MATH 151. Course Summary. Winter 2000.

Section 1. Basic Concepts.

Definitions of events, outcome spaces Ω, probabilities ($P(A)$).

Equally likely outcomes.

Counting definition of probability: $P(A) = \#(A)/\#(\Omega)$.

Counting rules, combinations and permutations.

Axioms of probability:

(1) Non-negativity: $P(A) \geq 0$ for all events A;

(2) Total one: $P(\Omega) = 1$;

(3) Addition: for a partition $B_1, \cdots, B_n$ of $B, P(B) = \sum_{i=1}^{n} P(B_i)$.

Complement Rule: $P(A^c) = 1 - P(A)$.

Difference rule: If $A \subseteq B$ then $P(A) \leq P(B)$ and $P(B \cap A^c) = P(B) - P(A)$.

Inclusion-Exclusion rule: $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.

DeMorgan's Rules: $(A \cap B)^c = A^c \cup B^c, \quad (A \cup B)^c = A^c \cap B^c$.

Boole's inequality.

Definition of conditional probability: $P(A \mid B) = P(A \cap B)/P(B)$.

Multiplication rule: $P(A \cap B) = P(A \mid B)P(B)$.

Rule of average proportions: For a partition $B_1, \cdots, B_n$ of Ω,

$$P(A) = P(A|B_1)P(B_1) + \cdots + P(A|B_n)P(B_n).$$

Bayes' Rule: If $B_1, \cdots, B_n$ partition Ω then

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}.$$

Independence: Events A and B are independent if $P(A \cap B) = P(A)P(B)$.

Tree diagrams.

Multiplication rule for n events.

Pairwise independence and mutual independence.

Section 2. Binomial distribution.

Definition of binomial distribution: If X is distributed binomial (n, p) then

$P(X = k) = \binom{n}{k} p^k q^{n-k}$ for $k = 0, \cdots, n$; $q = 1 - p$.

Definition of binomial coefficients and Pascal's triangle.

Expected number of successes for binomial (n, p) is np.

Normal approximation to the binomial:

• Standard Normal density: $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$, for $-\infty < x < \infty$.

• Standard Normal distribution: $\Phi(z) = \int_{-\infty}^{z} \phi(x)dx$.

• If X is distributed as Normal (μ, σ^2) than $P(a \leq X \leq b) = \Phi(\frac{b-\mu}{\sigma}) - \Phi(\frac{a-\mu}{\sigma})$.

• If X is distributed as Normal (μ, σ^2) then $A = (X - \mu)/\sigma$ has a Standard Normal distribution.

• $\Phi(z) + \Phi(-z) = 1$.

If X has a binomial (n, p) distribution and np is large, then $P(a \leq X \leq b) \approx \Phi(\frac{b+\frac{1}{2}-\mu}{\sigma}) - \Phi(\frac{a-\frac{1}{2}-\mu}{\sigma})$.

Poisson Approximation to the binomial:

• Poisson distribution with mean $\mu : P(X = k) = \frac{e^{-\mu}\mu^k}{k!}$.

If X has a binomial (n, p) distribution and for large n, np is small, then X has an approximately Poisson distribution with mean np.

Hypergeometric distribution: Used for sampling without replacement. In a population of N individuals containing G "good" individuals and B "bad" individuals, if a sample of size n is chosen without replacement the probability of getting g "good" elements and $b = n - g$ "bad" elements is $\frac{\binom{G}{g}\binom{B}{b}}{\binom{N}{n}}$.

Normal Approximation to the hypergeometric.

Section 3. Random Variables.

Definition of random variables, univariate distributions $P(X = x)$, joint distributions $P(X = x, Y = y) = P(x, y)$.

Marginal distributions: $P(X = x) = \sum_{\text{all } y} P(x, y), P(Y = y) = \sum_{\text{all } x} P(x, y)$.

Independence: $P(X = x, Y = y) = P(X = x)P(Y = y)$.

Properties of expectation:

• Expectation of a constant is constant.

• If I_A is an indicator function for the set A then $E(I_A) = P(A)$.

• Definition: $E\{g(X)\} = \sum_{\text{all } x} g(x)P(X = x)$.

• $E(aX + b) = aE(X) + b$.

• $E(X + Y) = E(X) + E(Y)$.

• If X and Y are independent $E(XY) = E(X)E(Y)$.

Tail Sum Formula: If X takes on values $\{0, \cdots, n\}$ then $E(X) = \sum_{k=1}^{n} P(X \geq k)$.

Markov's Inequality: If $X \geq 0$ then $P(X \geq a) \leq E(X)/a$ for every $a > 0$.

Definition of Variance: $Var(X) = E\{(X - \mu)^2\}$.

Computational formula for variance: $Var(X) = E(X^2) - \{E(X)\}^2$.

Scaling and Shifting: $Var(aX + b) = a^2 Var(X)$.

Standardized Random Variable: If X has mean μ and variance σ^2 then $Z = (X - \mu)/\sigma$ has mean 0 and variance 1.

Chebychev's inequality: For any random variable X with mean μ and variance σ^2, and for any $k \geq 0, P(|X - \mu| \geq k\sigma) \leq 1/k^2$.

Addition Rule for variances: If X and Y are independent, $Var(X+Y) = Var(X) + Var(Y)$.

If $X_1, \cdots, X_n$ are i.i.d. random variables with mean μ and variance σ^2 and $\bar{X}_n = S_n/n = (X_1 + \cdots + X_n)/n$, then $E(\bar{X}_n) = \mu$ and $Var(\bar{X}_n) = \sigma^2/n$.

Law of Large Numbers: If $X_1, \cdots, X_n$ is a sequence of i.i.d. random variables with mean μ and variance σ^2, then for every $\epsilon > 0, P(|\bar{X}_n - \mu| < \epsilon) \to 1$ as $n \to \infty$.

Normal Approximation: $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ has an approximate Standard Normal distribution, provided $\sigma^2 < \infty$.

Moment generating function $g(s) = E(e^{sX})$, $g'(0) = E(X)$, $g''(0) = E(X^2)$ provided $g(s)$ finite for small $|s|$.

Discrete distributions:

Infinite sum rule: If an event A partitions into an infinite sequence of mutually exclusive sets $\{A_i\}_{i=1}^{\infty}$ then $P(A) = \sum_i P(A_i)$.

Expectation: $E(X) = \sum_x xP(X = x)$ provided the absolute sum converges.

• Geometric distribution (p). Time until first success in a sequence of Bernoulli trials.

⊙ $P(T = i) = q^{i-1}p$ for $i = 1, 2, \cdots; \ q = 1 - p$.

⊙ $E(T) = 1/p, Var(T) = q/p^2$.

• Negative Binomial distribution (r, p): Time to the r'th success in a sequence of Bernoulli trials.

⊙ $P(T_r = t) = \binom{t-1}{r-1} p^r q^{t-r}$ for $t = r, r+1, \cdots$.

⊙ $E(T_r) = r/p, Var(T_r) = rq/p^2$.

• Poisson distribution. See Section 2. $E(N) = \mu, Var(N) = \mu$.

⊙ Sum of independent Poissons is Poisson.

⊙ N_1, N_2 independent Poissons, N_1 conditioned on $N_1 + N_2 = n$ is Binomial.

• Poisson Random Scatter (with intensity λ).

⊙ Thinning a Poisson Scatter (new intensity λp).

Section 4. Continuous distributions.

Definition of a density $f(x) = \lim_{dx \to 0} P(X \in dx)/dx$.

If X is a continuous r.v. with density f then $P(a < X < b) = \int_a^b f(x)dx$.

$\int_{-\infty}^{\infty} f(x)dx = 1, f(x) \geq 0$.

$E\{g(X)\} = \int_{-\infty}^{\infty} g(x)f(x)dx$ provided $g(x)f(x)$ is absolutely integrable.

• Uniform distribution on $(a, b) : f(x) = 1/(b - a)$ if $x \in (a, b)$ and 0 otherwise:

⊙ $P(c < X < d) = (d - c)/(b - a)$ if $c, d \in (a, b)$.

⊙ $E(U) = \frac{a+b}{2}, Var(U) = \frac{(b-a)^2}{12}$.

• Normal distribution. See Section 2.

⊙ Central Limit Theorem. If $X_1, \cdots, X_n$ are i.i.d. random variables with mean μ and variance $\sigma^2 < \infty$, and if $S_n = X_1 + \cdots + X_n$ then $E(S_n) = n\mu$ and $Var(S_n) = n\sigma^2$. If $Z_n = (S_n - n\mu)/(\sqrt{n}\sigma)$ is the Standardized sum, then $\lim_{n\to\infty} P(a < Z_n < b) = \Phi(b) - \Phi(a)$ for $a < b$.

• Exponential distribution with rate λ:

⊙ $f(t) = \lambda e^{-\lambda t}$, if $t \geq 0$ and 0 otherwise.

⊙ $E(T) = 1/\lambda, Var(T) = 1/\lambda^2$.

⊙ Memoryless property. If T is distributed as exponential(λ) then

$$P(T > t + s | T > t) = P(T > s), \quad t \geq 0, \quad s \geq 0.$$

• Poisson arrival processes. The number of particles $N(I)$ arriving in a time interval I of length t is a Poisson random variable with mean λt. This is called a Poisson arrival process with rate λ. The times between arrivals are independent exponential random variables with same rate λ.

• Gamma distribution (r, λ). Time until r'th arrival in a Poisson arrival process with rate λ. Sum of r i.i.d. exponential(λ).

⊙ $f(t) = e^{-\lambda t}(\lambda t)^{r-1}\lambda/(r-1)!$ for $t \geq 0$, and 0 otherwise.

⊙ $E(T_r) = r/\lambda, Var(T_r) = r/\lambda^2$.

Definition of Distribution Functions: $F(x) = P(X \leq x)$.

Interval Probabilities: $P(a < X \leq b) = F(b) - f(a)$.

Rules for distribution functions:

(1) $F(x) \geq 0, F(-\infty) = 0, F(\infty) = 1$;

(2) $F(x)$ is a non-decreasing function of x;

(3) $F(x)$ is a right-continuous function of x.

For continuous distributions, $F(x)$ is the area under the density curve $f(x)$ up to $x : F(x) = \int_{-\infty}^{x} f(x)dx$. Conversely, $f(x) = \frac{d}{dx}F(x)$.

Change of variable formula: Let X be a continuous random variable with density $f_X(x)$.

(1) Linear functions. If $Y = aX + b$ then the density $f_Y(y)$ of Y is $f_Y(y) = f_X(\frac{y-b}{a})/|a|$.

(2) One-to-one differentiable functions. If $Y = g(X)$ where g is strictly monotonic then $f_Y(y) = f_X(g^{-1}(y))/|\frac{dy}{dx}|$. If g is strictly increasing $F_Y(y) = F_X(g^{-1}(y))$.

(3) Functions that are not one-to-one. $f_Y(y) = \sum_{\{x:g(x)=y\}} f_X(x)/|\frac{dy}{dx}|$.

(4) Maxima and minima. If $X_1, \cdots, X_n$ are independent random variables with distribution functions $F_{X_1}, \cdots, F_{X_n}$, and $X_{max} = \max\{X_1, \cdots, X_n\}$, $X_{min} = \min\{X_1, \cdots, X_n\}$, then $F_{X_{max}}(x) = F_{X_1}(x) \cdots F_{X_n}(x)$, and $F_{X_{min}}(x) = 1 - \{1 - F_{X_1}(x)\} \cdots \{1 - F_{X_n}(x)\}$.

If the X_i's are i.i.d. with distribution F, then $F_{X_{max}} = \{F(x)\}^n$ and $F_{X_{min}} = 1-\{1-F(x)\}^n$. Therefore, $f_{X_{max}}(x) = n\{F(x)\}^{n-1}f(x)$ and $f_{X_{min}}(x) = n\{1 - F(x)\}^{n-1}f(x)$.

• Lognormal distribution. $Y = e^X$, X is Normal random variable.

⊙ If $X_1, \cdots, X_n$ are positive i.i.d. random variables, $\mu = E(\ln\ X)$ and $\sigma^2 = VAR(\ln\ X) < \infty$ and if $W_n = X_1 \cdot X_2 \cdot X_n$ then $E(\ln\ W_n) = n\mu$ and $Var(\ln\ W_n) = n\sigma^2$. For $\ln X$ Standardized random variable $\lim_{n\to\infty} P(a < W_n < b) = \Phi(\ln b) - \Phi(\ln a)$ for $a < b$.

Section 5. Continuous Joint Distributions.

• Uniform distribution on the plane. A random point (X, Y) in the plane has a uniform distribution on D, where D is a finite region of the plane with finite area if

(i) (X, Y) is certain to lie in D;

(ii) For a subregion C of D, $P((X, Y) \in C) = area(C)/area(D)$.

Definition of joint density: $f(x, y) = \lim_{dx\to 0, dy\to 0} P(X \in dx, Y \in dy)/(dx\ dy)$.

Probability formula. The probability that (X, Y) lies in a region B of the plane is the volume under the joint density confined by B; that is $P((X, Y) \in B) = \iint_B f(x, y)dx\ dy$.

Marginal densities: $f_X(x) = \int_{-\infty}^{\infty} f(x, y)dy$, $f_Y(y) = \int_{\infty}^{\infty} f(x, y)dx$.

Independence: If X and Y are independent then $f(x, y) = f_X(x)f_Y(y)$.

Expectation: $E\{g(X, Y)\} = \iint g(x, y)f(x, y)dx\ dy$.

Independent Normal variables. If X and Y are independent Standard Normal random variables then $\phi(x, y) = \frac{1}{2\pi}e^{-\frac{1}{2}(x^2+y^2)}$, for $(x, y) \in (-\infty, \infty)^2$.

Rayleigh distribution: $R^2 = X^2 + Y^2$. $f_R(r) = re^{-\frac{1}{2}r^2}$ for $r \geq 0$ and 0 otherwise. R^2 is exponential(1/2).

Linear combination of Normals: If X and Y are independent Normal random variables with means λ and μ and variances σ^2 and τ^2 respectively, then $Z = \alpha X + \beta Y$ is Normally distributed with mean $\alpha\lambda + \beta\mu$ and variance $\alpha^2\sigma^2 + \beta^2\tau^2$.

Operations: If X has density f_X and Y has density f_Y then $Z = X + Y$ has density on the line $f_Z(z) = \int_{-\infty}^{\infty} f(x, z - x)dx$.

If X and Y are independent then $f_Z(z) = \int_{-\infty}^{\infty} f_X(x) f_Y(z - x)dx$.

Section 6. Conditional distributions and expectations.

Conditional distribution (discrete case): $P(Y = y|X = x) = P(X = x, Y = y)/P(X = x)$.

Conditional density: $f_Y(y|X = x) = f_{X,Y}(x, y)/f_X(x)$.

Multiplication rules: $P(X = x, Y = y) = P(X = x)P(Y = y|X = x)$

and $f_{X,Y}(x, y) = f_X(x) f_Y(y|X = x)$.

Average conditional probability:

$P(B) = \sum_{all\ x} P(B|X = x)P(X = x)$, $P(Y = y) = \sum_{all\ x} P(Y = y|X = x)P(X = x)$

Or $P(B) = \int P(B|X = x) f_X(x)dx$, $f_Y(y) = \int f_Y(y|X = x) f_X(x)dx$.

Bayes' Rule: $P(X = x|Y = y) = P(Y = y|X = x)P(X = x)/P(Y = y)$

and $f_X(x|Y = y) = f_Y(y|X = x) f_X(x)/f_Y(y)$.

Independence: Random variables X and Y are independent if for all subsets B in the range of Y, $P(Y \in B|X = x) = P(Y \in B)$.

Conditional expectation: $E(g(Y)|X = x) = \sum_{all\ y} g(y)P(Y = y|X = x)$,

or $E(g(Y)|X = x) = \int g(y) f_Y(y|X = x)dy$.

Conditional expectation $E(g(Y)|X)$ as a random variable.

Average conditional expectation: $E(Y) = E[E(Y|X)]$.

That is $E(Y) = \sum_{all\ x} E(Y|X = x)P(X = x)$, or $E(Y) = \int E(Y|X = x) f_X(x)dx$.