\{x(\cdot) : x(t_i) \in B_i, i = 1,\dots,n\}$ for a positive integer n, $B_i \in \mathcal{B}$ and $t_i \in \mathbb{T}$, $i = 1,\dots,n$. The* cylindrical σ-algebra *$\mathcal{B}^{\mathbb{T}}$ is the σ-algebra generated by the collection of all finite dimensional measurable rectangles.*

Note that in case $\mathbb{T} = \{1,2,\dots\}$, the σ-algebra $\mathcal{B}^{\mathbb{T}}$ is precisely the product σ-algebra $\mathcal{B}_c$ used in stating and proving Kolmogorov's extension theorem. Further, enumerating $\mathbb{C} = \{t_k\}$, it is not hard to see that $\mathcal{B}^{\mathbb{C}}$ is in one to one correspondence with $\mathcal{B}_c$ for any infinite, countable $\mathbb{C} \subseteq \mathbb{T}$.

The next concept is handy in studying the structure of $\mathcal{B}^{\mathbb{T}}$ for uncountable $\mathbb{T}$.

Definition 7.1.6. *We say that $A \subseteq \mathbb{R}^{\mathbb{T}}$ has a* countable representation *if*

$$A = \{x(\cdot) \in \mathbb{R}^{\mathbb{T}} : (x(t_1), x(t_2), \dots) \in D\}\,,$$

for some $D \in \mathcal{B}_c$ and $\mathbb{C} = \{t_k\} \subseteq \mathbb{T}$. The set $\mathbb{C}$ is then called the (countable) base *of the (countable) representation $(\mathbb{C}, D)$ of A.*

Indeed, $\mathcal{B}^{\mathbb{T}}$ consists of the sets in $\mathbb{R}^{\mathbb{T}}$ having a countable representation and is further image of $\mathcal{F}^{\mathbf{X}} = \sigma(X_t, t \in \mathbb{T})$ via the mapping $X_{\cdot} : \Omega \mapsto \mathbb{R}^{\mathbb{T}}$.

Lemma 7.1.7. *The σ-algebra $\mathcal{B}^{\mathbb{T}}$ is the collection $\mathcal{C}$ of all subsets of $\mathbb{R}^{\mathbb{T}}$ that have a countable representation. Further, for any S.P. $\{X_t, t \in \mathbb{T}\}$, the σ-algebra $\mathcal{F}^{\mathbf{X}}$ is the collection $\mathcal{G}$ of sets of the form $\{\omega \in \Omega : X_{\cdot}(\omega) \in A\}$ with $A \in \mathcal{B}^{\mathbb{T}}$.*

Proof. First note that for any subsets $\mathbb{T}_1 \subseteq \mathbb{T}_2$ of $\mathbb{T}$, the restriction to $\mathbb{T}_1$ of functions on $\mathbb{T}_2$ induces a measurable projection $p : (\mathbb{R}^{\mathbb{T}_2}, \mathcal{B}^{\mathbb{T}_2}) \mapsto (\mathbb{R}^{\mathbb{T}_1}, \mathcal{B}^{\mathbb{T}_1})$. Further, enumerating over a countable $\mathbb{C}$ maps the corresponding cylindrical σ-algebra $\mathcal{B}^{\mathbb{C}}$ in a one to one manner into the product σ-algebra $\mathcal{B}_c$. Thus, if $A \in \mathcal{C}$ has the countable representation $(\mathbb{C}, D)$ then $A = p^{-1}(D)$ for the measurable projection p from $\mathbb{R}^{\mathbb{T}}$ to $\mathbb{R}^{\mathbb{C}}$, hence $A \in \mathcal{B}^{\mathbb{T}}$. Having just shown that $\mathcal{C} \subseteq \mathcal{B}^{\mathbb{T}}$ we turn to show that conversely $\mathcal{B}^{\mathbb{T}} \subseteq \mathcal{C}$. Since each finite dimensional measurable rectangle has a countable representation (of a finite base), this is an immediate consequence of the fact that $\mathcal{C}$ is a σ-algebra. Indeed, $\mathbb{R}^{\mathbb{T}}$ has a countable representation (of empty base), and if $A \in \mathcal{C}$ has the countable representation $(\mathbb{C}, D)$ then A^c has the countable representation $(\mathbb{C}, D^c)$. Finally, if $A_k \in \mathcal{C}$ has a countable representation $(\mathbb{C}_k, D_k)$ for $k = 1,2,\dots$ then the subset $\mathbb{C} = \cup_k \mathbb{C}_k$ of $\mathbb{T}$ serves as a common countable base for these sets. That is, A_k has the countable representation $(\mathbb{C}, \widetilde{D}_k)$, for $k = 1,2,\dots$ and $\widetilde{D}_k = p_k^{-1}(D_k) \in \mathcal{B}_c$, where p_k denotes the measurable projection from $\mathbb{R}^{\mathbb{C}}$ to

$\mathbb{R}^{\mathbb{C}_k}$. Consequently, as claimed $\cup_k A_k \in \mathcal{C}$ for it has the countable representation $(\mathbb{C}, \cup_k \widetilde{D}_k)$.

As for the second part of the lemma, temporarily imposing on Ω the σ-algebra 2^Ω makes $X_. : \Omega \mapsto \mathbb{R}^{\mathbb{T}}$ an $(\mathbb{S}, \mathcal{S})$-valued R.V. for $\mathbb{S} = \mathbb{R}^{\mathbb{T}}$ and $\mathcal{S} = \mathcal{B}^{\mathbb{T}}$. From Exercises 1.2.10 and 1.2.11 we thus deduce that $\mathcal{G}$ is the σ-algebra generated by the sets of the form $\{\omega \in \Omega : X_{t_i}(\omega) \in B_i, i = 1, \ldots, n\}$ for $B_i \in \mathcal{B}$, $t_i \in \mathbb{T}$ and finite n, which is precisely the σ-algebra $\mathcal{F}^{\mathbf{X}}$. $\square$

Combining Lemma 7.1.7 and Kolmogorov's extension theorem, we proceed with the promised canonical construction, yielding the following conclusion.

Proposition 7.1.8. *For any consistent collection of f.d.d., there exists a probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and a stochastic process $\omega \mapsto \{X_t(\omega), t \in \mathbb{T}\}$ on it, whose f.d.d. are in agreement with the given collection. Further, the restriction of the probability measure $\mathbf{P}$ to the σ-algebra $\mathcal{F}^{\mathbf{X}}$ is uniquely determined by the specified f.d.d.*

Proof. Starting with the existence of the probability space, suppose first that $\mathbb{T} = \mathbb{C}$ is countable. In this case, enumerating over $\mathbb{C} = \{s_j\}$ we further have from the consistency condition (7.1.3) of Lemma 7.1.4 that it suffices to consider the sequence of f.d.d. $\mu_{s_1,\ldots,s_n}$ for $n = 1, 2, \ldots$ and the existence of a probability measure $\mathbf{P}_{\mathbb{C}}$ on $(\mathbb{R}^{\mathbb{C}}, \mathcal{B}_c)$ that agrees with the given f.d.d. follows by Kolmogorov's extension theorem (i.e. Theorem 1.4.22). Moving to deal with uncountable $\mathbb{T}$, take $\Omega = \mathbb{R}^{\mathbb{T}}$ and $\mathcal{F} = \mathcal{B}^{\mathbb{T}}$ with $X_t(\omega) = \omega_t$. Recall Lemma 7.1.7, that any $A \in \mathcal{B}^{\mathbb{T}}$ has a countable representation $(\mathbb{C}, D)$ so we can assign $\mathbf{P}(A) = \mathbf{P}_{\mathbb{C}}(D)$, where $\mathbf{P}_{\mathbb{C}}$ is defined through Kolmogorov's extension theorem for the countable subset $\mathbb{C}$ of $\mathbb{T}$. We proceed to show that $\mathbf{P}(\cdot)$ is well defined. That is, $\mathbf{P}_{\mathbb{C}_1}(D_1) = \mathbf{P}_{\mathbb{C}_2}(D_2)$ for any two countable representations $(\mathbb{C}_1, D_1)$ and $(\mathbb{C}_2, D_2)$ of the same set $A \in \mathcal{B}^{\mathbb{T}}$. Since $\mathbb{C} = \mathbb{C}_1 \cup \mathbb{C}_2$ is then also a countable base for A, we may and shall assume that $\mathbb{C}_1 \subset \mathbb{C}_2$ in which case necessarily $D_2 = p_{21}^{-1}(D_1)$ for the measurable projection p_{21} from $\mathbb{R}^{\mathbb{C}_2}$ to $\mathbb{R}^{\mathbb{C}_1}$. By their construction, $\mathbf{P}_{\mathbb{C}_i}$ for $i = 1, 2$ coincide on all finite dimensional measurable rectangles with a base in $\mathbb{C}_1$. Hence, $\mathbf{P}_{\mathbb{C}_1} = \mathbf{P}_{\mathbb{C}_2} \circ p_{21}^{-1}$ and in particular $\mathbf{P}_{\mathbb{C}_2}(D_2) = \mathbf{P}_{\mathbb{C}_1}(D_1)$. By construction the non-negative set function $\mathbf{P}$ on $(\mathbb{R}^{\mathbb{T}}, \mathcal{B}^{\mathbb{T}})$ has the specified f.d.d. for $X_t(\omega) = \omega_t$ so we complete the proof of existence by showing that $\mathbf{P}$ is countably additive. To this end, as shown in the proof of Lemma 7.1.7, any sequence of disjoint sets $A_k \in \mathcal{B}^{\mathbb{T}}$ admits countable representations $(\mathbb{C}, \widetilde{D}_k)$, $k = 1, 2, \ldots$ with a common base $\mathbb{C}$ and disjoint $\widetilde{D}_k \in \mathcal{B}_c$. Hence, by the countable additivity of $\mathbf{P}_{\mathbb{C}}$,

$$\mathbf{P}(\cup_k A_k) = \mathbf{P}_{\mathbb{C}}(\cup_k \widetilde{D}_k) = \sum_k \mathbf{P}_{\mathbb{C}}(\widetilde{D}_k) = \sum_k \mathbf{P}(A_k)\,.$$

As for uniqueness, recall Lemma 7.1.7 that every set in $\mathcal{F}^{\mathbf{X}}$ is of the form $\{\omega : (X_{t_1}(\omega), X_{t_2}(\omega), \ldots) \in D\}$ for some $D \in \mathcal{B}_c$ and $\mathbb{C} = \{t_j\}$ a countable subset of $\mathbb{T}$. Fixing such $\mathbb{C}$, recall Kolmogorov's extension theorem, that the law of $(X_{t_1}, X_{t_2}, \ldots)$ on $\mathcal{B}_c$ is uniquely determined by the specified laws of $(X_{t_1}, \ldots, X_{t_n})$ for $n = 1, 2, \ldots$. Since this applies for any countable $\mathbb{C}$, we see that the whole restriction of $\mathbf{P}$ to $\mathcal{F}^{\mathbf{X}}$ is uniquely determined by the given collection of f.d.d. $\square$

Remark. Recall Corollary 1.4.25 that Kolmogorov's extension theorem holds when $(\mathbb{R}, \mathcal{B})$ is replaced by any $\mathcal{B}$-isomorphic measurable space $(\mathbb{S}, \mathcal{S})$. Check that thus, the same applies for the preceding proof, hence Proposition 7.1.8 holds for any

$(\mathbb{S}, \mathcal{S})$-valued S.P. $\{X_t\}$ provided $(\mathbb{S}, \mathcal{S})$ is $\mathcal{B}$-isomorphic (c.f. [**Dud89**, Theorem 12.1.2] for an even more general setting in which the same applies).

Motivated by Proposition 7.1.8 our definition of the law of the S.P. is as follows.

Definition 7.1.9. *The* law *(or* distribution*) of a S.P. is the probability measure $\mathcal{P}_{\mathbf{X}}$ on $\mathcal{B}^{\mathbb{T}}$ such that for all $A \in \mathcal{B}^{\mathbb{T}}$,*

$$\mathcal{P}_{\mathbf{X}}(A) = \mathbf{P}(\{\omega : X_\cdot(\omega) \in A\})\,.$$

Proposition 7.1.8 tells us that the f.d.d. uniquely determine the law of any S.P. and provide the probability of any event in $\mathcal{F}^{\mathbf{X}}$. However, for our construction to be considered a success story, we want most events of interest be in $\mathcal{F}^{\mathbf{X}}$. That is, their image via the sample function should be in $\mathcal{B}^{\mathbb{T}}$. Unfortunately, as we show next, this is *certainly not the case* for uncountable $\mathbb{T}$.

Lemma 7.1.10. *Fixing $\gamma \in \mathbb{R}$ and $\mathbb{I} = [a, b)$ for some $a < b$, the following sets*

$$A_\gamma = \{x \in \mathbb{R}^{\mathbb{I}} : x(t) \le \gamma \text{ for all } t \in \mathbb{I}\}\,,$$
$$C(\mathbb{I}) = \{x \in \mathbb{R}^{\mathbb{I}} : t \mapsto x(t) \text{ is continuous on } \mathbb{I}\}\,,$$

are not in $\mathcal{B}^{\mathbb{I}}$.

Proof. In view of Lemma 7.1.7, if $A_\gamma \in \mathcal{B}^{\mathbb{I}}$ then A_γ has a countable base $\mathbb{C} = \{t_k\}$ and in particular the values of $x(t_k)$ determine whether $x(\cdot) \in A_\gamma$ or not. But $\mathbb{C}$ is a *strict* subset of the uncountable index set $\mathbb{I}$, so fixing some values $x(t_k) \le \gamma$ for all $t_k \in \mathbb{C}$, the function $x(\cdot)$ on $\mathbb{I}$ still may or may not be in A_γ, as by definition the latter further requires that $x(t) \le \gamma$ for all $t \in \mathbb{I} \setminus \mathbb{C}$. Similarly, if $C(\mathbb{I}) \in \mathcal{B}^{\mathbb{I}}$ then it has a countable base $\mathbb{C} = \{t_k\}$ and the values of $x(t_k)$ determine whether $x(\cdot) \in C(\mathbb{I})$. However, since $\mathbb{C} \neq \mathbb{I}$, fixing $x(\cdot)$ continuous on $\mathbb{I} \setminus \{t\}$ with $t \in \mathbb{I} \setminus \mathbb{C}$, the function $x(\cdot)$ may or may not be continuous on $\mathbb{I}$, depending on the value of $x(t)$. $\square$

Remark. With $A_\gamma \notin \mathcal{B}^{\mathbb{I}}$, the canonical construction provides minimal information about $M_{\mathbb{I}} = \sup_{t \in \mathbb{I}} X_t$, which typically is not even measurable with respect to $\mathcal{F}^{\mathbf{X}}$. However, note that $A_\gamma \in \mathcal{F}^{\mathbf{X}}$ in case all sample functions of $\{X_t\}$ are right-continuous. That is, for such S.P. the law of $M_{\mathbb{I}}$ is uniquely determined by the f.d.d. We return to this point in Section 7.2 when considering separable modifications.

Similarly, since $C(\mathbb{I}) \notin \mathcal{B}^{\mathbb{I}}$, the canonical construction does not assign a probability for continuity of the sample function. To further demonstrate that this type of difficulty is generic, recall that by our ad-hoc construction of the Poisson process out of its jump times, all sample functions of this process are in

$$Z_\uparrow = \{x \in \mathbb{Z}_+^{\mathbb{I}} : t \mapsto x(t) \text{ is non-decreasing}\}\,,$$

where $\mathbb{I} = [0, \infty)$. However, convince yourself that $Z_\uparrow \notin \mathcal{B}^{\mathbb{I}}$, so had we applied the canonical construction starting from the f.d.d. of the Poisson process, we would not have had any probability assigned to this key property of its sample functions.

You are next to extend the phenomena illustrated by Lemma 7.1.10, providing a host of relevant subsets of $\mathbb{R}^{\mathbb{I}}$ which are not in $\mathcal{B}^{\mathbb{I}}$.

Exercise 7.1.11. *Let $\mathbb{I} \subseteq \mathbb{R}$ denote an interval of positive length.*

(a) *Show that none of the following collections of functions is in $\mathcal{B}^{\mathbb{I}}$: all linear functions, all polynomials, all constants, all non-decreasing functions, all functions of bounded variation, all differentiable functions, all analytic functions, all functions continuous at a fixed $t \in \mathbb{I}$.*
(b) *Show that $\mathcal{B}^{\mathbb{I}}$ fails to contain the collection of functions that vanish somewhere in $\mathbb{I}$, the collection of functions such that $x(s) < x(t)$ for some $s < t$, and the collection of functions with at least one local maximum.*
(c) *Show that $C(\mathbb{I})$ has no non-empty subset $A \in \mathcal{B}^{\mathbb{I}}$, but the complement of $C(\mathbb{I})$ in $\mathbb{R}^{\mathbb{I}}$ has a non-empty subset $A \in \mathcal{B}^{\mathbb{I}}$.*
(d) *Show that the completion $\overline{\mathcal{B}}^{\mathbb{I}}$ of $\mathcal{B}^{\mathbb{I}}$ with respect to* any *probability measure $\mathbf{P}$ on $\mathcal{B}^{\mathbb{I}}$ fails to contain the set $A = \mathcal{B}(\mathbb{I})$ of all Borel measurable functions $x : \mathbb{I} \mapsto \mathbb{R}$.*
Hint: *Consider A and A^c.*

In contrast to the preceding exercise, independence of the increments of a S.P. is determined by its f.d.d.

Exercise 7.1.12. *A continuous time S.P. $\{X_t, t \geq 0\}$ has* independent increments *if $X_{t+h} - X_t$ is independent of $\mathcal{F}_t^{\mathbf{X}} = \sigma(X_s, 0 \leq s \leq t)$ for any $h > 0$ and all $t \geq 0$.*
Show that if $X_{t_1}, X_{t_2} - X_{t_1}, \ldots, X_{t_n} - X_{t_{n-1}}$ are mutually independent, for all $n < \infty$ and $0 \leq t_1 < t_2 < \cdots < t_n < \infty$, then $\{X_t\}$ has independent increments. Hence, this property is determined by the f.d.d. of $\{X_t\}$.

Here is the canonical construction for Poisson random measures, where $\mathbb{T}$ is not a subset of $\mathbb{R}$ (for example, the Poisson point processes where $\mathbb{T} = \mathcal{B}_{\mathbb{R}^d}$).

Exercise 7.1.13. *Let $\mathbb{T} = \{A \in \mathcal{X} : \mu(A) < \infty\}$ for a given measure space $(\mathbb{X}, \mathcal{X}, \mu)$. Construct a S.P. $\{N_A : A \in \mathbb{T}\}$ such that N_A has the Poisson($\mu(A)$) law for each $A \in \mathbb{T}$ and N_{A_k}, $k = 1, \ldots, n$ are $\mathbf{P}$-mutually independent whenever A_k, $k = 1, \ldots, n$ are disjoint sets.*
Hint: *Given $A_j \in \mathbb{T}$, $j = 1, 2$, let $B_{j1} = A_j = B_{j0}^c$ and N_{b_1, b_2}, for $b_1, b_2 \in \{0, 1\}$ such that $(b_1, b_2) \neq (0, 0)$, be independent R.V. of Poisson($\mu(B_{1b_1} \cap B_{2b_2})$) law. As the distribution of (N_{A_1}, N_{A_2}) take the joint law of $(N_{1,1} + N_{1,0}, N_{1,1} + N_{0,1})$.*

Remark. The Poisson process $\widehat{N}_t$ of rate one is merely the restriction to sets $A = [0, t]$, $t \geq 0$, of the Poisson random measure $\{N_A\}$ in case $\mu(\cdot)$ is Lebesgue's measure on $[0, \infty)$. More generally, in case $\mu(\cdot)$ has density $f(\cdot)$ with respect to Lebesgue's measure on $[0, \infty)$, we call such restriction $X_t = N_{[0,t]}$ the *inhomogeneous Poisson process* of rate function $f(t) \geq 0$, $t \geq 0$. It is a counting process of independent increments, which is a non-random *time change* $X_t = \widehat{N}_{\mu([0,t])}$ of a Poisson process of rate one, but in general the gaps between jump times of $\{X_t\}$ are neither i.i.d. nor of exponential distribution.

7.2. Continuous and separable modifications

The canonical construction of Section 7.1 determines the law of a S.P. $\{X_t\}$ on the image $\mathcal{B}^{\mathbb{T}}$ of $\mathcal{F}^{\mathbf{X}}$. Recall that $\mathcal{F}^{\mathbf{X}}$ is inadequate as far as properties of the sample functions $t \mapsto X_t(\omega)$ are concerned. Nevertheless, a typical patch of this approach is to choose among S.P. with the given f.d.d. one that has regular enough sample functions. To illustrate this, we start with a simple explicit example in which path properties are not entirely determined by the f.d.d.

Example 7.2.1. *consider the S.P.*

$$Y_t(\omega) = 0, \quad \forall t, \omega \qquad\qquad X_t(\omega) = \begin{cases} 1, t = \omega \\ 0, \ \textit{otherwise} \end{cases}$$

on the probability space $([0,1], \mathcal{B}_{[0,1]}, U)$*, with* U *the uniform measure on* $\mathbb{I} = [0,1]$*. Since* $A_t = \{\omega : X_t(\omega) \neq Y_t(\omega)\} = \{t\}$*, clearly* $\mathbf{P}(X_t = Y_t) = 1$ *for each fixed* $t \in \mathbb{I}$*. Moreover,* $\mathbf{P}(\bigcup_{i=1}^n A_{t_i}) = 0$ *for any* $t_1, \ldots, t_n \in \mathbb{I}$*, hence* $\{X_t\}$ *has the same f.d.d. as* $\{Y_t\}$*. However,* $\mathbf{P}(\{\omega : \sup_{t\in\mathbb{I}} X_t(\omega) \neq 0\}) = 1$*, whereas* $\mathbf{P}(\{\omega : \sup_{t\in\mathbb{I}} Y_t(\omega) \neq 0\}) = 0$*. Similarly,* $\mathbf{P}(\{\omega : X_\cdot(\omega) \in C(\mathbb{I})\}) = 0$*, whereas* $\mathbf{P}(\{\omega : Y_\cdot(\omega) \in C(\mathbb{I})\}) = 1$*.*

While the two S.P. of Example 7.2.1 have different maximal value and differ in their sample path continuity, we would typically consider one to be merely a (small) modification of the other, motivating our next definition.

Definition 7.2.2. *Stochastic processes* $\{X_t\}$ *and* $\{Y_t\}$ *are called* versions *of one another if they have the same f.d.d. A S.P.* $\{Y_t, t \in \mathbb{T}\}$ *is further called a* modification *of* $\{X_t, t \in \mathbb{T}\}$ *if* $\mathbf{P}(X_t \neq Y_t) = 0$ *for all* $t \in \mathbb{T}$ *and two such S.P. are called* indistinguishable *if* $\{\omega : X_t(\omega) \neq Y_t(\omega)$ *for some* $t \in \mathbb{T}\}$ *is a* $\mathbf{P}$*-null set (hence, upon completing the space,* $\overline{\mathbf{P}}(X_t \neq Y_t$ *for some* $t \in \mathbb{T}) = 0$*). Similarly to Definition 1.2.8, throughout we consider two indistinguishable S.P.-s to be the same process, hence often omit the qualifier "a.s." in reference to sample function properties that apply* for all $t \in \mathbb{T}$.

For example, $\{Y_t\}$ is the *continuous modification* of $\{X_t\}$ in Example 7.2.1 but these two processes are clearly distinguishable. In contrast, modifications with a.s. right-continuous sample functions are indistinguishable.

Exercise 7.2.3. *Show that continuous time S.P.-s* $\{X_t\}$ *and* $\{Y_t\}$ *which are modifications of each other and have w.p.1. right-continuous sample functions, must also be indistinguishable.*

You should also convince yourself at this point that as we have implied, if $\{Y_t\}$ is a modification of $\{X_t\}$, then $\{Y_t\}$ is also a version of $\{X_t\}$. The converse fails, for while a modification has to be defined on the same probability space as the original S.P. this is not required of versions. Even on the same probability space it is easy to find a pair of versions which are not modifications of each other.

Example 7.2.4. *For the uniform probability measure on the finite set* $\Omega = \{H, T\}$*, the constant in time S.P.-s* $X_t(\omega) = I_H(\omega)$ *and* $Y_t(\omega) = 1 - X_t(\omega)$ *are clearly versions of each other but not modifications of each other.*

We proceed to derive a relatively easy to check sufficient condition for the existence of a (continuous) modification of the S.P. which has *Hölder continuous* sample functions, as defined next.

Definition 7.2.5. *Recall that a function* $f(t)$ *on a metric space* $(\mathbb{T}, d(\cdot,\cdot))$ *is* locally γ-Hölder continuous *if*

$$\sup_{\{t\neq s, d(t,u)\vee d(s,u)<h_u\}} \frac{|f(t)-f(s)|}{d(t,s)^\gamma} \le c_u \,,$$

for $\gamma > 0$*, some* $c : \mathbb{T} \mapsto [0,\infty)$ *and* $h : \mathbb{T} \mapsto (0,\infty]$*, and is* uniformly γ-Hölder continuous *if the same applies for constant* $c < h = \infty$*. In case* $\gamma = 1$ *such functions are also called* locally (or uniformly) Lipschitz continuous*, respectively. We say* *that a S.P. $\{Y_t, t \in \mathbb{T}\}$ is locally/uniformly γ-Hölder/Lipschitz continuous, with respect to a metric $d(\cdot,\cdot)$ on $\mathbb{T}$ if its sample functions $t \mapsto Y_t(\omega)$ have the corresponding property (for some S.P $c_u(\omega) < \infty$ and $h_u(\omega) > 0$, further requiring c to be a non-random constant for uniform continuity). Since local Hölder continuity implies continuity, clearly then $\mathbf{P}(\{\omega : Y_t(\omega) \in C(\mathbb{T})\}) = 1$. That is, such processes have* continuous sample functions. *We also use the term* continuous modification *to denote a modification $\{\widetilde{X}_t\}$ of a given S.P. $\{X_t\}$ such that $\{\widetilde{X}_t\}$ has continuous sample functions (and similarly define* locally/uniformly γ-Hölder continuous modifications*).*

Remark. The Euclidean norm $d(t,s) = \|t-s\|$ is used for sample path continuity of a *random field*, namely, where $\mathbb{T} \subseteq \mathbb{R}^r$ for some finite r, taking $d(t,s) = |t-s|$ for a continuous time S.P. Also, recall that for compact metric space $(\mathbb{T}, d)$ there is no difference between local and uniform Hölder continuity of $f : \mathbb{T} \mapsto \mathbb{R}$, so in this case local γ-Hölder continuity of S.P. $\{Y_t, t \in \mathbb{T}\}$ is equivalent to

$$\mathbf{P}(\{\omega : \sup_{s \neq t \in \mathbb{T}} \frac{|Y_t(\omega) - Y_s(\omega)|}{d(t,s)^\gamma} \leq c(\omega)\}) = 1\,,$$

for some finite R.V. $c(\omega)$.

Theorem 7.2.6 (Kolmogorov-Centsov continuity theorem)**.** *Suppose $\{X_t\}$ is a S.P. indexed on $\mathbb{T} = \mathbb{I}^r$, with $\mathbb{I}$ a compact interval. If there exist positive constants α, β and finite c such that*

$$\mathbf{E}[|X_t - X_s|^\alpha] \leq c\|t-s\|^{r+\beta}, \qquad \text{for all} \quad s,t \in \mathbb{T}\,, \tag{7.2.1}$$

then there exists a continuous modification of $\{X_t, t \in \mathbb{T}\}$ which is also locally γ-Hölder continuous for any $0 < \gamma < \beta/\alpha$.

Remark. Since condition (7.2.1) involves only the joint distribution of (X_s, X_t), it is determined by the f.d.d. of the process. Consequently, either all versions of the given S.P. satisfy (7.2.1) or none of them does.

PROOF. We consider hereafter the case of $r = 1$, assuming with no loss of generality that $\mathbb{T} = [0,1]$, and leave to the reader the adaptation of the proof to $r \geq 2$ (to this end, see [**KaS97**, Solution of Problem 2.2.9]).

Our starting point is the bound

$$\mathbf{P}(|X_t - X_s| \geq \varepsilon) \leq \varepsilon^{-\alpha}\mathbf{E}[|X_t - X_s|^\alpha] \leq c\varepsilon^{-\alpha}|t-s|^{1+\beta}\,, \tag{7.2.2}$$

which holds for any $\varepsilon > 0$, $t, s \in \mathbb{I}$, where the first inequality follows from Markov's inequality and the second from (7.2.1). From this bound we establish the a.s. local Hölder continuity of the sample function of $\{X_t\}$ over the collection $\mathbb{Q}_1^{(2)} = \bigcup_{\ell \geq 1} \mathbb{Q}_1^{(2,\ell)}$ of dyadic rationals in $[0,1]$, where $\mathbb{Q}_T^{(2,\ell)} = \{j2^{-\ell} \leq T, j \in \mathbb{Z}_+\}$. To this end, fixing $\gamma < \beta/\alpha$ and considering (7.2.2) for $\varepsilon = 2^{-\gamma\ell}$, we have by finite sub-additivity that

$$\mathbf{P}(\max_{j=0}^{2^\ell - 1} |X_{(j+1)2^{-\ell}} - X_{j2^{-\ell}}| \geq 2^{-\gamma\ell}) \leq c2^{-\ell\eta}\,,$$

for $\eta = \beta - \alpha\gamma > 0$. Since $\sum_\ell 2^{-\ell\eta}$ is finite, it then follows by Borel-Cantelli I that

$$\max_{j=0}^{2^\ell - 1} |X_{(j+1)2^{-\ell}} - X_{j2^{-\ell}}| < 2^{-\gamma\ell}\,, \quad \forall \ell \geq n_\gamma(\omega)\,,$$

where $n_\gamma(\omega)$ is finite for all $\omega \notin N_\gamma$ and $N_\gamma \in \mathcal{F}$ has zero probability.

As you show in Exercise 7.2.7 this implies the local γ-Hölder continuity of $t \mapsto X_t(\omega)$ over the dyadic rationals. That is,

$$|X_t(\omega) - X_s(\omega)| \le c(\gamma)|t-s|^{\gamma}\,, \tag{7.2.3}$$

for $c(\gamma) = 2/(1-2^{-\gamma})$ finite and any $t, s \in \mathbb{Q}_1^{(2)}$ such that $|t-s| < h_\gamma(\omega) = 2^{-n_\gamma(\omega)}$.

Turning to construct the S.P. $\{\widetilde{X}_t, t \in \mathbb{T}\}$, we fix $\gamma_k \uparrow \beta/\alpha$ and set $N_\star = \cup_k N_{\gamma_k}$. Considering the R.V. $\widetilde{X}_s(\omega) = X_s(\omega) I_{N_\star^c}(\omega)$ for $s \in \mathbb{Q}_1^{(2)}$, we further set $\widetilde{X}_t = \lim_{n\to\infty} \widetilde{X}_{s_n}$ for some non-random $\{s_n\} \subset \mathbb{Q}_1^{(2)}$ such that $s_n \to t \in [0,1] \setminus \mathbb{Q}_1^{(2)}$. Indeed, in view of (7.2.3), by the uniform continuity of $s \mapsto \widetilde{X}_s(\omega)$ over $\mathbb{Q}_1^{(2)}$, the sequence $n \mapsto \widetilde{X}_{s_n}(\omega)$ is Cauchy, hence convergent, per $\omega \in \Omega$. By construction, the S.P. $\{\widetilde{X}_t, t \in [0,1]\}$ is such that

$$|\widetilde{X}_t(\omega) - \widetilde{X}_s(\omega)| \le c_k |t-s|^{\gamma_k}\,,$$

for any k and $t, s \in [0,1]$ such that $|t-s| < \widetilde{h}_k(\omega)$, where $\widetilde{h}_k = I_{N_\star} + I_{N_\star^c} h_{\gamma_k}$ is positive for all $\omega \in \Omega$ and $c_k = c(\gamma_k)$ is finite. That is, $\{\widetilde{X}_t\}$ is locally γ-Hölder continuous on $\mathbb{T} = [0,1]$ for any γ_k, hence also for all $\gamma < \beta/\alpha$ (and in particular, $\{\widetilde{X}_t, t \in [0,1]\}$ has continuous sample functions).

It thus remains only to verify that $\{\widetilde{X}_t\}$ is a modification of $\{X_t\}$. To this end, observe first that since $\mathbf{P}(N_{\gamma_k}) = 0$ for all k, also $\mathbf{P}(N_\star) = 0$. Further, $\widetilde{X}_s(\omega) = X_s(\omega)$ for all $s \in \mathbb{Q}_1^{(2)}$ and $\omega \notin N_\star$. Next, from (7.2.2) we have that $\mathbf{P}(|X_t - X_{s_n}| \ge \varepsilon) \to 0$ for any fixed $\varepsilon > 0$ and $s_n \to t$, that is, $X_{s_n} \xrightarrow{p} X_t$. Hence, recall Theorem 2.2.10, also $X_{s_{n(k)}} \xrightarrow{a.s.} X_t$ along some subsequence $k \mapsto n(k)$. Considering an arbitrary $t \in [0,1] \setminus \mathbb{Q}_1^{(2)}$ and the sequence $s_n \in \mathbb{Q}_1^{(2)}$ as in the construction of $\{\widetilde{X}_t\}$, we have in addition that $\widetilde{X}_{s_{n(k)}} \to \widetilde{X}_t$. Consequently, $\mathbf{P}(\widetilde{X}_t \ne X_t) \le \mathbf{P}(N_\star) = 0$ from which we conclude that $\{\widetilde{X}_t\}$ is a modification of $\{X_t\}$ on $\mathbb{T} = [0,1]$. $\square$

Exercise 7.2.7. *Fixing $x \in \mathbb{R}^{[0,1]}$, let*

$$\Delta_{\ell,r}(x) = \max_{j=0}^{2^\ell - r} |x((j+r)2^{-\ell}) - x(j2^{-\ell})|\,.$$

(a) *Show that for any integers $k > m \ge 0$,*

$$\sup_{\substack{t,s \in \mathbb{Q}_1^{(2,k)} \\ |t-s| < 2^{-m}}} |x(t) - x(s)| \le 2 \sum_{\ell=m+1}^{k} \Delta_{\ell,1}(x)\,.$$

Hint: *Applying induction on k consider $s < t$ and $s \le s' \le t' \le t$, where $s' = \min\{u \in \mathbb{Q}_1^{(2,k-1)} : u \ge s\}$ and $t' = \max\{u \in \mathbb{Q}_1^{(2,k-1)} : u \le t\}$.*

(b) *Fixing $\gamma > 0$, let $c_\gamma = 2/(1-2^{-\gamma})$ and deduce that if $\Delta_{\ell,1}(x) \le 2^{-\gamma\ell}$ for all $\ell \ge n$, then*

$$|x(t) - x(s)| \le c_\gamma |t-s|^\gamma \quad \textit{for all } t, s \in \mathbb{Q}_1^{(2)} \textit{ such that } |t-s| < 2^{-n}\,.$$

Hint: *Apply part (a) for $m \ge n$ such that $2^{-(m+1)} \le |t-s| < 2^{-m}$.*

We next identify the restriction of the cylindrical σ-algebra of $\mathbb{R}^{\mathbb{T}}$ to $C(\mathbb{T})$ as the Borel σ-algebra on the space of continuous functions, starting with $\mathbb{T} = \mathbb{I}^r$ for $\mathbb{I}$ a compact interval.

Lemma 7.2.8. *For $\mathbb{T} = \mathbb{I}^r$ and $\mathbb{I} \subset \mathbb{R}$ a compact interval, consider the topological space $(C(\mathbb{T}), \|\cdot\|_\infty)$ of continuous functions on $\mathbb{T}$, equipped with the topology induced by the supremum norm $\|x\|_\infty = \sup_{t\in\mathbb{T}} |x(t)|$. The corresponding Borel σ-algebra, denoted hereafter $\mathcal{B}_{C(\mathbb{T})}$ coincides with $\{A \cap C(\mathbb{T}) : A \in \mathcal{B}^{\mathbb{T}}\}$.*

Proof. Recall that for any $z \in C(\mathbb{T})$,

$$\|z\|_\infty = \sup_{t\in\mathbb{T}\cap\mathbb{Q}^r} |z(t)| \,.$$

Hence, each open ball

$$B(x, r) = \{y \in C(\mathbb{T}) : \|y - x\|_\infty < r\}$$

in $\mathbb{S} = (C(\mathbb{T}), \|\cdot\|_\infty)$ is the countable intersection of $R_t \cap C(\mathbb{T})$ for the corresponding one dimensional measurable rectangles $R_t \in \mathcal{B}^{\mathbb{T}}$ indexed by $t \in \mathbb{T}\cap\mathbb{Q}^r$. Consequently, each open ball $B(x, r)$ is in the σ-algebra $\mathcal{C} = \{A \cap C(\mathbb{T}) : A \in \mathcal{B}^{\mathbb{T}}\}$. With Γ denoting a countable dense subset of the separable metric space $\mathbb{S}$, it readily follows that $\mathbb{S}$ has a countable base $\mathcal{U}$, consisting of the balls $B(x, 1/n)$ for positive integers n and centers $x \in \Gamma$. With every open set thus being a countable union of elements from $\mathcal{U}$, it follows that $\mathcal{B}_{\mathbb{S}} = \sigma(\mathcal{U})$. Further, $\mathcal{U} \subseteq \mathcal{C}$, hence also $\mathcal{B}_{\mathbb{S}} \subseteq \mathcal{C}$.

Conversely, recall that $\mathcal{C} = \sigma(\mathcal{O})$ for the collection $\mathcal{O}$ of sets of the form

$$O = \{x \in C(\mathbb{T}) : x(t_i) \in O_i, i = 1, \ldots, n\} \,,$$

with n finite, $t_i \in \mathbb{T}$ and open $O_i \subseteq \mathbb{R}$, $i = 1, \ldots, n$. Clearly, each $O \in \mathcal{O}$ is an open subset of $\mathbb{S}$ and it follows that $\mathcal{C} \subseteq \mathcal{B}_{\mathbb{S}}$. □

In the next exercise, you adapt the proof of Lemma 7.2.8 for $\mathbb{T} = [0, \infty)$ (and the same would apply for $\mathbb{T} \subseteq \mathbb{R}^r$ which is the product of one-dimensional intervals).

Exercise 7.2.9. *For $\mathbb{T} = [0, \infty)$, equip the set $C(\mathbb{T})$ of continuous functions on $\mathbb{T}$ with the topology of uniform convergence on compact subsets of $\mathbb{T}$. Show that the corresponding Borel σ-algebra $\mathcal{B}_{C(\mathbb{T})}$ coincides with $\{A \cap C(\mathbb{T}) : A \in \mathcal{B}^{\mathbb{T}}\}$.*
Hint: *Uniform convergence on compacts is equivalent to convergence in the complete, separable metric space $\mathbb{S} = (C([0, \infty)), \rho(\cdot, \cdot))$, where $\rho(x, y) = \sum_{j=1}^\infty 2^{-j}\varphi(\|x - y\|_j)$ for $\|x\|_t = \sup_{s\in[0,t]} |x(s)|$ and $\varphi(r) = r/(1 + r)$ (c.f.* [**Dud89**, Page 355]*).*

Combining Proposition 7.1.8, Exercise 7.2.3, Theorem 7.2.6 and Lemma 7.2.8, yields the following useful canonical construction for continuous-time processes of a.s. continuous sample path.

Corollary 7.2.10. *Given a consistent collection of f.d.d. indexed on $\mathbb{T} = \mathbb{I}^r$ (with $\mathbb{I} \subseteq \mathbb{R}$ a compact interval), such that (7.2.1) holds (for some positive α, β and finite c), there exists a S.P. $\widetilde{X}_\cdot(\omega) : \Omega \mapsto (C(\mathbb{T}), \|\cdot\|_\infty)$, measurable with respect to $\mathcal{B}_{C(\mathbb{T})}$, which has the specified f.d.d. and is indistinguishable from any of its continuous modifications.*

Remark. An alternative approach is to directly construct the sample functions of stochastic processes of interest. That is, to view the process from the start as a random variable $\omega \mapsto X_\cdot(\omega)$ taking values in certain topological space of functions equipped with its Borel σ-algebra (for example, the space $C(\mathbb{T})$ with a suitable metric). In dealing with the Brownian motion, we pursue both approaches, first relying on the canonical construction (namely, Corollary 7.2.10), and then proving instead an *invariance principle* via weak convergence in $C(\mathbb{T})$ (c.f. Section 9.2).

In contrast with Theorem 7.2.6, here is an example of a S.P. with no continuous modification, for which (7.2.1) holds with $\beta = 0$.

Example 7.2.11. *Consider the S.P. $X_t(\omega) = I_{\{\omega > t\}}$, for $t \in [0,1]$ and the uniform probability measure on $\Omega = (0,1]$. Then, $\mathbf{E}[|X_t - X_s|^\alpha] = U((s,t]) = |t-s|$ for all $0 < s < t \le 1$, so $\{X_t, t \in [0,1]\}$ satisfies (7.2.1) with $c = 1$, $\beta = 0$ and any $\alpha > 0$. However, if $\{\widetilde{X}_t\}$ is a modification of $\{X_t\}$ then a.s. $\widetilde{X}_t(\omega) = X_t(\omega)$ at all $t \in (0,1] \cap \mathbb{Q}$, from which it follows that $s \mapsto \widetilde{X}_s(\omega)$ is discontinuous at $s = \omega$.*

While direct application of Theorem 7.2.6 is limited to (locally γ-Hölder) continuous modifications on *compact* intervals, say $[0,T]$, it is easy to combine these to *one* (locally γ-Hölder) continuous modification, valid on $[0,\infty)$.

Lemma 7.2.12. *Suppose there exist $T_n \uparrow \infty$ such that the continuous time S.P. $\{X_t, t \ge 0\}$ has (locally γ-Hölder) continuous modifications $\{\widetilde{X}^{(n)}_t, t \in [0,T_n]\}$. Then, the S.P. $\{X_t, t \ge 0\}$ also has such modification on $[0,\infty)$.*

Proof. By assumption, for each positive integer n, the event

$$A_n = \{\omega : \widetilde{X}^{(n)}_t(\omega) = X_t(\omega), \quad \forall t \in \mathbb{Q} \cap [0,T_n]\}\,,$$

has probability one. The event $A_\star = \cap_n A_n$ of probability one is then such that $\widetilde{X}^{(n)}_t(\omega) = \widetilde{X}^{(m)}_t(\omega)$ for all $\omega \in A_\star$, positive integers n,m and any $t \in \mathbb{Q} \cap [0, T_n \wedge T_m]$. By continuity of $t \mapsto \widetilde{X}^{(n)}_t(\omega)$ and $t \mapsto \widetilde{X}^{(m)}_t(\omega)$ it follows that for all $\omega \in A_\star$,

$$\widetilde{X}^{(n)}_t(\omega) = \widetilde{X}^{(m)}_t(\omega)\,, \qquad \forall n,m,\, t \in [0, T_n \wedge T_m]\,.$$

Consequently, for such ω there exists a function $t \mapsto \widetilde{X}_t(\omega)$ on $[0,\infty)$ that coincides with each of the functions $\widetilde{X}^{(n)}_t(\omega)$ on its interval of definition $[0,T_n]$. By assumption the latter are (locally γ-Hölder) continuous, so the same applies for the sample function $t \mapsto \widetilde{X}_t(\omega)$ on $[0,\infty)$. Setting $\widetilde{X}_t(\omega) \equiv 0$ in case $\omega \notin A_\star$ completes the construction of the S.P. $\{\widetilde{X}_t, t \ge 0\}$ with (locally γ-Hölder) continuous sample functions, such that for any $t \in [0,T_n]$,

$$\mathbf{P}(X_t \ne \widetilde{X}_t) \le \mathbf{P}(A^c_\star) + \mathbf{P}(X_t \ne \widetilde{X}^{(n)}_t) = 0\,.$$

Since $T_n \to \infty$, we conclude that this S.P. is a (locally γ-Hölder) continuous modification of $\{X_t, t \ge 0\}$. $\square$

The following application of Kolmogorov-Centsov theorem demonstrates the importance of its free parameter α.

Exercise 7.2.13. *Suppose $\{X_t, t \in \mathbb{I}\}$ is a continuous time S.P. such that $\mathbf{E}(X_t) = 0$ and $\mathbf{E}(X_t^2) = 1$ for all $t \in \mathbb{I}$, a compact interval on the line.*

(a) *Show that if for some finite c, $p > 1$ and $h > 0$,*

$$\mathbf{E}[X_t X_s] \ge 1 - c(t-s)^p \qquad \text{for all } s < t \le s+h,\ t,s \in \mathbb{I}\,, \tag{7.2.4}$$

then there exists a continuous modification of $\{X_t, t \in \mathbb{I}\}$ which is also locally γ-Hölder continuous, for $\gamma < (p-1)/2$.

(b) *Show that if (X_s, X_t) is a multivariate normal for each $t > s$, then it suffices for the conclusion of part (a) to have $\mathbf{E}[X_t X_s] \ge 1 - c(t-s)^{p-1}$ instead of (7.2.4).*

Hint: *In part (a) use $\alpha = 2$ while for part (b) try $\alpha = 2k$ and $k \gg 1$.*

Example 7.2.14. *There exist S.P.-s satisfying (7.2.4) with $p = 1$ for which there is no continuous modification. One such process is the* random telegraph signal $R_t = (-1)^{N_t} R_0$, *where* $\mathbf{P}(R_0 = 1) = \mathbf{P}(R_0 = -1) = 1/2$ *and* R_0 *is independent of the Poisson process* $\{N_t\}$ *of rate one. The process* $\{R_t\}$ *alternately jumps between* -1 *and* $+1$ *at the random jump times* $\{T_k\}$ *of the Poisson process* $\{N_t\}$. *Hence, by the same argument as in Example 7.2.11 it does not have a continuous modification. Further, for any* $t > s \geq 0$,

$$\mathbf{E}[R_s R_t] = 1 - 2\mathbf{P}(R_s \neq R_t) \geq 1 - 2\mathbf{P}(N_s < N_t) \geq 1 - 2(t-s)\,,$$

so $\{R_t\}$ *satisfies (7.2.4) with* $p = 1$ *and* $c = 2$.

Remark. The S.P. $\{R_t\}$ of Example 7.2.14 is a special instance of the continuous-time *Markov jump processes*, which we study in Section 8.3.3. Though the sample function of this process is a.s. discontinuous, it has the following RCLL property, as is the case for all continuous-time Markov jump processes.

Definition 7.2.15. *Given a countable* $\mathbb{C} \subset \mathbb{I}$ *we say that a function* $x \in \mathbb{R}^{\mathbb{I}}$ *is* $\mathbb{C}$-separable at t *if there exists a sequence* $s_k \in \mathbb{C}$ *that converges to* t *such that* $x(s_k) \to x(t)$. *If this holds* at all $t \in \mathbb{I}$, *we call* $x(\cdot)$ *a* $\mathbb{C}$-separable function. *A continuous time S.P.* $\{X_t, t \in \mathbb{I}\}$ *is* separable *if there exists a non-random, countable* $\mathbb{C} \subset \mathbb{I}$ *such that* all *sample functions* $t \mapsto X_t(\omega)$ *are* $\mathbb{C}$*-separable. Such a process is further* right-continuous with left-limits *(in short, RCLL), if the sample function* $t \mapsto X_t(\omega)$ *is right-continuous and of left-limits at any* $t \in \mathbb{I}$ *(that is, for* $h \downarrow 0$ *both* $X_{t+h}(\omega) \to X_t(\omega)$ *and the limit of* $X_{t-h}(\omega)$ *exists). Similarly, a modification which is a separable S.P. or one having RCLL sample functions is called a* separable modification, *or* RCLL modification *of the S.P., respectively. As usual, suffices to have any of these properties w.p.1 (for we do not differentiate between a pair of indistinguishable S.P.).*

Remark. Clearly, a S.P. of continuous sample functions is also RCLL and a S.P. having right-continuous sample functions (in particular, any RCLL process), is further separable. To summarize,

$$\text{Hölder continuity} \;\Rightarrow\; \text{Continuity} \;\Rightarrow\; \text{RCLL} \;\Rightarrow\; \text{Separable}$$

But, the S.P. $\{X_t\}$ of Example 7.2.1 is non-separable. Indeed, $\mathbb{C}$-separability of $t \mapsto X_t(\omega)$ at $t = \omega$ requires that $\omega \in \mathbb{C}$, so for any countable subset $\mathbb{C}$ of $[0,1]$ we have that $\mathbf{P}(t \mapsto X_t \text{ is } \mathbb{C}\text{-separable}) \leq \mathbf{P}(\mathbb{C}) = 0$.

One motivation for the notion of separability is its prevalence. Namely, to any consistent collection of f.d.d. indexed on an interval $\mathbb{I}$, corresponds a separable S.P. with these f.d.d. This is achieved at the small cost of possibly moving from real-valued variables to $\overline{\mathbb{R}}$-valued variables (each of which is nevertheless a.s. real-valued).

Proposition 7.2.16. *Any continuous time S.P.* $\{X_t, t \in \mathbb{I}\}$ *admits a separable modification (consisting possibly of* $\overline{\mathbb{R}}$*-valued variables). Hence, to any consistent collection of f.d.d. indexed on* $\mathbb{I}$ *corresponds an* $(\overline{\mathbb{R}}^{\mathbb{I}}, (\mathcal{B}_{\overline{\mathbb{R}}})^{\mathbb{I}})$*-valued separable S.P. with these f.d.d.*

We prove this proposition following [**Bil95**, Theorem 38.1], but leave its technical engine (i.e. [**Bil95**, Lemma 1, Page 529]), as your next exercise.

Exercise 7.2.17. *Suppose* $\{Y_t, t \in \mathbb{I}\}$ *is a continuous time S.P.*

(a) *Fixing $B \in \mathcal{B}$, consider the probabilities $p(D) = \mathbf{P}(Y_s \in B$ for all $s \in D)$, for countable $D \subset \mathbb{I}$. Show that for any $A \subseteq \mathbb{I}$ there exists a countable subset $D_\star = D_\star(A, B)$ of A such that $p(D_\star) = \inf\{p(D) :$ countable $D \subset A\}$.*

Hint: *Let $D_\star = \cup_k D_k$ where $p(D_k) \le k^{-1} + \inf\{p(D) :$ countable $D \subset A\}$.*

(b) *Deduce that if $t \in A$ then $N_t(A, B) = \{\omega : Y_s(\omega) \in B$ for all $s \in D_\star(A, B)$ and $Y_t(\omega) \notin B\}$ has zero probability.*

(c) *Let $\mathbb{C}$ denote the union of $D_\star(A, B)$ over all $A = \mathbb{I} \cap (q_1, q_2)$ and $B = (q_3, q_4)^c$, with $q_i \in \mathbb{Q}$. Show that at any $t \in \mathbb{I}$ there exists $N_t \in \mathcal{F}$ such that $\mathbf{P}(N_t) = 0$ and the sample functions $t \mapsto Y_t(\omega)$ are $\mathbb{C}$-separable at t for every $\omega \notin N_t$.*

Hint: *Let N_t denote the union of $N_t(A, B)$ over the sets (A, B) as in the definition of $\mathbb{C}$, such that further $t \in A$.*

Proof. Assuming first that $\{Y_t, t \in \mathbb{I}\}$ is a $(0, 1)$-valued S.P. set $\widetilde{Y}_\cdot = Y_\cdot$ on the countable, dense $\mathbb{C} \subseteq \mathbb{I}$ of part (c) of Exercise 7.2.17. Then, fixing non-random $\{s_n\} \subseteq \mathbb{C}$ such that $s_n \to t \in \mathbb{I} \setminus \mathbb{C}$ we define the R.V.-s

$$\widetilde{Y}_t = Y_t I_{N_t^c} + I_{N_t} \limsup_{n\to\infty} Y_{s_n} ,$$

for the events N_t of zero probability from part (c) of Exercise 7.2.17. The resulting S.P. $\{\widetilde{Y}_t, t \in \mathbb{I}\}$ is a $[0, 1]$-valued modification of $\{Y_t\}$ (since $\mathbf{P}(\widetilde{Y}_t \neq Y_t) \le \mathbf{P}(N_t) = 0$ for each $t \in \mathbb{I}$). It clearly suffices to check $\mathbb{C}$-separability of $t \mapsto \widetilde{Y}_t(\omega)$ at each fixed $t \notin \mathbb{C}$ and this holds by our construction if $\omega \in N_t$ and by part (c) of Exercise 7.2.17 in case $\omega \in N_t^c$. For any $(0, 1)$-valued S.P. $\{Y_t\}$ we have thus constructed a separable $[0, 1]$-valued modification $\{\widetilde{Y}_t\}$. To handle an $\mathbb{R}$-valued S.P. $\{X_t, t \in \mathbb{I}\}$, let $\{\widetilde{Y}_t, t \in \mathbb{I}\}$ denote the $[0, 1]$-valued, separable modification of the $(0, 1)$-valued S.P. $Y_t = F_G(X_t)$, with $F_G(\cdot)$ denoting the standard normal distribution function. Since $F_G(\cdot)$ has a continuous inverse $F_G^{-1} : [0, 1] \mapsto \overline{\mathbb{R}}$ (where $F_G^{-1}(0) = -\infty$ and $F_G^{-1}(1) = \infty$), it directly follows that $\widetilde{X}_t = F_G^{-1}(\widetilde{Y}_t)$ is an $\overline{\mathbb{R}}$-valued separable modification of the S.P. $\{X_t\}$. $\square$

Here are few elementary and useful consequences of separability.

Exercise 7.2.18. *Suppose S.P. $\{X_t, t \in \mathbb{I}\}$ is $\mathbb{C}$-separable and $\mathbb{J} \subseteq \mathbb{I}$ with $\mathbb{J}$ an open interval.*

(a) *Show that*

$$\sup_{t\in\mathbb{J}} X_t = \sup_{t\in\mathbb{J}\cap\mathbb{C}} X_t$$

is in $m\mathcal{F}^{\mathbf{X}}$, hence its law is determined by the f.d.d.

(b) *Similarly, show that for any $h > 0$ and $s \in \mathbb{I}$,*

$$\sup_{t\in[s,s+h)} |X_t - X_s| = \sup_{t\in[s,s+h)\cap\mathbb{C}} |X_t - X_s|$$

is in $m\mathcal{F}^{\mathbf{X}}$ with its law determined by the f.d.d.

The joint measurability of sample functions is an important property to have.

Definition 7.2.19. *A continuous time S.P. $\{X_t, t \in \mathbb{I}\}$ is* measurable *if $X_t(\omega) : \mathbb{I} \times \Omega \mapsto \mathbb{R}$ is measurable with respect to $\overline{\mathcal{B}}_{\mathbb{I}} \times \overline{\mathcal{F}}$ (that is, for any $B \in \mathcal{B}$, the subset $\{(t, \omega) : X_t(\omega) \in B\}$ of $\mathbb{I} \times \Omega$ is in $\overline{\mathcal{B}}_{\mathbb{I}} \times \overline{\mathcal{F}}$, where as usual $\overline{\mathcal{B}}_{\mathbb{I}}$ denotes the* *completion of the Borel σ-algebra with respect to Lebesgue's measure on $\mathbb{I}$ and $\overline{\mathcal{F}}$ is the completion of $\mathcal{F}$ with respect to $\mathbf{P}$).*

As we show in Proposition 8.1.8, any right-continuous S.P. (and in particular, RCLL), is also measurable. While separability does not imply measurability, building on the obvious measurability of (simple) RCLL processes, following the proof of [**Doo53**, Theorem II.2.6] we show next that to any consistent and continuous in probability collection of f.d.d. corresponds a *both separable and measurable* S.P. having the specified f.d.d.

Definition 7.2.20. *A S.P. $\{X_t, t \in \mathbb{I}\}$ is* continuous in probability *if for any $t \in \mathbb{I}$ and $\varepsilon > 0$,*

$$\lim_{s\to t} \mathbf{P}(|X_s - X_t| > \varepsilon) = 0 \,.$$

Remark. Continuity in probability is a very mild property, which is completely determined by the f.d.d. and has little to do with the sample functions of the process. For example, note that the Poisson process is continuous in probability, as are the random telegraph noise $\{R_t\}$ of Example 7.2.14 and even the non-separable S.P. $\{X_t\}$ of Example 7.2.1.

Proposition 7.2.21. *Any continuous in probability process $\{X_t, t \in \mathbb{I}\}$ has an $(\overline{\mathbb{R}}^{\mathbb{I}}, (\mathcal{B}_{\overline{\mathbb{R}}})^{\mathbb{I}})$-valued separable modification which is further a measurable process.*

Proof. It suffices to consider $\mathbb{I} = [0, 1]$. Indeed, by an affine time change the same proof then applies for any compact interval $\mathbb{I}$, and if $\mathbb{I}$ is unbounded, simply decompose it to countably many disjoint bounded intervals and glue together the corresponding separable and measurable modifications of the given process.

Further, in view of Proposition 7.2.16 and the transformation via $F_G(\cdot)$ we have utilized in its proof, we consider with no loss of generality a $(0, 1)$-valued $\{s_k\}$-separable, continuous in probability S.P. $\{Y_t, t \in [0, 1]\}$ and provide a $[0, 1]$-valued measurable modification $\{\widetilde{Y}_t\}$ of $\{Y_t\}$, which we then verify to be also a separable process. To this end, with no loss of generality, assume further that $s_1 = 0$. Then, for any $n \in \mathbb{N}$ set $t_{n+1} = 2$ and with $0 = t_1 < \cdots < t_n$ the monotone increasing rearrangement of $\{s_k, k = 1, \ldots, n\}$, consider the $[0, 1]$-valued, RCLL stochastic process

$$Y_t^{(n)} = \sum_{j=1}^{n} Y_{t_j} I_{[t_j, t_{j+1})}(t) \,,$$

which is clearly also a measurable S.P. By the denseness of $\{s_k\}$ in $[0, 1]$, it follows from the continuity in probability of $\{Y_t\}$ that $Y_t^{(n)} \xrightarrow{p} Y_t$ as $n \to \infty$, for any fixed $t \in [0, 1]$. Hence, by bounded convergence $\mathbf{E}[|Y_t^{(n)} - Y_t^{(m)}|] \to 0$ as $n, m \to \infty$ for each $t \in [0, 1]$. Then, by yet another application of bounded convergence

$$\lim_{m,n\to\infty} \mathbf{E}[|Y_T^{(n)} - Y_T^{(m)}|] = 0 \,,$$

where the R.V. $T \in [0, 1]$ is chosen independently of $\mathbf{P}$, according to the uniform probability measure U corresponding to Lebesgue's measure $\overline{\lambda}(\cdot)$ restricted to $([0, 1], \overline{\mathcal{B}}_{[0,1]})$. By Fubini's theorem, this amounts to $\{Y_t^{(n)}(\omega)\}$ being a Cauchy, hence convergent, sequence in $L^1([0, 1] \times \Omega, \overline{\mathcal{B}}_{[0,1]} \times \overline{\mathcal{F}}, U \times \mathbf{P})$ (recall Proposition 4.3.7 that the latter is a Banach space). In view of Theorem 2.2.10, upon passing to a suitable subsequence n_j we thus have that $(t, \omega) \mapsto Y_t^{(n_j)}(\omega)$ converges to some

$\overline{\mathcal{B}}_{[0,1]} \times \overline{\mathcal{F}}$-measurable function $(t,\omega) \mapsto Y_t^{(\infty)}(\omega)$ for all $(t,\omega) \notin N$, where we may and shall assume that $\{s_k\} \times \Omega \subseteq N \in \overline{\mathcal{B}}_{[0,1]} \times \overline{\mathcal{F}}$ and $U \times \mathbf{P}(N) = 0$. Taking now

$$\widetilde{Y}_t(\omega) = I_{N^c}(t,\omega) Y_t^{(\infty)}(\omega) + I_N(t,\omega) Y_t(\omega)\,,$$

note that $\widetilde{Y}_t(\omega) = Y_t^{(\infty)}(\omega)$ for a.e. (t,ω), so with $Y_t^{(\infty)}(\omega)$ a measurable process, by the completeness of our product σ-algebra, the S.P. $\{\widetilde{Y}_t, t \in [0,1]\}$ is also measurable. Further, fixing $t \in [0,1]$, if $\{\widetilde{Y}_t(\omega) \neq Y_t(\omega)\}$ then $\omega \in A_t = \{\omega : Y_t^{(n_j)}(\omega) \to Y_t^{(\infty)}(\omega) \neq Y_t(\omega)\}$. But, recall that $Y_t^{(n_j)} \xrightarrow{p} Y_t$ for all $t \in [0,1]$, hence $\mathbf{P}(A_t) = 0$, i.e. $\{\widetilde{Y}_t, t \in [0,1]\}$ is a modification of the given process $\{Y_t, t \in [0,1]\}$.

Finally, since $\{\widetilde{Y}_t\}$ coincides with the $\{s_k\}$-separable S.P. $\{Y_t\}$ on the set $\{s_k\}$, the sample function $t \mapsto \widetilde{Y}_t(\omega)$ is, by our construction, $\{s_k\}$-separable at any $t \in [0,1]$ such that $(t,\omega) \in N$. Moreover, $Y_t^{(n_j)} = Y_{s_k} = \widetilde{Y}_{s_k}$ for some $k = k(j,t)$, with $s_{k(j,t)} \to t$ by the denseness of $\{s_k\}$ in $[0,1]$. Hence, if $(t,\omega) \notin N$ then

$$\widetilde{Y}_t(\omega) = \lim_{j\to\infty} Y_t^{(n_j)}(\omega) = \lim_{j\to\infty} \widetilde{Y}_{s_{k(j,t)}}(\omega)\,.$$

Thus, $\{\widetilde{Y}_t, t \in [0,1]\}$ is $\{s_k\}$-separable and as claimed, it is a separable, measurable modification of $\{Y_t, t \in [0,1]\}$. $\square$

Recall (1.4.7) that the measurability of the process, namely of $(t,\omega) \mapsto X_t(\omega)$, implies that all its sample functions $t \mapsto X_t(\omega)$ are Lebesgue measurable functions on $\mathbb{I}$. Measurability of a S.P. also results with well defined integrals of its sample function. For example, if a Borel function $h(t,x)$ is such that $\int_{\mathbb{I}} \mathbf{E}[|h(t,X_t)|]dt$ is finite, then by Fubini's theorem $t \mapsto \mathbf{E}[h(t,X_t)]$ is in $L^1(\mathbb{I}, \mathcal{B}_{\mathbb{I}}, \lambda)$, the integral $\int_{\mathbb{I}} h(s,X_s)ds$ is an a.s. finite R.V. and

$$\int_{\mathbb{I}} \mathbf{E}[h(s,X_s)]\; ds = \mathbf{E}[\int_{\mathbb{I}} h(s,X_s)\; ds]\,.$$

Conversely, as you are to show next, under mild conditions the differentiability of sample functions $t \mapsto X_t$ implies the differentiability of $t \mapsto \mathbf{E}[X_t]$.

Exercise 7.2.22. *Suppose each sample function $t \mapsto X_t(\omega)$ of a continuous time S.P. $\{X_t, t \in \mathbb{I}\}$ is differentiable at any $t \in \mathbb{I}$.*

(a) *Verify that $\frac{\partial}{\partial t} X_t$ is a random variable for each fixed $t \in \mathbb{I}$.*

(b) *Show that if $|X_t - X_s| \le |t-s|Y$ for some integrable random variable Y, a.e. $\omega \in \Omega$ and all $t,s \in \mathbb{I}$, then $t \mapsto \mathbf{E}[X_t]$ has a finite derivative and for any $t \in \mathbb{I}$,*

$$\frac{d}{dt}\mathbf{E}(X_t) = \mathbf{E}\Big(\frac{\partial}{\partial t} X_t\Big).$$

We next generalize the lack of correlation of independent R.V. to the setting of continuous time S.P.-s.

Exercise 7.2.23. *Suppose square-integrable, continuous time S.P.-s $\{X_t, t \in \mathbb{I}\}$ and $\{Y_t, t \in \mathbb{I}\}$ are* **P**-independent. *That is, both processes are defined on the same probability space and the σ-algebras $\mathcal{F}^{\mathbf{X}}$ and $\mathcal{F}^{\mathbf{Y}}$ are* **P***-independent. Show that in this case,*

$$\mathbf{E}[X_t Y_t | \mathcal{F}_s^{\mathbf{Z}}] = \mathbf{E}[X_t | \mathcal{F}_s^{\mathbf{X}}]\mathbf{E}[Y_t | \mathcal{F}_s^{\mathbf{Y}}]\,,$$

for any $s \le t \in \mathbb{I}$, where $Z_t = (X_t, Y_t) \in \mathbb{R}^2$, $\mathcal{F}_s^{\mathbf{X}} = \sigma(X_u, u \in \mathbb{I}, u \le s)$ and $\mathcal{F}_s^{\mathbf{Y}}$, $\mathcal{F}_s^{\mathbf{Z}}$ are similarly defined.

7.3. Gaussian and stationary processes

Building on Definition 3.5.13 of Gaussian random vectors, we have the following important class of (centered) Gaussian (stochastic) processes, which plays a key role in our construction of the Brownian motion.

Definition 7.3.1. *A S.P. $\{X_t,\ t \in \mathbb{T}\}$ is a* Gaussian process *(or Gaussian S.P.), if $(X_{t_1}, \dots, X_{t_n})$ is a Gaussian random vector for any n finite and $t_k \in \mathbb{T}$, $k = 1, \dots, n$. Alternatively, a S.P. is Gaussian if and only if it has multivariate normal f.d.d. We further say that a Gaussian S.P. is* centered *if its* mean *function $m(t) = \mathbf{E}[X_t]$ is zero.*

Recall the following notion of non-negative definiteness, based on Definition 3.5.12.

Definition 7.3.2. *A symmetric function $c(t,s) = c(s,t)$ on a product set $\mathbb{T} \times \mathbb{T}$ is called* non-negative definite *(or positive semidefinite) if for any finite n and $t_k \in \mathbb{T}$, $k = 1, \dots, n$, the $n \times n$ matrix of entries $c(t_j, t_k)$ is non-negative definite. That is, for any $a_k \in \mathbb{R}$, $k = 1, \dots, n$,*

$$\sum_{j=1}^{n} \sum_{k=1}^{n} a_j c(t_j, t_k) a_k \geq 0. \tag{7.3.1}$$

Example 7.3.3. *Note that the* auto-covariance *function $c(t,s) = \mathsf{Cov}(X_t, X_s)$ of a square-integrable S.P. $\{X_t, t \in \mathbb{T}\}$ is non-negative definite. Indeed, the left side of (7.3.1) is in this case precisely the non-negative $\mathsf{Var}(\sum_{j=1}^{n} a_j X_{t_j})$.*

Convince yourself that non-negative definiteness is the only property that the auto-covariance function of a Gaussian S.P. must have and further that the following is an immediate corollary of the canonical construction and the definitions of Gaussian random vectors and stochastic processes.

Exercise 7.3.4.

(a) *Show that for any index set $\mathbb{T}$, the law of a Gaussian S.P. is uniquely determined by its mean and auto-covariance functions.*
(b) *Show that a Gaussian S.P. exists for any mean function and any non-negative definite auto-covariance function.*

Remark. An interesting consequence of Exercise 7.3.4 is the existence of an *isonormal process* on any vector space $\mathbb{H}$ equipped with an inner product as in Definition 4.3.5. That is, a centered Gaussian process $\{X_h, h \in \mathbb{H}\}$ indexed by elements of $\mathbb{H}$ whose auto-covariance function is given by the inner product $(h_1, h_2) : \mathbb{H} \times \mathbb{H} \mapsto \mathbb{R}$. Indeed, the latter is non-negative definite on $\mathbb{H} \times \mathbb{H}$ since for $h = \sum_{j=1}^{n} a_j h_j \in \mathbb{H}$,

$$\sum_{j=1}^{n} \sum_{k=1}^{n} a_j (h_j, h_k) a_k = (h, h) \geq 0\,.$$

One of the useful properties of Gaussian processes is their closure with respect to L^2-convergence (as a consequence of Proposition 3.5.15).

Proposition 7.3.5. *If the S.P. $\{X_t,\ t \in \mathbb{T}\}$ and the Gaussian S.P. $\{X_t^{(k)},\ t \in \mathbb{T}\}$ are such that $\mathbf{E}[(X_t - X_t^{(k)})^2] \to 0$ as $k \to \infty$, for each fixed $t \in \mathbb{T}$, then $\{X_t, t \in \mathbb{T}\}$ is a Gaussian S.P. whose mean and auto-covariance functions are the pointwise limits of those for the processes $\{X_t^{(k)}, t \in \mathbb{T}\}$.*

PROOF. Fix n finite and $t_k \in \mathbb{T}$, $k = 1, \ldots, n$. Applying Proposition 3.5.15 for the sequence of Gaussian random vectors $\underline{X}_k = (X^{(k)}_{t_1}, \ldots, X^{(k)}_{t_n})$, we deduce that $\underline{X}_\infty = (X_{t_1}, \ldots, X_{t_n})$ is also a Gaussian random vector whose mean and covariance parameters $(\underline{\mu}, \mathbf{V})$ are the element-wise limits of the parameters of the sequence of random vectors $\{\underline{X}_k\}$. With this holding for all f.d.d. of the S.P. $\{X_t,\ t \in \mathbb{T}\}$, by Definition 7.3.1 the latter is a Gaussian S.P. (of the stated mean and auto-correlation functions). $\square$

Recall Exercise 5.1.9, that for a Gaussian random vector $(Y_{t_2}-Y_{t_1}, \ldots, Y_{t_n}-Y_{t_{n-1}})$, with n finite and $t_1 < t_2 < \cdots < t_n$, having independent coordinates is equivalent to having uncorrelated coordinates. Hence, from Exercise 7.1.12 we deduce that

Corollary 7.3.6. *A continuous time, Gaussian S.P. $\{Y_t, t \in \mathbb{I}\}$ has independent increments if and only if $\mathsf{Cov}(Y_t - Y_u, Y_s) = 0$ for all $s \le u < t \in \mathbb{I}$.*

Remark. Check that the zero covariance condition in this corollary is equivalent to the Gaussian process having auto-covariance function of the form $c(t, s) = g(t \wedge s)$.

Recall Definition 6.1.20 that a discrete time stochastic process $\{X_n\}$ with a $\mathcal{B}$-isomorphic state space $(\mathbb{S}, \mathcal{S})$, is (strictly) *stationary* if its law $\mathcal{P}_\mathbf{X}$ is shift invariant, namely, $\mathcal{P}_\mathbf{X} \circ \theta^{-1} = \mathcal{P}_\mathbf{X}$ for the shift operator $(\theta\omega)_k = \omega_{k+1}$ on $\mathbb{S}_\infty$. This concept of invariance of the law of the process to translation of time, extends naturally to continuous time S.P.

Definition 7.3.7. *The (time) shifts $\theta_s : \mathbb{S}^{[0,\infty)} \to \mathbb{S}^{[0,\infty)}$ are defined for $s \ge 0$ via $\theta_s(x)(\cdot) = x(\cdot + s)$ and a continuous time S.P. $\{X_t, t \ge 0\}$ is called* stationary *(or strictly stationary), if its law $\mathcal{P}_\mathbf{X}$ is invariant under* any *time shift θ_s, $s \ge 0$. That is, $\mathcal{P}_\mathbf{X} \circ (\theta_s)^{-1} = \mathcal{P}_\mathbf{X}$ for* all $s \ge 0$. *For two-sided continuous time S.P. $\{X_t, t \in \mathbb{R}\}$ the definition of time shifts extends to $s \in \mathbb{R}$ and stationarity is then the invariance of the law under θ_s for any $s \in \mathbb{R}$.*

Recall Proposition 7.1.8 that the law of a continuous time S.P. is uniquely determined by its f.d.d. Consequently, such process is (strictly) stationary if and only if its f.d.d. are invariant to translation of time. That is, if and only if

$$(X_{t_1}, \ldots, X_{t_n}) \overset{\mathcal{D}}{=} (X_{t_1+s}, \ldots, X_{t_n+s}) \tag{7.3.2}$$

for any n finite and $s, t_i \ge 0$ (or for any $s, t_i \in \mathbb{R}$ in case of a two-sided continuous time S.P.). In contrast, here is a much weaker concept of stationarity.

Definition 7.3.8. *A square-integrable continuous time S.P. of constant mean function and auto-covariance function of the form $c(t, s) = r(|t - s|)$ is called* weakly stationary *(or L^2-stationary).*

Indeed, considering (7.3.2) for $n = 1$ and $n = 2$, clearly any square-integrable stationary S.P. is also weakly stationary. As you show next, the converse fails in general, but applies for all Gaussian S.P.

Exercise 7.3.9. *Show that any weakly stationary Gaussian S.P. is also (strictly) stationary. In contrast, provide an example of a (non-Gaussian) weakly stationary process which is not stationary.*

To gain more insight about stationary processes solve the following exercise.

Exercise 7.3.10. *Suppose $\{X_t, t \geq 0\}$ is a centered weakly stationary S.P. of auto-covariance function $r(t)$.*

(a) *Show that $|r(t)| \leq r(0)$ for all $t > 0$ and further, if $r(h) = r(0)$ for some $h > 0$ then $X_{t+h} \stackrel{a.s.}{=} X_t$ for each $t \geq 0$.*
(b) *Deduce that any centered, weakly stationary process of independent increments must be a modification of the trivial process having constant sample functions $X_t(\omega) = X_0(\omega)$ for all $t \geq 0$ and $\omega \in \Omega$.*

Definition 7.3.11. *We say that a continuous time S.P. $\{X_t, t \in \mathbb{I}\}$ has* stationary increments *if for $t, s \in \mathbb{I}$ the law of the increment $X_t - X_s$ depends only on $t - s$.*

We conclude this chapter with the definition and construction of the celebrated *Brownian motion* which is the most fundamental continuous time stochastic process.

Definition 7.3.12. *A S.P. $\{W_t, t \geq 0\}$ is called a* Brownian motion *(or a* Wiener process*) starting at $x \in \mathbb{R}$, if it is a Gaussian process of mean function $m(t) = x$ and auto-covariance $c(t, s) = \mathsf{Cov}(W_t, W_s) = t \wedge s$, whose sample functions $t \mapsto W_t(\omega)$ are continuous. The case of $x = 0$ is called the* standard *Brownian motion (or* standard *Wiener process).*

In addition to constructing the Brownian motion, you are to show next that it has stationary, independent increments.

Exercise 7.3.13.

(a) *Construct a continuous time Gaussian S.P. $\{B_t, t \geq 0\}$ of the mean and auto-covariance functions of Definition 7.3.12*
Hint: *Look for f.d.d. such that $B_0 = x$ and having independent increments $B_t - B_s$ of zero mean and variance $t - s$.*
(b) *Show that there exists a Wiener process, namely a continuous modification $\{W_t, t \geq 0\}$ of $\{B_t, t \geq 0\}$.*
Hint: *Try Kolmogorov-Centsov theorem for $\alpha = 4$.*
(c) *Deduce that for any T finite, the S.P. $\{W_t, t \in [0, T]\}$ can be viewed as the random variable $W_\cdot : (\Omega, \mathcal{F}) \mapsto (C([0, T]), \|\cdot\|_\infty)$, which is measurable with respect to the Borel σ-algebra on $C([0, T])$ and is further a.s. locally γ-Hölder continuous for any $\gamma < 1/2$.*
Hint: *As in Exercise 7.2.13 try $\alpha = 2k$ with $k \gg 1$ in (7.2.1).*
(d) *Show that the S.P. $\{B_t, t \geq 0\}$ is non-stationary, but it is a process of stationary, independent increments.*

Example 7.3.14. *Convince yourself that every stationary process has stationary increments while the Brownian motion of Exercise 7.3.13 is an example of a non-stationary process with stationary (independent) increments. The same phenomena applies for discrete time S.P. (in which case the symmetric* SRW *serves as an example of a non-stationary process with stationary, independent increments).*

An alternative construction of the Wiener process on $\mathbb{I} = [0, T]$ is as the infinite series

$$W_t = x + \sum_{k=0}^{\infty} a_k(t) G_k \,,$$

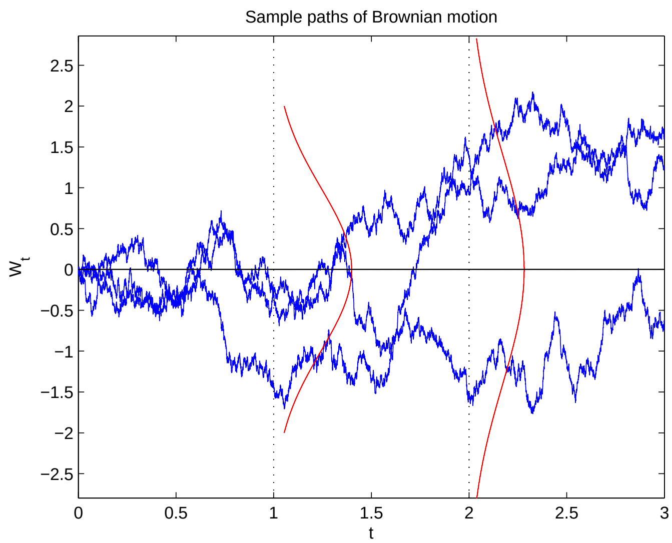

FIGURE 2. Three sample functions of Brownian motion. The density curves illustrate that the random variable W_1 has a $\mathcal{N}(0,1)$ law, while W_2 has a $\mathcal{N}(0,2)$ law.

with $\{G_k\}$ i.i.d. standard normal random variables and $a_k(\cdot)$ continuous functions on $\mathbb{I}$ such that

$$\sum_{k=0}^{\infty} a_k(t)a_k(s) = t \wedge s = \frac{1}{2}(|t+s| - |t-s|)\,. \tag{7.3.3}$$

For example, taking $T = 1/2$ and expanding $f(x) = |x|$ for $|x| \le 1$ into a Fourier series, one finds that

$$|x| = \frac{1}{2} - \sum_{k=0}^{\infty} \frac{4}{(2k+1)^2\pi^2} \cos((2k+1)\pi x)\,.$$

Hence, by the trigonometric identity $\cos(a-b) - \cos(a+b) = 2\sin(a)\sin(b)$ it follows that (7.3.3) holds for

$$a_k(t) = \frac{2}{(2k+1)\pi} \sin((2k+1)\pi t)\,.$$

Though we shall not do so, the continuity w.p.1. of $t \mapsto W_t$ is then obtained by showing that for any $\varepsilon > 0$

$$\mathbf{P}(\|\sum_{k=n}^{\infty} a_k(t)G_k\|_\infty \ge \varepsilon) \to 0$$

as $n \to \infty$ (see [**Bry95**, Theorem 8.1.3]).

We turn to explore some interesting Gaussian processes of continuous sample functions that are derived out of the Wiener process $\{W_t, t \ge 0\}$.

Exercise 7.3.15. *With* $\{W_t, t \ge 0\}$ *a* standard *Wiener process, show that each of the following is a Gaussian S.P. of continuous sample functions, compute its* *mean and auto-covariance functions and determine whether or not it is a stationary process.*

(a) *The* standard Brownian bridge $\widehat{B}_t = W_t - \min(t,1)W_1$.
(b) *The* Ornstein-Uhlenbeck process $U_t = e^{-t/2}W_{e^t}$.
(c) *The* Brownian motion with drift $Z_t^{(r,\sigma)} = \sigma W_t + rt + x$, *with non-random drift* $r \in \mathbb{R}$ *and diffusion coefficient* $\sigma > 0$.
(d) *The* integrated Brownian motion $I_t = \int_0^t W_s ds$.

Exercise 7.3.16. *Suppose* $\{W_t, t \geq 0\}$ *is a standard Wiener process.*

(a) *Compute* $\mathbf{E}(W_s|W_t)$ *and* $\mathsf{Var}(W_s|W_t)$, *first for* $s > t$, *then for* $s < t$.
(b) *Show that* $t^{-1}W_t \stackrel{a.s.}{\to} 0$ *when* $t \to \infty$.
Hint: *As we show in the sequel, the martingale* $\{W_t, t \geq 0\}$ *satisfies Doob's* L^2 *maximal inequality.*
(c) *Show that for* $t \in [0,1]$ *the S.P.* $\widetilde{B}_t = (1-t)W_{t/(1-t)}$ *(with* $\widetilde{B}_1 = 0$*), has the same law as the standard Brownian bridge and its sample functions are continuous w.p.1.*
(d) *Show that restricted to* $[0,1]$, *the law of the standard Brownian bridge matches that of* $\{W_t, t \in [0,1]$, *conditioned upon* $W_1 = 0\}$ *(hence the name Brownian bridge).*

The *fractional Brownian motion* is another Gaussian S.P. of considerable interest in financial mathematics and in the analysis of computer and queuing networks.

Exercise 7.3.17. *For* $H \in (0,1)$, *the* fractional Brownian motion *(or in short, fBM), of Hurst parameter* H *is the centered Gaussian S.P.* $\{X_t, t \geq 0\}$, *of auto-covariance function*

$$c(t,s) = \frac{1}{2}[|t|^{2H} + |s|^{2H} - |t-s|^{2H}]\ ,\ s,t \geq 0.$$

(a) *Show that the square-integrability with respect to Lebesgue's measure of* $g(u) = |1-u|^{H-1/2}\mathrm{sgn}(1-u) + |u|^{H-1/2}\mathrm{sgn}(u)$ *(which you need not verify), implies that* $c(t,s) = \int g_t(x)g_s(x)dx$ *for* $g_t(x) = \|g\|_2^{-1}|t|^{H-1/2}g(x/t)$ *in case* $t > 0$ *and* $g_0(x) = 0$.
Hint: $g_t(s+x) - g_s(s+x) = g_{t-s}(x)$, *hence* $\|g_t - g_s\|_2^2 = \|g_{t-s}\|_2^2$.
(b) *Deduce that the fBM* $\{X_t, t \geq 0\}$ *exists and has a continuous modification which is also locally* γ*-Hölder continuous for any* $0 < \gamma < H$.
(c) *Verify that for* $H = \frac{1}{2}$ *this modification is the standard Wiener process.*
(d) *Show that for any non-random* $b > 0$, *the S.P.* $\{b^{-H}X_{bt}, t \geq 0\}$ *is an fBM of the same Hurst parameter* H.
(e) *For which values of* H *are the increments of the fBM stationary and for which values are they independent?*

Exercise 7.3.18. *Let* $\mathbf{S}$ *denote the unit circle on the plane endowed with the Borel* σ*-algebra and uniform probability measure* $\mathbf{Q}$ *(which is just the image of* $[0,1]$ *equipped with Lebesgue's measure, under* $t \mapsto e^{i2\pi t} \in \mathbf{S}$*).*

Construct a centered Gaussian stochastic process $G(\cdot)$, *indexed on the collection* $\mathcal{A}$ *of sub-arcs of* $\mathbf{S}$, *of auto-covariance function* $\mathbf{E}[G(A)G(B)] = \mathbf{Q}(A \cap B) - \mathbf{Q}(A)\mathbf{Q}(B)$ *such that its sample functions* $(s,u) \mapsto G((s,u))(\omega)$ *are continuous with respect to the Euclidean topology of* $\mathbb{T} = [0,1]^2 \setminus \{(s,s) : s \in [0,1]\}$, *where* (s,u) *denotes the sub-arc* $A = \{e^{i2\pi t} : s < t \leq u\}$ *in case* $s < u$, *while* (u,s) *stands for* $A^c := \mathbf{S} \setminus A$.
Hint: *Try* $G(A) = -G(A^c) = \widehat{B}_u - \widehat{B}_s$ *for* $A = (s,u)$ *such that* $0 \leq s < u \leq 1$.

CHAPTER 8

Continuous time martingales and Markov processes

Continuous time filtrations and stopping times are introduced in Section 8.1, emphasizing the differences with the corresponding notions for discrete time processes and the connections to sample path continuity. Building upon it and Chapter 5 about discrete time martingales, we review in Section 8.2 the theory of continuous time martingales. Similarly, Section 8.3 builds upon Chapter 6 about Markov chains, in providing a short introduction to the rich theory of *strong* Markov processes.

8.1. Continuous time filtrations and stopping times

We start with the definitions of continuous time filtrations and S.P. adapted to them (compare with Definitions 5.1.1 and 5.1.2, respectively).

Definition 8.1.1. *A* (continuous time) filtration *is a non-decreasing family of sub-σ-algebras $\{\mathcal{F}_t\}$ of the measurable space $(\Omega, \mathcal{F})$, indexed by $t \geq 0$. By $\mathcal{F}_t \uparrow \mathcal{F}_\infty$ we denote such filtration $\{\mathcal{F}_t\}$ and the associated minimal σ-algebra $\mathcal{F}_\infty = \sigma(\bigcup_t \mathcal{F}_t)$ such that $\mathcal{F}_s \subseteq \mathcal{F}_t$ for all $0 \leq s \leq t \leq \infty$.*

Definition 8.1.2. *A (continuous time) S.P. $\{X_t, t \geq 0\}$ is* adapted *to a (continuous time) filtration $\{\mathcal{F}_t\}$, or in short $\mathcal{F}_t$*-adapted, *if $X_t \in m\mathcal{F}_t$ for each $t \geq 0$ or equivalently, if $\mathcal{F}_t^{\mathbf{X}} \subseteq \mathcal{F}_t$ for all $t \geq 0$.*

Remark 8.1.3. To avoid cumbersome technical difficulties, we assume throughout that the filtration is *augmented* so that every $\mathbf{P}$-null set is in $\mathcal{F}_0$. That is, if $N \subseteq A$ for some $A \in \mathcal{F}$ with $\mathbf{P}(A) = 0$ then $N \in \mathcal{F}_0$ (which is a somewhat stronger assumption than the completion of both $\mathcal{F}$ and $\mathcal{F}_0$). In particular, this assures that any modification of an $\mathcal{F}_t$-adapted continuous time S.P. remains $\mathcal{F}_t$-adapted.

When dealing with continuous time processes it helps if each new piece of information has a definite first time of arrival, as captured mathematically by the concept of right-continuous filtration.

Definition 8.1.4. *To any continuous time filtration $\{\mathcal{F}_t\}$ we associate the corresponding* left-filtration *$\mathcal{F}_{t^-} = \sigma(\mathcal{F}_s, s < t)$ at time t, consisting of all events prior to t (where we set $\mathcal{F}_{0^-} = \mathcal{F}_0$), and* right-filtration *$\mathcal{F}_{t^+} = \bigcap_{\varepsilon>0} \mathcal{F}_{t+\varepsilon}$ at time t, consisting of all events* immediately *after t. A filtration $\{\mathcal{F}_t\}$ is called* right-continuous *if it coincides with its right-filtration, that is $\mathcal{F}_t = \mathcal{F}_{t^+}$ for all $t \geq 0$.*

The next example ties the preceding definitions to those in the much simpler setting of Chapter 5.

Example 8.1.5. *To each discrete time filtration $\{\mathcal{G}_n, n \in \mathbb{Z}_+\}$ corresponds the* interpolated *(continuous time) filtration $\mathcal{F}_t = \mathcal{G}_{[t]}$, where $[t]$ denotes the integer part of $t \geq 0$. Convince yourself that any interpolated filtration is right-continuous, but usually not* left-continuous. *That is, $\mathcal{F}_t \neq \mathcal{F}_{t^-}$ (at any $t = n$ integer in which $\mathcal{G}_n \neq \mathcal{G}_{n-1}$), with each jump in the filtration accounting for a new piece of information arriving at that time.*

Similarly, we associate with any $\mathcal{G}_n$-adapted discrete time S.P. $\{Y_n\}$ an interpolated *continuous time S.P. $X_t = Y_{[t]}$, $t \geq 0$, noting that $\{X_t, t \geq 0\}$ is then $\mathcal{F}_t$-adapted if and only if $\{Y_n\}$ is $\mathcal{G}_n$-adapted.*

Example 8.1.6. *In analogy with Definition 5.1.3, another generic continuous time filtration is the* canonical filtration *$\mathcal{F}_t^{\mathbf{X}} = \sigma(X_s, 0 \leq s \leq t)$ associated with each continuous time S.P. $\{X_t, t \geq 0\}$.*

Unfortunately, sample path continuity of a S.P. $\{X_t\}$ does not guarantee the right-continuity of its canonical filtration $\{\mathcal{F}_t^{\mathbf{X}}\}$. Indeed, considering the uniform probability measure on $\Omega = \{-1, 1\}$ note that the canonical filtration $\{\mathcal{F}_t^{\mathbf{X}}\}$ of the S.P. $X_t(\omega) = \omega t$, which has continuous sample functions, is evidently not right-continuous at $t = 0$ (as $\mathcal{F}_0^{\mathbf{X}} = \{\emptyset, \Omega\}$ while $\mathcal{F}_t^{\mathbf{X}} = \mathcal{F} = 2^\Omega$ for all $t > 0$).

When S.P. $\{X_s, s \geq 0\}$ is $\mathcal{F}_t$-adapted, we can view $\{X_s, s \in [0, t]\}$ as a S.P. on the smaller measurable space $(\Omega, \mathcal{F}_t)$, for each $t \geq 0$. However, as seen in Section 7.2, more is required in order to have Borel sample functions, prompting the following extension of Definition 7.2.19 (and refinement of Definition 8.1.2).

Definition 8.1.7. *An $\mathcal{F}_t$-adapted S.P. $\{X_t, t \geq 0\}$ is called $\mathcal{F}_t$-*progressively measurable *if $X_s(\omega) : [0, t] \times \Omega \mapsto \mathbb{R}$ is measurable with respect to $\mathcal{B}_{[0,t]} \times \mathcal{F}_t$, for each $t \geq 0$.*

Remark. In contrast to Definition 7.2.19, we have dropped the completion of the relevant σ-algebras in the preceding definition. Indeed, the standing assumption of Remark 8.1.3 guarantees the completeness of each σ-algebra of the filtration $\mathcal{F}_t$ and as we see next, progressive measurability is in any case equivalent to adaptedness for all RCLL processes.

Proposition 8.1.8. *An $\mathcal{F}_t$-adapted S.P. $\{X_s, s \geq 0\}$ of right-continuous sample functions is also $\mathcal{F}_t$-progressively measurable.*

Proof. Fixing $t > 0$, let $\mathbb{Q}_{t^+}^{(2,\ell)}$ denote the finite set of dyadic rationals of the form $j2^{-\ell} \in [0, t]$ augmented by $\{t\}$ and arranged in increasing order $0 = t_0 < t_1 < \cdots < t_{k_\ell} = t$ (where $k_\ell = \lceil t2^\ell \rceil$). The ℓ-th approximation of the sample function $X_s(\omega)$ for $s \in [0, t]$, is then

$$X_s^{(\ell)}(\omega) = X_0 I_{\{0\}}(s) + \sum_{j=1}^{k_\ell} X_{t_j}(\omega) I_{(t_{j-1}, t_j]}(s) .$$

Note that per positive integer ℓ and $B \in \mathcal{B}$,

$$\{(s, \omega) \in [0, t] \times \Omega : X_s^{(\ell)}(\omega) \in B\} = \{0\} \times X_0^{-1}(B) \bigcup_{j=1}^{k_\ell} (t_{j-1}, t_j] \times X_{t_j}^{-1}(B) ,$$

which is in the product σ-algebra $\mathcal{B}_{[0,t]} \times \mathcal{F}_t$, since each of the sets $X_{t_j}^{-1}(B)$ is in $\mathcal{F}_t$ (recall that $\{X_s, s \geq 0\}$ is $\mathcal{F}_t$-adapted and $t_j \in [0, t]$). Consequently, each of the maps $(s,\omega) \mapsto X_s^{(\ell)}(\omega)$ is a real-valued R.V. on the product measurable space $([0,t] \times \Omega, \mathcal{B}_{[0,t]} \times \mathcal{F}_t)$. Further, by right-continuity of the sample functions $s \mapsto X_s(\omega)$, for each fixed $(s,\omega) \in [0,t] \times \Omega$ the sequence $X_s^{(\ell)}(\omega)$ converges as $\ell \to \infty$ to $X_s(\omega)$, which is thus a R.V. on the same (product) measurable space (recall Corollary 1.2.23). □

Associated with any filtration $\{\mathcal{F}_t\}$ is the collection of all $\mathcal{F}_t$-stopping times and the corresponding stopped σ-algebras (compare with Definitions 5.1.11 and 5.1.34).

Definition 8.1.9. *A random variable $\tau : \Omega \mapsto [0,\infty]$ is called a* stopping time *for the (continuous time) filtration $\{\mathcal{F}_t\}$, or in short $\mathcal{F}_t$-*stopping time*, if $\{\omega : \tau(\omega) \le t\} \in \mathcal{F}_t$ for all $t \ge 0$. Associated with each $\mathcal{F}_t$-stopping time τ is the* stopped σ-algebra $\mathcal{F}_\tau = \{A \in \mathcal{F}_\infty : A \cap \{\tau \le t\} \in \mathcal{F}_t$ *for all* $t \ge 0\}$ *(which quantifies the information in the filtration at the stopping time τ).*

*The $\mathcal{F}_{t^+}$-stopping times are also called $\mathcal{F}_t$-*Markov times *(or $\mathcal{F}_t$-*optional times*), with the corresponding* Markov σ-algebras $\mathcal{F}_{\tau^+} = \{A \in \mathcal{F}_\infty : A \cap \{\tau \le t\} \in \mathcal{F}_{t^+}$ *for all $t \ge 0\}$.*

Remark. As their name suggest, Markov/optional times appear both in the context of Doob's optional stopping theorem (in Section 8.2.3), and in that of the strong Markov property (see Section 8.3.2).

Obviously, any non-random constant $t \ge 0$ is a stopping time. Further, by definition, every $\mathcal{F}_t$-stopping time is also an $\mathcal{F}_t$-Markov time and the two concepts coincide for right-continuous filtrations. Similarly, the Markov σ-algebra $\mathcal{F}_{\tau^+}$ contains the stopped σ-algebra $\mathcal{F}_\tau$ for any $\mathcal{F}_t$-stopping time (and they coincide in case of right-continuous filtrations).

Your next exercise provides more explicit characterization of Markov times and closure properties of Markov and stopping times (some of which you saw before in Exercise 5.1.12).

Exercise 8.1.10.

(a) *Show that τ is an $\mathcal{F}_t$-Markov time if and only if $\{\omega : \tau(\omega) < t\} \in \mathcal{F}_t$ for all $t \ge 0$.*

(b) *Show that if $\{\tau_n, n \in \mathbb{Z}_+\}$ are $\mathcal{F}_t$-stopping times, then so are $\tau_1 \wedge \tau_2$, $\tau_1 + \tau_2$ and $\sup_n \tau_n$.*

(c) *Show that if $\{\tau_n, n \in \mathbb{Z}_+\}$ are $\mathcal{F}_t$-Markov times, then in addition to $\tau_1 + \tau_2$ and $\sup_n \tau_n$, also $\inf_n \tau_n$, $\liminf_n \tau_n$ and $\limsup_n \tau_n$ are $\mathcal{F}_t$-Markov times.*

(d) *In the setting of part (c) show that $\tau_1 + \tau_2$ is an $\mathcal{F}_t$-stopping time when either both τ_1 and τ_2 are strictly positive, or alternatively, when τ_1 is a strictly positive $\mathcal{F}_t$-stopping time.*

Similarly, here are some of the basic properties of stopped σ-algebras (compare with Exercise 5.1.35), followed by additional properties of Markov σ-algebras.

Exercise 8.1.11. *Suppose θ and τ are $\mathcal{F}_t$-stopping times.*

(a) *Verify that $\sigma(\tau) \subseteq \mathcal{F}_\tau$, that $\mathcal{F}_\tau$ is a σ-algebra, and if $\tau(\omega) = t$ is non-random then $\mathcal{F}_\tau = \mathcal{F}_t$.*

(b) *Show that $\mathcal{F}_{\theta \wedge \tau} = \mathcal{F}_\theta \cap \mathcal{F}_\tau$ and deduce that each of the events $\{\theta < \tau\}$, $\{\theta \le \tau\}$, $\{\theta = \tau\}$ belongs to $\mathcal{F}_{\theta \wedge \tau}$.*
Hint: *Show first that if $A \in \mathcal{F}_\theta$ then $A \cap \{\theta \le \tau\} \in \mathcal{F}_\tau$.*

(c) *Show that for any integrable R.V. Z,*

$$\mathbf{E}[Z|\mathcal{F}_\theta]I_{\theta\le\tau} = \mathbf{E}[Z|\mathcal{F}_{\theta\wedge\tau}]I_{\theta\le\tau}\,,$$

and deduce that

$$\mathbf{E}[\mathbf{E}(Z|\mathcal{F}_\theta)|\mathcal{F}_\tau] = \mathbf{E}[Z|\mathcal{F}_{\theta\wedge\tau}]\,.$$

(d) *Show that if $\theta \le \xi$ and $\xi \in m\mathcal{F}_\theta$ then ξ is an $\mathcal{F}_t$-stopping time.*

Exercise 8.1.12. *Suppose τ, τ_n are $\mathcal{F}_t$-Markov times.*

(a) *Verify that $\mathcal{F}_{\tau^+} = \{A \in \mathcal{F}_\infty : A \cap \{\tau < t\} \in \mathcal{F}_t \text{ for all } t \ge 0\}$.*
(b) *Suppose τ_1 is further $\mathcal{F}_t$-stopping time and $\tau \le \tau_1$ with a strict inequality whenever τ is finite. Show that then $\mathcal{F}_{\tau^+} \subseteq \mathcal{F}_{\tau_1}$.*
(c) *Setting $\tau = \inf_n \tau_n$, show that $\mathcal{F}_{\tau^+} = \bigcap_n \mathcal{F}_{\tau_n^+}$. Deduce that if τ_n are $\mathcal{F}_t$-stopping times and $\tau < \tau_n$ whenever τ is finite, then $\mathcal{F}_{\tau^+} = \bigcap_n \mathcal{F}_{\tau_n}$.*

In contrast to adaptedness, progressive measurability transfers to stopped processes (i.e. the continuous time extension of Definition 5.1.31), which is essential when dealing in Section 8.2 with stopped sub-martingales (i.e. the continuous time extension of Theorem 5.1.32).

Proposition 8.1.13. *Given $\mathcal{F}_t$-progressively measurable S.P. $\{X_s, s \ge 0\}$, the* stopped at ($\mathcal{F}_t$-stopping time) τ S.P. $\{X_{s\wedge\tau(\omega)}(\omega), s \ge 0\}$ *is also $\mathcal{F}_t$-progressively measurable. In particular, if either $\tau < \infty$ or there exists $X_\infty \in m\mathcal{F}_\infty$, then $X_\tau \in m\mathcal{F}_\tau$.*

Proof. Fixing $t > 0$, denote by $\mathcal{S}$ the product σ-algebra $\mathcal{B}_{[0,t]} \times \mathcal{F}_t$ on the product space $\mathbb{S} = [0,t] \times \Omega$. The assumed $\mathcal{F}_t$-progressive measurability of $\{X_s, s \ge 0\}$ amounts to the measurability of $g_1 : (\mathbb{S}, \mathcal{S}) \mapsto (\mathbb{R}, \mathcal{B})$ such that $g_1(s,\omega) = X_s(\omega)$. Further, as $(s,\omega) \mapsto X_{s\wedge\tau(\omega)}(\omega)$ is the composition $g_1(g_2(s,\omega))$ for the mapping $g_2(s,\omega) = (s\wedge\tau(\omega), \omega)$ from $(\mathbb{S}, \mathcal{S})$ to itself, by Proposition 1.2.18 the $\mathcal{F}_t$-progressive measurability of the stopped S.P. follows form our claim that g_2 is measurable. Indeed, recall that τ is an $\mathcal{F}_t$-stopping time, so $\{\omega : \tau(\omega) > u\} \in \mathcal{F}_t$ for any $u \in [0,t]$. Hence, for any fixed $u \in [0,t]$ and $A \in \mathcal{F}_t$,

$$g_2^{-1}((u,t] \times A) = (u,t] \times (A \cap \{\omega : \tau(\omega) > u\}) \in \mathcal{S}\,,$$

which suffices for measurability of g_2 (since the product σ-algebra $\mathcal{S}$ is generated by the collection $\{(u,t] \times A : u \in [0,t], A \in \mathcal{F}_t\}$).

Turning to the second claim, since $\{X_{s\wedge\tau}, s \ge 0\}$ is $\mathcal{F}_t$-progressively measurable, we have that for any fixed $B \in \mathcal{B}$ and finite $t \ge 0$,

$$X_\tau^{-1}(B) \cap \tau^{-1}([0,t]) = \{\omega : X_{t\wedge\tau(\omega)}(\omega) \in B\} \cap \{\omega : \tau(\omega) \le t\} \in \mathcal{F}_t$$

(recall (1.4.7) that $\{\omega : (t,\omega) \in A\}$ is in $\mathcal{F}_t$ for any set $A \in \mathcal{S}$). Moreover, by our assumptions $X_\infty^{-1}(B) \cap \tau^{-1}(\{\infty\})$ is in $\mathcal{F}_\infty$, hence so is its union with the sets $X_\tau^{-1}(B) \cap \tau^{-1}([0,n])$, $n \in \mathbb{Z}_+$, which is precisely $X_\tau^{-1}(B)$. We have thus shown that $X_\tau^{-1}(B) \in \mathcal{F}_\tau$ for any $B \in \mathcal{B}$, namely, that $X_\tau \in m\mathcal{F}_\tau$. $\square$

Recall Exercise 5.1.13 that for discrete time S.P. and filtrations, the first hitting time τ_B of a Borel set B by $\mathcal{F}_n$-adapted process is an $\mathcal{F}_n$-stopping time. Unfortunately, this may fail in the continuous time setting, even when considering an open set B and the canonical filtration $\mathcal{F}_t^{\mathbf{X}}$ of a S.P. of continuous sample functions.

Example 8.1.14. *Indeed, consider $B = (0, \infty)$ and the S.P. $X_t(\omega) = \omega t$ of Example 8.1.6. In this case, $\tau_B(1) = 0$ while $\tau_B(-1) = \infty$, so the event $\{\omega : \tau_B(\omega) \leq 0\} = \{1\}$ is not in $\mathcal{F}_0^{\mathbf{X}} = \{\emptyset, \Omega\}$ (hence τ_B is not an $\mathcal{F}_t^{\mathbf{X}}$-stopping time). As shown next, this problem is only due to the lack of right-continuity in the filtration $\{\mathcal{F}_t^{\mathbf{X}}\}$.*

Proposition 8.1.15. *Consider an $\mathcal{F}_t$-adapted, right-continuous S.P. $\{X_s, s \geq 0\}$. Then, the* first hitting time *$\tau_B(\omega) = \inf\{t \geq 0 : X_t(\omega) \in B\}$ is an $\mathcal{F}_t$-Markov time for an open set B and further an $\mathcal{F}_t$-stopping time when B is a closed set and $\{X_s, s \geq 0\}$ has continuous sample functions.*

Proof. Fixing $t > 0$, by definition of τ_B the set $\tau_B^{-1}([0,t))$ is the union of $X_s^{-1}(B)$ over all $s \in [0,t)$. Further, if the right-continuous function $s \mapsto X_s(\omega)$ intersects an open set B at some $s \in [0,t)$ then necessarily $X_q(\omega) \in B$ at some $q \in \mathbb{Q}_{t^-} = \mathbb{Q} \cap [0,t)$. Consequently,

$$\tau_B^{-1}([0,t)) = \bigcup_{s \in \mathbb{Q}_{t^-}} X_s^{-1}(B) \,. \tag{8.1.1}$$

Now, the $\mathcal{F}_t$-adaptedness of $\{X_s\}$ implies that $X_s^{-1}(B) \in \mathcal{F}_s \subseteq \mathcal{F}_t$ for all $s \leq t$, and in particular for any s in the countable collection $\mathbb{Q}_{t^-}$. We thus deduce from (8.1.1) that $\{\tau_B < t\} \in \mathcal{F}_t$ for all $t \geq 0$, and in view of part (a) of Exercise 8.1.10, conclude that τ_B is an $\mathcal{F}_t$-Markov time in case B is open.

Assuming hereafter that B is closed and $u \mapsto X_u$ continuous, we claim that for any $t > 0$,

$$\{\tau_B \leq t\} = \bigcup_{0 \leq s \leq t} X_s^{-1}(B) = \bigcap_{k=1}^{\infty} \{\tau_{B_k} < t\} := A_t \,, \tag{8.1.2}$$

where $B_k = \{x \in \mathbb{R} : |x - y| < k^{-1}$, for some $y \in B\}$, and that the left identity in (8.1.2) further holds for $t = 0$. Clearly, $X_0^{-1}(B) \in \mathcal{F}_0$ and for B_k open, by the preceding proof $\{\tau_{B_k} < t\} \in \mathcal{F}_t$. Hence, (8.1.2) implies that $\{\tau_B \leq t\} \in \mathcal{F}_t$ for all $t \geq 0$, namely, that τ_B is an $\mathcal{F}_t$-stopping time.

Turning to verify (8.1.2), fix $t > 0$ and recall that if $\omega \in A_t$ then $|X_{s_k}(\omega) - y_k| < k^{-1}$ for some $s_k \in [0,t)$ and $y_k \in B$. Upon passing to a sub-sequence, $s_k \to s \in [0,t]$, hence by continuity of the sample function $X_{s_k}(\omega) \to X_s(\omega)$. This in turn implies that $y_k \to X_s(\omega) \in B$ (because B is a closed set). Conversely, if $X_s(\omega) \in B$ for some $s \in [0,t)$ then also $\tau_{B_k} \leq s < t$ for all $k \geq 1$, whereas even if only $X_t(\omega) = y \in B$, by continuity of the sample function also $X_s(\omega) \to y$ for $0 \leq s \uparrow t$ (and once again $\tau_{B_k} < t$ for all $k \geq 1$). To summarize, $\omega \in A_t$ if and only if there exists $s \in [0,t]$ such that $X_s(\omega) \in B$, as claimed. Considering hereafter $t \geq 0$ (possibly $t = 0$), the existence of $s \in [0,t]$ such that $X_s \in B$ results with $\{\tau_B \leq t\}$. Conversely, if $\tau_B(\omega) \leq t$ then $X_{s_n}(\omega) \in B$ for some $s_n(\omega) \leq t + n^{-1}$ and all n. But then $s_{n_k} \to s \leq t$ along some sub-sequence $n_k \to \infty$, so for B closed, by continuity of the sample function also $X_{s_{n_k}}(\omega) \to X_s(\omega) \in B$. □

We conclude with a technical result on which we shall later rely, for example, in proving the optional stopping theorem and in the study of the strong Markov property.

Lemma 8.1.16. *Given an $\mathcal{F}_t$-Markov time τ, let $\tau_\ell = 2^{-\ell}([2^\ell \tau] + 1)$ for $\ell \geq 1$. Then, τ_ℓ are $\mathcal{F}_t$-stopping times and $A \cap \{\omega : \tau_\ell(\omega) = q\} \in \mathcal{F}_q$ for any $A \in \mathcal{F}_{\tau^+}$, $\ell \geq 1$ and $q \in \mathbb{Q}^{(2,\ell)} = \{k2^{-\ell}, k \in \mathbb{Z}_+\}$.*

PROOF. By its construction, τ_ℓ takes values in the discrete set $\mathbb{Q}^{(2,\ell)} \cup \{\infty\}$. Moreover, with $\{\omega : \tau(\omega) < t\} \in \mathcal{F}_t$ for any $t \geq 0$ (see part (a) of Exercise 8.1.10), it follows that for any $q \in \mathbb{Q}^{(2,\ell)}$,

$$\{\omega : \tau_\ell(\omega) = q\} = \{\omega : \tau(\omega) \in [q - 2^{-\ell}, q)\} \in \mathcal{F}_q \,.$$

Hence, τ_ℓ is an $\mathcal{F}_t$-stopping time, as claimed. Next, fixing $A \in \mathcal{F}_{\tau^+}$, in view of Definitions 8.1.4 and 8.1.9, the sets $A_{t,m} = A \cap \{\omega : \tau(\omega) \leq t - m^{-1}\}$ are in $\mathcal{F}_t$ for any $m \geq 1$. Further, by the preceding, fixing $q \in \mathbb{Q}^{(2,\ell)}$ and $q' = q - 2^{-\ell}$ you have that

$$A \cap \{\omega : \tau_\ell(\omega) = q\} = A \cap \{\omega : \tau(\omega) \in [q', q)\} = (\cup_{m\geq 1} A_{q,m}) \setminus (\cup_{m\geq 1} A_{q',m})$$

is the difference between an element of $\mathcal{F}_q$ and one of $\mathcal{F}_{q'} \subseteq \mathcal{F}_q$. Consequently, $A \cap \{\omega : \tau_\ell(\omega) = q\}$ is in $\mathcal{F}_q$, as claimed. $\square$

8.2. Continuous time martingales

As we show in this section, once the technical challenges involved with the continuity of time are taken care off, the results of Chapter 5 extend in a natural way to the collection of continuous time (sub and super) martingales. Similar to the break-up of Chapter 5, we devote Subsection 8.2.1 to the definition, examples and closure properties of this collection of S.P. (compare with Section 5.1), followed by Subsection 8.2.2 about tail (and upcrossing) inequalities and convergence properties of such processes (compare with Sections 5.2.2 and 5.3, respectively). The statement, proof and applications of Doob's optional stopping theorem are explored in Subsection 8.2.3 (compare with Section 5.4), with martingale representations being the focus of Subsection 8.2.4 (compare with Sections 5.2.1 and 5.3.2).

8.2.1. Definition, examples and closure properties. For a continuous filtration, it is not enough to consider the martingale property one step ahead, so we replace Definitions 5.1.4 and 5.1.16 by the following continuous time analog of Proposition 5.1.20.

Definition 8.2.1. *The pair $(X_t, \mathcal{F}_t, t \geq 0)$ is called a* continuous time martingale *(in short MG), if the integrable (continuous time) S.P. $\{X_t, t \geq 0\}$ is adapted to the (continuous time) filtration $\{\mathcal{F}_t, t \geq 0\}$ and for any fixed $t \geq s \geq 0$, the identity $\mathbf{E}[X_t|\mathcal{F}_s] = X_s$ holds a.s. Replacing the preceding identity with $\mathbf{E}[X_t|\mathcal{F}_s] \geq X_s$ a.s. for each $t \geq s \geq 0$, or with $\mathbf{E}[X_t|\mathcal{F}_s] \leq X_s$ a.s. for each $t \geq s \geq 0$, defines the* continuous time sub-MG *and* continuous time sup-MG*, respectively. These three classes of continuous time S.P. are related in the same manner as in the discrete time setting (c.f. Remark 5.1.17).*

It immediately follows from the preceding definition that $t \mapsto \mathbf{E}X_t$ is non-decreasing for a sub-MG, non-increasing for a sup-MG, and constant (in time), for a MG. Further, unless explicitly stated otherwise, one uses the canonical filtration when studying MGs (or sub/sup-MGs).

Exercise 8.2.2. *Suppose $(X_t, \mathcal{F}_t, t \geq 0)$ is a continuous time sub-MG.*

(a) *Show that $(X_t, \mathcal{F}^{\mathbf{X}}_t, t \geq 0)$ is also a sub-MG.*
(b) *Show that if $\mathbf{E}X_t = \mathbf{E}X_0$ for all $t \geq 0$, then $(X_t, \mathcal{F}_t, t \geq 0)$ is also a martingale.*

The decomposition of conditional second moments, as in part (b) of Exercise 5.1.8, applies for all continuous time square-integrable MGs.

Exercise 8.2.3. *Suppose $(X_t, \mathcal{F}_t, t \geq 0)$ is a square-integrable MG. Verify that*

$$\mathbf{E}[X_t^2|\mathcal{F}_s] - X_s^2 = \mathbf{E}[(X_t - X_s)^2|\mathcal{F}_s] \qquad \text{for any } t \geq s \geq 0\,, \tag{8.2.1}$$

and deduce that $t \mapsto \mathbf{E}X_t^2$ is non-decreasing.

As you see next, the Wiener process and the compensated Poisson process play the same role that the random walk of zero-mean increments plays in the discrete time setting (with Wiener process being the prototypical MG of continuous sample functions, and compensated Poisson process the prototypical MG of discontinuous RCLL sample functions).

Proposition 8.2.4. *Any integrable S.P. $\{X_t, t \geq 0\}$ of independent increments (see Exercise 7.1.12), and constant mean function is a MG.*

Proof. Recall that a S.P. X_t has independent increments if $X_{t+h} - X_t$ is independent of $\mathcal{F}_t^{\mathbf{X}}$, for all $h > 0$ and $t \geq 0$. We have also assumed that $\mathbf{E}|X_t| < \infty$ and $\mathbf{E}X_t = \mathbf{E}X_0$ for all $t \geq 0$. Therefore, $\mathbf{E}[X_{t+h} - X_t|\mathcal{F}_t^{\mathbf{X}}] = \mathbf{E}[X_{t+h} - X_t] = 0$. Further, $X_t \in m\mathcal{F}_t^{\mathbf{X}}$ and hence $\mathbf{E}[X_{t+h}|\mathcal{F}_t^{\mathbf{X}}] = X_t$. That is, $\{X_t, \mathcal{F}_t^{\mathbf{X}}, t \geq 0\}$ is a MG, as claimed. $\square$

Example 8.2.5. *In view of Exercise 7.3.13 and Proposition 8.2.4 we have that the* Wiener process/Brownian motion $(W_t, t \geq 0)$ *of Definition 7.3.12 is a martingale.*

Combining Proposition 3.4.9 and Exercise 7.1.12, we see that the Poisson process N_t of rate λ has independent increments and mean function $\mathbf{E}N_t = \lambda t$. Consequently, by Proposition 8.2.4 the compensated Poisson process $M_t = N_t - \lambda t$ *is also a martingale (and $\mathcal{F}_t^{\mathbf{M}} = \mathcal{F}_t^{\mathbf{N}}$).*

Similarly to Exercise 5.1.9, as you check next, a Gaussian martingale $\{X_t, t \geq 0\}$ is necessarily square-integrable and of independent increments, in which case $M_t = X_t^2 - \langle X \rangle_t$ is also a martingale.

Exercise 8.2.6.

(a) *Show that if $\{X_t, t \geq 0\}$ is a square-integrable S.P. having zero-mean independent increments, then $(X_t^2 - \langle X \rangle_t, \mathcal{F}_t^{\mathbf{X}}, t \geq 0)$ is a MG with $\langle X \rangle_t = \mathbf{E}X_t^2 - \mathbf{E}X_0^2$ a non-random, non-decreasing function.*

(b) *Prove that the conclusion of part (a) applies to any martingale $\{X_t, t \geq 0\}$ which is a Gaussian S.P.*

(c) *Deduce that if $\{X_t, t \geq 0\}$ is square-integrable, with $X_0 = 0$ and zero-mean, stationary independent increments, then $(X_t^2 - t\mathbf{E}X_1^2, \mathcal{F}_t^{\mathbf{X}}, t \geq 0)$ is a MG.*

In the context of the Brownian motion $\{B_t, t \geq 0\}$, we deduce from part (b) of Exercise 8.2.6 that $\{B_t^2 - t, t \geq 0\}$ is a MG. This is merely a special case of the following collection of MGs associated with the standard Brownian motion.

Exercise 8.2.7. *Let $u_{k+1}(t, y, \theta) = \frac{\partial}{\partial \theta} u_k(t, y, \theta)$ for $k \geq 0$ and $u_0(t, y, \theta) = \exp(\theta y - \theta^2 t/2)$.*

(a) *Show that for any $\theta \in \mathbb{R}$ the S.P. $(u_0(t, B_t, \theta), t \geq 0)$ is a martingale with respect to $\mathcal{F}_t^{\mathbf{B}}$.*

(b) *Check that for* $k = 1, 2, \ldots$,

$$u_k(t, y, 0) = \sum_{r=0}^{[k/2]} \frac{k!}{(k-2r)!r!} y^{k-2r} (-t/2)^r .$$

(c) *Deduce that the S.P.* $(u_k(t, B_t, \theta), t \geq 0)$, $k = 1, 2, \ldots$ *are also MGs with respect to* $\mathcal{F}_t^{\mathbf{B}}$, *as are* $B_t^2 - t$, $B_t^3 - 3tB_t$, $B_t^4 - 6tB_t^2 + 3t^2$ *and* $B_t^6 - 15tB_t^4 + 45t^2 B_t^2 - 15t^3$.

(d) *Verify that for each* $k \in \mathbb{Z}_+$ *and* $\theta \in \mathbb{R}$ *the function* $u_k(t, y, \theta)$ *solves the heat equation* $u_t(t, y) + \frac{1}{2} u_{yy}(t, y) = 0$.

The collection of sub-MG (equivalently, sup-MG or MG), is closed under the addition of S.P. (compare with Exercise 5.1.19).

Exercise 8.2.8. *Suppose* $(X_t, \mathcal{F}_t)$ *and* $(Y_t, \mathcal{F}_t)$ *are sub-MGs and* $t \mapsto f(t)$ *a non-decreasing, non-random function.*

(a) *Verify that* $(X_t + Y_t, \mathcal{F}_t)$ *is a sub-MG and hence so is* $(X_t + f(t), \mathcal{F}_t)$.

(b) *Rewrite this, first for sup-MGs* X_t *and* Y_t, *then in case of MGs.*

With the same proof as in Proposition 5.1.22, you are next to verify that the collection of sub-MGs (and that of sup-MGs), is also closed under the application of a non-decreasing convex (concave, respectively), function (c.f. Example 5.1.23 for the most common choices of this function).

Exercise 8.2.9. *Suppose the integrable S.P.* $\{X_t, t \geq 0\}$ *and convex function* $\Phi : \mathbb{R} \mapsto \mathbb{R}$ *are such that* $\mathbf{E}[|\Phi(X_t)|] < \infty$ *for all* $t \geq 0$. *Show that if* $(X_t, \mathcal{F}_t)$ *is a MG then* $(\Phi(X_t), \mathcal{F}_t)$ *is a sub-MG and the same applies even when* $(X_t, \mathcal{F}_t)$ *is only a sub-MG, provided* $\Phi(\cdot)$ *is also non-decreasing.*

As you show next, the *martingale Bayes rule* of Exercise 5.5.16 applies also for a positive, continuous time martingale $(Z_t, \mathcal{F}_t, t \geq 0)$.

Exercise 8.2.10. *Suppose* $(Z_t, \mathcal{F}_t, t \geq 0)$ *is a (strictly) positive MG on* $(\Omega, \mathcal{F}, \mathbf{P})$, *normalized so that* $\mathbf{E}Z_0 = 1$. *For each* $t > 0$, *let* $\mathbf{P}_t = \mathbf{P}\big|_{\mathcal{F}_t}$ *and consider the equivalent probability measure* $\mathbf{Q}_t$ *on* $(\Omega, \mathcal{F}_t)$ *of Radon-Nikodym derivative* $d\mathbf{Q}_t / d\mathbf{P}_t = Z_t$.

(a) *Show that* $\mathbf{Q}_s = \mathbf{Q}_t\big|_{\mathcal{F}_s}$ *for any* $s \in [0, t]$.

(b) *Fixing* $u \leq s \in [0, t]$ *and* $Y \in L^1(\Omega, \mathcal{F}_s, \mathbf{Q}_t)$ *show that* $\mathbf{Q}_t$*-a.s. (hence also* $\mathbf{P}$*-a.s.),* $\mathbf{E}_{\mathbf{Q}_t}[Y | \mathcal{F}_u] = \mathbf{E}[Y Z_s | \mathcal{F}_u] / Z_u$.

(c) *Verify that if* $\widetilde{\lambda} > 0$ *and* N_t *is a Poisson Process of rate* $\lambda > 0$ *then* $Z_t = e^{(\lambda - \widetilde{\lambda})t} (\widetilde{\lambda}/\lambda)^{N_t}$ *is a strictly positive martingale with* $\mathbf{E}Z_0 = 1$ *and show that* $\{N_t, t \in [0, T]\}$ *is a Poisson process of rate* $\widetilde{\lambda}$ *under the measure* $\mathbf{Q}_T$, *for any finite* T.

Remark. Up to the re-parametrization $\theta = \log(\widetilde{\lambda}/\lambda)$, the martingale Z_t of part (c) of the preceding exercise is of the form $Z_t = u_0(t, N_t, \theta)$ for $u_0(t, y, \theta) = \exp(\theta y - \lambda t(e^\theta - 1))$. Building on it and following the line of reasoning of Exercise 8.2.7 yields the analogous collection of martingales for the Poisson process $\{N_t, t \geq 0\}$. For example, here the functions $u_k(t, y, \theta)$ on $(t, y) \in \mathbb{R}_+ \times \mathbb{Z}_+$ solve the equation $u_t(t, y) + \lambda[u(t, y) - u(t, y+1)] = 0$, with $M_t = u_1(t, N_t, 0)$ being the compensated Poisson process of Example 8.2.5 while $u_2(t, N_t, 0)$ is the martingale $M_t^2 - \lambda t$.

Remark. While beyond our scope, we note in passing that in continuous time the martingale transform of Definition 5.1.27 is replaced by the *stochastic integral* $Y_t = \int_0^t V_s \, dX_s$. This stochastic integral results with stochastic differential equations and is the main object of study of *stochastic calculus* (to which many texts are devoted, among them [**KaS97**]). In case $V_s = X_s$ is the Wiener process W_s, the analog of Example 5.1.29 is $Y_t = \int_0^t W_s dW_s$, which for the appropriate definition of the stochastic integral (due to Itô), is merely the martingale $Y_t = \frac{1}{2}(W_t^2 - t)$. Indeed, Itô's stochastic integral is defined via martingale theory, at the cost of deviating from the standard integration by parts formula. The latter would have applied if the sample functions $t \mapsto W_t(\omega)$ were differentiable w.p.1., which is definitely not the case (as we shall see in Section 9.3).

Exercise 8.2.11. *Suppose S.P. $\{X_t, t \geq 0\}$ is integrable and $\mathcal{F}_t$-adapted. Show that if $\mathbf{E}[X_u] \geq \mathbf{E}[X_\tau]$ for any $u \geq 0$ and $\mathcal{F}_t$-stopping time τ whose range $\tau(\Omega)$ is a finite subset of $[0, u]$, then $(X_t, \mathcal{F}_t, t \geq 0)$ is a sub-MG.*
Hint: *Consider $\tau = sI_A + uI_{A^c}$ with $s \in [0, u]$ and $A \in \mathcal{F}_s$.*

We conclude this sub-section with the relations between continuous and discrete time (sub/super) martingales.

Example 8.2.12. *Convince yourself that to any discrete time sub-MG $(Y_n, \mathcal{G}_n, n \in \mathbb{Z}_+)$ corresponds the* interpolated *continuous time sub-MG $(X_t, \mathcal{F}_t, t \geq 0)$ of the interpolated right-continuous filtration $\mathcal{F}_t = \mathcal{G}_{[t]}$ and RCLL S.P. $X_t = Y_{[t]}$ of Example 8.1.5.*

Remark 8.2.13. In proving results about continuous time MGs (or sub-MGs/sup-MGs), we often rely on the converse of Example 8.2.12. Namely, for any non-random, non-decreasing sequence $\{s_k\} \subset [0, \infty)$, if $(X_t, \mathcal{F}_t)$ is a continuous time MG (or sub-MG/sup-MG), then clearly $(X_{s_k}, \mathcal{F}_{s_k}, k \in \mathbb{Z}_+)$ is a discrete time MG (or sub-MG/sup-MG, respectively), while $(X_{s_k}, \mathcal{F}_{s_k}, k \in \mathbb{Z}_-)$ is a RMG (or reversed subMG/supMG, respectively), where $s_0 \geq s_{-1} \geq \cdots \geq s_{-k} \geq \cdots$.

8.2.2. Inequalities and convergence.

In this section we extend the tail inequalities and convergence properties of discrete time sub-MGs (or sup-MGs), to the corresponding results for sub-MGs (and sup-MGs) of right-continuous sample functions, which we call hereafter in short *right-continuous* sub-MGs (or sup-MGs). We start with Doob's inequality (compare with Theorem 5.2.6).

Theorem 8.2.14 (Doob's inequality)**.** *If $\{X_s, s \geq 0\}$ is a right-continuous sub-MG, then for $t \geq 0$ finite, $M_t = \sup_{0 \leq s \leq t} X_s$, and any $x > 0$*

$$\mathbf{P}(M_t \geq x) \leq x^{-1}\mathbf{E}[X_t I_{\{M_t \geq x\}}] \leq x^{-1}\mathbf{E}[(X_t)_+] \,. \tag{8.2.2}$$

Proof. It suffices to show that for any $y > 0$

$$y\mathbf{P}(M_t > y) \leq \mathbf{E}[X_t I_{\{M_t > y\}}] \,. \tag{8.2.3}$$

Indeed, by dominated convergence, taking $y \uparrow x$ yields the left inequality in (8.2.2), and the proof is then complete since $\mathbf{E}[ZI_A] \leq \mathbf{E}[(Z)_+]$ for any event A and integrable R.V. Z.

Turning to prove (8.2.3), fix hereafter $t \geq 0$ and let $\mathbb{Q}_{t^+}^{(2,\ell)}$ denote the finite set of dyadic rationals of the form $j2^{-\ell} \in [0, t]$ augmented by $\{t\}$. Recall Remark 8.2.13 that enumerating $\mathbb{Q}_{t^+}^{(2,\ell)}$ in a non-decreasing order produces a discrete time sub-MG $\{X_{s_k}\}$. Applying Doob's inequality (5.2.1) for this sub-MG, we find that $x\mathbf{P}(M(\ell) \geq x) \leq \mathbf{E}[X_t I_{\{M(\ell)\geq x\}}]$ for

$$M(\ell) = \max_{s\in\mathbb{Q}^{(2,\ell)}_{t+}} X_s \,,$$

and any $x > 0$. Considering $x \downarrow y$, it then follows by dominated convergence that

$$y\mathbf{P}(M(\ell) > y) \leq \mathbf{E}[X_t I_{\{M(\ell)>y\}}] \,,$$

for any $\ell \geq 1$. Next note that as $\ell \uparrow \infty$,

$$M(\ell) \uparrow M(\infty) = \sup_{s\in\mathbb{Q}^{(2)}_t \cup\{t\}} X_s \,,$$

and moreover, $M(\infty) = M_t$ by the right-continuity of the sample function $t \mapsto X_t$ (compare with part (a) of Exercise 7.2.18). Consequently, for $\ell \uparrow \infty$ both $\mathbf{P}(M(\ell) > y) \uparrow \mathbf{P}(M_t > y)$ and $\mathbf{E}[X_t I_{\{M(\ell)>y\}}] \to \mathbf{E}[X_t I_{\{M_t>y\}}]$, thus completing the proof of (8.2.3). □

With $(y)^p_+$ denoting hereafter the function $(\max(y, 0))^p$, we proceed with the refinement of Doob's inequality for MGs or when the positive part of a sub-MG has finite p-th moment for some $p > 1$ (compare to Exercise 5.2.11).

Exercise 8.2.15.

(a) *Show that in the setting of Theorem 8.2.14, for any $p \geq 1$, finite $t \geq 0$ and $x > 0$,*

$$\mathbf{P}(M_t \geq x) \leq x^{-p}\mathbf{E}\big[(X_t)^p_+\big] \,,$$

(b) *Show that if $\{Y_s, s \geq 0\}$ is a right-continuous MG, then*

$$\mathbf{P}(\sup_{0\leq s\leq t} |Y_s| \geq y) \leq y^{-p}\mathbf{E}\Big[|Y_t|^p\Big].$$

By integrating Doob's inequality (8.2.2) you bound the moments of the supremum of a right-continuous sub-MG over a compact time interval.

Corollary 8.2.16 (L^p MAXIMAL INEQUALITIES)**.** *With $q = q(p) = p/(p-1)$, for any $p > 1$, $t \geq 0$ and a right-continuous sub-MG $\{X_s, s \geq 0\}$,*

$$\mathbf{E}\big[(\sup_{0\leq u\leq t} X_u)^p_+\big] \leq q^p\mathbf{E}[(X_t)^p_+] \,, \tag{8.2.4}$$

and if $\{Y_s, s \geq 0\}$ is a right-continuous MG then also

$$\mathbf{E}\big[(\sup_{0\leq u\leq t} |Y_u|)^p\big] \leq q^p\mathbf{E}[|Y_t|^p] \,. \tag{8.2.5}$$

PROOF. Adapting the proof of Corollary 5.2.13, the bound (8.2.4) is just the conclusion of part (b) of Lemma 1.4.31 for the non-negative variables $X = (X_t)_+$ and $Y = (M_t)_+$, with the left inequality in (8.2.2) providing its hypothesis. We are thus done, as the bound (8.2.5) is merely (8.2.4) in case of the non-negative sub-MG $X_t = |Y_t|$. □

In case $p = 1$ we have the following extension of Exercise 5.2.15.

Exercise 8.2.17. *Suppose $\{X_s, s \geq 0\}$ is a non-negative, right-continuous sub-MG. Show that for any $t \geq 0$,*

$$\mathbf{E}\big[\sup_{0\leq u\leq t} X_u\big] \leq (1 - e^{-1})^{-1}\{1 + \mathbf{E}[X_t(\log X_t)_+]\} \,.$$

Hint: *Relying on Exercise 5.2.15, interpolate as in our derivation of (8.2.2).*

Doob's fundamental up-crossing inequality (see Lemma 5.2.18), extends to the number of up-crossings in dyadic-times, as defined next.

Definition 8.2.18. *The number of* up-crossings *(in dyadic-times), of the interval* $[a, b]$ *by the continuous time S.P.* $\{X_u, u \in [0, t]\}$*, is the random variable* $U_t[a, b] = \sup_\ell U_{t,\ell}[a, b]$*, where* $U_{t,\ell}[a, b](\omega)$ *denotes the number of up-crossings of* $[a, b]$ *by the finite sequence* $\{X_{s_k}(\omega), s_k \in \mathbb{Q}^{(2,\ell)}_{t^+}\}$*, as in Definition 5.2.17.*

Remark. It is easy to check that for any right-continuous S.P. the number of up-crossings in dyadic-times coincides with the natural definition of the number of up-crossings $U^\star_t[a, b] = \sup\{U_F[a, b] : F \text{ a finite subset of } [0, t]\}$, where $U_F[a, b]$ is the number of up-crossings of $[a, b]$ by $\{X_s(\omega), s \in F\}$ However, for example the non-random S.P. $X_t = I_{t\in\mathbb{Q}}$ has zero up-crossings in dyadic-times, while $U^\star_t[a, b] = \infty$ for any $1 > b > a > 0$ and $t > 0$. Also, $U^\star_t[a, b]$ may be non-measurable on $(\Omega, \mathcal{F})$ in the absence of right continuity of the underlying S.P.

Lemma 8.2.19 (Doob's up-crossing inequality)**.** *If* $\{X_s, s \geq 0\}$ *is a sup-MG, then for any* $t \geq 0$*,*

$$(b - a)\mathbf{E}(U_t[a, b]) \leq \mathbf{E}[(X_t - a)_-] - \mathbf{E}[(X_0 - a)_-] \qquad \forall a < b\,. \tag{8.2.6}$$

Proof. Fix $b > a$ and $t \geq 0$. Since $\{\mathbb{Q}^{(2,\ell)}_{t^+}\}$ is a non-decreasing sequence of finite sets, by definition $\ell \mapsto U_{t,\ell}$ is non-decreasing and by monotone convergence it suffices to show that for all ℓ,

$$(b - a)\mathbf{E}(\, U_{t,\ell}[a, b]\,) \leq \mathbf{E}[(X_t - a)_-] - \mathbf{E}[(X_0 - a)_-]\,.$$

Recall Remark 8.2.13 that enumerating $s_k \in \mathbb{Q}^{(2,\ell)}_{t^+}$ in a non-decreasing order produces a discrete time sup-MG $\{X_{s_k}, k = 0, \dots, n\}$ with $s_0 = 0$ and $s_n = t$, so this is merely Doob's up-crossing inequality (5.2.6). $\square$

Since Doob's maximal and up-crossing inequalities apply for any right-continuous sub-MG (and sup-MG), so do most convergence results we have deduced from them in Section 5.3. For completeness, we provide a short summary of these results (and briefly outline how to adapt their proofs), starting with Doob's a.s. convergence theorem.

Theorem 8.2.20 (Doob's convergence theorem)**.** *Suppose right-continuous sup-MG* $\{X_t, t \geq 0\}$ *is such that* $\sup_t\{\mathbf{E}[(X_t)_-]\} < \infty$*. Then,* $X_t \overset{a.s.}{\to} X_\infty$ *and* $\mathbf{E}|X_\infty| \leq \liminf_t \mathbf{E}|X_t|$ *is finite.*

Proof. Let $U_\infty[a, b] = \sup_{n\in\mathbb{Z}_+} U_n[a, b]$. Paralleling the proof of Theorem 5.3.2, in view of our assumption that $\sup_t\{\mathbf{E}[(X_t)_-]\}$ is finite, it follows from Lemma 8.2.19 and monotone convergence that $\mathbf{E}(U_\infty[a, b])$ is finite for each $b > a$. Hence, w.p.1. the variables $U_\infty[a, b](\omega)$ are finite for all $a, b \in \mathbb{Q}$, $a < b$. By sample path right-continuity and diagonal selection, in the set

$$\Gamma_{a,b} = \{\omega : \liminf_{t\to\infty} X_t(\omega) < a < b < \limsup_{t\to\infty} X_t(\omega)\}\,,$$

it suffices to consider $t \in \mathbb{Q}^{(2)}$, hence $\Gamma_{a,b} \in \mathcal{F}$. Further, if $\omega \in \Gamma_{a,b}$, then

$$X_{q_{2k-1}}(\omega) < a < b < X_{q_{2k}}(\omega)\,,$$

for some dyadic rationals $q_k \uparrow \infty$, hence $U_\infty[a, b](\omega) = \infty$. Consequently, the a.s. convergence of X_t to X_∞ follows as in the proof of Lemma 5.3.1. Finally, the stated bound on $\mathbf{E}|X_\infty|$ is then derived exactly as in the proof of Theorem 5.3.2. $\square$

Remark. Similarly to Exercise 5.3.3, for right-continuous sub-MG $\{X_t\}$ the finiteness of $\sup_t \mathbf{E}|X_t|$, of $\sup_t \mathbf{E}[(X_t)_+]$ and of $\liminf_t \mathbf{E}|X_t|$ are equivalent to each other and to the existence of a finite limit for $\mathbf{E}|X_t|$ (or equivalently, for $\lim_t \mathbf{E}[(X_t)_+]$), each of which further implies that $X_t \stackrel{a.s.}{\to} X_\infty$ integrable. Replacing $(X_t)_+$ by $(X_t)_-$ the same applies for sup-MGs. In particular, any non-negative, right-continuous, sup-MG $\{X_t, t \geq 0\}$ converges a.s. to integrable X_∞ such that $\mathbf{E}X_\infty \leq \mathbf{E}X_0$.

Note that Doob's convergence theorem does not apply for the Wiener process $\{W_t, t \geq 0\}$ (as $\mathbf{E}[(W_t)_+] = \sqrt{t/(2\pi)}$ is unbounded). Indeed, as we see in Exercise 8.2.35, almost surely, $\limsup_{t\to\infty} W_t = \infty$ and $\liminf_{t\to\infty} W_t = -\infty$. That is, the magnitude of oscillations of the Brownian sample path grows indefinitely.

In contrast, Doob's convergence theorem allows you to extend Doob's inequality (8.2.2) to the maximal value of a U.I. right-continuous sub-MG over *all* $t \geq 0$.

Exercise 8.2.21. *Let $M_\infty = \sup_{s\geq 0} X_s$ for a U.I. right-continuous sub-MG $\{X_t, t \geq 0\}$. Show that $X_t \stackrel{a.s.}{\to} X_\infty$ integrable and for any $x > 0$,*

$$\mathbf{P}(M_\infty \geq x) \leq x^{-1}\mathbf{E}[X_\infty I_{\{M_\infty \geq x\}}] \leq x^{-1}\mathbf{E}[(X_\infty)_+]\,. \tag{8.2.7}$$

Hint: *Start with (8.2.3) and adapt the proof of Corollary 5.3.4.*

The following integrability condition is closely related to L^1 convergence of right-continuous sub-MGs (and sup-MGs).

Definition 8.2.22. *We say that a sub-MG $(X_t, \mathcal{F}_t, t \geq 0)$ is* right closable*, or has a* last element *$(X_\infty, \mathcal{F}_\infty)$ if $\mathcal{F}_t \uparrow \mathcal{F}_\infty$ and $X_\infty \in L^1(\Omega, \mathcal{F}_\infty, \mathbf{P})$ is such that for any $t \geq 0$, almost surely $\mathbf{E}[X_\infty|\mathcal{F}_t] \geq X_t$. A similar definition applies for a sup-MG, but with $\mathbf{E}[X_\infty|\mathcal{F}_t] \leq X_t$ and for a MG, in which case we require that $\mathbf{E}[X_\infty|\mathcal{F}_t] = X_t$, namely, that $\{X_t\}$ is a* Doob's martingale *of X_∞ with respect to $\{\mathcal{F}_t\}$ (see Definition 5.3.13).*

Building upon Doob's convergence theorem, we extend Theorem 5.3.12 and Corollary 5.3.14, showing that for right-continuous MGs the properties of having a last element, uniform integrability and L^1 convergence, are equivalent to each other.

Proposition 8.2.23. *The following conditions are equivalent for a right-continuous non-negative sub-MG $\{X_t, t \geq 0\}$:*

(a) *$\{X_t\}$ is U.I.;*
(b) *$X_t \stackrel{L^1}{\to} X_\infty$;*
(c) *$X_t \stackrel{a.s.}{\to} X_\infty$ a last element of $\{X_t\}$.*

Further, even without non-negativity (a) $\iff$ (b) $\implies$ (c) and a right-continuous MG has any, hence all, of these properties, if and only if it is a Doob martingale.

Remark. By definition, any non-negative sup-MG has a last element $X_\infty = 0$ (and obviously, the same applies for any non-positive sub-MG), but many non-negative sup-MGs are not U.I. (for example, any non-degenerate critical branching process is such, as explained in the remark following the proof of Proposition 5.5.5). So, whereas a MG with a last element is U.I. this is not always the case for sub-MGs (and sup-MGs).

PROOF. *(a) $\iff$ (b):* U.I. implies L^1-boundedness which for a right-continuous sub-MG yields by Doob's convergence theorem the convergence a.s., and hence in probability of X_t to integrable X_∞. Clearly, the L^1 convergence of X_t to X_∞

also implies such convergence in probability. Either way, recall Vitali's convergence theorem (i.e. Theorem 1.3.49), that U.I. is equivalent to L^1 convergence when $X_t \xrightarrow{p} X_\infty$. We thus deduce the equivalence of (a) and (b) for right-continuous sub-MGs, where either (a) or (b) yields the corresponding a.s. convergence.
(a) and (b) yield a last element: With X_∞ denoting the a.s. and L^1 limit of the U.I. collection $\{X_t\}$, it is left to show that $\mathbf{E}[X_\infty|\mathcal{F}_s] \geq X_s$ for any $s \geq 0$. Fixing $t > s$ and $A \in \mathcal{F}_s$, by the definition of sub-MG we have $\mathbf{E}[X_t I_A] \geq \mathbf{E}[X_s I_A]$. Further, $\mathbf{E}[X_t I_A] \to \mathbf{E}[X_\infty I_A]$ (recall part (c) of Exercise 1.3.55). Consequently, $\mathbf{E}[X_\infty I_A] \geq \mathbf{E}[X_s I_A]$ for all $A \in \mathcal{F}_s$. That is, $\mathbf{E}[X_\infty|\mathcal{F}_s] \geq X_s$.
Last element and non-negative $\Longrightarrow$ (a): Since $X_t \geq 0$ and $\mathbf{E}X_t \leq \mathbf{E}X_\infty$ finite, it follows that for any finite $t \geq 0$ and $M > 0$, by Markov's inequality $\mathbf{P}(X_t > M) \leq M^{-1}\mathbf{E}X_t \leq M^{-1}\mathbf{E}X_\infty \to 0$ as $M \uparrow \infty$. It then follows that $\mathbf{E}[X_\infty I_{\{X_t>M\}}]$ converges to zero as $M \uparrow \infty$, uniformly in t (recall part (b) of Exercise 1.3.43). Further, by definition of the last element we have that $\mathbf{E}[X_t I_{\{X_t>M\}}] \leq \mathbf{E}[X_\infty I_{\{X_t>M\}}]$. Therefore, $\mathbf{E}[X_t I_{\{X_t>M\}}]$ also converges to zero as $M \uparrow \infty$, uniformly in t, i.e. $\{X_t\}$ is U.I.
Equivalence for MGs: For a right-continuous MG the equivalent properties (a) and (b) imply the a.s. convergence to X_∞ such that for any fixed $t \geq 0$, a.s. $X_t \leq \mathbf{E}[X_\infty|\mathcal{F}_t]$. Applying this also for the right-continuous MG $\{-X_t\}$ we deduce that X_∞ is a last element of the Doob's martingale $X_t = \mathbf{E}[X_\infty|\mathcal{F}_t]$. To complete the proof recall that any Doob's martingale is U.I. (see Proposition 4.2.33). $\square$

Finally, paralleling the proof of Proposition 5.3.21, upon combining Doob's convergence theorem 8.2.20 with Doob's L^p maximal inequality (8.2.5) we arrive at Doob's L^p MG convergence.

Proposition 8.2.24 (Doob's L^p martingale convergence)**.**
If right-continuous MG $\{X_t, t \geq 0\}$ is L^p-bounded for some $p > 1$, then $X_t \to X_\infty$ a.s. and in L^p (in particular, $\|X_t\|_p \to \|X_\infty\|_p$).

Throughout we rely on right continuity of the sample functions to control the tails of continuous time sub/sup-MGs and thereby deduce convergence properties. Of course, the interpolated MGs of Example 8.2.12 and the MGs derived in Exercise 8.2.7 out of the Wiener process are right-continuous. More generally, as shown next, for any MG the right-continuity of the filtration translates (after a modification) into RCLL sample functions, and only a little more is required for an RCLL modification in case of a sup-MG (or a sub-MG).

Theorem 8.2.25. *Suppose $(X_t, \mathcal{F}_t, t \geq 0)$ is a sup-MG with right-continuous filtration $\{\mathcal{F}_t, t \geq 0\}$ and $t \mapsto \mathbf{E}X_t$ is right-continuous. Then, there exists an RCLL modification $\{\widetilde{X}_t, t \geq 0\}$ of $\{X_t, t \geq 0\}$ such that $(\widetilde{X}_t, \mathcal{F}_t, t \geq 0)$ is a sup-MG.*

Proof. *Step 1.* To construct $\{\widetilde{X}_t, t \geq 0\}$ recall Lemma 8.2.19 that any sup-MG $\{X_t, t \geq 0\}$ has a finite expected number of up-crossings $\mathbf{E}(U_n[a,b])$ for each $b > a$ and $n \in \mathbb{Z}_+$. Hence, $\mathbf{P}(\Gamma) = 0$, where

$$\Gamma = \{\omega : U_n[a,b](\omega) = \infty, \text{ for some } n \in \mathbb{Z}_+, a, b \in \mathbb{Q}, b > a\}.$$

Further, if $\omega \in \Omega$ is such that for some $0 \leq t < n$,

$$\liminf_{q \downarrow t, q \in \mathbb{Q}^{(2)}} X_q(\omega) < \limsup_{q \downarrow t, q \in \mathbb{Q}^{(2)}} X_q(\omega),$$

then there exist $a, b \in \mathbb{Q}$, $b > a$, and a decreasing sequence $q_k \in \mathbb{Q}_n^{(2)}$ such that $X_{q_{2k}}(\omega) < a < b < X_{q_{2k-1}}(\omega)$, which in turn implies that $U_n[a,b](\omega)$ is infinite. Thus, if $\omega \notin \Gamma$ then the limits $X_{t^+}(\omega)$ of $X_q(\omega)$ over dyadic rationals $q \downarrow t$ exist at all $t \geq 0$. Considering the R.V.-s $M_n^{\pm} = \sup\{(X_q)_{\pm} : q \in \mathbb{Q}_n^{(2)}\}$ and the event

$$\Gamma_\star = \Gamma \bigcup \{\omega : M_n^{\pm}(\omega) = \infty, \text{ for some } n \in \mathbb{Z}_+\}\,,$$

observe that if $\omega \notin \Gamma_\star$ then $X_{t^+}(\omega)$ are finite valued for all $t \geq 0$. Further, setting $\widetilde{X}_t(\omega) = X_{t^+}(\omega) I_{\Gamma_\star^c}(\omega)$, note that $\omega \mapsto \widetilde{X}_t(\omega)$ is measurable and finite for each $t \geq 0$. We conclude the construction by verifying that $\mathbf{P}(\Gamma_\star) = 0$. Indeed, recall that right-continuity was applied only at the end of the proof of Doob's inequality (8.2.2), so using only the sub-MG property of $(X_t)_-$ we have that for all $y > 0$,

$$\mathbf{P}(M_n^- > y) \leq y^{-1}\mathbf{E}[(X_n)_-]\,.$$

Hence, M_n^- is a.s. finite. Starting with Doob's second inequality (5.2.3) for the sub-MG $\{-X_t\}$, by the same reasoning $\mathbf{P}(M_n^+ > y) \leq y^{-1}(\mathbf{E}[(X_n)_-] + \mathbf{E}[X_0])$ for all $y > 0$. Thus, M_n^+ is also a.s. finite and as claimed $\mathbf{P}(\Gamma_\star) = 0$.

Step 2. Recall that our convention, as in Remark 8.1.3, implies that the $\mathbf{P}$-null event $\Gamma_\star \in \mathcal{F}_0$. It then follows by the $\mathcal{F}_t$-adaptedness of $\{X_t\}$ and the preceding construction of X_{t^+}, that $\{\widetilde{X}_t\}$ is $\mathcal{F}_{t^+}$-adapted, namely $\mathcal{F}_t$-adapted (by the assumed right-continuity of $\{\mathcal{F}_t, t \geq 0\}$). Clearly, our construction of X_{t^+} yields right-continuous sample functions $t \mapsto \widetilde{X}_t(\omega)$. Further, a re-run of part of Step 1 yields the RCLL property, by showing that for any $\omega \in \Gamma_\star^c$ the sample function $t \mapsto X_{t^+}(\omega)$ has finite left limits at each $t > 0$. Indeed, otherwise there exist $a, b \in \mathbb{Q}$, $b > a$ and $s_k \uparrow t$ such that $X_{s_{2k-1}^+}(\omega) < a < b < X_{s_{2k}^+}(\omega)$. By construction of X_{t^+} this implies the existence of $q_k \in \mathbb{Q}_n^{(2)}$ such that $q_k \uparrow t$ and $X_{q_{2k-1}}(\omega) < a < b < X_{q_{2k}}(\omega)$. Consequently, in this case $U_n[a,b](\omega) = \infty$, in contradiction with $\omega \in \Gamma_\star^c$.

Step 3. Fixing $s \geq 0$, we show that $X_{s^+} = X_s$ for a.e. $\omega \notin \Gamma_\star$, hence the $\mathcal{F}_t$-adapted S.P. $\{\widetilde{X}_t, t \geq 0\}$ is a modification of the sup-MG $(X_t, \mathcal{F}_t, t \geq 0)$ and as such, $(\widetilde{X}_t, \mathcal{F}_t, t \geq 0)$ is also a sup-MG. Turning to show that $X_{s^+} \overset{a.s.}{=} X_s$, fix non-random dyadic rationals $q_k \downarrow s$ as $k \downarrow -\infty$ and recall Remark 8.2.13 that $(X_{q_k}, \mathcal{F}_{q_k}, k \in \mathbb{Z}_-)$ is a reversed sup-MG. Further, from the sup-MG property, for any $A \in \mathcal{F}_s$,

$$\sup_k \mathbf{E}[X_{q_k} I_A] \leq \mathbf{E}[X_s I_A] < \infty\,.$$

Considering $A = \Omega$, we deduce by Exercise 5.5.21 that the collection $\{X_{q_k}\}$ is U.I. and thus, the a.s. convergence of X_{q_k} to X_{s^+} yields that $\mathbf{E}[X_{q_k} I_A] \to \mathbf{E}[X_{s^+} I_A]$ (recall part (c) of Exercise 1.3.55). Moreover, $\mathbf{E}[X_{q_k}] \to \mathbf{E}[X_s]$ in view of the assumed right-continuity of $t \mapsto \mathbf{E}[X_t]$. Consequently, taking $k \downarrow -\infty$ we deduce that $\mathbf{E}[X_{s^+} I_A] \leq \mathbf{E}[X_s I_A]$ for all $A \in \mathcal{F}_s$, with equality in case $A = \Omega$. With both X_{s^+} and X_s measurable on $\mathcal{F}_s$, it thus follows that a.s. $X_{s^+} = X_s$, as claimed. $\square$

8.2.3. The optional stopping theorem. We are ready to extend the very useful Doob's optional stopping theorem (see Theorem 5.4.1), to the setting of right-continuous sub-MGs.

Theorem 8.2.26 (Doob's optional stopping).
If $(X_t, \mathcal{F}_t, t \in [0,\infty])$ is a right-continuous sub-MG with a last element $(X_\infty, \mathcal{F}_\infty)$ *in the sense of Definition 8.2.22, then for any $\mathcal{F}_t$-Markov times $\tau \geq \theta$, the integrable X_θ and X_τ are such that $\mathbf{E}X_\tau \geq \mathbf{E}X_\theta$, with equality in case of a MG.*

Remark. Recall Proposition 8.1.8 that right-continuous, $\mathcal{F}_t$-adapted $\{X_s, s \geq 0\}$ is $\mathcal{F}_t$-progressively measurable, hence also $\mathcal{F}_{t^+}$-progressively measurable. With the existence of $X_\infty \in m\mathcal{F}_\infty$, it then follows from Proposition 8.1.13 that $X_\theta \in m\mathcal{F}_{\theta^+}$ is a R.V. for any $\mathcal{F}_t$-Markov time θ (and by the same argument $X_\theta \in m\mathcal{F}_\theta$ in case θ is an $\mathcal{F}_t$-stopping time).

PROOF. Fixing $\ell \geq 1$ and setting $s_k = k2^{-\ell}$ for $k \in \mathbb{Z}_+ \cup \{\infty\}$, recall Remark 8.2.13 that $(X_{s_k}, \mathcal{F}_{s_k}, k \in \mathbb{Z}_+)$ is a discrete time sub-MG. Further, the assumed existence of a last element $(X_\infty, \mathcal{F}_\infty)$ for the sub-MG $(X_t, \mathcal{F}_t, t \geq 0)$ implies that a.s. $\mathbf{E}[X_{s_\infty}|\mathcal{F}_{s_k}] \geq X_{s_k}$ for any $k \in \mathbb{Z}_+$. With a slight abuse of notations we call $\{s_k\}$-valued R.V. τ an $\mathcal{F}_{s_k}$-stopping time if $\{\tau \leq s_k\} \in \mathcal{F}_{s_k}$ for all $k \in \mathbb{Z}_+$. Then, as explained in Remark 5.4.2, it thus follows from Theorem 5.4.1 that for any $\mathcal{F}_{s_k}$-stopping times $\tau_\ell \geq \theta_\ell$, the R.V. X_{τ_ℓ} and X_{θ_ℓ} are integrable, with

$$\mathbf{E}[X_{\tau_\ell}] \geq \mathbf{E}[X_{\theta_\ell}] \geq \mathbf{E}[X_0]\,. \tag{8.2.8}$$

In Lemma 8.1.16 we have constructed an $\mathcal{F}_{s_k}$-stopping time $\tau_\ell = 2^{-\ell}([2^\ell \tau] + 1)$ for the given $\mathcal{F}_t$-Markov time τ. Similarly, we have the $\mathcal{F}_{s_k}$-stopping time $\theta_\ell = 2^{-\ell}([2^\ell \theta] + 1)$ corresponding to the $\mathcal{F}_t$-Markov time θ. Our assumption that $\tau \geq \theta$ translates to $\tau_\ell \geq \theta_\ell$, hence the inequality (8.2.8) holds for any positive integer ℓ. By their construction $\tau_\ell(\omega) \downarrow \tau(\omega)$ and $\theta_\ell(\omega) \downarrow \theta(\omega)$ as $\ell \uparrow \infty$. Thus, by the assumed right-continuity of $t \mapsto X_t(\omega)$, we have the a.s. convergence of X_{θ_ℓ} to X_θ and of X_{τ_ℓ} to X_τ (when $\ell \to \infty$).

We claim that $(X_{\tau_{-n}}, \mathcal{F}_{\tau_{-n}}, n \in \mathbb{Z}_-)$ is a discrete time *reversed sub-MG*. Indeed, fixing $\ell \geq 2$, note that $\mathbb{Q}^{(2,\ell-1)} \cup \{\infty\}$ is a *subset* of $\mathbb{Q}^{(2,\ell)} \cup \{\infty\} = \{s_k\}$. Appealing once more to Remark 5.4.2, we can thus apply Lemma 5.4.3 for the pair $\tau_{\ell-1} \geq \tau_\ell$ of $\mathcal{F}_{s_k}$-stopping times and deduce that a.s.

$$\mathbf{E}[X_{\tau_{\ell-1}}|\mathcal{F}_{\tau_\ell}] \geq X_{\tau_\ell}\,.$$

The latter inequality holds for all $\ell \geq 2$, amounting to the claimed reversed sub-MG property. Since in addition $\inf_n \mathbf{E}X_{\tau_{-n}} \geq \mathbf{E}X_0$ is finite (see (8.2.8)), we deduce from Exercise 5.5.21 that the sequence $\{X_{\tau_\ell}\}_{\ell=1}^\infty$ is U.I. The same argument shows that $\{X_{\theta_\ell}\}_{\ell=1}^\infty$ is U.I. Hence, both sequences converge in L^1 to their respective limits X_τ and X_θ. In particular, both variables are integrable and in view of (8.2.8) they further satisfy the stated inequality $\mathbf{E}X_\tau \geq \mathbf{E}X_\theta$. $\square$

We proceed with a few of the consequences of Doob's optional theorem, starting with the extension of Lemma 5.4.3 to our setting.

Corollary 8.2.27. *If $(X_t, \mathcal{F}_t, t \in [0,\infty])$ is a right-continuous sub-MG with a* last element*, then $\mathbf{E}[X_\tau|\mathcal{F}_{\theta^+}] \geq X_\theta$ w.p.1. for any $\mathcal{F}_t$-Markov times $\tau \geq \theta$ (with equality in case of a MG), and if θ is an $\mathcal{F}_t$-stopping time, then further $\mathbf{E}[X_\tau|\mathcal{F}_\theta] \geq X_\theta$ w.p.1. (again with equality in case of a MG).*

PROOF. Fixing $A \in \mathcal{F}_{\theta^+}$, it follows as in the proof of Lemma 5.4.3 that $\eta = \theta I_A + \tau I_{A^c}$ is an $\mathcal{F}_{t^+}$-stopping time. Thus, applying Theorem 8.2.26 for $\tau \geq \eta$ we deduce that $\mathbf{E}[X_\tau] \geq \mathbf{E}[X_\eta]$. Further, X_τ and X_θ are integrable, so proceeding as in the proof of Lemma 5.4.3 we get that $\mathbf{E}[(Z^+ - X_\theta)I_A] \geq 0$ for $Z^+ = \mathbf{E}[X_\tau|\mathcal{F}_{\theta^+}]$ and all $A \in \mathcal{F}_{\theta^+}$. Recall as noted just after the statement Theorem 8.2.26, that $X_\theta \in m\mathcal{F}_{\theta^+}$ for the $\mathcal{F}_{t^+}$-stopping time θ, and consequently, a.s. $Z^+ \geq X_\theta$, as claimed.

In case θ is an $\mathcal{F}_t$-stopping time, note that by the tower property (and taking out the known I_A), also $\mathbf{E}[(Z - X_\theta)I_A] \geq 0$ for $Z = \mathbf{E}[Z^+|\mathcal{F}_\theta] = \mathbf{E}[X_\tau|\mathcal{F}_\theta]$ and all $A \in \mathcal{F}_\theta$. Here, as noted before, we further have that $X_\theta \in m\mathcal{F}_\theta$ and consequently, in this case, a.s. $Z \geq X_\theta$ as well. Finally, if $(X_t, \mathcal{F}_t, t \geq 0)$ is further a MG, combine the statement of the corollary for sub-MGs $(X_t, \mathcal{F}_t)$ and $(-X_t, \mathcal{F}_t)$ to find that a.s. $X_\theta = Z^+$ (and $X_\theta = Z$ for an $\mathcal{F}_t$-stopping time θ). $\square$

Remark 8.2.28. We refer hereafter to both Theorem 8.2.26 and its refinement in Corollary 8.2.27 as Doob's optional stopping. Clearly, both apply if the right-continuous sub-MG $(X_t, \mathcal{F}_t, t \geq 0)$ is such that a.s. $\mathbf{E}[Y|\mathcal{F}_t] \geq X_t$ for some integrable R.V. Y and each $t \geq 0$ (for by the tower property, such sub-MG has the last element $X_\infty = \mathbf{E}[Y|\mathcal{F}_\infty]$). Further, note that if τ is a bounded $\mathcal{F}_t$-Markov time, namely $\tau \in [0, T]$ for some non-random finite T, then you dispense of the requirement of a last element by considering these results for $Y_t = X_{t\wedge T}$ (whose last element Y_∞ is the integrable $X_T \in m\mathcal{F}_T \subseteq m\mathcal{F}_\infty$ and where $Y_\tau = X_\tau$, $Y_\theta = X_\theta$). As this applies whenever both τ and θ are non-random, we deduce from Corollary 8.2.27 that if X_t is a right-continuous sub-MG (or MG) for some filtration $\{\mathcal{F}_t\}$, then it is also a right-continuous sub-MG (or MG, respectively), for the corresponding filtration $\{\mathcal{F}_{t^+}\}$.

The latter observation leads to the following result about the stopped continuous time sub-MG (compare to Theorem 5.1.32).

Corollary 8.2.29. *If η is an $\mathcal{F}_t$-stopping time and $(X_t, \mathcal{F}_t, t \geq 0)$ is a right-continuous subMG (or supMG or a MG), then $X_{t\wedge\eta} = X_{t\wedge\eta(\omega)}(\omega)$ is also a right-continuous subMG (or supMG or MG, respectively), for this filtration.*

Proof. Recall part (b) of Exercise 8.1.10, that $\tau = u \wedge \eta$ is a bounded $\mathcal{F}_t$-stopping time for each $u \in [0, \infty)$. Further, fixing $s \leq u$, note that for any $A \in \mathcal{F}_s$,

$$\theta = (s \wedge \eta)I_A + (u \wedge \eta)I_{A^c} \,,$$

is an $\mathcal{F}_t$-stopping time such that $\theta \leq \tau$. Indeed, as $\mathcal{F}_s \subseteq \mathcal{F}_t$ when $s \leq t$, clearly

$$\{\theta \leq t\} = \{\eta \leq t\} \cup (A \cap \{s \leq t\}) \cup (A^c \cap \{s \leq u \leq t\}) \in \mathcal{F}_t \,,$$

for all $t \geq 0$. In view of Remark 8.2.28 we thus deduce, upon applying Theorem 8.2.26, that $\mathbf{E}[I_A X_{u\wedge\eta}] \geq \mathbf{E}[I_A X_{s\wedge\eta}]$ for all $A \in \mathcal{F}_s$. From this we conclude that the sub-MG condition $\mathbf{E}[X_{u\wedge\eta} \,|\, \mathcal{F}_s] \geq X_{s\wedge\eta}$ holds a.s. whereas the right-continuity of $t \mapsto X_{t\wedge\eta}$ is an immediate consequence of right-continuity of $t \mapsto X_t$. $\square$

In the discrete time setting we have derived Theorem 5.4.1 also for U.I. $\{X_{n\wedge\tau}\}$ and mostly used it in this form (see Remark 5.4.2). Similarly, you now prove Doob's optional stopping theorem for right-continuous sub-MG $(X_t, \mathcal{F}_t, t \geq 0)$ and $\mathcal{F}_t$-stopping time τ such that $\{X_{t\wedge\tau}\}$ is U.I.

Exercise 8.2.30. *Suppose $(X_t, \mathcal{F}_t, t \geq 0)$ is a right-continuous sub-MG.*

(a) *Fixing finite, non-random $u \geq 0$, show that for any $\mathcal{F}_t$-stopping times $\tau \geq \theta$, a.s. $\mathbf{E}[X_{u\wedge\tau}|\mathcal{F}_\theta] \geq X_{u\wedge\theta}$ (with equality in case of a MG).*
Hint: *Apply Corollary 8.2.27 for the stopped sub-MG $(X_{t\wedge u}, \mathcal{F}_t, t \geq 0)$.*
(b) *Show that if $(X_{u\wedge\tau}, u \geq 0)$ is U.I. then further X_θ and X_τ (defined as $\limsup_t X_t$ in case $\tau = \infty$), are integrable and $\mathbf{E}[X_\tau|\mathcal{F}_\theta] \geq X_\theta$ a.s. (again with equality for a MG).*
Hint: *Show that $Y_u = X_{u\wedge\tau}$ has a last element.*

Relying on Corollary 8.2.27 you can now also extend Corollary 5.4.5.

Exercise 8.2.31. *Suppose $(X_t, \mathcal{F}_t, t \geq 0)$ is a right-continuous sub-MG and $\{\tau_k\}$ is a non-decreasing sequence of $\mathcal{F}_t$-stopping times. Show that if $(X_t, \mathcal{F}_t, t \geq 0)$ has a last element or $\sup_k \tau_k \leq T$ for some non-random finite T, then $(X_{\tau_k}, \mathcal{F}_{\tau_k}, k \in \mathbb{Z}_+)$ is a discrete time sub-MG.*

Next, restarting a right-continuous sub-MG at a stopping time yields another sub-MG and an interesting formula for the distribution of the supremum of certain non-negative MGs.

Exercise 8.2.32. *Suppose $(X_t, \mathcal{F}_t, t \geq 0)$ is a right-continuous sub-MG and that τ is a bounded $\mathcal{F}_t$-stopping time.*

(a) *Verify that if $\mathcal{F}_t$ is a right-continuous filtration, then so is $\mathcal{G}_t = \mathcal{F}_{t+\tau}$.*
(b) *Taking $Y_t = X_{\tau+t} - X_\tau$, show that $(Y_t, \mathcal{G}_t, t \geq 0)$ is a right-continuous sub-MG.*

Exercise 8.2.33. *Consider a non-negative MG $\{Z_t, t \geq 0\}$ of continuous sample functions, such that $Z_0 = 1$ and $Z_t \stackrel{a.s.}{\to} 0$ as $t \to \infty$. Show that for any $x > 1$,*

$$\mathbf{P}(\sup_{t>0} Z_t \geq x) = x^{-1} .$$

Exercise 8.2.34. *Using Doob's optional stopping theorem re-derive Doob's inequality. Namely, show that for $t, x > 0$ and right-continuous sub-MG $\{X_s, s \geq 0\}$,*

$$\mathbf{P}(\sup_{0 \leq s \leq t} X_s > x) \leq x^{-1} \mathbf{E}[(X_t)_+] .$$

Hint: *Consider the sub-MG $((X_{u \wedge t})_+, \mathcal{F}^{\mathbf{X}}_u)$, the $\mathcal{F}^{\mathbf{X}}_u$-Markov time $\theta = \inf\{s \geq 0 : X_s > x\}$, and $\tau = \infty$.*

We conclude this sub-section with concrete applications of Doob's optional stopping theorem in the context of first hitting times for the Wiener process $(W_t, t \geq 0)$ of Definition 7.3.12.

Exercise 8.2.35. *For $r \leq 0$ consider $Z^{(r)}_t = W_t + rt$, the Brownian motion with drift (and continuous sample functions), starting at $Z^{(r)}_0 = 0$.*

(a) *Check that the first hitting time $\tau^{(r)}_b = \inf\{t \geq 0 : Z^{(r)}_t \geq b\}$ of level $b > 0$, is an $\mathcal{F}^{\mathbf{W}}_t$-stopping time.*
(b) *For $s > 0$ set $\theta(r, s) = \sqrt{r^2 + 2s} - r$ and show that*

$$\mathbf{E}[\exp(-s\tau^{(r)}_b)] = \exp(-\theta(r, s)b) .$$

Hint: *Check that $\frac{1}{2}\theta^2 + r\theta - s = 0$ at $\theta = \theta(r, s)$, then stop the martingale $u_0(t, W_t, \theta(r, s))$ of Exercise 8.2.7 at $\tau^{(r)}_b$.*
(c) *Letting $s \downarrow 0$ deduce that $\mathbf{P}(\tau^{(r)}_b < \infty) = \exp(2rb)$.*
(d) *Considering now $r = 0$ and $b \uparrow \infty$, deduce that a.s. $\limsup_{t\to\infty} W_t = \infty$ and $\liminf_{t\to\infty} W_t = -\infty$.*

Exercise 8.2.36. *Consider the exit time $\tau^{(r)}_{a,b} = \inf\{t \geq 0 : Z^{(r)}_t \notin (-a, b)\}$ of an interval, for the S.P. $Z^{(r)}_t = W_t + rt$ of continuous sample functions, where $W_0 = 0$, $r \in \mathbb{R}$ and $a, b > 0$ are finite non-random.*

(a) *Check that* $\tau_{a,b}^{(r)}$ *is a.s. finite* $\mathcal{F}_t^{\mathbf{W}}$*-stopping time and show that for any* $r \neq 0$,

$$\mathbf{P}(Z^{(r)}_{\tau^{(r)}_{a,b}} = -a) = 1 - \mathbf{P}(Z^{(r)}_{\tau^{(r)}_{a,b}} = b) = \frac{1 - e^{-2rb}}{e^{2ra} - e^{-2rb}}\,,$$

while $\mathbf{P}(W_{\tau^{(0)}_{a,b}} = -a) = b/(a+b)$.

Hint: *For* $r \neq 0$ *consider* $u_0(t, W_t, -2r)$ *of Exercise 8.2.7 stopped at* $\tau^{(r)}_{a,b}$.

(b) *Show that for all* $s \geq 0$

$$\mathbf{E}(e^{-s\tau^{(0)}_{a,b}}) = \frac{\sinh(a\sqrt{2s}) + \sinh(b\sqrt{2s})}{\sinh((a+b)\sqrt{2s})}\,.$$

Hint: *Stop the MGs* $u_0(t, W_t, \pm\sqrt{2s})$ *of Exercise 8.2.7 at* $\tau_{a,b} = \tau^{(0)}_{a,b}$.

(c) *Deduce that* $\mathbf{E}\tau_{a,b} = ab$ *and* $\mathsf{Var}(\tau_{a,b}) = \frac{ab}{3}(a^2 + b^2)$.
Hint: *Recall part (b) of Exercise 3.2.40.*

Here is a related result about first hitting time of spheres by a standard d-dimensional Brownian motion.

Definition 8.2.37. *The* standard d-dimensional Brownian motion *is the* $\mathbb{R}^d$*-valued S.P.* $\{\underline{W}(t), t \geq 0\}$ *such that* $\underline{W}(t) = (W_1(t), \ldots, W_d(t))$ *with* $\{W_i(t), t \geq 0\}$, $i = 1, 2, \ldots, d$ *mutually independent, standard (one-dimensional) Wiener processes. It is clearly a MG and a centered* $\mathbb{R}^d$*-valued Gaussian S.P. of continuous sample functions and stationary, independent increments.*

Exercise 8.2.38. *Let* $\mathcal{F}_t^{\mathbf{W}} = \sigma(\underline{W}(s), s \leq t)$ *denote the canonical filtration of a standard* k*-dimensional Brownian motion,* $R_t = \|\underline{W}(t)\|_2$ *its Euclidean distance from the origin and* $\theta_b = \inf\{t \geq 0 : R_t \geq b\}$ *the corresponding first hitting time of a sphere of radius* $b > 0$ *centered at the origin.*

(a) *Show that* $M_t = R_t^2 - kt$ *is an* $\mathcal{F}_t^{\mathbf{W}}$*-martingale of continuous sample functions and that* θ_b *is an a.s. finite* $\mathcal{F}_t^{\mathbf{W}}$*-stopping time.*

(b) *Deduce that* $\mathbf{E}[\theta_b] = b^2/k$.

Remark. The S.P. $\{R_t, t \geq 0\}$ of the preceding exercise is called the *Bessel process with dimension* k. Though we shall not do so, it can be shown that the S.P. $B_t = R_t - \nu\int_0^t R_s^{-1} ds$ is well-defined and in fact is a standard Wiener process (c.f. [**KaS97**, Proposition 3.3.21]), with $\nu = (k-1)/2$ the corresponding *index* of the Bessel process. The Bessel process is thus defined for all $\nu \geq 1/2$ (and starting at $R_0 = r > 0$, also for $0 < \nu < 1/2$). One can then further show that if $R_0 = r > 0$ then $\mathbf{P}_r(\inf_{t \geq 0} R_t > 0) = I_{\{\nu > 1/2\}}$ (hence the k-dimensional Brownian motion is O-transient for $k \geq 3$, see Definition 6.3.21), and $\mathbf{P}_r(R_t > 0, \text{ for all } t \geq 0) = 1$ even for the critical case of $\nu = 1/2$ (so by translation, for any given point $z \in \mathbb{R}^2$, the two-dimensional Brownian path, starting at any position other than $z \in \mathbb{R}^2$ w.p.1. enters every disc of positive radius centered at z but never reaches the point z).

8.2.4. Doob-Meyer decomposition and square-integrable martingales. In this section we study the structure of *square-integrable martingales* and in particular the roles of the corresponding predictable compensator and quadratic variation. In doing so, we fix throughout the probability space $(\Omega, \mathcal{F}, \mathbf{P})$ and a right-continuous filtration $\{\mathcal{F}_t\}$ on it, augmented so that every $\mathbf{P}$-null set is in $\mathcal{F}_0$ (see Remark 8.1.3).

Definition 8.2.39. *We denote by $\mathcal{M}_2$ the vector space of all square-integrable martingales $\{X_t, t \geq 0\}$ for the fixed right-continuous filtration, which start at $X_0 = 0$ and have right-continuous sample functions. We further denote by $\mathcal{M}_2^c$ the linear subspace of $\mathcal{M}_2$ consisting of those square-integrable martingales whose sample functions are* continuous *(and as before $X_0 = 0$).*

As in the discrete time setting of Section 5.3.2, the key to the study of a square-integrable martingale $X \in \mathcal{M}_2$ is the *Doob-Meyer decomposition* of X_t^2 to the sum of a martingale and the predictable quadratic variation $\langle X \rangle_t$. More generally, the Doob-Meyer decomposition is the continuous time analog of Doob's decomposition of any discrete time integrable process as the sum of a martingale and a predictable sequence. The extension of the concept of predictable S.P. to the continuous time setting is quite subtle and outside our scope, but recall Exercise 5.2.2 that when decomposing a sub-MG, the non-martingale component should be an *increasing process*, as defined next.

Definition 8.2.40. *An $\mathcal{F}_t$-adapted, integrable S.P. $\{A_t, t \geq 0\}$ of right-continuous, non-decreasing sample functions starting at $A_0 = 0$, is called an* increasing process *(or more precisely, an $\mathcal{F}_t$-increasing process).*

Remark. An increasing process is obviously a non-negative, right-continuous, sub-MG. By monotonicity, $A_\infty = \lim_{t\to\infty} A_t$ is a well defined random variable, and due to Proposition 8.2.23, integrability of A_∞ is equivalent to $\{A_t\}$ being U.I. which in turn is equivalent to this sub-MG having a last-element (i.e. being right closable).

Recall the notion of q-th variation of a function $f : [a, b] \mapsto \mathbb{R}$, with $q > 0$ a parameter, which we next extend to the q-th variation of continuous time S.P.-s.

Definition 8.2.41. *For any finite partition $\pi = \{a = s_0^{(\pi)} < s_1^{(\pi)} < \cdots < s_k^{(\pi)} = b\}$ of $[a, b]$, let $\|\pi\| = \max_{i=1}^k \{s_i^{(\pi)} - s_{i-1}^{(\pi)}\}$ denote the length of the longest interval in π and*

$$V_{(\pi)}^{(q)}(f) = \sum_{i=1}^{k} |f(s_i^{(\pi)}) - f(s_{i-1}^{(\pi)})|^q$$

denote the q-th variation of the function $f(\cdot)$ on the partition π. The q-th variation *of $f(\cdot)$ on $[a, b]$ is then the $[0, \infty]$-valued*

$$V^{(q)}(f) = \lim_{\|\pi\|\to 0} V_{(\pi)}^{(q)}(f) \ , \tag{8.2.9}$$

provided such limit exists (namely, the same *$\overline{\mathbb{R}}$-valued limit exists along each sequence $\{\pi_n, n \geq 1\}$ such that $\|\pi_n\| \to 0$). Similarly, the q-th variation on $[a, b]$ of a S.P. $\{X_t, t \geq 0\}$, denoted $V^{(q)}(X)$ is the limit* in probability *of $V_{(\pi)}^{(q)}(X_\cdot(\omega))$ per (8.2.9), if such a limit exists, and when this occurs for any compact interval $[0, t]$, we have the q-th variation, denoted $V^{(q)}(X)_t$, as a* stochastic process *with non-negative, non-decreasing sample functions, such that $V^{(q)}(X)_0 = 0$.*

Remark. As you are soon to find out, of most relevance here is the case of q-th variation for $q = 2$, which is also called the *quadratic variation*. Note also that $V_{(\pi)}^{(1)}(f)$ is bounded above by the *total variation* of the function f, namely $V(f) = \sup\{V_{(\pi)}^{(1)}(f) : \pi \text{ a finite partition of } [a, b]\}$ (which induces at each interval a norm on the linear subspace of functions of finite total variation, see also the related

Definition 3.2.22 of total variation norm for finite signed measures). Further, as you show next, if $V^{(1)}(f)$ exists then it equals to $V(f)$ (but beware that $V^{(1)}(f)$ may not exist, for example, in case $f(t) = \mathbf{1}_{\mathbb{Q}}(t)$).

Exercise 8.2.42.

(a) *Show that if* $f : [a,b] \mapsto \mathbb{R}$ *is* monotone *then* $V^{(1)}_{(\pi)}(f) = \max_{t\in[a,b]}\{f(t)\} - \min_{t\in[a,b]}\{f(t)\}$ *for any finite partition* π*, so in this case* $V^{(1)}(f) = V(f)$ *is finite.*

(b) *Show that* $\pi \mapsto V^{(1)}_{(\pi)}(\cdot)$ *is non-decreasing with respect to a refinement of the finite partition* π *of* $[a,b]$ *and hence, for each* f *there exist finite partitions* π_n *such that* $\|\pi_n\| \downarrow 0$ *and* $V^{(1)}_{(\pi_n)}(f) \uparrow V(f)$.

(c) *For* $f : [0,\infty) \mapsto \mathbb{R}$ *and* $t \geq 0$ *let* $V(f)_t$ *denote the value of* $V(f)$ *for the interval* $[0,t]$*. Show that if* $f(\cdot)$ *is left-continuous, then so is the non-decreasing function* $t \mapsto V(f)_t$.

(d) *Show that if* $t \mapsto X_t$ *is left-continuous, then* $V(X)_t$ *is* $\mathcal{F}^{\mathbf{X}}_t$*-progressively measurable, and* $\tau_n = \inf\{t \geq 0 : V(X)_t \geq n\}$ *are non-decreasing* $\mathcal{F}^{\mathbf{X}}_t$*-Markov times such that* $V(X)_{t\wedge\tau_n} \leq n$ *for all* n *and* t.
Hint: *Show that enough to consider for* $V(X)_t$ *the countable collection of finite partitions for which* $s^{(\pi)}_i \in \mathbb{Q}^{(2)}_{t^+}$*, then note that* $\{\tau_n < t\} = \cup_{k\geq 1}\{V(X)_{t-k^{-1}} \geq n\}$.

From the preceding exercise we see that any increasing process A_t has finite total variation, with $V(A)_t = V^{(1)}(A)_t = A_t$ for all t. This is certainly not the case for non-constant continuous martingales, as shown in the next lemma (which is also key to the uniqueness of the Doob-Meyer decomposition for sub-MGs of continuous sample path).

Lemma 8.2.43. *A martingale* M_t *of continuous sample functions and finite total variation on each compact interval, is indistinguishable from a constant.*

Remark. Sample path continuity is necessary here, for in its absence we have the compensated Poisson process $M_t = N_t - \lambda t$ which is a martingale (see Example 8.2.5), of finite total variation on compact intervals (since $V(M)_t \leq V(N)_t + V(\lambda t)_t = N_t + \lambda t$ by part (a) of Exercise 8.2.42).

Proof. Considering the martingale $\widetilde{M}_t = M_t - M_0$ such that $V(\widetilde{M})_t = V(M)_t$ for all t, we may and shall assume hereafter that $M_0 = 0$. Suppose first that $V(M)_t \leq K$ is bounded, uniformly in t and ω by a non-random finite constant. In particular, $|M_t| \leq K$ for all $t \geq 0$ and fixing a finite partition $\pi = \{0 = s_0 < s_1 < \cdots < s_k = t\}$, the discrete time martingale M_{s_i} is square integrable and as shown in part (b) of Exercise 5.1.8

$$\mathbf{E}[M_t^2] = \mathbf{E}[\sum_{i=1}^k (M_{s_i} - M_{s_{i-1}})^2] = \mathbf{E}[V^{(2)}_{(\pi)}(M)]\,.$$

By the definition of the q-th variation and our assumption that $V(M)_t \leq K$, it follows that

$$V^{(2)}_{(\pi)}(M) \leq K \sup_{i=1}^{k} |M_{s_i} - M_{s_{i-1}}| =: K D_\pi\,.$$

Taking expectation on both sides we deduce in view of the preceding identity that $\mathbf{E}[M_t^2] \leq K\mathbf{E}D_\pi$ where $0 \leq D_\pi \leq V(M)_t \leq K$ for all finite partitions π of $[0,t]$.

Further, by the uniform continuity of $t \mapsto M_t(\omega)$ on $[0,T]$ we have that $D_\pi(\omega) \to 0$ when $\|\pi\| \downarrow 0$, hence $\mathbf{E}[D_\pi] \to 0$ as $\|\pi\| \downarrow 0$ and consequently $\mathbf{E}[M_t^2] = 0$.

We have thus shown that if the continuous martingale M_t is such that $\sup_t V(M)_t$ is bounded by a non-random constant, then $M_t(\omega) = 0$ for any $t \geq 0$ and a.e. $\omega \in \Omega$. To deal with the general case, recall Remark 8.2.28 that $(M_t, \mathcal{F}^{\mathbf{M}}_{t^+}, t \geq 0)$ is a continuous martingale, hence by Corollary 8.2.29 and part (d) of Exercise 8.2.42 so is $(M_{t\wedge\tau_n}, \mathcal{F}^{\mathbf{M}}_{t^+}, t \geq 0)$, where $\tau_n = \inf\{t \geq 0 : V(M)_t \geq n\}$ are non-decreasing and $V(M)_{t\wedge\tau_n} \leq n$ for all n and t. Consequently, for any $t \geq 0$, w.p.1. $M_{t\wedge\tau_n} = 0$ for $n = 1, 2, \ldots$. The assumed finiteness of $V(M)_t(\omega)$ implies that $\tau_n \uparrow \infty$, hence $M_{t\wedge\tau_n} \to M_t$ as $n \to \infty$, resulting with $M_t(\omega) = 0$ for a.e. $\omega \in \Omega$. Finally, by the continuity of $t \mapsto M_t(\omega)$, the martingale M must then be indistinguishable from the zero stochastic process (see Exercise 7.2.3). □

Considering a bounded, continuous martingale X_t, the next lemma allows us to conclude in the sequel that $V^{(2)}_{(\pi)}(X)$ converges in L^2 as $\|\pi\| \downarrow 0$ and its limit can be set to be an increasing process.

Lemma 8.2.44. *Suppose $X \in \mathcal{M}_2^c$. For any partition $\pi = \{0 = s_0 < s_1 < \cdots\}$ of $[0,\infty)$ with a finite number of points on each compact interval, the S.P. $M_t^{(\pi)} = X_t^2 - V_t^{(\pi)}(X)$ is an $\mathcal{F}_t$-martingale of continuous sample path, where*

$$V_t^{(\pi)}(X) = \sum_{i=1}^{k} (X_{s_i} - X_{s_{i-1}})^2 + (X_t - X_{s_k})^2\,, \quad \forall t \in [s_k, s_{k+1})\,. \tag{8.2.10}$$

If in addition $\sup_t |X_t| \leq K$ for some finite, non-random constant K, then $V^{(2)}_{(\pi_n)}(X)$ is a Cauchy sequence in $L^2(\Omega, \mathcal{F}, \mathbf{P})$ for any fixed b and finite partitions π_n of $[0,b]$ such that $\|\pi_n\| \to 0$.

Proof. (a). With $\mathcal{F}^{\mathbf{X}}_t \subseteq \mathcal{F}_t$, the $\mathcal{F}_t$-adapted process $M_t = M_t^{(\pi)}$ of continuous sample paths is integrable (by the assumed square integrability of X_t). Noting that for any $k \geq 0$ and all $s_k \leq s < t \leq s_{k+1}$,

$$M_t - M_s = X_t^2 - X_s^2 - (X_t - X_{s_k})^2 + (X_s - X_{s_k})^2 = 2X_{s_k}(X_t - X_s)\,,$$

clearly then $\mathbf{E}[M_t - M_s|\mathcal{F}_s] = 2X_{s_k}\mathbf{E}[X_t - X_s|\mathcal{F}_s] = 0$, by the martingale property of $(X_t, \mathcal{F}_t)$, which suffices for verifying that $\{M_t, \mathcal{F}_t, t \geq 0\}$ is a martingale.

(b). Utilizing these martingales, we now turn to prove the second claim of the lemma. To this end, fix two finite partitions π and π' of $[0,b]$ and let $\widehat{\pi}$ denote the partition based on the collection of points $\pi \cup \pi'$. With $U'_t = V_t^{(\pi')}(X)$ and $U_t = V_t^{(\pi)}(X)$, applying part (a) of the proof for the martingale $Z_t = M_t^{(\pi)} - M_t^{(\pi')} = U'_t - U_t$ (which is square-integrable by the assumed boundedness of X_t), we deduce that $Z_t^2 - V_t^{(\widehat{\pi})}(Z)$, $t \in [0,b]$ is a martingale whose value at $t = 0$ is zero. Noting that $Z_b = V^{(2)}_{(\pi')}(X) - V^{(2)}_{(\pi)}(X)$, it then follows that

$$\mathbf{E}\Big[\big(V^{(2)}_{(\pi')}(X) - V^{(2)}_{(\pi)}(X)\big)^2\Big] = \mathbf{E}[Z_b^2] = \mathbf{E}[V_b^{(\widehat{\pi})}(Z)]\,.$$

Next, recall (8.2.10) that $V_b^{(\widehat{\pi})}(Z)$ is a finite sum of terms of the form $(U'_u - U'_s - U_u + U_s)^2 \leq 2(U'_u - U'_s)^2 + 2(U_u - U_s)^2$. Consequently, $V^{(\widehat{\pi})}(Z) \leq 2V^{(\widehat{\pi})}(U') + 2V^{(\widehat{\pi})}(U)$ and to conclude that $V^{(2)}_{(\pi_n)}(X)$ is a Cauchy sequence in $L^2(\Omega, \mathcal{F}, \mathbf{P})$ for any finite partitions π_n of $[0,b]$ such that $\|\pi_n\| \to 0$, it suffices to show that $\mathbf{E}[V_b^{(\widehat{\pi})}(U)] \to 0$ as $\|\pi'\| \vee \|\pi\| \to 0$.

To establish the latter claim, note first that since $\widehat{\pi}$ is a refinement of π, each interval $[t_j, t_{j+1}]$ of $\widehat{\pi}$ is contained within some interval $[s_i, s_{i+1}]$ of π, and then

$$U_{t_{j+1}} - U_{t_j} = (X_{t_{j+1}} - X_{s_i})^2 - (X_{t_j} - X_{s_i})^2 = (X_{t_{j+1}} - X_{t_j})(X_{t_{j+1}} + X_{t_j} - 2X_{s_i})$$

(see (8.2.10)). Since $t_{j+1} - s_i \le \|\pi\|$, this implies in turn that

$$(U_{t_{j+1}} - U_{t_j})^2 \le 4(X_{t_{j+1}} - X_{t_j})^2 [\mathsf{osc}_{\|\pi\|}(X)]^2 \,,$$

where $\mathsf{osc}_\delta(X) = \sup\{|X_t - X_s| : |t-s| \le \delta,\, t,s \in [0,b]\}$. Consequently,

$$V_b^{(\widehat{\pi})}(U) \le 4V_b^{(\widehat{\pi})}(X)[\mathsf{osc}_{\|\pi\|}(X)]^2$$

and by the Cauchy-Schwarz inequality

$$\Big[\mathbf{E}V_b^{(\widehat{\pi})}(U)\Big]^2 \le 16\mathbf{E}\Big[\big(V_b^{(\widehat{\pi})}(X)\big)^2\Big]\mathbf{E}\Big[\big(\,\mathsf{osc}_{\|\pi\|}(X)\big)^4\Big]\,.$$

The random variables $\mathsf{osc}_\delta(X)$ are uniformly (in δ and ω) bounded (by $2K$) and converge to zero as $\delta \downarrow 0$ (in view of the uniform continuity of $t \mapsto X_t$ on $[0,b]$). Thus, by bounded convergence the right-most expectation in the preceding inequality goes to zero as $\|\pi\| \to 0$. To complete the proof simply note that $V_b^{(\widehat{\pi})}(X)$ is of the form $\sum_{j=1}^{\ell} D_j^2$ for the differences $D_j = X_{t_j} - X_{t_{j-1}}$ of the uniformly bounded discrete time martingale $\{X_{t_j}\}$, hence $\mathbf{E}[\big(V_b^{(\widehat{\pi})}(X)\big)^2] \le 6K^4$ by part (c) of Exercise 5.1.8. $\square$

Building on the preceding lemma, the following decomposition is an important special case of the more general Doob-Meyer decomposition and a key ingredient in the theory of stochastic integration.

Theorem 8.2.45. *For $X \in \mathcal{M}_2^c$, the continuous modification of $V^{(2)}(X)_t$ is the unique $\mathcal{F}_t$-increasing process $A_t = \langle X \rangle_t$ of continuous sample functions, such that $M_t = X_t^2 - A_t$ is an $\mathcal{F}_t$-martingale (also of continuous sample functions), and any two such decompositions of X_t^2 as the sum of a martingale and increasing process are indistinguishable.*

Proof. *Step 1. Uniqueness.* If $X_t^2 = M_t + A_t = N_t + B_t$ with A_t, B_t increasing processes of continuous sample paths and M_t, N_t martingales, then $Y_t = N_t - M_t = A_t - B_t$ is a martingale of continuous sample paths, starting at $Y_0 = A_0 - B_0 = 0$, such that $V(Y)_t \le V(A)_t + V(B)_t = A_t + B_t$ is finite for any t finite. From Lemma 8.2.43 we then deduce that w.p.1. $Y_t = 0$ for all $t \ge 0$ (i.e. $\{A_t\}$ is indistinguishable of $\{B_t\}$), proving the stated uniqueness of the decomposition.

Step 2. Existence of $V^{(2)}(X)_t$ when X is uniformly bounded.

Turning to construct such a decomposition, assume first that $X \in \mathcal{M}_2^c$ is uniformly (in t and ω) bounded by a non-random finite constant. Let $V_\ell(t) = V_t^{(\pi_\ell)}(X)$ of (8.2.10) for the partitions π_ℓ of $[0,\infty)$ whose elements are the dyadic $\mathbb{Q}^{(2,\ell)} = \{k2^{-\ell}, k \in \mathbb{Z}_+\}$. By definition, $V_\ell(t) = V_{(\pi'_\ell)}^{(2)}(X)$ for the partitions π'_ℓ of $[0,t]$ whose elements are the finite collections $\mathbb{Q}_{t^+}^{(2,\ell)}$ of dyadic from $\pi_\ell \cap [0,t]$ augmented by $\{t\}$. Since $\|\pi'_\ell\| \le \|\pi_\ell\| = 2^{-\ell}$, we deduce from Lemma 8.2.44 that per $t \ge 0$ fixed, $\{V_\ell(t), \ell \ge 1\}$ is a Cauchy sequence in $L^2(\Omega, \mathcal{F}, \mathbf{P})$. Recall Proposition 4.3.7 that any Cauchy sequence in $L^2(\Omega, \mathcal{F}, \mathbf{P})$ has a limit. So, in particular $V_\ell(t)$ converges in L^2, for $\ell \to \infty$, to some $U(t,\omega)$. For any (other) sequence $\widetilde{\pi}_{2n}$ of finite partitions of $[0,t]$ such that $\|\widetilde{\pi}_{2n}\| \downarrow 0$, upon interlacing $\widetilde{\pi}_{2n+1} = \pi'_n$ we further have by Lemma 8.2.44 that $V^{(2)}_{(\widetilde{\pi}_n)}(X)$ is a Cauchy, hence convergent in L^2, sequence. Its limit coincides with the sub-sequential limit $U(t,\omega)$ along $n_\ell = 2\ell+1$, which also matches the L^2 limit of $V^{(2)}_{(\widetilde{\pi}_{2n})}(X)$. As this applies for any finite partitions $\widetilde{\pi}_{2n}$ of $[0,t]$ such that $\|\widetilde{\pi}_{2n}\| \downarrow 0$, we conclude that $(t,\omega) \mapsto U(t,\omega)$ is the quadratic variation of $\{X_t\}$.

Step 3. Constructing A_t. Turning to produce a continuous modification $A_t(\omega)$ of $U(t,\omega)$, recall Lemma 8.2.44 that for each ℓ the process $M_{\ell,t} = X_t^2 - V_\ell(t)$ is an $\mathcal{F}_t$-martingale of continuous sample path. The same applies for $V_n(t) - V_m(t) = M_{m,t} - M_{n,t}$, so fixing an integer $j \geq 1$ we deduce by Doob's L^2 maximal inequality (see (8.2.5) of Corollary 8.2.16), that

$$\mathbf{E}[\|V_n - V_m\|_j^2] \leq 4\mathbf{E}[(V_n(j) - V_m(j))^2]\,,$$

where $\|f\|_j = \sup\{|f(t)| : t \in [0,j]\}$ makes $\mathbb{Y} = C([0,j])$ into a Banach space (see part (b) of Exercise 4.3.8). In view of the L^2 convergence of $V_n(j)$ we have that $\mathbf{E}[(V_n(j) - V_m(j))^2] \to 0$ as $n,m \to \infty$, hence $V_n : \Omega \mapsto \mathbb{Y}$ is a Cauchy sequence in $L^2(\Omega, \mathcal{F}, \mathbf{P}; \mathbb{Y})$, which by part (a) of Exercise 4.3.8 converges in this space to some $U_j(\cdot,\omega) \in C([0,j])$. By the preceding we further deduce that $U_j(t,\omega)$ is a continuous modification on $[0,j]$ of the pointwise L^2 limit function $U(t,\omega)$. In view of Exercise 7.2.3, for any $j' > j$ the S.P.-s U_j and $U_{j'}$ are indistinguishable on $[0,j]$, so there exists one square-integrable, continuous modification $A : \Omega \mapsto C([0,\infty))$ of $U(t,\omega)$ whose restriction to each $[0,j]$ coincides with U_j (up to one $\mathbf{P}$-null set).

Step 4. The decomposition: A_t increasing and $M_t = X_t^2 - A_t$ martingale.

First, as $V_\ell(0) = 0$ for all ℓ, also $A_0 = U(0,\omega) = 0$. We saw in Step 3 that $\|V_\ell - A\|_j \to 0$ in L^2, hence also in L^1 and consequently $\mathbf{E}[\varphi(\|V_\ell - A\|_j)] \to 0$ when $\ell \to \infty$, for $\varphi(r) = r/(1+r) \leq r$ and any fixed positive integer j. Hence,

$$\mathbf{E}[\rho(V_\ell, A)] = \sum_{j=1}^{\infty} 2^{-j}\mathbf{E}[\varphi(\|V_\ell - A\|_j)] \to 0\,,$$

as $\ell \to \infty$, where $\rho(\cdot,\cdot)$ is a metric on $C([0,\infty))$ for the topology of uniform convergence on compact intervals (see Exercise 7.2.9). To verify that A_t is an $\mathcal{F}_t$-increasing process, recall Theorem 2.2.10 that $\rho(V_{n_k}, A) \overset{a.s.}{\to} 0$ along some non-random subsequence n_k. That is, with $\mathcal{F}_0$ augmented as usual by all $\mathbf{P}$-null sets, $V_{n_k}(t,\omega) \to A_t(\omega)$ as $k \to \infty$, for *all* $t \geq 0$ and $\omega \notin N$, where $N \in \mathcal{F}_0$ is such that $\mathbf{P}(N) = 0$. Setting $A_t \equiv 0$ when $\omega \in N$, the $\mathcal{F}_t$-adaptedness of $V_{n_k}(t)$ transfers to A_t. Also, by construction $t \mapsto V_t^{(\pi)}$ is non-decreasing when restricted to the times in π. Moreover, if $q < q' \in \mathbb{Q}^{(2)}$ then for all k large enough $q, q' \in \pi_{n_k}$ implying that $V_{n_k}(q) \leq V_{n_k}(q')$. Taking $k \to \infty$ it follows that $A_q(\omega) \leq A_{q'}(\omega)$ for all $\omega \in \Omega$, thus by sample path continuity, A_t is an $\mathcal{F}_t$-increasing process.

Finally, since the $\mathcal{F}_t$-martingales $M_{\ell,t}$ converge in L^1 for $\ell \to \infty$ (and $t \geq 0$ fixed), to the $\mathcal{F}_t$-adapted process $M_t = X_t^2 - A_t$, it is easy to check that $\{M_t, \mathcal{F}_t, t \geq 0\}$ is a martingale.

Step 5. Localization. Having established the stated decomposition in case $X \in \mathcal{M}_2^c$ is uniformly bounded by a non-random constant, we remove the latter condition by *localizing* via the stopping times $\tau_r = \inf\{t \geq 0 : |X_t| \geq r\}$ for positive integers r. Indeed, note that since $X_t(\omega)$ is bounded on any compact time interval, $\tau_r \uparrow \infty$ when $r \to \infty$ (for each $\omega \in \Omega$). Further, with $X_t^{(r)} = X_{t\wedge\tau_r}$ a uniformly bounded (by r), continuous martingale (see Corollary 8.2.29), by the preceding proof we have $\mathcal{F}_t$-increasing processes $A_t^{(r)}$, each of which is the continuous modification of the quadratic variation $V^{(2)}(X^{(r)})_t$, such that $M_t^{(r)} = X_{t\wedge\tau_r}^2 - A_t^{(r)}$ are continuous $\mathcal{F}_t$-martingales. Since $\mathbf{E}[\rho(V^{(\pi_\ell)}(X^{(r)}), A^{(r)})] \to 0$ for $\ell \to \infty$ and each positive integer r, in view of Theorem 2.2.10 we get by diagonal selection the existence of a non-random sub-sequence $n_k \to \infty$ and a $\mathbf{P}$-null set $N_\star$ such that $\rho(V^{(\pi_{n_k})}(X^{(r)}), A^{(r)}) \to 0$ for $k \to \infty$, *all* r and $\omega \notin N_\star$. From (8.2.10) we note that $V_t^{(\pi)}(X^{(r)}) = V_{t\wedge\tau_r}^{(\pi)}(X)$ for any t, ω, r and π. Consequently, if $\omega \notin N_\star$ then $A_t^{(r)} = A_{t\wedge\tau_r}^{(r)}$ for all $t \ge 0$ and $A_t^{(r)} = A_t^{(r')}$ as long as $r \le r'$ and $t \le \tau_r$. Since $t \mapsto A_t^{(r)}(\omega)$ are non-decreasing, necessarily $A_t^{(r)} \le A_t^{(r')}$ for all $t \ge 0$. We thus deduce that $A_t^{(r)} \uparrow A_t$ for any $\omega \notin N_\star$ and all $t \ge 0$. Further, with $A_t^{(r)}$ independent of r as soon as $\tau_r(\omega) \ge t$, the non-decreasing sample function $t \mapsto A_t(\omega)$ inherits the continuity of $t \mapsto A_t^{(r)}(\omega)$. Taking $A_t(\omega) \equiv 0$ for $\omega \in N_\star$ we proceed to show that A_t is integrable, hence an $\mathcal{F}_t$-increasing process of continuous sample functions. To this end, fixing $u \ge 0$ and setting $Z_r = X_{u\wedge\tau_r}^2$, by monotone convergence $\mathbf{E}Z_r = \mathbf{E}M_u^{(r)} + \mathbf{E}A_u^{(r)} \uparrow \mathbf{E}A_u$ when $r \to \infty$ (as $M_t^{(r)}$ are martingales, starting at $M_0^{(r)} = 0$). Since $u \wedge \tau_r \uparrow u$ and the sample functions $t \mapsto X_t$ are continuous, clearly $Z_r \to X_u^2$. Moreover, $\sup_r |Z_r| \le (\sup_{0\le s\le u} |X_s|)^2$ is integrable (by Doob's L^2 maximal inequality (8.2.5)), so by dominated convergence $\mathbf{E}Z_r \to \mathbf{E}X_u^2$. Consequently, $\mathbf{E}A_u = \mathbf{E}X_u^2$ is finite, as claimed.

Next, fixing $t \ge 0$, $\varepsilon > 0$, $r \in \mathbb{Z}_+$ and a finite partition π of $[0,t]$, since $V_{(\pi)}^{(2)}(X) = V_{(\pi)}^{(2)}(X^{(r)})$ whenever $\tau_r \ge t$, clearly,

$$\{|V_{(\pi)}^{(2)}(X) - A_t| \ge 2\varepsilon\} \subseteq \{\tau_r < t\} \cup \{|V_{(\pi)}^{(2)}(X^{(r)}) - A_t^{(r)}| \ge \varepsilon\} \cup \{|A_t^{(r)} - A_t| \ge \varepsilon\}\,.$$

We have shown already that $V_{(\pi)}^{(2)}(X^{(r)}) \xrightarrow{p} A_t^{(r)}$ as $\|\pi\| \to 0$. Hence,

$$\limsup_{\|\pi\|\to 0} \mathbf{P}(|V_{(\pi)}^{(2)}(X) - A_t| \ge 2\varepsilon) \le \mathbf{P}(\tau_r < t) + \mathbf{P}(|A_t^{(r)} - A_t| \ge \varepsilon)$$

and considering $r \to \infty$ we deduce that $V_{(\pi)}^{(2)}(X) \xrightarrow{p} A_t$. That is, the process $\{A_t\}$ is a modification of the quadratic variation of $\{X_t\}$.

We complete the proof by verifying that the integrable, $\mathcal{F}_t$-adapted process $M_t = X_t^2 - A_t$ of continuous sample functions satisfies the martingale condition. Indeed, since $M_t^{(r)}$ are $\mathcal{F}_t$-martingales, we have for each $s \le u$ and all r that w.p.1

$$\mathbf{E}[X_{u\wedge\tau_r}^2|\mathcal{F}_s] = \mathbf{E}[A_u^{(r)}|\mathcal{F}_s] + M_s^{(r)}\,.$$

Considering $r \to \infty$ we have already seen that $X_{u\wedge\tau_r}^2 \to X_u^2$ and a.s. $A_u^{(r)} \uparrow A_u$, hence also $M_s^{(r)} \xrightarrow{a.s.} M_s$. With $\sup_r\{X_{u\wedge\tau_r}^2\}$ integrable, we get by dominated convergence of C.E. that $\mathbf{E}[X_{u\wedge\tau_r}^2|\mathcal{F}_s] \to \mathbf{E}[X_u^2|\mathcal{F}_s]$ (see Theorem 4.2.26). Similarly, $\mathbf{E}[A_u^{(r)}|\mathcal{F}_s] \uparrow \mathbf{E}[A_u|\mathcal{F}_s]$ by monotone convergence of C.E. hence w.p.1 $\mathbf{E}[X_u^2|\mathcal{F}_s] = \mathbf{E}[A_u|\mathcal{F}_s] + M_s$ for each $s \le u$, namely, $(M_t, \mathcal{F}_t)$ is a martingale. □

The following exercise shows that $X \in \mathcal{M}_2^c$ has zero q-th variation for all $q > 2$. Moreover, unless $X_t \in \mathcal{M}_2^c$ is zero throughout an interval of positive length, its q-th variation for $0 < q < 2$ is infinite with positive probability and its sample path are *then* not locally γ-Hölder continuous for any $\gamma > 1/2$.

Exercise 8.2.46.

(a) *Suppose S.P. $\{X_t, t \geq 0\}$ of continuous sample functions has an a.s. finite r-th variation $V^{(r)}(X)_t$ for each fixed $t > 0$. Show that then for each $t > 0$ and $q > r$ a.s. $V^{(q)}(X)_t = 0$ whereas if $0 < q < r$, then $V^{(q)}(X)_t = \infty$ for a.e. ω for which $V^{(r)}(X)_t > 0$.*

(b) *Show that if $X \in \mathcal{M}_2^c$ and $\widetilde{A}_t$ is a S.P. of continuous sample path and finite total variation on compact intervals, then the quadratic variation of $X_t + \widetilde{A}_t$ is $\langle X \rangle_t$.*

(c) *Suppose $X \in \mathcal{M}_2^c$ and $\mathcal{F}_t$-stopping time τ are such that $\langle X \rangle_\tau = 0$. Show that $\mathbf{P}(X_{t\wedge\tau} = 0$ for all $t \geq 0) = 1$.*

(d) *Show that if a S.P. $\{X_t, t \geq 0\}$ is locally γ-Hölder continuous on $[0,T]$ for some $\gamma > 1/2$, then its quadratic variation on this interval is zero.*

Remark. You may have noticed that so far we did not need the assumed right-continuity of $\mathcal{F}_t$. In contrast, the latter assumption plays a key role in our proof of the more general Doob-Meyer decomposition, which is to follow next.

We start by stating the necessary and sufficient condition under which a sub-MG has a Doob-Meyer decomposition, namely, it is the sum of a martingale and increasing part.

Definition 8.2.47. *An $\mathcal{F}_t$-progressively measurable (and in particular $\mathcal{F}_t$-adapted, right-continuous), S.P. $\{Y_t, t \geq 0\}$ is of* class DL *if the collection $\{Y_{u\wedge\theta}, \theta$ an $\mathcal{F}_t$-stopping time$\}$ is U.I. for each finite, non-random u.*

Theorem 8.2.48 (Doob-Meyer decomposition)**.** *A right-continuous, sub-MG $\{Y_t, t \geq 0\}$ for $\{\mathcal{F}_t\}$ admits the decomposition $Y_t = M_t + A_t$ with M_t a right-continuous $\mathcal{F}_t$-martingale and A_t an $\mathcal{F}_t$-increasing process, if and only if $\{Y_t, t \geq 0\}$ is of class DL.*

Remark 8.2.49. To extend the uniqueness of Doob-Meyer decomposition beyond sub-MGs with continuous sample functions, one has to require A_t to be a *natural process*. While we do not define this concept here, we note in passing that every continuous increasing process is a natural process and a natural process is also an increasing process (c.f. [**KaS97**, Definition 1.4.5]), whereas the uniqueness is attained since if a finite linear combination of natural processes is a martingale, then it is indistinguishable from zero (c.f. proof of [**KaS97**, Theorem 1.4.10]).

Proof outline. We focus on constructing the Doob-Meyer decomposition for $\{Y_t, t \in \mathbb{I}\}$ in case $\mathbb{I} = [0,1]$. To this end, start with the right-continuous modification of the non-positive $\mathcal{F}_t$-sub-martingale $Z_t = Y_t - \mathbf{E}[Y_1|\mathcal{F}_t]$, which exists since $t \mapsto \mathbf{E}Z_t$ is right-continuous (see Theorem 8.2.25). Suppose you can find $A_1 \in L^1(\Omega, \mathcal{F}_1, \mathbf{P})$ such that

$$A_t = Z_t + \mathbf{E}[A_1|\mathcal{F}_t]\,, \tag{8.2.11}$$

is $\mathcal{F}_t$-increasing on $\mathbb{I}$. Then, $M_t = Y_t - A_t$ must be right-continuous, integrable and $\mathcal{F}_t$-adapted. Moreover, for any $t \in \mathbb{I}$,

$$M_t = Y_t - A_t = Y_t - Z_t - \mathbf{E}[A_1|\mathcal{F}_t] = \mathbf{E}[M_1|\mathcal{F}_t]\,.$$

So, by the tower property $(M_t, \mathcal{F}_t, t \in \mathbb{I})$ satisfies the martingale condition and we are done.

Proceeding to construct such A_1, fix $\ell \ge 1$ and for the (ordered) finite set $\mathbb{Q}_1^{(2,\ell)}$ of dyadic rationals recall Doob's decomposition (in Theorem 5.2.1), of the discrete time sub-MG $\{Z_{s_j}, \mathcal{F}_{s_j}, s_j \in \mathbb{Q}_1^{(2,\ell)}\}$ as the sum of a discrete time U.I. martingale $\{M_{s_j}^{(\ell)}, s_j \in \mathbb{Q}_1^{(2,\ell)}\}$ and the predictable, non-decreasing (in view of Exercise 5.2.2), finite sequence $\{A_{s_j}^{(\ell)}, s_j \in \mathbb{Q}_1^{(2,\ell)}\}$, starting with $A_0^{(\ell)} = 0$. Noting that $Z_1 = 0$, or equivalently $M_1^{(\ell)} = -A_1^{(\ell)}$, it follows that for any $q \in \mathbb{Q}_1^{(2,\ell)}$

$$A_q^{(\ell)} = Z_q - M_q^{(\ell)} = Z_q - \mathbf{E}[M_1^{(\ell)}|\mathcal{F}_q] = Z_q + \mathbf{E}[A_1^{(\ell)}|\mathcal{F}_q]\,. \tag{8.2.12}$$

Relying on the fact that the sub-MG $\{Y_t, t \in \mathbb{I}\}$ is of class DL, this representation allows one to deduce that the collection $\{A_1^{(\ell)}, \ell \ge 1\}$ is U.I. (for details see [**KaS97**, proof of Theorem 1.4.10]). This in turn implies by the Dunford-Pettis compactness criterion that there exists an integrable A_1 and a non-random sub-sequence $n_k \to \infty$ such that $A_1^{(n_k)} \stackrel{wL^1}{\to} A_1$, as in Definition 4.2.31. Now consider the $\mathcal{F}_t$-adapted, integrable S.P. defined via (8.2.11), where by Theorem 8.2.25 (and the assumed right continuity of the filtration $\{\mathcal{F}_t\}$), we may and shall assume that the U.I. MG $\mathbf{E}[A_1|\mathcal{F}_t]$ has right-continuous sample functions (and hence, so does $t \mapsto A_t$). Since $\mathbb{Q}_1^{(2,\ell)} \uparrow \mathbb{Q}_1^{(2)}$, upon comparing (8.2.11) and (8.2.12) we find that for any $q \in \mathbb{Q}_1^{(2)}$ and all ℓ large enough

$$A_q^{(\ell)} - A_q = \mathbf{E}[A_1^{(\ell)} - A_1|\mathcal{F}_q]\,.$$

Consequently, $A_q^{(n_k)} \stackrel{wL^1}{\to} A_q$ for all $q \in \mathbb{Q}_1^{(2)}$ (see Exercise 4.2.32). In particular, $A_0 = 0$ and setting $q < q' \in \mathbb{Q}_1^{(2)}$, $V = I_{\{A_q > A_{q'}\}}$ we deduce by the monotonicity of $j \mapsto A_{s_j}^{(\ell)}$ for each ℓ and ω, that

$$\mathbf{E}[(A_q - A_{q'})V] = \lim_{k\to\infty} \mathbf{E}[(A_q^{(n_k)} - A_{q'}^{(n_k)})V] \le 0\,.$$

So, by our choice of V necessarily $\mathbf{P}(A_q > A_{q'}) = 0$ and consequently, w.p.1. the sample functions $t \mapsto A_t(\omega)$ are non-decreasing over $\mathbb{Q}_1^{(2)}$. By right-continuity the same applies over $\mathbb{I}$ and we are done, for $\{A_t, t \in \mathbb{I}\}$ of (8.2.11) is thus indistinguishable from an $\mathcal{F}_t$-increasing process.

The same argument applies for $\mathbb{I} = [0, r]$ and any $r \in \mathbb{Z}_+$. While we do not do so here, the $\mathcal{F}_t$-increasing process $\{A_t, t \in \mathbb{I}\}$ can be further shown to be a natural process. By the uniqueness of such decompositions, as alluded to in Remark 8.2.49, it then follows that the restriction of the process $\{A_t\}$ constructed on $[0, r']$ to a smaller interval $[0, r]$ is indistinguishable from the increasing process one constructed directly on $[0, r]$. Thus, concatenating the processes $\{A_t, t \le r\}$ and $\{M_t, t \le r\}$ yields the stated Doob-Meyer decomposition on $[0, \infty)$.

As for the much easier converse, fixing non-random $u \in \mathbb{R}$, by monotonicity of $t \mapsto A_t$ the collection $\{A_{u\wedge\theta}, \theta \text{ an } \mathcal{F}_t\text{-stopping time}\}$ is dominated by the integrable A_u hence U.I. Applying Doob's optional stopping theorem for the right-continuous MG $(M_t, \mathcal{F}_t)$, you further have that $M_{u\wedge\theta} = \mathbf{E}[M_u|\mathcal{F}_\theta]$ for any $\mathcal{F}_t$-stopping time θ (see part (a) of Exercise 8.2.30), so by Proposition 4.2.33 the collection $\{M_{u\wedge\theta}, \theta$ an $\mathcal{F}_t$-stopping time$\}$ is also U.I. In conclusion, the existence of such Doob-Meyer decomposition $Y_t = M_t + A_t$ implies that the right-continuous sub-MG $\{Y_t, t \ge 0\}$ is of class DL (recall part (b) of Exercise 1.3.55). $\square$

Your next exercise provides a concrete instance in which Doob-Meyer decomposition applies, connecting it with the decomposition in Theorem 8.2.45 of the non-negative sub-MG $Y_t = X_t^2$ of continuous sample path, as the sum of the quadratic variation $\langle X\rangle_t$ and the continuous martingale $X_t^2 - \langle X\rangle_t$.

Exercise 8.2.50. *Suppose $\{Y_t, t \geq 0\}$ is a non-negative, right-continuous, sub-MG for $\{\mathcal{F}_t\}$.*

(a) *Show that Y_t is in class DL.*

(b) *Show that if Y_t further has continuous sample functions then the processes M_t and A_t in its Doob-Meyer decomposition also have continuous sample functions (and are thus unique).*

Remark. From the preceding exercise and Remark 8.2.49, we associate to each $X \in \mathcal{M}_2$ a unique natural process, denoted $\langle X\rangle_t$ and called the *predictable quadratic variation* of X, such that $X_t^2 - \langle X\rangle_t$ is a right-continuous martingale. However, when $X \notin \mathcal{M}_2^c$, it is no longer the case that the predictable quadratic variation matches the quadratic variation of Definition 8.2.41 (as a matter of fact, the latter may not exist).

Example 8.2.51. *A* standard Brownian Markov process *consists of a standard Wiener process $\{W_t, t \geq 0\}$ and filtration $\{\mathcal{F}_t, t \geq 0\}$ such that $\mathcal{F}_s^{\mathbf{W}} \subseteq \mathcal{F}_s$ for any $s \geq 0$ while $\sigma(W_t - W_s, t \geq s)$ is independent of $\mathcal{F}_s$ (see also Definition 8.3.7 for its Markov property). For right-continuous augmented filtration $\mathcal{F}_t$, such process W_t is in $\mathcal{M}_2^c$ and further, $M_t = W_t^2 - t$ is a martingale of continuous sample path. We thus deduce from Theorem 8.2.45 that its (predictable) quadratic variation is the non-random $\langle W\rangle_t = t$, which by Exercise 8.2.46 implies that the total variation of the Brownian sample path is a.s. infinite on any interval of positive length. More generally, recall part (b) of Exercise 8.2.6 that $\langle X\rangle_t$ is non-random for any Gaussian martingale, hence so is the quadratic variation of any Gaussian martingale of continuous sample functions.*

As you show next, the type of convergence to the quadratic variation may be strengthened (e.g. to convergence in L^2 or a.s.) for certain S.P. by imposing some restrictions on the partitions considered.

Exercise 8.2.52. *Let $V^{(2)}_{(\pi_n)}(W)$ denote the quadratic variations of the Wiener process on a sequence of finite partitions π_n of $[0,t]$ such that $\|\pi_n\| \to 0$ as $n \to \infty$.*

(a) *Show that $V^{(2)}_{(\pi_n)}(W) \xrightarrow{L^2} t$.*

(b) *Show that $V^{(2)}_{(\pi_n)}(W) \xrightarrow{a.s.} t$ whenever $\sum_{n=1}^{\infty} \|\pi_n\| < \infty$.*

Remark. However, beware that for a.e. $\omega \in \Omega$ there exist *random* finite partitions π_n of $[0,1]$ such that $\|\pi_n\| \to 0$ and $V^{(2)}_{(\pi_n)}(W) \to \infty$ (see [**Fre71**, Page 48]).

Example 8.2.53. *While we shall not prove it,* Lévy's martingale characterization of the Brownian motion *states the converse of Example 8.2.51, that any $X \in \mathcal{M}_2^c$ of quadratic variation $\langle X\rangle_t = t$ must be a standard Brownian Markov process (c.f.* [**KaS97**, Theorem 3.3.16]*). However, recall Example 8.2.5 that for a Poisson process N_t of rate λ, the compensated process $M_t = N_t - \lambda t$ is in $\mathcal{M}_2$ and you can easily check that $M_t^2 - \lambda t$ is then a right-continuous martingale. Since the continuous increasing process λt is natural, we deduce from the uniqueness of the Doob-Meyer decomposition that $\langle M\rangle_t = \lambda t$. More generally, by the same argument we deduce* *from part (c) of Exercise 8.2.6 that* $\langle X\rangle_t = t\mathbf{E}(X_1^2)$ *for any square-integrable S.P. with* $X_0 = 0$ *and zero-mean, stationary independent increments. In particular, this shows that sample path* continuity is necessary *for Lévy's characterization of the Brownian motion and that the standard Wiener process is the only zero-mean, square-integrable stochastic process* X_t *of continuous sample path and stationary independent increments, such that* $X_0 = 0$.

Building upon Lévy's characterization, you can now prove the following special case of the extremely useful *Girsanov's theorem*.

Exercise 8.2.54. *Suppose* $(W_t, \mathcal{F}_t, t \geq 0)$ *is a standard Brownian Markov process on a probability space* $(\Omega, \mathcal{F}, \mathbf{P})$ *and fixing a non-random parameters* $\theta \in \mathbb{R}$ *and* $T > 0$ *consider the exponential* $\mathcal{F}_t$*-martingale* $Z_t = \exp(\theta W_t - \theta^2 t/2)$ *and the corresponding probability measure* $\mathbf{Q}_T(A) = \mathbf{E}(I_A Z_T)$ *on* $(\Omega, \mathcal{F}_T)$.

(a) *Show that* $V^{(2)}(Z)_t = \theta^2 \int_0^t Z_u^2 du$.

(b) *Show that* $\widetilde{W}_u = W_u - \theta u$ *is for* $u \in [0, T]$ *an* $\mathcal{F}_u$*-martingale on the probability space* $(\Omega, \mathcal{F}_T, \mathbf{Q}_T)$.

(c) *Deduce that* $(\widetilde{W}_t, \mathcal{F}_t, t \leq T)$ *is a standard Brownian Markov process on* $(\Omega, \mathcal{F}_T, \mathbf{Q}_T)$.

Here is the extension to the continuous time setting of Lemma 5.2.7 and Proposition 5.3.30.

Exercise 8.2.55. *Let* $V_t = \sup_{s\in[0,t]} Y_s$ *and* A_t *be the increasing process of continuous sample functions in the Doob-Meyer decomposition of a non-negative, continuous,* $\mathcal{F}_t$*-submartingale* $\{Y_t, t \geq 0\}$ *with* $Y_0 = 0$.

(a) *Show that* $\mathbf{P}(V_\tau \geq x, A_\tau < y) \leq x^{-1}\mathbf{E}(A_\tau \wedge y)$ *for all* $x, y > 0$ *and any* $\mathcal{F}_t$*-stopping time* τ.

(b) *Setting* $c_1 = 4$ *and* $c_q = (2-q)/(1-q)$ *for* $q \in (0,1)$, *conclude that* $\mathbf{E}[\sup_s |X_s|^{2q}] \leq c_q \mathbf{E}[\langle X\rangle_\infty^q]$ *for any* $X \in \mathcal{M}_2^c$ *and* $q \in (0,1]$, *hence* $\{|X_t|^{2q}, t \geq 0\}$ *is U.I. when* $\langle X\rangle_\infty^q$ *is integrable.*

Taking $X, Y \in \mathcal{M}_2$ we deduce from the Doob-Meyer decomposition of $X \pm Y \in \mathcal{M}_2$ that $(X \pm Y)_t^2 - \langle X \pm Y\rangle_t$ are right-continuous $\mathcal{F}_t$-martingales. Considering their difference we deduce that $XY - \langle X, Y\rangle$ is a martingale for

$$\langle X, Y\rangle_t = \frac{1}{4}\Big[\langle X+Y\rangle_t - \langle X-Y\rangle_t\Big]$$

(this is an instance of the more general *polarization* technique). In particular, XY is a right-continuous martingale whenever $\langle X, Y\rangle = 0$, prompting our next definition.

Definition 8.2.56. *For any pair* $X, Y \in \mathcal{M}_2$, *we call the S.P.* $\langle X, Y\rangle_t$ *the* bracket *of* X *and* Y *and say that* $X, Y \in \mathcal{M}_2$ *are* orthogonal *if for any* $t \geq 0$ *the bracket* $\langle X, Y\rangle_t$ *is a.s. zero.*

Remark. It is easy to check that $\langle X, X\rangle = \langle X\rangle$ for any $X \in \mathcal{M}_2$. Further, for any $s \in [0,t]$, w.p.1.

$$\mathbf{E}[(X_t - X_s)(Y_t - Y_s)|\mathcal{F}_s] = \mathbf{E}[X_tY_t - X_sY_s|\mathcal{F}_s] = \mathbf{E}[\langle X, Y\rangle_t - \langle X, Y\rangle_s|\mathcal{F}_s],$$

so the orthogonality of $X, Y \in \mathcal{M}_2$ amounts to X and Y having uncorrelated increments over $[s,t]$, conditionally on $\mathcal{F}_s$. Here is more on the structure of the bracket as a bi-linear form on $\mathcal{M}_2$, which on $\mathcal{M}_2^c$ coincides with the *cross variation* of X and Y.

Exercise 8.2.57. *Show that for all $X, X_i, Y \in \mathcal{M}_2$:*

(a) $\langle c_1 X_1 + c_2 X_2, Y\rangle = c_1\langle X_1, Y\rangle + c_2\langle X_2, Y\rangle$ *for any $c_i \in \mathbb{R}$, $i = 1, 2$.*
Hint: *Recall Remark 8.2.49 that a martingale of the form $\sum_{j=1}^{\ell} \pm\langle U_j\rangle$ for $U_j \in \mathcal{M}_2$ and ℓ finite, is zero.*
(b) $\langle X, Y\rangle = \langle Y, X\rangle$.
(c) $|\langle X, Y\rangle|^2 \le \langle X\rangle\langle Y\rangle$.
(d) *With $Z_t = V(\langle X, Y\rangle)_t$, for a.e. $\omega \in \Omega$ and all $0 \le s < t < \infty$,*

$$Z_t - Z_s \le \frac{1}{2}[\langle X\rangle_t - \langle X\rangle_s + \langle Y\rangle_t - \langle Y\rangle_s] .$$

(e) *Show that for $X, Y \in \mathcal{M}_2^c$ the bracket $\langle X, Y\rangle_t$ is also the limit in probability as $\|\pi\| \to 0$ of*

$$\sum_{i=1}^{k} [X_{t_i^{(\pi)}} - X_{t_{i-1}^{(\pi)}}][Y_{t_i^{(\pi)}} - Y_{t_{i-1}^{(\pi)}}] ,$$

where $\pi = \{0 = t_0^{(\pi)} < t_1^{(\pi)} < \cdots < t_k^{(\pi)} = t\}$ is a finite partition of $[0, t]$.

We conclude with a brief introduction to stochastic integration (for more on this topic, see [**KaS97**, Section 3.2]). Following our general approach to integration, the *Itô stochastic integral* $I_t = \int_0^t X_s dW_s$ is constructed first for *simple processes* X_t, i.e. those having sample path that are piecewise constant on non-random intervals, as you are to do next.

Exercise 8.2.58. *Suppose $(W_t, \mathcal{F}_t)$ is a standard Brownian Markov process and X_t is a bounded, $\mathcal{F}_t$-adapted, left-continuous* simple process. *That is,*

$$X_t(\omega) = \eta_0(\omega)\mathbf{1}_{\{0\}}(t) + \sum_{i=0}^{\infty} \eta_i(\omega)\mathbf{1}_{(t_i, t_{i+1}]}(t) ,$$

where $0 = t_0 < t_1 < \cdots < t_k < \cdots$ is a non-random unbounded sequence and the $\mathcal{F}_{t_n}$-adapted sequence $\{\eta_n(\omega)\}$ is bounded uniformly in n and ω.

(a) *With $A_t = \int_0^t X_u^2 du$, show that both*

$$I_t = \sum_{j=0}^{k-1} \eta_j (W_{t_{j+1}} - W_{t_j}) + \eta_k(W_t - W_{t_k}), \text{ when } t \in [t_k, t_{k+1}) ,$$

and $I_t^2 - A_t$ are martingales with respect to $\mathcal{F}_t$.
(b) *Deduce that $I_t \in \mathcal{M}_2^c$ with $A_t = \langle I\rangle_t$ being its quadratic variation, and in particular $\mathbf{E} I_t^2 = \int_0^t \mathbf{E}[X_u^2]du$.*

8.3. Markov and Strong Markov processes

In Subsection 8.3.1 we define Markov semi-groups and the corresponding Markov processes. We also extend the construction of Markov chains from Subsection 6.1 to deal with these S.P. This is followed in Subsection 8.3.2 with the study of the strong Markov property and the related Feller property, showing in particular that both the Brownian motion and the Poisson process are strong Markov processes. We then devote Subsection 8.3.3 to the study of Markov jump processes, which are the natural extension of both Markov chains and (compound) Poisson processes.

8.3.1. Markov semi-groups, processes and the Markov property. We start with the definition of a Markov process, focusing on (time) homogeneous processes having (stationary, regular) transition probabilities (compare with Definitions 6.1.1 and 6.1.2).

Definition 8.3.1 (MARKOV PROCESSES). *A collection* $\{p_{s,t}(\cdot,\cdot), t \geq s \geq 0\}$ *of* transition probabilities *on a measurable space* $(\mathbb{S}, \mathcal{S})$ *(as in Definition 6.1.2), is* consistent *if it satisfies the* Chapman-Kolmogorov equations

$$p_{t_1,t_3}(x, B) = p_{t_1,t_2} p_{t_2,t_3}(x, B)\,, \quad \forall x \in \mathbb{S},\ B \in \mathcal{S}, \tag{8.3.1}$$

for any $t_3 \geq t_2 \geq t_1 \geq 0$ *(c.f. Corollary 6.3.3 for the composition of transition probabilities). In particular,* $p_{t,t}(x, B) = I_B(x) = \delta_x(B)$ *for any* $t \geq 0$*. Such collection is called a* Markov semi-group *(of stationary transition probabilities), if in addition* $p_{s,t} = p_{t-s}$ *for all* $t \geq s \geq 0$*. The Chapman-Kolmogorov equations are then*

$$p_{s+u}(x, B) = p_s p_u(x, B)\,, \quad \forall x \in \mathbb{S},\ B \in \mathcal{S},\ u, s \geq 0\,, \tag{8.3.2}$$

with $p_0(x, B) = I_B(x) = \delta_x(B)$ *being the semi-group identity element.*

An $\mathcal{F}_t$*-adapted S.P.* $\{X_t, t \geq 0\}$ *taking values in* $(\mathbb{S}, \mathcal{S})$ *is an* $\mathcal{F}_t$-Markov process *of (consistent) transition probabilities* $\{p_{s,t}, t \geq s \geq 0\}$ *and* state space $(\mathbb{S}, \mathcal{S})$ *if for any* $t \geq s \geq 0$ *and* $B \in \mathcal{S}$*, almost surely*

$$\mathbf{P}(X_t \in B | \mathcal{F}_s) = p_{s,t}(X_s, B)\,. \tag{8.3.3}$$

It is further a (time) homogeneous $\mathcal{F}_t$-Markov process *of semi-group* $\{p_u, u \geq 0\}$ *if for any* $u, s \geq 0$ *and* $B \in \mathcal{S}$*, almost surely*

$$\mathbf{P}(X_{s+u} \in B | \mathcal{F}_s) = p_u(X_s, B)\,. \tag{8.3.4}$$

Remark. Recall that a $\mathcal{G}_n$-Markov chain $\{Y_n\}$ is a discrete time S.P. Hence, in this case one considers only $t, s \in \mathbb{Z}_+$ and (8.3.1) is automatically satisfied by setting $p_{s,t} = p_{s,s+1} p_{s+1,s+2} \cdots p_{t-1,t}$ to be the composition of the (one-step) transition probabilities of the Markov chain (see Definition 6.1.2, with $p_{s,s+1} = p$ independent of s when the chain is homogeneous). Further, the interpolated process $X_t = Y_{[t]}$ is then a right-continuous $\mathcal{F}_t$-Markov process for the right-continuous interpolated filtration $\mathcal{F}_t = \mathcal{G}_{[t]}$ of Example 8.1.5, but $\{X_t\}$ is in general an inhomogeneous Markov process, even in case the Markov chain $\{Y_n\}$ is homogeneous.

Similarly, if $\mathcal{F}_t$-adapted S.P. $(X_t, t \geq 0)$ satisfies (8.3.4) for $p_t(x, B) = \mathbf{P}(X_t \in B | X_0 = x)$ and $x \mapsto p_t(x, B)$ is measurable per fixed $t \geq 0$ and $B \in \mathcal{S}$, then considering the tower property for $I_B(X_{s+u}) I_{\{x\}}(X_0)$ and $\sigma(X_0) \subseteq \mathcal{F}_s$, one easily verifies that (8.3.2) holds, hence $(X_t, t \geq 0)$ is a homogeneous $\mathcal{F}_t$-Markov process. More generally, in analogy with our definition of Markov chains via (6.1.1), one may opt to say that $\mathcal{F}_t$-adapted S.P. $(X_t, t \geq 0)$ is an $\mathcal{F}_t$-Markov process provided for each $B \in \mathcal{S}$ and $t \geq s \geq 0$,

$$\mathbf{P}(X_t \in B | \mathcal{F}_s) \overset{a.s.}{=} \mathbf{P}(X_t \in B | X_s)\,.$$

Indeed, as noted in Remark 6.1.6 (in view of Exercise 4.4.5), for $\mathcal{B}$-isomorphic $(\mathbb{S}, \mathcal{S})$ this suffices for the existence of transition probabilities which satisfy (8.3.3). However, this simpler to verify plausible definition of Markov processes results with Chapman-Kolmogorov equations holding only up to a null set *per fixed* $t_3 \geq t_2 \geq t_1 \geq 0$. The study of such processes is consequently made more cumbersome, which is precisely why we, like most texts, do not take this route.

By Lemma 6.1.3 we deduce from Definition 8.3.1 that for any $f \in b\mathcal{S}$ and all $t \geq s \geq 0$,

$$\mathbf{E}[f(X_t)|\mathcal{F}_s] = (p_{s,t}f)(X_s)\,, \tag{8.3.5}$$

where $f \mapsto (p_{s,t}f) : b\mathcal{S} \mapsto b\mathcal{S}$ and $(p_{s,t}f)(x) = \int p_{s,t}(x,dy)f(y)$ denotes the Lebesgue integral of $f(\cdot)$ under the probability measure $p_{s,t}(x,\cdot)$ per fixed $x \in \mathbb{S}$.

The Chapman-Kolmogorov equations are necessary and sufficient for generating consistent Markovian f.d.d. out of a given collection of transition probabilities and a specified initial probability distribution. As outlined next, we thus canonically construct the Markov process, following the same approach as in proving Theorem 6.1.8 (for Markov chain), and Proposition 7.1.8 (for continuous time S.P.).

Theorem 8.3.2. *Suppose $(\mathbb{S},\mathcal{S})$ is $\mathcal{B}$-isomorphic. Given any $(\mathbb{S},\mathcal{S})$-valued consistent transition probabilities $\{p_{s,t}, t \geq s \geq 0\}$, the probability distribution ν on $(\mathbb{S},\mathcal{S})$ uniquely determines the linearly ordered consistent f.d.d.*

$$\mu_{0,s_1,\ldots,s_n} = \nu \otimes p_{0,s_1} \otimes \cdots \otimes p_{s_{n-1},s_n} \tag{8.3.6}$$

for $0 = s_0 < s_1 < \cdots < s_n$, and there exists a Markov process of state space $(\mathbb{S},\mathcal{S})$ having these f.d.d. Conversely, the f.d.d. of any Markov process having initial probability distribution $\nu(B) = \mathbf{P}(X_0 \in B)$ *and satisfying (8.3.3), are given by (8.3.6).*

Proof. Recall Proposition 6.1.5 that $\nu \otimes p_{0,s_1} \otimes \cdots \otimes p_{s_{n-1},s_n}$ denotes the Markov-product-like measures, whose evaluation on product sets is by iterated integrations over the transition probabilities p_{s_{k-1},s_k}, in reverse order $k = n,\ldots,1$, followed by a final integration over the initial measure ν. As shown in this proposition, given any transition probabilities $\{p_{s,t}, t \geq s \geq 0\}$, the probability distribution ν on $(\mathbb{S},\mathcal{S})$ uniquely determines $\nu \otimes p_{0,s_1} \otimes \cdots \otimes p_{s_{n-1},s_n}$, namely, the f.d.d. specified in (8.3.6). We then uniquely specify the remaining f.d.d. as the probability measures $\mu_{s_1,\ldots,s_n}(D) = \mu_{s_0,s_1,\ldots,s_n}(\mathbb{S} \times D)$. Proceeding to check the consistency of these f.d.d. note that $p_{s,u} \otimes p_{u,t}(\cdot, \mathbb{S} \times \cdot) = p_{s,t}(\cdot,\cdot)$ for any $s < u < t$ (by the Chapman-Kolmogorov identity (8.3.1)). Thus, considering $s = s_{k-1}$, $u = s_k$ and $t = s_{k+1}$ we deduce that if $D = A_0 \times \cdots \times A_n$ with $A_k = \mathbb{S}$ for some $k = 1,\ldots,n-1$, then

$$\nu \otimes p_{s_0,s_1} \cdots \otimes p_{s_{n-1},s_n}(D) = \nu \otimes \cdots \otimes p_{s_{k-1},s_{k+1}} \otimes \cdots \otimes p_{s_{n-1},s_n}(D_k)$$

for $D_k = A_0 \times \cdots \times A_{k-1} \times A_{k+1} \times \cdots \times A_n$, which are precisely the consistency conditions of (7.1.3) for the f.d.d. $\{\mu_{s_0,\ldots,s_n}\}$. These consistency requirements are further handled in case of a product set D with $A_n = \mathbb{S}$ by observing that for all $x \in \mathbb{S}$ and any transition probability $p_{s_{n-1},s_n}(x,\mathbb{S}) = 1$, whereas our definition of $\mu_{s_1,\ldots,s_n}$ already dealt with $A_0 = \mathbb{S}$. Having shown that this collection of f.d.d. is consistent, recall that Proposition 7.1.8 applies even with $(\mathbb{R},\mathcal{B})$ replaced by the $\mathcal{B}$-isomorphic measurable space $(\mathbb{S},\mathcal{S})$. Setting $\mathbb{T} = [0,\infty)$, it provides the construction of a S.P. $\{Y_t(\omega) = \omega(t), t \in \mathbb{T}\}$ via the coordinate maps on the canonical probability space $(\mathbb{S}^{\mathbb{T}}, \mathcal{S}^{\mathbb{T}}, \mathbf{P}_\nu)$ with the f.d.d. of (8.3.6). Turning next to verify that $(Y_t, \mathcal{F}^{\mathbf{Y}}_t, t \in \mathbb{T})$ satisfies the Markov condition (8.3.3), fix $t \geq s \geq 0$, $B \in \mathcal{S}$ and recall that, for $t > s$ as in the proof of Theorem 6.1.8, and by definition in case $t = s$,

$$\mathbf{E}[I_{\{Y_\cdot \in A\}} I_B(Y_t)] = \mathbf{E}[I_{\{Y_\cdot \in A\}} p_{s,t}(Y_s, B)] \tag{8.3.7}$$

for any finite dimensional measurable rectangle $A = \{x(\cdot) : x(t_i) \in B_i, i = 1, \dots, n\}$ such that $t_i \in [0, s]$ and $B_i \in \mathcal{S}$. Thus, the collection

$$\mathcal{L} = \{A \in \mathcal{S}^{[0,s]} : (8.3.7) \text{ holds for } A\},$$

contains the π-system of finite dimensional measurable rectangles which generates $\mathcal{S}^{[0,s]}$, and in particular, $\mathbb{S} \in \mathcal{L}$. Further, by linearity of the expectation $\mathcal{L}$ is closed under proper difference and by monotone convergence if $A_n \in \mathcal{L}$ is such that $A_n \uparrow A$ then $A \in \mathcal{L}$ as well. Consequently, $\mathcal{L}$ is a λ-system and by Dynkin's $\pi - \lambda$ theorem, (8.3.7) holds for every set in $\mathcal{S}^{[0,s]} = \mathcal{F}_s^{\mathbf{Y}}$ (see Lemma 7.1.7). It then follows that $\mathbf{P}(Y_t \in B | \mathcal{F}_s^{\mathbf{Y}}) = p_{s,t}(Y_s, B)$ a.s. for each $t \geq s \geq 0$ and $B \in \mathcal{S}$. That is, $\{Y_t, t \geq 0\}$ is an $\mathcal{F}_t^{\mathbf{Y}}$-Markov process.

Conversely, suppose $\{X_t, t \geq 0\}$ satisfies (8.3.3) and has initial probability distribution $\nu(\cdot)$. Then, for any $t_0 > \cdots > t_n \geq s \geq 0$, and $f_\ell \in b\mathcal{S}$, $\ell = 0, \dots, n$, almost surely,

$$\mathbf{E}[\prod_{\ell=0}^{n} f_\ell(X_{t_\ell}) | \mathcal{F}_s] = \int p_{s,t_n}(X_s, dy_n) f_n(y_n) \cdots \int p_{t_1,t_0}(y_1, dy_0) f_0(y_0). \tag{8.3.8}$$

The latter identity is proved by induction on n, where denoting its right side by $g_{n+1,s}(X_s)$, we see that $g_{n+1,s} = p_{s,t_n}(f_n g_{n,t_n})$ and the case $n = 0$ is merely (8.3.5). In the induction step we have from the tower property and $\mathcal{F}_t$-adaptedness of $\{X_t\}$ that

$$\begin{aligned}\mathbf{E}[\prod_{\ell=0}^{n} f_\ell(X_{t_\ell}) | \mathcal{F}_s] &= \mathbf{E}[f_n(X_{t_n}) \mathbf{E}[\prod_{\ell=0}^{n-1} f_\ell(X_{t_\ell}) | \mathcal{F}_{t_n}] | \mathcal{F}_s] \\ &= \mathbf{E}[f_n(X_{t_n}) g_{n,t_n}(X_{t_n}) | \mathcal{F}_s] = g_{n+1,s}(X_s),\end{aligned}$$

where the induction hypothesis is used in the second equality and (8.3.5) in the third. In particular, considering the expected value of (8.3.8) for $s = 0$ and indicator functions $f_\ell(\cdot)$ it follows that the f.d.d. of this process are given by (8.3.6), as claimed. $\square$

Remark 8.3.3. As in Lemma 7.1.7, for $\mathcal{B}$-isomorphic state space $(\mathbb{S}, \mathcal{S})$ any $F \in \mathcal{F}^{\mathbf{X}}$ is of the form $F = (X_\cdot)^{-1}(A)$ for some $A \in \mathcal{S}^{\mathbb{T}}$, where $X_\cdot(\omega) : \Omega \mapsto \mathbb{S}^{\mathbb{T}}$ denote the collection of sample functions of the given Markov process $\{X_t, t \geq 0\}$. Then, $\mathbf{P}(F) = \mathbf{P}_\nu(A)$, so while proving Theorem 8.3.2 we have defined the *law* $\mathbf{P}_\nu(\cdot)$ of Markov process $\{X_t, t \geq 0\}$ as the unique probability measure on $\mathcal{S}^{[0,\infty)}$ such that

$$\mathbf{P}_\nu(\{\omega : \omega(s_\ell) \in B_\ell, \ell = 0, \dots, n\}) = \mathbf{P}(X_{s_0} \in B_0, \dots, X_{s_n} \in B_n),$$

for $B_\ell \in \mathcal{S}$ and distinct $s_\ell \geq 0$ (compare with Definition 6.1.7 for the law of a Markov chain). We denote by $\mathbf{P}_x$ the law $\mathbf{P}_\nu$ in case $\nu(B) = I_{x \in B}$, namely, when $X_0 = x$ is non-random and note that $\mathbf{P}_\nu(A) = \int_{\mathbb{S}} \mathbf{P}_x(A) \nu(dx)$ for any probability measure ν on $(\mathbb{S}, \mathcal{S})$ and all $A \in \mathcal{S}^{\mathbb{T}}$, with $\mathbf{P}_x$ uniquely determined by the specified (consistent) transition probabilities $\{p_{s,t}, t \geq s \geq 0\}$.

The evaluation of the f.d.d. of a Markov process is more explicit when $\mathbb{S}$ is a countable set, as then $p_{s,t}(x, B) = \sum_{y \in B} p_{s,t}(x, y)$ for any $B \subseteq \mathbb{S}$ (and all Lebesgue integrals are merely sums). Likewise, in case $\mathbb{S} = \mathbb{R}^d$ (equipped with $\mathcal{S} = \mathcal{B}_{\mathbb{S}}$), computations are relatively explicit if for each $t > s \geq 0$ and $x \in \mathbb{S}$ the probability measure $p_{s,t}(x, \cdot)$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{S}$, in which case $(p_{s,t} f)(x) = \int p_{s,t}(x, y) f(y) dy$ and the right side of (8.3.8) amounts to iterated integrations of the *transition probability kernel* $p_{s,t}(x,y)$ of the process with respect to Lebesgue measure on $\mathbb{S}$.

The next exercise is about the closure of the collection of Markov processes under certain invertible non-random measurable mappings.

Exercise 8.3.4. *Suppose* $(X_t, \mathcal{F}_t^{\mathbf{X}}, t \geq 0)$ *is a Markov process of state space* $(\mathbb{S}, \mathcal{S})$, $u : [0,\infty) \mapsto [0,\infty)$ *is an invertible, strictly increasing function and for each* $t \geq 0$ *the measurable mapping* $\Phi_t : (\mathbb{S}, \mathcal{S}) \mapsto (\widetilde{\mathbb{S}}, \widetilde{\mathcal{S}})$ *is invertible, with* Φ_t^{-1} *measurable.*

(a) *Setting* $Y_t = \Phi_t(X_{u(t)})$, *verify that* $\mathcal{F}_t^{\mathbf{Y}} = \mathcal{F}_{u(t)}^{\mathbf{X}}$ *and that* $(Y_t, \mathcal{F}_t^{\mathbf{Y}}, t \geq 0)$ *is a Markov process of state space* $(\widetilde{\mathbb{S}}, \widetilde{\mathcal{S}})$.

(b) *Show that if* $(X_t, \mathcal{F}_t^{\mathbf{X}}, t \geq 0)$ *is a homogeneous Markov process then so is* $Z_t = \Phi_0(X_t)$.

Of particular note is the following collection of Markov processes.

Proposition 8.3.5. *If real-valued S.P.* $\{X_t, t \geq 0\}$ *has independent increments, then* $(X_t, \mathcal{F}_t^{\mathbf{X}}, t \geq 0)$ *is a Markov process of transition probabilities* $p_{s,t}(y, B) = \mathcal{P}_{X_t - X_s}(\{z : y + z \in B\})$, *and if* $\{X_t, t \geq 0\}$ *further has stationary, independent increments, then this Markov process is homogeneous.*

Proof. Considering Exercise 4.2.2 for $\mathcal{G} = \mathcal{F}_s^{\mathbf{X}}$, $Y = X_s \in m\mathcal{G}$ and the R.V. $Z = Z_{t,s} = X_t - X_s$ which is independent of $\mathcal{G}$, you find that (8.3.3) holds for $p_{s,t}(y,B) = \mathbf{P}(y + Z \in B)$, which in case of stationary increments depends only on $t - s$. Clearly, $B \mapsto \mathbf{P}(y + Z \in B)$ is a probability measure on $(\mathbb{R}, \mathcal{B})$, for any $t \geq s$ and $y \in \mathbb{R}$. Further, if $B = (-\infty, b]$ then $p_{s,t}(y,B) = F_Z(b - y)$ is a Borel function of y (see Exercise 1.2.27). As the λ-system $\mathcal{L} = \{B \in \mathcal{B} : y \mapsto \mathbf{P}(y + Z \in B)$ is a Borel function$\}$ contains the π-system $\{(-\infty, b] : b \in \mathbb{R}\}$ generating $\mathcal{B}$, it follows that $\mathcal{L} = \mathcal{B}$, hence $p_{s,t}(\cdot,\cdot)$ is a transition probability for each $t \geq s \geq 0$. To verify that the Chapman-Kolmogorov equations hold, fix $u \in [s,t]$ noting that $Z_{s,t} = Z_{s,u} + Z_{u,t}$, with $Z_{u,t} = X_t - X_u$ independent of $Z_{s,u} = X_u - X_s$. Hence, by the tower property,

$$\begin{aligned} p_{s,t}(y,B) &= \mathbf{E}[\mathbf{P}(y + Z_{s,u} + Z_{u,t} \in B | Z_{s,u})] \\ &= \mathbf{E}[p_{u,t}(y + Z_{s,u}, B)] = (p_{s,u}(p_{u,t} I_B))(y) = p_{s,u} p_{u,t}(y,B)\,, \end{aligned}$$

and this relation, i.e. (8.3.1), holds for all $y \in \mathbb{R}$ and $B \in \mathcal{B}$, as claimed. $\square$

Among the consequences of Proposition 8.3.5 is the fact that both the Brownian motion and the Poisson process (potentially starting at $N_0 = x \in \mathbb{R}$), are homogeneous Markov processes of explicit Markov semi-groups.

Example 8.3.6. *Recall Proposition 3.4.9 and Exercise 7.3.13 that both the Poisson process and the Brownian motion are processes of stationary independent increments. Further, this property clearly extends to the* Brownian motion with drift $Z_t^{(r)} = W_t + rt + x$, *and to the* Poisson process with drift $N_t^{(r)} = N_t + rt + x$, *where the drift* $r \in \mathbb{R}$ *is a non-random constant,* $x \in \mathbb{R}$ *is the specified (under* $\mathbf{P}_x$*), initial value of* $N_0^{(r)}$ *(or* $Z_0^{(r)}$*), and* $N_t - N_0$ *is a Poisson process of rate* λ. *Consequently, both* $\{Z_t^{(r)}, t \geq 0\}$ *and* $\{N_t^{(r)}, t \geq 0\}$ *are real-valued homogeneous Markov processes. Specifically, from the preceding proposition we have that the Markov semi-group of the Brownian motion with drift is* $p_t(x + rt, B)$, *where for* $t > 0$,

$$p_t(x,B) = \int_B \frac{e^{-(y-x)^2/2t}}{\sqrt{2\pi t}} dy\,, \tag{8.3.9}$$

having the transition probability kernel $p_t(x,y) = \exp(-(y-x)^2/2t)/\sqrt{2\pi t}$. *Similarly, the Markov semi-group of the Poisson process with drift is* $q_t(x+rt, B)$, *where*

$$q_t(x,B) = e^{-\lambda t} \sum_{k=0}^{\infty} \frac{(\lambda t)^k}{k!} I_B(x+k)\,. \tag{8.3.10}$$

Remark. Homogeneous Markov chains are characterized by their (one-step) transition probabilities, whereas each homogeneous Markov process has a full semi-group $p_t(\cdot)$, $t \geq 0$. While outside our scope, we note in passing that the semi-group relation (8.3.2) can be rearranged as $s^{-1}(p_{s+t} - p_t) = s^{-1}(p_s - p_0)p_t$, which subject to the appropriate regularity conditions should yield for $s \downarrow 0$ the celebrated *backward Kolmogorov equation* $\partial_t p_t = \mathbf{L} p_t$. The operator $\mathbf{L} = \lim_{s\downarrow 0} s^{-1}(p_s - p_0)$ is then called the *generator* of the Markov process (or its semi-group). For example, the transition probability kernel $p_t(x+rt, y)$ of the Brownian motion with drift solves the partial differential equation (PDE), $u_t = \frac{1}{2}u_{xx} + ru_x$ and the generator of this semi-group is $\mathbf{L}u = \frac{1}{2}u_{xx} + ru_x$ (c.f. [**KaS97**, Chapter 5]). For this reason, many computations about Brownian motion can also be done by solving rather simple elliptic or parabolic PDE-s.

We saw in Proposition 8.3.5 that the Wiener process $(W_t, t \geq 0)$ is a homogeneous $\mathcal{F}_t^{\mathbf{W}}$-Markov process of continuous sample functions and the Markov semi-group of (8.3.9). This motivates the following definition of a *Brownian Markov process* $(W_t, \mathcal{F}_t)$, where our accommodation of possible enlargements of the filtration and different initial distributions will be useful in future applications.

Definition 8.3.7 (Brownian Markov process)**.** *We call* $(W_t, \mathcal{F}_t)$ *a* Brownian Markov process *if* $\{W_t, t \geq 0\}$ *of continuous sample functions is a homogeneous* $\mathcal{F}_t$*-Markov process with the Brownian semi-group* $\{p_t, t \geq 0\}$ *of (8.3.9). If in addition* $W_0 = 0$, *we call such process a* standard Brownian Markov process.

Stationarity of Markov processes, in the sense of Definition 7.3.7, is related to the important concept of *invariant probability measures* which we define next (compare with Definition 6.1.20).

Definition 8.3.8. *A probability measure* ν *on a* $\mathcal{B}$*-isomorphic space* $(\mathbb{S}, \mathcal{S})$ *is called an* invariant (probability) measure *for a semi-group of transition probabilities* $\{p_u, u \geq 0\}$, *if the induced law* $\mathbf{P}_\nu(\cdot) = \int_{\mathbb{S}} \mathbf{P}_x(\cdot)\nu(dx)$ *(see Remark 8.3.3), is invariant under any time shift* θ_s, $s \geq 0$.

You can easily check that if Markov process is also a stationary process under an initial probability measure ν, then it is effectively a homogeneous Markov process, in the sense that $p_{s,t}(x,\cdot) = p_{t-s}(x,\cdot)$ for any $t \geq s \geq 0$ and ν-a.e. $x \in \mathbb{S}$. However, many homogeneous Markov processes are non-stationary (for example, recall Examples 7.3.14 and 8.3.6, that the Brownian motion is non-stationary yet homogeneous, Markov process).

Here is the explicit characterization of invariant measures for a given Markov semi-group and their connection to stationary Markov processes.

Exercise 8.3.9. *Adapting the proof of Proposition 6.1.23 show that a probability measure* ν *on* $\mathcal{B}$*-isomorphic* $(\mathbb{S}, \mathcal{S})$ *is an invariant measure for a Markov semi-group* $\{p_u, u \geq 0\}$, *if and only if* $\nu p_t = \nu$ *for any* $t \geq 0$ *(note that a homogeneous Markov*

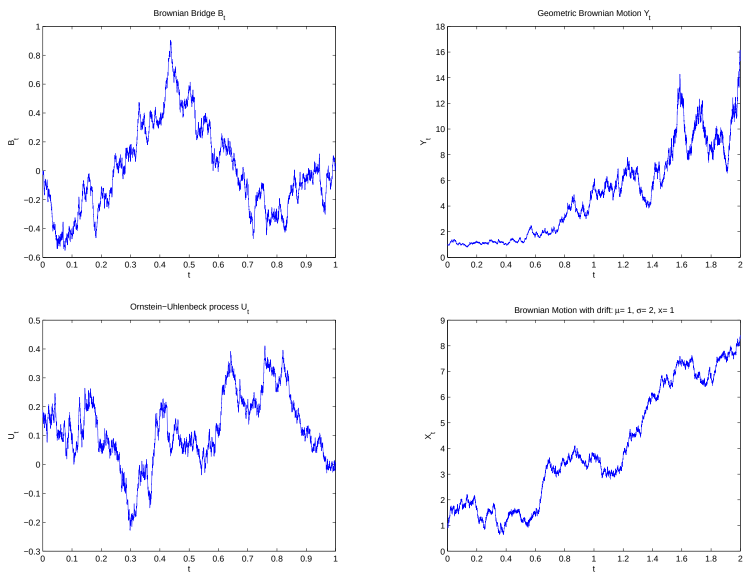

FIGURE 1. Illustration of sample paths for processes in Exercise 8.3.10.

process $\{X_t, t \geq 0\}$ *is a stationary S.P. if and only if the initial distribution* $\nu(B) = \mathbf{P}(X_0 \in B)$ *is an invariant probability measure for the corresponding Markov semi-group).*

Pursuing similar themes, your next exercise examines some of the most fundamental S.P. one derives out of the Brownian motion.

Exercise 8.3.10. *With* $\{W_t, t \geq 0\}$ *a Wiener process, consider the* Geometric Brownian motion $Y_t = e^{W_t}$, *Ornstein-Uhlenbeck process* $U_t = e^{-t/2} W_{e^t}$, *Brownian motion with drift* $Z_t^{(r,\sigma)} = \sigma W_t + rt$ *and the standard Brownian bridge on* $[0, 1]$ *(as in Exercises 7.3.15-7.3.16).*

(a) *Determine which of these four S.P. is a Markov process with respect to* its *canonical filtration, and among those, which are also homogeneous.*
(b) *Find among these S.P. a homogeneous Markov process whose increments are neither independent nor stationary.*
(c) *Find among these S.P. a Markov process of stationary increments, which is not a homogeneous Markov process.*

Homogeneous Markov processes possess the following *Markov property*, extending the invariance (8.3.4) of the process under the time shifts θ_s of Definition 7.3.7 to any bounded $\mathcal{S}^{[0,\infty)}$-measurable function of its sample path (compare to (6.1.8) in case of homogeneous Markov chains).

Proposition 8.3.11 (MARKOV PROPERTY)**.** *Suppose* $(X_t, t \geq 0)$ *is a homogeneous* $\mathcal{F}_t$*-Markov process on a* $\mathcal{B}$*-isomorphic state space* $(\mathbb{S}, \mathcal{S})$ *and let* $\mathbf{P}_x$ *denote the corresponding family of laws associated with its semi-group. Then,* $x \mapsto \mathbf{E}_x[h]$ *is measurable on* $(\mathbb{S}, \mathcal{S})$ *for any* $h \in b\mathcal{S}^{[0,\infty)}$*, and further for any* $s \geq 0$*, almost surely*

$$\mathbf{E}[h \circ \theta_s(X_\cdot(\omega)) | \mathcal{F}_s] = \mathbf{E}_{X_s}[h]\,. \tag{8.3.11}$$

Remark 8.3.12. From Lemma 7.1.7 you can easily deduce that any $V \in b\mathcal{F}^{\mathbf{X}}$ is of the form $V = h(X.)$ with $h \in b\mathcal{S}^{[0,\infty)}$. Further, in view of Exercises 1.2.32 and 7.2.9, any bounded Borel function $h(\cdot)$ on the space $C([0,\infty))$ of continuous functions equipped with the topology of uniform convergence on compact intervals is the restriction to $C([0,\infty))$ of some $\widetilde{h} \in b\mathbb{R}^{[0,\infty)}$. In particular, for a real-valued Markov process $\{X_t, t \geq 0\}$ of continuous sample functions, $\mathbf{E}_x[\widetilde{h}] = \mathbf{E}_x[h]$ and $h \circ \theta_s(X.) = \widetilde{h} \circ \theta_s(X.)$, hence (8.3.11) applies for any bounded, $\mathcal{B}_{C([0,\infty))}$-measurable function h.

PROOF. Fixing $s \geq 0$, in case $h(x(\cdot)) = \prod_{\ell=0}^{n} f_\ell(x(u_\ell))$ for finite n, $f_\ell \in b\mathcal{S}$ and $u_0 > \cdots > u_n \geq 0$, we have by (8.3.8) for $t_\ell = s + u_\ell$ and the semi-group $p_{r,t} = p_{t-r}$ of $(X_t, t \geq 0)$, that

$$\mathbf{E}[\prod_{\ell=0}^{n} f_\ell(X_{t_\ell}) | \mathcal{F}_s] = p_{u_n}(f_n p_{u_{n-1}-u_n}(\cdots(f_1 p_{u_0-u_1} f_0)))(X_s) = \mathbf{E}_{X_s}[\prod_{\ell=0}^{n} f_\ell(X_{u_\ell})] \,.$$

The measurability of $x \mapsto \mathbf{E}_x[h]$ for such functionals $h(\cdot)$ is verified by induction on n, where if $n = 0$ then for $f_0 \in b\mathcal{S}$ by Lemma 6.1.3 also $\mathbf{E}_x h = g_1(x) = p_{u_0} f_0(x)$ is in $b\mathcal{S}$ and by the same argument, in the induction step $g_{n+1}(x) = p_{u_n}(f_n g_n)(x)$ are also in $b\mathcal{S}$.

To complete the proof consider the collection $\mathcal{H}$ of functionals $h \in b\mathcal{S}^{[0,\infty)}$ such that $x \mapsto \mathbf{E}_x[h]$ is $\mathcal{S}$-measurable and (8.3.11) holds. The linearity of the (conditional) expectation and the monotone convergence theorem result with $\mathcal{H}$ a vector-space that is closed under monotone limits, respectively. Further, as already shown, $\mathcal{H}$ contains the indicators $h(\cdot) = I_A(\cdot)$ with $A = \{x(\cdot) : x(u_\ell) \in B_\ell \in \mathcal{S}, \ell = 0, \ldots, n\}$ a finite dimensional measurable rectangle. Thus, $\mathcal{H}$ satisfies the conditions of the monotone class theorem. Consequently $\mathcal{H} = b\mathcal{S}^{[0,\infty)}$, that is, for each $h \in \mathcal{S}^{[0,\infty)}$ both $x \mapsto \mathbf{E}_x[h] \in b\mathcal{S}$ and (8.3.11) holds w.p.1. $\square$

8.3.2. Strong Markov processes and Feller semi-groups. Given a homogeneous $\mathcal{F}_t$-Markov process $(X_t, t \geq 0)$, we seek to strengthen its Markov property about the shift of the sample path by non-random $s \geq 0$ (see Proposition 8.3.11), to the *strong Markov property*, whereby shifting by any $\mathcal{F}_t$-Markov time τ is accommodated (see Proposition 6.1.16 about Markov chains having this property).

Definition 8.3.13 (STRONG MARKOV PROCESS)**.** *We say that an $\mathcal{F}_t$-progressively measurable, homogeneous Markov process $\{X_t, t \geq 0\}$ on $\mathcal{B}$-isomorphic state space $(\mathbb{S}, \mathcal{S})$, has the* strong Markov property *(or that $(X_t, \mathcal{F}_t)$ is a* strong Markov process*), if for any bounded $h(s, x(\cdot))$ measurable on the product σ-algebra $\mathcal{U} = \mathcal{B}_{[0,\infty)} \times \mathcal{S}^{[0,\infty)}$, and any $\mathcal{F}_t$-Markov time τ, almost surely*

$$I_{\{\tau<\infty\}} \mathbf{E}[h(\tau, X_{\tau+\cdot}(\omega)) | \mathcal{F}_{\tau^+}] = g_h(\tau, X_\tau) I_{\{\tau<\infty\}} \,, \tag{8.3.12}$$

where $g_h(s, x) = \mathbf{E}_x[h(s, \cdot)]$ is bounded and measurable on $\mathcal{B}_{[0,\infty)} \times \mathcal{S}$, $x \mapsto \mathbf{P}_x$ are the laws associated with the semi-group of $(X_t, \mathcal{F}_t)$, $\mathcal{F}_{\tau^+}$ is the Markov σ-algebra associated with τ (c.f. Definition 8.1.9), and both sides of (8.3.12) are set to zero when $\tau(\omega) = \infty$.

As noted in Remark 8.3.12, every $V \in b\mathcal{F}^{\mathbf{X}}$ is of the form $V = h(X.)$, with $h(\cdot)$ in the scope of the strong Markov property, which for a real-valued homogeneous Markov process $\{X_t, t \geq 0\}$ of continuous sample functions, contains all bounded

Borel functions $h(\cdot,\cdot)$ on $[0,\infty)\times C([0,\infty))$. In applications it is often handy to further have a time varying functional $h(s,x(\cdot))$ (for example, see our proof of the *reflection principle*, in Proposition 9.1.10).

Remark. Recall that the Markov time τ is an $\mathcal{F}_{t^+}$-stopping time (see Definition 8.1.9), hence the assumed $\mathcal{F}_t$-progressive measurability of $\{X_t\}$ guarantees that on the event $\{\tau<\infty\}$ the R.V. τ and X_τ are measurable on $\mathcal{F}_{\tau^+}$ (see Proposition 8.1.13), hence by our definition so is $g_h(\tau,X_\tau)$. While we have stated and proved Proposition 8.1.13 only in case of real-valued S.P. $\{X_t,t\ge 0\}$, the same proof (and conclusion), applies for any state space $(\mathbb{S},\mathcal{S})$. We lose nothing by assuming progressive measurability of $\{X_t\}$ since for a right-continuous process this is equivalent to its adaptedness (see Proposition 8.1.8, whose proof and conclusion extend to any topological state space).

Here is an immediate consequence of Definition 8.3.13.

Corollary 8.3.14. *If $(X_t,\mathcal{F}_t)$ is a strong Markov process and τ is an $\mathcal{F}_t$-stopping time, then for any $h\in b\mathcal{U}$, almost surely*

$$I_{\{\tau<\infty\}}\mathbf{E}[h(\tau,X_{\tau+\cdot}(\omega))|\mathcal{F}_\tau]=g_h(\tau,X_\tau)I_{\{\tau<\infty\}}\,. \tag{8.3.13}$$

In particular, if $(X_t,\mathcal{F}_t)$ is a strong Markov process, then $\{X_t,t\ge0\}$ is a homogeneous $\mathcal{F}_{t^+}$-Markov process and for any $s\ge0$ and $h\in b\mathcal{S}^{[0,\infty)}$, almost surely

$$\mathbf{E}[h(X_\cdot)|\mathcal{F}_{s^+}]=\mathbf{E}[h(X_\cdot)|\mathcal{F}_s]\,. \tag{8.3.14}$$

Proof. By the preceding remark, having an $\mathcal{F}_t$-stopping time τ results with $g_h(\tau,X_\tau)I_{\{\tau<\infty\}}$ which is measurable on $\mathcal{F}_\tau$. Thus, applying the tower property for the expectation of (8.3.12) conditional on $\mathcal{F}_\tau\subseteq\mathcal{F}_{\tau^+}$, results with (8.3.13). Comparing (8.3.13) and (8.3.12) for constant in time $h(x(\cdot))$ and the non-random, finite stopping time $\tau=s$ we deduce that (8.3.14) holds whenever $h=h_0\circ\theta_s$ for some $h_0\in b\mathcal{S}^{[0,\infty)}$. Since $(X_t,\mathcal{F}_t)$ is a homogeneous Markov process, considering $h_0(x(\cdot))=I_B(x(u))$ for $u\ge0$ and $B\in\mathcal{S}$, it follows that (8.3.4) holds also for $(X_t,\mathcal{F}_{t^+})$, namely, that $\{X_t,t\ge0\}$ is a homogeneous $\mathcal{F}_{t^+}$-Markov process. With $\mathcal{H}$ denoting the collection of functionals $h\in b\mathcal{S}^{[0,\infty)}$ for which (8.3.14) holds, by the monotone class theorem it suffices to check that this is the case when $h(x(\cdot))=\prod_{m=1}^k I_{B_m}(x(u_m))$, with k finite, $u_m\ge0$ and $B_m\in\mathcal{S}$. Representing such functionals as $h(\cdot)=h_1(\cdot)h_0\circ\theta_s(\cdot)$ with $h_0(x(\cdot))=\prod_{u_m\ge s}I_{B_m}(x(u_m-s))$ and $h_1(x(\cdot))=\prod_{u_m<s}I_{B_m}(x(u_m))$, we complete the proof by noting that $h_1(X_\cdot)$ is measurable with respect to $\mathcal{F}_s\subseteq\mathcal{F}_{s^+}$, so can be taken out of both conditional expectations in (8.3.14) and thus eliminated. □

To make the most use of the strong Markov property, Definition 8.3.13 calls for an arbitrary $h\in b\mathcal{U}$. As we show next, for checking that a specific S.P. is a strong Markov process, it suffices to verify (8.3.12) only for $h(s,x(\cdot))=I_B(x(u))$ and bounded Markov times (compare with the definition of a homogeneous Markov process via (8.3.4)), which is way more manageable task.

Proposition 8.3.15. *An $\mathcal{F}_t$-progressively measurable, homogeneous Markov process $\{X_t,t\ge0\}$ with a semi-group $\{p_u,u\ge0\}$ on $\mathcal{B}$-isomorphic state space $(\mathbb{S},\mathcal{S})$, has the strong Markov property if for any $u\ge0$, $B\in\mathcal{S}$ and* bounded *$\mathcal{F}_t$-Markov times τ, almost surely*

$$\mathbf{P}[X_{\tau+u}\in B|\mathcal{F}_{\tau^+}]=p_u(X_\tau,B)\,. \tag{8.3.15}$$

PROOF. *Step 1.* We start by extending the validity of (8.3.15) to any a.s. finite Markov time. To this end, fixing $u \geq 0$, $B \in \mathcal{S}$, $n \in \mathbb{Z}_+$ and a $[0,\infty]$-valued $\mathcal{F}_t$-Markov time τ, recall that $\tau_n = \tau \wedge n$ is a bounded $\mathcal{F}_{t^+}$-stopping time (c.f. part (c) of Exercise 8.1.10). Further, the bounded $I_{\{\tau \leq n\}}$ and $p_u(X_{\tau_n}, B)$ are both measurable on $\mathcal{F}_{\tau_n^+}$ (see part (a) of Exercise 8.1.11 and Proposition 8.1.13, respectively). Hence, multiplying the identity (8.3.15) in case of τ_n by $I_{\{\tau \leq n\}}$, and taking in, then out, what is known, we find that a.s.

$$0 = \mathbf{E}[I_{\{\tau \leq n\}}(I_B(X_{\tau_n+u}) - p_u(X_{\tau_n}, B))|\mathcal{F}_{\tau_n^+}] = I_{\{\tau \leq n\}}\mathbf{E}[Z|\mathcal{F}_{\tau_n^+}]\,,$$

for the bounded R.V.

$$Z = I_{\{\tau < \infty\}}[I_B(X_{\tau+u}) - p_u(X_\tau, B)]\,.$$

By part (c) of Exercise 8.1.11 it then follows that w.p.1. $I_{\{\tau \leq n\}}\mathbf{E}[Z|\mathcal{F}_{\tau^+}] = 0$. Taking $n \uparrow \infty$ we deduce that a.s. $\mathbf{E}[Z|\mathcal{F}_{\tau^+}] = 0$. Upon taking out the known $I_{\{\tau<\infty\}}p_u(X_\tau, B)$ we represent this as

$$\mathbf{E}[I_{\{\tau<\infty\}}f(X_{\tau+u})|\mathcal{F}_{\tau^+}] = I_{\{\tau<\infty\}}(p_u f)(X_\tau)\,, \quad \text{almost surely} \tag{8.3.16}$$

for $f(\cdot) = I_B(\cdot)$. By linearity of the expectation and conditional expectation, this identity extends from indicators to all $\mathcal{S}$-measurable simple functions, whereby it follows by monotone convergence that it holds for all $f \in b\mathcal{S}$.
Step 2. We are ready to prove that (8.3.12) holds for any $\mathcal{F}_t$-Markov time τ, in case $h(s, x(\cdot)) = f_0(s)\prod_{\ell=1}^n f_\ell(x(u_\ell))$, with bounded Borel $f_0 : [0,\infty) \to \mathbb{R}$, $f_\ell \in b\mathcal{S}$ and $u_1 > \cdots > u_n \geq 0 = u_{n+1}$. As $f_0(\tau) \in b\mathcal{F}_{\tau^+}$, one can always take this (known) part of $h(\cdot,\cdot)$ out of the conditional expectation in (8.3.12) and thereafter eliminate it. Thus, setting $f_0 = 1$ we proceed to prove by induction on n that (8.3.12) holds, namely, that for *any* $\mathcal{F}_t$-Markov time τ, $f_\ell \in b\mathcal{S}$ and $u_1 > \cdots > u_n \geq 0$, almost surely,

$$\mathbf{E}[I_{\{\tau<\infty\}}\prod_{\ell=1}^n f_\ell(X_{\tau+u_\ell})|\mathcal{F}_{\tau^+}] = I_{\{\tau<\infty\}}g_n(X_\tau)\,,$$

for the bounded, $\mathcal{S}$-measurable functions $g_1 = p_{u_1-u_2}f_1$ and

$$g_\ell = p_{u_\ell - u_{\ell+1}}(f_\ell g_{\ell-1})\,, \quad \ell = 2, \ldots, n.$$

The identity (8.3.16) is the $n = 1$ basis of the proof. To carry out the induction step, recall part (c) of Exercise 8.1.10 that $\tau_\ell = \tau + u_\ell \geq \tau$ is a decreasing sequence of $\mathcal{F}_t$-Markov times, which are finite if and only if τ is, and further, $\mathcal{F}_{\tau^+} \subseteq \mathcal{F}_{\tau_n^+}$ (see part (b) of Exercise 8.1.11). It thus follows by the tower property and taking out the known term $f_n(X_{\tau_n^+}) \in b\mathcal{F}_{\tau_n^+}$ (when $\tau < \infty$, see Proposition 8.1.13), that

$$\begin{aligned}\mathbf{E}[I_{\{\tau<\infty\}}\prod_{\ell=1}^n f_\ell(X_{\tau_\ell})|\mathcal{F}_{\tau^+}] &= \mathbf{E}[f_n(X_{\tau_n})\mathbf{E}[I_{\{\tau_n<\infty\}}\prod_{\ell=1}^{n-1} f_\ell(X_{\tau_\ell})|\mathcal{F}_{\tau_n^+}]|\mathcal{F}_{\tau^+}] \\ &= \mathbf{E}[I_{\{\tau<\infty\}}f_n(X_{\tau_n})g_{n-1}(X_{\tau_n})|\mathcal{F}_\tau^+] = I_{\{\tau<\infty\}}g_n(X_\tau)\,.\end{aligned}$$

Indeed, since $\tau_\ell - \tau_n = u_\ell - u_n$ are non-random and positive, the induction hypothesis applies for the $\mathcal{F}_t$-Markov time τ_n to yield the second equality, whereas the third equality is established by considering the identity (8.3.16) for $f = f_n g_{n-1}$ and $u = u_n$.
Step 3. Similarly to the proof of Proposition 8.3.11, fixing $A \in \mathcal{F}_{\tau^+}$, yet another application of the monotone class theorem shows that any $h \in b\mathcal{U}$ is in the collection

$\mathcal{H} \subseteq b\mathcal{U}$ for which $g_h(s,x) = \mathbf{E}_x[h(s,\cdot)]$ is measurable on $\mathcal{B}_{[0,\infty)} \times \mathcal{S}$ and

$$\mathbf{E}[I_{\{\tau<\infty\}} I_A h(\tau, X_{\tau+\cdot})] = \mathbf{E}[g_h(\tau, X_\tau) I_{\{\tau<\infty\}} I_A]\,. \tag{8.3.17}$$

Indeed, in Step 2 we have shown that $\mathcal{H}$ contains the indicators on the π-system

$$\mathcal{P} = \{B \times D : B \in \mathcal{B}_{[0,\infty)}, D \in \mathcal{S}^{[0,\infty)} \text{ a finite dimensional measurable rectangle}\}\,,$$

such that $\mathcal{U} = \sigma(\mathcal{P})$. Further, constants are in $\mathcal{H}$ which by the linearity of the expectation (and hence of $h \mapsto g_h$), is a vector space. Finally, if $h_n \uparrow h$ bounded and $h_n \in \mathcal{H}$ are non-negative, then $h \in b\mathcal{U}$ and by monotone convergence $g_{h_n} \uparrow g_h$ bounded and measurable, with the pair (h, g_h) also satisfying (8.3.17). Since $g_h(\tau, X_\tau) I_{\{\tau<\infty\}}$ is in $b\mathcal{F}_{\tau_+}$ and the preceding argument applies for all $A \in \mathcal{F}_{\tau^+}$, we conclude that per τ and h the identity (8.3.12) holds w.p.1., as claimed. $\square$

As you are to show now, both Markov and strong Markov properties apply for product laws of finitely many independent processes, each of which has the corresponding property.

Exercise 8.3.16. *Suppose on some probability space $(\Omega, \mathcal{F}, \mathbf{P})$ we have homogeneous Markov processes $(X_t^{(i)}, \mathcal{F}_t^{(i)})$ of $\mathcal{B}$-isomorphic state spaces $(\mathbb{S}_i, \mathcal{S}_i)$ and Markov semi-groups $p_t^{(i)}(\cdot,\cdot)$, such that $\mathcal{F}_\infty^{(i)}$, $i = 1, \ldots, \ell$ are $\mathbf{P}$-mutually independent.*

(a) *Let $\underline{X}_t = (X_t^{(1)}, \ldots, X_t^{(\ell)})$ and $\mathcal{F}_t = \sigma(\mathcal{F}_t^{(1)}, \ldots, \mathcal{F}_t^{(\ell)})$. Show that $(\underline{X}_t, \mathcal{F}_t)$ is a homogeneous Markov process, of the Markov semi-group*

$$p_t(\underline{x}, B_1 \times \cdots \times B_\ell) = \prod_{i=1}^{\ell} p_t^{(i)}(x_i, B_i)$$

on the $\mathcal{B}$-isomorphic state space $(\mathbb{S}, \mathcal{S})$, where $\mathbb{S} = \mathbb{S}_1 \times \cdots \times \mathbb{S}_\ell$ and $\mathcal{S} = \mathcal{S}_1 \times \cdots \times \mathcal{S}_\ell$.

(b) *Suppose in addition that for each $i = 1, \ldots, k$ the $\mathcal{F}_t^{(i)}$-progressively measurable process $X_t^{(i)}$ has the strong Markov property. Show that the strong Markov property then holds for the $\mathcal{F}_t$-progressively measurable $(\mathbb{S}, \mathcal{S})$-valued stochastic process $\underline{X}_t$.*

Recall Proposition 6.1.16 that *every* homogeneous Markov chain of a $\mathcal{B}$-isomorphic state space has the strong Markov property and that in this context every Markov time is a stopping time and takes only countably many possible values. As expected, you are to show next that any homogeneous Markov process has the strong Markov property (8.3.13) for such stopping times.

Exercise 8.3.17. *Suppose $(X_t, \mathcal{F}_t)$ is a homogeneous Markov process, $(\mathbb{S}, \mathcal{S})$ its $\mathcal{B}$-isomorphic state space and $\tau : \Omega \mapsto \mathbb{C}$ is an $\mathcal{F}_t$-stopping time with countable $\mathbb{C} = \{s_k\} \subset [0, \infty]$.*

(a) *Show that $A \cap \{\omega : \tau(\omega) = s_k\} \in \mathcal{F}_{s_k}$ for any finite $s_k \in \mathbb{C}$ and $A \in \mathcal{F}_\tau$.*

(b) *Deduce that $h(\tau, X_{\tau+\cdot}) I_{\{\tau<\infty\}}$ is a R.V. and $g_h(\tau, X_\tau) I_{\{\tau<\infty\}} \in b\mathcal{F}_\tau$ provided $h(s_k, \cdot)$ are $\mathcal{S}^{[0,\infty)}$-measurable and uniformly bounded on $\mathbb{C} \times \mathbb{S}^{[0,\infty)}$.*

(c) *Conclude that (8.3.13) holds a.s. for any such τ and h.*

For the Feller semi-groups we define next (compare with the strong Feller property of Remark 6.3.12), the right-continuity of sample functions yields the strong Markov property.

Definition 8.3.18. *A* Feller semi-group *is a Markov semi-group $\{p_u, u \geq 0\}$ on $(\mathbb{R}, \mathcal{B})$ such that $p_t : C_b(\mathbb{R}) \mapsto C_b(\mathbb{R})$ for any $t \geq 0$. That is, $x \mapsto (p_t f)(x)$ is continuous for any fixed bounded, continuous function f and $t \geq 0$.*

Proposition 8.3.19. *Any right-continuous homogeneous Markov process $(X_t, \mathcal{F}_t)$ with a Feller semi-group (of transition probabilities), is a strong Markov process.*

PROOF. Fixing $u \geq 0$, a bounded $\mathcal{F}_t$-Markov time τ, $A \in \mathcal{F}_{\tau^+}$ and $f \in C_b(\mathbb{R})$, we proceed to show that

$$\mathbf{E}[f(X_{\tau+u}) I_A] = \mathbf{E}[(p_u f)(X_\tau) I_A] \,. \tag{8.3.18}$$

Indeed, recall that in Lemma 8.1.16 we have constructed a sequence of finite $\mathcal{F}_t$-stopping times $\tau_\ell = 2^{-\ell}([2^\ell \tau]+1)$ taking values in the countable set of non-negative dyadic rationals, such that $\tau_\ell \downarrow \tau$. Further, for any ℓ we have that $A \in \mathcal{F}_{\tau_\ell}$ (see part (b) of Exercise 8.1.12), hence as shown in Exercise 8.3.17,

$$\mathbf{E}[f(X_{\tau_\ell+u}) I_A] = \mathbf{E}[(p_u f)(X_{\tau_\ell}) I_A] \,.$$

Due to the sample path right-continuity, both $X_{\tau_\ell+u} \to X_{\tau+u}$ and $X_{\tau_\ell} \to X_\tau$. Since $f \in C_b(\mathbb{R})$ and $p_u f \in C_b(\mathbb{R})$ (by the assumed Feller property), as $\ell \to \infty$ both $f(X_{\tau_\ell+u}) \to f(X_{\tau+u})$ and $(p_u f)(X_{\tau_\ell}) \to (p_u f)(X_\tau)$. We thus deduce by bounded convergence that (8.3.18) holds.

Next, consider non-negative $f_k \in C_b(\mathbb{R})$ such that $f_k \uparrow I_{(-\infty,b)}$ (see Lemma 3.1.6 for an explicit construction of such). By monotone convergence $p_u f_k \uparrow p_u I_{(-\infty,b)}$ and hence

$$\mathbf{E}[I_B(X_{\tau+u}) I_A] = \mathbf{E}[p_u(X_\tau, B) I_A] \,, \tag{8.3.19}$$

for any B in the π-system $\{(-\infty, b) : b \in \mathbb{R}\}$ which generates the Borel σ-algebra $\mathcal{B}$. The collection of $\mathcal{L}$ of sets $B \in \mathcal{B}$ for which the preceding identity holds is a λ-system (by linearity of the expectation and monotone convergence), so by Dynkin's $\pi - \lambda$ theorem it holds for any Borel set B. Since this applies for any $A \in \mathcal{F}_{\tau^+}$, the strong Markov property of $(X_t, \mathcal{F}_t)$ follows from Proposition 8.3.15, upon noting that the right-continuity of $t \mapsto X_t$ implies that X_t is $\mathcal{F}_t$-progressively measurable, with $p_u(X_\tau, B) \in m\mathcal{F}_{\tau^+}$ (see Propositions 8.1.8 and 8.1.13, respectively). $\square$

Taking advantage of the preceding result, you can now verify that any right-continuous S.P. of stationary, independent increments is a strong Markov process.

Exercise 8.3.20. *Suppose $\{X_t, t \geq 0\}$ is a real-valued process of stationary, independent increments.*

(a) *Show that $\{X_t, t \geq 0\}$ has a Feller semi-group.*
(b) *Show that if $\{X_t, t \geq 0\}$ is also right-continuous, then it is a strong Markov process. Deduce that this applies in particular for the Poisson process (starting at $N_0 = x \in \mathbb{R}$ as in Example 8.3.6), as well as for any Brownian Markov process $(X_t, \mathcal{F}_t)$.*
(c) *Suppose the right-continuous $\{X_t, t \geq 0\}$ is such that $\lim_{t \downarrow 0} \mathbf{E}|X_t| = 0$ and $X_0 = 0$. Show that X_t is integrable for all $t \geq 0$ and $M_t = X_t - t\mathbf{E}X_1$ is a martingale. Deduce that then $\mathbf{E}[X_\tau] = \mathbf{E}[\tau]\mathbf{E}[X_1]$ for any integrable $\mathcal{F}_t^{\mathbf{X}}$-stopping time τ.*
Hint: *Establish the last claim first for the $\mathcal{F}_t^{\mathbf{X}}$-stopping times $\tau_\ell = 2^{-\ell}([2^\ell \tau]+ 1)$.*

Our next example demonstrates that some regularity of the semi-group is needed when aiming at the strong Markov property (i.e., merely considering the canonical filtration of a homogeneous Markov process with continuous sample functions is not enough).

Example 8.3.21. *Suppose X_0 is independent of the standard Wiener process $\{W_t, t \geq 0\}$ and $q = \mathbf{P}(X_0 = 0) \in (0,1)$. The S.P. $X_t = X_0 + W_t I_{\{X_0 \neq 0\}}$ has continuous sample functions and for any fixed $s \geq 0$, a.s. $I_{\{X_0=0\}} = I_{\{X_s=0\}}$ (as the difference occurs on the event $\{W_s = -X_0 \neq 0\}$ which is of zero probability). Further, the independence of increments of $\{W_t\}$ implies the same for $\{X_t\}$* conditioned on X_0, *hence for any $u \geq 0$ and Borel set B, almost surely,*

$$\begin{aligned}\mathbf{P}(X_{s+u} \in B|\mathcal{F}_s^{\mathbf{X}}) &= I_{0\in B} I_{\{X_0=0\}} + \mathbf{P}(W_{s+u} - W_s + X_s \in B|X_s) I_{\{X_0\neq 0\}}\\ &= \widehat{p}_u(X_s, B)\,,\end{aligned}$$

where $\widehat{p}_u(x,B) = p_0(x,B)I_{x=0} + p_u(x,B)I_{x\neq 0}$ for the Brownian semi-group $p_u(\cdot)$. Clearly, per u fixed, $\widehat{p}_u(\cdot,\cdot)$ is a transition probability on $(\mathbb{R},\mathcal{B})$ and $\widehat{p}_0(x,B)$ is the identity element for the semi-group relation $\widehat{p}_{u+s} = \widehat{p}_u\widehat{p}_s$ which is easily shown to hold (but this is not a Feller semi-group, since $x \mapsto (\widehat{p}_t f)(x)$ is discontinuous at $x = 0$ whenever $f(0) \neq \mathbf{E}f(W_t)$). In view of Definition 8.3.1, we have just shown that $\widehat{p}_u(\cdot,\cdot)$ is the Markov semi-group associated with the $\mathcal{F}_t^{\mathbf{X}}$-progressively measurable homogeneous Markov process $\{X_t, t \geq 0\}$ (regardless of the distribution of X_0). However, $(X_t, \mathcal{F}_t^{\mathbf{X}})$ is not a strong Markov process. *Indeed, note that $\tau = \inf\{t \geq 0 : X_t = 0\}$ is an $\mathcal{F}_t^{\mathbf{X}}$-stopping time (see Proposition 8.1.15), which is finite a.s. (since if $X_0 \neq 0$ then $X_t = W_t + X_0$ and $\tau = \tau^{(0)}_{-X_0}$ of Exercise 8.2.35, whereas for $X_0 = 0$ obviously $\tau = 0$). Further, by continuity of the sample functions, $X_\tau = 0$ whenever $\tau < \infty$, so if $(X_t, \mathcal{F}_t^{\mathbf{X}})$ was a strong Markov process, then in particular, a.s.*

$$\mathbf{P}(X_{\tau+1} > 0|\mathcal{F}_\tau^{\mathbf{X}}) = \widehat{p}_1(0,(0,\infty)) = 0$$

(this is merely (8.3.15) for the stopping time τ, $u = 1$ and $B = (0,\infty)$). However, the latter identity fails whenever $X_0 \neq 0$ (i.e. with probability $1 - q > 0$), for then the left side is merely $p_1(0,(0,\infty)) = 1/2$ (since $\{W_t, \mathcal{F}_t^{\mathbf{X}}\}$ is a Brownian Markov process, hence a strong Markov process, see Exercise 8.3.20).

Here is an alternative, martingale based, proof that any Brownian Markov process is a strong Markov process.

Exercise 8.3.22. *Suppose $(X_t, \mathcal{F}_t)$ is a Brownian Markov process.*

(a) *Let R_t and I_t denote the real and imaginary parts of the complex-valued S.P. $M_t = \exp(i\theta X_t + t\theta^2/2)$. Show that both $(R_t, \mathcal{F}_t)$ and $(I_t, \mathcal{F}_t)$ are MG-s.*
(b) *Fixing a bounded $\mathcal{F}_t$-Markov time τ, show that $\mathbf{E}[M_{\tau+u}|\mathcal{F}_{\tau^+}] = M_\tau$ w.p.1.*
(c) *Deduce that w.p.1. the R.C.P.D. of $X_{\tau+u}$ given $\mathcal{F}_{\tau^+}$ matches the normal distribution of mean $X_\tau(\omega)$ and variance u.*
(d) *Conclude that the $\mathcal{F}_t$-progressively measurable homogeneous Markov process $\{X_t, t \geq 0\}$ is a strong Markov process.*

8.3.3. Markov jump processes. This section is about the following Markov processes which in many respects are very close to Markov chains.

Definition 8.3.23. *A function* $x : \mathbb{R}_+ \mapsto \mathbb{S}$ *is called a* step function *if it is constant on each of the intervals* $[s_{k-1}, s_k)$*, for some countable (possibly finite), set of* isolated *points* $0 = s_0 < s_1 < s_2 < \cdots$. *A continuous-time stochastic process* $(X_t, t \geq 0)$ *taking values in some measurable space* $(\mathbb{S}, \mathcal{S})$ *is called a* pure jump process *if its sample functions are step functions. A* Markov pure jump process *is a homogeneous Markov process which, starting at any non-random* $X_0 = x \in \mathbb{S}$*, is also a pure jump process on its* $\mathcal{B}$*-isomorphic state space* $(\mathbb{S}, \mathcal{S})$.

Remark. We often use *Markov jump process* for Markov pure jump process and note in passing that these processes are sometimes also called *continuous time Markov chains.*

The relatively explicit analysis of Markov jump processes, as provided here, owes much to the fact that the jump times in their sample functions are *isolated*. Many interesting, and harder to analyze Markov processes have piecewise constant sample functions, but with accumulation points of jump times.

We start by showing that the strong Markov property applies for *all* Markov jump processes.

Proposition 8.3.24. *Any Markov jump process* $(X_t, \mathcal{F}_t)$ *is a strong Markov process.*

Proof. Though we did not even endow the state space $(\mathbb{S}, \mathcal{S})$ with a topology, the sample functions $t \mapsto X_t$, being step functions, are trivially right continuous, hence the Markov jump process is $\mathcal{F}_t$-progressively measurable (see Proposition 8.1.8). Fixing $u \geq 0$, a bounded $\mathcal{F}_t$-Markov time τ, $A \in \mathcal{F}_{\tau^+}$ and $B \in \mathcal{S}$, as in the proof of Proposition 8.3.19 the identity (8.3.19) holds for some sequence $\{\tau_\ell\}$ of $\mathcal{F}_t$-stopping times such that $\tau_\ell \downarrow \tau$. Since the right-continuous sample functions $t \mapsto X_t$ of a jump process are constant except possibly for isolated jump times, both $X_\tau = X_{\tau_\ell}$ and $X_{\tau+u} = X_{\tau_\ell+u}$ for all ℓ large enough. Consequently, $I_B(X_{\tau+u}) = I_B(X_{\tau_\ell+u})$ and $p_u(X_\tau, B) = p_u(X_{\tau_\ell}, B)$ for all ℓ large enough, so by bounded convergence the identity (8.3.19) also holds for the $\mathcal{F}_t$-Markov time τ. Since this applies for any $A \in \mathcal{F}_{\tau^+}$, as explained while proving Proposition 8.3.19, the strong Markov property of $(X_t, \mathcal{F}_t)$ then follows from Proposition 8.3.15. $\square$

Example 8.3.25. *The semi-group of a Markov jump process is often not a Feller semi-group (so Proposition 8.3.24 is not a special case of Proposition 8.3.19). For example, setting* $\text{sgn}(0) = 0$ *it is easy to check that* $p_t(x, A) = e^{-t} I_{\{x \in A\}} + (1 - e^{-t}) I_{\{\text{sgn}(x) \in A\}}$ *is a Markov semi-group on* $\mathbb{R}$*, which is not a Feller semi-group (as* $(p_1 h)(x) = e^{-1} h(x) + (1 - e^{-1}) \mathbf{1}_{x \neq 0}$ *is discontinuous for* $h(x) = x^2 \wedge 1 \in C_b(\mathbb{R})$*). This semi-group corresponds to a Markov jump process* $\{X_t\}$ *with at most one jump per sample function, such that starting at any state* X_0 *other than the (absorbing) states* -1, 0 *and* 1*, it jumps to* $\text{sgn}(X_0) \in \{-1, 0, 1\}$ *at a random time* τ *having the exponential distribution of parameter one.*

In view of Lemma 7.1.7, the law of a homogeneous Markov process does not tell us directly whether or not it is a Markov jump process. In fact, a Markov jump process corresponds to the piecewise constant RCLL modification of the given Markov law (and such modification is essentially unique, see Exercise 7.2.3), so one of the central issues here is to determine when such a modification exists.

With the Poisson process as our prototypical example of a Markov jump process, we borrow from the treatment of the Poisson process (in Subsection 3.4.2), and proceed to describe the jump parameters of Markov jump processes. These parameters then serve as a convenient alternative to the general characterization of a homogeneous Markov process via its (Markov) semi-group.

Proposition 8.3.26. *Suppose $(X_t, t \geq 0)$ is a right-continuous, homogeneous Markov process.*

(a) *Under $\mathbf{P}_y$, the $\mathcal{F}_t^{\mathbf{X}}$-Markov time $\tau = \inf\{t \geq 0 : X_t \neq X_0\}$ has the exponential distribution of parameter λ_y, for all $y \in \mathbb{S}$ and some measurable $\lambda : \mathbb{S} \mapsto [0, \infty]$.*

(b) *If $\lambda_y > 0$ then τ is $\mathbf{P}_y$-almost-surely finite and $\mathbf{P}_y$-independent of the $\mathbb{S}$-valued random variable X_τ.*

(c) *If $(X_t, t \geq 0)$ is a strong Markov process and $\lambda_y > 0$ is finite, then $\mathbf{P}_y$-almost-surely $X_\tau \neq y$.*

(d) *If $(X_t, t \geq 0)$ is a Markov jump process, then τ is a strictly positive, $\mathcal{F}_t^{\mathbf{X}}$-stopping time.*

Proof. (a). From Proposition 8.1.15 we know that $\tau \geq 0$ is an $\mathcal{F}_t^{\mathbf{X}}$-Markov time, as are $\tau_{u,z} = \inf\{t \geq u : X_t \neq z\}$ for any $z \in \mathbb{S}$ and $u \geq 0$. Under $\mathbf{P}_y$ the event $\{\tau \geq u + t\}$ for $t > 0$ implies that $X_u = y$ and $\tau = \tau_{u,X_u}$. Thus, applying the Markov property for $h = I_{\tau \geq t}$ (so $h(X_{u+\cdot}) = I_{\{\tau_{u,X_u} \geq u+t\}}$), we have that

$$
\begin{aligned}
&\mathbf{P}_y(\tau \geq u + t) = \mathbf{P}_y(\tau_{u,X_u} \geq u + t, X_u = y, \tau \geq u) \\
&= \mathbf{E}_y[\mathbf{E}(\tau_{u,X_u} \geq u + t | \mathcal{F}_u^{\mathbf{X}}) I_{\{\tau \geq u, X_u = y\}}] = \mathbf{P}_y(\tau \geq t)\mathbf{P}_y(\tau \geq u, X_u = y)\,.
\end{aligned}
$$

Considering this identity for $t = s + n^{-1}$ and $u = v + n^{-1}$ with $n \to \infty$, we find that the $[0, 1]$-valued function $g(t) = \mathbf{P}_y(\tau > t)$ is such that $g(s + v) = g(s)g(v)$ for all $s, v \geq 0$. Setting $g(1) = \exp(-\lambda_y)$ for $\lambda : \mathbb{S} \mapsto [0, \infty]$ which is measurable, by elementary algebra we have that $g(q) = \exp(-\lambda_y q)$ for any positive $q \in \mathbb{Q}$. Considering rational $q_n \downarrow t$ it thus follows that $\mathbf{P}_y(\tau > t) = \exp(-\lambda_y t)$ for all $t \geq 0$.
(b). If $\lambda_y = \infty$, then $\mathbf{P}_y$-a.s. both $\tau = 0$ and $X_\tau = X_0 = y$ are non-random, hence independent of each other. Suppose now that $\lambda_y > 0$ is finite, in which case τ is finite and positive $\mathbf{P}_y$-almost surely. Then, X_τ is well defined and applying the Markov property for $h = I_B(X_\tau)I_{\tau \geq t}$ (so $h(X_{u+\cdot}) = I_B(X_{\tau_{u,X_u}})I_{\{\tau_{u,X_u} \geq t+u\}}$), we have that for any $B \in \mathcal{S}$, $u \geq 0$ and $t > 0$,

$$
\begin{aligned}
\mathbf{P}_y(X_\tau \in B, \tau \geq t + u) &= \mathbf{P}_y(X_{\tau_{u,y}} \in B, \tau_{u,y} \geq t + u, \tau \geq u, X_u = y) \\
&= \mathbf{E}_y[\mathbf{P}(X_{\tau_{u,X_u}} \in B, \tau_{u,X_u} \geq t + u | \mathcal{F}_u^{\mathbf{X}}) I_{\{\tau \geq u, X_u = y\}}] \\
&= \mathbf{P}_y(X_\tau \in B, \tau \geq t)\mathbf{P}_y(\tau \geq u, X_u = y)\,.
\end{aligned}
$$

Considering this identity for $t = n^{-1}$ and $u = v + n^{-1}$ with $n \to \infty$, we find that

$$
\mathbf{P}_y(X_\tau \in B, \tau > v) = \mathbf{P}_y(X_\tau \in B, \tau > 0)\mathbf{P}_y(\tau > v) = \mathbf{P}_y(X_\tau \in B)\mathbf{P}_y(\tau > v)\,.
$$

Since $\{\tau > v\}$ and $\{X_\tau \in B, \tau < \infty\}$ are $\mathbf{P}_y$-independent for any $v \geq 0$ and $B \in \mathcal{S}$, it follows that the random variables τ and X_τ are $\mathbf{P}_y$-independent, as claimed.
(c). With $A = \{\exists q_n \in \mathbb{Q}, q_n \downarrow 0 : x(q_n) \neq x(0) = y\} \in \mathcal{S}^{[0,\infty)}$, note that by right-continuity of the sample functions $t \mapsto X_t(\omega)$ the event $\{X_\cdot \in A\}$ is merely $\{\tau = 0, X_0 = y\}$ and with λ_y finite, further $\mathbf{P}_z(X_\cdot \in A) = 0$ for all $z \in \mathbb{S}$. Since $\lambda_y > 0$, by the strong Markov property of $(X_t, t \geq 0)$ for $h(s, x(\cdot)) = I_A(x(\cdot))$ and the $\mathbf{P}_y$-a.s. finite $\mathcal{F}_t^{\mathbf{X}}$-Markov time τ, we find that $\mathbf{P}_y$-a.s. $X_{\tau+\cdot} \notin A$. Since the event $\{X_\tau = X_0\}$ implies, by definition of τ and sample path right-continuity, the existence of rational $q_n \downarrow 0$ such that $X_{\tau+q_n} \neq X_0$, we conclude that $\mathbf{P}_y$-a.s.

$X_\tau \neq y$.
(d). Here $t \mapsto X_t(\omega)$ is a step function, hence clearly, for each $t \geq 0$

$$\{\tau \leq t\} = \bigcup_{q \in \mathbb{Q}^{(2)}_{t+}} \{X_q \neq X_0\} \in \mathcal{F}^{\mathbf{X}}_t$$

and τ is a strictly positive, $\mathcal{F}^{\mathbf{X}}_t$-stopping time. □

Markov jump processes have the following parameters.

Definition 8.3.27. *We call* $p(x, A)$ *and* $\{\lambda_x\}$ *the* jump transition probability *and* jump rates *of a Markov jump process* $\{X_t, t \geq 0\}$*, if* $p(x, A) = \mathbf{P}_x(X_\tau \in A)$ *for* $A \in \mathcal{S}$ *and* $x \in \mathbb{S}$ *of positive jump rate* λ_x*, while* $p(x, A) = I_{x \in A}$ *in case* $\lambda_x = 0$*. More generally, a pair* (λ, p) *with* $\lambda : \mathbb{S} \mapsto \mathbb{R}_+$ *measurable and* $p(\cdot, \cdot)$ *a transition probability on* $(\mathbb{S}, \mathcal{S})$ *such that* $p(x, \{x\}) = I_{\lambda_x = 0}$ *is called* jump parameters.

The jump parameters provide the following canonical construction of Markov jump processes.

Theorem 8.3.28. *Suppose* (λ, p) *are jump parameters on a* $\mathcal{B}$*-isomorphic space* $(\mathbb{S}, \mathcal{S})$*. Let* $\{Z_n, n \geq 0\}$ *be the homogeneous Markov chain of transition probability* $p(\cdot, \cdot)$ *and initial state* $Z_0 = x \in \mathbb{S}$*. For each* $y \in \mathbb{S}$ *let* $\{\tau_j(y), j \geq 1\}$ *be i.i.d. random variables, independent of* $\{Z_n\}$ *and having each the exponential distribution of parameter* λ_y*. Set* $T_0 = 0$*,* $T_k = \sum_{j=1}^k \tau_j(Z_{j-1})$*,* $k \geq 1$ *and* $X_t = Z_k$ *for all* $t \in [T_k, T_{k+1})$*,* $k \geq 0$*. Assuming* $\mathbf{P}_x(T_\infty < \infty) = 0$ *for all* $x \in \mathbb{S}$*, the process* $\{X_t, t \geq 0\}$ *thus constructed is the unique Markov jump process with the given jump parameters. Conversely,* (λ, p) *are the parameters of a Markov jump process if and only if* $\mathbf{P}_x(T_\infty < \infty) = 0$ *for all* $x \in \mathbb{S}$*.*

Remark. The random time $\tau_j(Z_{j-1})$ is often called the *holding time* at state Z_{j-1} (or alternatively, the j-th holding time), along the sample path of the Markov jump process. However, recall part (c) of Proposition 8.3.26, that strong Markov processes of continuous sample path have trivial jump parameters, i.e. their holding times are either zero or infinite.

PROOF. Part I. *Existence.*
Starting from jump parameters (λ, p) and $X_0 = x$, if $\mathbf{P}_x(T_\infty < \infty) = 0$ then our construction produces $\{X_t, t \geq 0\}$ which is indistinguishable from a pure jump process and whose parameters coincide with the specified (λ, p). So, assuming hereafter with no loss of generality that $T_\infty(\omega) = \infty$ for *all* $\omega \in \Omega$, we proceed to show that $\{X_t, \mathcal{F}^{\mathbf{X}}_t\}$ is a homogeneous Markov process. Indeed, since $p_t(x, B) = \mathbf{P}_x(X_t \in B)$ is per $t \geq 0$ a transition probability on $(\mathbb{S}, \mathcal{S})$, this follows as soon as we show that $\mathbf{P}_x(X_{s+u} \in B | \mathcal{F}^{\mathbf{X}}_s) = \mathbf{P}_{X_s}(X_u \in B)$ for any fixed $s, u \geq 0$, $x \in \mathbb{S}$ and $B \in \mathcal{S}$.

Turning to prove the latter identity, fix s, u, x, B and note that

$$\{X_u \in B\} = \bigcup_{\ell \geq 0} \{Z_\ell \in B, T_{\ell+1} > u \geq T_\ell\},$$

is of the form $\{X_u \in B\} = \{(Z_\cdot, T_\cdot) \in A_u\}$ where $A_u \in (\mathbb{S} \times [0, \infty])_c$. Hence, this event is determined by the law of the homogeneous Markov chain $\{Z_n, T_n, n \geq 0\}$ on $\mathbb{S} \times [0, \infty]$. With $Y_t = \sup\{k \geq 0 : T_k \leq t\}$ counting the number of jumps in the interval $[0, t]$, we further have that if $\{Y_s = k\}$, then $X_s = Z_k$ and $\{X_{s+u} \in$

$B\} = \{(Z_{k+\cdot}, T_{k+\cdot} - s) \in A_u\}$. Moreover, since $t \mapsto X_t(\omega)$ is a step function, $\mathcal{F}_s^{\mathbf{X}} = \sigma(Y_s, Z_k, T_k, k \le Y_s)$. Thus, decomposing Ω as the union of disjoint events $\{Y_s = k\}$ it suffices to show that under $\mathbf{P}_x$, the law of $(Z_{k+\cdot}, T_{k+\cdot} - s)$ conditional on (Z_k, T_k) and the event $\{Y_s = k\} = \{\tau_{k+1}(Z_k) > s - T_k \ge 0\}$, is the same as the law of $(Z_\cdot, T_\cdot)$ under $\mathbf{P}_{Z_k}$. In our construction, given $Z_k \in \mathbb{S}$, the random variable $\tau = \tau_{k+1}(Z_k) = T_{k+1} - T_k$ is independent of T_k and follows the exponential distribution of parameter λ_{Z_k}. Hence, setting $\xi = s - T_k \ge 0$, by the lack of memory of this exponential distribution, for any $t, s, k \ge 0$,

$$\mathbf{P}_x(T_{k+1} > t + s | T_k, Z_k, \{Y_s = k\}) = \mathbf{P}_x(\tau > t + \xi | T_k, Z_k, \{\tau > \xi\}) = \mathbf{P}(\tau > t | Z_k) \, .$$

That is, under $\mathbf{P}_x$, the law of $T_{k+1} - s$ conditional on $\mathcal{F}_s^{\mathbf{X}}$ and the event $\{Y_s = k\}$, is the same as the law of T_1 under $\mathbf{P}_{Z_k}$. With $\{Z_n, n \ge 0\}$ a homogeneous Markov chain whose transition probabilities are independent of $\{T_n, n \ge 0\}$, it follows that further the joint law of $(Z_{k+1}, T_{k+1} - s)$ conditional on $\mathcal{F}_s^{\mathbf{X}}$ and the event $\{Y_s = k\}$ is the same as the joint law of (Z_1, T_1) under $\mathbf{P}_{Z_k}$. This completes our proof that $\{X_t, \mathcal{F}_t^{\mathbf{X}}\}$ is a homogeneous Markov process, since for any $z \in \mathbb{S}$, conditional on $Z_{k+1} = z$ the value of $(Z_{k+1+\cdot}, T_{k+1+\cdot} - T_{k+1})$ is independent of T_{k+1} and by the Markov property has the same joint law as $(Z_{1+\cdot}, T_{1+\cdot} - T_1)$ given $Z_1 = z$.

Part II. *Uniqueness*. Start conversely with a Markov pure jump process $(\widehat{X}_t, t \ge 0)$ such that $\widehat{X}_0 = x$ and whose jump parameters per Definition 8.3.27 are (λ, p). In the sequel we show that with probability one we can embed within its sample function $t \mapsto \widehat{X}_t(\omega)$ a realization of the Markov chain $\{Z_n, n \ge 0\}$ of transition probability $p(\cdot, \cdot)$, starting at $Z_0 = x$, such that $\widehat{X}_t = Z_k$ for all $t \in [T_k, T_{k+1})$, $k \ge 0$ and with $T_0 = 0$, show that for any $k \ge 0$, conditionally on $\{Z_j, T_j, j \le k\}$, the variables $\tau_{k+1} = T_{k+1} - T_k$ and Z_{k+1} are independent of each other, with τ_{k+1} having the exponential distribution of parameter λ_{Z_k}.

This of course implies that even conditionally on the infinite sequence $\{Z_n, n \ge 0\}$, the holding times $\{\tau_{k+1}, k \ge 0\}$ are independent of each other, with τ_{k+1} maintaining its exponential distribution of parameter λ_{Z_k}. Further, since $t \mapsto \widehat{X}_t(\omega)$ is a step function (see Definition 8.3.23), necessarily here $T_\infty(\omega) = \infty$ for all $\omega \in \Omega$. This applies for any non-random $x \in \mathbb{S}$, thus showing that any Markov pure jump process can be constructed as in the statement of the theorem, provided (λ, p) are such that $\mathbf{P}_x(T_\infty < \infty) = 0$ for all $x \in \mathbb{S}$, with the latter condition also necessary for (λ, p) to be the jump parameters of *any* Markov pure jump process (and a moment thought will convince you that this completes the proof of the theorem).

Turning to the promised embedding, let $T_0 = 0$, $Z_0 = \widehat{X}_0 = x$ and $T_1 = T_0 + \tau_1$ for $\tau_1 = \inf\{t \ge 0 : \widehat{X}_t \neq Z_0\}$. Recall Proposition 8.3.26 that τ_1 has the exponential distribution of parameter λ_x and is an $\mathcal{F}_t^{\widehat{\mathbf{X}}}$-stopping time. In case $\lambda_x = 0$ we are done for then $T_1 = \infty$ and $\widehat{X}_t(\omega) = Z_0(\omega) = x$ for all $t \ge 0$. Otherwise, recall Proposition 8.3.26 that T_1 is finite w.p.1. in which case $Z_1 = X_{T_1}$ is well defined and independent of T_1, with the law of Z_1 being $p(Z_0, \cdot)$. Further, since $(\widehat{X}_t, \mathcal{F}_t^{\widehat{\mathbf{X}}})$ is a strong Markov process (see Proposition 8.3.24) and excluding the null set $\{\omega : T_1(\omega) = \infty\}$, upon applying the strong Markov property at the finite stopping time T_1 we deduce that conditional on $\mathcal{F}_{T_1}^{\widehat{\mathbf{X}}}$ the process $\{\widehat{X}_{T_1+t}, t \ge 0\}$ is a Markov pure jump process, of the same jump parameters, but now starting at Z_1. We can thus repeat this procedure and w.p.1. construct the sequence $\tau_{k+1} = \inf\{t \ge 0 : \widehat{X}_{t+T_k} \neq \widehat{X}_{T_k}\}$, $k \ge 1$, where $T_{k+1} = T_k + \tau_{k+1}$ are $\mathcal{F}_t^{\widehat{\mathbf{X}}}$-stopping times and $Z_k = \widehat{X}_{T_k}$ (terminating at $T_{k+1} = \infty$ if $\lambda_{Z_k} = 0$). This is the embedding described before, for indeed $\widehat{X}_t = Z_k$ for all $t \in [T_k, T_{k+1})$, the sequence $\{Z_n, n \geq 0\}$ has the law of a homogeneous Markov chain of transition probability $p(\cdot,\cdot)$ (starting at $Z_0 = x$), and conditionally on $\sigma(Z_j, T_j, j \leq k) \subseteq \mathcal{F}^{\widehat{\mathbf{X}}}_{T_k}$, the variables τ_{k+1} and Z_{k+1} are independent of each other, with τ_{k+1} having the exponential distribution of parameter λ_{Z_k}. $\square$

Remark 8.3.29. When the jump rates $\lambda_x = \lambda$ are constant, the corresponding jump times T_k are those of a Poisson process N_t of rate λ, which is independent of the Markov chain $\{Z_n\}$. Hence, in this case the Markov jump process has the particularly simple structure $X_t = Z_{N_t}$.

Here is a more explicit, equivalent condition for existence of a Markov pure jump process with the specified jump parameters (λ, p). It implies in particular that such a process exists whenever the jump rates are bounded (i.e. $\sup_x \lambda_x$ finite).

Exercise 8.3.30. *Suppose (λ, p) are jump parameters on the $\mathcal{B}$-isomorphic state space $(\mathbb{S}, \mathcal{S})$.*

(a) *Show that $\mathbf{P}_x(T_\infty < \infty) = 0$ if and only if $\mathbf{P}_x(\sum_n \lambda_{Z_n}^{-1} < \infty) = 0$.*
Hint: *Upon Conditioning on $\{Z_n\}$ consider part (d) of Exercise 2.3.25.*

(b) *Conclude that to any jump parameters $p(\cdot,\cdot)$ and $\lambda \in b\mathcal{S}$ corresponds a well defined, unique Markov jump process, constructed as in Theorem 8.3.28.*

Remark. The event $\{\omega : T_\infty(\omega) < \infty\}$ is often called an *explosion*. A further distinction can then be made between the pure (or non-explosive) Markov jump processes we consider here, and the *explosive Markov jump processes* such that $\mathbf{P}_x(T_\infty < \infty) > 0$ for some $x \in \mathbb{S}$, which nevertheless can be constructed as in Theorem 8.3.28 to have step sample functions, but only up to the time T_∞ of explosion.

Example 8.3.31 (BIRTH PROCESSES). *Markov (jump) processes which are also counting processes, are called* birth processes. *The state space of such processes is $\mathbb{S} = \{0, 1, 2, \ldots\}$ and in view of Theorem 8.3.28 they correspond to jump transitions $p(x, x+1) = 1$. Specifically, these processes are of the form $X_t = \sup\{k \geq 0 : \sum_{j=X_0}^{k-1} \tau_j \leq t\}$, where the holding times τ_j, $j \geq 1$, are independent Exponential(λ_j) random variables. In view of Exercise 8.3.30 such processes are non-explosive if and only if $\sum_{j \geq k} \lambda_j^{-1} = \infty$ for all $k \geq 0$. For example, this is the case when $\lambda_j = j\Delta + \lambda_0$ with $\lambda_0 \geq 0$ and $\Delta > 0$, and such a process is then called* simple birth with immigration process *if also $\lambda_0 > 0$, or merely* simple birth process *if $\lambda_0 = 0$ (in contrast, the* Poisson process *corresponds to $\Delta = 0$ and $\lambda_0 > 0$). The latter processes serve in modeling the growth in time of a population composed of individuals who independently give birth at rate Δ (following an exponentially distributed holding time between consecutive birth events), with additional immigration into the population at rate λ_0, independently of birth events.*

Remark. In the context of Example 8.3.31, $\mathbf{E}_x T_k = \sum_{j=x}^{x+k-1} \lambda_j^{-1}$ for the arrival time T_k to state $k+x$, so taking for example $\lambda_j = j^\alpha$ for some $\alpha > 1$ results with an explosive Markov jump process. Indeed, then $\mathbf{E}_x T_k \leq c$ for finite $c = \sum_{j \geq 1} j^{-\alpha}$ and any $x, k \geq 1$. By monotone convergence $\mathbf{E}_x T_\infty \leq c$, so within an integrable, hence a.s. finite time T_∞ the sample function $t \mapsto X_t$ escapes to infinity, hence the name explosion given to such phenomena. But, observe that unbounded jump rates do not necessarily imply an explosion (as for example, in case of simple birth processes), and explosion may occur for one initial state but not for another (for example here $\lambda_0 = 0$ so there is no explosion if starting at $x = 0$).

As you are to verify now, the jump parameters characterize the relatively explicit generator for the semi-group of a Markov jump process, which in particular satisfies Kolmogorov's forward (in case of bounded jump rates), and backward equations.

Definition 8.3.32. *The linear operator* $\mathbf{L} : b\mathcal{S} \mapsto m\mathcal{S}$ *such that* $(\mathbf{L}h)(x) = \lambda_x \int (h(y) - h(x))p(x, dy)$ *for* $h \in b\mathcal{S}$ *is called the* generator *of the Markov jump process corresponding to jump parameters* (λ, p)*. In particular,* $(\mathbf{L}I_{\{x\}^c})(x) = \lambda_x$ *and more generally* $(\mathbf{L}I_B)(x) = \lambda_x p(x, B)$ *for any* $B \subseteq \{x\}^c$ *(so specifying the generator is in this context equivalent to specifying the jump parameters).*

Exercise 8.3.33. *Consider a Markov jump process* $(X_t, t \geq 0)$ *of semi-group* $p_t(\cdot, \cdot)$ *and jump parameters* (λ,p) *as in Definition 8.3.32. Let* $T_k = \sum_{j=1}^k \tau_j$ *denote the jump times of the sample function* $s \mapsto X_s(\omega)$ *and* $Y_t = \sum_{k \geq 1} I_{\{T_k \leq t\}}$ *the number of such jumps in the interval* $[0, t]$*.*

(a) *Show that if* $\lambda_x > 0$ *then*

$$\mathbf{P}_x(\tau_2 \leq t|\tau_1) = \int (1 - e^{-\lambda_y t})p(x, dy)\,,$$

and deduce that $t^{-1}\mathbf{P}_x(Y_t \geq 2) \to 0$ *as* $t \downarrow 0$*, for any* $x \in \mathbb{S}$*.*

(b) *Fixing* $x \in \mathbb{S}$ *and* $h \in b\mathcal{S}$*, show that*

$$|(p_s h)(x) - (p_0 h)(x) - \mathbf{E}_x[(h(X_\tau) - h(x))I_{\tau \leq s}]| \leq 2\|h\|_\infty \mathbf{P}_x(Y_s \geq 2)\,,$$

and deduce that for $\mathbf{L}$ *per Definition 8.3.32,*

$$\lim_{s \downarrow 0} s^{-1}((p_s h)(x) - (p_0 h)(x)) = (\mathbf{L}h)(x)\,, \tag{8.3.20}$$

where the convergence in (8.3.20) is uniform within $\{h \in b\mathcal{S} : \|h\|_\infty \leq K\}$*, for any* K *finite.*

(c) *Verify that* $t \mapsto (\mathbf{L}p_t h)(x)$ *is continuous and* $t \mapsto (p_t h)(x)$ *is differentiable for any* $x \in \mathbb{S}$*,* $h \in b\mathcal{S}$*,* $t \geq 0$ *and conclude that the backward Kolmogorov equation holds. Specifically, show that*

$$\partial_t (p_t h)(x) = (\mathbf{L}p_t h)(x) \qquad \forall t \geq 0, x \in \mathbb{S}, h \in b\mathcal{S}. \tag{8.3.21}$$

(d) *Show that if* $\sup_{x \in \mathbb{S}} \lambda_x$ *is finite, then* $\mathbf{L} : b\mathcal{S} \mapsto b\mathcal{S}$*, the convergence in (8.3.20) is also uniform in* x *and* Kolmogorov's forward equation *(also known as the* Fokker-Planck equation*), holds. That is,*

$$\partial_t (p_t h)(x) = (p_t(\mathbf{L}h))(x) \qquad \forall t \geq 0, x \in \mathbb{S}, h \in b\mathcal{S}. \tag{8.3.22}$$

Remark. Exercise 8.3.33 relates the Markov semi-group with the corresponding jump parameters, showing that a Markov semi-group $p_t(\cdot, \cdot)$ corresponds to a Markov jump process only if for any $x \in \mathbb{S}$, the limit

$$\lim_{t \downarrow 0} t^{-1}(1 - p_t(x, \{x\}) = \lambda_x \tag{8.3.23}$$

exists, is finite and $\mathcal{S}$-measurable. Moreover, necessarily then also

$$\lim_{t \downarrow 0} t^{-1} p_t(x, B) = \lambda_x p(x, B) \qquad \forall B \subseteq \{x\}^c\,, \tag{8.3.24}$$

for some transition probability $p(\cdot,\cdot)$. Recall Theorem 8.3.28 that with the exception of possible explosion, the converse applies, namely whenever (8.3.23) and (8.3.24) hold, the semi-group $p_t(\cdot,\cdot)$ corresponds to a (possibly explosive) Markov jump process. We note in passing that while Kolmogorov's backward equation (8.3.21) is well defined for *any* jump parameters, the existence of solution which is a Markov semi-group, is equivalent to non-explosion of the corresponding Markov jump process.

In particular, in case of bounded jump rates the conditions (8.3.23) and (8.3.24) are equivalent to the Markov process being a Markov pure jump process and in this setting you are now to characterize the invariant measures for the Markov jump process in terms of its jump parameters (or equivalently, in terms of its generator).

Exercise 8.3.34. *Suppose (λ, p) are jump parameters on $\mathcal{B}$-isomorphic state space $(\mathbb{S}, \mathcal{S})$ such that $\sup_{x\in\mathbb{S}} \lambda_x$ is finite.*

(a) *Show that probability measure ν is invariant for the corresponding Markov jump process if and only if $\nu(\mathbf{L}h) = 0$ for the generator $\mathbf{L} : b\mathcal{S} \mapsto b\mathcal{S}$ of these jump parameters and all $h \in b\mathcal{S}$.*
Hint: *Combine Exercises 8.3.9 and 8.3.33 (utilizing the boundedness of $x \mapsto (\mathbf{L}h)(x)$).*

(b) *Deduce that ν is an invariant probability measure for (λ, p) if and only if $\nu(\lambda p) = \lambda\nu$.*

In particular, the invariant probability measures of a Markov jump process with constant jump rates are precisely the invariant measures for its jump transition probability.

Of particular interest is the following special family of Markov jump processes.

Definition 8.3.35. *Real-valued Markov pure jump processes with a constant jump rate λ whose jump transition probability is of the form $p(x, B) = \mathcal{P}_\xi(\{z : x+z \in B\})$ for some law $\mathcal{P}_\xi$ on $(\mathbb{R}, \mathcal{B})$, are called* compound Poisson processes. *Recall Remark 8.3.29 that a compound Poisson process is of the form $X_t = S_{N_t}$ for a random walk $S_n = S_0 + \sum_{k=1}^n \xi_k$ with i.i.d. $\{\xi, \xi_k\}$ which are independent of the Poisson process N_t of rate λ.*

Remark. The random telegraph signal $R_t = (-1)^{N_t} R_0$ of Example 7.2.14 is a Markov jump process on $\mathbb{S} = \{-1, 1\}$ with constant jump rate λ, which is not a compound Poisson process (as its transition probabilities $p(1,-1) = p(-1,1) = 1$ do not correspond to a random walk).

As we see next, compound Poisson processes retain many of the properties of the Poisson process.

Proposition 8.3.36. *A compound Poisson process $\{X_t, t \geq 0\}$ has stationary, independent increments and the characteristic function of its Markov semi-group $p_t(x, \cdot)$ is*

$$\mathbf{E}_x[e^{i\theta X_t}] = e^{i\theta x + \lambda t(\Phi_\xi(\theta)-1)}\,, \tag{8.3.25}$$

where $\Phi_\xi(\cdot)$ denotes the characteristic function of the corresponding jump sizes ξ_k.

Proof. We start by proving that $\{X_t, t \geq 0\}$ has independent increments, where by Exercise 7.1.12 it suffices to fix $0 = t_0 < t_1 < t_2 < \cdots < t_n$ and show that the random variables $D_i = X_{t_i} - X_{t_{i-1}}$, $i = 1, \ldots, n$, are mutually independent. To this end, note that $N_{t_0} = 0$ and conditional on the event $N_{t_i} = m_i$ for $m_i = \sum_{j=1}^{i} r_j$ and fixed $\underline{r} = (r_1, \dots, r_n) \in \mathbb{Z}_+^n$, we have that $D_i = \sum_{k=m_{i-1}+1}^{m_i} \xi_k$ are mutually independent with D_i then having the same distribution as the random walk S_{r_i} starting at $S_0 = 0$. So, for any $f_i \in b\mathcal{B}$, by the tower property and the mutual independence of $\{N_{t_i} - N_{t_{i-1}}, 1 \le i \le n\}$,

$$\mathbf{E}_x[\prod_{i=1}^n f_i(D_i)] = \mathbf{E}[\mathbf{E}_x(\prod_{i=1}^n f_i(D_i)|\mathcal{F}^{\mathbf{N}})] = \sum_{\underline{r}\in\mathbb{Z}_+^n} \prod_{i=1}^n \Big\{\mathbf{P}(N_{t_i} - N_{t_{i-1}} = r_i)\mathbf{E}_0[f_i(S_{r_i})]\Big\}$$

$$= \prod_{i=1}^n \Big\{ \sum_{r_i=0}^{\infty} \mathbf{P}(N_{t_i} - N_{t_{i-1}} = r_i)\mathbf{E}_0[f_i(S_{r_i})]\Big\} = \prod_{i=1}^n \mathbf{E}_x[f_i(D_i)]\,,$$

yielding the mutual independence of D_i, $i = 1, \dots, n$.

We have just seen that for each $t > s$ the increment $X_t - X_s$ has under $\mathbf{P}_x$ the same law as $S_{N_t - N_s}$ has under $\mathbf{P}_0$. Since $N_t - N_s \overset{\mathcal{D}}{=} N_{t-s}$, it follows by the independence of the random walk $\{S_r\}$ and the Poisson process $\{N_t, t \ge 0\}$ that $\mathbf{P}_x(X_t - X_s \in \cdot) = \mathbf{P}_0(S_{N_{t-s}} \in \cdot)$ depends only on $t - s$, which by Definition 7.3.11 amounts to $\{X_t, t \ge 0\}$ having stationary increments.

Finally, the identity (3.3.3) extends to $\mathbf{E}[z^{N_t}] = \exp(\lambda t(z - 1))$ for N_t having a Poisson distribution with parameter λt and any complex variable z. Thus, as $\mathbf{E}_x[e^{i\theta S_r}] = e^{i\theta x}\Phi_\xi(\theta)^r$ (see Lemma 3.3.8), utilizing the independence of $\{S_r\}$ from N_t, we conclude that

$$\mathbf{E}_x[e^{i\theta X_t}] = \mathbf{E}[\mathbf{E}_x(e^{i\theta S_{N_t}}|N_t)] = e^{i\theta x}\mathbf{E}[\Phi_\xi(\theta)^{N_t}] = e^{i\theta x + \lambda t(\Phi_\xi(\theta)-1)}\,,$$

for all $t \ge 0$ and $x, \theta \in \mathbb{R}$, as claimed. $\square$

Exercise 8.3.37. *Let $\{X_t, t \ge 0\}$ be a compound Poisson process of jump rate λ.*

(a) *Show that if the corresponding jump sizes $\{\xi_k\}$ are square integrable then $\mathbf{E}_0 X_t = \lambda t \mathbf{E}\xi_1$ and $\mathsf{Var}(X_t) = \lambda t \mathbf{E}\xi_1^2$.*

(b) *Show that if $\mathbf{E}\xi_1 = 0$ then $\{X_t, t \ge 0\}$ is a martingale. More generally, $u_0(t, X_t, \theta)$ is a martingale for $u_0(t, y, \theta) = \exp(\theta y - \lambda t(M_\xi(\theta) - 1))$ and any $\theta \in \mathbb{R}$ for which the moment generating function $M_\xi(\theta) = \mathbf{E}[e^{\theta\xi_1}]$ is finite.*

Here is the analog for compound Poisson processes of the thinning of Poisson variables.

Proposition 8.3.38. *Suppose $\{X_t, t \ge 0\}$ is a compound Poisson process of jump rate λ and jump size law $\mathcal{P}_\xi$. Fixing a disjoint finite partition of $\mathbb{R} \setminus \{0\}$ to Borel sets B_j, $j = 1, \dots, m$, consider the decomposition $X_t = X_0 + \sum_{j=1}^{m} X_t^{(j)}$ in terms of the contributions*

$$X_t^{(j)} = \sum_{k=1}^{N_t} \xi_k I_{B_j}(\xi_k)$$

to X_t by jumps whose size belong to B_j. Then $\{X_t^{(j)}, t \ge 0\}$ for $j = 1, \dots, m$ are independent compound Poisson processes of jump rates $\lambda^{(j)} = \lambda\mathbf{P}(\xi \in B_j)$ and i.i.d. jump sizes $\{\xi^{(j)}, \xi_\ell^{(j)}\}$ such that $\mathbf{P}(\xi^{(j)} \in \cdot) = \mathbf{P}(\xi \in \cdot|\xi \in B_j)$, starting at $X_0^{(j)} = 0$.

Proof. While one can directly prove this result along the lines of Exercise 3.4.16, we resort to an indirect alternative, whereby we set $\widehat{X}_t = X_0 + \sum_{j=1}^m Y_t^{(j)}$ for the independent compound Poisson processes $Y_t^{(j)}$ of jump rates $\lambda^{(j)}$ and i.i.d. jump sizes $\{\xi_k^{(j)}\}$, starting at $Y_0^{(j)} = 0$. By construction, $\widehat{X}_t$ is a pure jump process whose jump times $\{T_k(\omega)\}$ are contained in the union over $j = 1, \ldots, m$ of the isolated jump times $\{T_k^{(j)}(\omega)\}$ of $t \mapsto Y_t^{(j)}(\omega)$. Recall that each $T_k^{(j)}$ has the gamma density of parameters $\alpha = k$ and $\lambda^{(j)}$ (see Exercise 1.4.46 and Definition 3.4.8). Therefore, by the independence of $\{Y_t^{(j)}, t \geq 0\}$ w.p.1. no two jump times among $\{T_k^{(j)}, j, k \geq 1\}$ are the same, in which case $\widehat{X}_t^{(j)} = Y_t^{(j)}$ for all j and $t \geq 0$ (as the jump sizes of each $Y_t^{(j)}$ are in the disjoint element B_j of the specified finite partition of $\mathbb{R} \setminus \{0\}$). With the $\mathbb{R}^m$-valued process $(\widehat{X}^{(1)}, \ldots, \widehat{X}^{(m)})$ being indistinguishable from $(Y^{(1)}, \ldots, Y^{(m)})$, it thus suffices to show that $\{\widehat{X}_t, t \geq 0\}$ is a compound Poisson process of the specified jump rate λ and jump size law $\mathcal{P}_\xi$.

To this end, recall Proposition 8.3.36 that each of the processes $Y_t^{(j)}$ has stationary independent increments and due to their independence, the same applies for $\{\widehat{X}_t, t \geq 0\}$, which is thus a real-valued homogeneous Markov process (see Proposition 8.3.5). Next, note that since $\lambda = \sum_{j=1}^m \lambda^{(j)}$ and for all $\theta \in \mathbb{R}$,

$$\sum_{j=1}^m \lambda^{(j)} \Phi_{\xi^{(j)}}(\theta) = \lambda \sum_{j=1}^m \mathbf{E}[e^{i\theta\xi} I_{B_j}(\xi)] = \Phi_\xi(\theta)\,,$$

we have from (8.3.25) and Lemma 3.3.8 that for any $\theta \in \mathbb{R}$,

$$\begin{aligned}
\mathbf{E}_x[e^{i\theta \widehat{X}_t}] &= e^{i\theta x} \prod_{j=1}^m \mathbf{E}[e^{i\theta Y_t^{(j)}}] \\
&= e^{i\theta x} \prod_{j=1}^m e^{\lambda^{(j)} t(\Phi_{\xi^{(j)}}(\theta)-1)} = e^{i\theta x + \lambda t(\Phi_\xi(\theta)-1)} = \mathbf{E}_x[e^{i\theta X_t}]\,.
\end{aligned}$$

That is, denoting by $p_t(\cdot,\cdot)$ and $\widehat{p}_t(\cdot,\cdot)$ the Markov semi-groups of $\{X_t, t \geq 0\}$ and $\{\widehat{X}_t, t \geq 0\}$ respectively, we found that per fixed $x \in \mathbb{R}$ and $t \geq 0$ the transition probabilities $p_t(x,\cdot)$ and $\widehat{p}_t(x,\cdot)$ have the same characteristic function. Consequently, by Lévy's inversion theorem $p_t(\cdot,\cdot) = \widehat{p}_t(\cdot,\cdot)$ for all $t \geq 0$, i.e., these two semi-groups are identical. Obviously, this implies that the Markov pure jump processes X_t and $\widehat{X}_t$ have the same jump parameters (see (8.3.23) and (8.3.24)), so as claimed $\{\widehat{X}_t, t \geq 0\}$ is a compound Poisson process of jump rate λ and jump size law $\mathcal{P}_\xi$. □

Exercise 8.3.39. *Suppose $\{Y_t, t \geq 0\}$ is a compound Poisson process of jump rate $\lambda > 0$, integrable Y_0 and jump size law $\mathcal{P}_\xi$ for some integrable $\xi > 0$.*

(a) *Show that for any integrable $\mathcal{F}_t^{\mathbf{Y}}$-Markov time τ, $\mathbf{E}Y_\tau = \mathbf{E}Y_0 + \lambda\mathbf{E}\xi\mathbf{E}\tau$.*
Hint: *Consider the $\mathcal{F}_t^{\mathbf{Y}}$-martingales $X_{t\wedge n}$, $n \geq 1$, where $X_t = Y_t - \lambda t\mathbf{E}\xi$.*

(b) *Suppose that Y_0 and ξ are square integrable and let $\theta_r = \inf\{t \geq 0 : Z_t^{(r)} > Y_t\}$, where $Z_t^{(r)} = B_t + rt$ for a standard Brownian Markov process $(B_t, t \geq 0)$, independent of $(Y_t, t \geq 0)$. Show that for any $r \in \mathbb{R}$,*

$$\mathbf{E}\theta_r = \frac{\mathbf{E}(Y_0)_+}{(r - \lambda\mathbf{E}\xi)_+}$$

(where trivially $\theta_r = 0$ w.p.1. in case $\mathbf{E}(Y_0)_+ = 0$).
Hint: *Consider filtration $\mathcal{F}_t = \sigma(Y_s, B_s, s \le t)$ and MG $M_t = X_t - B_t$.*

As in the case of Markov chains, the jump transition probability of a Markov jump process with countable state space $\mathbb{S}$ is of the form $p(x, A) = \sum_{y \in A} p(x, y)$. In this case, accessibility and intercommunication of states, as well as irreducible, transient and recurrent classes of states, are defined according to the transition probability $\{p(x, y)\}$ and obey the relations explored already in Subsection 6.2.1. Moreover, as you are to check next, Kolmogorov's equations (8.3.21) and (8.3.22) are more explicit in this setting.

Exercise 8.3.40. *Suppose (λ, p) are the parameters of a Markov jump process on a countable state space $\mathbb{S}$.*

(a) *Check that $p_s(x, z) = \mathbf{P}_x(X_s = z)$ are then the solution of the countable system of linear ODEs*

$$\frac{dp_s(x, z)}{ds} = \sum_{y \in \mathbb{S}} q(x, y) p_s(y, z) \qquad \forall s \ge 0, x, z \in \mathbb{S},$$

starting at $p_0(x, z) = I_{x=z}$, where $q(x, x) = -\lambda_x$ and $q(x, y) = \lambda_x p(x, y)$ for $x \neq y$.

(b) *Show that if $\sup_x \lambda_x$ is finite then $p_s(x, z)$ must also satisfy the corresponding forward equation*

$$\frac{dp_s(x, z)}{ds} = \sum_{y \in \mathbb{S}} p_s(x, y) q(y, z) \qquad \forall s \ge 0, x, z \in \mathbb{S}.$$

(c) *In case $\mathbb{S}$ is a finite set, show that the matrix $\mathbf{P}_s$ of entries $p_s(x, z)$ is given by $\mathbf{P}_s = e^{s\mathbf{Q}} = \sum_{k=0}^{\infty} \frac{s^k}{k!} \mathbf{Q}^k$, where $\mathbf{Q}$ is the matrix of entries $q(x, y)$.*

The formula $\mathbf{P}_s = e^{s\mathbf{Q}}$ explains why $\mathbf{Q}$, and more generally $\mathbf{L}$, is called the generator *of the semi-group $\mathbf{P}_s$.*

From Exercise 8.3.34 we further deduce that, at least for bounded jump rates, an invariant probability measure for the Markov jump process is uniquely determined by the function $\pi : \mathbb{S} \mapsto [0, 1]$ such that $\sum_x \pi(x) = 1$ and

$$\lambda_y \pi(y) = \sum_{x \in \mathbb{S}} \pi(x) \lambda_x p(x, y) \qquad \forall y \in \mathbb{S}\,. \tag{8.3.26}$$

For constant positive jump rates this condition coincides with the characterization (6.2.5) of invariant probability measures for the jump transition probability. Consequently, for such jump processes the invariant, reversible and excessive measures as well as positive and null recurrent states are defined as the corresponding objects for the jump transition probability and obey the relations explored already in Subsection 6.2.2.

Remark. While we do not pursue this further, we note in passing that more generally, a measure $\mu(\cdot)$ is reversible for a Markov jump process with countable state space $\mathbb{S}$ if and only if $\lambda_y \mu(y) p(y, x) = \mu(x) \lambda_x p(x, y)$ for any $x, y \in \mathbb{S}$ (so any reversible probability measure is by (8.3.26) invariant for the Markov jump process). Similarly, in general we call $x \in \mathbb{S}$ with $\lambda_x = 0$ an absorbing, hence positive recurrent, state and say that a non-absorbing state is positive recurrent if it has finite mean return time. That is, if $\mathbf{E}_x T_x < \infty$ for the *first return time* $T_x = \inf\{t \geq \tau : X_t = x\}$ to state x. It can then be shown, in analogy with Proposition 6.2.41, that any invariant probability measure $\pi(\cdot)$ is zero outside the positive recurrent states and if its support is an irreducible class $\mathbf{R}$ of non-absorbing positive recurrent states, then $\pi(z) = 1/(\lambda_z \mathbf{E}_z[T_z])$ (see, [**GS01**, Section 6.9] for more details).

To practice your understanding, the next exercise explores in more depth the important family of *birth and death Markov jump processes* (or in short, birth and death processes).

Exercise 8.3.41 (Birth and death processes)**.** *A birth and death process is a Markov jump process $\{X_t\}$ on $\mathbb{S} = \{0, 1, 2, \ldots\}$ for which $\{Z_n\}$ is a birth and death chain. That is, $p(x, x+1) = p_x = 1 - p(x, x-1)$ for all $x \in \mathbb{S}$ (where of course $p_0 = 1$). Assuming $\lambda_x > 0$ for all x and $p_x \in (0,1)$ for all $x > 0$, let*

$$\widehat{\pi}(k) = \frac{\lambda_0}{\lambda_k} \prod_{i=1}^{k} \frac{p_{i-1}}{1 - p_i} .$$

Show that $\{X_t\}$ is irreducible and has an invariant probability measure if and only if $c = \sum_{k \geq 0} \widehat{\pi}(k)$ is finite, in which case its invariant measure is $\pi(k) = \widehat{\pi}(k)/c$.

The next exercise deals with independent random sampling along the path of a Markov pure jump process.

Exercise 8.3.42. *Let $Y_k = X_{\widetilde{T}_k}$, $k = 0, 1, \ldots$, where $\widetilde{T}_k = \sum_{i=1}^{k} \widetilde{\tau}_i$ and the i.i.d. $\widetilde{\tau}_i \geq 0$ are independent of the Markov pure jump process $\{X_t, t \geq 0\}$.*

(a) *Show that $\{Y_k\}$ is a homogeneous Markov chain and verify that any invariant probability measure for $\{X_t\}$ is also an invariant measure for $\{Y_k\}$.*
(b) *Show that in case of constant jump rates $\lambda_x = \lambda$ and each $\widetilde{\tau}_i$ having the exponential distribution of parameter $\widetilde{\lambda} > 0$, one has the representation $Y_k = Z_{L_k}$ of sampling the embedded chain $\{Z_n\}$ at $L_k = \sum_{i=1}^{k} (\eta_i - 1)$ for i.i.d. $\eta_i \geq 1$, each having the Geometric distribution of success probability $p = \widetilde{\lambda}/(\lambda + \widetilde{\lambda})$.*
(c) *Conclude that if $\{\widetilde{T}_k\}$ are the jump times of a Poisson process of rate $\widetilde{\lambda} > 0$ which is independent of the compound Poisson process $\{X_t\}$, then $\{Y_k\}$ is a random walk, the increment of which has the law of $\sum_{i=1}^{\eta-1} \xi_i$.*

Compare your next result with part (a) of Exercise 8.2.46.

Exercise 8.3.43. *Suppose $\{X_t, t \geq 0\}$ is a real-valued Markov pure jump process, with $0 = T_0 < T_1 < T_2 < \cdots$ denoting the jump times of its sample function. Show that for any $q > 0$ its finite q-th variation $V^{(q)}(X)_t$ exists, and is given by*

$$V^{(q)}(X)_t = \sum_{k \geq 1} I_{\{T_k \leq t\}} |X_{T_k} - X_{T_{k-1}}|^q .$$

CHAPTER 9

The Brownian motion

The Brownian motion is the most fundamental continuous time stochastic process. We have seen already in Section 7.3 that it is a Gaussian process of continuous sample functions and independent, stationary increments. In addition, it is a martingale of the type considered in Section 8.2 and has the strong Markov property of Section 8.3. Having all these beautiful properties allows for a rich mathematical theory. For example, many probabilistic computations involving the Brownian motion can be made explicit by solving partial differential equations. Further, the Brownian motion is the corner stone of diffusion theory and of stochastic integration. As such it is the most fundamental object in applications to and modeling of natural and man-made phenomena.

This chapter deals with some of the most interesting properties of the Brownian motion. Specifically, in Section 9.1 we use stopping time, Markov and martingale theory to study path properties of this process, focusing on passage times and running maxima. Expressing in Section 9.2 random walks and discrete time MGs as time-changed Brownian motion, we prove Donsker's celebrated invariance principle. It then provides fundamental results about these discrete time S.P.-s, such as the law of the iterated logarithm (in short LIL), and the martingale CLT. Finally, the fascinating aspects of the (lack of) regularity of the Brownian sample path are the focus of Section 9.3.

9.1. Brownian transformations, hitting times and maxima

We start with a few elementary path transformations under which the Wiener process of Definition 7.3.12 is invariant (see also Figure 2 illustrating its sample functions).

Exercise 9.1.1. *For $\{W_t, t \geq 0\}$ a standard Wiener process, show that the S.P. $\widetilde{W}_t^{(i)}$, $i = 1, \ldots, 6$ are also standard Wiener processes.*

(a) *(Symmetry)* $\widetilde{W}_t^{(1)} = -W_t$, $t \geq 0$.
(b) *(Time-homogeneity)* $\widetilde{W}_t^{(2)} = W_{T+t} - W_T$, $t \geq 0$ *with* $T > 0$ *a non-random constant.*
(c) *(Time-reversal)* $\widetilde{W}_t^{(3)} = W_T - W_{T-t}$, *for* $t \in [0, T]$, *with* $T > 0$ *a non-random constant.*
(d) *(Scaling)* $\widetilde{W}_t^{(4)} = \alpha^{-1/2} W_{\alpha t}$, $t \geq 0$, *with* $\alpha > 0$ *a non-random constant.*
(e) *(Time-inversion)* $\widetilde{W}_t^{(5)} = tW_{1/t}$ *for* $t > 0$ *and* $\widetilde{W}_0^{(5)} = 0$.
(f) *(Averaging)* $\widetilde{W}_t^{(6)} = \sum_{k=1}^n c_k W_t^{(k)}$, $t \geq 0$, *where* $W_t^{(k)}$ *are independent copies of the Wiener process and* c_k *non-random such that* $\sum_{k=1}^n c_k^2 = 1$.
(g) *Show that* $\widetilde{W}_t^{(2)}$ *and* $\widetilde{W}_t^{(3)}$ *are independent Wiener processes and evaluate* $q_t = \mathbf{P}_x(W_T > W_{T-t} > W_{T+t})$, *where* $t \in [0, T]$.

Remark. These invariance transformations are extensively used in the study of the Wiener process. As a token demonstration, note that since time-inversion maps $L_{a,b} = \sup\{t \geq 0 : W_t \notin (-at, bt)\}$ to the stopping time $\tau_{a,b}$ of Exercise 8.2.36, it follows that $L_{a,b}$ is a.s. finite and $\mathbf{P}(W_{L_{a,b}} = bL_{a,b}) = a/(a+b)$.

Recall Exercise 8.3.20 (or Exercise 8.3.22), that any Brownian Markov process $(W_t, \mathcal{F}_t)$ is a strong Markov process, yielding the following consequence of Corollary 8.3.14 (and of the identification of the Borel σ-algebra of $C([0,\infty))$ as the restriction of the cylindrical σ-algebra $\mathcal{B}^{[0,\infty)}$ to $C([0,\infty))$, see Exercise 7.2.9).

Corollary 9.1.2. *If $(W_t, \mathcal{F}_t)$ is a Brownian Markov process, then $\{W_t, t \geq 0\}$ is a homogeneous $\mathcal{F}_{t^+}$-Markov process and further, for any $s \geq 0$ and Borel measurable functional $h : C([0,\infty)) \mapsto \mathbb{R}$, almost surely*

$$\mathbf{E}[h(W.)|\mathcal{F}_{s^+}] = \mathbf{E}[h(W.)|\mathcal{F}_s]\,. \tag{9.1.1}$$

From this corollary and the Brownian time-inversion property we further deduce both *Blumenthal's 0-1 law* about the $\mathbf{P}_x$-triviality of the σ-algebra $\mathcal{F}^{\mathbf{W}}_{0^+}$ and its analog about the $\mathbf{P}_x$-triviality of the *tail σ-algebra* of the Wiener process (compare the latter with Kolmogorov's 0-1 law). To this end, we first extend the definition of the tail σ-algebra, as in Definition 1.4.9, to continuous time S.P.-s.

Definition 9.1.3. *Associate with any continuous time S.P. $\{X_t, t \geq 0\}$ the canonical future σ-algebras $\mathcal{T}^{\mathbf{X}}_t = \sigma(X_s, s \geq t)$, with the corresponding* tail σ-algebra *of the process being $\mathcal{T}^{\mathbf{X}} = \bigcap_{t\geq 0} \mathcal{T}^{\mathbf{X}}_t$.*

Proposition 9.1.4 (Blumenthal's 0-1 law)**.** *Let $\mathbf{P}_x$ denote the law of the Wiener process $\{W_t, t \geq 0\}$ starting at $W_0 = x$ (identifying $(\Omega, \mathcal{F}^{\mathbf{W}})$ with $C([0,\infty))$ and its Borel σ-algebra). Then, $\mathbf{P}_x(A) \in \{0,1\}$ for each $A \in \mathcal{F}^{\mathbf{W}}_{0^+}$ and $x \in \mathbb{R}$. Further, if $A \in \mathcal{T}^{\mathbf{W}}$ then either $\mathbf{P}_x(A) = 0$ for all x or $\mathbf{P}_x(A) = 1$ for all x.*

Proof. Applying Corollary 9.1.2 for the Wiener process starting at $W_0 = x$ and its canonical filtration, we have by the $\mathbf{P}_x$-triviality of $\mathcal{F}^{\mathbf{W}}_0$ that for each $A \in \mathcal{F}^{\mathbf{W}}_{0^+}$,

$$I_A = \mathbf{E}_x[I_A|\mathcal{F}^{\mathbf{W}}_{0^+}] = \mathbf{E}_x[I_A|\mathcal{F}^{\mathbf{W}}_0] = \mathbf{P}_x(A) \qquad \mathbf{P}_x - \text{a.s.}$$

Hence, $\mathbf{P}_x(A) \in \{0,1\}$. Proceeding to prove our second claim, set $X_0 = 0$ and $X_t = tW_{1/t}$ for $t > 0$, noting that $\{X_t, t \geq 0\}$ is a standard Wiener process (see part (e) of Exercise 9.1.1). Further, $\mathcal{T}^{\mathbf{W}}_t = \mathcal{F}^{\mathbf{X}}_{1/t}$ for any $t > 0$, hence

$$\mathcal{T}^{\mathbf{W}} = \bigcap_{t>0} \mathcal{T}^{\mathbf{W}}_t = \bigcap_{t>0} \mathcal{F}^{\mathbf{X}}_{1/t} = \mathcal{F}^{\mathbf{X}}_{0^+}\,.$$

Consequently, applying our first claim for the canonical filtration of the standard Wiener processes $\{X_t\}$ we see that $\mathbf{P}_0(A) \in \{0,1\}$ for any $A \in \mathcal{F}^{\mathbf{X}}_{0^+} = \mathcal{T}^{\mathbf{W}}$. Moreover, since $A \in \mathcal{T}^{\mathbf{W}}_1$, it is of the form $I_A = I_D \circ \theta_1$ for some $D \in \mathcal{F}^{\mathbf{W}}$, so by the tower and Markov properties,

$$\mathbf{P}_x(A) = \mathbf{E}_x[I_D \circ \theta_1(\omega(\cdot))] = \mathbf{E}_x[\mathbf{P}_{W_1}(D)] = \int p_1(x,y)\mathbf{P}_y(D)dy\,,$$

for the strictly positive Brownian transition kernel $p_1(x,y) = \exp(-(x-y)^2/2)/\sqrt{2\pi}$. If $\mathbf{P}_0(A) = 0$ then necessarily $\mathbf{P}_y(D) = 0$ for Lebesgue almost every y, hence also $\mathbf{P}_x(A) = 0$ for all $x \in \mathbb{R}$. Conversely, if $\mathbf{P}_0(A) = 1$ then $\mathbf{P}_0(A^c) = 0$ and with $A^c \in \mathcal{T}^{\mathbf{W}}$, by the preceding argument $1 - \mathbf{P}_x(A) = \mathbf{P}_x(A^c) = 0$ for all $x \in \mathbb{R}$. $\square$

Blumenthal's 0-1 law is very useful in determining properties of the Brownian sample function in the limits $t \downarrow 0$ and $t \uparrow \infty$. Here are few of its many consequences.

Corollary 9.1.5. *Let $\tau_{0^+} = \inf\{t \geq 0 : W_t > 0\}$, $\tau_{0^-} = \inf\{t \geq 0 : W_t < 0\}$ and $T_0 = \inf\{t > 0 : W_t = 0\}$. Then, $\mathbf{P}_0(\tau_{0^+} = 0) = \mathbf{P}_0(\tau_{0^-} = 0) = \mathbf{P}_0(T_0 = 0) = 1$ and w.p.1. the standard Wiener process changes sign infinitely many times in any time interval $[0, \varepsilon]$, $\varepsilon > 0$. Further, for any $x \in \mathbb{R}$, with $\mathbf{P}_x$-probability one,*

$$\limsup_{t\to\infty} \frac{1}{\sqrt{t}} W_t = \infty\,, \qquad \liminf_{t\to\infty} \frac{1}{\sqrt{t}} W_t = -\infty\,, \qquad W_{u_n} = 0 \text{ for some } u_n(\omega) \uparrow \infty\,.$$

Proof. Since $\mathbf{P}_0(\tau_{0^+} \leq t) \geq \mathbf{P}_0(W_t > 0) = 1/2$ for all $t > 0$, also $\mathbf{P}_0(\tau_{0^+} = 0) \geq 1/2$. Further, τ_{0^+} is an $\mathcal{F}_t^{\mathbf{W}}$-Markov time (see Proposition 8.1.15). Hence, $\{\tau_{0^+} = 0\} = \{\tau_{0^+} \leq 0\} \in \mathcal{F}_{0^+}^{\mathbf{W}}$ and from Blumenthal's 0-1 law it follows that $\mathbf{P}_0(\tau_{0^+} = 0) = 1$. By the symmetry property of the standard Wiener process (see part (a) of Exercise 9.1.1), also $\mathbf{P}_0(\tau_{0^-} = 0) = 1$. Combining these two facts we deduce that $\mathbf{P}_0$-a.s. there exist $t_n \downarrow 0$ and $s_n \downarrow 0$ such that $W_{t_n} > 0 > W_{s_n}$ for all n. By sample path continuity, this implies the existence of $u_n \downarrow 0$ such that $W_{u_n} = 0$ for all n. Hence, $\mathbf{P}_0(T_0 = 0) = 1$. As for the second claim, note that for any $r > 0$,

$$\mathbf{P}_0(W_n \geq r\sqrt{n} \text{ i.o.}) \geq \limsup_{n\to\infty} \mathbf{P}_0(W_n \geq r\sqrt{n}) = \mathbf{P}_0(W_1 \geq r) > 0$$

where the first inequality is due to Exercise 2.2.2 and the equality holds by the scaling property of $\{W_t\}$ (see part (d) of Exercise 9.1.1). Since $\{W_n \geq r\sqrt{n} \text{ i.o.}\} \in \mathcal{T}^{\mathbf{W}}$ we thus deduce from Blumenthal's 0-1 law that $\mathbf{P}_x(W_n \geq r\sqrt{n} \text{ i.o.}) = 1$ for any $x \in \mathbb{R}$. Considering $r_k \uparrow \infty$ this implies that $\limsup_{t\to\infty} W_t/\sqrt{t} = \infty$ with $\mathbf{P}_x$-probability one. Further, by the symmetry property of the standard Wiener process,

$$\mathbf{P}_0(W_n \leq -r\sqrt{n}, \text{ i.o.}) = \mathbf{P}_0(W_n \geq r\sqrt{n}, \text{ i.o.}) > 0\,,$$

so the preceding argument leads to $\liminf_{t\to\infty} W_t/\sqrt{t} = -\infty$ with $\mathbf{P}_x$-probability one. In particular, $\mathbf{P}_x$-a.s. there exist $t_n \uparrow \infty$ and $s_n \uparrow \infty$ such that $W_{t_n} > 0 > W_{s_n}$ which by sample path continuity implies the existence of $u_n \uparrow \infty$ such that $W_{u_n} = 0$ for all n. $\square$

Combining the strong Markov property of the Brownian Markov process and the independence of its increments, we deduce next that each a.s. finite Markov time τ is a *regeneration time* for this process, where it "starts afresh" independently of the path it took up to this (random) time.

Corollary 9.1.6. *If $(W_t, \mathcal{F}_t)$ is a Brownian Markov process and τ is an a.s. finite $\mathcal{F}_t$-Markov time, then the S.P. $\{W_{\tau+t} - W_\tau, t \geq 0\}$ is a standard Wiener process, which is independent of $\mathcal{F}_{\tau^+}$.*

Proof. With τ a.s. finite, $\mathcal{F}_t$-Markov time and $\{W_t, t \geq 0\}$ an $\mathcal{F}_t$-progressively measurable process, it follows that $B_t = W_{t+\tau} - W_\tau$ is a R.V. on our probability space and $\{B_t, t \geq 0\}$ is a well defined S.P. whose sample functions inherit the continuity of those of $\{W_t, t \geq 0\}$. Since the S.P. $\widetilde{W}_t = W_t - W_0$ has the f.d.d. hence the law of the *standard* Wiener process, fixing $h \in b\mathcal{B}^{[0,\infty)}$ and $\widetilde{h}(x(\cdot)) = h(x(\cdot) - x(0))$, the value of $g_{\widetilde{h}}(y) = \mathbf{E}_y[\widetilde{h}(W_\cdot)] = \mathbf{E}[h(\widetilde{W}_\cdot)]$ is independent of y. Consequently, fixing $A \in \mathcal{F}_{\tau^+}$, by the tower property and the strong Markov property (8.3.12) of the Brownian Markov process $(W_t, \mathcal{F}_t)$ we have that

$$\mathbf{E}[I_A h(B.)] = \mathbf{E}[I_A \widetilde{h}(W_{\tau+\cdot})] = \mathbf{E}[I_A g_{\widetilde{h}}(W_\tau)] = \mathbf{P}(A)\mathbf{E}[h(\widetilde{W}.)]\,.$$

In particular, considering $A = \Omega$ we deduce that the S.P. $\{B_t, t \geq 0\}$ has the f.d.d. and hence the law of the standard Wiener process $\{\widetilde{W}_t\}$. Further, recall Lemma 7.1.7 that for any $F \in \mathcal{F}^{\mathbf{B}}$, the indicator I_F is of the form $I_F = h(B.)$ for some $h \in b\mathcal{B}^{[0,\infty)}$, in which case by the preceding $\mathbf{P}(A \cap F) = \mathbf{P}(A)\mathbf{P}(F)$. Since this applies for any $F \in \mathcal{F}^{\mathbf{B}}$ and $A \in \mathcal{F}_{\tau^+}$ we have established the $\mathbf{P}$-independence of the two σ-algebras, namely, the stated independence of $\{B_t, t \geq 0\}$ and $\mathcal{F}_{\tau^+}$. $\square$

Beware that to get such a regeneration it is imperative to start with a Markov time τ. To convince yourself, solve the following exercise.

Exercise 9.1.7. *Suppose $\{W_t, t \geq 0\}$ is a standard Wiener process.*

(a) *Provide an example of a finite a.s. random variable $\tau \geq 0$ such that $\{W_{\tau+t} - W_\tau, t \geq 0\}$ does not have the law of a standard Brownian motion.*
(b) *Provide an example of a finite $\mathcal{F}^{\mathbf{W}}_t$-stopping time τ such that $[\tau]$ is not an $\mathcal{F}^{\mathbf{W}}_t$-stopping time.*

Combining Corollary 9.1.6 with the fact that w.p.1. $\tau_{0^+} = 0$, you are next to prove the somewhat surprising fact that w.p.1. a Brownian Markov process enters (b, ∞) as soon as it exits $(-\infty, b)$.

Exercise 9.1.8. *Let $\tau_{b^+} = \inf\{t \geq 0 : W_t > b\}$ for $b \geq 0$ and a Brownian Markov process $(W_t, \mathcal{F}_t)$.*

(a) *Show that $\mathbf{P}_0(\tau_b \neq \tau_{b^+}) = 0$.*
(b) *Suppose $W_0 = 0$ and a finite random-variable $H \geq 0$ is independent of $\mathcal{F}^{\mathbf{W}}$. Show that $\{\tau_H \neq \tau_{H^+}\} \in \mathcal{F}$ has probability zero.*

The strong Markov property of the Wiener process also provides the probability that starting at $x \in (c, d)$ it reaches level d before level c (i.e., the event $W_{\tau^{(0)}_{a,b}} = b$ of Exercise 8.2.36, with $b = d - x$ and $a = x - c$).

Exercise 9.1.9. *Consider the stopping time $\tau = \inf\{t \geq 0 : W_t \notin (c, d)\}$ for a Wiener process $\{W_t, t \geq 0\}$ starting at $x \in (c, d)$.*

(a) *Using the strong Markov property of W_t show that $u(x) = \mathbf{P}_x(W_\tau = d)$ is an* harmonic function, *namely, $u(x) = (u(x + r) + u(x - r))/2$ for any $c \leq x - r < x < x + r \leq d$, with boundary conditions $u(c) = 0$ and $u(d) = 1$.*
(b) *Check that $v(x) = (x - c)/(d - c)$ is an harmonic function satisfying the same boundary conditions as $u(x)$.*

Since boundary conditions at $x = c$ and $x = d$ uniquely determine the value of a harmonic function in (c, d) (a fact you do not need to prove), you thus showed that $\mathbf{P}_x(W_\tau = d) = (x - c)/(d - c)$.

We proceed to derive some of the many classical explicit formulas involving Brownian hitting times, starting with the celebrated *reflection principle*, which provides among other things the probability density functions of the *passage times* for a standard Wiener process and of their dual, the running maxima of this process.

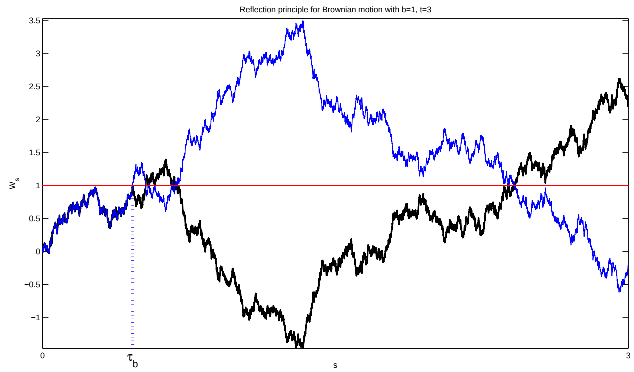

FIGURE 1. Illustration of the reflection principle for Brownian motion.

Proposition 9.1.10 (REFLECTION PRINCIPLE)**.** *With $\{W_t, t \geq 0\}$ the standard Wiener process, let $M_t = \sup_{s\in[0,t]} W_s$ denote its* running maxima *and $T_b = \inf\{t \geq 0 : W_t = b\}$ the corresponding* passage times. *Then, for any $t, b > 0$,*

$$\mathbf{P}(M_t \geq b) = \mathbf{P}(\tau_b \leq t) = \mathbf{P}(T_b \leq t) = 2\mathbf{P}(W_t \geq b)\,. \tag{9.1.2}$$

Remark. The reflection principle was stated by P. Lévy [**Lev39**] and first rigorously proved by Hunt [**Hun56**]. It is attributed to D. André [**And1887**] who solved the ballot problem of Exercise 5.5.30 by a similar symmetry argument (leading also to the reflection principle for symmetric random walks, as in Exercise 6.1.19).

PROOF. Recall Proposition 8.1.15 that τ_b is a stopping time for $\mathcal{F}_t^{\mathbf{W}}$. Further, since $b > 0 = W_0$ and $s \mapsto W_s$ is continuous, clearly $\tau_b = T_b$ and $W_{T_b} = b$ whenever T_b is finite. Heuristically, given that $T_b = s < u$ we have that $W_s = b$ and by reflection symmetry of the Brownian motion, expect the conditional law of $W_u - W_s$ to retain its symmetry around zero, as illustrated in Figure 1. This of course leads to the prediction that for any $u, b > 0$,

$$\mathbf{P}(T_b < u, W_u > b) = \frac{1}{2}\mathbf{P}(T_b < u)\,. \tag{9.1.3}$$

With $W_0 = 0$, by sample path continuity $\{W_u > b\} \subseteq \{T_b < u\}$, so the preceding prediction implies that

$$\mathbf{P}(T_b < u) = 2\mathbf{P}(T_b < u, W_u > b) = 2\mathbf{P}(W_u > b)\,.$$

The supremum $M_t(\omega)$ of the continuous function $s \mapsto W_s(\omega)$ over the compact interval $[0,t]$ is attained at some $s \in [0,t]$, hence the identity $\{M_t \geq b\} = \{\tau_b \leq t\}$ holds for all $t, b > 0$. Thus, considering $u \downarrow t > 0$ leads in view of the continuity of $(u, b) \mapsto \mathbf{P}(W_u > b)$ to the statement (9.1.2) of the proposition. Turning to rigorously prove (9.1.3), we rely on the strong Markov property of the standard Wiener process for the $\mathcal{F}_t^{\mathbf{W}}$-stopping time T_b and the functional $h(s, x(\cdot)) = I_A(s, x(\cdot))$, where $A = \{(s, x(\cdot)) : x(\cdot) \in C(\mathbb{R}_+), s \in [0, u) \text{ and } x(u - s) > b\}$. To this end, note that $F_{y,a,a'} = \{x \in C([0,\infty)) : x(u - s) \geq y \text{ for all } s \in [a, a']\}$ is closed (with respect to uniform convergence on compact subsets of $[0,\infty)$), and $x(u - s) > b$ for some $s \in [0, u)$ if and only if $x(\cdot) \in F_{b_k,q,q'}$ for some $b_k = b + 1/k$, $k \geq 1$ and $q < q' \in \mathbb{Q}_u^{(2)}$. So, A is the countable union of closed sets $[q, q'] \times F_{b_k,q,q'}$, hence

Borel measurable on $[0,\infty) \times C([0,\infty))$. Next recall that by the definition of the set A,

$$g_h(s,b) = \mathbf{E}_b[I_A(s,W_\cdot)] = I_{[0,u)}(s)\mathbf{P}_b(W_{u-s} > b) = \frac{1}{2}I_{[0,u)}(s)\,.$$

Further, $h(s,x(s+\cdot)) = I_{[0,u)}(s)I_{x(u)>b}$ and $W_{T_b} = b$ whenever T_b is finite, so taking the expectation of (8.3.13) yields (for our choices of $h(\cdot,\cdot)$ and τ), the identity,

$$\begin{aligned}\mathbf{E}[I_{\{T_b<u\}}I_{\{W_u>b\}}] &= \mathbf{E}[h(T_b,W_{T_b+\cdot})] = \mathbf{E}[g_h(T_b,W_{T_b})] \\ &= \mathbf{E}[g_h(T_b,b)] = \frac{1}{2}\mathbf{E}[I_{\{T_b<u\}}]\,,\end{aligned}$$

which is precisely (9.1.3). □

Since $t^{-1/2}W_t \overset{\mathcal{D}}{=} G$, a standard normal variable of continuous distribution function, we deduce from the reflection principle that the distribution functions of T_b and M_t are continuous and such that $F_{T_b}(t) = 1 - F_{M_t}(b) = 2(1 - F_G(b/\sqrt{t}))$. In particular, $\mathbf{P}(T_b > t) \to 0$ as $t \to \infty$, hence T_b is a.s. finite. We further have the corresponding explicit *probability density functions* on $[0,\infty)$,

$$f_{T_b}(t) = \frac{\partial}{\partial t}F_{T_b}(t) = \frac{b}{\sqrt{2\pi t^3}}e^{-\frac{b^2}{2t}}\,, \tag{9.1.4}$$

$$f_{M_t}(b) = \frac{\partial}{\partial b}F_{M_t}(b) = \frac{2}{\sqrt{2\pi t}}e^{-\frac{b^2}{2t}}\,. \tag{9.1.5}$$

Remark. From the preceding formula for the density of $T_b = \tau_b$ you can easily check that it has infinite expected value, in contrast with the exit times $\tau_{a,b}$ of bounded intervals $(-a,b)$, which have finite moments (see part (c) of Exercise 8.2.36 for finiteness of the second moment and note that the same method extends to all moments). Recall that in part (b) of Exercise 8.2.35 you have already found that the Laplace transform of the density of T_b is

$$L_{f_{T_b}}(s) = \int_0^\infty e^{-st}f_{T_b}(t)dt = e^{-\sqrt{2s}b}$$

(and for inverting Laplace transforms, see Exercise 2.2.15). Further, using the density of passage times, you can now derive the well-known arc-sine law for the last exit of the Brownian motion from zero by time one.

Exercise 9.1.11. *For the standard Wiener process $\{W_t\}$ and any $t > 0$, consider the time $L_t = \sup\{s \in [0,t] : W_s = 0\}$ of last exit from zero by t, and the Markov time $R_t = \inf\{s > t : W_s = 0\}$ of first return to zero after t.*

(a) *Verify that $\mathbf{P}_x(T_y > u) = \mathbf{P}(T_{|y-x|} > u)$ for any $x,y \in \mathbb{R}$, and with $p_t(x,y)$ denoting the Brownian transition probability kernel, show that for $u > 0$ and $0 < u < t$, respectively,*

$$\mathbf{P}(R_t > t+u) = \int_{-\infty}^{\infty} p_t(0,y)\mathbf{P}(T_{|y|} > u)dy\,,$$

$$\mathbf{P}(L_t \le u) = \int_{-\infty}^{\infty} p_u(0,y)\mathbf{P}(T_{|y|} > t-u)dy\,.$$

(b) *Deduce from (9.1.4) that the probability density function of $R_t - t$ is $f_{R_t-t}(u) = \sqrt{t}/(\pi\sqrt{u}(t+u))$.*
Hint: *Express $\partial\mathbf{P}(R_t > t+u)/\partial u$ as one integral over $y \in \mathbb{R}$, then change variables to $z^2 = y^2(1/u + 1/t)$.*

(c) *Show that* L_t *has the* arc-sine law $\mathbf{P}(L_t \leq u) = (2/\pi)\arcsin(\sqrt{u/t})$ *and hence the density* $f_{L_t}(u) = 1/(\pi\sqrt{u(t-u)})$ *on* $[0,t]$.

(d) *Find the joint probability density function of* (L_t, R_t).

Remark. Knowing the law of L_t is quite useful, for $\{L_t > u\}$ is just the event $\{W_s = 0 \text{ for some } s \in (u,t]\}$. You have encountered the arc-sine law in Exercise 3.2.16 (where you proved the discrete reflection principle for the path of the symmetric SRW). Indeed, as shown in Section 9.2 by Donsker's invariance principle, these two arc-sine laws are equivalent.

Here are a few additional results about passage times and running maxima.

Exercise 9.1.12. *Generalizing the proof of (9.1.3), deduce that for a standard Wiener process, any* $u > 0$ *and* $a_1 < a_2 \leq b$,

$$\mathbf{P}(T_b < u, a_1 < W_u < a_2) = \mathbf{P}(2b - a_2 < W_u < 2b - a_1)\,, \tag{9.1.6}$$

and conclude that the joint density of (M_t, W_t) *is*

$$f_{W_t,M_t}(a,b) = \frac{2(2b-a)}{\sqrt{2\pi t^3}} e^{-\frac{(2b-a)^2}{2t}}\,, \tag{9.1.7}$$

for $b \geq a \vee 0$ *and zero otherwise.*

Exercise 9.1.13. *Let* $\underline{W}(t) = (W_1(t), W_2(t))$ *denote the two-dimensional Brownian motion of Definition 8.2.37, starting at a non-random* $\underline{W}(0) = (x_1, x_2)$ *with* $x_1 > 0$ *and* $x_2 > 0$.

(a) *Find the density of* $\tau = \inf\{t \geq 0 : W_1(t) = 0 \text{ or } W_2(t) = 0\}$.

(b) *Find the joint density of* $(\tau, W_1(\tau), W_2(\tau))$ *with respect to Lebesgue measure on* $\{(t,x,y) \in (0,\infty)^3 : x = 0 \text{ or } y = 0\}$.

Hint: *The identity (9.1.6) might be handy here.*

Exercise 9.1.14. *Consider a Brownian Markov process* $(W_t, \mathcal{F}_t)$ *with* $W_0 \geq 0$ *and* $p_t(x,B) = \mathbf{P}_x(W_t \in B)$ *its Brownian semi-group of transition probabilities.*

(a) *Show that* $(W_{t\wedge T_0}, \mathcal{F}_t)$ *is a homogeneous Markov process on* $[0,\infty)$ *whose transition probabilities are:* $p_{-,t}(0,\{0\}) = 1$, *and if* $x > 0$ *then* $p_{-,t}(x,B) = p_t(x,B) - p_t(x,-B)$ *for* $B \subseteq (0,\infty)$, *while* $p_{-,t}(x,\{0\}) = 2p_t(x,(-\infty,0])$.

(b) *Show that* $(|W_t|, \mathcal{F}_t)$ *is a homogeneous Markov process on* $[0,\infty)$ *whose transition probabilities are* $p_{+,t}(x,B) = p_t(x,B) + p_t(x,-B)$ *(for* $x \geq 0$ *and* $B \subseteq [0,\infty)$*).*

Remark. We call $(W_{t\wedge T_0}, \mathcal{F}_t)$ the *Brownian motion absorbed at zero* and $(|W_t|, \mathcal{F}_t)$ the *reflected Brownian motion.* These are the simplest possible ways of constraining the Brownian motion to have state space $[0,\infty)$.

Exercise 9.1.15. *The Brownian Markov process* $(W_t, \mathcal{F}_t)$ *starts at* $W_0 = 0$.

(a) *Show that* $Y_t = M_t - W_t$ *is an* $\mathcal{F}_t$*-Markov process, of the same transition probabilities* $\{p_{+,t}, t \geq 0\}$ *on* $[0,\infty)$ *as the reflected Brownian motion.*

(b) *Deduce that* $\{Y_t, t \geq 0\}$ *has the same law as the reflected Brownian motion.*

Exercise 9.1.16. *For a Brownian Markov process* $(W_t, \mathcal{F}_t)$ *starting at* $W_0 = 0$, *show that* $L_{\downarrow,t} = \sup\{s \in [0,t] : W_s = M_t\}$ *has the same arc-sine law as* L_t.

Solving the next exercise you first show that $\{T_b, b \geq 0\}$ is a strictly increasing, left-continuous process, whose sample path is a.s. purely discontinuous.

Exercise 9.1.17. *Consider the passage times $\{T_b, b \geq 0\}$ for a Brownian Markov process $(W_t, \mathcal{F}_t)$ starting at $W_0 = 0$.*

(a) *Show that $b \mapsto T_b(\omega)$, is left-continuous, strictly increasing and w.p.1.* purely discontinuous *(i.e. there is no interval of positive length on which $b \mapsto T_b$ is continuous).*
Hint: *Check that $\mathbf{P}(t \mapsto M_t$ is strictly increasing on $[0, \epsilon]) = 0$ for any $\epsilon > 0$, and relying on the strong Markov property of W_t deduce that w.p.1. $b \mapsto T_b$ must be purely discontinuous.*

(b) *Show that for any $h \in b\mathcal{B}$ and $0 \leq b < c$,*
$$\mathbf{E}[h(T_c - T_b)|\mathcal{F}_{T_b^+}] = \mathbf{E}_b[h(T_c)] = \mathbf{E}_0[h(T_{c-b})]\,,$$

(c) *Deduce that $\{T_b, b \geq 0\}$ is a S.P. of stationary, non-negative independent increments, whose Markov semi-group has the transition probability kernel*
$$\widehat{q}_t(x, y) = \frac{t}{\sqrt{2\pi(y-x)_+^3}} e^{-\frac{t^2}{2(y-x)_+}}\,,$$
corresponding to the one-sided $1/2$-stable *density of (9.1.4).*

(d) *Show that $\{\tau_{b^+}, b \geq 0\}$ of Exercise 9.1.8 is a right-continuous modification of $\{T_b, b \geq 0\}$, hence a strong Markov process of same transition probabilities.*

(e) *Show that $T_c \overset{\mathcal{D}}{=} c^2 T_1$ for any $c \in \mathbb{R}$.*

Exercise 9.1.18. *Suppose that for some $b > 0$ fixed, $\{\xi_k\}$ are i.i.d. each having the probability density function (9.1.4) of T_b.*

(a) *Show that $n^{-2}\sum_{k=1}^n \xi_k \overset{\mathcal{D}}{=} T_b$ (which is why we say that the law of T_b is α-stable for $\alpha = 1/2$).*

(b) *Show that $\mathbf{P}(n^{-2}\max_{k=1}^n \xi_k \leq y) \to \exp(-b\sqrt{2/(\pi y)})$ for all $y \geq 0$ (compare with part (b) of Exercise 3.2.13).*

Exercise 9.1.19. *Consider a standard Wiener process $\{W_t, t \geq 0\}$.*

(a) *Fixing $b, t > 0$ let $\theta_{b,t} = \inf\{s \geq t : V_s \geq b\}$ for $V_s = |W_s|/\sqrt{s}$. Check that $\theta_{b,t} \overset{\mathcal{D}}{=} t\theta_{b,1}$, then show that for $b < 1$,*
$$\mathbf{E}\theta_{b,1} = \frac{1}{1-b^2}\mathbf{E}[(V_1^2 - b^2)I_{\{V_1 \geq b\}}]\,,$$
whereas $\mathbf{E}\theta_{b,1} = \infty$ in case $b \geq 1$.

(b) *Considering now $V_s = \int_0^s \exp[c(W_s - W_u) - c^2(s-u)/2]du$ for $c \in \mathbb{R}$ non-random, verify that $V_s - s$ is a martingale and deduce that in this case $\mathbf{E}\theta_{b,0} = b$ for any $b > 0$.*

9.2. Weak convergence and invariance principles

Consider the linearly interpolated, time-space rescaled random walk $\widehat{S}_n(t) = n^{-1/2}S(nt)$ (as depicted in Figure 2, for the symmetric SRW), where

$$S(t) = \sum_{k=1}^{[t]} \xi_k + (t - [t])\xi_{[t]+1}\,, \tag{9.2.1}$$

and $\{\xi_k\}$ are i.i.d. Recall Exercise 3.5.18 that by the CLT, if $\mathbf{E}\xi_1 = 0$ and $\mathbf{E}\xi_1^2 = 1$,

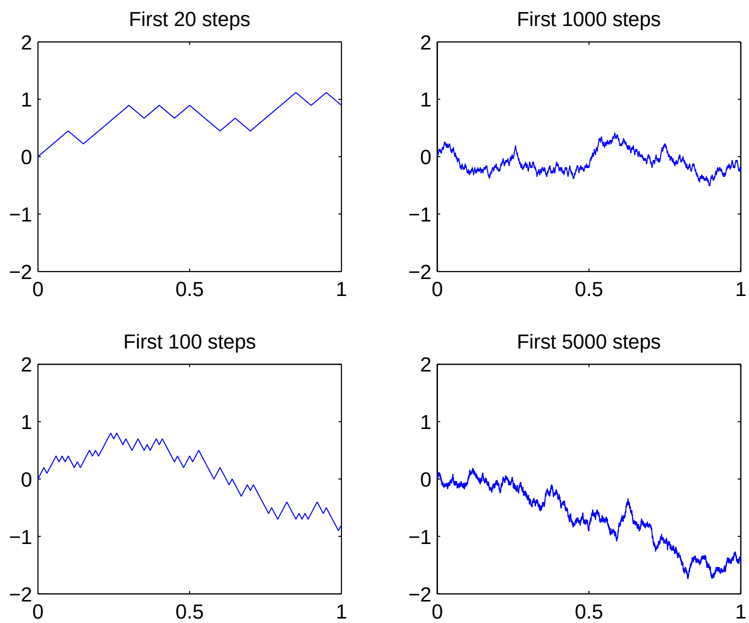

FIGURE 2. Scaled SRW for different values of n.

then as $n \to \infty$ the f.d.d. of the S.P. $\widehat{S}_n(\cdot)$ of continuous sample path, converge weakly to those of the standard Wiener process. Since f.d.d. uniquely determine the law of a S.P. it is thus natural to expect also to have the stronger, convergence in distribution, as defined next.

Definition 9.2.1. *We say that S.P. $\{X_n(t), t \geq 0\}$ of continuous sample functions* converge in distribution *to a S.P. $\{X_\infty(t), t \geq 0\}$, denoted $X_n(\cdot) \xrightarrow{\mathcal{D}} X_\infty(\cdot)$, if the corresponding laws converge weakly in the topological space $\mathbb{S}$ consisting of $C([0,\infty))$ equipped with the topology of uniform convergence on compact subsets of $[0,\infty)$. That is, if $g(X_n(\cdot)) \xrightarrow{\mathcal{D}} g(X_\infty(\cdot))$ whenever $g : C([0,\infty)) \mapsto \mathbb{R}$ Borel measurable, is such that w.p.1. the sample function of $X_\infty(\cdot)$ is not in the set $\mathbf{D}_g$ of points of discontinuity of g (with respect to uniform convergence on compact subsets of $[0,\infty)$).*

As we state now and prove in the sequel, such *functional* CLT, also known as Donsker's *invariance principle*, indeed holds.

Theorem 9.2.2 (DONSKER'S INVARIANCE PRINCIPLE)**.** *If $\{\xi_k\}$ are i.i.d. with $\mathbf{E}\xi_1 = 0$ and $\mathbf{E}\xi_1^2 = 1$, then for $S(\cdot)$ of (9.2.1), the S.P. $\widehat{S}_n(\cdot) = n^{-1/2}S(n\cdot)$ converge in distribution, as $n \to \infty$, to the standard Wiener process.*

Remark. The preceding theorem is called an *invariance principle* because the limiting process does not depend on the law of the summands $\{\xi_k\}$ of the random walk. However, the condition $\mathbf{E}\xi_1^2 < \infty$ is almost necessary for the $n^{-1/2}$ scaling and for having a Brownian limit process. Indeed, note Remark 3.1.13 that both fail as soon as $\mathbf{E}|\xi_1|^\alpha = \infty$ for some $0 < \alpha < 2$.

Since $h(x(\cdot)) = f(x(t_1), \ldots, x(t_k))$ is continuous and bounded on $C([0,\infty))$ for any $f \in C_b(\mathbb{R}^k)$ and each finite subset $\{t_1, \ldots, t_k\}$ of $[0,\infty)$, convergence in distribution of S.P. of continuous sample path implies the *weak convergence* of their f.d.d. But, beware that the convergence of f.d.d. does not necessarily imply convergence in distribution, even for S.P. of continuous sample functions.

Exercise 9.2.3. *Give a counter-example to show that weak convergence of the f.d.d. of S.P. $\{X_n(\cdot)\}$ of continuous sample functions to those of S.P. $\{X_\infty(\cdot)\}$ of continuous sample functions, does not imply that $X_n(\cdot) \xrightarrow{\mathcal{D}} X_\infty(\cdot)$.*
Hint: Try $X_n(t) = nt\mathbf{1}_{[0,1/n]}(t) + (2-nt)\mathbf{1}_{(1/n,2/n]}(t)$.

Nevertheless, with $\mathbb{S} = (C([0,\infty),\rho)$ a complete, separable metric space (c.f. Exercise 7.2.9), we have the following useful partial converse as an immediate consequence of Prohorov's theorem.

Proposition 9.2.4. *If the laws of S.P. $\{X_n(\cdot)\}$ of continuous sample functions are uniformly tight in $C([0,\infty))$ and for $n \to \infty$ the f.d.d. of $\{X_n(\cdot)\}$ converge weakly to the f.d.d. of $\{X_\infty(\cdot)\}$, then $X_n(\cdot) \xrightarrow{\mathcal{D}} X_\infty(\cdot)$.*

Proof. Recall part (e) of Theorem 3.5.2, that by the Portmanteau theorem $X_n(\cdot) \xrightarrow{\mathcal{D}} X_\infty(\cdot)$ as in Definition 9.2.1, if and only if the corresponding laws $\nu_n = \mathcal{P}_{\mathbf{X_n}}$ converge weakly on the metric space $\mathbb{S} = C([0,\infty))$ (and its Borel σ-algebra). That is, if and only if $\mathbf{E}h(X_n(\cdot)) \to \mathbf{E}h(X_\infty(\cdot))$ for each h continuous and bounded on $\mathbb{S}$ (also denoted by $\nu_n \stackrel{w}{\Rightarrow} \nu_\infty$, see Definition 3.2.17). Let $\{\nu_n^{(m)}\}$ be a subsequence of $\{\nu_n\}$. Since $\{\nu_n\}$ is uniformly tight, so is $\{\nu_n^{(m)}\}$. Thus, by Prohorov's theorem, there exists a further sub-subsequence $\{\nu_n^{(m_k)}\}$ such that $\nu_n^{(m_k)}$ converges weakly to a probability measure $\widetilde{\nu}_\infty$ on $\mathbb{S}$. Recall Proposition 7.1.8 that the f.d.d. uniquely determine the law of S.P. of continuous sample functions. Hence, from the assumed convergence of f.d.d. of $\{X_n(\cdot)\}$ to those of $\{X_\infty(\cdot)\}$, we deduce that $\widetilde{\nu}_\infty = \mathcal{P}_{\mathbf{X}_\infty} = \nu_\infty$. Consequently, $\mathbf{E}h(X_n^{(m_k)}(\cdot)) \to \mathbf{E}h(X_\infty(\cdot))$ for each $h \in C_b([0,\infty))$ (see Exercise 7.2.9). Fixing $h \in C_b([0,\infty))$ note that we have just shown that every subsequence $y_n^{(m)}$ of the sequence $y_n = \mathbf{E}h(X_n(\cdot))$ has a further sub-subsequence $y_n^{(m_k)}$ that converges to y_∞. Hence, we deduce by Lemma 2.2.11 that $y_n \to y_\infty$. Since this holds for all $h \in C_b([0,\infty))$, we conclude that $X_n(\cdot) \xrightarrow{\mathcal{D}} X_\infty(\cdot)$. □

Having Proposition 9.2.4 and the convergence of f.d.d. of $\widehat{S}_n(\cdot)$, Donsker's invariance principle is a consequence of the uniform tightness in $\mathbb{S}$ of the laws of these S.P.-s. In view of Definition 3.2.31, we prove this uniform tightness by exhibiting compact sets $\mathbb{K}_\ell$ such that $\sup_n \mathbf{P}(\widehat{S}_n \notin \mathbb{K}_\ell) \to 0$ as $\ell \to \infty$. To this end, recall the following classical result of functional analysis (for a proof see [**KaS97**, Theorem 2.4.9] or the more general version provided in [**Dud89**, Theorem 2.4.7]).

Theorem 9.2.5 (Arzelà-Ascoli theorem). *A set $\mathbb{K} \subset C([0,\infty))$ has compact closure with respect to uniform convergence on compact intervals, if and only if $\sup_{x\in\mathbb{K}} |x(0)|$ is finite and for $t > 0$ fixed, $\sup_{x\in\mathbb{K}} \mathsf{osc}_{t,\delta}(x(\cdot)) \to 0$ as $\delta \downarrow 0$, where*

$$\mathsf{osc}_{t,\delta}(x(\cdot)) = \sup_{0\le h\le\delta} \sup_{0\le s\le s+h\le t} |x(s+h) - x(s)|\,, \tag{9.2.2}$$

is just the maximal absolute increment of $x(\cdot)$ over all pairs of times $[0,t]$ which are within distance δ of each other.

The Arzelà-Ascoli theorem suggests the following strategy for proving uniform tightness.

Exercise 9.2.6. *Let $\mathbb{S}$ denote the set $C([0,\infty))$ equipped with the topology of uniform convergence on compact intervals, and consider its subsets $F_{r,\delta} = \{x(\cdot) : x(0) = 0, \mathsf{osc}_{r,\delta}(x(\cdot)) \le 1/r\}$ for $\delta > 0$ and integer $r \ge 1$.*

(a) *Verify that the functional $x(\cdot) \mapsto \mathsf{osc}_{t,\delta}(x(\cdot))$ is continuous on $\mathbb{S}$ per fixed t and δ and further that per $x(\cdot)$ fixed, the function $\mathsf{osc}_{t,\delta}(x(\cdot))$ is non-decreasing in t and in δ. Deduce that $F_{r,\delta}$ are closed sets and for any $\delta_r \downarrow 0$, the intersection $\cap_r F_{r,\delta_r}$ is a compact subset of $\mathbb{S}$.*

(b) *Show that if S.P.-s $\{X_n(t), t \ge 0\}$ of continuous sample functions are such that $X_n(0) = 0$ for all n and for any $r \ge 1$,*

$$\lim_{\delta \downarrow 0} \sup_{n \ge 1} \mathbf{P}(\mathsf{osc}_{r,\delta}(X_n(\cdot)) > r^{-1}) = 0\,,$$

then the corresponding laws are uniformly tight in $\mathbb{S}$.
Hint: *Let $\mathbb{K}_\ell = \cap_r F_{r,\delta_r}$ with $\delta_r \downarrow 0$ such that $\mathbf{P}(X_n \notin F_{r,\delta_r}) \le 2^{-\ell-r}$.*

Since $\widehat{S}_n(\cdot) = n^{-1/2} S(n\cdot)$ and $S(0) = 0$, by the preceding exercise the uniform tightness of the laws of $\widehat{S}_n(\cdot)$, and hence Donsker's invariance principle, is an immediate consequence of the following bound.

Proposition 9.2.7. *If $\{\xi_k\}$ are i.i.d. with $\mathbf{E}\xi_1 = 0$ and $\mathbf{E}\xi_1^2$ finite, then*

$$\lim_{\delta \downarrow 0} \sup_{n \ge 1} \mathbf{P}(\mathsf{osc}_{nr,n\delta}(S(\cdot)) > r^{-1}\sqrt{n}) = 0\,,$$

for $S(t) = \sum_{k=1}^{[t]} \xi_k + (t - [t])\xi_{[t]+1}$ and any integer $r \ge 1$.

Proof. Fixing $r \ge 1$, let $q_{n,\delta} = \mathbf{P}(\mathsf{osc}_{nr,n\delta}(S(\cdot)) > r^{-1}\sqrt{n})$. Since $t \mapsto S(t)$ is uniformly continuous on compacts, $\mathsf{osc}_{nr,n\delta}(S(\cdot))(\omega) \downarrow 0$ when $\delta \downarrow 0$ (for each $\omega \in \Omega$). Consequently, $q_{n,\delta} \downarrow 0$ for each fixed n, hence uniformly over $n \le n_0$ and any fixed n_0. With $\delta \mapsto q_{n,\delta}$ non-decreasing, this implies that $b = b(n_0) = \inf_{\delta > 0} \sup_{n \ge n_0} q_{n,\delta}$ is independent of n_0, hence $b(1) = 0$ provided $\inf_{n_0 \ge 1} b(n_0) = \lim_{\delta \to 0} \limsup_{k \to \infty} q_{k,\delta} = 0$. To show the latter, observe that since the piecewise linear $S(t)$ changes slope only at integer values of t,

$$\mathsf{osc}_{kr,k\delta}(S(\cdot)) \le \mathsf{osc}_{kr,m}(S(\cdot)) \le M_{m,\ell}\,,$$

for $m = [k\delta] + 1$, $\ell = rk/m$ and

$$M_{m,\ell} = \max_{\substack{1 \le i \le m \\ 0 \le j \le \ell m - 1}} |S(i+j) - S(j)|\,. \tag{9.2.3}$$

Thus, for any $\delta > 0$,

$$\limsup_{k \to \infty} q_{k,\delta} \le \limsup_{m \to \infty} \mathbf{P}(M_{m,\ell(v)} > v\sqrt{m})\,,$$

where $v = r^{-1}\sqrt{k/m} \to 1/(r\sqrt{\delta})$ as $k \to \infty$ and $\ell(v) = r^3 v^2$. Since $v \to \infty$ when $\delta \downarrow 0$, we complete the proof by appealing to part (c) of Exercise 9.2.8. $\square$

As you have just seen, the key to the proof of Proposition 9.2.7 is the following bound on maximal fluctuations of increments of the random walk.

Exercise 9.2.8. *Suppose $S_m = \sum_{k=1}^m \xi_k$ for i.i.d. ξ_k such that $\mathbf{E}\xi_1 = 0$ and $\mathbf{E}\xi_1^2 = 1$. For integers $m, \ell \ge 1$, let $S(m) = S_m$ and $M_{m,\ell}$ be as in (9.2.3), with $M_{m,0} = \max_{i=1}^m |S_i|$.*

(a) *Show that for any $m \geq 1$ and $t \geq 0$,*

$$\mathbf{P}(M_{m,0} \geq t + \sqrt{2m}) \leq 2\mathbf{P}(|S_m| \geq t)\,.$$

Hint: *Use Ottaviani's inequality (see part (a) of Exercise 5.2.16).*

(b) *Show that for any $m, \ell \geq 1$ and $x \geq 0$,*

$$\mathbf{P}(M_{m,\ell} > 2x) \leq \ell\mathbf{P}(M_{m,1} > 2x) \leq \ell\mathbf{P}(M_{2m,0} > x)\,.$$

(c) *Deduce that if $v^{-2}\log \ell(v) \to 0$, then*

$$\limsup_{v\to\infty} \limsup_{m\to\infty} \mathbf{P}(M_{m,\ell(v)} > v\sqrt{m}) = 0\,.$$

Hint: *Recall that $m^{-1/2}S_m \xrightarrow{\mathcal{D}} G$ by the* CLT.

Applying Donsker's invariance principle, you can induce limiting results for random walks out of the corresponding facts about the standard Brownian motion, which we have found already in Subsection 9.1.

Example 9.2.9. *Recall the running maxima $M_t = \sup_{s\in[0,t]} W_s$, whose density we got in (9.1.5) out of the reflection principle. Since $h_0(x(\cdot)) = \sup\{x(s) : s \in [0,1]\}$ is continuous with respect to uniform convergence on $C([0,1])$, we have from Donsker's invariance principle that as $n \to \infty$,*

$$h_0(\widehat{S}_n) = \frac{1}{\sqrt{n}} \max_{k=0}^{n} S_k \xrightarrow{\mathcal{D}} M_1$$

(where we have used the fact that the maximum of the linearly interpolated function $S(t)$ must be obtained at some integer value of t). The functions $h_\ell(x(\cdot)) = \int_0^1 x(s)^\ell ds$ for $\ell = 1, 2, \ldots$ are also continuous on $C([0,1])$, so by same reasoning,

$$h_\ell(\widehat{S}_n) = n^{-(1+\ell/2)} \int_0^n S(u)^\ell du \xrightarrow{\mathcal{D}} \int_0^1 (W_u)^\ell du\,.$$

Similar limits can be obtained by considering $h_\ell(|x(\cdot)|)$.

Exercise 9.2.10.

(a) *Building on Example 9.2.9, show that for any integer $\ell \geq 1$,*

$$n^{-(1+\ell/2)} \sum_{k=1}^{n} (S_k)^\ell \xrightarrow{\mathcal{D}} \int_0^1 (W_u)^\ell du\,,$$

as soon as $\mathbf{E}\xi_1 = 0$ and $\mathbf{E}\xi_1^2 = 1$ (i.e. there is no need to assume finiteness of the ℓ-th moment of ξ_1), and in case $\ell = 1$ the limit law is merely a normal of zero mean and variance $1/3$.

(b) *The cardinality of the set $\{S_0, \ldots, S_n\}$ is called the* range *of the walk by time n and denoted rng_n. Show that for the symmetric* SRW *on $\mathbb{Z}$,*

$$n^{-1/2}\,\mathsf{rng}_n \xrightarrow{\mathcal{D}} \sup_{s\leq 1} W_s - \inf_{s\leq 1} W_s\,.$$

We continue in the spirit of Example 9.2.9, except for dealing with functionals that are no longer continuous throughout $C([0,\infty))$.

Example 9.2.11. *Let $\widehat{\tau}_n = n^{-1}\inf\{k \geq 1 : S_k \geq \sqrt{n}\}$. As shown in Exercise 9.2.12, considering the function $g_1(x(\cdot)) = \inf\{t \geq 0 : x(t) \geq 1\}$ we find that $\widehat{\tau}_n \xrightarrow{\mathcal{D}} T_1$ as $n \to \infty$, where the density of $T_1 = \inf\{t \geq 0 : W_t \geq 1\}$ is given in (9.1.4).*

Similarly, let $A_t(b) = \int_0^t I_{W_s>b} ds$ *denote the* occupation time *of* $B = (b, \infty)$ *by the standard Brownian motion up to time* t. *Then, considering* $g(x(\cdot), B) = \int_0^1 I_{\{x(s)\in B\}} ds$, *we find that for* $n \to \infty$ *and* $b \in \mathbb{R}$ *fixed,*

$$\widehat{A}_n(b) = \frac{1}{n} \sum_{k=1}^n I_{\{S_k > b\sqrt{n}\}} \xrightarrow{\mathcal{D}} A_1(b)\,, \tag{9.2.4}$$

as you are also to justify upon solving Exercise 9.2.12. Of particular note is the case of $b = 0$, *where Lévy's* arc-sine law *tells us that* $A_t(0) \overset{\mathcal{D}}{=} L_t$ *of Exercise 9.1.11 (as shown for example in* [**KaS97**, Proposition 4.4.11]*).*

Recall the arc-sine limiting law of Exercise 3.2.16 for $n^{-1} \sup\{\ell \le n : S_{\ell-1}S_\ell \le 0\}$ *in case of the symmetric* SRW. *In view of Exercise 9.1.11, working with* $g_0(x(\cdot)) = \sup\{s \in [0,1] : x(s) = 0\}$ *one can extend the validity of this limit law to any random walk with increments of mean zero and variance one (c.f.* [**Dur10**, Example 8.6.3]*).*

Exercise 9.2.12.

(a) *Let* $g_{1^+}(x(\cdot)) = \inf\{t \ge 0 : x(t) > 1\}$. *Show that* $\mathbf{P}(W_\cdot \in \mathbb{G}) = 1$ *for the subset* $\mathbb{G} = \{x(\cdot) : x(0) = 0$ *and* $g_1(x(\cdot)) = g_{1^+}(x(\cdot)) < \infty\}$ *of* $C([0,\infty))$, *and that* $g_1(x_n(\cdot)) \to g_1(x(\cdot))$ *for any sequence* $\{x_n(\cdot)\} \subseteq C([0,\infty))$ *which converges uniformly on compacts to* $x(\cdot) \in \mathbb{G}$. *Further, show that* $\widehat{\tau}_n - n^{-1} \le g_1(\widehat{S}_n) \le \widehat{\tau}_n$ *and deduce that* $\widehat{\tau}_n \xrightarrow{\mathcal{D}} T_1$.

(b) *To justify (9.2.4), first verify that the non-negative* $g(x(\cdot), (b,\infty))$ *is continuous on any sequence whose limit is in* $\mathbb{G} = \{x(\cdot) : g(x(\cdot), \{b\}) = 0\}$ *and that* $\mathbf{E}[g(W_\cdot, \{b\})] = 0$, *hence* $g(\widehat{S}_n, (b,\infty)) \xrightarrow{\mathcal{D}} A_1(b)$. *Then, deduce that* $\widehat{A}_n(b) \xrightarrow{\mathcal{D}} A_1(b)$ *by showing that for any* $\delta > 0$ *and* $n \ge 1$,

$$g(\widehat{S}_n, (b+\delta,\infty)) - \Delta_n(\delta) \le \widehat{A}_n(b) \le g(\widehat{S}_n, (b-\delta,\infty)) + \Delta_n(\delta)\,,$$

with $\Delta_n(\delta) = n^{-1}\sum_{k=1}^n I_{\{|\xi_k| > \delta\sqrt{n}\}}$ *converging in probability to zero when* $n \to \infty$.

Our next result is a refinement due to Kolmogorov and Smirnov, of the *Glivenko-Cantelli theorem* (which states that for i.i.d. $\{X, X_k\}$ the empirical distribution functions $F_n(x) = n^{-1}\sum_{i=1}^n I_{(-\infty,x]}(X_i)$, converge w.p.1., uniformly in x, to the distribution function of X, whichever it may be, see Theorem 2.3.6).

Corollary 9.2.13. *Suppose* $\{X_k, X\}$ *are i.i.d. and* $x \mapsto F_X(x)$ *is a continuous function. Then, setting* $D_n = \sup_{x\in\mathbb{R}} |F_X(x) - F_n(x)|$, *as* $n \to \infty$,

$$n^{1/2} D_n \xrightarrow{\mathcal{D}} \sup_{t\in[0,1]} |\widehat{B}_t|\,, \tag{9.2.5}$$

for the standard Brownian bridge $\widehat{B}_t = W_t - tW_1$ *on* $[0,1]$.

Proof. Recall the Skorokhod construction $X_k = X^-(U_k)$ of Theorem 1.2.37, with i.i.d. uniform variables $\{U_k\}$ on $(0,1]$, such that $\{X_k \le x\} = \{U_k \le F_X(x)\}$ for all $x \in \mathbb{R}$ (see (1.2.1)), by which it follows that

$$D_n = \sup_{u\in F_X(\mathbb{R})} |u - n^{-1}\sum_{i=1}^n I_{(0,u]}(U_i)|\,.$$

The assumed continuity of the distribution function $F_X(\cdot)$ further implies that $F_X(\mathbb{R}) = (0,1)$. Throwing away the sample U_n, let $V_{n-1,k}$, $k = 1, \ldots, n-1$, denote the k-th smallest number in $\{U_1,\dots,U_{n-1}\}$. It is not hard to check that $|D_n - \widetilde{D}_n| \le 2n^{-1}$ for

$$\widetilde{D}_n = \max_{k=1}^{n-1} |V_{n-1,k} - \frac{k}{n}| .$$

Now, from Exercise 3.4.11 we have the representation $V_{n-1,k} = T_k/T_n$, where $T_k = \sum_{j=1}^k \tau_j$ for i.i.d. $\{\tau_j\}$, each having the exponential distribution of parameter one. Next, note that $S_k = T_k - k$ is a random walk $S_k = \sum_{j=1}^k \xi_k$ with i.i.d. $\xi_j = \tau_j - 1$ such that $\mathbf{E}\xi_1 = 0$ and $\mathbf{E}\xi_1^2 = 1$. Hence, setting $Z_n = n/T_n$ one easily checks that

$$n^{1/2}\widetilde{D}_n = Z_n \max_{k=1}^{n} |n^{-1/2}S_k - \frac{k}{n}(n^{-1/2}S_n)| = Z_n \sup_{t\in[0,1]} |\widehat{S}_n(t) - t\widehat{S}_n(1)| ,$$

where $\widehat{S}_n(t) = n^{-1/2}S(nt)$ for $S(\cdot)$ of (9.2.1), so $t \mapsto \widehat{S}_n(t) - t\widehat{S}_n(1)$ is linear on each of the intervals $[k/n,(k+1)/n]$, $k \ge 0$. Consequently,

$$n^{1/2}\widetilde{D}_n = Z_n g(\widehat{S}_n(\cdot)) ,$$

for $g(x(\cdot)) = \sup_{t\in[0,1]} |x(t) - tx(1)|$. By the strong law of large numbers, $Z_n \xrightarrow{a.s.} \frac{1}{\mathbf{E}\tau_1} = 1$. Moreover, $g(\cdot)$ is continuous on $C([0,1])$, so by Donsker's invariance principle $g(\widehat{S}_n(\cdot)) \xrightarrow{\mathcal{D}} g(W_\cdot) = \sup_{t\in[0,1]} |\widehat{B}_t|$ (see Definition 9.2.1), and by Slutsky's lemma, first $n^{1/2}\widetilde{D}_n = Z_n g(\widehat{S}_n(\cdot))$ and then $n^{1/2}D_n$ (which is within $2n^{-1/2}$ of $n^{1/2}\widetilde{D}_n$), have the same limit in distribution (see part (c), then part (b) of Exercise 3.2.8). □

Remark. Defining $F_n^{-1}(t) = \inf\{x\in\mathbb{R} : F_n(x) \ge t\}$, with minor modifications the preceeding proof also shows that in case $F_X(\cdot)$ is continuous, $n^{1/2}(F_X(F_n^{-1}(t)) - t) \xrightarrow{\mathcal{D}} \widehat{B}_t$ on $[0,1]$. Further, with little additional work one finds that $n^{1/2}D_n \xrightarrow{\mathcal{D}} \sup_{x\in\mathbb{R}} |\widehat{B}_{F_X(x)}|$ even in case $F_X(\cdot)$ is discontinuous (and which for continuous $F_X(\cdot)$ coincides with (9.2.5)).

You are now to provide an explicit formula for the distribution function $F_{\mathrm{KS}}(\cdot)$ of $\sup_{t\in[0,1]} |\widehat{B}_t|$.

Exercise 9.2.14. *Consider the standard Brownian bridge $\widehat{B}_t$ on $[0,1]$, as in Exercises 7.3.15-7.3.16.*

(a) *Show that $q_b = \mathbf{P}(\sup_{t\in[0,1]} \widehat{B}_t \ge b) = \exp(-2b^2)$ for any $b > 0$.*
Hint: *Argue that $q_b = \mathbf{P}(\tau_b^{(b)} < \infty)$ for $\tau_b^{(r)}$ of Exercise 8.2.35.*

(b) *Deduce that for any non-random $a, c > 0$,*

$$\mathbf{P}(\inf_{t\in[0,1]} \widehat{B}_t \le -a \textit{ or } \sup_{t\in[0,1]} \widehat{B}_t \ge c) = \sum_{n\ge 1} (-1)^{n-1}(p_n + r_n) ,$$

where $p_{2n} = r_{2n} = q_{na+nc}$, $r_{2n+1} = q_{na+nc+c}$ and $p_{2n+1} = q_{na+nc+a}$.
Hint: *Using inclusion-exclusion prove this expression for $p_n = \mathbf{P}($ for some $0 < t_1 < \cdots < t_n \le 1$, $\widehat{B}_{t_i} = -a$ for odd i and $\widehat{B}_{t_i} = c$ for even $i)$ and r_n similarly defined, just with $\widehat{B}_{t_i} = c$ at odd i and $\widehat{B}_{t_i} = -a$ at even i. Then use the reflection principle for Brownian motion W_t such that $|W_1| \le \varepsilon$ to equate these with the relevant q_b (in the limit $\varepsilon \downarrow 0$).*

(c) *Conclude that for any non-random $b > 0$,*

$$F_{\rm KS}(b) = \mathbf{P}(\sup_{t\in[0,1]} |\widehat{B}_t| \le b) = 1 - 2\sum_{n=1}^{\infty}(-1)^{n-1}e^{-2n^2b^2}\,. \tag{9.2.6}$$

Remark. The typical approach to accepting/rejecting the hypothesis that i.i.d. observations $\{X_k\}$ have been generated according to a specified continuous distribution F_X is by thresholding the value of $F_{KS}(b)$ at the observed *Kolmogorov-Smirnov statistic* $b = n^{1/2}D_n$, per (9.2.6). To this end, while outside our scope, using the so called *Hungarian construction*, which is a much sharper coupling alternative to Corollary 9.2.21, one can further find the rate (in n) of the convergence in distribution in (9.2.5) (for details, see [**SW86**, Chapter 12.1]).

We conclude with a sufficient condition for convergence in distribution on $C([0,\infty))$.

Exercise 9.2.15. *Suppose $C([0,\infty))$-valued random variables X_n, $1 \le n \le \infty$, defined on the same probability space, are such that $\|X_n - X_\infty\|_\ell \xrightarrow{p} 0$ for $n \to \infty$ and any $\ell \ge 1$ fixed (where $\|x\|_t = \sup_{s\in[0,t]}|x(s)|$). Show that $X_n \xrightarrow{\mathcal{D}} X_\infty$ in the topological space $\mathbb{S}$ consisting of $C([0,\infty))$ equipped with the topology of uniform convergence on compact subsets of $[0,\infty)$.*
Hint: *Consider Exercise 7.2.9 and Corollary 3.5.3.*

9.2.1. Skorokhod's representation and the martingale CLT. We pursue here an alternative approach for proving invariance principles, which is better suited to deal with dependence, culminating with a Lindeberg type, martingale CLT. We first utilize the continuity of the Brownian path to deduce, in view of Corollary 3.5.3, that linear interpolation of random discrete samples along the path converges in distribution to the Brownian motion, provided the sample times approach a uniform density.

Lemma 9.2.16. *Suppose $\{W(t), t \ge 0\}$ is a standard Wiener process and $k \mapsto T_{n,k}$ are non-decreasing, such that $T_{n,[nt]} \xrightarrow{p} t$ when $n \to \infty$, for each fixed $t \in [0,\ell]$. Then, $\|\widehat{S}_n - W\| \xrightarrow{p} 0$ for the norm $\|x(\cdot)\| = \sup_{t\in[0,\ell]}|x(t)|$ of $C([0,\ell])$ and $\widehat{S}_n(t) = S_n(nt)$, where*

$$S_n(t) = W(T_{n,[t]}) + (t-[t])(W(T_{n,[t]+1}) - W(T_{n,[t]}))\,. \tag{9.2.7}$$

Remark. In view of Exercise 9.2.15, the preceding lemma implies in particular that if $T_{n,[nt]} \xrightarrow{p} t$ for each fixed $t \ge 0$, then $\widehat{S}_n(\cdot) \xrightarrow{\mathcal{D}} W(\cdot)$ in $C([0,\infty))$.

PROOF. Recall that each sample function $t \mapsto W(t)(\omega)$ is uniformly continuous on $[0,\ell]$, hence $\mathsf{osc}_{\ell,\delta}(W(\cdot))(\omega) \downarrow 0$ as $\delta \downarrow 0$ (see (9.2.2) for the definition of $\mathsf{osc}_{\ell,\delta}(x(\cdot))$). Fixing $\varepsilon > 0$ note that as $r \uparrow \infty$,

$$G_r = \{\omega \in \Omega : \mathsf{osc}_{\ell,3/r}(W(\cdot))(\omega) \le \varepsilon\} \uparrow \Omega\,,$$

so by continuity from below of probability measures, $\mathbf{P}(G_r) \ge 1-\varepsilon$ for some integer r. Setting $s_j = j/r$ for $j = 0,1,\ldots,\ell r+1$, our hypothesis that $T_{n,[nt]} \xrightarrow{p} t$ per fixed $t \ge 0$, hence uniformly on any *finite* collection of times, implies that for some finite $n_0 = n_0(\varepsilon, r)$ and all $n \ge n_0$,

$$\mathbf{P}(\max_{j=0}^{\ell r}|T_{n,[ns_j]} - s_j| \le r^{-1}) \ge 1-\varepsilon\,.$$

Further, by the monotonicity of $k \mapsto T_{n,k}$, if $t \in [s_{j-1}, s_j)$ and $n \geq r$, then

$$T_{n,[ns_{j-1}]} - s_j \leq T_{n,[nt]} - t \leq T_{n,[nt]+1} - t \leq T_{n,[ns_{j+1}]} - s_{j-1}$$

and since $s_{j+1} - s_{j-1} = 2/r$, it follows that for any $n \geq \max(n_0, r)$,

$$\mathbf{P}(\sup_{b\in\{0,1\}, t\in[0,\ell)} |T_{n,[nt]+b} - t| \leq 3r^{-1}) \geq 1 - \varepsilon\,. \tag{9.2.8}$$

Recall that for any $n \geq 1$ and $t \geq 0$,

$$|\widehat{S}_n(t) - W(t)| \leq (1-\eta)|W(T_{n,[nt]}) - W(t)| + \eta|W(T_{n,[nt]+1}) - W(t)|\,,$$

where $\eta = nt - [nt] \in [0, 1)$. Observe that if both G_r and the event in (9.2.8) occur, then by definition of G_r each of the two terms on the right-side of the last inequality is at most ε. We thus see that $\|\widehat{S}_n - W\|_\ell \leq \varepsilon$ whenever both G_r and the event in (9.2.8) occur. That is, $\mathbf{P}(\|\widehat{S}_n - W\| \leq \varepsilon) \geq 1 - 2\varepsilon$. Since this applies for all $\varepsilon > 0$, we have just shown that $\|\widehat{S}_n - W\| \xrightarrow{p} 0$, as claimed. $\square$

The key tool in our program is an alternative Skorokhod representation of random variables. Whereas in Theorem 1.2.37 we applied the inverse of the desired distribution function to a uniform random variable on $[0, 1]$, here we construct a stopping time τ of the form $\tau_{A,B} = \inf\{t \geq 0 : W_t \notin (-A, B)\}$ such that W_τ has the stated, mean zero law. To this end, your next exercise exhibits the appropriate *random* levels (A, B) to be used in this construction.

Definition 9.2.17. *Given a random variable $V \geq 0$ of positive, finite mean, we say that $Z \geq 0$ is a* size-biased sample *of V if $\mathbf{E}[g(Z)] = \mathbf{E}[Vg(V)]/\mathbf{E}V$ for all $g \in b\mathcal{B}$. Alternatively, the Radon-Nikodym derivative between the corresponding laws is $\frac{d\mathcal{P}_Z}{d\mathcal{P}_V}(v) = v/\mathbf{E}V$.*

Exercise 9.2.18. *Suppose X is an integrable random variable, such that $\mathbf{E}X = 0$ (so $\mathbf{E}X_+ = \mathbf{E}X_-$ is finite). Consider the $[0,\infty)^2$-valued random vector*

$$(A, B) = (0, 0)I_{\{X=0\}} + (Z, X)I_{\{X>0\}} + (-X, Y)I_{\{X<0\}}\,,$$

where Y and Z are size-biased samples of X_+ and X_-, respectively, which are further independent of X. Show that then for any $f \in b\mathcal{B}$,

$$\mathbf{E}[r(A,B)f(-A) + (1 - r(A,B))f(B)] = \mathbf{E}[f(X)]\,, \tag{9.2.9}$$

where $r(a, b) = b/(a + b)$ for $a > 0$ and $r(0, b) = 1$.
Hint: *It suffices to show that $\mathbf{E}[Xh(Z,X)I_{\{X>0\}}] = \mathbf{E}[(-X)h(-X,Y)I_{\{X<0\}}]$ for $h(a,b) = (f(b) - f(-a))/(a + b)$.*

Theorem 9.2.19 (Skorokhod's representation)**.** *Suppose $(W_t, \mathcal{F}_t, t \geq 0)$ is a Brownian Markov process such that $W_0 = 0$ and $\mathcal{F}_\infty$ is independent of the $[0, 1]$-valued independent uniform variables U_i, $i = 1, 2$. With $\mathcal{G}_t = \sigma(\sigma(U_1, U_2), \mathcal{F}_t)$, given the law $\mathcal{P}_X$ of an integrable X such that $\mathbf{E}X = 0$, there exists an a.s. finite $\mathcal{G}_t$-stopping time τ such that $W_\tau \stackrel{\mathcal{D}}{=} X$, $\mathbf{E}\tau = \mathbf{E}X^2$ and $\mathbf{E}\tau^2 \leq 2\mathbf{E}X^4$.*

Remark. The extra randomness in the form of U_i, $i = 1, 2$ is not needed when $X = \{-a, b\}$ takes only two values (for you can easily check that then Exercise 9.2.18 simply sets $A = a$ and $B = b$). In fact, one can eliminate it altogether, at the cost of a more involved proof (see [**Bil95**, Theorem 37.6] or [**MP09**, Sections 5.3.1-5.3.2] for how this is done).

PROOF. We first construct (X, Y, Z) of Exercise 9.2.18 from the independent pair of uniformly distributed random variables (U_1, U_2). That is, given the specified distribution function $F_X(\cdot)$ of X, we set $X(\omega) = \sup\{x : F_X(x) < U_1(\omega)\}$ as in Theorem 1.2.37 so that $X \in m\sigma(U_1)$. Noting that $F_X(\cdot)$ uniquely determines the distribution functions $F_Y(y) = I_{\{y>0\}} \int_0^y v dF_X(v)/\mathbf{E}X_+$ and $F_Z(z) = I_{\{z>0\}} \int_{-z}^0 (-v) dF_X(v)/\mathbf{E}X_-$, we apply the same procedure to construct the strictly positive Y and Z out of U_2. With the resulting pair (Y, Z) measurable on $\sigma(U_2)$ and independent of X, we proceed to have the $[0,\infty)^2$-valued random vector (A, B) measurable on $\sigma(U_1, U_2)$ as in Exercise 9.2.18.

We claim that $\tau = \inf\{t \geq 0 : W_t \notin (-A, B)\}$ is an a.s. finite $\mathcal{G}_t$-stopping time. Indeed, considering the $\mathcal{F}_t$-adapted $M_t^+ = \sup_{s\in[0,t]} W_s$ and $M_t^- = \inf_{s\in[0,t]} W_s$, note that by continuity of the Brownian sample function, $\tau \leq t$ if and only if either $M_t^+ \geq B$ or $M_t^- \leq -A$. With (A, B) measurable on $\sigma(U_1, U_2) \subseteq \mathcal{G}_0$ it follows that $\{\omega : \tau(\omega) \leq t\} \in \mathcal{G}_t$ for each $t \geq 0$. Further, recall part (d) of Exercise 8.2.35 that w.p.1. $M_t^+ \uparrow \infty$, in which case the $\mathcal{G}_t$-stopping time $\tau(\omega)$ is finite.

Setting $\mathbb{Y} = [0,\infty)^2$, we deduce from the preceding analysis of the events $\{\tau \leq t\}$ that $\tau(\omega) = h(\omega, (A(\omega), B(\omega))$ with $h(\omega, (a, b)) = \tau_{a,b}(\omega)$ measurable on the product space $(\Omega, \mathcal{F}_\infty) \times (\mathbb{Y}, \mathcal{B}_\mathbb{Y})$. Since $\{W_t\}$ is $\mathcal{F}_t$-progressively measurable, the same applies for $V = I_{\{\tau<\infty\}} f(W_\tau, \tau)$ and any bounded below Borel function f on $\mathbb{R}^2$. With $\sigma(A, B) \subseteq \sigma(U_1, U_2)$ independent of $\mathcal{F}_\infty$, it thus follows from Fubini's theorem that $\mathbf{E}[V] = \mathbf{E}[g(A, B)]$ for

$$g(a, b) = \mathbf{E}[I_{\{\tau_{a,b}<\infty\}} f(W_{\tau_{a,b}}, \tau_{a,b})] .$$

In particular, considering $V = I_{\{\tau<\infty\}} f(W_\tau)$, since $\{W_t\}$ is a Brownian Markov process starting at $W_0 = 0$, we have from part (a) of Exercise 8.2.36 that

$$g(a, b) = r(a, b)f(-a) + (1 - r(a, b))f(b) ,$$

where $r(a, b) = b/(a + b)$ for $a > 0$ and $r(0, b) = 1$. In view of the identity (9.2.9) we thus deduce that for any $f \in b\mathcal{B}$,

$$\mathbf{E}[f(W_\tau)] = E[V] = \mathbf{E}[g(A, B)] = \mathbf{E}[f(X)] .$$

That is, $W_\tau \overset{\mathcal{D}}{=} X$, as claimed. Recall part (c) of Exercise 8.2.36 that $\mathbf{E}[\tau_{a,b}] = ab$. Since $ab = r(a, b)f(-a) + (1 - r(a, b))f(b)$ for $f(x) = x^2$ and any $(a, b) \in \mathbb{Y}$, we deduce by the same reasoning that $\mathbf{E}[\tau] = \mathbf{E}[AB] = \mathbf{E}[X^2]$. Similarly, the bound $\mathbf{E}\tau^2 \leq \frac{5}{3}\mathbf{E}X^4$ follows from the identity $\mathbf{E}[\tau_{a,b}^2] = (ab)^2 + ab(a^2 + b^2)/3$ of part (c) of Exercise 8.2.36, and the inequality

$$ab(a^2 + b^2 + 3ab) \leq 5ab(a^2 + b^2 - ab) = 5[r(a, b)a^4 + (1 - r(a, b))b^4]$$

for all $(a, b) \in \mathbb{Y}$. $\square$

Building on Theorem 9.2.19 we have the following representation, due to Strassen, of any discrete time martingale via sampling along the path of the Brownian motion.

Theorem 9.2.20 (STRASSEN'S MARTINGALE REPRESENTATION)**.** *Suppose the probability space contains a martingale $(M_\ell, \mathcal{F}_\ell)$ such that $M_0 = 0$, the i.i.d. $[0,1]$-valued uniform $\{U_i\}$ and a standard Wiener process $\{W(t), t \geq 0\}$, both independent of $\mathcal{F}_\infty$ and of each other. Then:*

(a) *The filtrations $\mathcal{F}_{k,t} = \sigma(\mathcal{F}_k, \sigma(U_i, i \leq 2k), \mathcal{F}_t^{\mathbf{W}})$ are such that $(W(t), \mathcal{F}_{k,t})$ is a Brownian Markov process for any $1 \leq k \leq \infty$.*

(b) *There exist non-decreasing a.s. finite $\mathcal{F}_{k,t}$-stopping times $\{T_k\}$, starting with $T_0 = 0$, where $\tau_k = T_k - T_{k-1}$ and the filtration $\mathcal{H}_k = \mathcal{F}_{k,T_k}$ are such that w.p.1. $\mathbf{E}[\tau_k|\mathcal{H}_{k-1}] = \mathbf{E}[D_k^2|\mathcal{F}_{k-1}]$ and $\mathbf{E}[\tau_k^2|\mathcal{H}_{k-1}] \le 2\mathbf{E}[D_k^4|\mathcal{F}_{k-1}]$ for the martingale differences $D_k = M_k - M_{k-1}$ and all $k \ge 1$.*

(c) *The discrete time process $\{W(T_\ell)\}$ has the same f.d.d. as $\{M_\ell\}$.*

Remark. In case $\mathcal{F}_\ell = \sigma(M_k, k \le \ell)$, our proof constructs the martingale as samples $M_\ell = W(T_\ell)$ of the standard Wiener process $\{W(t), t \ge 0\}$ at a.s. finite, non-decreasing $\sigma(\sigma(U_i, i \le 2k), \mathcal{F}_t^{\mathbf{W}})$-stopping times T_k. Upon eliminating the extra randomness (U_1, U_2) in Theorem 9.2.19, we thus get the *embedding* of $\{M_n\}$ inside the path $t \mapsto W_t$, where $\{T_k\}$ are $\mathcal{F}_t^{\mathbf{W}}$-stopping times and part (b) of the theorem applies for the corresponding stopped σ-algebras $\mathcal{H}_k = \mathcal{F}_{T_k}^{\mathbf{W}}$. Indeed, we have stipulated the a-apriori existence of $(M_\ell, \mathcal{F}_\ell)$ independently of the Wiener process only in order to accommodate non-canonical filtrations $\{\mathcal{F}_\ell\}$.

Proof. (a). Fixing $1 \le k \le \infty$, since $\mathcal{H} = \mathcal{F}_{k,0}$ is independent of $\mathcal{F}^{\mathbf{W}}$, we have from Proposition 4.2.3 that for any $B \in \mathcal{B}$ and $u, s \ge 0$,

$$\mathbf{E}[I_B(W(u+s))|\mathcal{F}_{k,s}] = \mathbf{E}[I_B(W(u+s))|\sigma(\mathcal{H}, \mathcal{F}_s^{\mathbf{W}})] = \mathbf{E}[I_B(W(u+s))|\mathcal{F}_s^{\mathbf{W}}] .$$

With $(W(t), \mathcal{F}_t^{\mathbf{W}})$ a Brownian Markov process, it thus follows that so are $(W(t), \mathcal{F}_{k,t})$.
(b). Starting at $T_0 = 0$ we sequentially construct the non-decreasing $\mathcal{F}_{k,t}$-stopping times T_k. Assuming T_{k-1} have been constructed already, consider Corollary 9.1.6 for the Brownian Markov process $(W(t), \mathcal{F}_{k-1,t})$ and the $\mathcal{F}_{k-1,t}$-stopping time T_{k-1}, to deduce that $W_k(t) = W(T_{k-1}+t) - W(T_{k-1})$ is a standard Wiener process which is independent of $\mathcal{H}_{k-1}$. The pair (U_{2k-1}, U_{2k}) of $[0,1]$-valued independent uniform variables is by assumption independent of $\mathcal{F}_{k-1,\infty}$, hence of both $\mathcal{H}_{k-1}$ and the standard Wiener process $\{W_k(t)\}$. Recall that $D_k = M_k - M_{k-1}$ is integrable and by the martingale property $\mathbf{E}[D_k|\mathcal{F}_{k-1}] = 0$. With $\mathcal{F}_{k-1} \subseteq \mathcal{H}_{k-1}$ and representing our probability measure as a product measure on $(\Omega \times \Omega', \mathcal{F}_\infty \times \mathcal{G}_\infty)$, we thus apply Theorem 9.2.19 for the continuous filtration $\mathcal{G}_t = \sigma(U_{2k-1}, U_{2k}, W_k(s), s \le t)$ which is independent of $\mathcal{H}_{k-1}$ and the *random* distribution function

$$F_{X_k}(x;\omega) = \widehat{\mathbf{P}}_{D_k|\mathcal{F}_{k-1}}((-\infty, x], \omega) ,$$

corresponding to the zero-mean R.C.P.D. of D_k given $\mathcal{F}_{k-1}$. The resulting a.s. finite $\mathcal{G}_t$-stopping time τ_k is such that

$$\mathbf{E}[\tau_k|\mathcal{H}_{k-1}] = \int x^2 dF_{X_k}(x;\omega) = \mathbf{E}[D_k^2|\mathcal{F}_{k-1}]$$

(see Exercise 4.4.6 for the second identity), and by the same reasoning $\mathbf{E}[\tau_k^2|\mathcal{H}_{k-1}] \le 2\mathbf{E}[D_k^4|\mathcal{F}_{k-1}]$, while the R.C.P.D. of $W_k(\tau_k)$ given $\mathcal{H}_{k-1}$ matches the law $\widehat{\mathbf{P}}_{D_k|\mathcal{F}_{k-1}}$ of X_k. Note that the threshold levels (A_k, B_k) of Exercise 9.2.18 are measurable on $\mathcal{F}_{k,0}$ since by right-continuity of distribution functions their construction requires only the values of (U_{2k-1}, U_{2k}) and $\{F_{X_k}(q;\omega), q \in \mathbb{Q}\}$. For example, for any $x \in \mathbb{R}$,

$$\{\omega : X_k(\omega) \le x\} = \{\omega : U_1(\omega) \le F_{X_k}(q;\omega) \text{ for all } q \in \mathbb{Q}, q > x\} ,$$

with similar identities for $\{Y_k \le y\}$ and $\{Z_k \le z\}$, whose distribution functions at $q \in \mathbb{Q}$ are ratios of integrals of the type $\int_0^q v dF_{X_k}(v;\omega)$, each of which is the limit as $n \to \infty$ of the $\mathcal{F}_{k-1}$-measurable

$$\sum_{\ell \le qn} (\ell/n)(F_{X_k}(\ell/n + 1/n;\omega) - F_{X_k}(\ell/n;\omega)) .$$

Further, setting $T_k = T_{k-1} + \tau_k$, from the proof of Theorem 9.2.19 we have that $\{T_k \le t\}$ if and only if $\{T_{k-1} \le t\}$ and either $\sup_{u\in[0,t]}\{W(u) - W(u \wedge T_{k-1})\} \ge B_k$ or $\inf_{u\in[0,t]}\{W(u) - W(u \wedge T_{k-1})\} \le -A_k$. Consequently, the event $\{T_k \le t\}$ is in

$$\sigma(A_k, B_k, t \wedge T_{k-1}, I_{\{T_{k-1}\le t\}}, W(s), s \le t) \subseteq \mathcal{F}_{k,t}\ ,$$

by the $\mathcal{F}_{k,0}$-measurability of (A_k, B_k) and our hypothesis that T_{k-1} is an $\mathcal{F}_{k-1,t}$-stopping time.

(c). With $W(T_0) = M_0 = 0$, it clearly suffices to show that the f.d.d. of $\{W_\ell(\tau_\ell) = W(T_\ell) - W(T_{\ell-1})\}$ match those of $\{D_\ell = M_\ell - M_{\ell-1}\}$. To this end, recall that $\mathcal{H}_k = \mathcal{F}_{k,T_k}$ is a filtration (see part (b) of Exercise 8.1.11), and in part (b) we saw that $W_k(\tau_k)$ is $\mathcal{H}_k$-adapted such that its R.C.P.D. given $\mathcal{H}_{k-1}$ matches the R.C.P.D. of the $\mathcal{F}_k$-adapted D_k, given $\mathcal{F}_{k-1}$. Hence, for any $f_\ell \in b\mathcal{B}$, we have from the tower property that

$$\begin{aligned}
\mathbf{E}[\prod_{\ell=1}^{n} f_\ell(W_\ell(\tau_\ell))] &= \mathbf{E}[\mathbf{E}[f_n(W_n(\tau_n))|\mathcal{H}_{n-1}] \prod_{\ell=1}^{n-1} f_\ell(W_\ell(\tau_\ell))] \\
&= \mathbf{E}[\mathbf{E}[f_n(D_n)|\mathcal{F}_{n-1}] \prod_{\ell=1}^{n-1} f_\ell(W_\ell(\tau_\ell))] \\
&= \mathbf{E}[\mathbf{E}[f_n(D_n)|\mathcal{F}_{n-1}] \prod_{\ell=1}^{n-1} f_\ell(D_\ell)] = \mathbf{E}[\prod_{\ell=1}^{n} f_\ell(D_\ell)],
\end{aligned}$$

where the third equality is from the induction assumption that $\{D_\ell\}_{\ell=1}^{n-1}$ has the same law as $\{W_\ell(\tau_\ell)\}_{\ell=1}^{n-1}$. $\square$

The following corollary of Strassen's representation recovers Skorokhod's representation of the random walk $\{S_n\}$ as the samples of Brownian motion at a sequence of stopping times with i.i.d. increments.

Corollary 9.2.21 (Skorokhod's representation for random walks)**.** *Suppose ξ_1 is integrable and of zero mean. The random walk $S_n = \sum_{k=1}^n \xi_k$ of i.i.d. $\{\xi_k\}$ can be represented as $S_n = W(T_n)$ for $T_0 = 0$, i.i.d. $\tau_k = T_k - T_{k-1} \ge 0$ such that $\mathbf{E}[\tau_1] = \mathbf{E}[\xi_1^2]$ and standard Wiener process $\{W(t), t \ge 0\}$. Further, each T_k is a stopping time for $\mathcal{F}_{k,t} = \sigma(\sigma(U_i, i \le 2k), \mathcal{F}_t^{\mathbf{W}})$ (with i.i.d. $[0,1]$-valued uniform $\{U_i\}$ that are independent of $\{W_t, t \ge 0\}$).*

Proof. The construction we provided in proving Theorem 9.2.20 is based on inductively applying Theorem 9.2.19 for $k = 1, 2, \ldots$, where X_k follows the R.C.P.D. of the MG difference D_k given $\mathcal{F}_{k-1}$. For a martingale of independent differences, such as the random walk S_n, we can *a-apriori* produce the independent thresholds (A_k, B_k), $k \ge 1$, out of the given pairs of $\{U_i\}$, independently of the Wiener process. Then, in view of Corollary 9.1.6, for $k = 1, 2, \ldots$ both (A_k, B_k) and the standard Wiener process $W_k(\cdot) = W(T_{k-1} + \cdot) - W(T_{k-1})$ are independent of the stopped at T_{k-1} element of the continuous time filtration $\sigma(A_i, B_i, i < k, W(s), s \le t)$. Consequently, so is the stopping time $\tau_k = \inf\{t \ge 0 : W_k(t) \notin (-A_k, B_k)\}$ with respect to the continuous time filtration $\mathcal{G}_{k,t} = \sigma(A_k, B_k, W_k(s), s \le t)$, from which we conclude that $\{\tau_k\}$ are in this case i.i.d. $\square$

Combining Strassen's martingale representation with Lemma 9.2.16, we are now in position to prove a Lindeberg type martingale CLT.

Theorem 9.2.22 (MARTINGALE CLT, LINDEBERG'S). *Suppose that for any $n \geq 1$ fixed, $(M_{n,\ell}, \mathcal{F}_{n,\ell})$ is a (discrete time) L^2-martingale, starting at $M_{n,0} = 0$, and the corresponding martingale differences $D_{n,k} = M_{n,k} - M_{n,k-1}$ and* predictable compensators

$$\langle M_n \rangle_\ell = \sum_{k=1}^{\ell} \mathbf{E}[D_{n,k}^2 | \mathcal{F}_{n,k-1}]\,,$$

are such that for any fixed $t \in [0,1]$, as $n \to \infty$,

$$\langle M_n \rangle_{[nt]} \xrightarrow{p} t\,. \tag{9.2.10}$$

If in addition, for each $\varepsilon > 0$,

$$g_n(\varepsilon) = \sum_{k=1}^{n} \mathbf{E}[D_{n,k}^2 I_{\{|D_{n,k}| \geq \varepsilon\}} | \mathcal{F}_{n,k-1}] \xrightarrow{p} 0\,, \tag{9.2.11}$$

then as $n \to \infty$, the linearly interpolated, time-scaled S.P.

$$\widehat{S}_n(t) = M_{n,[nt]} + (nt - [nt]) D_{n,[nt]+1}\,, \tag{9.2.12}$$

converges in distribution on $C([0,1])$ to the standard Wiener process, $\{W(t), t \in [0,1]\}$.

Remark. For martingale differences $\{D_{n,k}, k = 1, 2, \ldots\}$ that are mutually independent per fixed n, our assumption (9.2.11) reduces to Lindeberg's condition (3.1.4) and the predictable compensators $v_{n,t} = \langle M_n \rangle_{[nt]}$ are then non-random. In particular, $v_{n,t} = [nt]/n \to t$ in case $D_{n,k} = n^{-1/2}\xi_k$ for i.i.d. $\{\xi_k\}$ of zero mean and unit variance, in which case $g_n(\varepsilon) = \mathbf{E}[\xi_1^2; |\xi_1| \geq \varepsilon\sqrt{n}] \to 0$ (see Remark 3.1.4), and we recover Donsker's invariance principle as a special case of Lindeberg's martingale CLT.

PROOF. *Step 1.* We first prove a somewhat stronger convergence statement for martingales of uniformly bounded differences and predictable compensators. Specifically, using $\|\cdot\|$ for the supremum norm $\|x(\cdot)\| = \sup_{t\in[0,1]} |x(t)|$ on $C([0,1])$, we proceed under the *additional* assumption that for some non-random $\varepsilon_n \to 0$,

$$\langle M_n \rangle_n \leq 2\,, \qquad \max_{k=1}^{n} |D_{n,k}| \leq 2\varepsilon_n\,,$$

to construct a *coupling* of $\widehat{S}_n(\cdot)$ of (9.2.12) and the standard Wiener process $W(\cdot)$ in the same probability space, such that $\|\widehat{S}_n - W\| \xrightarrow{p} 0$. To this end, apply Theorem 9.2.20 simultaneously for the martingales $(M_{n,\ell}, \mathcal{F}_{n,\ell})$ with the same standard Wiener process $\{W(t), t \geq 0\}$ and auxiliary i.i.d. $[0,1]$-valued uniform $\{U_i\}$, to get the representation $M_{n,\ell} = W(T_{n,\ell})$. Recall that $T_{n,\ell} = \sum_{k=1}^{\ell} \tau_{n,k}$ where for each n, the non-negative $\tau_{n,k}$ are adapted to the filtration $\{\mathcal{H}_{n,k}, k \geq 1\}$ and such that w.p.1. for $k = 1, \ldots, n$,

$$\mathbf{E}[\tau_{n,k} | \mathcal{H}_{n,k-1}] = \mathbf{E}[D_{n,k}^2 | \mathcal{F}_{n,k-1}]\,, \tag{9.2.13}$$

$$\mathbf{E}[\tau_{n,k}^2 | \mathcal{H}_{n,k-1}] \leq 2\mathbf{E}[D_{n,k}^4 | \mathcal{F}_{n,k-1}]\,. \tag{9.2.14}$$

Under this representation, the process $\widehat{S}_n(\cdot)$ of (9.2.12) is of the form considered in Lemma 9.2.16, and as shown there, $\|\widehat{S}_n - W\| \xrightarrow{p} 0$ provided $T_{n,[nt]} \xrightarrow{p} t$ for each fixed $t \in [0,1]$.

To verify the latter convergence in probability, set $\widehat{T}_{n,\ell} = T_{n,\ell} - \langle M_n \rangle_\ell$ and $\widehat{\tau}_{n,k} = \tau_{n,k} - \mathbf{E}[\tau_{n,k}|\mathcal{H}_{n,k-1}]$. Note that by the identities of (9.2.13),

$$\langle M_n \rangle_\ell = \sum_{k=1}^{\ell} \mathbf{E}[D_{n,k}^2|\mathcal{F}_{n,k-1}] = \sum_{k=1}^{\ell} \mathbf{E}[\tau_{n,k}|\mathcal{H}_{n,k-1}]\,,$$

hence $\widehat{T}_{n,\ell} = \sum_{k=1}^{\ell} \widehat{\tau}_{n,k}$ is for each n, the $\mathcal{H}_{n,\ell}$-martingale part in Doob's decomposition of the integrable, $\mathcal{H}_{n,\ell}$-adapted sequence $\{T_{n,\ell}, \ell \geq 0\}$. Further, considering the expectation in both sides of (9.2.14), by our assumed uniform bound $|D_{n,k}| \leq 2\varepsilon_n$ it follows that for any $k = 1, \ldots, n$,

$$\mathbf{E}[\widehat{\tau}_{n,k}^2] \leq \mathbf{E}[\tau_{n,k}^2] \leq 2\mathbf{E}[D_{n,k}^4] \leq 8\varepsilon_n^2 \mathbf{E}[D_{n,k}^2]\,,$$

where the left-most inequality is just the L^2-reduction of conditional centering (namely, $\mathbf{E}[\mathsf{Var}(X|\mathcal{H})] \leq \mathbf{E}[X^2]$, see part (a) of Exercise 4.2.16). Consequently, the martingale $\{\widehat{T}_{n,\ell}, \ell \geq 0\}$ is square-integrable and since its differences $\widehat{\tau}_{n,k}$ are uncorrelated (see part (a) of Exercise 5.1.8), we deduce that for any $\ell \leq n$,

$$\mathbf{E}[\widehat{T}_{n,\ell}^2] = \sum_{k=1}^{\ell} \mathbf{E}[\widehat{\tau}_{n,k}^2] \leq 8\varepsilon_n^2 \sum_{k=1}^{n} \mathbf{E}[D_{n,k}^2] = 8\varepsilon_n^2 \mathbf{E}[\langle M_n \rangle_n]\,.$$

Recall our assumption that $\langle M_n \rangle_n$ is uniformly bounded, hence fixing $t \in [0,1]$, we conclude that $\widehat{T}_{n,[nt]} \stackrel{L^2}{\to} 0$ as $n \to \infty$. This of course implies the convergence to zero in probability of $\widehat{T}_{n,[nt]}$, and in view of assumption (9.2.10) and Slutsky's lemma, also $T_{n,[nt]} = \widehat{T}_{n,[nt]} + \langle M_n \rangle_{[nt]} \stackrel{p}{\to} t$ as $n \to \infty$.

Step 2. We next eliminate the superfluous assumption $\langle M_n \rangle_n \leq 2$ via the strategy employed in proving part (a) of Theorem 5.3.32. That is, consider the $\mathcal{F}_{n,\ell}$-stopped martingales $\widetilde{M}_{n,\ell} = M_{n,\ell\wedge\theta_n}$ for stopping times $\theta_n = n \wedge \min\{\ell < n : \langle M_n \rangle_{\ell+1} > 2\}$, such that $\langle M_n \rangle_{\theta_n} \leq 2$. As the corresponding martingale differences are $\widetilde{D}_{n,k} = D_{n,k} I_{\{k \leq \theta_n\}}$, you can easily verify that for all $\ell \leq n$

$$\langle \widetilde{M}_n \rangle_\ell = \langle M_n \rangle_{\ell\wedge\theta_n}\,.$$

In particular, $\langle \widetilde{M}_n \rangle_n = \langle M_n \rangle_{\theta_n} \leq 2$. Further, if $\{\theta_n = n\}$ then $\langle \widetilde{M}_n \rangle_{[nt]} = \langle M_n \rangle_{[nt]}$ for all $t \in [0,1]$ and the function

$$\widetilde{S}_n(t) = \widetilde{M}_{n,[nt]} + (nt - [nt])\widetilde{D}_{n,[nt]+1}\,,$$

coincides on $[0,1]$ with $\widehat{S}_n(\cdot)$ of (9.2.12). Due to the monotonicity of $\ell \mapsto \langle M_n \rangle_\ell$ we have that $\{\theta_n < n\} = \{\langle M_n \rangle_n > 2\}$, so our assumption (9.2.10) that $\langle M_n \rangle_n \stackrel{p}{\to} 1$ implies that $\mathbf{P}(\theta_n < n) \to 0$. Consequently, $\langle \widetilde{M}_n \rangle_{[nt]} \stackrel{p}{\to} t$ and applying Step 1 for the martingales $\{\widetilde{M}_{n,\ell}, \ell \leq n\}$ we have the coupling of $\widetilde{S}_n(\cdot)$ and the standard Wiener process $W(\cdot)$ such that $\|\widetilde{S}_n - W\| \stackrel{p}{\to} 0$. Combining it with the (natural) coupling of $\widetilde{S}_n(\cdot)$ and $\widehat{S}_n(\cdot)$ such that $\mathbf{P}(\widehat{S}_n(\cdot) \neq \widetilde{S}_n(\cdot)) \leq \mathbf{P}(\theta_n < n) \to 0$, we arrive at the coupling of $\widehat{S}_n(\cdot)$ and $W(\cdot)$ such that $\|\widehat{S}_n - W\| \stackrel{p}{\to} 0$.

Step 3. We establish the CLT under the condition (9.2.11), by reducing the problem to the setting we have already handled in Step 2. This is done by a truncation argument similar to the one we used in proving Theorem 2.1.11, except that now we need to re-center the martingale differences after truncation is done (to convince yourself that some truncation is required, consider the special case of i.i.d. $\{D_{n,k}\}$

with infinite forth moment, and note that a stopping argument as in Step 2 is not feasible here because unlike $\langle M_n\rangle_\ell$, the martingale differences are not $\mathcal{F}_{n,\ell}$-predictable).

Specifically, our assumption (9.2.11) implies the existence of finite $n_j \uparrow \infty$ such that $\mathbf{P}(g_n(j^{-1}) \ge j^{-3}) \le j^{-1}$ for all $j \ge 1$ and $n \ge n_j \ge n_1 = 1$. Setting $\varepsilon_n = j^{-1}$ for $n \in [n_j, n_{j+1})$ it then follows that as $n \to \infty$ both $\varepsilon_n \to 0$ and for $\delta > 0$ fixed,

$$\mathbf{P}(\varepsilon_n^{-2} g_n(\varepsilon_n) \ge \delta) \le \mathbf{P}(\varepsilon_n^{-2} g_n(\varepsilon_n) \ge \varepsilon_n) \to 0\,.$$

In conclusion, there exist non-random $\varepsilon_n \to 0$ such that $\varepsilon_n^{-2} g_n(\varepsilon_n) \xrightarrow{p} 0$, which we use hereafter as the truncation level for the martingale differences $\{D_{n,k}, k \le n\}$. That is, consider for each n, the new martingale $\widetilde{M}_{n,\ell} = \sum_{k=1}^{\ell} \widetilde{D}_{n,k}$, where

$$\widetilde{D}_{n,k} = \overline{D}_{n,k} - \mathbf{E}[\overline{D}_{n,k}|\mathcal{F}_{n,k-1}]\,, \qquad \overline{D}_{n,k} = D_{n,k} I_{\{|D_{n,k}|<\varepsilon_n\}}\,.$$

By construction, $|\widetilde{D}_{n,k}| \le 2\varepsilon_n$ for all $k \le n$. Hence, with $\widehat{D}_{n,k} = D_{n,k} I_{\{|D_{n,k}|\ge\varepsilon_n\}}$, by Slutsky's lemma and the preceding steps of the proof we have a coupling such that $\|\widetilde{S}_n - W\| \xrightarrow{p} 0$, as soon as we show that for all $\ell \le n$,

$$0 \le \langle M_n\rangle_\ell - \langle \widetilde{M}_n\rangle_\ell \le 2\sum_{k=1}^{\ell} \mathbf{E}[\widehat{D}_{n,k}^2|\mathcal{F}_{n,k-1}]$$

(for the right hand side is bounded by $2g_n(\varepsilon_n)$ which by our choice of ε_n converges to zero in probability, so the convergence (9.2.10) of the predictable compensators transfers from M_n to $\widetilde{M}_n$). These inequalities are in turn a direct consequence of the bounds

$$\mathbf{E}[\widetilde{D}_{n,k}^2|\mathcal{F}] \le \mathbf{E}[\overline{D}_{n,k}^2|\mathcal{F}] \le \mathbf{E}[D_{n,k}^2|\mathcal{F}] \le \mathbf{E}[\widetilde{D}_{n,k}^2|\mathcal{F}] + 2\mathbf{E}[\widehat{D}_{n,k}^2|\mathcal{F}] \tag{9.2.15}$$

holding for $\mathcal{F} = \mathcal{F}_{n,k-1}$ and all $1 \le k \le n$. The left-most inequality in (9.2.15) is merely an instance of the L^2-reduction of conditional centering, while the middle one follows from the identity $D_{n,k} = \overline{D}_{n,k} + \widehat{D}_{n,k}$ upon realizing that by definition $\overline{D}_{n,k} \ne 0$ if and only if $\widehat{D}_{n,k} = 0$, so

$$\mathbf{E}[D_{n,k}^2|\mathcal{F}] = \mathbf{E}[\overline{D}_{n,k}^2|\mathcal{F}] + \mathbf{E}[\widehat{D}_{n,k}^2|\mathcal{F}]\,.$$

The latter identity also yields the right-most inequality in (9.2.15), for $\mathbf{E}[\overline{D}_{n,k}|\mathcal{F}] = -\mathbf{E}[\widehat{D}_{n,k}|\mathcal{F}]$ (due to the martingale condition $\mathbf{E}[D_{n,k}|\mathcal{F}_{n,k-1}] = 0$), hence

$$\mathbf{E}[\overline{D}_{n,k}^2|\mathcal{F}] - \mathbf{E}[\widetilde{D}_{n,k}^2|\mathcal{F}] = \big(\mathbf{E}[\overline{D}_{n,k}|\mathcal{F}]\big)^2 = \big(\mathbf{E}[\widehat{D}_{n,k}|\mathcal{F}]\big)^2 \le \mathbf{E}[\widehat{D}_{n,k}^2|\mathcal{F}]\,.$$

Now that we have exhibited a coupling for which $\|\widetilde{S}_n - W\| \xrightarrow{p} 0$, if $\|\widehat{S}_n - \widetilde{S}_n\| \xrightarrow{p} 0$ then by the triangle inequality for the supremum norm $\|\cdot\|$ there also exists a coupling with $\|\widehat{S}_n - W\| \xrightarrow{p} 0$ (to construct the latter coupling, since there exist non-random $\eta_n \to 0$ such that $\mathbf{P}(\|\widehat{S}_n - \widetilde{S}_n\| \ge \eta_n) \to 0$, given the coupled $\widetilde{S}_n(\cdot)$ and $W(\cdot)$ you simply construct $\widehat{S}_n(\cdot)$ per n, conditional on the value of $\widetilde{S}_n(\cdot)$ in such a way that the joint law of $(\widehat{S}_n, \widetilde{S}_n)$ minimizes $\mathbf{P}(\|\widehat{S}_n - \widetilde{S}_n\| \ge \eta_n)$ subject to the specified laws of $\widehat{S}_n(\cdot)$ and of $\widetilde{S}_n(\cdot)$). In view of Corollary 3.5.3 this implies the convergence in distribution of $\widehat{S}_n(\cdot)$ to $W(\cdot)$ (on the metric space $(C([0,1]), \|\cdot\|)$). Turning to verify that $\|\widehat{S}_n - \widetilde{S}_n\| \xrightarrow{p} 0$, recall first that $|\widehat{D}_{n,k}| \le |\widehat{D}_{n,k}|^2/\varepsilon_n$, hence for $\mathcal{F} = \mathcal{F}_{n,k-1}$,

$$|\mathbf{E}(\overline{D}_{n,k}|\mathcal{F})| = |\mathbf{E}(\widehat{D}_{n,k}|\mathcal{F})| \le \mathbf{E}(\,|\widehat{D}_{n,k}|\,|\mathcal{F}) \le \varepsilon_n^{-1}\mathbf{E}[\widehat{D}_{n,k}^2|\mathcal{F}]\,.$$

Note also that if the event $\Gamma_n = \{|D_{n,k}| < \varepsilon_n, \text{ for all } k \le n\}$ occurs, then $D_{n,k} - \widetilde{D}_{n,k} = \mathbf{E}[\overline{D}_{n,k}|\mathcal{F}_{n,k-1}]$ for all k. Therefore,

$$I_{\Gamma_n}\|\widehat{S}_n - \widetilde{S}_n\| \le I_{\Gamma_n}\sum_{k=1}^n |D_{n,k} - \widetilde{D}_{n,k}| \le \sum_{k=1}^n |\mathbf{E}(\overline{D}_{n,k}|\mathcal{F}_{n,k-1})| \le \varepsilon_n^{-1} g_n(\varepsilon_n)\,,$$

and our choice of $\varepsilon_n \to 0$ such that $\varepsilon_n^{-2} g_n(\varepsilon_n) \xrightarrow{p} 0$ implies that $I_{\Gamma_n}\|\widehat{S}_n - \widetilde{S}_n\| \xrightarrow{p} 0$.

We thus complete the proof by showing that $\mathbf{P}(\Gamma_n^c) \to 0$. Indeed, fixing n and $r > 0$, we apply Exercise 5.3.39 for the events $A_k = \{|D_{n,k}| \ge \varepsilon_n\}$ adapted to the filtration $\{\mathcal{F}_{n,k}, k \ge 0\}$, and by Markov's inequality for C.E. (see part (b) of Exercise 4.2.22), arrive at

$$\begin{aligned}\mathbf{P}(\Gamma_n^c) &= \mathbf{P}\Big(\bigcup_{k=1}^n A_k\Big) \le er + \mathbf{P}\Big(\sum_{k=1}^n \mathbf{P}(|D_{n,k}| \ge \varepsilon_n|\mathcal{F}_{n,k-1}) > r\Big) \\ &\le er + \mathbf{P}\Big(\varepsilon_n^{-2}\sum_{k=1}^n \mathbf{E}[\widehat{D}_{n,k}^2|\mathcal{F}_{n,k-1}] > r\Big) = er + \mathbf{P}(\varepsilon_n^{-2} g_n(\varepsilon_n) > r)\,.\end{aligned}$$

Consequently, our choice of ε_n implies that $\mathbf{P}(\Gamma_n^c) \le 3r$ for any $r > 0$ and all n large enough. So, upon considering $r \downarrow 0$ we deduce that $\mathbf{P}(\Gamma_n^c) \to 0$. As explained before this concludes our proof of the martingale CLT. ∎

Specializing Theorem 9.2.22 to the case of a *single* martingale $(M_\ell, \mathcal{F}_\ell)$ leads to the following corollary.

Corollary 9.2.23. *Suppose an L^2-martingale $(M_\ell, \mathcal{F}_\ell)$ starting at $M_0 = 0$, is of $\mathcal{F}_\ell$-predictable compensators such that $n^{-1}\langle M\rangle_n \xrightarrow{p} 1$ and as $n \to \infty$,*

$$n^{-1}\sum_{k=1}^n \mathbf{E}[(M_k - M_{k-1})^2; |M_k - M_{k-1}| \ge \varepsilon\sqrt{n}] \to 0\,,$$

for any fixed $\varepsilon > 0$. Then, as $n \to \infty$, the linearly interpolated, time-scaled S.P.

$$\widehat{S}_n(t) = n^{-1/2}\{M_{[nt]} + (nt - [nt])(M_{[nt]+1} - M_{[nt]})\}\,, \tag{9.2.16}$$

converge in distribution on $C([0,1])$ to the standard Wiener process.

Proof. Simply consider Theorem 9.2.22 for $M_{n,\ell} = n^{-1/2}M_\ell$ and $\mathcal{F}_{n,\ell} = \mathcal{F}_\ell$. In this case $\langle M_n\rangle_\ell = n^{-1}\langle M\rangle_\ell$ so (9.2.10) amounts to $n^{-1}\langle M\rangle_n \xrightarrow{p} 1$ and in stating the corollary we merely replaced the condition (9.2.11) by the stronger assumption that $\mathbf{E}[g_n(\varepsilon)] \to 0$ as $n \to \infty$. ∎

Further specializing Theorem 9.2.22, you are to derive next the martingale extension of Lyapunov's CLT.

Exercise 9.2.24. *Let $Z_k = \sum_{i=1}^k X_i$, where the $\mathcal{F}_k$-adapted, square-integrable $\{X_k\}$ are such that w.p.1. $\mathbf{E}[X_k|\mathcal{F}_{k-1}] = \mu$ for some non-random μ and all $k \ge 1$. Setting $V_{n,q} = n^{-q/2}\sum_{k=1}^n \mathbf{E}[|X_k - \mu|^q\,|\mathcal{F}_{k-1}]$ suppose further that $V_{n,2} \xrightarrow{p} 1$, while for some $q > 2$ non-random, $V_{n,q} \xrightarrow{p} 0$.*

(a) *Setting $M_k = Z_k - k\mu$ show that $\widehat{S}_n(\cdot)$ of (9.2.16) converges in distribution on $C([0,1])$ to the standard Wiener process.*

(b) *Deduce that* $L_n \xrightarrow{\mathcal{D}} L_\infty$, *where*

$$L_n = n^{-1/2} \max_{0 \le k \le n} \{Z_k - \frac{k}{n} Z_n\}$$

and $\mathbf{P}(L_\infty \ge b) = \exp(-2b^2)$ *for any* $b > 0$.

(c) *In case* $\mu > 0$, *set* $T_b = \inf\{k \ge 1 : Z_k > b\}$ *and show that* $b^{-1} T_b \xrightarrow{p} 1/\mu$ *when* $b \uparrow \infty$.

The following exercises present typical applications of the martingale CLT, starting with the least-squares parameter estimation for first order auto regressive processes (see Exercises 6.1.15 and 6.3.30 for other aspects of these processes).

Exercise 9.2.25 (First order auto regressive process)**.** *Consider the* $\mathbb{R}$*-valued S.P.* $Y_0 = 0$ *and* $Y_n = \alpha Y_{n-1} + D_n$ *for* $n \ge 1$, *with* $\{D_n\}$ *a uniformly bounded* $\mathcal{F}_n$*-adapted martingale difference sequence such that a.s.* $\mathbf{E}[D_k^2 | \mathcal{F}_{k-1}] = 1$ *for all* $k \ge 1$, *and* $\alpha \in (-1, 1)$ *is a non-random parameter.*

(a) *Check that* $\{Y_n\}$ *is uniformly bounded. Deduce that* $n^{-1} \sum_{k=1}^n D_k^2 \xrightarrow{a.s.} 1$ *and* $n^{-1} Z_n \xrightarrow{a.s.} 0$, *where* $Z_n = \sum_{k=1}^n Y_{k-1} D_k$.
Hint: *See part (c) of Exercise 5.3.40.*

(b) *Let* $V_n = \sum_{k=1}^n Y_{k-1}^2$. *Considering the estimator* $\widehat{\alpha}_n = \frac{1}{V_n} \sum_{k=1}^n Y_k Y_{k-1}$ *of the parameter* α, *conclude that* $\sqrt{V_n}(\widehat{\alpha}_n - \alpha) \xrightarrow{\mathcal{D}} \mathcal{N}(0, 1)$ *as* $n \to \infty$.

(c) *Suppose now that* $\alpha = 1$. *Show that in this case* $(Y_n, \mathcal{F}_n)$ *is a martingale of uniformly bounded differences and deduce from the martingale* CLT *that the two-dimensional random vectors* $(n^{-1} Z_n, n^{-2} V_n)$ *converge in distribution to* $(\frac{1}{2}(W_1^2 - 1), \int_0^1 W_t^2 dt)$ *with* $\{W_t, t \ge 0\}$ *a standard Wiener process. Conclude that in this case*

$$\sqrt{V_n}(\widehat{\alpha}_n - \alpha) \xrightarrow{\mathcal{D}} \frac{W_1^2 - 1}{2\sqrt{\int_0^1 W_t^2 dt}}.$$

(d) *Show that the conclusion of part (c) applies in case* $\alpha = -1$, *except for multiplying the limiting variable by* -1.
Hint: *Consider the sequence* $(-1)^k Y_k'$ *with* Y_k' *corresponding to* $D_k' = (-1)^k D_k$ *and* $\alpha = 1$.

Exercise 9.2.26. *Let* $L_n = n^{-1/2} \max_{0 \le k \le n} \{\sum_{i=1}^k c_i Y_i\}$, *where the* $\mathcal{F}_k$*-adapted, square-integrable* $\{Y_k\}$ *are such that w.p.1.* $\mathbf{E}[Y_k | \mathcal{F}_{k-1}] = 0$ *and* $\mathbf{E}[Y_k^2 | \mathcal{F}_{k-1}] = 1$ *for all* $k \ge 1$. *Suppose further that* $\sup_{k \ge 1} \mathbf{E}[|Y_k|^q | \mathcal{F}_{k-1}]$ *is finite a.s. for some* $q > 2$ *and* $c_k \in m\mathcal{F}_{k-1}$ *are such that* $n^{-1} \sum_{k=1}^n c_k^2 \xrightarrow{p} 1$. *Show that* $L_n \xrightarrow{\mathcal{D}} L_\infty$ *with* $\mathbf{P}(L_\infty \ge b) = 2\mathbf{P}(G \ge b)$ *for a standard normal variable* G *and all* $b \ge 0$.
Hint: *Show that* $k^{-1/2} c_k \xrightarrow{p} 0$, *then consider part (a) of Exercise 9.2.24.*

9.2.2. Law of the iterated logarithm. With W_t a centered normal variable of variance t, one expects the Brownian sample function to grow as $\sqrt{t}$ for $t \to \infty$ and $t \downarrow 0$. While this is true for fixed, non-random times, such reasoning ignores the random fluctuations of the path (as we have discussed before in the context of random walks, see Exercise 2.2.24). Accounting for these we obtain the following *law of the iterated logarithm* (LIL).

Theorem 9.2.27 (Kinchin's lil)**.** *Set* $h(t) = \sqrt{2t \log\log(1/t)}$ *for* $t < 1/e$ *and* $\widetilde{h}(t) = th(1/t)$*. Then, for standard Wiener processes* $\{W_t, t \ge 0\}$ *and* $\{\widetilde{W}_t, t \ge 0\}$*, w.p.1. the following hold:*

$$\limsup_{t \downarrow 0} \frac{W_t}{h(t)} = 1, \qquad \liminf_{t \downarrow 0} \frac{W_t}{h(t)} = -1, \tag{9.2.17}$$

$$\limsup_{t \to \infty} \frac{\widetilde{W}_t}{\widetilde{h}(t)} = 1, \qquad \liminf_{t \to \infty} \frac{\widetilde{W}_t}{\widetilde{h}(t)} = -1. \tag{9.2.18}$$

Remark. To determine the scale $h(t)$ recall the estimate $\mathbf{P}(G \ge x) = e^{-\frac{x^2}{2}(1+o(1))}$ which implies that for $t_n = \alpha^{2n}$, $\alpha \in (0,1)$ the sequence $\mathbf{P}(W_{t_n} \ge bh(t_n)) = n^{-b^2+o(1)}$ is summable when $b > 1$ but not summable when $b < 1$. Indeed, using such tail bounds we prove the lil in the form of (9.2.17) by the subsequence method you have seen already in the proof of Proposition 2.3.1. Specifically, we consider such time skeleton $\{t_n\}$ and apply Borel-Cantelli I for $b > 1$ and α near one (where Doob's inequality controls the fluctuations of $t \mapsto W_t$ by those at $\{t_n\}$), en-route to the upper bound. To get a matching lower bound we use Borel-Cantelli II for $b < 1$ and the independent increments $W_{t_n} - W_{t_{n+1}}$ (which are near W_{t_n} when α is small).

Proof. Since $\widetilde{h}(t) = th(1/t)$, by the time-inversion invariance of the standard Wiener process, it follows upon considering $\widetilde{W}_t = tW_{1/t}$ that (9.2.18) is equivalent to (9.2.17). Further, by the symmetry of this process, it suffices to prove the statement about the lim sup in (9.2.17).

Proceeding to upper bound $W_t/h(t)$, applying Doob's inequality (8.2.2) for the non-negative martingale $X_s = \exp(\theta(W_s - \theta s/2))$ (see part (a) of Exercise 8.2.7), such that $\mathbf{E}[X_0] = 1$, we find that for any $t, \theta, y \ge 0$,

$$\mathbf{P}(\sup_{s \in [0,t]} \{W_s - \theta s/2\} \ge y) = \mathbf{P}(\sup_{s \in [0,t]} X_s \ge e^{\theta y}) \le e^{-\theta y}.$$

Fixing $\delta, \alpha \in (0,1)$, consider this inequality for $t_n = \alpha^{2n}$, $y_n = h(t_n)/2$, $\theta_n = (1+2\delta)h(t_n)/t_n$ and $n > n_0(\alpha) = 1/(2\log(1/\alpha))$. Since $\exp(h(t)^2/2t) = \log(1/t)$, it follows that $e^{-\theta_n y_n} = (n_0/n)^{1+2\delta}$ is summable. Thus, by Borel-Cantelli I, for some $n_1 = n_1(\omega, \alpha, \delta) \ge n_0$ finite, w.p.1. $\sup_{s \in [0,t_n]}\{W_s - \theta_n s/2\} \le y_n$ for all $n \ge n_1$. With $\log\log(1/t)$ non-increasing on $[0, 1/e]$, we then have that for every $s \in (t_{n+1}, t_n]$ and $n \ge n_1$,

$$W_s \le y_n + \theta_n t_n/2 = (1+\delta)h(t_n) \le \frac{1+\delta}{\alpha} h(s)$$

Therefore, w.p.1. $\limsup_{s \downarrow 0} W_s/h(s) \le (1+\delta)/\alpha$. Considering $\delta_k = 1/k = 1 - \alpha_k$ and $k \uparrow \infty$ we conclude that

$$\limsup_{s \downarrow 0} \frac{W_s}{h(s)} \le 1, \quad \text{w.p.1.} \tag{9.2.19}$$

To bound below the left side of (9.2.19), consider the independent events $A_n = \{W_{t_n} - W_{t_{n+1}} \ge (1-\alpha^2)h(t_n)\}$, where as before $t_n = \alpha^{2n}$, $n > n_0(\alpha)$ and $\alpha \in (0,1)$ is fixed. Setting $x_n = (1-\alpha^2)h(t_n)/\sqrt{t_n - t_{n+1}}$, we have by the time-homogeneity and scaling properties of the standard Wiener process (see parts (b) and (d) of Exercise 9.1.1), that $\mathbf{P}(A_n) = \mathbf{P}(W_1 \ge x_n)$. Further, noting that both $x_n^2/2 = (1-\alpha^2)\log\log(1/t_n) \uparrow \infty$ and $nx_n^{-1}\exp(-x_n^2/2) \to \infty$ as $n \to \infty$, by the lower bound on the tail of the standard normal distribution (see part (a) of Exercise 2.2.24), we have that for some $\kappa_\alpha > 0$ and all n large enough,

$$\mathbf{P}(A_n) = \mathbf{P}(W_1 \geq x_n) \geq \frac{1 - x_n^{-2}}{\sqrt{2\pi} x_n} e^{-x_n^2/2} \geq \kappa_\alpha n^{-1} .$$

Now, by Borel-Cantelli II the divergence of the series $\sum_n \mathbf{P}(A_n)$ implies that w.p.1. $W_{t_n} - W_{t_{n+1}} \geq (1 - \alpha^2) h(t_n)$ for infinitely many values of n. Further, applying the bound (9.2.19) for the standard Wiener process $\{-W_s, s \geq 0\}$, we know that w.p.1. $W_{t_{n+1}} \geq -2h(t_{n+1}) \geq -4\alpha h(t_n)$ for all n large enough. Upon adding these two bounds, we have that w.p.1. $W_{t_n} \geq (1 - 4\alpha - \alpha^2) h(t_n)$ for infinitely many values of n. Finally, considering $\alpha_k = 1/k$ and $k \uparrow \infty$, we conclude that w.p.1. $\limsup_{t \downarrow 0} W_t / h(t) \geq 1$, which completes the proof. $\square$

As illustrated by the next exercise, restricted to a sufficiently sparsely spaced $\{t_n\}$, the a.s. maximal fluctuations of the Wiener process are closer to the fixed time CLT scale of $O(\sqrt{t})$, than the LIL scale $\widetilde{h}(t)$.

Exercise 9.2.28. *Show that for $t_n = \exp(\exp(n))$ and a Brownian Markov process $\{W_t, t \geq 0\}$, almost surely,*

$$\limsup_{n \to \infty} W_{t_n} / \sqrt{2 t_n \log \log \log t_n} = 1 .$$

Combining Kinchin's LIL and the representation of the random walk as samples of the Brownian motion at random times, we have the corresponding LIL of Hartman-Wintner for the random walk.

Proposition 9.2.29 (Hartman-Wintner's LIL)**.** *Suppose $S_n = \sum_{k=1}^n \xi_k$, where ξ_k are i.i.d. with $\mathbf{E}\xi_1 = 0$ and $\mathbf{E}\xi_1^2 = 1$. Then, w.p.1.*

$$\limsup_{n \to \infty} S_n / \widetilde{h}(n) = 1 ,$$

where $\widetilde{h}(n) = n h(1/n) = \sqrt{2n \log \log n}$.

Remark. Recall part (c) of Exercise 2.3.4 that if $\mathbf{E}[|\xi_1|^\alpha] = \infty$ for some $0 < \alpha < 2$ then w.p.1. $n^{-1/\alpha} |S_n|$ is unbounded, hence so is $|S_n| / \widetilde{h}(n)$ and in particular the LIL fails.

Proof. In Corollary 9.2.21 we represent the random walk as $S_n = W_{T_n}$ for the standard Wiener process $\{W_t, t \geq 0\}$ and $T_n = \sum_{k=1}^n \tau_k$ with non-negative i.i.d. $\{\tau_k\}$ of mean one. By the strong law of large numbers, $n^{-1} T_n \stackrel{a.s.}{\to} 1$. Thus, fixing $\varepsilon > 0$, w.p.1. there exists $n_0(\omega)$ finite such that $t/(1+\varepsilon) \leq T_{\lfloor t \rfloor} \leq t(1+\varepsilon)$ for all $t \geq n_0$. With $t_\ell = e^e (1+\varepsilon)^\ell$ for $\ell \geq 0$, and $V_\ell = \sup\{|W_s - W_t| : s, t \in [t_\ell, t_{\ell+3}]\}$, note that if $t \in [t_{\ell+1}, t_{\ell+2}]$ and $t \geq n_0(\omega)$, then $|W_{T_{\lfloor t \rfloor}} - W_t| \leq V_\ell$. Further, $t \mapsto \widetilde{h}(t)$ is non-decreasing, so w.p.1.

$$\limsup_{t \to \infty} \Big| \frac{S_{\lfloor t \rfloor}}{\widetilde{h}(\lfloor t \rfloor)} - \frac{W_t}{\widetilde{h}(t)} \Big| \leq \limsup_{\ell \to \infty} \frac{V_\ell}{\widetilde{h}(t_\ell)}$$

and in view of (9.2.18), it suffices to show that for some non-random $\eta_\varepsilon \downarrow 0$ as $\varepsilon \downarrow 0$,

$$\limsup_{\ell \to \infty} \frac{V_\ell}{\widetilde{h}(t_\ell)} \leq \eta_\varepsilon , \qquad \text{w.p.1.} \tag{9.2.20}$$

To this end, by the triangle inequality,

$$V_\ell \leq 2\sup\{|W_s - W_{t_\ell}| : s \in [t_\ell, t_{\ell+3}]\}\,.$$

It thus follows that for $\delta_\varepsilon = (t_{\ell+3} - t_\ell)/t_\ell = (1+\varepsilon)^3 - 1$ and all $\ell \geq 0$,

$$\begin{aligned}\mathbf{P}(V_\ell \geq \sqrt{8\delta_\varepsilon}\widetilde{h}(t_\ell)) &\leq \mathbf{P}\big(\sup_{s\in[t_\ell,t_{\ell+3}]} |W_s - W_{t_\ell}| \geq \sqrt{2\delta_\varepsilon}\widetilde{h}(t_\ell)\big)\\ &= \mathbf{P}(\sup_{u\in[0,1]} |W_u| \geq x_\ell) \leq 4\mathbf{P}(W_1 \geq x_\ell)\,,\end{aligned}$$

where $x_\ell = \sqrt{2\delta_\varepsilon}\widetilde{h}(t_\ell)/\sqrt{t_{\ell+3} - t_\ell} = \sqrt{4\log\log t_\ell} \geq 2$, the equality is by time-homogeneity and scaling of the Brownian motion and the last inequality is a consequence of the symmetry of Brownian motion and the reflection principle (see (9.1.2)). With $\ell^2\exp(-x_\ell^2/2) = (\ell/\log t_\ell)^2$ bounded above (in ℓ), applying the upper bound of part (a) of Exercise 2.2.24 for the standard normal distribution of W_1 we find that for some finite κ_ε and all $\ell \geq 0$,

$$\mathbf{P}(W_1 \geq x_\ell) \leq (2\pi)^{-1/2}x_\ell^{-1}e^{-x_\ell^2/2} \leq \kappa_\varepsilon\ell^{-2}\,.$$

Having just shown that $\sum_\ell \mathbf{P}(V_\ell \geq \sqrt{8\delta_\varepsilon}\widetilde{h}(t_\ell))$ is finite, we deduce by Borel-Cantelli I that (9.2.20) holds for $\eta_\varepsilon = \sqrt{8\delta_\varepsilon}$, which as explained before, completes the proof. $\square$

Remark. Strassen's LIL goes further than Hartman-Wintner's LIL, in characterizing the almost sure limit set (i.e., the collection of all limits of convergent subsequences in $C([0,1])$), for $\{S(n\cdot)/\widetilde{h}(n)\}$ and $S(\cdot)$ of (9.2.1), as

$$\mathcal{K} = \{x(\cdot) \in C([0,1]) : x(t) = \int_0^t y(s)ds : \int_0^1 y(s)^2 ds \leq 1\}\,.$$

While Strassen's LIL is outside our scope, here is a small step in this direction.

Exercise 9.2.30. *Show that w.p.1.* $[-1,1]$ *is the limit set of the* $\mathbb{R}$*-valued sequence* $\{S_n/\widetilde{h}(n)\}$.

9.3. Brownian path: regularity, local maxima and level sets

Recall Exercise 7.3.13 that the Brownian sample function is a.s. locally γ-Hölder continuous for any $\gamma < 1/2$ and Kinchin's LIL tells us that it is not γ-Hölder continuous for any $\gamma \geq 1/2$ and any interval $[0,t]$. Generalizing the latter irregularity property, we first state and prove the classical result of Paley, Wiener and Zygmund (see [**PWZ33**]), showing that a.s. a Brownian Markov process has *nowhere differentiable* sample functions (not even at a random time $t = t(\omega)$).

Definition 9.3.1. *For a continuous function* $f : [0,\infty) \mapsto \mathbb{R}$ *and* $\gamma \in (0,1]$*, the* upper *and* lower *(right)* γ-derivatives *at* $s \geq 0$ *are the* $\overline{\mathbb{R}}$*-valued*

$$D^\gamma f(s) = \limsup_{u\downarrow 0} u^{-\gamma}[f(s+u) - f(s)] \quad and \quad D_\gamma f(s) = \liminf_{u\downarrow 0} u^{-\gamma}[f(s+u) - f(s)]\,,$$

which always exist. The Dini derivatives *correspond to* $\gamma = 1$ *and denoted by* $D^1 f(s)$ *and* $D_1 f(s)$*. Indeed, a continuous function* f *is* differentiable from the right *at* s *if* $D^1 f(s) = D_1 f(s)$ *is finite.*

Proposition 9.3.2 (Paley-Wiener-Zygmund). *With probability one, the sample function of a Wiener process $t \mapsto W_t(\omega)$ is* nowhere differentiable. *More precisely, for $\gamma = 1$ and any $T \le \infty$,*

$$\mathbf{P}(\{\omega \in \Omega : -\infty < D_\gamma W_t(\omega) \le D^\gamma W_t(\omega) < \infty \text{ for some } t \in [0,T]\}) = 0\,. \tag{9.3.1}$$

Proof. Fixing integers $k, r \ge 1$, let

$$A_{kr} = \bigcup_{s\in[0,1]} \bigcap_{u\in[0,1/r]} \{\omega \in \Omega : |W_{s+u}(\omega) - W_s(\omega)| \le ku\}\,.$$

Note that if $-c \le D_1 W_t(\omega) \le D^1 W_t(\omega) \le c$ for some $t \in [0,1]$ and $c < \infty$, then by definition of the Dini derivatives, $\omega \in A_{kr}$ for $k = [c]+1$ and some $r \ge 1$. We thus establish (9.3.1) for $\gamma = 1$ and $T = 1$, as soon as we show that $A_{kr} \subseteq C$ for some $C \in \mathcal{F}^{\mathbf{W}}$ such that $\mathbf{P}(C) = 0$ (due to the uncountable union/intersection in the definition of A_{kr}, it is a-apriori not in $\mathcal{F}^{\mathbf{W}}$, but recall Remark 8.1.3 that we add all $\mathbf{P}$-null sets to $\mathcal{F}_0 \subset \mathcal{F}$, hence a-posteriori $A_{kr} \in \mathcal{F}$ and $\mathbf{P}(A_{kr}) = 0$). To this end we set

$$C = \bigcap_{n\ge 4r} \bigcup_{i=1}^{n} C_{n,i}$$

in $\mathcal{F}^{\mathbf{W}}$, where for $i = 1, \dots, n$,

$$C_{n,i} = \{\omega \in \Omega : |W_{(i+j)/n}(\omega) - W_{(i+j-1)/n}(\omega)| \le 8k/n \text{ for } j = 1,2,3\}\,.$$

To see that $A_{kr} \subseteq C$ note that for any $n \ge 4r$, if $\omega \in A_{kr}$ then for some integer $1 \le i \le n$ there exists $s \in [(i-1)/n, i/n]$ such that $|W_t(\omega) - W_s(\omega)| \le k(t-s)$ for all $t \in [s, s+1/r]$. This applies in particular for $t = (i+j)/n$, $j = 0,1,2,3$, in which case $0 \le t - s \le 4/n \le 1/r$ and consequently, $|W_t(\omega) - W_s(\omega)| \le 4k/n$. Then, by the triangle inequality necessarily also $\omega \in C_{n,i}$.

We next show that $\mathbf{P}(C) = 0$. Indeed, note that for each i, n the random variables $G_j = \sqrt{n}(W_{(i+j)/n} - W_{(i+j-1)/n})$, $j = 1, 2, \dots$, are independent, each having the standard normal distribution. With their density bounded by $1/\sqrt{2\pi} \le 1/2$, it follows that $\mathbf{P}(|G_j| \le \varepsilon) \le \varepsilon$ for all $\varepsilon > 0$ and consequently,

$$\mathbf{P}(C_{n,i}) = \prod_{j=1}^{3} \mathbf{P}(|G_j| \le 8kn^{-1/2}) \le (8k)^3 n^{-3/2}\,.$$

This in turn implies that $\mathbf{P}(C) \le \sum_{i\le n} \mathbf{P}(C_{n,i}) \le (8k)^3/\sqrt{n}$ for any $n \ge 4r$ and upon taking $n \to \infty$, results with $\mathbf{P}(C) = 0$, as claimed.

Having established (9.3.1) for $T = 1$, we note that by the scaling property of the Wiener process, the same applies for any finite T. Finally, the subset of Ω considered there in case $T = \infty$ is merely the increasing limit as $n \uparrow \infty$ of such subsets for $T = n$, hence also of zero probability. □

You can even improve upon this negative result as follows.

Exercise 9.3.3. *Adapting the proof of Proposition 9.3.2 show that for any fixed $\gamma > \frac{1}{2}$, w.p.1. the sample function $t \mapsto W_t(\omega)$ is nowhere γ-Hölder continuous. That is, (9.3.1) holds for any $\gamma > 1/2$.*

Exercise 9.3.4. *Let $\{X_t, t \ge 0\}$ be a stochastic process of stationary increments which satisfies for some $H \in (0,1)$ the self-similarity property $X_{ct} \overset{\mathcal{D}}{=} c^H X_t$, for all $t \ge 0$ and $c > 0$. Show that if $\mathbf{P}(X_1 = 0) = 0$ then for any $t \ge 0$ fixed, a.s.*

$\limsup_{u\downarrow 0} u^{-1}|X_{t+u} - X_t| = \infty$. *Hence, w.p.1. the sample functions $t \mapsto X_t$ are not differentiable at any fixed $t \geq 0$.*

Almost surely the sample path of the Brownian motion is locally γ-Hölder continuous for any $\gamma < \frac{1}{2}$ but not for any $\gamma > \frac{1}{2}$. Further, by the LIL its *modulus of continuity* is at least $h(\delta) = \sqrt{2\delta \log\log(1/\delta)}$. The *exact* modulus of continuity of the Brownian path is provided by the following theorem, due to Paul Lévy (see [**Lev37**]).

Theorem 9.3.5 (Lévy's modulus of continuity). *Setting $g(\delta) = \sqrt{2\delta\log(\frac{1}{\delta})}$ for $\delta \in (0,1]$, with $(W_t, t \in [0,T])$ a Wiener process, for any $0 < T < \infty$, almost surely,*

$$\limsup_{\delta\downarrow 0} \frac{\mathsf{osc}_{T,\delta}(W_\cdot)}{g(\delta)} = 1\,, \tag{9.3.2}$$

where $\mathsf{osc}_{T,\delta}(x(\cdot))$ is defined in (9.2.2).

Remark. This result does not extend to $T = \infty$, as by independence of the unbounded Brownian increments, for any $\delta > 0$, with probability one

$$\mathsf{osc}_{\infty,\delta}(W_\cdot) \geq \max_{k\geq 1}\{|W_{k\delta} - W_{(k-1)\delta}|\} = \infty\,.$$

Proof. Fixing $0 < T < \infty$, note that $g(T\delta)/(\sqrt{T}g(\delta)) \to 1$ as $\delta \downarrow 0$. Further, $\mathsf{osc}_{1,\delta}(\widetilde{W}_\cdot) = T^{-1/2}\,\mathsf{osc}_{T,T\delta}(W_\cdot)$ where $\widetilde{W}_s = T^{-1/2}W_{Ts}$ is a standard Wiener process on $[0,1]$. Consequently, it suffices to prove (9.3.2) only for $T = 1$.

Setting hereafter $T = 1$, we start with the easier lower bound on the left side of (9.3.2). To this end, fix $\varepsilon \in (0,1)$ and note that by independence of the increments of the Wiener process,

$$\mathbf{P}(\Delta_{\ell,1}(W_\cdot) \leq (1-\varepsilon)g(2^{-\ell})) = (1-q_\ell)^{2^\ell} \leq \exp(-2^\ell q_\ell)\,,$$

where $q_\ell = \mathbf{P}(|W_{2^{-\ell}}| > (1-\varepsilon)g(2^{-\ell}))$ and

$$\Delta_{\ell,r}(x(\cdot)) = \max_{j=0}^{2^\ell - r} |x((j+r)2^{-\ell}) - x(j2^{-\ell})|\,. \tag{9.3.3}$$

Further, by scaling and the lower bound of part (a) of Exercise 2.2.24, it is easy to check that for some $\ell_0 = \ell_0(\varepsilon)$ and all $\ell \geq \ell_0$,

$$q_\ell = \mathbf{P}(|G| > (1-\varepsilon)\sqrt{2\ell\log 2}) \geq 2^{-\ell(1-\varepsilon)}\,.$$

By definition $\mathsf{osc}_{1,2^{-\ell}}(x(\cdot)) \geq \Delta_{\ell,1}(x(\cdot))$ for any $x : [0,1] \mapsto \mathbb{R}$ and with $\exp(-2^\ell q_\ell) \leq \exp(-2^{\ell\varepsilon})$ summable, it follows by Borel-Cantelli I that w.p.1.

$$\mathsf{osc}_{1,2^{-\ell}}(W_\cdot) \geq \Delta_{\ell,1}(W_\cdot) > (1-\varepsilon)g(2^{-\ell})$$

for all $\ell \geq \ell_1(\varepsilon,\omega)$ finite. In particular, for any $\varepsilon > 0$ fixed, w.p.1.

$$\limsup_{\delta\downarrow 0} \frac{\mathsf{osc}_{1,\delta}(W_\cdot)}{g(\delta)} \geq 1 - \varepsilon\,,$$

and considering $\varepsilon_k = 1/k \downarrow 0$, we conclude that

$$\limsup_{\delta\downarrow 0} \frac{\mathsf{osc}_{1,\delta}(W_\cdot)}{g(\delta)} \geq 1 \quad \text{w.p.1.}$$

To show the matching upper bound, we fix $\eta \in (0,1)$ and $b = b(\eta) = (1+2\eta)/(1-\eta)$ and consider the events

$$A_\ell = \bigcap_{r \le 2^{\eta\ell}} \{\Delta_{\ell,r}(W_\cdot) < \sqrt{b}g(r2^{-\ell})\}\,.$$

By the sub-additivity of probabilities,

$$\mathbf{P}(A_\ell^c) \le \sum_{r \le 2^{\eta\ell}} \mathbf{P}(\Delta_{\ell,r}(W_\cdot) \ge \sqrt{b}g(r2^{-\ell})) \le \sum_{r \le 2^{\eta\ell}} \sum_{j=0}^{2^\ell - r} \mathbf{P}(|G_{r,j}| \ge x_{\ell,r})$$

where $x_{\ell,r} = \sqrt{2b\log(2^\ell/r)}$ and $G_{r,j} = (W_{(j+r)2^{-\ell}} - W_{j2^{-\ell}})/\sqrt{r2^{-\ell}}$ have the standard normal distribution. Since $\eta > 0$, clearly $x_{\ell,r}$ is bounded away from zero for $r \le 2^{\eta\ell}$ and $\exp(-x_{\ell,r}^2/2) = (r2^{-\ell})^b$. Hence, from the upper bound of part (a) of Exercise 2.2.24 we deduce that the r-th term of the outer sum is at most $2^\ell C(r2^{-\ell})^b$ for some finite constant $C = C(\eta)$. Further, for some finite $\kappa = \kappa(\eta)$ and all $\ell \ge 1$,

$$\sum_{r \le 2^{\eta\ell}} r^b \le \int_0^{2^{\eta\ell}+1} t^b dt \le \kappa 2^{\eta\ell(b+1)}\,.$$

Therefore, as $(1-\eta)b(\eta) - (1+\eta) = \eta > 0$,

$$\mathbf{P}(A_\ell^c) \le C\kappa 2^\ell 2^{-b\ell} 2^{\eta\ell(b+1)} = C\kappa 2^{-\eta\ell}\,,$$

and since $\sum_\ell \mathbf{P}(A_\ell^c)$ is finite, by the first Borel-Cantelli lemma, on a set Ω_η of probability one, $\omega \in A_\ell$ for all $\ell \ge \ell_0(\eta,\omega)$ finite. As you show in Exercise 9.3.6, it then follows from the continuity of $t \mapsto W_t$ that on Ω_η,

$$|W_{s+h}(\omega) - W_s(\omega)| \le \sqrt{b}g(h)[1 + \varepsilon(\eta, \ell_0(\eta,\omega), h)]\,,$$

where $\varepsilon(\eta,\ell,h) \downarrow 0$ as $h \downarrow 0$. Consequently, for any $\omega \in \Omega_\eta$,

$$\limsup_{\delta\downarrow 0} g(\delta)^{-1} \sup_{0 \le s \le s+h \le 1, h=\delta} |W_{s+h}(\omega) - W_s(\omega)| \le \sqrt{b}\,.$$

Since $g(\cdot)$ is non-decreasing on $[0,1/e]$, we can further replace the condition $h = \delta$ in the preceding inequality by $h \in [0,\delta]$ and deduce that on Ω_η

$$\limsup_{\delta\downarrow 0} \frac{\mathsf{osc}_{1,\delta}(W_\cdot)}{g(\delta)} \le \sqrt{b(\eta)}\,.$$

Taking $\eta_k = 1/k$ for which $b(1/k) \downarrow 1$ we conclude that w.p.1. the same bound also holds with $b = 1$. □

Exercise 9.3.6. *Suppose $x \in C([0,1])$ and $\Delta_{m,r}(x)$ are as in (9.3.3).*

(a) *Show that for any $m, r \ge 0$,*

$$\sup_{r2^{-m} \le |t-s| < (r+1)2^{-m}} |x(t) - x(s)| \le 4\sum_{\ell=m+1}^\infty \Delta_{\ell,1}(x) + \Delta_{m,r}(x)\,.$$

Hint: *Deduce from part (a) of Exercise 7.2.7 that this holds if in addition $t, s \in Q_1^{(2,k)}$ for some $k > m$.*

(b) *Show that for some c finite, if $2^{-(m+1)(1-\eta)} \leq h \leq 1/e$ with $m \geq 0$ and $\eta \in (0,1)$, then*

$$\sum_{\ell=m+1}^{\infty} g(2^{-\ell}) \leq cg(2^{-m-1}) \leq \frac{c}{\sqrt{1-\eta}} 2^{-\eta(m+1)/2} g(h) .$$

Hint: *Recall that $g(h) = \sqrt{2h \log(1/h)}$ is non-decreasing on $[0, 1/e]$.*

(c) *Conclude that there exists $\varepsilon(\eta, \ell_0, h) \downarrow 0$ as $h \downarrow 0$, such that if $\Delta_{\ell,r}(x) \leq \sqrt{b} g(r2^{-\ell})$ for some $\eta \in (0,1)$ and all $1 \leq r \leq 2^{\eta\ell}$, $\ell \geq \ell_0$, then*

$$\sup_{0 \leq s \leq s+h \leq 1} |x(s+h) - x(s)| \leq \sqrt{b} g(h) [1 + \varepsilon(\eta, \ell_0, h)] .$$

We take up now the study of *level sets* of the standard Wiener process

$$\mathcal{Z}_\omega(b) = \{t \geq 0 : W_t(\omega) = b\} , \tag{9.3.4}$$

for non-random $b \in \mathbb{R}$, starting with its *zero set* $\mathcal{Z}_\omega = \mathcal{Z}_\omega(0)$.

Proposition 9.3.7. *For a.e. $\omega \in \Omega$, the zero set $\mathcal{Z}_\omega$ of the standard Wiener process, is closed, unbounded, of zero Lebesgue measure and having no isolated points.*

Remark. Recall that by Baire's category theorem, any closed subset of $\mathbb{R}$ having no isolated points, must be uncountable (c.f. [**Dud89**, Theorem 2.5.2]).

Proof. First note that $(t, \omega) \mapsto W_t(\omega)$ is measurable with respect to $\mathcal{B}_{[0,\infty)} \times \mathcal{F}$ and hence so is the set $\mathcal{Z} = \cup_\omega \mathcal{Z}_\omega \times \{\omega\}$. Applying Fubini's theorem for the product measure $Leb \times \mathbf{P}$ and $h(t,\omega) = I_{\mathcal{Z}}(t,\omega) = I_{\{W_t(\omega)=0\}}$ we find that $\mathbf{E}[Leb(\mathcal{Z}_\omega)] = (Leb \times \mathbf{P})(\mathcal{Z}) = \int_0^\infty \mathbf{P}(W_t = 0) dt = 0$. Thus, the set $\mathcal{Z}_\omega$ is w.p.1. of zero Lebesgue measure, as claimed. The set $\mathcal{Z}_\omega$ is closed since it is the inverse image of the closed set $\{0\}$ under the continuous mapping $t \mapsto W_t$. In Corollary 9.1.5 we have further shown that w.p.1. $\mathcal{Z}_\omega$ is unbounded and that the continuous function $t \mapsto W_t$ changes sign infinitely many times in any interval $[0, \varepsilon]$, $\varepsilon > 0$, from which it follows that zero is an accumulation point of $\mathcal{Z}_\omega$.

Next, with $A_{s,t} = \{\omega : \mathcal{Z}_\omega \cap (s,t) \text{ is a single point}\}$, note that the event that $\mathcal{Z}_\omega$ has an isolated point in $(0,\infty)$ is the countable union of $A_{s,t}$ over $s, t \in \mathbb{Q}$ such that $0 < s < t$. Consequently, to show that w.p.1. $\mathcal{Z}_\omega$ has no isolated point, it suffices to show that $\mathbf{P}(A_{s,t}) = 0$ for any $0 < s < t$. To this end, consider the a.s. finite $\mathcal{F}_t^{\mathbf{W}}$-Markov times $R_r = \inf\{u > r : W_u = 0\}$, $r \geq 0$. Fixing $0 < s < t$, let $\tau = R_s$ noting that $A_{s,t} = \{\tau < t \leq R_\tau\}$ and consequently $\mathbf{P}(A_{s,t}) \leq \mathbf{P}(R_\tau > \tau)$. By continuity of $t \mapsto W_t$ we know that $W_\tau = 0$, hence $R_\tau - \tau = \inf\{u > 0 : W_{\tau+u} - W_\tau = 0\}$. Recall Corollary 9.1.6 that $\{W_{\tau+u} - W_\tau, u \geq 0\}$ is a standard Wiener process and therefore $\mathbf{P}(R_\tau > \tau) = \mathbf{P}(R_0 > 0) = \mathbf{P}_0(T_0 > 0) = 0$ (as shown already in Corollary 9.1.5). □

In view of its strong Markov property, the level sets of the Wiener process inherit the properties of its zero set.

Corollary 9.3.8. *For any fixed $b \in \mathbb{R}$ and a.e. $\omega \in \Omega$, the level set $\mathcal{Z}_\omega(b)$ is closed, unbounded, of zero Lebesgue measure and having no isolated points.*

Proof. Fixing $b \in \mathbb{R}$, $b \neq 0$, consider the $\mathcal{F}_t^{\mathbf{W}}$-Markov time $T_b = \inf\{s > 0 : W_s = b\}$. While proving the reflection principle we have seen that w.p.1. T_b

is finite and $W_{T_b} = b$, in which case it follows from (9.3.4) that $t \in \mathcal{Z}_\omega(b)$ if and only if $t = T_b + u$ for $u \geq 0$ such that $\widetilde{W}_u(\omega) = 0$, where $\widetilde{W}_u = W_{T_b+u} - W_{T_b}$ is, by Corollary 9.1.6, a standard Wiener process. That is, up to a translation by $T_b(\omega)$ the level set $\mathcal{Z}_\omega(b)$ is merely the zero set of $\widetilde{W}_t$ and we conclude the proof by applying Proposition 9.3.7 for the latter zero set. □

Remark 9.3.9. Recall Example 8.2.51 that for a.e. ω the sample function $W_t(\omega)$ is of unbounded *total variation* on each finite interval $[s,t]$ with $s < t$. Thus, from part (a) of Exercise 8.2.42 we deduce that on any such interval w.p.1. the sample function $W_\cdot(\omega)$ is non-monotone. Since every nonempty interval includes one with rational endpoints, of which there are only countably many, we conclude that for a.e. $\omega \in \Omega$, the sample path $t \mapsto W_t(\omega)$ of the Wiener process is monotone in no interval. Here is an alternative, direct proof of this fact.

Exercise 9.3.10. *Let* $A_n = \bigcap_{i=1}^n \{\omega \in \Omega : W_{i/n}(\omega) - W_{(i-1)/n}(\omega) \geq 0\}$ *and* $A = \{\omega \in \Omega : t \mapsto W_t(\omega)$ *is non-decreasing on* $[0,1]\}$.

(a) *Show that* $\mathbf{P}(A_n) = 2^{-n}$ *for all* $n \geq 1$ *and that* $A = \cap_n A_n \in \mathcal{F}$ *has zero probability.*

(b) *Deduce that for any interval* $[s,t]$ *with* $0 \leq s < t$ *non-random the probability that* $W_\cdot(\omega)$ *is monotone on* $[s,t]$ *is zero and conclude that the event* $F \in \mathcal{F}$ *where* $t \mapsto W_t(\omega)$ *is monotone on some non-empty open interval, has zero probability.*
Hint: *Recall the invariance transformations of Exercise 9.1.1 and that* F *can be expressed as a countable union of events indexed by* $s < t \in \mathbb{Q}$.

Our next objects of interest are the collections of local maxima and points of increase along the Brownian sample path.

Definition 9.3.11. *We say that* $t \geq 0$ *is a* point of local maximum *of* $f : [0,\infty) \mapsto \mathbb{R}$ *if there exists* $\delta > 0$ *such that* $f(t) \geq f(s)$ *for all* $s \in [(t-\delta)_+, t+\delta]$, $s \neq t$, *and a* point of strict local maximum *if further* $f(t) > f(s)$ *for any such* s. *Similarly, we say that* $t \geq 0$ *is a* point of increase *of* $f : [0,\infty) \mapsto \mathbb{R}$ *if there exists* $\delta > 0$ *such that* $f((t-h)_+) \leq f(t) \leq f(t+h)$ *for all* $h \in (0,\delta]$.

The irregularity of the Brownian sample path suggests that it has many local maxima, as we shall indeed show, based on the following exercise in real analysis.

Exercise 9.3.12. *Fix* $f : [0,\infty) \mapsto \mathbb{R}$.

(a) *Show that the set of strict local maxima of* f *is countable.*
Hint: *For any* $\delta > 0$, *the points of* $M_\delta = \{t \geq 0 : f(t) > f(s)$, *for all* $s \in [(t-\delta)_+, t+\delta], s \neq t\}$ *are isolated.*

(b) *Suppose* f *is continuous but* f *is monotone on no interval. Show that if* $f(b) > f(a)$ *for* $b > a \geq 0$, *then there exist* $b > u_3 > u_2 > u_1 \geq a$ *such that* $f(u_2) > f(u_3) > f(u_1) = f(a)$, *and deduce that* f *has a local maximum in* $[u_1, u_3]$.
Hint: *Set* $u_1 = \sup\{t \in [a,b) : f(t) = f(a)\}$.

(c) *Conclude that for a continuous function* f *which is monotone on no interval, the set of local maxima of* f *is dense in* $[0,\infty)$.

Proposition 9.3.13. *For a.e.* $\omega \in \Omega$, *the set of points of local maximum for the Wiener sample path* $W_t(\omega)$ *is a countable, dense subset of* $[0,\infty)$ *and all local maxima are strict.*

Remark. Recall that the upper Dini derivative $D^1 f(t)$ of Definition 9.3.1 is non-positive, hence finite, at every point t of local maximum of $f(\cdot)$. Thus, Proposition 9.3.13 provides a dense set of points $t \geq 0$ where $D^1 W_t(\omega) < \infty$ and by symmetry of the Brownian motion, another dense set where $D_1 W_t(\omega) > -\infty$, though as we have seen in Proposition 9.3.2, w.p.1. there is no point $t(\omega) \geq 0$ for which both apply.

PROOF. If a continuous function f has a non-strict local maximum then there exist rational numbers $0 \leq q_1 < q_4$ such that the set $\mathcal{M} = \{u \in (q_1, q_4) : f(u) = \sup_{t \in [q_1, q_4]} f(t)\}$ has an accumulation point in $[q_1, q_4)$. In particular, for some rational numbers $0 \leq q_1 < q_2 < q_3 < q_4$ the set $\mathcal{M}$ intersects both intervals (q_1, q_2) and (q_3, q_4). Thus, setting $M_{s,r} = \sup_{t \in [s,r]} W_t$, if $\mathbf{P}(M_{s_3,s_4} = M_{s_1,s_2}) = 0$ for each $0 \leq s_1 < s_2 < s_3 < s_4$, then w.p.1. every local maximum of $W_t(\omega)$ is strict. This is all we need to show, since in view of Remark 9.3.9, Exercise 9.3.12 and the continuity of Brownian motion, w.p.1. the (countable) set of (strict) local maxima of $W_t(\omega)$ is dense on $[0, \infty)$. Now, fixing $0 \leq s_1 < s_2 < s_3 < s_4$ note that $M_{s_3,s_4} - M_{s_1,s_2} = Z - Y + X$ for the mutually independent $Z = \sup_{t \in [s_3,s_4]}\{W_t - W_{s_3}\}$, $Y = \sup_{t \in [s_1,s_2]}\{W_t - W_{s_2}\}$ and $X = W_{s_3} - W_{s_2}$. Since $g(x) = \mathbf{P}(X = x) = 0$ for all $x \in \mathbb{R}$, we are done as by Fubini's theorem, $\mathbf{P}(M_{s_3,s_4} = M_{s_1,s_2}) = \mathbf{P}(X - Y + Z = 0) = \mathbf{E}[g(Y - Z)] = 0$. □

Remark. While proving Proposition 9.3.13 we have shown that for any countable collection of disjoint intervals $\{I_i\}$, w.p.1. the corresponding maximal values $\sup_{t \in I_i} W_t$ of the Brownian motion must all be distinct. In particular, $\mathbf{P}(W_q = W_{q'}$ for some $q \neq q' \in \mathbb{Q}) = 0$, which of course does not contradict the fact that $\mathbf{P}(W_0 = W_t$ for uncountably many $t \geq 0) = 1$ (as implied by Proposition 9.3.7).

Here is a remarkable contrast with Proposition 9.3.13, showing that the Brownian sample path has no point of increase (try to imagine such a path!).

Theorem 9.3.14 (DVORETZKY, ERDÖS, KAKUTANI)**.** *Almost every sample path of the Wiener process has no point of increase (or decrease).*

For the proof of this result, see [**MP09**, Theorem 5.14].

Bibliography

[And1887] Désiré André, *Solution directe du probléme résolu par M. Bertrand*, Comptes Rendus Acad. Sci. Paris, **105**, (1887), 436–437.

[Bil95] Patrick Billingsley, *Probability and measure*, third edition, John Wiley and Sons, 1995.

[Bre92] Leo Breiman, *Probability*, Classics in Applied Mathematics, Society for Industrial and Applied Mathematics, 1992.

[Bry95] Wlodzimierz Bryc, *The normal distribution*, Springer-Verlag, 1995.

[Doo53] Joseph Doob, *Stochastic processes*, Wiley, 1953.

[Dud89] Richard Dudley, *Real analysis and probability*, Chapman and Hall, 1989.

[Dur10] Rick Durrett, *Probability: Theory and Examples*, fourth edition, Cambridge U Press, 2010.

[DKW56] Aryeh Dvortzky, Jack Kiefer and Jacob Wolfowitz, *Asymptotic minimax character of the sample distribution function and of the classical multinomial estimator*, Ann. Math. Stat., **27**, (1956), 642–669.

[Dyn65] Eugene Dynkin, *Markov processes*, volumes 1-2, Springer-Verlag, 1965.

[Fel71] William Feller, *An introduction to probability theory and its applications, volume II*, second edition, John Wiley and sons, 1971.

[Fel68] William Feller, *An introduction to probability theory and its applications, volume I*, third edition, John Wiley and sons, 1968.

[Fre71] David Freedman, *Brownian motion and diffusion*, Holden-Day, 1971.

[GS01] Geoffrey Grimmett and David Stirzaker, *Probability and random processes*, 3rd ed., Oxford University Press, 2001.

[Hun56] Gilbert Hunt, *Some theorems concerning Brownian motion*, Trans. Amer. Math. Soc., **81**, (1956), 294–319.

[KaS97] Ioannis Karatzas and Steven E. Shreve, *Brownian motion and stochastic calculus*, Springer Verlag, third edition, 1997.

[Lev37] Paul Lévy, *Théorie de l'addition des variables aléatoires*, Gauthier-Villars, Paris, (1937).

[Lev39] Paul Lévy, *Sur certains processus stochastiques homogénes*, Compositio Math., **7**, (1939), 283–339.

[KT75] Samuel Karlin and Howard M. Taylor, *A first course in stochastic processes*, 2nd ed., Academic Press, 1975.

[Mas90] Pascal Massart, *The tight constant in the Dvoretzky-Kiefer-Wolfowitz inequality*, Ann. Probab. **18**, (1990), 1269–1283.

[MP09] Peter Mörters and Yuval Peres, *Brownian motion*, Cambridge University Press, 2010.

[Num84] Esa Nummelin, *General irreducible Markov chains and non-negative operators*, Cambridge University Press, 1984.

[Oks03] Bernt Oksendal, *Stochastic differential equations: An introduction with applications*, 6th ed., Universitext, Springer Verlag, 2003.

[PWZ33] Raymond E.A.C. Paley, Norbert Wiener and Antoni Zygmund, *Note on random functions*, Math. Z. **37**, (1933), 647–668.

[Pit56] E. J. G. Pitman, *On the derivative of a characteristic function at the origin*, Ann. Math. Stat. **27** (1956), 1156–1160.

[SW86] Galen R. Shorak and Jon A. Wellner, *Empirical processes with applications to statistics*, Wiley, 1986.

[Str93] Daniel W. Stroock, *Probability theory: an analytic view*, Cambridge university press, 1993.

[Wil91] David Williams, *Probability with martingales*, Cambridge university press, 1991.

Index