Lecture 1

Casella and Berger

Chapters 1, 2.1, 3.1, 3.2, 3.3

Basic Concepts

Probability space: $(S, \mathcal{B}, P)$

- S: sample space
- $\mathcal{B}$: σ-algebra (or Borel field)
- P: probability function

Def 1.1 Sample space S is the set of all possible outcomes of a particular experiment.

Def 1.2 A collection of subsets of S is called a σ-algebra or Borel field, denoted by $\mathcal{B}$, if

a. $\varnothing \in \mathcal{B}$;

b. If $A \in \mathcal{B}$, then $A^c \in \mathcal{B}$;

c. If $A_1, A_2, \cdots \in \mathcal{B}$, then $\cup_{i=1}^{\infty} A_i \in \mathcal{B}$.

Def 1.3 Given S and $\mathcal{B}$, a *probability function* is a function P with domains $\mathcal{B}$ that satisfies

a. $P(A) \geq 0$ for all $A \in \mathcal{B}$.

b. $P(S) = 1$.

c. If $A_1, A_2, \cdots \in \mathcal{B}$ are pairwise disjoint, then $P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$.

The Calculus of Probabilities

Theorem 1.1 If P is a probability function and A is any set in $\mathcal{B}$, then

a. $P(\varnothing) = 0$.

b. $P(A) \leq 1$.

c. $P(A^c) = 1 - P(A)$.

Theorem 1.2 If P is a probability function, then

a. $P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$ for any partition $C_1, C_2, ...$;

b. $P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$ for any sets $A_1, A_2, ...$

Conditional Probability and Independence

Def 1.4 If A and B are events in S, and $P(B) > 0$, then the *conditional probability* of A given B, written $P(A|B)$, is

$$P(A|B) = \frac{P(A \cap B)}{P(B)}.$$

Theorem 1.3 (Bayes' Rule) Let $A_1, A_2, ...$ be a partition of S and let B be any set. Then, for each $i = 1, 2, ...,$

$$P(A_i|B) = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^{\infty} P(B|A_j)P(A_j)}.$$

Def 1.5 Two events, A and B, are *statistically independent* if

$$P(A \cap B) = P(A)P(B).$$

Theorem 1.4 If A and B are independent events, then the following are also independent:

a. A and B^c,

b. A^c and B,

c. A^c and B^c.

Randome Variables

Def 1.6 A *random variable* is a function from a sample space S into real numbers.

Def 1.7 The *cumulative distribution function* or *cdf* of a random variable X, denoted by $F_X(x)$, is defined by

$$F_X(x) = P_X(X \leq x), \text{for all } x.$$

How to determine whether a function is a cdf?

The function $F(x)$ is a cdf if and only if:

a. $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$.

b. $F(x)$ is a nondecreasing function of x.

c. $F(x)$ is right-continuous; that is, for every number x_0, $\lim_{x\downarrow x_0} F(x) = F(x_0)$.

Continuous r.v. vs. Discrete r.v.

Def 1.8 A random variable X is *continuous* if $F_X(x)$ is a continuous function of x. A random variable is *discrete* if $F_X(x)$ is a step function of x.

Def 1.9 The random variables X and Y are *identically distributed* if, for every set $A \in \mathcal{B}$, $P(X \in A) = P(Y \in A)$.

Theorem 1.5 The following two statements are equivalent:
a. The random variables X and Y are identically distributed.
b. $F_X(x) = F_Y(x)$ for every x.

pmf vs. pdf

Def 1.10 The *probability mass function (pmf)* of a discrete random variable X is given by

$$f_X(x) = P(X = x) \text{ for all } x.$$

Def 1.11 The *probability density function (pdf)*, $f_X(x)$, of a continuous random variable X is the function that satisfies

$$F_X(x) = \int_{-\infty}^{x} f_X(t)dt \text{ for all } x.$$

How to determine whether a function is a pmf (or pdf)?

A function $f_X(x)$ is a pdf (or pmf) of a random variable X if and only if

a. $f_X(x) \geq 0$ for all x.

b. $\sum_x f_X(x) = 1$ (pmf) or $\int_{-\infty}^{\infty} f_X(x)dx = 1$ (pdf).

Examples of Discrete Distributions

Binomial distribution

- *Binomial(n,p)*:

$$P(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, k = 0, ..., n.$$

- *Bernoulli distribution*: a special case of binomial distribution, $Binomial(1, p)$

Discrete Uniform Distribution

$$P(X = k|N) = \frac{1}{N}, k = 1, 2, ..., N,$$

where N is a specified integer.

Hypergeometric Distribution

$$P(X = k|N, M, K) = \frac{\binom{M}{k}\binom{N-M}{K-x}}{\binom{N}{K}}, k = 0, 1, ..., K.$$

Poisson Distribution

$$P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, ...,$$

where λ is the intensity parameter.

Negative Binomial Distribution

$$P(X = x|r, p) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, x = r, r+1, ...,$$

where r is a positive fixed integer.

Geometric Distribution

$$P(X = x|p) = p(1-p)^{x-1}, x = 1, 2, ...$$

Examples of Continuous Distributions

Uniform Distribution

$$f(x|a,b) = \begin{cases} \frac{1}{b-a} & \text{if } x \in [a,b] \\ 0 & \text{otherwise} \end{cases}$$

Normal Distribution

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(x-\mu)^2/(2\sigma^2)}, -\infty < x < \infty.$$

Gamma Distribution

$$f(x|\alpha, \beta) = \frac{1}{\Gamma(\alpha)\beta^{\alpha}} x^{\alpha-1} e^{-x/\beta}, 0 < x < \infty, \alpha > 0, \beta > 0.$$

Special cases

- *exponential distribution* with scale parameter β when $\alpha = 1$.
- *chi square distribution with p degrees of freedom* when $\alpha = p/2$ and $\beta = 2$.

Beta Distribution

$$f(x|\alpha,\beta) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1}, 0 < x < 1, \alpha > 0, \beta > 0,$$

where $B(\alpha,\beta)$ denotes the beta function

$$B(\alpha,\beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}dx.$$

Cauchy Distribution

$$f(x|a,b) = \frac{1}{b\pi}\frac{1}{1+(x-a)^2/b^2}, -\infty < x < \infty, \infty < a < \infty, b > 0.$$

Double Exponential Distribution

$$f(x|\mu,\sigma) = \frac{1}{2\sigma} e^{-|x-\mu|/\sigma}, -\infty < x < \infty, -\infty < \mu < \infty, \sigma > 0.$$

Log-normal Distribution

$$f(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{x} e^{-(\log x - u)^2/(2\sigma^2)}, 0 < x < \infty, -\infty < \mu < \infty, \sigma > 0.$$

Transformations

If X is a random variable with cdf $F_X(x)$, for a function of X, say $g(X)$, what is the cdf of the new random variable $Y = g(X)$?

For any set A, we have

$$\begin{aligned} P(Y \in A) &= P(g(X) \in A) \\ &= P(\{x \in \mathcal{X} : g(x) \in A\}) \\ &= P(X \in g^{-1}(A)), \end{aligned}$$

where $\mathcal{X}$ is the sample space of X.

Theorem 1.6 Let X have cdf $F_X(x)$, let $Y = g(X)$, and let $\mathcal{X}$and $\mathcal{Y}$ be the sample space of X and Y.

a. If g is an increasing function on $\mathcal{X}$, $F_Y(y) = F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$.

b. If g is a decreasing function on $\mathcal{X}$ and X is a continuous random variable, $F_Y(y) = 1 - F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$.

Theorem 1.7 Let X have pdf $f_X(x)$ and let $Y = g(X)$, where g is a monotone function. Suppose that $f_X(x)$ is continuous on $\mathcal{X}$ and that $g^{-1}(y)$ has a continuous derivatives on $\mathcal{Y}$. Then

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| & y \in \mathcal{Y} \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 1.8 (extension of Theorem 1.7) Suppose there exists a partition, $A_0, A_1, ..., A_k$, of $\mathcal{X}$ such that $P(X \in A_0) = 0$ and $f_X(x)$ is continuous in each A_i. Further, suppose there exist functions $g_1(x), ..., g_k(x)$, defined on $A_1, ..., A_k$, respectively, satisfying

a. $g(x) = g_i(x)$, for $x \in A_i$,

b. $g_i(x)$ is monotone on A_i,

c. the set $\mathcal{Y} = \{y : y = g_i(x) \text{ for some } x \in A_i\}$ is the same for each $i = 1, ..., k$,

d. $g^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$.

Then

$$f_Y(y) = \begin{cases} \sum_{i=1}^{k} f_X(g_i^{-1}(y)) \left| \frac{d}{dy} g_i^{-1}(y) \right| & y \in \mathcal{Y} \\ 0 & \text{otherwise.} \end{cases}$$

Theorem 1.9 (Probability integral transformation) Let X have continuous cdf $F_X(x)$ and define the random variable Y as $Y = F_X(X)$. Then $Y \sim uniform(0, 1)$.

Application in generating a random number

If Y is a continuous random variable with cdf F_Y, then the random variable $F_Y^{-1}(U)$, where $U \sim uniform(0, 1)$, has distribution F_Y.