Introduction to Biostatistical Theory

Robert Parker

Principle aim of this course

To understand the basics about parametric estimation and hypothesis testing and their applications.

- Briefly review probability and random variables.
- Properties of statistics, in particular sufficient statistics;
- Point estimators and measures for goodness-of-estimation.
- Hypothesis testing and confidence intervals.
- Asymptotic (large sample size) theory

Definition (1.1)

A **set** is a collection of objects, which are called elements of the set. A set is typically denoted by capital letters A, B, C,...

Definition (1.2)

Let A and B be sets. A is said to be a **subset** of B if and only if every element of A is an element if B, i.e. $x \in A \implies x \in B$. This is denoted by $A \subseteq B$.

Operations on Sets.

- Union: $A \cup B = \{x : x \in A \text{ or } x \in B\}$.
- Intersection: $A \cap B = \{x : x \in A \text{ and } x \in B\}$
- Complementation: $A^c = \{x : x \notin A\}$ ($\bar{A}$ is often used, too.)

Some equations:

- Commutativity: $A \cup B = B \cup A$, $A \cap B = B \cap A$.
- Associativity: $$A \cup (B \cup C) = (A \cup B) \cup C,$$ $$A \cap (B \cap C) = (A \cap B) \cap C.$$
- Distributive Law: $$A \cap (B \cup C) = (A \cap B) \cup (A \cap C),$$ $$A \cup (B \cap C) = (A \cup B) \cap (A \cup C).$$
- DeMorgan's Law: $(A \cup B)^c = A^c \cap B^c$, $(A \cap B)^c = A^c \cup B^c$.

Proof for $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$

Definition (1.3)

Two sets A and B are disjoint (or mutually exclusive) if $A \cap B = \emptyset$. Events A_1, A_2, ... are pairwise disjoint if $A_i \cap A_j = \emptyset$ for all $i \neq j$.

Definition (1.4)

$A_1, A_2, \ldots$ form a *partition* of S if they are pairwise disjoint and $\cup_{i=1}^{\infty} A_i = S$.

Definition (1.5)

The limit of a sequence of sets $A_1, A_2, \ldots$ (subsets of S):

- Infimum: $\liminf_{n\to\infty} A_n = \cup_{n=1}^{\infty} \cap_{i=n}^{\infty} A_i$.
- Supremum: $\limsup_{n\to\infty} A_n = \cap_{n=1}^{\infty} \cup_{i=n}^{\infty} A_i$.
- $\liminf_{n\to\infty} A_n \subseteq \limsup_{n\to\infty} A_n$.
- If $\liminf_{n\to\infty} A_n = \limsup_{n\to\infty} A_n$, then $\lim_{n\to\infty} A_n$ exists.

Definition (1.6)

The sequence of sets B_n is said to be decreasing (or nonincreasing) if $B_{n+1} \subseteq B_n$ for all n.

Claim: $\lim_{n\to\infty} A_n = \cap_{n=1}^{\infty} A_n$

Definition

The sequence of sets B_n is said to be increasing (or nondecreasing) if $B_n \subseteq B_{n+1}$ for all n.

Claim: $\lim_{n\to\infty} A_n = \cup_{n=1}^{\infty} A_n$

Limit of a Sequence of Sets

What are the limits of the following sequence of sets?

1. $A_n = (0, 2 + 1/n),\ n = 1, 2, \ldots$
2. $A_n = (0, 2 - 1/n),\ n = 1, 2, \ldots$

Probability Model

A **random experiment** is an experiment whose outcome is uncertain

Definition (1.7)

A **sample space** is the set, S, of all possible outcomes of a particular experiment. An **event** is a subset of S.

Example 1.2 Toss a coin 3 times

- $S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$
- An example of an event is $A = \{\text{obtain exactly 2 heads}\}$ $= \{HHT, HTH, THH\}$ $(A \subseteq S)$

In probability, the set of all events of a sample space is a σ-algebra.

Definition (1.8)

A collection of subsets of S is called a *sigma algebra* (or *sigma field*), denoted by $\mathscr{B}$, if it satisfies the following:

(a) $\emptyset \in \mathscr{B}$.

(b) If $A \in \mathscr{B}$, then $A^c \in \mathscr{B}$.

(c) If $A_1, A_2, \ldots \in \mathscr{B}$, then $\cup_{i=1}^{\infty} A_i \in \mathscr{B}$ ($\mathscr{B}$ is closed under countable unions).

- (a) and (b) implies $S \in \mathscr{B}$.
- (b) and (c) also implies $\cap_{i=1}^{\infty} A_i \in \mathscr{B}$.
- (c) implies $\mathscr{B}$ is closed under finite unions.

Definition (1.9)

Given a sample space S and a σ-algebra $\mathscr{B}$ of subsets of S, a **probability model** (or probability assignment, function or more formally a probability measure) is a function $P : \mathscr{B} \to \mathbb{R}$ such that

(i) $0 \leq P(A) \leq 1$ for all $A \in \mathscr{B}$

(ii) $P(S) = 1$

(iii) if $A_1, A_2, A_3, \ldots$ are pairwise disjoint events, i.e. $A_i \cap A_j = \emptyset$ for all $i \neq j$ and $A_i \in \mathscr{B}$ for all i, then

$$P\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i).$$

Properties of a Probability Model

1. $P(\emptyset) = 0$

Properties of a Probability Model

2. $P(A \cup B) = P(A) + P(B)$ whenever $A \cap B = \emptyset$

Properties of a Probability Model

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

Properties of a Probability Model

4. $P(A) = P(A \cap B) + P(A \cap B^c)$

Properties of a Probability Model

5. (Law of total probability) For any *partition* $C_1, C_2, \ldots$ of S and any event A, we have

$$P(A) = \sum_{i=1}^{\infty} P(A \cap C_i).$$

Properties of a Probability Model

6. If $A \subseteq B$, then $P(A) \leq P(B)$

Properties of a Probability Model

7. (Boole's inequality) For any events $A_1, A_2, \ldots,$

$$P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i).$$

Properties of a Probability Model

8. If $A_1, A_2, \ldots$ are events with $A_n \uparrow$, then

$$P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n). \quad \left(\text{Recall } \lim_{n\to\infty} A_n = \bigcup_{n=1}^{\infty} A_n\right)$$

Properties of a Probability Model

9. If $A_1, A_2, \ldots$ are events with $A_n \downarrow$, then

$$P(\lim_{n\to\infty} A_n) = \lim_{n\to\infty} P(A_n). \quad \left(\text{Recall } \lim_{n\to\infty} A_n = \bigcap_{n=1}^{\infty} A_n\right)$$

Conditional Probability and Independence

Definition (1.10)

Let $A, B \subseteq S$ such that $P(B) > 0$. Then the **conditional probability of A given B** is

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

Definition (1.11)

Events A and B are **independent** if

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

Note: If A and B are independent, then so are A and B^c, A^c and B and A^c and B^c.

If A and B are independent, then A and B^c are independent

Bayes Rule

Let $B_1, B_2, \ldots, B_k$ be a partition of S. Then for any event A

$$
\begin{aligned}
P(A) &= \sum_{i=1}^{k} P(A \cap B_i) \qquad \text{(by the law of total probability)} \\
&= \sum_{i=1}^{k} P(A|B_i) \cdot P(B_i) \qquad \text{(by Defintion 10)}
\end{aligned}
$$

Definition (1.12)

Bayes Rule: Let A and B be events. Then

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)P(B|A)}{P(A)P(B|A) + P(A^c)P(B|A^c)}$$

Three Prisoners Problem

Three prisoners A, B, and C are on death row. The governor decided to pardon one of them and randomly picked the one to be pardoned. He told the warden about his choice but asked him to keep it secret. Prisoner A asked the warden who had been pardoned, and the warden wouldn't tell. A then asked which of B and C would be executed. The warden though for a while and then told A that B would be executed.

- Warden's reasoning: each prisoner has a $1/3$ chance to be pardoned. Clearly, either B or C will be executed. I gave A no information about whether he will be pardoned.
- A's reasoning: given that B will be executed, then either A or C will be pardoned and the chances are equal. My chance of being pardoned has increased from $1/3$ to $1/2$.

Who is correct?

Three Prisoners Problem - Warden

Three Prisoners Problem - Prisoner's Mistake

Random Variables

Definition (1.13)

A **random variable** is a function (mapping) from a sample space S into the real numbers. These are usually denoted by the capital letters X, Y, Z, U, V, W.

Given a random variable X on a probability space $(S, \mathscr{B}, P)$, the **cumulative distribution function** of X is

$$F_X(x) = P(X \leq x).$$

Properties of a CDF

Let $F(x) = P(X \leq x)$ be the CDF of a random variable X

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

(ii) $F(x) \leq F(y)$ whenever $x \leq y$

(iii) F is right continuous (i.e. $F(b+) = \lim_{x \to b^+} F(x) = F(b)$

(iv) $F(b-) = \lim_{x \to b^-} F(x)$ exists and is finite.

Discrete Random Variable

A discrete random variable $X : S \mapsto \mathbb{R}$ takes only a countable number of values $X(S) = \{X(\omega) : \omega \in S\} = \{x_1, x_2, x_2, \ldots\}$ and has a **probability mass function** (pmf)

$$p(x_i) = p_i = P(X = x_i), i = 1, 2, \ldots,$$

where $\sum_i p_i = 1$. The **cumulative distribution function** (cdf) is $F(x) = \sum_{i:x_i \leq x} p(x_i)$.

A function $p(x)$ is a pmf iff

(i) $p(x) \geq 0$ for all x

(ii) $\sum_{\text{all } x} p(x) = 1$

Examples: Binomial, Poisson, Negative Binomial, Geometric, Hypergeometric

Continuous Random Variable

A random variable X is said to be continuous if its CDF $F(x) = P(X \leq x)$ is a continuous function. F is continuous at b iff

$$P(X < b) = \lim_{y \to b^-} F(y) = \lim_{y \to b^+} F(y) = P(X \leq b),\ \forall b \in \mathbb{R}$$

iff $P(X = b) = 0$ for all b. The probability density function *pdf* of a continuous random variable X is defined as $f(x) = \frac{d}{dy} F(y)\Big|_{y=x}$.

A function f is a pdf iff

(i) $f(x) \geq 0$ for all $x \in \mathbb{R}$

(ii) $\int_{-\infty}^{\infty} f(x) dx = 1$

Examples: Exponential, Uniform, Normal, Beta, Gamma

Expected Value

Definition (1.15)

The **expected value** or mean of a random variable X, denoted $E[X]$, is given by

1. Discrete: $E[X] = \sum_{x \in X(S)} x \cdot P(X = x) = \sum_{\text{all } x} xp(x)$
2. Continuous: $E[X] = \int_{-\infty}^{\infty} xf(x)dx$

if $E[|X|] < \infty$. The **variance** is given by

$$\text{Var}(X) = E[(X - E[X])^2] = E[X^2] - (E[X])^2.$$

Note: For any function $g : \mathbb{R} \to \mathbb{R}$

$$E[g(X)] = \begin{cases} \sum_{\text{all } x} g(x)P(X = x), & \text{(X is discrete)} \\ \int_{-\infty}^{\infty} g(x)f(x)dx, & \text{(X is continuous)} \end{cases}$$

For example, taking $g(x) = x^n$ for any integer $n > 0$ yields the n^{th} moment of X, $E[X^n]$.

Mean of a Poisson RV

Let $X \sim \text{Poisson}(\lambda)$ with *pmf* of

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, 2, \ldots.$$

Find the expected value of X.

Let X be a random variable with pmf

$$p(x) = C/x^2,\ x = 1, 2, 3, \ldots.$$

Find C such that this is a valid pmf and calculate $E[X]$ if it exists.

Variance of an Exponential RV

Let $X \sim \text{Exp}(\theta)$ with pdf

$$f(x) = \begin{cases} \dfrac{e^{-x/\theta}}{\theta}, & x > 0 \\ 0, & x \leq 0 \end{cases}.$$

Calculate the variance of X.

Mean of a Cauchy RV

Find the mean of the Cauchy distribution if it exists. Let $X \sim$ Cauchy with pdf

$$f(x) = \frac{1}{\pi(1 + x^2)}, -\infty < x < \infty.$$

Moment Generating Functions

Definition (1.16)

For a random variable X, the **moment generating function** (mgf) is defined to be

$$M_X(t) = E[e^{tX}]$$

provided the expectation exists for $t \in (-h, h)$ for some $h > 0$.

Theorem (1.2)

Let $M_X(t)$ be the mgf of X and assume that $M_X(t) < \infty$ for all $t \in (-h, h)$ for some $h > 0$. Then

$$\frac{d^n}{dt^n} M_X(t)\Big|_{t=0} = E[X^n], n = 1, 2, 3, \ldots.$$

Binomial MGF

Let $X \sim \text{Binomial}(n, p)$ with pmf

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x}, x = 0, 1, 2, \ldots, n.$$

Poisson MGF

Let $Y \sim \text{Poisson}(\lambda)$ with pmf

$$p(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \; x = 0, 1, 2, \ldots.$$

MGFs completely determine the distribution of a random variable when they exist and can be used to establish convergence in distribution for sequence of random variables (also called weak convergence. more on this later). This is formally stated in the next theorems.

Theorem (1.3)

1. *Let X and Y be random variables with mgfs M_X and M_Y, respectively. If $M_X(t) = M_Y(t)$ for all $t \in (-h, h)$ for some $h > 0$ and are finite, then $F_X = F_Y$ (they have the same distribution).*
2. *Let $\{X_i, i \geq 1\}$ be a sequence of random variables with mgfs M_{X_n} and X a random variable with mgf M_X. If $\lim_{n\to\infty} M_{X_n}(t) = M_X(t)$ for all $t \in (-h, h)$ for some $h > 0$, then $\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ whenever F_X is continuous at x.*

Square of a Standard Normal

Let $Z \sim N(0,1)$ with pdf $f(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$, $-\infty < z < \infty$, and let $Y \sim \text{Gamma}(\alpha, \beta)$ with mgf $M_Y(t) = \left(\frac{1}{1-\beta t}\right)^{\alpha}$, $t < 1/\beta$.

What is the distribution of Z^2?

Binomial Approximation of a Poisson

Let $X_n \sim \text{Binomial}(n, p_n)$ and $Y \sim \text{Poisson}(\lambda)$ be such that $np_n \to \lambda$.

Univariate Transformations

Let X be a continuous random variable with pdf f_X. Suppose we want to know the pdf of the random variable $Y = g(X)$, f_Y. We can find this pdf the method of cdf.

1. Express $\{Y \leq y\}$ in terms of X
2. Express $F_Y(y) = P(Y \leq y)$ in terms of F_X
3. $f_Y(y) = \dfrac{d}{dy} F_Y(y)$

Univariate Transformations

Let $Y = g(X)$ where X is a random variable with pdf $f_X(x)$. The cdf of Y is

$$\begin{aligned} F_Y(y) &= P(Y \leq y) = P(g(X) \leq y) = P(\{x : g(x) \leq y\}) \\ &= \int_{\{x:g(x)\leq y\}} f_X(x)dx \end{aligned}$$

- If g is an increasing function, $g(X) \leq y \iff X \leq g^{-1}(y)$ and $F_Y(y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y))$
- If g is an decreasing function, $g(X) \leq y \iff X \geq g^{-1}(y)$ and
  $F_Y(y) = P(g(X) \leq y) = P(X \geq g^{-1}(y)) = 1 - F_X(g^{-1}(y))$
- Generally, if g is monotonic and the derivative of g^{-1} is continuous, then

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

Square of a Standard Normal

Let $Z \sim N(0,1)$ with pdf

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

Find the distribution of $Y = Z^2$.

Poisson Distribution

Poisson Variance

Lemma

Suppose the $\sum_{x=1}^{\infty} h(p,x)$ converges for some $p \in (a,b)$ and

1. for all x, $\dfrac{\partial}{\partial p} h(p,x)$ is continuous for $p \in (a,b)$.
2. $\sum_{x=1}^{\infty} \dfrac{\partial}{\partial p} h(p,x)$ converges uniformly for $p \in [c,d] \subseteq (a,b)$.

Then $\sum_{x=1}^{\infty} \dfrac{\partial}{\partial p} h(p,x) = \dfrac{d}{dp} \sum_{x=1}^{\infty} h(p,x)$.

Geometric Distribution: Mean

Memoryless Property

Lemma

If $X \sim \text{Geometric}(p)$, then $P(X \geq t + s | X \geq s) = P(X \geq t)$ for all $t, s \geq 0$.

Memoryless Property

Lemma

Suppose that X is a random variable such that

1. Supp(X)$\subseteq \mathbb{Z}$.
2. $P(X \geq 0) > 0$
3. $P(X \geq t + s | X \geq s) = P(X \geq t)$ for all non-negative integers s and t.

Then $X \sim$ Geometric(p), where $p = P(X = 0)$.

Memoryless Property

Uniform Distribution

Memoryless Property

Lemma

Suppose X is a continuous random variable. Then $P(X \geq t + s | X \geq t) = P(X \geq s)$ for all $s, t \geq 0$ and $P(X \geq 0) > 0$ if and only if $X \sim \text{Exp}(\theta)$ for some $\theta > 0$.

Memoryless Property

Memoryless Property

Memoryless Property

Chi-Square Distribution

Joint Distributions

Lemma

If $X \sim \text{Geometric}(p)$, then $P(X \geq t + s | X \geq s) = P(X \geq t)$ for all $t, s \geq 0$.

Definition (1.17)

The **joint distribution function** of two random variables X and Y id denoted by $F_{X,Y}$ and is defined as

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y), \text{ for all } x, y \in \mathbb{R}.$$

Definition (Joint mass functions and densities (1.18))

- If X and Y are both discrete, then their **joint mass function** is defined as $p_{X,Y}(x, y) = P(X = x, Y = y)$ for all x, y
- If X and Y are jointly continuous, then there exists a function $f_{X,Y}(x, y)$ called a **joint probability density function** such that

$$P(X \leq x, Y \leq y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{X,Y}(s, t) dt ds, \text{ for all } x, y \in \mathbb{R}$$

Marginal Distributions

(i) To obtain the marginal density function for X or Y in the discrete case sum over the range of the other variable:
$p_X(x) = \sum_{y \in \mathcal{Y}} p_{X,Y}(x,y)$ and $p_y(y) = \sum_{x \in \mathcal{X}} p_{X,Y}(x,y)$

(ii) To obtain the marginal densities in the jointly continuous case, integrate over the range of the other variable:
$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy$ and $f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$

Note: To calculate joint probabilities over non-rectangular sets, sum the joint pmf (discrete) or integrate the joint pdf (continuous) over the desired region.

$$P((X,Y) \in A) = \begin{cases} \sum_{(x,y) \in A} p_{X,Y}(x,y), & \text{(discrete)} \\ \int \int_A f_{X,Y}(x,y)dxdy, & \text{(continuous)} \end{cases}$$

Calculating a Joint Probability

Consider the joint density

$$f_{X,Y}(x,y) = \begin{cases} e^{-y}, & 0 \leq x \leq y < \infty \\ 0, & \text{otherwise} \end{cases}$$

Find $P(X \leq 1, Y \leq 2)$.

Conditional Distributions

The support of a random variable X, $\text{Supp}(X) = \{x : f_X(x) > 0\}$.

Definition (1.19)

The **conditional mass function** of X given $Y = y$ is defined as

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \quad \text{for all } x \in \text{Supp}(X)$$

and is defined to be 0 for all $y \notin \text{Supp}(Y)$ (i.e. for all y such that $f_Y(y) = 0$).

Definition (1.20)

Suppose that (X, Y) are continuous R.V.'s. The **conditional density** of X given $Y = y$ is defined as

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \quad \text{for all } x \in \text{Supp}(X)$$

and is defined to be 0 for all $y \notin \text{Supp}(Y)$.

$X|Y = y$

Suppose (X, Y) are jointly continuous with density

$$f_{X,Y}(x,y) = \begin{cases} e^{-y}, & 0 \leq x \leq y < \infty \\ 0, & \text{otherwise} \end{cases}.$$

$Y|X = x$

Suppose (X, Y) are jointly continuous with density

$$f_{X,Y}(x,y) = \begin{cases} e^{-y}, & 0 \leq x \leq y < \infty \\ 0, & \text{otherwise} \end{cases}.$$

Joint and Condtional Expectations

Definition (1.21)

Let X and Y be R.V.'s with joint density $f_{X,Y}$ and $g : \mathbb{R}^2 \to \mathbb{R}$, then the expected value of $g(X, Y)$ exists if

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |g(x, y)| f_{X,Y}(x, y) dxdy < \infty.$$

In such a case $E[g(X, Y)] = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dxdy$. (In the case that X and Y are discrete, integrals are replaced with sums).

Definition (1.22)

Let X and Y be jointly continuous R.V.'s and let $h : \mathbb{R} \to \mathbb{R}$. The conditional expectation of $h(X)$ given $Y = y$ is

$$E[h(X)|Y = y] = \int_{-\infty}^{\infty} h(x) f_{X|Y=y}(x) dx$$

provided $\int_{-\infty}^{\infty} |h(x)| f_{X|Y=y}(x) dx < \infty$.

$E[X^2 + Y]$

Let X and Y be jointly continuous R.V.'s with joint density

$$f_{X,Y}(x,y) = \begin{cases} e^{-y}, & 0 \le x \le y < \infty \\ 0, & \text{otherwise} \end{cases}$$

$E[X^2|Y=y]$

Let X and Y be jointly continuous R.V.'s with joint density

$$f_{X,Y}(x,y) = \begin{cases} e^{-y}, & 0 \leq x \leq y < \infty \\ 0, & \text{otherwise} \end{cases}$$

Tower Property

Definition (1.23)

The conditional expectation of X given Y is a function of Y, denoted $E[X|Y]$. This function, say h, is specified as $h(y) = E[X|Y = y]$.

Theorem (Tower Property)

$E[X] = E[E[X|Y]]$ *provided* $E[X]$ *exists.*

Proof of the Tower Property

Conditional Variance Formula

Definition (1.24)

The conditional variance of X given Y, denoted $V(X|Y)$, is defined as

$$V(X|Y) = E[(X - E[X|Y])^2|Y] = E[X^2|Y] - (E[X|Y])^2$$

Theorem (Conditional Variance Formula)

$$V(X) = V(E[X|Y]) + E[V(X|Y)]$$

E[X] using the Tower Property

Toss a coin n times. Suppose $p = P(\text{heads})$ is unknown. Let

$$X = \text{the number of heads observed out of } n \text{ tosses,}$$

then $X|p \sim \text{Binomial}(n, p)$. Suppose $p \sim \text{Beta}(a, b)$.

V(X) using the Conditional Variance

Toss a coin n times. Suppose $p = P(\text{heads})$ is unknown. Let

$$X = \text{the number of heads observed out of } n \text{ tosses},$$

then $X|p \sim \text{Binomial}(n, p)$. Suppose $p \sim \text{Beta}(a, b)$.

V(X) using the Conditional Variance

Independence

Definition (1.25)

Two random variables are said to be **independent** if

$$P(X \leq x, Y \leq y) = P(X \leq x)P(Y \leq y), \text{ for all } x, y \in \mathbb{R}.$$

Theorem

If X and Y are jointly continuous R.V.'s, then X and Y are independent if and only if

$$f_{X,Y}(x, y) = f_X(x)f_Y(y), \text{ for all } x, y \in \mathbb{R}.$$

Proof

Independence

Theorem

If there exists integrable functions g and h such that $f_{X,Y}(x,y) = g(x)h(y)$ for all $x, y \in \mathbb{R}$, then X and Y are independent.

Theorem

If X and Y are independent, then for any functions g and h $E[g(X)h(Y)] = E[g(X)]E[h(Y)]$.

Covariance

Definition (1.26)

The **covariance** of two R.V.'s X and Y is defined as

$$\mathrm{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])].$$

Properties:

1. (i) $\mathrm{Cov}(X, Y) = E[XY] - E[X]E[Y]$
2. (ii) $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
3. (iii) $\mathrm{Cov}(X + Y, Z) = \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z)$
4. (iv) $\mathrm{Cov}(aX, Y) = a\mathrm{Cov}(X, Y)$ for any $a \in \mathbb{R}$
5. (v) $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$
6. (vi) For all $a_i, b_i \in \mathbb{R}$,

$$\mathrm{Cov}\left(\sum_{i=1}^{n} a_i X_i, \sum_{j=1}^{m} b_i Y_i\right) = \sum_{i=1}^{n}\sum_{j=1}^{m} a_i b_i \mathrm{Cov}(X_i, Y_i)$$

Correlation

Definition (1.27)

The **correlation** between two R.V.'s X and Y is defined as

$$\rho_{X,Y} = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}}.$$

Properties of $\rho_{X,Y}$:

1. $\rho_{X,Y}$ is unitless
2. $|\rho_{X,Y}| \leq 1$
3. If $\rho_{X,Y} = 1$, then $Y = a + bX$ where $b > 0$. If $\rho_{X,Y} = -1$, then $Y = a + bX$ where $b < 0$. If $\rho_{X,Y} = 0$, then there is no linear association between X and Y (this is not the same as no association).

Independence and Covariance

Theorem

If X and Y are independent, then $Cov(X, Y) = 0$.

Independence and Covariance

Does $\text{Cov}(X, Y) = 0$ imply X and Y are independent? No!

Let $X \sim N(0, 1)$ and $Y = I_{|X|<2}$. X and Y are clearly not independent, but

$$\begin{aligned}
\text{Cov}(X, Y) &= E[XY] - E[X]E[Y] \\
&= E[XI_{|X|<2}] && \text{(since } E[X] = 0\text{)} \\
&= \int_{-\infty}^{\infty} xI_{|x|<2} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx \\
&= \int_{-2}^{2} \frac{1}{\sqrt{2\pi}} xe^{-x^2/2}\, dx \\
&= 0. && \text{(since } xe^{-x^2/2} \text{ is odd)}
\end{aligned}$$

Bivariate Transformations - Discrete Case

Let $X \sim \text{Poisson}(\theta)$ and $Y \sim \text{Poisson}(\lambda)$ with X independent of Y. Find the distribution of $U = X + Y$.

Bivariate Transformations - Discrete Case

Bivariate Transformations - Continous Case

Theorem

Let X and Y be two continuous random variables with joint density $f(x, y)$. Define two new random variables $U = g_1(X, Y)$ and $V = g_2(X, Y)$ with one-to-one functions g_1 and g_2. The joint density of U and V is given by $f_{U,V}(u, v) = f_{X,Y}(h_1(u, v), h_2(u, v))|\boldsymbol{J}|$, where $x = h_1(u, v)$ and $y = h_2(u, v)$ are the inverse transformations associated with g_1 and g_2, and $\boldsymbol{J} = \begin{vmatrix} \partial x/\partial u & \partial x/\partial v \\ \partial y/\partial u & \partial y/\partial v \end{vmatrix}$.

Let X and Y be independent R.V.'s with $X \sim \text{Gamma}(\alpha, 1)$ and $Y \sim \text{Gamma}(\beta, 1)$. Find the distribution of

$$(U, V) = \left(\frac{X}{X + Y}, X + Y \right).$$

Inequalities

1. Markov's Inequality: Let g be a nonnegative, increasing function such that $E[g(X)]$ exists. Then,

$$g(a)P(X > a) \leq E[g(X)], \text{ for every } a \in \mathbb{R}.$$

2. Chebychev's Inequality: $P(|X - E[X]| > k) \leq \text{Var}(X)/k^2$, for all $k > 0$.

3. Cauchy-Schwarz Inequality:

$$|\mathbb{E}XY| \leq \mathbb{E}|XY| \leq \left(\mathbb{E}|X|^2\right)^{1/2}\left(\mathbb{E}|Y|^2\right)^{1/2}.$$

4. Jensen's Inequality: If $g(x)$ is a convex function, then $\mathbb{E}g(X) \geq g(\mathbb{E}X)$.
   - A function $g(x)$ is convex if $g(\lambda x + (1-\lambda)y) \leq \lambda g(x) + (1-\lambda)g(y)$ for all x and y, where $\lambda \in (0,1)$. A function $g(x)$ is concave if $-g(x)$ is convex.

Inequalities

Prove that $|\rho_{X,Y}| \leq 1$.

Convergence in Probability

Motivation: We want to estimate a population parameter θ. Suppose that we have independent and identically distributed (i.i.d.) data, $X_1, X_2, \dots, X_n$. We want to find a sequence of estimators $\widehat{\theta}_n$ such that $\widehat{\theta}_n$ "converges" to θ as the sample size n increases to infinity. This property is called **consistency**.

Definition (1.28)

Let $\{X_n, n \geq 1\}$ be a sequence of R.V.'s. We say that the sequence $\{X_n, n \geq 1\}$ **converges in probability** to a R.V. X if for all $\varepsilon > 0$,

$$P(|X_n - X| > \varepsilon) \to 0 \text{ as } n \to \infty.$$

This is denoted by $X_n \xrightarrow{P} X$ as $n \to \infty$.

Definition (1.29)

A sequence of estimators $\{\widehat{\theta}_n, n \geq 1\}$ is said to be **weakly consistent** for θ if $\widehat{\theta}_n \xrightarrow{P} \theta$.

Weak Consistency of $\bar{X}_n$

Suppose that $X_1, X_2, \ldots$ are iid with $E[X_1] = \theta$ and $\text{Var}(X_1) = \sigma^2 < \infty$. Let $\widehat{\theta}_n = \sum_{i=1}^n X_i/n$. Is the sequence of estimators, $\{\widehat{\theta}_n, n \geq 1\}$, weakly consistent for θ?

Convergence in Probability

Theorem (Weak Law of Large Numbers)

Let $\{X_n, n \geq 1\}$ be a sequence of iid R.V.'s such that $E[X_1]$ exists. Then,

$$\frac{1}{n}\sum_{i=1}^{n} X_i \xrightarrow{P} E[X_1].$$

Properties of Convergence in Probability:

1. If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then $X_n \pm Y_n \xrightarrow{P} X \pm Y$.
2. If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then $X_n Y_n \xrightarrow{P} XY$.
3. If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then $X_n / Y_n \xrightarrow{P} X/Y$, (provided $Y_n, Y \neq 0$ with probability 1).
4. If $X_n \xrightarrow{P} X$ and f is a continuous function, then $f(X_n) \xrightarrow{P} f(X)$.

Suppose $\{X_n, n \geq 1\}$ are iid with mean $E[X_1] = \theta$ and $\text{Var}(X_1) = \sigma^2 < \infty$. Let $\widehat{\theta}_n = \bar{X}$. Show that $\widehat{\sigma}_n^2 = \sum_{i=1}^n (X_i - \widehat{\theta}_n)^2 / n$ is a weakly consistent estimator of σ^2.

Theorem

T_n is a weakly consistent sequence of estimators for $g(\theta)$ if $ET_n \to g(\theta)$ and $Var(T_n) \to 0$.

Convergence in Distribution

Definition (1.30)

A sequence of R.V.'s $\{X_n, n \geq 1\}$ is said to **converge in distribution** to a R.V. X ($X_n \xrightarrow{D} X$) if

$$F_{X_n}(x) \to F_X(x) \text{ as } n \to \infty$$

for every $x \in \mathbb{R}$ such that F_X is continuous at x.

Motivation: If $\widehat{\theta}_n$ is a consistent estimator of θ, then how accurate is it? We want to know what the value of $P(|\widehat{\theta}_n - \theta| > \varepsilon)$ is (What is the probability that $\widehat{\theta}_n$ is beyond some threshold from θ?)

Theorem (Central Limit Theorem (CLT))

Let $X_1, X_2, X_3, \ldots$ be iid R.V.'s with $E[X_1] = \theta$ and $Var(X_1) = \sigma^2 < \infty$. Then

$$\frac{\frac{1}{n}\sum_{i=1}^{n} X_i - \theta}{\sigma/\sqrt{n}} \xrightarrow{D} Z \text{ where } Z \sim N(0,1).$$

Suppose $\{X_n, n \geq 1\}$ are iid with mean $E[X_1] = \theta$ and $\text{Var}(X_1) = \sigma^2 < \infty$. Let $\widehat{\theta}_n = \sum_{i=1}^n X_i/n$. Then by the CLT

$$\frac{\widehat{\theta}_n - \theta}{\sigma/\sqrt{n}} \xrightarrow{D} Z \text{ where } Z \sim N(0,1),$$

so that for large n

$$P(|\widehat{\theta}_n - \theta| > \varepsilon) = P\left(\left|\frac{\widehat{\theta}_n - \theta}{\sigma/\sqrt{n}}\right| > \frac{\varepsilon}{\sigma/\sqrt{n}}\right) \approx P\left(|Z| > \frac{\varepsilon}{\sigma/\sqrt{n}}\right).$$

Thus for a large sample size, we can approximate these probabilities using the standard normal distribution.

Delta Method

Theorem (Slutsky's theorem)

If $X_n \xrightarrow{D} X$ *and* $Y_n \xrightarrow{P} y \in \mathbb{R}$, *then* $X_n Y_n \xrightarrow{D} X \cdot y$.

Theorem

If $X_n \xrightarrow{P} X$, *then* $X_n \xrightarrow{D} X$.

Theorem

For a constant $a \in \mathbb{R}$, $X_n \xrightarrow{P} a$ *iff* $X_n \xrightarrow{D} a$.

Theorem (Delta Method)

Suppose $\sqrt{n}(Y_n - \mu) \xrightarrow{D} N(0, \sigma^2)$ *and let* $g : \mathbb{R} \to \mathbb{R}$ *be a function with non-zero derivative at* μ . *Then*

$$\sqrt{n}(g(Y_n) - g(\mu)) \xrightarrow{D} N(0, [g'(\mu)]^2\sigma^2).$$

"Proof" Of Delta Method

Suppose g is differentiable in an open neigborhood of μ with $g'(\mu) \neq 0$ and g' continuous at μ. Then we have the following Taylor expansion about μ: for any x near μ there exists an x^* between x and μ such that

$$g(x) = g(\mu) + g'(x^*)(x - \mu).$$

By assumption, $\sqrt{n}(Y_n - \mu) \xrightarrow{D} N(0, \sigma^2)$, so

$$Y_n - \mu = \frac{1}{\sqrt{n}}\sqrt{n}(Y_n - \mu) \xrightarrow{D} 0 \cdot Z = 0,$$

by Slutsky's theorem, where $Z \sim N(0, \sigma^2)$. Then by Lemma 44, $Y_n - \mu \xrightarrow{P} 0$. Define Y_n^* to be the random variable such that

$$g(Y_n) = g(\mu) + g'(Y_n^*)(Y_n - \mu).$$

Then $Y_n^* \xrightarrow{P} \mu$, since $Y_n \xrightarrow{P} \mu$ and $|Y_n^* - \mu| \leq |Y_n - \mu|$. Thus,

$$\sqrt{n}(g(Y_n) - g(\mu)) = \underbrace{g'(Y_n^*)}_{\xrightarrow{P} g'(\mu)} \underbrace{\sqrt{n}(Y_n - \mu)}_{\xrightarrow{D} N(0,\sigma^2)} \xrightarrow{D} N(0, [g'(\mu)]^2\sigma^2).$$

Example

Let $X_1, X_2, \ldots$ be iid Uniform$(0, \theta)$. Consider the estimator $X_{(n)} = \max\{X_1, \ldots, X_n\}$ for θ (which is the MLE of θ). Show that $n(\theta - X_{(n)}) \xrightarrow{D} X$ where $X \sim \text{Exp}(\theta)$.

$$
\begin{aligned}
F_{n(\theta - X_{(n)})}(x) &= 1 - P\left(X_1 \le -\frac{x}{n} + \theta\right)^n \\
&= \begin{cases} 1, & \text{if } -x/n + \theta \le 0 \\ 1 - \left(\dfrac{-x/n + \theta}{\theta}\right)^n, & \text{if } 0 < -x/n + \theta < \theta \\ 0, & \text{if } -x/n + \theta \ge \theta \end{cases} \\
&= \begin{cases} 1, & \text{if } -x/n + \theta \le 0 \\ 1 - \left(1 - \dfrac{x/\theta}{n}\right)^n, & \text{if } 0 < -x/n + \theta < \theta \\ 0, & \text{if } -x/n + \theta \ge \theta \end{cases} \\
&\to \begin{cases} 1 - e^{-x/\theta}, & x \ge 0 \\ 0, & x < 0 \end{cases}
\end{aligned}
$$

Variance Stabilizing Transformation

Let $X_1, X_2, \ldots, X_n, \ldots$ be iid Poisson(λ), $\lambda > 0$. Find the asymptotic distribution of $\sqrt{n}(2\sqrt{\bar{X}} - 2\sqrt{\lambda})$.