STATISTICS: textbooks and monographs volume 172

Computational Methods in Statistics and Econometrics

Hisashi Tanizaki
Graduate School of Economics
Kobe University, Kobe 657-8501, Japan
(tanizaki@kobe-u.ac.jp)

Contents

List of Tables

List of Figures

Chapter 1

Elements of Statistics

In this chapter, the statistical methods used in the proceeding chapters are summarized. Mood, Graybill and Bose (1974), Hogg and Craig (1995) and Stuart and Ord (1991, 1994) are good references in Sections 1.1 – 1.8, while Judge, Hill, Griffiths and Lee (1980) and Greene (1993, 1997, 2000) are representative textbooks in Section 1.9.

1.1 Event and Probability

1.1.1 Event

We consider an **experiment** whose outcome is not known in advance but an event occurs with probability, which is sometimes called a **random experiment**. The **sample space** of an experiment is the set of all possible outcomes. Each element of a sample space is called an **element** of the sample space or a **sample point**, which represents each outcome obtained by the experiment. An **event** is any collection of outcomes contained in the sample space, or equivalently a subset of the sample space. A **simple event** consists of exactly one element and a **compound event** consists of more than one element. Sample space is denoted by Ω and sample point is given by ω.

Suppose that event A is a subset of sample space Ω. Let ω be a sample point in event A. Then, we say that a sample point ω is contained in a sample space A, which is denoted by $\omega \in A$.

A set of the sample points which does not belong to event A is called the **complementary event** of A, which is denoted by A^c. An event which do not have any sample point is called the **empty event**, denoted by $\emptyset$. Conversely, an event which includes all possible sample points is called the **whole event**, represented by Ω.

Next, consider two events A and B. A set consisting of the whole sample points which belong to either event A or event B is called the **sum event**, which is denoted by $A \cap B$. A set consisting of the whole sample points which belong to both event A and event B is called the **product event**, denoted by $A \cap B$. When $A \cap B = \emptyset$, we say that events A and B are **mutually exclusive**.

Example 1.1: Consider an experiment of casting a die. We have six sample points, which are denoted by $\omega_1 = \{1\}$, $\omega_2 = \{2\}$, $\omega_3 = \{3\}$, $\omega_4 = \{4\}$, $\omega_5 = \{5\}$ and $\omega_6 = \{6\}$, where ω_i represents the sample point that we have i. In this experiment, the sample space is given by $\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6\}$. Let A be the event that we have even numbers and B be the event that we have multiples of three. Then, we can write as $A = \{\omega_2, \omega_4, \omega_6\}$ and $B = \{\omega_3, \omega_6\}$. The complementary event of A is given by $A^c = \{\omega_1, \omega_3, \omega_5\}$, which is the event that we have odd numbers. The sum event of A and B is written as $A \cup B = \{\omega_2, \omega_3, \omega_4, \omega_6\}$, while the product event is $A \cap B = \{\omega_6\}$. Since $A \cap A^c = \emptyset$, we have the fact that A and A^c are mutually exclusive.

Example 1.2: Cast a coin three times. In this case, we have the following eight sample points:

$$\begin{array}{llll} \omega_1 = (\mathrm{H,H,H}), & \omega_2 = (\mathrm{H,H,T}), & \omega_3 = (\mathrm{H,T,H}), & \omega_4 = (\mathrm{H,T,T}), \\ \omega_5 = (\mathrm{T,H,H}), & \omega_6 = (\mathrm{T,H,T}), & \omega_7 = (\mathrm{T,T,H}), & \omega_8 = (\mathrm{T,T,T}), \end{array}$$

where H represents head while T indicates tail. For example, (H,T,H) means that the first flip lands head, the second flip is tail and the third one is head. Therefore, the sample space of this experiment can be written as:

$$\Omega = \{\omega_1, \omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8\}.$$

Let A be an event that we have two heads, B be an event that we obtain at least one tail, C be an event that we have head in the second flip, and D be an event that we obtain tail in the third flip. Then, the events A, B, C and D are give by:

$$\begin{array}{ll} A = \{\omega_2, \omega_3, \omega_5\}, & B = \{\omega_2, \omega_3, \omega_4, \omega_5, \omega_6, \omega_7, \omega_8\}, \\ C = \{\omega_1, \omega_2, \omega_5, \omega_6\}, & D = \{\omega_2, \omega_4, \omega_6, \omega_8\}. \end{array}$$

Since A is a subset of B, denoted by $A \subset B$, a sum event is $A \cup B = B$, while a product event is $A \cap B = A$. Moreover, we obtain $C \cap D = \{\omega_2, \omega_6\}$ and $C \cup D = \{\omega_1, \omega_2, \omega_4, \omega_5, \omega_6, \omega_8\}$.

1.1.2 Probability

Let $n(A)$ be the number of sample points in A. We have $n(A) \le n(B)$ when $A \subseteq B$. Each sample point is equally likely to occur. In the case of Example 1.1 (Section 1.1.1), each of the six possible outcomes has probability 1/6 and in Example 1.2 (Section 1.1.1), each of the eight possible outcomes has probability 1/8. Thus, the probability which the event A occurs is defined as:

$$P(A) = \frac{n(A)}{n(\Omega)}.$$

In Example 1.1, $P(A) = 3/6$ and $P(A \cap B) = 1/6$ are obtained, because $n(\Omega) = 6$, $n(A) = 3$ and $n(A \cap B) = 1$. Similarly, in Example 1.2, we have $P(C) = 4/8$,

$P(A \cap B) = P(A) = 3/8$ and so on. Note that we obtain $P(A) \leq P(B)$ because of $A \subseteq B$.

It is known that we have the following three properties on probability:

$$0 \leq P(A) \leq 1, \tag{1.1}$$

$$P(\Omega) = 1, \tag{1.2}$$

$$P(\emptyset) = 0. \tag{1.3}$$

$\emptyset \subseteq A \subseteq \Omega$ implies $n(\emptyset) \leq n(A) \leq n(\Omega)$. Therefore, we have:

$$\frac{n(\emptyset)}{n(\Omega)} \leq \frac{n(A)}{n(\Omega)} \leq \frac{n(\Omega)}{n(\Omega)} = 1.$$

Dividing by $n(\Omega)$, we obtain:

$$P(\emptyset) \leq P(A) \leq P(\Omega) = 1.$$

Because $\emptyset$ has no sample point, the number of the sample point is given by $n(\emptyset) = 0$ and accordingly we have $P(\emptyset) = 0$. Therefore, $0 \leq P(A) \leq 1$ is obtained as in (1.1). Thus, (1.1) – (1.3) are obtained.

When events A and B are mutually exclusive, i.e., when $A \cap B = \emptyset$, then $P(A \cup B) = P(A) + P(B)$ holds. Moreover, since A and A^c are mutually exclusive, $P(A^c) = 1 - P(A)$ is obtained. Note that $P(A \cup A^c) = P(\Omega) = 1$ holds. Generally, unless A and B are not exclusive, we have the following formula:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B),$$

which is known as the **addition rule**. In Example 1.1, each probability is given by $P(A \cup B) = 2/3$, $P(A) = 1/2$, $P(B) = 1/3$ and $P(A \cap B) = 1/6$. Thus, in the example we can verify that the above addition rule holds.

The probability which event A occurs, given that event B has occurred, is called the **conditional probability**, i.e.,

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

or equivalently,

$$P(A \cap B) = P(A|B)P(B),$$

which is called the **multiplication rule**. When event A is **independent** of event B, we have $P(A \cap B) = P(A)P(B)$, which implies that $P(A|B) = P(A)$. Conversely, $P(A \cap B) = P(A)P(B)$ implies that A is independent of B. In Example 1.2, because of $P(A \cap C) = 1/4$ and $P(C) = 1/2$, the conditional probability $P(A|C) = 1/2$ is obtained. From $P(A) = 3/8$, we have $P(A \cap C) \neq P(A)P(C)$. Therefore, A is not independent of C. As for C and D, since we have $P(C) = 1/2$, $P(D) = 1/2$ and $P(C \cap D) = 1/4$, we can show that C is independent of D.

1.2 Random Variable and Distribution

1.2.1 Univariate Random Variable and Distribution

The **random variable** X is defined as the real value function on sample space Ω. Since X is a function of a sample point ω, it is written as $X = X(\omega)$. Suppose that $X(\omega)$ takes a real value on the interval I. That is, X depends on a set of the sample point ω, i.e., $\{\omega; X(\omega) \in I\}$, which is simply written as $\{X \in I\}$.

In Example 1.1 (Section 1.1.1), suppose that X is a random variable which takes the number of spots up on the die. Then, X is a function of ω and takes the following values:

$$\begin{aligned} &X(\omega_1) = 1, \quad X(\omega_2) = 2, \quad X(\omega_3) = 3, \quad X(\omega_4) = 4, \\ &X(\omega_5) = 5, \quad X(\omega_6) = 6. \end{aligned}$$

In Example 1.2 (Section 1.1.1), suppose that X is a random variable which takes the number of heads. Depending on the sample point ω_i, X takes the following values:

$$\begin{aligned} &X(\omega_1) = 3, \quad X(\omega_2) = 2, \quad X(\omega_3) = 2, \quad X(\omega_4) = 1, \\ &X(\omega_5) = 2, \quad X(\omega_6) = 1, \quad X(\omega_7) = 1, \quad X(\omega_8) = 0. \end{aligned}$$

Thus, the random variable depends on a sample point.

There are two kinds of random variables. One is a **discrete random variable**, while another is a **continuous random variable**.

Discrete Random Variable and Probability Function: Suppose that the discrete random variable X takes x_1, x_2, $\cdots$, where $x_1 < x_2 < \cdots$ is assumed. Consider the probability that X takes x_i, i.e., $P(X = x_i) = p_i$, which is a function of x_i. That is, a function of x_i, say $f(x_i)$, is associated with $P(X = x_i) = p_i$. The function $f(x_i)$ represents the probability in the case where X takes x_i. Therefore, we have the following relation:

$$P(X = x_i) = p_i = f(x_i), \quad i = 1, 2, \cdots,$$

where $f(x_i)$ is called the **probability function** of X.

More formally, the function $f(x_i)$ which has the following properties is defined as the probability function.

$$\begin{aligned} &f(x_i) \geq 0, \quad i = 1, 2, \cdots, \\ &\sum_i f(x_i) = 1. \end{aligned}$$

Furthermore, for an event A, we can write a probability as the following equation:

$$P(X \in A) = \sum_{x_i \in A} f(x_i).$$

Several functional forms of $f(x_i)$ are shown in Section 2.4.

In Example 1.2 (Section 1.1.1), all the possible values of X are 0, 1, 2 and 3. (note that X denotes the number of heads when a die is cast three times). That is, $x_1 = 0$, $x_2 = 1$, $x_3 = 2$ and $x_4 = 3$ are assigned in this case. The probability that X takes x_1, x_2, x_3 or x_4 is given by:

$$\begin{aligned}
&P(X = 0) = f(0) = P(\{\omega_8\}) = \frac{1}{8}, \\
&P(X = 1) = f(1) = P(\{\omega_4, \omega_6, \omega_7\}) = P(\{\omega_4\}) + P(\{\omega_6\}) + P(\{\omega_7\}) = \frac{3}{8}, \\
&P(X = 2) = f(2) = P(\{\omega_2, \omega_3, \omega_5\}) = P(\{\omega_2\}) + P(\{\omega_3\}) + P(\{\omega_5\}) = \frac{3}{8}, \\
&P(X = 3) = f(3) = P(\{\omega_1\}) = \frac{1}{8},
\end{aligned}$$

which can be written as:

$$P(X = x) = f(x) = \frac{3!}{x!(3-x)!}\Big(\frac{1}{2}\Big)^3, \qquad x = 0, 1, 2, 3.$$

For $P(X = 1)$ and $P(X = 2)$, note that each sample point is mutually exclusive. The above probability function is called the **binomial distribution** discussed in Section 2.4.5. Thus, it is easy to check $f(x) \geq 0$ and $\sum_x f(x) = 1$ in Example 1.2.

Continuous Random Variable and Probability Density Function: Whereas a discrete random variable assumes at most a countable set of possible values, a continuous random variable X takes any real number within an interval I. For the interval I, the probability which X is contained in A is defined as:

$$P(X \in I) = \int_I f(x)\,\mathrm{d}x.$$

For example, let I be the interval between a and b for $a < b$. Then, we can rewrite $P(X \in I)$ as follows:

$$P(a < X < b) = \int_a^b f(x)\,\mathrm{d}x,$$

where $f(x)$ is called the **probability density function** of X, or simply the **density function** of X.

In order for $f(x)$ to be a probability density function, $f(x)$ has to satisfy the following properties:

$$\begin{aligned}
&f(x) \geq 0, \\
&\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1.
\end{aligned}$$

Some functional forms of $f(x)$ are discussed in Sections 2.1 – 2.3.

For a continuous random variable, note as follows:

$$P(X = x) = \int_x^x f(t)\,\mathrm{d}t = 0.$$

In the case of discrete random variables, $P(X = x_i)$ represents the probability which X takes x_i, i.e., $p_i = f(x_i)$. Thus, the probability function $f(x_i)$ itself implies probability. However, in the case of continuous random variables, $P(a < X < b)$ indicates the probability which X lies on the interval (a, b).

Example 1.3: As an example, consider the following function:

$$f(x) = \begin{cases} 1, & \text{for } 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Clearly, since $f(x) \geq 0$ for $-\infty < x < \infty$ and $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \int_0^1 f(x)\,\mathrm{d}x = [x]_0^1 = 1$, the above function can be a probability density function. In fact, it is called a **uniform distribution**. See Section 2.1 for the uniform distribution.

Example 1.4: As another example, consider the following function:

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

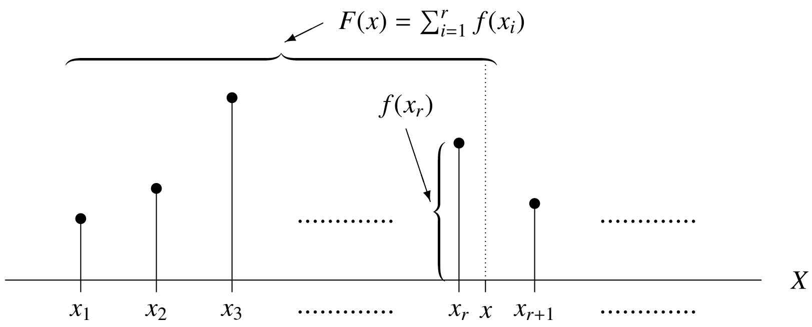
for $-\infty < x < \infty$. Clearly, we have $f(x) \geq 0$ for all x. We check whether $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$. First of all, we define I as $I = \int_{-\infty}^{\infty} f(x)\,\mathrm{d}x$.

To show $I = 1$, we may prove $I^2 = 1$ because of $f(x) > 0$ for all x, which is shown as follows:

$$\begin{aligned}
I^2 &= \Big(\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x\Big)^2 = \Big(\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x\Big)\Big(\int_{-\infty}^{\infty} f(y)\,\mathrm{d}y\Big) \\
&= \Big(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2)\,\mathrm{d}x\Big)\Big(\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}y^2)\,\mathrm{d}y\Big) \\
&= \frac{1}{2\pi} \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\Big(-\frac{1}{2}(x^2 + y^2)\Big)\,\mathrm{d}x\,\mathrm{d}y \\
&= \frac{1}{2\pi} \int_0^{2\pi}\int_0^{\infty} \exp(-\frac{1}{2}r^2) r\,\mathrm{d}r\,\mathrm{d}\theta \\
&= \frac{1}{2\pi} \int_0^{2\pi}\int_0^{\infty} \exp(-s)\,\mathrm{d}s\,\mathrm{d}\theta = \frac{1}{2\pi} 2\pi[-\exp(-s)]_0^{\infty} = 1.
\end{aligned}$$

In the fifth equality, integration by substitution is used. See Appendix 1.1 for the integration by substitution. $x = r\cos\theta$ and $y = r\sin\theta$ are taken for transformation,

Figure 1.1: Probability Function $f(x)$ and Distribution Function $F(x)$
— Discrete Random Variable —



Note that r is the integer which satisfies $x_r \leq x < x_{r+1}$.

which is a one-to-one transformation from (x, y) to (r, θ). Note that $0 < r < +\infty$ and $0 < \theta < 2\pi$. The Jacobian is given by:

$$J = \begin{vmatrix} \dfrac{\partial x}{\partial r} & \dfrac{\partial x}{\partial \theta} \\ \dfrac{\partial y}{\partial r} & \dfrac{\partial y}{\partial \theta} \end{vmatrix} = \begin{vmatrix} \cos\theta & -r\sin\theta \\ \sin\theta & r\cos\theta \end{vmatrix} = r.$$

In the inner integration of the sixth equality, again, integration by substitution is utilized, where transformation is $s = \dfrac{1}{2}r^2$.

Thus, we obtain the result $I^2 = 1$ and accordingly we have $I = 1$ because of $f(x) \geq 0$. Therefore, $f(x) = e^{-\frac{1}{2}x^2}/\sqrt{2\pi}$ is also taken as a probability density function. Actually, this density function is called the **standard normal probability density function**, discussed in Section 2.2.1.

Distribution Function: The **distribution function** (or the **cumulative distribution function**), denoted by $F(x)$, is defined as:

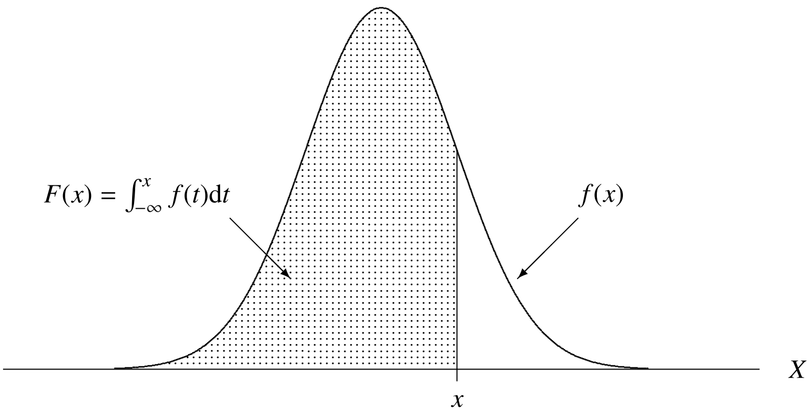
$$P(X \leq x) = F(x),$$

which represents the probability less than x. The properties of the distribution function $F(x)$ are given by:

$$F(x_1) \leq F(x_2), \quad \text{for } x_1 < x_2,$$
$$P(a < X \leq b) = F(b) - F(a), \quad \text{for } a < b,$$
$$F(-\infty) = 0, \quad F(+\infty) = 1.$$

The difference between the discrete and continuous random variables is given by:

Figure 1.2: Density Function $f(x)$ and Distribution Function $F(x)$
— Continuous Random Variable —



1. Discrete random variable (Figure 1.1):
   - $F(x) = \displaystyle\sum_{i=1}^{r} f(x_i) = \sum_{i=1}^{r} p_i$,
   where r denotes the integer which satisfies $x_r \leq x < x_{r+1}$.
   - $F(x_i) - F(x_i - \epsilon) = f(x_i) = p_i$,
   where ϵ is a small positive number less than $x_i - x_{i-1}$.
2. Continuous random variable (Figure 1.2):
   - $F(x) = \displaystyle\int_{-\infty}^{x} f(t)\,\mathrm{d}t$,
   - $F'(x) = f(x)$.

$f(x)$ and $F(x)$ are displayed in Figure 1.1 for a discrete random variable and Figure 1.2 for a continuous random variable.

1.2.2 Multivariate Random Variable and Distribution

We consider two random variables X and Y in this section. It is easy to extend to more than two random variables.

Discrete Random Variables: Suppose that discrete random variables X and Y take $x_1, x_2, \cdots$ and $y_1, y_2, \cdots$, respectively. The probability which event $\{\omega; X(\omega) = x_i$ and $Y(\omega) = y_j\}$ occurs is given by:

$$P(X = x_i, Y = y_j) = f_{xy}(x_i, y_j),$$

where $f_{xy}(x_i, y_j)$ represents the **joint probability function** of X and Y. In order for $f_{xy}(x_i, y_j)$ to be a joint probability function, $f_{xy}(x_i, y_j)$ has to satisfies the following properties:

$$
\begin{aligned}
& f_{xy}(x_i, y_j) \geq 0, \quad i, j = 1, 2, \cdots \\
& \sum_i \sum_j f_{xy}(x_i, y_j) = 1.
\end{aligned}
$$

Define $f_x(x_i)$ and $f_y(y_j)$ as:

$$
\begin{aligned}
f_x(x_i) &= \sum_j f_{xy}(x_i, y_j), \quad i = 1, 2, \cdots, \\
f_y(y_j) &= \sum_i f_{xy}(x_i, y_j), \quad j = 1, 2, \cdots.
\end{aligned}
$$

Then, $f_x(x_i)$ and $f_y(y_j)$ are called the **marginal probability functions** of X and Y. $f_x(x_i)$ and $f_y(y_j)$ also have the properties of the probability functions, i.e., $f_x(x_i) \geq 0$ and $\sum_i f_x(x_i) = 1$, and $f_y(y_j) \geq 0$ and $\sum_i f_y(y_j) = 1$.

Continuous Random Variables: Consider two continuous random variables X and Y. For a domain D, the probability which event $\{\omega; (X(\omega), Y(\omega)) \in D\}$ occurs is given by:

$$
P((X, Y) \in D) = \iint_D f_{xy}(x, y) \, \mathrm{d}x \, \mathrm{d}y,
$$

where $f_{xy}(x, y)$ is called the **joint probability density function** of X and Y or the **joint density function** of X and Y. $f_{xy}(x, y)$ has to satisfy the following properties:

$$
\begin{aligned}
& f_{xy}(x, y) \geq 0, \\
& \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f_{xy}(x, y) \, \mathrm{d}x \, \mathrm{d}y = 1.
\end{aligned}
$$

Define $f_x(x)$ and $f_y(y)$ as:

$$
\begin{aligned}
f_x(x) &= \int_{-\infty}^{\infty} f_{xy}(x, y) \, \mathrm{d}y, \qquad \text{for all } x \text{ and } y, \\
f_y(y) &= \int_{-\infty}^{\infty} f_{xy}(x, y) \, \mathrm{d}x,
\end{aligned}
$$

where $f_x(x)$ and $f_y(y)$ are called the **marginal probability density functions** of X and Y or the **marginal density functions** of X and Y.

For example, consider the event $\{\omega; a < X(\omega) < b, \ c < Y(\omega) < d\}$, which is a specific case of the domain D. Then, the probability that we have the event $\{\omega; a < X(\omega) < b, \ c < Y(\omega) < d\}$ is written as:

$$
P(a < X < b, \ c < Y < d) = \int_a^b \int_c^d f_{xy}(x, y) \, \mathrm{d}x \, \mathrm{d}y.
$$

The mixture of discrete and continuous random variables is also possible. For example, let X be a discrete random variable and Y be a continuous random variable. X takes x_1, x_2, $\cdots$. The probability which both X takes x_i and Y takes real numbers within the interval I is given by:

$$P(X = x_i, Y \in I) = \int_I f_{xy}(x_i, y)\,\mathrm{d}y.$$

Then, we have the following properties:

$$f_{xy}(x_i, y) \geq 0, \quad \text{for all } y \text{ and } i = 1, 2, \cdots,$$
$$\sum_i \int_{-\infty}^{\infty} f_{xy}(x_i, y)\,\mathrm{d}y = 1.$$

The marginal probability function of X is given by:

$$f_x(x_i) = \int_{-\infty}^{\infty} f_{xy}(x_i, y)\,\mathrm{d}y,$$

for $i = 1, 2, \cdots$. The marginal probability density function of Y is:

$$f_y(y) = \sum_i f_{xy}(x_i, y).$$

1.2.3 Conditional Distribution

Discrete Random Variable: The **conditional probability function** of X given $Y = y_j$ is represented as:

$$P(X = x_i | Y = y_j) = f_{x|y}(x_i|y_j) = \frac{f_{xy}(x_i, y_j)}{f_y(y_j)} = \frac{f_{xy}(x_i, y_j)}{\sum_i f_{xy}(x_i, y_j)}.$$

The second equality indicates the definition of the conditional probability, which is shown in Section 1.1.2. The features of the conditional probability function $f_{x|y}(x_i|y_j)$ are:

$$f_{x|y}(x_i|y_j) \geq 0, \quad i = 1, 2, \cdots,$$
$$\sum_i f_{x|y}(x_i|y_j) = 1, \quad \text{for any } j.$$

Continuous Random Variable: The **conditional probability density function** of X given $Y = y$ (or the **conditional density function** of X given $Y = y$) is:

$$f_{x|y}(x|y) = \frac{f_{xy}(x, y)}{f_y(y)} = \frac{f_{xy}(x, y)}{\int_{-\infty}^{\infty} f_{xy}(x, y)\,\mathrm{d}x}.$$

The properties of the conditional probability density function $f_{x|y}(x|y)$ are given by:

$$f_{x|y}(x|y) \geq 0,$$
$$\int_{-\infty}^{\infty} f_{x|y}(x|y)\,\mathrm{d}x = 1, \quad \text{for any } Y = y.$$

Independence of Random Variables: For discrete random variables X and Y, we say that X is **independent** (or **stochastically independent**) of Y if and only if $f_{xy}(x_i, y_j) = f_x(x_i) f_y(y_j)$. Similarly, for continuous random variables X and Y, we say that X is independent of Y if and only if $f_{xy}(x, y) = f_x(x) f_y(y)$.

When X and Y are stochastically independent, $g(X)$ and $h(Y)$ are also stochastically independent, where $g(X)$ and $h(Y)$ are functions of X and Y.

1.3 Mathematical Expectation

1.3.1 Univariate Random Variable

Definition of Mathematical Expectation: Let $g(X)$ be a function of random variable X. The mathematical expectation of $g(X)$, denoted by $\mathrm{E}(g(X))$, is defined as follows:

$$\mathrm{E}(g(X)) = \begin{cases} \displaystyle\sum_i g(x_i) p_i = \sum_i g(x_i) f(x_i), & \text{(Discrete Random Variable)}, \\ \displaystyle\int_{-\infty}^{\infty} g(x) f(x)\,\mathrm{d}x, & \text{(Continuous Random Variable)}. \end{cases}$$

The following three functional forms of $g(X)$ are important.

1. $\boldsymbol{g(X) = X}$.

   The expectation of X, $\mathrm{E}(X)$, is known as **mean** of random variable X.

   $$\begin{aligned} \mathrm{E}(X) &= \begin{cases} \displaystyle\sum_i x_i f(x_i), & \text{(Discrete Random Variable)}, \\ \displaystyle\int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x, & \text{(Continuous Random Variable)}, \end{cases} \\ &= \mu, \quad (\text{or } \mu_x). \end{aligned}$$

   When a distribution of X is symmetric, mean indicates the center of the distribution.

2. $\boldsymbol{g(X) = (X - \mu)^2}$.

   The expectation of $(X - \mu)^2$ is known as **variance** of random variable X, which is denoted by $\mathrm{V}(X)$.

   $$\begin{aligned} \mathrm{V}(X) &= \mathrm{E}((X - \mu)^2) \\ &= \begin{cases} \displaystyle\sum_i (x_i - \mu)^2 f(x_i), & \text{(Discrete Random Variable)}, \\ \displaystyle\int_{-\infty}^{\infty} (x - \mu)^2 f(x)\,\mathrm{d}x, & \text{(Continuous Random Variable)}, \end{cases} \\ &= \sigma^2, \quad (\text{or } \sigma_x^2). \end{aligned}$$

If X is broadly distributed, $\sigma^2 = \mathrm{V}(X)$ becomes large. Conversely, if the distribution is concentrated on the center, σ^2 becomes small. Note that $\sigma = \sqrt{V(X)}$ is called the **standard deviation**.

3. $\boldsymbol{g(X) = e^{\theta X}}$.

   The expectation of $e^{\theta X}$ is called the **moment-generating function**, which is denoted by $\phi(\theta)$.

$$\begin{aligned}\phi(\theta) &= \mathrm{E}(e^{\theta X}) \\ &= \begin{cases} \displaystyle\sum_i e^{\theta x_i} f(x_i), & \text{(Discrete Random Variable)}, \\ \displaystyle\int_{-\infty}^{\infty} e^{\theta x} f(x)\,\mathrm{d}x, & \text{(Continuous Random Variable)}. \end{cases}\end{aligned}$$

   Note that the definition of e is given by:

$$\begin{aligned} e &= \lim_{x\to 0}(1+x)^{\frac{1}{x}} = \lim_{h\to\infty}\Big(1+\frac{1}{h}\Big)^h \\ &= 2.71828182845905. \end{aligned}$$

   The moment-generating function plays an important roll in statistics, which is discussed in Section 1.5.

In Examples 1.5 – 1.8, mean, variance and the moment-generating function are computed.

Example 1.5: In Example 1.2 of flipping a coin three times (Section 1.1.1), we see in Section 1.2.1 that the probability function is written as the following binomial distribution:

$$P(X=x) = f(x) = \frac{n!}{x!(n-x)!}p^x(1-p)^{n-x}, \quad \text{for } x = 0, 1, 2, \cdots, n,$$

where $n = 3$ and $p = 1/2$. When X has the binomial distribution above, we obtain $\mathrm{E}(X)$, $\mathrm{V}(X)$ and $\phi(\theta)$ as follows.

First, $\mu = \mathrm{E}(X)$ is computed as:

$$\begin{aligned}\mu = \mathrm{E}(X) &= \sum_x x f(x) = \sum_x x\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x} \\ &= \sum_x \frac{n!}{(x-1)!(n-x)!}p^x(1-p)^{n-x} = np\sum_x \frac{(n-1)!}{(x-1)!(n-x)!}p^{x-1}(1-p)^{n-x} \\ &= np\sum_{x'}\frac{n'!}{x'!(n'-x')!}p^{x'}(1-p)^{n'-x'} = np,\end{aligned}$$

where $n' = n-1$ and $x' = x-1$ are set.

Second, in order to obtain $\sigma^2 = \mathrm{V}(X)$, we rewrite $\mathrm{V}(X)$ as:

$$\sigma^2 = \mathrm{V}(X) = \mathrm{E}(X^2) - \mu^2 = \mathrm{E}(X(X-1)) + \mu - \mu^2.$$

$\mathrm{E}(X(X-1))$ is given by:

$$\begin{aligned}
\mathrm{E}(X(X-1)) &= \sum_x x(x-1)f(x) = \sum_x x(x-1)\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x} \\
&= \sum_x \frac{n!}{(x-2)!(n-x)!}p^x(1-p)^{n-x} \\
&= n(n-1)p^2 \sum_x \frac{(n-2)!}{(x-2)!(n-x)!}p^{x-2}(1-p)^{n-x} \\
&= n(n-1)p^2 \sum_{x'} \frac{n'!}{x'!(n'-x')!}p^{x'}(1-p)^{n'-x'} = n(n-1)p^2,
\end{aligned}$$

where $n' = n-2$ and $x' = x-2$ are re-defined. Therefore, $\sigma^2 = \mathrm{V}(X)$ is obtained as:

$$\begin{aligned}
\sigma^2 = \mathrm{V}(X) &= \mathrm{E}(X(X-1)) + \mu - \mu^2 \\
&= n(n-1)p^2 + np - n^2p^2 = -np^2 + np = np(1-p).
\end{aligned}$$

Finally, the moment-generating function $\phi(\theta)$ is represented as:

$$\begin{aligned}
\phi(\theta) = \mathrm{E}(e^{\theta X}) &= \sum_x e^{\theta x}\frac{n!}{x!(n-x)!}p^x(1-p)^{n-p} \\
&= \sum_x \frac{n!}{x!(n-x)!}(pe^\theta)^x(1-p)^{n-p} = (pe^\theta + 1 - p)^n.
\end{aligned}$$

In the last equality, we utilize the following formula:

$$(a+b)^n = \sum_{x=0}^{n} \frac{n!}{x!(n-x)!}a^x b^{n-x},$$

which is called the **binomial theorem**.

Example 1.6: As an example of continuous random variables, in Section 1.2.1 the uniform distribution is introduced, which is given by:

$$f(x) = \begin{cases} 1, & \text{for } 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

When X has the uniform distribution above, $\mathrm{E}(X)$, $\mathrm{V}(X)$ and $\phi(\theta)$ are computed as follows:

$$\mu = \mathrm{E}(X) = \int_{-\infty}^{\infty} xf(x)\,\mathrm{d}x = \int_0^1 x\,\mathrm{d}x = \Big[\frac{1}{2}x^2\Big]_0^1 = \frac{1}{2},$$

$$\begin{aligned}
\sigma^2 &= \mathrm{V}(X) = \mathrm{E}(X^2) - \mu^2 \\
&= \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x - \mu^2 = \int_0^1 x^2\,\mathrm{d}x - \mu^2 = \Big[\frac{1}{3}x^3\Big]_0^1 - \Big(\frac{1}{2}\Big)^2 = \frac{1}{12},
\end{aligned}$$

$$\phi(\theta) = \mathrm{E}(e^{\theta X}) = \int_{-\infty}^{\infty} e^{\theta x} f(x)\,\mathrm{d}x = \int_0^1 e^{\theta x}\,\mathrm{d}x = [\frac{1}{\theta}e^{\theta x}]_0^1 = \frac{1}{\theta}(e^{\theta} - 1).$$

Example 1.7: As another example of continuous random variables, we take the standard normal distribution:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}, \qquad \text{for } -\infty < x < \infty,$$

which is discussed in Section 2.2.1. When X has a standard normal distribution, i.e., when $X \sim N(0,1)$, $\mathrm{E}(X)$, $\mathrm{V}(X)$ and $\phi(\theta)$ are as follows.

$\mathrm{E}(X)$ is obtained as:

$$\mathrm{E}(X) = \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x e^{-\frac{1}{2}x^2}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \Big[-e^{-\frac{1}{2}x^2}\Big]_{-\infty}^{\infty} = 0,$$

because $\lim_{x\to\pm\infty} -e^{-\frac{1}{2}x^2} = 0$.

$\mathrm{V}(X)$ is computed as follows:

$$\begin{aligned}
\mathrm{V}(X) = \mathrm{E}(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} x \frac{\mathrm{d}(-e^{-\frac{1}{2}x^2})}{\mathrm{d}x}\,\mathrm{d}x \\
&= \frac{1}{\sqrt{2\pi}} \Big[x(-e^{-\frac{1}{2}x^2})\Big]_{-\infty}^{\infty} + \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}x^2}\,\mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,\mathrm{d}x = 1.
\end{aligned}$$

The first equality holds because of $\mathrm{E}(X) = 0$. In the fifth equality, use the following integration formula, called the **integration by parts**:

$$\int_a^b h(x) g'(x)\,\mathrm{d}x = \Big[h(x)g(x)\Big]_a^b - \int_a^b h'(x) g(x)\,\mathrm{d}x,$$

where we take $h(x) = x$ and $g(x) = -e^{-\frac{1}{2}x^2}$ in this case. See Appendix 1.2 for the integration by parts. In the sixth equality, $\lim_{x\to\pm\infty} -xe^{-\frac{1}{2}x^2} = 0$ is utilized. The last equality is because the integration of the standard normal probability density function is equal to one (see p.6 in Section 1.2.1 for the integration of the standard normal probability density function).

$\phi(\theta)$ is derived as follows:

$$\begin{aligned}
\phi(\theta) &= \int_{-\infty}^{\infty} e^{\theta x} f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} e^{\theta x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,\mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2 + \theta x}\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\left((x-\theta)^2 - \theta^2\right)}\,\mathrm{d}x = e^{\frac{1}{2}\theta^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-\theta)^2}\,\mathrm{d}x = e^{\frac{1}{2}\theta^2}.
\end{aligned}$$

The last equality holds because the integration indicates the normal density with mean θ and variance one. See Section 2.2.2 for the normal density.

Example 1.8: When the moment-generating function of X is given by $\phi_x(\theta) = e^{\frac{1}{2}\theta^2}$ (i.e., X has a standard normal distribution), we want to obtain the moment-generating function of $Y = \mu + \sigma X$.

Let $\phi_x(\theta)$ and $\phi_y(\theta)$ be the moment-generating functions of X and Y, respectively. Then, the moment-generating function of Y is obtained as follows:

$$\begin{aligned}\phi_y(\theta) &= \mathrm{E}(e^{\theta Y}) = \mathrm{E}(e^{\theta(\mu+\sigma X)}) = e^{\theta\mu}\mathrm{E}(e^{\theta\sigma X}) = e^{\theta\mu}\phi_x(\theta\sigma) = e^{\theta\mu}e^{\frac{1}{2}\sigma^2\theta^2} \\ &= \exp(\mu\theta + \frac{1}{2}\sigma^2\theta^2).\end{aligned}$$

Some Formulas of Mean and Variance:

1. **Theorem:** $\mathrm{E}(aX + b) = a\mathrm{E}(X) + b$, where a and b are constant.

   Proof:

   When X is a discrete random variable,

   $$\begin{aligned}\mathrm{E}(aX + b) &= \sum_i (ax_i + b)f(x_i) = a\sum_i x_i f(x_i) + b\sum_i f(x_i) \\ &= a\mathrm{E}(X) + b.\end{aligned}$$

   Note that we have $\sum_i x_i f(x_i) = \mathrm{E}(X)$ from the definition of mean and $\sum_i f(x_i) = 1$ because $f(x_i)$ is a probability function.

   If X is a continuous random variable,

   $$\begin{aligned}\mathrm{E}(aX + b) &= \int_{-\infty}^{\infty} (ax + b)f(x)\,\mathrm{d}x = a\int_{-\infty}^{\infty} xf(x)\,\mathrm{d}x + b\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x \\ &= a\mathrm{E}(X) + b.\end{aligned}$$

   Similarly, note that we have $\int_{-\infty}^{\infty} xf(x)\,\mathrm{d}x = \mathrm{E}(X)$ from the definition of mean and $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$ because $f(x)$ is a probability density function.

2. **Theorem:** $\mathrm{V}(X) = \mathrm{E}(X^2) - \mu^2$, where $\mu = \mathrm{E}(X)$.

   Proof:

   $\mathrm{V}(X)$ is rewritten as follows:

   $$\begin{aligned}\mathrm{V}(X) &= \mathrm{E}((X-\mu)^2) = \mathrm{E}(X^2 - 2\mu X - \mu^2) \\ &= \mathrm{E}(X^2) - 2\mu\mathrm{E}(X) + \mu^2 = \mathrm{E}(X^2) - \mu^2.\end{aligned}$$

   The first equality is due to the definition of variance.

3. **Theorem:** $\mathrm{V}(aX + b) = a^2\mathrm{V}(X)$, where a and b are constant.

Proof:

From the definition of the mathematical expectation, $\mathrm{V}(aX+b)$ is represented as:

$$\begin{aligned}\mathrm{V}(aX+b) &= \mathrm{E}\Big(((aX+b)-\mathrm{E}(aX+b))^2\Big) = \mathrm{E}\Big((aX-a\mu)^2\Big)\\ &= \mathrm{E}(a^2(X-\mu)^2) = a^2\mathrm{E}((X-\mu)^2) = a^2\mathrm{V}(X)\end{aligned}$$

The first and the fifth equalities are from the definition of variance. We use $\mathrm{E}(aX+b) = a\mu + b$ in the second equality.

4. **Theorem:** The random variable X is assumed to be distributed with mean $\mathrm{E}(X) = \mu$ and variance $\mathrm{V}(X) = \sigma^2$. Define $Z = (X-\mu)/\sigma$. Then, we have $\mathrm{E}(Z) = 0$ and $\mathrm{V}(Z) = 1$.

   Proof:

   $\mathrm{E}(X)$ and $\mathrm{V}(X)$ are obtained as:

$$\mathrm{E}(Z) = \mathrm{E}\Big(\frac{X-\mu}{\sigma}\Big) = \frac{\mathrm{E}(X)-\mu}{\sigma} = 0,$$

$$\mathrm{V}(Z) = \mathrm{V}\Big(\frac{1}{\sigma}X - \frac{\mu}{\sigma}\Big) = \frac{1}{\sigma^2}\mathrm{V}(X) = 1.$$

   The transformation from X to Z is known as normalization or standardization.

1.3.2 Bivariate Random Variable

Definition: Let $g(X,Y)$ be a function of random variables X and Y. The mathematical expectation of $g(X,Y)$, denoted by $\mathrm{E}(g(X,Y))$, is defined as:

$$\mathrm{E}(g(X,Y)) = \begin{cases} \displaystyle\sum_i\sum_j g(x_i,y_j)f(x_i,y_j), & \text{(Discrete Random Variables)},\\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} g(x,y)f(x,y)\,\mathrm{d}x\,\mathrm{d}y, & \text{(Continuous Random Variables)}.\end{cases}$$

The following four functional forms are important, i.e., mean, variance, covariance and the moment-generating function.

1. $\boldsymbol{g(X,Y) = X}$:

   The expectation of random variable X, i.e., $\mathrm{E}(X)$, is given by:

$$\begin{aligned}\mathrm{E}(X) &= \begin{cases} \displaystyle\sum_i\sum_j x_i f(x_i,y_j), & \text{(Discrete Random Variables)},\\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x f(x,y)\,\mathrm{d}x\,\mathrm{d}y, & \text{(Continuous Random Variables)},\end{cases}\\ &= \mu_x.\end{aligned}$$

   The case of $g(X,Y) = Y$ is exactly the same formulation as above, i.e., $\mathrm{E}(Y) = \mu_y$.

2. $\boldsymbol{g(X, Y) = (X - \mu_x)^2}$:

   The expectation of $(X - \mu_x)^2$ is known as variance of random variable X, which is denoted by $\mathrm{V}(X)$ and represented as follows:

$$
\begin{aligned}
\mathrm{V}(X) &= \mathrm{E}((X-\mu_x)^2) \\
&= \begin{cases} \displaystyle\sum_i \sum_j (x_i - \mu_x)^2 f(x_i, y_j), & \text{(Discrete Case)}, \\ \displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x, y)\,\mathrm{d}x\,\mathrm{d}y, & \text{(Continuous Case)}, \end{cases} \\
&= \sigma_x^2.
\end{aligned}
$$

   The variance of Y is also obtained in the same fashion, i.e., $\mathrm{V}(Y) = \sigma_y^2$.

3. $\boldsymbol{g(X, Y) = (X - \mu_x)(Y - \mu_y)}$:

   The expectation of $(X - \mu_x)(Y - \mu_y)$ is known as **covariance** of X and Y, which is denoted by $\mathrm{Cov}(X, Y)$ and written as:

$$
\begin{aligned}
\mathrm{Cov}(X, Y) &= \mathrm{E}((X-\mu_x)(Y-\mu_y)) \\
&= \begin{cases} \displaystyle\sum_i \sum_j (x_i - \mu_x)(y_j - \mu_y) f(x_i, y_j), & \text{(Discrete Case)}, \\ \displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mu_x)(y - \mu_y) f(x, y)\,\mathrm{d}x\,\mathrm{d}y, & \text{(Continuous Case)}. \end{cases}
\end{aligned}
$$

   Thus, covariance is defined in the case of bivariate random variables.

4. $\boldsymbol{g(X, Y) = e^{\theta_1 X + \theta_2 Y}}$:

   The mathematical expectation of $e^{\theta_1 X + \theta_2 Y}$ is called the moment-generating function, which is denoted by $\phi(\theta_1, \theta_2)$ and written as:

$$
\begin{aligned}
\phi(\theta_1, \theta_2) &= \mathrm{E}(e^{\theta_1 X + \theta_2 Y}) \\
&= \begin{cases} \displaystyle\sum_i \sum_j e^{\theta_1 x_i + \theta_2 y_j} f(x_i, y_j), & \text{(Discrete Case)}, \\ \displaystyle\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\theta_1 x + \theta_2 y} f(x, y)\,\mathrm{d}x\,\mathrm{d}y, & \text{(Continuous Case)}. \end{cases}
\end{aligned}
$$

   In Section 1.5, the moment-generating function in the multivariate cases is discussed in more detail.

Some Formulas of Mean and Variance: We consider two random variables X and Y. Some formulas are shown as follows.

1. **Theorem:** $\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y)$.

Proof:

For discrete random variables X and Y, it is given by:

$$\begin{aligned}
\mathrm{E}(X+Y) &= \sum_i \sum_j (x_i + y_j) f_{xy}(x_i, y_j) \\
&= \sum_i \sum_j x_i f_{xy}(x_i, y_j) + \sum_i \sum_j y_j f_{xy}(x_i, y_j) \\
&= \mathrm{E}(X) + \mathrm{E}(Y).
\end{aligned}$$

For continuous random variables X and Y, we can show:

$$\begin{aligned}
\mathrm{E}(X+Y) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x+y) f_{xy}(x, y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x f_{xy}(x, y)\,\mathrm{d}x\,\mathrm{d}y + \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} y f_{xy}(x, y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \mathrm{E}(X) + \mathrm{E}(Y).
\end{aligned}$$

2. **Theorem:** $\mathrm{E}(XY) = \mathrm{E}(X)\mathrm{E}(Y)$, when X is independent of Y.

   Proof:

   For discrete random variables X and Y,

$$\begin{aligned}
\mathrm{E}(XY) &= \sum_i \sum_j x_i y_j f_{xy}(x_i, y_j) = \sum_i \sum_j x_i y_j f_x(x_i) f_y(y_j) \\
&= \Big(\sum_i x_i f_x(x_i)\Big)\Big(\sum_j y_j f_y(y_j)\Big) = \mathrm{E}(X)\mathrm{E}(Y).
\end{aligned}$$

   If X is independent of Y, the second equality holds, i.e., $f_{xy}(x_i, y_j) = f_x(x_i) f_y(y_j)$.

   For continuous random variables X and Y,

$$\begin{aligned}
\mathrm{E}(XY) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_{xy}(x, y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} xy f_x(x) f_y(y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \Big(\int_{-\infty}^{\infty} x f_x(x)\,\mathrm{d}x\Big)\Big(\int_{-\infty}^{\infty} y f_y(y)\,\mathrm{d}y\Big) = \mathrm{E}(X)\mathrm{E}(Y).
\end{aligned}$$

   When X is independent of Y, we have $f_{xy}(x, y) = f_x(x) f_y(y)$ in the second equality.

3. **Theorem:** $\mathrm{Cov}(X, Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y)$.

Proof:

For both discrete and continuous random variables, we can rewrite as follows:

$$\begin{aligned}\mathrm{Cov}(X,Y) &= \mathrm{E}((X-\mu_x)(Y-\mu_y)) = \mathrm{E}(XY-\mu_xY-\mu_yX+\mu_x\mu_y)\\ &= \mathrm{E}(XY)-\mathrm{E}(\mu_xY)-\mathrm{E}(\mu_yX)+\mu_x\mu_y\\ &= \mathrm{E}(XY)-\mu_x\mathrm{E}(Y)-\mu_y\mathrm{E}(X)+\mu_x\mu_y\\ &= \mathrm{E}(XY)-\mu_x\mu_y-\mu_y\mu_x+\mu_x\mu_y = \mathrm{E}(XY)-\mu_x\mu_y\\ &= \mathrm{E}(XY)-\mathrm{E}(X)\mathrm{E}(Y).\end{aligned}$$

In the fourth equality, the theorem in Section 1.3.1 is used, i.e., $\mathrm{E}(\mu_xY)=\mu_x\mathrm{E}(Y)$ and $\mathrm{E}(\mu_yX)=\mu_y\mathrm{E}(X)$.

4. **Theorem:** $\mathrm{Cov}(X,Y)=0$, when X is independent of Y.

   Proof:

   From the above two theorems, we have $\mathrm{E}(XY)=\mathrm{E}(X)\mathrm{E}(Y)$ when X is independent of Y and $\mathrm{Cov}(X,Y)=\mathrm{E}(XY)-\mathrm{E}(X)\mathrm{E}(Y)$. Therefore, $\mathrm{Cov}(X,Y)=0$ is obtained when X is independent of Y.

5. **Definition:** The **correlation coefficient** between X and Y, denoted by ρ_{xy}, is defined as:

$$\rho_{xy}=\frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{V}(X)}\sqrt{\mathrm{V}(Y)}}=\frac{\mathrm{Cov}(X,Y)}{\sigma_x\sigma_y}.$$

   When $\rho_{xy}>0$, we say that there is a **positive correlation** between X and Y. As ρ_{xy} approaches 1, we say that there is a **strong positive correlation** between X and Y. When $\rho_{xy}<0$, we say that there is a **negative correlation** between X and Y. As ρ_{xy} approaches -1, we say that there is a **strong negative correlation** between X and Y.

6. **Theorem:** $\rho_{xy}=0$, when X is independent of Y.

   Proof:

   When X is independent of Y, we have $\mathrm{Cov}(X,Y)=0$. Therefore, we can obtain the result $\rho_{xy}=\dfrac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{V}(X)}\sqrt{\mathrm{V}(Y)}}=0$. However, note that $\rho_{xy}=0$ does not mean the independence between X and Y.

7. **Theorem:** $\mathrm{V}(X\pm Y)=\mathrm{V}(X)\pm 2\mathrm{Cov}(X,Y)+\mathrm{V}(Y)$.

   Proof:

   For both discrete and continuous random variables, $\mathrm{V}(X\pm Y)$ is rewritten as follows:

$$\mathrm{V}(X\pm Y)=\mathrm{E}\Big(((X\pm Y)-\mathrm{E}(X\pm Y))^2\Big)=\mathrm{E}\Big(((X-\mu_x)\pm(Y-\mu_y))^2\Big)$$

$$
\begin{aligned}
&= \mathrm{E}((X-\mu_x)^2 \pm 2(X-\mu_x)(Y-\mu_y) + (Y-\mu_y)^2) \\
&= \mathrm{E}((X-\mu_x)^2) \pm 2\mathrm{E}((X-\mu_x)(Y-\mu_y)) + \mathrm{E}((Y-\mu_y)^2) \\
&= \mathrm{V}(X) \pm 2\mathrm{Cov}(X,Y) + \mathrm{V}(Y).
\end{aligned}
$$

8. **Theorem:** $-1 \leq \rho_{xy} \leq 1$.

   Proof:

   Consider the following function of t: $f(t) = \mathrm{V}(Xt - Y)$, which is always greater than or equal to zero because of the definition of variance. Therefore, for all t, we have $f(t) \geq 0$. $f(t)$ is rewritten as follows:

$$
\begin{aligned}
f(t) &= \mathrm{V}(Xt - Y) = \mathrm{V}(Xt) - 2\mathrm{Cov}(Xt, Y) + \mathrm{V}(Y) \\
&= t^2\mathrm{V}(X) - 2t\mathrm{Cov}(X,Y) + \mathrm{V}(Y) \\
&= \mathrm{V}(X)\Big(t - \frac{\mathrm{Cov}(X,Y)}{\mathrm{V}(X)}\Big)^2 + \mathrm{V}(Y) - \frac{(\mathrm{Cov}(X,Y))^2}{\mathrm{V}(X)}.
\end{aligned}
$$

   In order to have $f(t) \geq 0$ for all t, we need the following condition:

$$
\mathrm{V}(Y) - \frac{(\mathrm{Cov}(X,Y))^2}{\mathrm{V}(X)} \geq 0,
$$

   because the first term in the last equality is nonnegative, which implies:

$$
\frac{(\mathrm{Cov}(X,Y))^2}{\mathrm{V}(X)\mathrm{V}(Y)} \leq 1.
$$

   Therefore, we have:

$$
-1 \leq \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{V}(X)}\sqrt{\mathrm{V}(Y)}} \leq 1.
$$

   From the definition of correlation coefficient, i.e., $\rho_{xy} = \dfrac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{V}(X)}\sqrt{\mathrm{V}(Y)}}$, we obtain the result: $-1 \leq \rho_{xy} \leq 1$.

9. **Theorem:** $\mathrm{V}(X \pm Y) = \mathrm{V}(X) + \mathrm{V}(Y)$, when X is independent of Y.

   Proof:

   From the theorem above, $\mathrm{V}(X \pm Y) = \mathrm{V}(X) \pm 2\mathrm{Cov}(X,Y) + \mathrm{V}(Y)$ generally holds. When random variables X and Y are independent, we have $\mathrm{Cov}(X,Y) = 0$. Therefore, $\mathrm{V}(X+Y) = \mathrm{V}(X) + \mathrm{V}(Y)$ holds, when X is independent of Y.

10. **Theorem:** For n random variables $X_1, X_2, \cdots, X_n$,

$$
\begin{aligned}
\mathrm{E}(\sum_i a_i X_i) &= \sum_i a_i \mu_i, \\
\mathrm{V}(\sum_i a_i X_i) &= \sum_i \sum_j a_i a_j \mathrm{Cov}(X_i, X_j),
\end{aligned}
$$

where $\mathrm{E}(X_i) = \mu_i$ and a_i is a constant value. Especially, when $X_1, X_2, \cdots, X_n$ are mutually independent, we have the following:

$$\mathrm{V}(\sum_i a_i X_i) = \sum_i a_i^2 \mathrm{V}(X_i).$$

Proof:

For mean of $\sum_i a_i X_i$, the following representation is obtained.

$$\mathrm{E}(\sum_i a_i X_i) = \sum_i \mathrm{E}(a_i X_i) = \sum_i a_i \mathrm{E}(X_i) = \sum_i a_i \mu_i.$$

The first and second equalities come from the previous theorems on mean.

For variance of $\sum_i a_i X_i$, we can rewrite as follows:

$$\begin{aligned}\mathrm{V}(\sum_i a_i X_i) &= \mathrm{E}\Big(\sum_i a_i (X_i - \mu_i)\Big)^2 = \mathrm{E}\Big(\sum_i a_i (X_i - \mu_i)\Big)\Big(\sum_j a_j (X_j - \mu_j)\Big) \\ &= \mathrm{E}\Big(\sum_i \sum_j a_i a_j (X_i - \mu_i)(X_j - \mu_j)\Big) \\ &= \sum_i \sum_j a_i a_j \mathrm{E}\Big((X_i - \mu_i)(X_j - \mu_j)\Big) = \sum_i \sum_j a_i a_j \mathrm{Cov}(X_i, X_j).\end{aligned}$$

When $X_1, X_2, \cdots, X_n$ are mutually independent, we obtain $\mathrm{Cov}(X_i, X_j) = 0$ for all $i \neq j$ from the previous theorem. Therefore, we obtain:

$$\mathrm{V}(\sum_i a_i X_i) = \sum_i a_i^2 \mathrm{V}(X_i).$$

Note that $\mathrm{Cov}(X_i, X_i) = \mathrm{E}((X_i - \mu)^2) = \mathrm{V}(X_i)$.

11. **Theorem:** n random variables $X_1, X_2, \cdots, X_n$ are mutually independently and identically distributed with mean μ and variance σ^2. That is, for all $i = 1, 2, \cdots, n$, $\mathrm{E}(X_i) = \mu$ and $\mathrm{V}(X_i) = \sigma^2$ are assumed. Consider arithmetic average $\overline{X} = (1/n) \sum_{i=1}^n X_i$. Then, mean and variance of $\overline{X}$ are given by:

$$\mathrm{E}(\overline{X}) = \mu, \qquad \mathrm{V}(\overline{X}) = \frac{\sigma^2}{n}.$$

Proof:

The mathematical expectation of $\overline{X}$ is given by:

$$\mathrm{E}(\overline{X}) = \mathrm{E}(\frac{1}{n} \sum_{i=1}^n X_i) = \frac{1}{n} \mathrm{E}(\sum_{i=1}^n X_i) = \frac{1}{n} \sum_{i=1}^n \mathrm{E}(X_i) = \frac{1}{n} \sum_{i=1}^n \mu = \frac{1}{n} n\mu = \mu.$$

$\mathrm{E}(aX) = a\mathrm{E}(X)$ in the second equality and $\mathrm{E}(X + Y) = \mathrm{E}(X) + \mathrm{E}(Y)$ in the third equality are utilized, where X and Y are random variables and a is a constant value. For these formulas, see p.15 in Section 1.3.1 and p.17 in this section.

The variance of $\overline{X}$ is computed as follows:

$$\begin{aligned}\mathrm{V}(\overline{X}) &= \mathrm{V}(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\mathrm{V}(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{V}(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}n\sigma^2 \\ &= \frac{\sigma^2}{n}.\end{aligned}$$

We use $\mathrm{V}(aX) = a^2\mathrm{V}(X)$ in the second equality and $\mathrm{V}(X + Y) = \mathrm{V}(X) + \mathrm{V}(Y)$ for X independent of Y in the third equality, where X and Y denote random variables and a is a constant value. For these formulas, see p.15 in Section 1.3.1 and p.20 in this section.

1.4 Transformation of Variables

Transformation of variables is used in the case of continuous random variables. Based on a distribution of a random variable, a distribution of the transformed random variable is derived. In other words, when a distribution of X is known, we can find a distribution of Y using the transformation of variables, where Y is a function of X.

1.4.1 Univariate Case

Distribution of $Y = \psi^{-1}(X)$: Let $f_x(x)$ be the probability density function of continuous random variable X and $X = \psi(Y)$ be a one-to-one transformation. Then, the probability density function of Y, i.e., $f_y(y)$, is given by:

$$f_y(y) = |\psi'(y)| f_x\big(\psi(y)\big).$$

We can derive the above transformation of variables from X to Y as follows. Let $f_x(x)$ and $F_x(x)$ be the probability density function and the distribution function of X, respectively. Note that $F_x(x) = P(X \le x)$ and $f_x(x) = F'_x(x)$.

When $X = \psi(Y)$, we want to obtain the probability density function of Y. Let $f_y(y)$ and $F_y(y)$ be the probability density function and the distribution function of Y, respectively.

In the case of $\psi'(X) > 0$, the distribution function of Y, $F_y(y)$, is rewritten as follows:

$$F_y(y) = P(Y \le y) = P\big(\psi(Y) \le \psi(y)\big) = P\big(X \le \psi(y)\big) = F_x\big(\psi(y)\big).$$

The first equality is the definition of the cumulative distribution function. The second equality holds because of $\psi'(Y) > 0$. Therefore, differentiating $F_y(y)$ with respect to y, we can obtain the following expression:

$$f_y(y) = F'_y(y) = \psi'(y)F'_x\big(\psi(y)\big) = \psi'(y)f_x\big(\psi(y)\big). \tag{1.4}$$

Next, in the case of $\psi'(X) < 0$, the distribution function of Y, $F_y(y)$, is rewritten as follows:

$$\begin{aligned} F_y(y) &= P(Y \le y) = P\big(\psi(Y) \ge \psi(y)\big) = P\big(X \ge \psi(y)\big) \\ &= 1 - P\big(X < \psi(y)\big) = 1 - F_x\big(\psi(y)\big). \end{aligned}$$

Thus, in the case of $\psi'(X) < 0$, pay attention to the second equality, where the inequality sign is reversed. Differentiating $F_y(y)$ with respect to y, we obtain the following result:

$$f_y(y) = F_y'(y) = -\psi'(y)F_x'\big(\psi(y)\big) = -\psi'(y)f_x\big(\psi(y)\big). \tag{1.5}$$

Note that $-\psi'(y) > 0$.

Thus, summarizing the above two cases, i.e., $\psi'(X) > 0$ and $\psi'(X) < 0$, equations (1.4) and (1.5) indicate the following result:

$$f_y(y) = |\psi'(y)| f_x\big(\psi(y)\big),$$

which is called the **transformation of variables**.

Example 1.9: When X has a standard normal density function, i.e., when $X \sim N(0,1)$, we derive the probability density function of Y, where $Y = \mu + \sigma X$.

Since we have:

$$X = \psi(Y) = \frac{Y - \mu}{\sigma},$$

$\psi'(y) = 1/\sigma$ is obtained. Therefore, the density function of Y, $f_y(y)$, is given by:

$$f_y(y) = |\psi'(y)| f_x\big(\psi(y)\big) = \frac{1}{\sigma\sqrt{2\pi}} \exp\Big(-\frac{1}{2\sigma^2}(y-\mu)^2\Big),$$

which indicates the normal distribution with mean μ and variance σ^2, denoted by $N(\mu, \sigma^2)$.

On Distribution of $Y = X^2$: As an example, when we know the distribution function of X as $F_x(x)$, we want to obtain the distribution function of Y, $F_y(y)$, where $Y = X^2$. Using $F_x(x)$, $F_y(y)$ is rewritten as follows:

$$\begin{aligned} F_y(y) &= P(Y \le y) = P(X^2 \le y) = P(-\sqrt{y} \le X \le \sqrt{y}) \\ &= F_x(\sqrt{y}) - F_x(-\sqrt{y}). \end{aligned}$$

Therefore, when we have $f_x(x)$ and $Y = X^2$, the probability density function of Y is obtained as follows:

$$f_y(y) = F_y'(y) = \frac{1}{2\sqrt{y}}\big(f_x(\sqrt{y}) + f_x(-\sqrt{y})\big).$$

1.4.2 Multivariate Cases

Bivariate Case: Let $f_{xy}(x, y)$ be a joint probability density function of X and Y. Let $X = \psi_1(U, V)$ and $Y = \psi_2(U, V)$ be a one-to-one transformation from (X, Y) to (U, V). Then, we obtain a joint probability density function of U and V, denoted by $f_{uv}(u, v)$, as follows:

$$f_{uv}(u, v) = |J| f_{xy}\Big(\psi_1(u, v), \psi_2(u, v)\Big),$$

where J is called the **Jacobian** of the transformation, which is defined as:

$$J = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix}.$$

Multivariate Case: Let $f_x(x_1, x_2, \cdots, x_n)$ be a joint probability density function of X_1, X_2, $\cdots$ X_n. Suppose that a one-to-one transformation from $(X_1, X_2, \cdots, X_n)$ to $(Y_1, Y_2, \cdots, Y_n)$ is given by:

$$\begin{aligned} X_1 &= \psi_1(Y_1, Y_2, \cdots, Y_n), \\ X_2 &= \psi_2(Y_1, Y_2, \cdots, Y_n), \\ &\vdots \\ X_n &= \psi_n(Y_1, Y_2, \cdots, Y_n). \end{aligned}$$

Then, we obtain a joint probability density function of Y_1, Y_2, $\cdots$, Y_n, denoted by $f_y(y_1, y_2, \cdots, y_n)$, as follows:

$$f_y(y_1, y_2, \cdots, y_n) = |J| f_x\Big(\psi_1(y_1, \cdots, y_n), \psi_2(y_1, \cdots, y_n), \cdots, \psi_n(y_1, \cdots, y_n)\Big),$$

where J is called the Jacobian of the transformation, which is defined as:

$$J = \begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} & \cdots & \dfrac{\partial x_1}{\partial y_n} \\ \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} & \cdots & \dfrac{\partial x_2}{\partial y_n} \\ \vdots & \vdots & \ddots & \vdots \\ \dfrac{\partial x_n}{\partial y_1} & \dfrac{\partial x_n}{\partial y_2} & \cdots & \dfrac{\partial x_n}{\partial y_n} \end{vmatrix}.$$

1.5 Moment-Generating Function

1.5.1 Univariate Case

As discussed in Section 1.3.1, the moment-generating function is defined as $\phi(\theta) = \mathrm{E}(e^{\theta X})$. In this section, several important theorems and remarks of the moment-generating function are summarized.

For a random variable X, $\mu'_n \equiv \mathrm{E}(X^n)$ is called the ***n*th moment** of X. Then, we have the following first theorem.

1. **Theorem:** $\phi^{(n)}(0) = \mu'_n \equiv \mathrm{E}(X^n)$.

   Proof:

   First, from the definition of the moment-generating function, $\phi(\theta)$ is written as:

   $$\phi(\theta) = \mathrm{E}(e^{\theta X}) = \int_{-\infty}^{\infty} e^{\theta x} f(x)\,\mathrm{d}x.$$

   The nth derivative of $\phi(\theta)$, denoted by $\phi^{(n)}(\theta)$, is:

   $$\phi^{(n)}(\theta) = \int_{-\infty}^{\infty} x^n e^{\theta x} f(x)\,\mathrm{d}x.$$

   Evaluating $\phi^{(n)}(\theta)$ at $\theta = 0$, we obtain:

   $$\phi^{(n)}(0) = \int_{-\infty}^{\infty} x^n f(x)\,\mathrm{d}x = \mathrm{E}(X^n) \equiv \mu'_n,$$

   where the second equality comes from the definition of the mathematical expectation.

2. **Remark:** Let X and Y be two random variables. When the moment-generating function of X is equivalent to that of Y, we have the fact that X has the same distribution as Y.

3. **Theorem:** Let $\phi(\theta)$ be the moment-generating function of X. Then, the moment-generating function of Y, where $Y = aX + b$, is given by $e^{b\theta}\phi(a\theta)$.

   Proof:

   Let $\phi_y(\theta)$ be the moment-generating function of Y. Then, $\phi_y(\theta)$ is rewritten as follows:

   $$\phi_y(\theta) = \mathrm{E}(e^{\theta Y}) = \mathrm{E}(e^{\theta(aX+b)}) = e^{b\theta}\mathrm{E}(e^{a\theta X}) = e^{b\theta}\phi(a\theta).$$

   Note that $\phi(\theta)$ represents the moment-generating function of X.

4. **Theorem:** Let $\phi_1(\theta)$, $\phi_2(\theta)$, $\cdots$, $\phi_n(\theta)$ be the moment-generating functions of $X_1, X_2, \cdots, X_n$, which are mutually independently distributed random variables. Define $Y = X_1 + X_2 + \cdots + X_n$. Then, the moment-generating function of Y is given by $\phi_1(\theta)\phi_2(\theta)\cdots\phi_n(\theta)$, i.e.,

   $$\phi_y(\theta) = \mathrm{E}(e^{\theta Y}) = \phi_1(\theta)\phi_2(\theta)\cdots\phi_n(\theta),$$

   where $\phi_y(\theta)$ represents the moment-generating function of Y.

Proof:

The moment-generating function of Y, i.e., $\phi_y(\theta)$, is rewritten as:

$$\begin{aligned}\phi_y(\theta) &= \mathrm{E}(e^{\theta Y}) = \mathrm{E}(e^{\theta(X_1+X_2+\cdots+X_n)}) = \mathrm{E}(e^{\theta X_1})\mathrm{E}(e^{\theta X_2})\cdots\mathrm{E}(e^{\theta X_n}) \\ &= \phi_1(\theta)\phi_2(\theta)\cdots\phi_n(\theta).\end{aligned}$$

The third equality holds because X_1, X_2, $\cdots$, X_n are mutually independently distributed random variables.

5. **Theorem:** When X_1, X_2, $\cdots$, X_n are mutually independently and identically distributed and the moment-generating function of X_i is given by $\phi(\theta)$ for all i, the moment-generating function of Y is represented by $\big(\phi(\theta)\big)^n$, where $Y = X_1 + X_2 + \cdots + X_n$.

   Proof:

   Using the above theorem, we have the following:

   $$\phi_y(\theta) = \phi_1(\theta)\phi_2(\theta)\cdots\phi_n(\theta) = \phi(\theta)\phi(\theta)\cdots\phi(\theta) = \big(\phi(\theta)\big)^n.$$

   Note that $\phi_i(\theta) = \phi(\theta)$ for all i.

6. **Theorem:** When X_1, X_2, $\cdots$, X_n are mutually independently and identically distributed and the moment-generating function of X_i is given by $\phi(\theta)$ for all i, the moment-generating function of $\overline{X}$ is represented by $\Big(\phi(\dfrac{\theta}{n})\Big)^n$, where $\overline{X} = (1/n)\sum_{i=1}^n X_i$.

   Proof:

   Let $\phi_{\overline{x}}(\theta)$ be the moment-generating function of $\overline{X}$.

   $$\phi_{\overline{x}}(\theta) = \mathrm{E}(e^{\theta\overline{X}}) = \mathrm{E}(e^{\frac{\theta}{n}\sum_{i=1}^n X_i}) = \prod_{i=1}^n \mathrm{E}(e^{\frac{\theta}{n}X_i}) = \prod_{i=1}^n \phi(\frac{\theta}{n}) = \Big(\phi(\frac{\theta}{n})\Big)^n.$$

Example 1.10: For the binomial random variable, the moment-generating function $\phi(\theta)$ is known as:

$$\phi(\theta) = (pe^\theta + 1 - p)^n,$$

which is discussed in Example 1.5 (Section 1.3.1). Using the moment-generating function, we check whether $\mathrm{E}(X) = np$ and $\mathrm{V}(X) = np(1-p)$ are obtained when X is a binomial random variable.

The first- and the second-derivatives with respect to θ are given by:

$$\begin{aligned}\phi'(\theta) &= npe^\theta(pe^\theta + 1 - p)^{n-1}, \\ \phi''(\theta) &= npe^\theta(pe^\theta + 1 - p)^{n-1} + n(n-1)p^2e^{2\theta}(pe^\theta + 1 - p)^{n-2}.\end{aligned}$$

Evaluating at $\theta = 0$, we have:

$$\mathrm{E}(X) = \phi'(0) = np, \quad \mathrm{E}(X^2) = \phi''(0) = np + n(n-1)p^2.$$

Therefore, $\mathrm{V}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = np(1-p)$ can be derived. Thus, we can make sure that $\mathrm{E}(X)$ and $\mathrm{V}(X)$ are obtained from $\phi(\theta)$.

1.5.2 Multivariate Cases

Bivariate Case: As discussed in Section 1.3.2, for two random variables X and Y, the moment-generating function is defined as $\phi(\theta_1, \theta_2) = \mathrm{E}(e^{\theta_1 X + \theta_2 Y})$. Some useful and important theorems and remarks are shown as follows.

1. **Theorem:** Consider two random variables X and Y. Let $\phi(\theta_1, \theta_2)$ be the moment-generating function of X and Y. Then, we have the following result:

$$\frac{\partial^{j+k}\phi(0,0)}{\partial\theta_1^j \partial\theta_2^k} = \mathrm{E}(X^j Y^k).$$

Proof:

Let $f_{xy}(x, y)$ be the probability density function of X and Y. From the definition, $\phi(\theta_1, \theta_2)$ is written as:

$$\phi(\theta_1, \theta_2) = \mathrm{E}(e^{\theta_1 X + \theta_2 Y}) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{\theta_1 x + \theta_2 y} f_{xy}(x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

Taking the jth derivative of $\phi(\theta_1, \theta_2)$ with respect to θ_1 and at the same time the kth derivative with respect to θ_2, we have the following expression:

$$\frac{\partial^{j+k}\phi(\theta_1, \theta_2)}{\partial\theta_1^j \partial\theta_2^k} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^j y^k e^{\theta_1 x + \theta_2 y} f_{xy}(x, y)\,\mathrm{d}x\,\mathrm{d}y.$$

Evaluating the above equation at $(\theta_1, \theta_2) = (0, 0)$, we can easily obtain:

$$\frac{\partial^{j+k}\phi(0,0)}{\partial\theta_1^j \partial\theta_2^k} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^j y^k f_{xy}(x, y)\,\mathrm{d}x\,\mathrm{d}y \equiv \mathrm{E}(X^j Y^k).$$

2. **Remark:** Let (X_i, Y_i) be a pair of random variables. Suppose that the moment-generating function of (X_1, Y_1) is equivalent to that of (X_2, Y_2). Then, (X_1, Y_1) has the same distribution function as (X_2, Y_2).

3. **Theorem:** Let $\phi(\theta_1, \theta_2)$ be the moment-generating function of (X, Y). The moment-generating function of X is given by $\phi_1(\theta_1)$ and that of Y is $\phi_2(\theta_2)$. Then, we have the following facts:

$$\phi_1(\theta_1) = \phi(\theta_1, 0), \quad \phi_2(\theta_2) = \phi(0, \theta_2).$$

Proof:

Again, the definition of the moment-generating function of X and Y is represented as:

$$\phi(\theta_1, \theta_2) = \mathrm{E}(e^{\theta_1 X + \theta_2 Y}) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\theta_1 x + \theta_2 y} f_{xy}(x, y) \, \mathrm{d}x \, \mathrm{d}y.$$

When $\phi(\theta_1, \theta_2)$ is evaluated at $\theta_2 = 0$, $\phi(\theta_1, 0)$ is rewritten as follows:

$$\begin{aligned}
\phi(\theta_1, 0) = \mathrm{E}(e^{\theta_1 X}) &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{\theta_1 x} f_{xy}(x, y) \, \mathrm{d}x \, \mathrm{d}y \\
&= \int_{-\infty}^{\infty} e^{\theta_1 x} \Big(\int_{-\infty}^{\infty} f_{xy}(x, y) \, \mathrm{d}y \Big) \, \mathrm{d}x \\
&= \int_{-\infty}^{\infty} e^{\theta_1 x} f_x(x) \, \mathrm{d}x = \mathrm{E}(e^{\theta_1 X}) = \phi_1(\theta_1).
\end{aligned}$$

Thus, we obtain the result: $\phi(\theta_1, 0) = \phi_1(\theta_1)$. Similarly, $\phi(0, \theta_2) = \phi_2(\theta_2)$ can be derived.

4. **Theorem:** The moment-generating function of (X, Y) is given by $\phi(\theta_1, \theta_2)$. Let $\phi_1(\theta_1)$ and $\phi_2(\theta_2)$ be the moment-generating functions of X and Y, respectively. If X is independent of Y, we have:

$$\phi(\theta_1, \theta_2) = \phi_1(\theta_1)\phi_2(\theta_2).$$

Proof:

From the definition of $\phi(\theta_1, \theta_2)$, the moment-generating function of X and Y is rewritten as follows:

$$\phi(\theta_1, \theta_2) = \mathrm{E}(e^{\theta_1 X + \theta_2 Y}) = \mathrm{E}(e^{\theta_1 X})\mathrm{E}(e^{\theta_2 Y}) = \phi_1(\theta_1)\phi_2(\theta_2).$$

The second equality holds because X is independent of Y.

Multivariate Case: For multivariate random variables $X_1, X_2, \cdots, X_n$, the moment-generating function is defined as:

$$\phi(\theta_1, \theta_2, \cdots, \theta_n) = \mathrm{E}(e^{\theta_1 X_1 + \theta_2 X_2 + \cdots + \theta_n X_n}).$$

1. **Theorem:** If the multivariate random variables X_1, X_2, $\cdots$, X_n are mutually independent, the moment-generating function of X_1, X_2, $\cdots$, X_n, denoted by $\phi(\theta_1, \theta_2, \cdots, \theta_n)$, is given by:

$$\phi(\theta_1, \theta_2, \cdots, \theta_n) = \phi_1(\theta_1)\phi_2(\theta_2) \cdots \phi_n(\theta_n),$$

where $\phi_i(\theta) = \mathrm{E}(e^{\theta X_i})$.

Proof:

From the definition of the moment-generating function in the multivariate cases, we obtain the following:

$$\begin{aligned}\phi(\theta_1, \theta_2, \cdots, \theta_n) &= \mathrm{E}(e^{\theta_1 X_1 + \theta_2 X_2 + \cdots + \theta_n X_n}) = \mathrm{E}(e^{\theta_1 X_1})\mathrm{E}(e^{\theta_2 X_2}) \cdots \mathrm{E}(e^{\theta_n X_n}) \\ &= \phi_1(\theta_1)\phi_2(\theta_2) \cdots \phi_n(\theta_n).\end{aligned}$$

2. **Theorem:** Suppose that the multivariate random variables X_1, X_2, $\cdots$, X_n are mutually independently and identically distributed. X_i has a normal distribution with mean μ and variance σ^2, i.e., $X_i \sim N(\mu, \sigma^2)$. Let us define $\hat{\mu} = \sum_{i=1}^n a_i X_i$, where a_i, $i = 1, 2, \cdots, n$, are assumed to be known. Then, $\hat{\mu}$ has a normal distribution with mean $\mu \sum_{i=1}^n a_i$ and variance $\sigma^2 \sum_{i=1}^n a_i^2$, i.e., $\hat{\mu} \sim N(\mu \sum_{i=1}^n a_i, \sigma^2 \sum_{i=1}^n a_i^2)$.

   Proof:

   From Example 1.8 (p.15) and Example 1.9 (p.23), it is shown that the moment-generating function of X is given by: $\phi_x(\theta) = \exp(\mu\theta + \frac{1}{2}\sigma^2\theta^2)$, when X is normally distributed as $X \sim N(\mu, \sigma^2)$.

   Let $\phi_{\hat{\mu}}$ be the moment-generating function of $\hat{\mu}$.

$$\begin{aligned}\phi_{\hat{\mu}}(\theta) &= \mathrm{E}(e^{\theta\hat{\mu}}) = \mathrm{E}(e^{\theta \sum_{i=1}^n a_i X_i}) = \prod_{i=1}^n \mathrm{E}(e^{\theta a_i X_i}) = \prod_{i=1}^n \phi_x(a_i\theta) \\ &= \prod_{i=1}^n \exp(\mu a_i \theta + \frac{1}{2}\sigma^2 a_i^2 \theta^2) = \exp(\mu \sum_{i=1}^n a_i \theta + \frac{1}{2}\sigma^2 \sum_{i=1}^n a_i^2 \theta^2)\end{aligned}$$

   which is equivalent to the moment-generating function of the normal distribution with mean $\mu \sum_{i=1}^n a_i$ and variance $\sigma^2 \sum_{i=1}^n a_i^2$, where μ and σ^2 in $\phi_x(\theta)$ is simply replaced by $\mu \sum_{i=1}^n a_i$ and $\sigma^2 \sum_{i=1}^n a_i^2$ in $\phi_{\hat{\mu}}(\theta)$, respectively.

   Moreover, note as follows. When $a_i = 1/n$ is taken for all $i = 1, 2, \cdots, n$, i.e., when $\hat{\mu} = \overline{X}$ is taken, $\hat{\mu} = \overline{X}$ is normally distributed as: $\overline{X} \sim N(\mu, \sigma^2/n)$. The readers should check difference between Theorem 11 on p.21 and this theorem.

1.6 Law of Large Numbers and Central Limit Theorem

1.6.1 Chebyshev's Inequality

In this section, we introduce Chebyshev's inequality, which enables us to find upper and lower bounds given a certain probability.

Theorem: Let $g(X)$ be a nonnegative function of the random variable X, i.e., $g(X) \geq 0$. If $\mathrm{E}(g(X))$ exists, then we have:

$$P(g(X) \geq k) \leq \frac{\mathrm{E}(g(X))}{k}, \tag{1.6}$$

for a positive constant value k.

Proof:
We define the discrete random variable U as follows:

$$U = \begin{cases} 1, & \text{if } g(X) \geq k, \\ 0, & \text{if } g(X) < k. \end{cases}$$

Thus, the discrete random variable U takes 0 or 1. Suppose that the probability function of U is given by:

$$f(u) = P(U = u),$$

where $P(U = u)$ is represented as:

$$P(U = 1) = P(g(X) \geq k),$$
$$P(U = 0) = P(g(X) < k).$$

Then, in spite of the value which U takes, the following equation always holds:

$$g(X) \geq kU,$$

which implies that we have $g(X) \geq k$ when $U = 1$ and $g(X) \geq 0$ when $U = 0$, where k is a positive constant value. Therefore, taking the expectation on both sides, we obtain:

$$\mathrm{E}(g(X)) \geq k\mathrm{E}(U), \tag{1.7}$$

where $\mathrm{E}(U)$ is given by:

$$\begin{aligned} \mathrm{E}(U) &= \sum_{u=0}^{1} uP(U = u) = 1 \times P(U = 1) + 0 \times P(U = 0) = P(U = 1) \\ &= P(g(X) \geq k). \end{aligned} \tag{1.8}$$

Accordingly, substituting equation (1.8) into equation (1.7), we have the following inequality:

$$P(g(X) \geq k) \leq \frac{\mathrm{E}(g(X))}{k}.$$

Chebyshev's Inequality: Assume that $\mathrm{E}(X) = \mu$, $\mathrm{V}(X) = \sigma^2$, and λ is a positive constant value. Then, we have the following inequality:

$$P(|X - \mu| \geq \lambda\sigma) \leq \frac{1}{\lambda^2},$$

or equivalently,

$$P(|X - \mu| < \lambda\sigma) \geq 1 - \frac{1}{\lambda^2},$$

which is called **Chebyshev's inequality**.

Proof:
Take $g(X) = (X - \mu)^2$ and $k = \lambda^2\sigma^2$. Then, we have:

$$P((X - \mu)^2 \geq \lambda^2\sigma^2) \leq \frac{\mathrm{E}(X - \mu)^2}{\lambda^2\sigma^2},$$

which implies

$$P(|X - \mu| \geq \lambda\sigma) \leq \frac{1}{\lambda^2}.$$

Note that $\mathrm{E}(X - \mu)^2 = \mathrm{V}(X) = \sigma^2$.

Since we have $P(|X-\mu| \geq \lambda\sigma) + P(|X-\mu| < \lambda\sigma) = 1$, we can derive the following inequality:

$$P(|X - \mu| < \lambda\sigma) \geq 1 - \frac{1}{\lambda^2}. \tag{1.9}$$

An Interpretation of Chebyshev's inequality: $1/\lambda^2$ is an upper bound for the probability $P(|X - \mu| \geq \lambda\sigma)$. Equation (1.9) is rewritten as:

$$P(\mu - \lambda\sigma < X < \mu + \lambda\sigma) \geq 1 - \frac{1}{\lambda^2}.$$

That is, the probability that X falls within $\lambda\sigma$ units of μ is greater than or equal to $1 - 1/\lambda^2$. Taking an example of $\lambda = 2$, the probability that X falls within two standard deviations of its mean is at least 0.75.

Furthermore, note as follows. Taking $\epsilon = \lambda\sigma$, we obtain as follows:

$$P(|X - \mu| \geq \epsilon) \leq \frac{\sigma^2}{\epsilon^2},$$

i.e.,

$$P(|X - \mathrm{E}(X)| \geq \epsilon) \leq \frac{\mathrm{V}(X)}{\epsilon^2}, \tag{1.10}$$

which inequality is used in the next section.

Remark: Equation (1.10) can be derived when we take $g(X) = (X - \mu)^2$, $\mu = \mathrm{E}(X)$ and $k = \epsilon^2$ in equation (1.6). Even when we have $\mu \neq \mathrm{E}(X)$, the following inequality still hold:

$$P(|X - \mu| \geq \epsilon) \leq \frac{\mathrm{E}((X - \mu)^2)}{\epsilon^2}.$$

Note that $\mathrm{E}((X - \mu)^2)$ represents the mean square error (MSE). When $\mu = \mathrm{E}(X)$, the mean square error reduces to the variance.

1.6.2 Law of Large Numbers (Convergence in probability)

Law of Large Numbers: Assume that $X_1, X_2, \cdots, X_n$ are mutually independently and identically distributed with mean $\mathrm{E}(X_i) = \mu$ and variance $\mathrm{V}(X_i) = \sigma^2 < \infty$ for all i. Then, for any positive value ϵ, as $n \longrightarrow \infty$, we have the following result:

$$P(|\overline{X}_n - \mu| > \epsilon) \longrightarrow 0,$$

where $\overline{X}_n = (1/n) \sum_{i=1}^n X_i$. We say that $\overline{X}_n$ converges to μ in probability.

Proof:
Using (1.10), Chebyshev's inequality is represented as follows:

$$P(|\overline{X}_n - \mathrm{E}(\overline{X}_n)| > \epsilon) \leq \frac{\mathrm{V}(\overline{X}_n)}{\epsilon^2},$$

where X in (1.10) is replaced by $\overline{X}_n$. As in Section 1.3.2 (p.21), we have $\mathrm{E}(\overline{X}_n) = \mu$ and $\mathrm{V}(\overline{X}_n) = \sigma^2/n$, which are substituted into the above inequality. Then, we obtain:

$$P(|\overline{X}_n - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2}.$$

Accordingly, when $n \longrightarrow \infty$, the following equation holds:

$$P(|\overline{X}_n - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \longrightarrow 0.$$

That is, $\overline{X}_n \longrightarrow \mu$ is obtained as $n \longrightarrow \infty$, which is written as: $\operatorname{plim} \overline{X}_n = \mu$. This theorem is called the **law of large numbers**.

The condition $P(|\overline{X}_n - \mu| > \epsilon) \longrightarrow 0$ or equivalently $P(|\overline{X}_n - \mu| < \epsilon) \longrightarrow 1$ is used as the definition of **convergence in probability**. In this case, we say that $\overline{X}_n$ converges to μ in probability.

Theorem: In the case where $X_1, X_2, \cdots, X_n$ are not identically distributed and they are not mutually independently distributed, we assume that

$$\begin{aligned} m_n &= \mathrm{E}(\sum_{i=1}^{n} X_i) < \infty, \\ V_n &= \mathrm{V}(\sum_{i=1}^{n} X_i) < \infty, \\ \frac{V_n}{n^2} &\longrightarrow 0, \quad \text{as } n \longrightarrow \infty. \end{aligned}$$

Then, we obtain the following result:

$$\frac{\sum_{i=1}^{n} X_i - m_n}{n} \longrightarrow 0.$$

That is, $\overline{X}_n$ converges to $\lim_{n\to\infty} \dfrac{m_n}{n}$ in probability. This theorem is also called the law of large numbers.

1.6.3 Central Limit Theorem

Central Limit Theorem: $X_1, X_2, \cdots, X_n$ are mutually independently and identically distributed with $\mathrm{E}(X_i) = \mu$ and $\mathrm{V}(X_i) = \sigma^2$ for all i. Both μ and σ^2 are finite. Under the above assumptions, when $n \longrightarrow \infty$, we have:

$$P\Big(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} < x\Big) \longrightarrow \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, \mathrm{d}u,$$

which is called the **central limit theorem**.

Proof:

Define $Y_i = \dfrac{X_i - \mu}{\sigma}$. We can rewrite as follows:

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - \mu}{\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i.$$

Since $Y_1, Y_2, \cdots, Y_n$ are mutually independently and identically distributed, the moment-generating function of Y_i is identical for all i, which is denoted by $\phi(\theta)$. Using $\mathrm{E}(Y_i) = 0$ and $\mathrm{V}(Y_i) = 1$, the moment-generating function of Y_i, $\phi(\theta)$, is rewritten as:

$$\begin{aligned} \phi(\theta) &= \mathrm{E}(e^{Y_i\theta}) = \mathrm{E}\Big(1 + Y_i\theta + \frac{1}{2}Y_i^2\theta^2 + \frac{1}{3!}Y_i^3\theta^3 \cdots\Big) \\ &= 1 + \frac{1}{2}\theta^2 + O(\theta^3). \end{aligned}$$

In the second equality, $e^{Y_i\theta}$ is approximated by the Taylor series expansion around $\theta = 0$. See Appendix 1.3 for the Taylor series expansion. $O(x)$ implies that it is a polynomial function of x and the higher-order terms but it is dominated by x. In this case, $O(\theta^3)$ is a function of $\theta^3, \theta^4, \cdots$. Since the moment-generating function is conventionally evaluated at $\theta = 0$, θ^3 is the largest value of $\theta^3, \theta^4, \cdots$ and accordingly $O(\theta^3)$ is dominated by θ^3 (in other words, $\theta^4, \theta^5, \cdots$ are small enough, compared with θ^3).

Define Z as:

$$Z = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i.$$

Then, the moment-generating function of Z, i.e., $\phi_z(\theta)$, is given by:

$$\begin{aligned}
\phi_z(\theta) &= \mathrm{E}(e^{Z\theta}) = \mathrm{E}\Big(e^{\frac{\theta}{\sqrt{n}}\sum_{i=1}^{n} Y_i}\Big) = \prod_{i=1}^{n} \mathrm{E}\Big(e^{\frac{\theta}{\sqrt{n}} Y_i}\Big) = \Big(\phi(\frac{\theta}{\sqrt{n}})\Big)^n \\
&= \Big(1 + \frac{1}{2}\frac{\theta^2}{n} + O(\frac{\theta^3}{n^{\frac{3}{2}}})\Big)^n = \Big(1 + \frac{1}{2}\frac{\theta^2}{n} + O(n^{-\frac{3}{2}})\Big)^n.
\end{aligned}$$

We consider that n goes to infinity. Therefore, $O(\frac{\theta^3}{n^{\frac{3}{2}}})$ indicates a function of $n^{-\frac{3}{2}}$.

Moreover, consider $x = \frac{1}{2}\frac{\theta^2}{n} + O(n^{-\frac{3}{2}})$. Multiply n/x on both sides of $x = \frac{1}{2}\frac{\theta^2}{n} + O(n^{-\frac{3}{2}})$. Then, we obtain $n = \frac{1}{x}\Big(\frac{1}{2}\theta^2 + O(n^{-\frac{1}{2}})\Big)$. Substitute $n = \frac{1}{x}\Big(\frac{1}{2}\theta^2 + O(n^{-\frac{1}{2}})\Big)$ into the moment-generating function of Z, i.e., $\phi_z(\theta)$. Then, we obtain:

$$\begin{aligned}
\phi_z(\theta) &= \Big(1 + \frac{1}{2}\frac{\theta^2}{n} + O(n^{-\frac{3}{2}})\Big)^n = (1 + x)^{\frac{1}{x}(\frac{\theta^2}{2} + O(n^{-\frac{1}{2}}))} \\
&= \Big((1 + x)^{\frac{1}{x}}\Big)^{\frac{\theta^2}{2} + O(n^{-\frac{1}{2}})} \longrightarrow e^{\frac{\theta^2}{2}}.
\end{aligned}$$

Note that $x \longrightarrow 0$ when $n \longrightarrow \infty$ and that $\lim_{x\to 0}(1 + x)^{1/x} = e$ as in Section 1.2.3 (p.12). Furthermore, we have $O(n^{-\frac{1}{2}}) \longrightarrow 0$ as $n \longrightarrow \infty$.

Since $\phi_z(\theta) = e^{\frac{\theta^2}{2}}$ is the moment-generating function of the standard normal distribution (see p.14 in Section 1.3.1 for the moment-generating function of the standard normal probability density), we have:

$$P\Big(\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} < x\Big) \longrightarrow \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\, \mathrm{d}u,$$

or equivalently,

$$\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \longrightarrow N(0, 1).$$

The following expression is also possible:

$$\sqrt{n}(\overline{X}_n - \mu) \longrightarrow N(0, \sigma^2). \tag{1.11}$$

Corollary 1: When $\mathrm{E}(X_i) = \mu$, $\mathrm{V}(X_i) = \sigma^2$ and $\overline{X}_n = (1/n)\sum_{i=1}^n X_i$, note that

$$\frac{\overline{X}_n - \mathrm{E}(\overline{X}_n)}{\sqrt{\mathrm{V}(\overline{X}_n)}} = \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}.$$

Therefore, we can rewrite the above theorem as:

$$P\Big(\frac{\overline{X}_n - \mathrm{E}(\overline{X}_n)}{\sqrt{\mathrm{V}(\overline{X}_n)}} < x\Big) \longrightarrow \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\,\mathrm{d}u.$$

Corollary 2: Consider the case where $X_1, X_2, \cdots, X_n$ are not identically distributed and they are not mutually independently distributed. Assume that

$$\lim_{n\to\infty} n\mathrm{V}(\overline{X}_n) = \sigma^2 < \infty,$$

where $\overline{X}_n = (1/n)\sum_{i=1}^n X_i$. Then, when $n \longrightarrow \infty$, we have:

$$P\Big(\frac{\overline{X}_n - \mathrm{E}(\overline{X}_n)}{\sqrt{\mathrm{V}(\overline{X}_n)}} < x\Big) = P\Big(\frac{\sum_{i=1}^n X_i - \mathrm{E}(\sum_{i=1}^n X_i)}{\sqrt{\mathrm{V}(\sum_{i=1}^n X_i)}} < x\Big) \longrightarrow \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}u^2}\,\mathrm{d}u.$$

1.7 Statistical Inference

1.7.1 Point Estimation

Suppose that the functional form of the underlying distribution on population is known but the parameter θ included in the distribution is not known. The distribution function of population is given by $f(x;\theta)$. Let $x_1, x_2, \cdots, x_n$ be the n observed data drawn from the population distribution. Consider estimating the parameter θ using the n observed data. Let $\hat{\theta}_n(x_1, x_2, \cdots, x_n)$ be a function of the observed data $x_1, x_2, \cdots, x_n$. Suppose that $\hat{\theta}_n(x_1, x_2, \cdots, x_n)$ is constructed from the purpose of estimating the parameter θ. $\hat{\theta}_n(x_1, x_2, \cdots, x_n)$ takes a certain value given the n observed data. Then, $\hat{\theta}_n(x_1, x_2, \cdots, x_n)$ is called the **point estimate** of θ, or simply the **estimate** of θ.

Example 1.11: Consider the case of $\theta = (\mu, \sigma^2)$, where the unknown parameters contained in population is given by mean and variance. A point estimate of population mean μ is given by:

$$\hat{\mu}_n(x_1, x_2, \cdots, x_n) \equiv \overline{x} = \frac{1}{n}\sum_{i=1}^n x_i.$$

A point estimate of population variance σ^2 is:

$$\hat{\sigma}_n^2(x_1, x_2, \cdots, x_n) \equiv s^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i - \overline{x})^2.$$

An alternative point estimate of population variance σ^2 is:

$$\widetilde{\sigma}_n^2(x_1, x_2, \cdots, x_n) \equiv s^{**2} = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

1.7.2 Statistic, Estimate and Estimator

The underlying distribution of population is assumed to be known, but the parameter θ, which characterizes the underlying distribution, is unknown. The probability density function of population is given by $f(x;\theta)$. Let X_1, X_2, $\cdots$, X_n be a subset of population, which are regarded as the random variables and are assumed to be mutually independent. x_1, x_2, $\cdots$, x_n are taken as the experimental values of the random variables X_1, X_2, $\cdots$, X_n. In statistics, we consider that n-variate random variables X_1, X_2, $\cdots$, X_n takes the experimental values x_1, x_2, $\cdots$, x_n by chance. There, the experimental values and the actually observed data series are used in the same meaning.

As discussed in Section 1.7.1, $\hat{\theta}_n(x_1, x_2, \cdots, x_n)$ denotes the point estimate of θ. In the case where the observed data x_1, x_2, $\cdots$, x_n are replaced by the corresponding random variables X_1, X_2, $\cdots$, X_n, a function of X_1, X_2, $\cdots$, X_n, i.e., $\hat{\theta}(X_1, X_2, \cdots, X_n)$, is called the **estimator** of θ, which should be distinguished from the **estimate** of θ, i.e., $\hat{\theta}(x_1, x_2, \cdots, x_n)$.

Example 1.12: Let X_1, X_2, $\cdots$, X_n denote a random sample of n from a given distribution $f(x;\theta)$. Consider the case of $\theta = (\mu, \sigma^2)$.

The estimator of μ is given by $\overline{X} = (1/n)\sum_{i=1}^n X_i$, while the estimate of μ is $\overline{x} = (1/n)\sum_{i=1}^n x_i$. The estimator of σ^2 is $S^2 = \sum_{i=1}^n (X_i - \overline{X})^2/(n-1)$ and the estimate of σ^2 is $s^2 = \sum_{i=1}^n (x_i - \overline{x})^2/(n-1)$.

There are numerous estimators and estimates of θ. All of $(1/n)\sum_{i=1}^n X_i$, $(X_1+X_n)/2$, median of $(X_1, X_2, \cdots, X_n)$ and so on are taken as the estimators of μ. Of course, they are called the estimates of θ when X_i is replaced by x_i for all i. Similarly, both $S^2 = \sum_{i=1}^n (X_i - \overline{X})^2/(n-1)$ and $S^{*2} = \sum_{i=1}^2 (X_i - \overline{X})^2/n$ are the estimators of σ^2. We need to choose one out of the numerous estimators of θ. The problem of choosing an optimal estimator out of the numerous estimators is discussed in Sections 1.7.4 and 1.7.5.

In addition, note as follows. A function of random variables is called a **statistic**. The statistic for estimation of the parameter is called an estimator. Therefore, an estimator is a family of a statistic.

1.7.3 Estimation of Mean and Variance

Suppose that the population distribution is given by $f(x;\theta)$. The random sample X_1, X_2, $\cdots$, X_n are assumed to be drawn from the population distribution $f(x;\theta)$, where $\theta = (\mu, \sigma^2)$. Therefore, we can assume that X_1, X_2, $\cdots$, X_n are mutually independently and identically distributed, where "identically" implies $\mathrm{E}(X_i) = \mu$ and $\mathrm{V}(X_i) = \sigma^2$ for all i.

Consider the estimators of $\theta = (\mu, \sigma^2)$ as follows.

1. The estimator of population mean μ is:
   - $\overline{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$.
2. The estimators of population variance σ^2 are:
   - $S^{*2} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \mu)^2$, when μ is known,
   - $S^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2$,
   - $S^{**2} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \overline{X})^2$,

Properties of $\overline{X}$: From Theorem on p.21, mean and variance of $\overline{X}$ are obtained as follows:

$$\mathrm{E}(\overline{X}) = \mu, \quad \mathrm{V}(\overline{X}) = \frac{\sigma^2}{n}.$$

Properties of S^{*2}, S^2 and S^{**2}: The expectation of S^{*2} is:

$$\begin{aligned}
\mathrm{E}(S^{*2}) &= \mathrm{E}\Big(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2\Big) = \frac{1}{n}\mathrm{E}\Big(\sum_{i=1}^{n}(X_i - \mu)^2\Big) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\big((X_i - \mu)^2\big) \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathrm{V}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n}n\sigma^2 = \sigma^2,
\end{aligned}$$

where $\mathrm{E}((X_i - \mu)^2) = \mathrm{V}(X_i) = \sigma^2$ is used in the fourth and fifth equalities.

Next, the expectation of S^2 is given by:

$$\begin{aligned}
\mathrm{E}(S^2) &= \mathrm{E}\Big(\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2\Big) = \frac{1}{n-1}\mathrm{E}\Big(\sum_{i=1}^{n}(X_i - \overline{X})^2\Big) \\
&= \frac{1}{n-1}\mathrm{E}\Big(\sum_{i=1}^{n}((X_i - \mu) - (\overline{X} - \mu))^2\Big) \\
&= \frac{1}{n-1}\mathrm{E}\Big(\sum_{i=1}^{n}((X_i - \mu)^2 - 2(X_i - \mu)(\overline{X} - \mu) + (\overline{X} - \mu)^2)\Big) \\
&= \frac{1}{n-1}\mathrm{E}\Big(\sum_{i=1}^{n}(X_i - \mu)^2 - 2(\overline{X} - \mu)\sum_{i=1}^{n}(X_i - \mu) + n(\overline{X} - \mu)^2\Big)
\end{aligned}$$

$$= \frac{1}{n-1}\mathrm{E}\Big(\sum_{i=1}^{n}(X_i-\mu)^2 - n(\overline{X}-\mu)^2\Big)$$
$$= \frac{n}{n-1}\mathrm{E}\Big(\frac{1}{n}\sum_{i=1}^{n}(X_i-\mu)^2\Big) - \frac{n}{n-1}\mathrm{E}((\overline{X}-\mu)^2)$$
$$= \frac{n}{n-1}\sigma^2 - \frac{n}{n-1}\frac{\sigma^2}{n} = \sigma^2.$$

$\sum_{i=1}^{n}(X_i-\mu) = n(\overline{X}-\mu)$ is used in the sixth equality. $\mathrm{E}\Big((1/n)\sum_{i=1}^{n}(X_i-\mu)^2\Big) = \mathrm{E}(S^{*2}) = \sigma^2$ and $\mathrm{E}((\overline{X}-\mu)^2) = \mathrm{V}(\overline{X}) = \sigma^2/n$ are required in the eighth equality.

Finally, the mathematical expectation of S^{**2} is represented by:

$$\mathrm{E}(S^{**2}) = \mathrm{E}\Big(\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\Big) = \mathrm{E}\Big(\frac{n-1}{n}\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2\Big)$$
$$= \mathrm{E}\Big(\frac{n-1}{n}S^2\Big) = \frac{n-1}{n}\mathrm{E}(S^2) = \frac{n-1}{n}\sigma^2 \neq \sigma^2.$$

Summarizing the above results, we obtain as follows:

$$\mathrm{E}(S^{*2}) = \sigma^2, \quad \mathrm{E}(S^2) = \sigma^2, \quad \mathrm{E}(S^{**2}) = \frac{n-1}{n}\sigma^2 \neq \sigma^2.$$

1.7.4 Point Estimation: Optimality

As mentioned in the previous sections, θ denotes the parameter to be estimated. $\hat{\theta}_n(X_1, X_2, \cdots, X_n)$ represents the estimator of θ, while $\hat{\theta}_n(x_1, x_2, \cdots, x_n)$ indicates the estimate of θ. Hereafter, in the case of no confusion, $\hat{\theta}_n(X_1, X_2, \cdots, X_n)$ is simply written as $\hat{\theta}_n$.

As discussed above, there are numerous candidates of the estimator $\hat{\theta}_n$. The desired properties which $\hat{\theta}_n$ have to satisfy include unbiasedness, efficiency and consistency.

Unbiasedness: One of the desirable features that the estimator of the parameter should have is given by:

$$\mathrm{E}(\hat{\theta}_n) = \theta, \tag{1.12}$$

which implies that $\hat{\theta}_n$ is distributed around θ. When the condition (1.12) holds, $\hat{\theta}_n$ is called the **unbiased estimator** of θ. $\mathrm{E}(\hat{\theta}_n) - \theta$ is defined as **bias**.

As an example of unbiasedness, consider the case of $\theta = (\mu, \sigma^2)$. Suppose that $X_1, X_2, \cdots, X_n$ are mutually independently and identically distributed with mean μ and variance σ^2. Consider the following estimators of μ and σ^2.

1. The estimator of μ is:
   - $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i.$

2. The estimators of σ^2 are:
   - $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$,
   - $S^{**2} = \dfrac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$.

Since we have obtained $\mathrm{E}(\overline{X}) = \mu$ and $\mathrm{E}(S^2) = \sigma^2$ in Section 1.7.3, $\overline{X}$ and S^2 are unbiased estimators of μ and σ^2. However, we have obtained the result $\mathrm{E}(S^{**2}) \neq \sigma^2$ in Section 1.7.3 and therefore S^{**2} is not an unbiased estimator of σ^2. Thus, according to the criterion of unbiasedness, S^2 is preferred to S^{**2} for estimation of σ^2.

Efficiency: Consider two estimators, i.e., $\hat{\theta}_n$ and $\widetilde{\theta}_n$. Both are assumed to be unbiased. That is, we have the following condition: $\mathrm{E}(\hat{\theta}_n) = \theta$ and $\mathrm{E}(\widetilde{\theta}_n) = \theta$. When $\mathrm{V}(\hat{\theta}_n) < \mathrm{V}(\widetilde{\theta}_n)$, we say that $\hat{\theta}_n$ is more efficient than $\widetilde{\theta}_n$. The estimator which is widely distributed is not preferred.

Consider as many unbiased estimators as possible. The unbiased estimator with the least variance is known as the **efficient estimator**. We have the case where an efficient estimator does not exist.

In order to obtain the efficient estimator, we utilize Cramer-Rao inequality. Suppose that X_i has the probability density function $f(x_i;\theta)$ for all i, i.e., X_1, X_2, $\cdots$, X_n are mutually independently and identically distributed. For any unbiased estimator of θ, denoted by $\hat{\theta}_n$, it is known that we have the following inequality:

$$\mathrm{V}(\hat{\theta}_n) \geq \frac{\sigma^2(\theta)}{n}, \tag{1.13}$$

where

$$\sigma^2(\theta) = \frac{1}{\mathrm{E}\Big(\big(\dfrac{\partial \log f(X;\theta)}{\partial\theta}\big)^2\Big)} = \frac{1}{\mathrm{V}\Big(\big(\dfrac{\partial \log f(X;\theta)}{\partial\theta}\big)\Big)} = -\frac{1}{\mathrm{E}\big(\dfrac{\partial^2 \log f(X;\theta)}{\partial\theta^2}\big)}, \tag{1.14}$$

which is known as the **Cramer-Rao inequality**. See Appendix 1.4 for proof of the Cramer-Rao inequality.

When there exists the unbiased estimator $\hat{\theta}_n$ such that the equality in (1.13) holds, $\hat{\theta}_n$ becomes the unbiased estimator with minimum variance, which is the efficient estimator. $\sigma^2(\theta)/n$ is called the **Cramer-Rao lower bound**.

Example 1.13 (Efficient Estimator): Suppose that X_1, X_2, $\cdots$, X_n are mutually independently, identically and normally distributed with mean μ and variance σ^2. Then, we show that $\overline{X}$ is an efficient estimator of μ.

When $\sigma^2 < \infty$, from Theorem on p.21, $\mathrm{V}(\overline{X})$ is given by σ^2/n in spite of the distribution of X_i, $i = 1, 2, \cdots, n$. ....(A)

On the other hand, because we assume that X_i is normally distributed with mean μ and variance σ^2, the probability density function of X_i is given by:

$$f(x;\mu) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}(x-\mu)^2\Big).$$

The Cramer-Rao inequality is represented as:

$$\mathrm{V}(\overline{X}) \geq \frac{1}{n\mathrm{E}\bigg(\Big(\dfrac{\partial \log f(X;\mu)}{\partial \mu}\Big)^2\bigg)},$$

where the logarithm of $f(X;\mu)$ is written as:

$$\log f(X;\mu) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}(X-\mu)^2.$$

Therefore, the partial derivative of $f(X;\mu)$ with respect to μ is:

$$\frac{\partial \log f(X;\mu)}{\partial \mu} = \frac{1}{\sigma^2}(X-\mu).$$

Accordingly, the Cramer-Rao inequality in this case is written as:

$$\mathrm{V}(\overline{X}) \geq \frac{1}{n\mathrm{E}\bigg(\Big(\dfrac{1}{\sigma^2}(X-\mu)\Big)^2\bigg)} = \frac{1}{n\dfrac{1}{\sigma^4}\mathrm{E}((X-\mu)^2)} = \frac{\sigma^2}{n}. \qquad \text{(B)}$$

From (A) and (B), variance of $\overline{X}$ is equal to the lower bound of Cramer-Rao inequality, i.e., $\mathrm{V}(\overline{X}) = \dfrac{\sigma^2}{n}$, which implies that the equality included in the Cramer-Rao inequality holds. Therefore, we can conclude that the sample mean $\overline{X}$ is an efficient estimator of μ.

Example 1.14 (Linear Unbiased Minimum Variance Estimator): Suppose that X_1, X_2, $\cdots$, X_n are mutually independently and identically distributed with mean μ and variance σ^2 (note that the normality assumption is excluded from Example 1.13). Consider the following linear estimator: $\hat{\mu} = \sum_{i=1}^n a_i X_i$. Then, we want to show $\hat{\mu}$ (i.e., $\overline{X}$) is a **linear unbiased minimum variance estimator** if $a_i = 1/n$ for all i, i.e., if $\hat{\mu} = \overline{X}$.

Utilizing Theorem on p.20, when $\mathrm{E}(X_i) = \mu$ and $\mathrm{V}(X_i) = \sigma^2$ for all i, we have: $\mathrm{E}(\hat{\mu}) = \mu \sum_{i=1}^n a_i$ and $\mathrm{V}(\hat{\mu}) = \sigma^2 \sum_{i=1}^n a_i^2$.

Since $\hat{\mu}$ is linear in X_i, $\hat{\mu}$ is called a **linear estimator** of μ. In order for $\hat{\mu}$ to be unbiased, we need to have the condition: $\mathrm{E}(\hat{\mu}) = \mu \sum_{i=1}^n a_i = \mu$. That is, if $\sum_{i=1}^n a_i = 1$ is satisfied, $\hat{\mu}$ gives us a **linear unbiased estimator**. Thus, as mentioned in Example 1.12 of Section 1.7.2, there are numerous unbiased estimators.

The variance of $\hat{\mu}$ is given by $\sigma^2 \sum_{i=1}^n a_i^2$. We obtain the value of a_i which minimizes $\sum_{i=1}^n a_i^2$ with the constraint $\sum_{i=1}^n a_i = 1$. Construct the Lagrange function as follows:

$$L = \frac{1}{2}\sum_{i=1}^n a_i^2 + \lambda(1 - \sum_{i=1}^n a_i),$$

where λ denotes the Lagrange multiplier. The $\frac{1}{2}$ in front of the first term appears to make life easier later on and does not affect the outcome. To determine the optimum values, we set the partial derivatives of L with respect to a_i and λ equal to zero, i.e.,

$$\frac{\partial L}{\partial a_i} = a_i - \lambda = 0, \qquad i = 1, 2, \cdots, n,$$
$$\frac{\partial L}{\partial \lambda} = 1 - \sum_{i=1}^n a_i = 0.$$

Solving the above equations, $a_i = \lambda = 1/n$ is obtained. Therefore, when $a_i = 1/n$ for all i, $\hat{\mu}$ has minimum variance in a class of linear unbiased estimators. That is, $\overline{X}$ is a **linear unbiased minimum variance estimator**.

The linear unbiased minimum variance estimator should be distinguished from the efficient estimator discussed in Example 1.13. The former does not requires the assumption on the underlying distribution. The latter gives us the unbiased estimator which variance is equal to the Cramer-Rao lower bound, which is not restricted to a class of the linear unbiased estimators. Under the assumption of normal population, the linear unbiased minimum variance estimator leads to the efficient estimator. However, both are different in general. In addition, note that the efficient estimator does not necessarily exist.

Consistency: Let $\hat{\theta}_n$ be an estimator of θ. Suppose that for any $\epsilon > 0$ we have the following:

$$P(|\hat{\theta}_n - \theta| > \epsilon) \longrightarrow 0, \qquad \text{as} \quad n \longrightarrow \infty,$$

which implies that $\hat{\theta} \longrightarrow \theta$ as $n \longrightarrow \infty$. Then, we say that $\hat{\theta}_n$ is a **consistent estimator** of θ. That is, the estimator which approaches the true parameter value as the sample size is large is called the consistent estimator of the parameter.

Example 1.15: Suppose that $X_1, X_2, \cdots, X_n$ are mutually independently and identically distributed with mean μ and variance σ^2. Assume that σ^2 is known. Then, it is shown that $\overline{X}$ is a consistent estimator of μ.

From (1.10), Chebyshev's inequality is given by:

$$P(|X - \mathrm{E}(X)| > \epsilon) \le \frac{\mathrm{V}(X)}{\epsilon^2},$$

for a random variable X. Here, replacing X by $\overline{X}$, we obtain $\mathrm{E}(\overline{X})$ and $\mathrm{V}(\overline{X})$ as follows:

$$\mathrm{E}(\overline{X}) = \mu, \quad \mathrm{V}(\overline{X}) = \frac{\sigma^2}{n},$$

because $\mathrm{E}(X_i) = \mu$ and $\mathrm{V}(X_i) = \sigma^2 < \infty$ are assumed for all i.

Then, when $n \longrightarrow \infty$, we obtain the following result:

$$P(|\overline{X} - \mu| > \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \longrightarrow 0,$$

which implies that $\overline{X} \longrightarrow \mu$ as $n \longrightarrow \infty$. Therefore, we can conclude that $\overline{X}$ is a consistent estimator of μ.

Summarizing the results up to now, $\overline{X}$ is an unbiased, minimum variance and consistent estimator of population mean μ. When the distribution of X_i is assumed to be normal for all i, $\overline{X}$ leads to an unbiased, efficient and consistent estimator of μ.

Example 1.16: Suppose that $X_1, X_2, \cdots, X_n$ are mutually independently, identically and normally distributed with mean μ and variance σ^2. Consider $S^{**2} = \frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2$, which is an estimate of σ^2.

In Remark on p.32, X and μ are replaced by S^{**2} and σ^2. Then, we obtain the following inequality:

$$P(|S^{**2} - \sigma^2| < \epsilon) \geq 1 - \frac{\mathrm{E}((S^{**2} - \sigma^2)^2)}{\epsilon^2}.$$

We compute $\mathrm{E}((S^{**2} - \sigma^2)^2)$. Since $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$, we obtain $\mathrm{E}((n-1)S^2/\sigma^2) = n-1$ and $\mathrm{V}((n-1)S^2/\sigma^2) = 2(n-1)$, where $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$. See Section 2.2.8 (p.146) for the chi-square distribution $\chi^2(n-1)$. Therefore, $\mathrm{E}(S^2) = \sigma^2$ and $\mathrm{V}(S^2) = 2\sigma^4/(n-1)$ can be derived. Using $S^{**2} = S^2(n-1)/n$, we have the following:

$$\begin{aligned}\mathrm{E}((S^{**2} - \sigma^2)^2) &= \mathrm{E}\Big(\Big(\frac{n-1}{n}S^2 - \sigma^2\Big)^2\Big) = \mathrm{E}\Big(\Big(\frac{n-1}{n}(S^2 - \sigma^2) - \frac{\sigma^2}{n}\Big)^2\Big) \\ &= \frac{(n-1)^2}{n^2}\mathrm{E}((S^2 - \sigma^2)^2) + \frac{\sigma^4}{n^2} = \frac{(n-1)^2}{n^2}\mathrm{V}(S^2) + \frac{\sigma^4}{n^2} = \frac{(2n-1)}{n^2}\sigma^4.\end{aligned}$$

Therefore, as $n \longrightarrow \infty$, we obtain:

$$P(|S^{**2} - \sigma^2| < \epsilon) \geq 1 - \frac{1}{\epsilon^2}\frac{(2n-1)}{n^2}\sigma^4 \longrightarrow 1.$$

Because $S^{**2} \longrightarrow \sigma^2$, S^{**2} is a consistent estimator of σ^2. Thus, S^{**2} is not unbiased (see Section 1.7.3, p.38), but is is consistent.

1.7.5 Maximum Likelihood Estimator

In Section 1.7.4, the properties of the estimators $\overline{X}$ and S^2 are discussed. It is shown that $\overline{X}$ is an unbiased, efficient and consistent estimator of μ under normality assumption and that S^2 is an unbiased estimator of σ^2. Note that S^2 is not efficient but consistent (we do not check these features of S^2 in this book).

The population parameter θ depends on a functional form of the population distribution $f(x;\theta)$. It corresponds to (μ, σ^2) in the case of the normal distribution and β in the case of the exponential distribution (Section 2.2.4). Now, in more general cases, we want to consider how to estimate θ. The maximum likelihood estimator gives us one of the solutions.

Let X_1, X_2, $\cdots$, X_n be mutually independently and identically distributed random samples. X_i has the probability density function $f(x;\theta)$. Under these assumptions, the joint density function of X_1, X_2, $\cdots$, X_n is given by:

$$f(x_1, x_2, \cdots, x_n; \theta) = \prod_{i=1}^{n} f(x_i; \theta),$$

where θ denotes the unknown parameter.

Given the actually observed data x_1, x_2, $\cdots$, x_n, the joint density $f(x_1, x_2, \cdots, x_n; \theta)$ is regarded as a function of θ, i.e.,

$$l(\theta) = l(\theta; x) = l(\theta; x_1, x_2, \cdots, x_n) = \prod_{i=1}^{n} f(x_i; \theta).$$

$l(\theta)$ is called the **likelihood function**.

Let $\hat{\theta}_n$ be the θ which maximizes the likelihood function. Replacing x_1, x_2, $\cdots$, x_n by X_1, X_2, $\cdots$, X_n, $\hat{\theta}_n = \hat{\theta}_n(X_1, X_2, \cdots, X_n)$ is called the **maximum likelihood estimator**, while $\hat{\theta}_n(x_1, x_2, \cdots, x_n)$ is called the **maximum likelihood estimate**.

That is, solving the following equation:

$$\frac{\partial l(\theta)}{\partial \theta} = 0,$$

the maximum likelihood estimator $\hat{\theta}_n \equiv \hat{\theta}_n(X_1, X_2, \cdots, X_n)$ is obtained.

Example 1.17: Suppose that X_1, X_2, $\cdots$, X_n are mutually independently, identically and normally distributed with mean μ and variance σ^2. We derive the maximum likelihood estimators of μ and σ^2. The joint density (or the likelihood function) of X_1, X_2, $\cdots$, X_n is written as:

$$\begin{aligned} f(x_1, x_2, \cdots, x_n; \mu, \sigma^2) &= \prod_{i=1}^{n} f(x_i; \mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}(x_i - \mu)^2\Big) \\ &= (2\pi\sigma^2)^{-n/2} \exp\Big(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2\Big) = l(\mu, \sigma^2). \end{aligned}$$

The logarithm of the likelihood function is given by:

$$\log l(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2,$$

which is called the **log-likelihood function**. For maximization of the likelihood function, differentiating the log-likelihood function $\log l(\mu, \sigma^2)$ with respect to μ and σ^2, the first derivatives should be equal to zero, i.e.,

$$\frac{\partial \log l(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i - \mu) = 0,$$

$$\frac{\partial \log l(\mu, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu)^2 = 0.$$

Let $\hat{\mu}$ and $\hat{\sigma}^2$ be the solution which satisfies the above two equations. Solving the two equations, we obtain the maximum likelihood estimates as follows:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x},$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})^2 = s^{**2}.$$

Replacing x_i by X_i for $i = 1, 2, \cdots, n$, the maximum likelihood estimators of μ and σ^2 are given by $\overline{X}$ and S^{**2}, respectively. Since $\mathrm{E}(\overline{X}) = \mu$, the maximum likelihood estimator of μ, $\overline{X}$, is an unbiased estimator. However, because of $\mathrm{E}(S^{**2}) = \dfrac{n-1}{n}\sigma^2 \neq \sigma^2$ as shown in Section 1.7.3, the maximum likelihood estimator of σ^2, S^{**2}, is not an unbiased estimator.

Properties of Maximum Likelihood Estimator: For small sample, the maximum likelihood estimator has the following properties.

- The maximum likelihood estimator is not necessarily unbiased in general, but we often have the case where we can construct the unbiased estimator by an appropriate transformation.

  For instance, in Example 1.17, we find that the maximum likelihood estimator of σ^2, S^{**2}, is not unbiased. However, $\dfrac{n}{n-1}S^{**2}$ is an unbiased estimator of σ^2.

- If the efficient estimator exists, i.e., if there exists the estimator which satisfies the equality in the Cramer-Rao inequality, the maximum likelihood estimator is efficient.

For large sample, as $n \longrightarrow \infty$, the maximum likelihood estimator of θ, $\hat{\theta}_n$, has the following property:

$$\sqrt{n}(\hat{\theta}_n - \theta) \longrightarrow N(0, \sigma^2(\theta)), \tag{1.15}$$

where

$$\sigma^2(\theta) = \frac{1}{\mathrm{E}\Big(\big(\frac{\partial \log f(X;\theta)}{\partial\theta}\big)^2\Big)}.$$

(1.15) indicates that the maximum likelihood estimator has consistency, asymptotic unbiasedness, asymptotic efficiency and asymptotic normality. Asymptotic normality of the maximum likelihood estimator comes from the central limit theorem discussed in Section 1.6.3. Even though the underlying distribution is not normal, i.e., even though $f(x;\theta)$ is not normal, the maximum likelihood estimator is asymptotically normally distributed. Note that the properties of $n \longrightarrow \infty$ are called the asymptotic properties, which include consistency, asymptotic normality and so on.

By normalizing, as $n \longrightarrow \infty$, we obtain as follows:

$$\frac{\sqrt{n}(\hat{\theta}_n - \theta)}{\sigma(\theta)} = \frac{\hat{\theta}_n - \theta}{\sigma(\theta)/\sqrt{n}} \longrightarrow N(0,1).$$

As another representation, when n is large, we can approximate the distribution of $\hat{\theta}_n$ as follows:

$$\hat{\theta}_n \sim N\Big(\theta, \frac{\sigma^2(\theta)}{n}\Big).$$

This implies that when $n \longrightarrow \infty$, $\hat{\theta}_n$ approaches the lower bound of Cramer-Rao inequality: $\sigma^2(\theta)/n$, which property is called an asymptotic efficiency.

Moreover, replacing θ in variance $\sigma^2(\theta)$ by $\hat{\theta}_n$, when $n \longrightarrow \infty$, we have the following property:

$$\frac{\hat{\theta}_n - \theta}{\sigma(\hat{\theta}_n)/\sqrt{n}} \longrightarrow N(0,1), \tag{1.16}$$

which also comes from the central limit theorem.

Practically, when n is large, we approximately use as follows:

$$\hat{\theta}_n \sim N\Big(\theta, \frac{\sigma^2(\hat{\theta}_n)}{n}\Big). \tag{1.17}$$

Proof of (1.15): By the central limit theorem (1.11) on p.34,

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial\theta} \longrightarrow N\Big(0, \frac{1}{\sigma^2(\theta)}\Big), \tag{1.18}$$

where $\sigma^2(\theta)$ is defined in (1.14), i.e., $\mathrm{V}(\partial \log f(X_i;\theta)/\partial\theta) = 1/\sigma^2(\theta)$. As shown in (1.46) of Appendix 1.4, note that $\mathrm{E}(\partial \log f(X_i;\theta)/\partial\theta) = 0$. We can apply the central limit theorem, taking $\partial \log f(X_i;\theta)/\partial\theta$ as the ith random variable.

By performing the first order Taylor series expansion around $\hat{\theta}_n = \theta$, we have the following approximation:

$$\begin{aligned} 0 &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \log f(X_i; \hat{\theta}_n)}{\partial \theta} \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \log f(X_i; \theta)}{\partial \theta} + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial^2 \log f(X_i; \theta)}{\partial \theta^2} (\hat{\theta}_n - \theta) + \cdots . \end{aligned}$$

Therefore, the following approximation also holds:

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \log f(X_i; \theta)}{\partial \theta} \approx -\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial^2 \log f(X_i; \theta)}{\partial \theta^2} (\hat{\theta}_n - \theta).$$

From (1.18) and the above equation, we obtain:

$$-\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log f(X_i; \theta)}{\partial \theta^2} \sqrt{n}(\hat{\theta}_n - \theta) \longrightarrow N\Big(0, \frac{1}{\sigma^2(\theta)}\Big).$$

The law of large numbers indicates as follows:

$$-\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log f(X_i; \theta)}{\partial \theta^2} \longrightarrow -\mathrm{E}\Big(\frac{\partial^2 \log f(X_i; \theta)}{\partial \theta^2}\Big) = \frac{1}{\sigma^2(\theta)},$$

where the last equality is from (1.14). Thus, we have the following relation:

$$-\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log f(X_i; \theta)}{\partial \theta^2} \sqrt{n}(\hat{\theta}_n - \theta) \longrightarrow \frac{1}{\sigma^2(\theta)} \sqrt{n}(\hat{\theta}_n - \theta) \longrightarrow N\Big(0, \frac{1}{\sigma^2(\theta)}\Big)$$

Therefore, the asymptotic normality of the maximum likelihood estimator is obtained as follows:

$$\sqrt{n}(\hat{\theta}_n - \theta) \longrightarrow N(0, \sigma^2(\theta)).$$

Thus, (1.15) is obtained.

1.7.6 Interval Estimation

In Sections 1.7.1 – 1.7.5, the point estimation is discussed. It is important to know where the true parameter value of θ is likely to lie.

Suppose that the population distribution is given by $f(x; \theta)$. Using the random sample X_1, X_2, $\cdots$, X_n drawn from the population distribution, we construct the two statistics, say, $\hat{\theta}_U(X_1, X_2, \cdots, X_n; \theta^*)$ and $\hat{\theta}_L(X_1, X_2, \cdots, X_n; \theta^{**})$, where θ^* and θ^{**} denote the constant values which satisfy:

$$P(\theta^* < \hat{\theta}_n < \theta^{**}) = 1 - \alpha, \tag{1.19}$$

for $\theta^{**} > \theta^*$. Note that $\hat{\theta}_n$ depends on X_1, X_2, $\cdots$, X_n as well as θ, i.e., $\hat{\theta}_n \equiv \hat{\theta}_n(X_1, X_2, \cdots, X_n; \theta)$. Now we assume that we can solve (1.19) with respect to θ, which is rewritten as follows:

$$P\Big(\hat{\theta}_L(X_1, X_2, \cdots, X_n; \theta^*) < \theta < \hat{\theta}_U(X_1, X_2, \cdots, X_n; \theta^{**})\Big) = 1 - \alpha. \qquad (1.20)$$

(1.20) implies that θ lies on the interval $\Big(\hat{\theta}_L(X_1, X_2, \cdots, X_n; \theta^*), \hat{\theta}_U(X_1, X_2, \cdots, X_n; \theta^{**})\Big)$ with probability $1 - \alpha$. Depending on a functional form of $\hat{\theta}_n(X_1, X_2, \cdots, X_n; \theta)$, we possibly have the situation that θ^* and θ^{**} are switched with each other.

Now, we replace the random variables X_1, X_2, $\cdots$, X_n by the experimental values x_1, x_2, $\cdots$, x_n. Then, we say that the interval:

$$\Big(\hat{\theta}_L(x_1, x_2, \cdots, x_n; \theta^*),\ \hat{\theta}_U(x_1, x_2, \cdots, x_n; \theta^{**})\Big)$$

is called the $100 \times (1 - \alpha)\%$ **confidence interval** of θ. Thus, estimating the interval is known as the **interval estimation**, which is distinguished from the point estimation. In the interval, $\hat{\theta}_L(x_1, x_2, \cdots, x_n; \theta^*)$ is known as the **lower bound** of the confidence interval, while $\hat{\theta}_U(x_1, x_2, \cdots, x_n; \theta^{**})$ is the **upper bound** of the confidence interval.

Given probability α, the $\hat{\theta}_L(X_1, X_2, \cdots, X_n; \theta^*)$ and $\hat{\theta}_U(X_1, X_2, \cdots, X_n; \theta^{**})$ which satisfies equation (1.20) are not unique. For estimation of the unknown parameter θ, it is more optimal to minimize the width of the confidence interval. Therefore, we should choose θ^* and θ^{**} which minimizes the width $\hat{\theta}_U(X_1, X_2, \cdots, X_n; \theta^{**}) - \hat{\theta}_L(X_1, X_2, \cdots, X_n; \theta^*)$.

Interval Estimation of $\overline{X}$: Let X_1, X_2, $\cdots$, X_n be mutually independently and identically distributed random variables. X_i has a distribution with mean μ and variance σ^2. From the central limit theorem,

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \longrightarrow N(0, 1).$$

Replacing σ^2 by its estimator S^2 (or S^{**2}),

$$\frac{\overline{X} - \mu}{S / \sqrt{n}} \longrightarrow N(0, 1).$$

Therefore, when n is large enough,

$$P(z^* < \frac{\overline{X} - \mu}{S / \sqrt{n}} < z^{**}) = 1 - \alpha,$$

where z^* and z^{**} ($z^* < z^{**}$) are percent points from the standard normal density function. Solving the inequality above with respect to μ, the following expression is obtained.

$$P\Big(\overline{X} - z^{**}\frac{S}{\sqrt{n}} < \mu < \overline{X} - z^*\frac{S}{\sqrt{n}}\Big) = 1 - \alpha,$$

where $\hat{\theta}_L$ and $\hat{\theta}_U$ correspond to $\overline{X} - z^{**}S/\sqrt{n}$ and $\overline{X} - z^{*}S/\sqrt{n}$, respectively.

The length of the confidence interval is given by:

$$\hat{\theta}_U - \hat{\theta}_L = \frac{S}{\sqrt{n}}(z^{**} - z^{*}),$$

which should be minimized subject to:

$$\int_{z^*}^{z^{**}} f(x)\,\mathrm{d}x = 1 - \alpha,$$

i.e.,

$$F(z^{**}) - F(z^{*}) = 1 - \alpha,$$

where $F(\cdot)$ denotes the standard normal cumulative distribution function.

Solving the minimization problem above, we can obtain the conditions that $f(z^*) = f(z^{**})$ for $z^* < z^{**}$ and that $f(x)$ is symmetric. Therefore, we have:

$$-z^* = z^{**} = z_{\alpha/2},$$

where $z_{\alpha/2}$ denotes the $100 \times \alpha/2$ percent point from the standard normal density function.

Accordingly, replacing the estimators $\overline{X}$ and S^2 by their estimates $\overline{x}$ and s^2, the $100 \times (1-\alpha)\%$ confidence interval of μ is approximately represented as:

$$\Big(\overline{x} - z_{\alpha/2}\frac{s}{\sqrt{n}},\ \overline{x} + z_{\alpha/2}\frac{s}{\sqrt{n}}\Big),$$

for large n.

For now, we do not impose any assumptions on the distribution of X_i. If we assume that X_i is normal, $\sqrt{n}(\overline{X} - \mu)/S$ has a t distribution with $n-1$ degrees of freedom for any n. Therefore, $100 \times (1-\alpha)\%$ confidence interval of μ is given by:

$$\Big(\overline{x} - t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}},\ \overline{x} + t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}\Big),$$

where $t_{\alpha/2}(n-1)$ denotes the $100 \times \alpha/2$ percent point of the t distribution with $n-1$ degrees of freedom. See Section 2.2.10, p.155 for the t distribution.

Interval Estimation of $\hat{\theta}_n$: Let $X_1, X_2, \cdots, X_n$ be mutually independently and identically distributed random variables. X_i has the probability density function $f(x_i;\theta)$. Suppose that $\hat{\theta}_n$ represents the maximum likelihood estimator of θ.

From (1.17), we can approximate the $100 \times (1-\alpha)\%$ confidence interval of θ as follows:

$$\Big(\hat{\theta}_n - z_{\alpha/2}\frac{\sigma(\hat{\theta}_n)}{\sqrt{n}},\ \hat{\theta}_n + z_{\alpha/2}\frac{\sigma(\hat{\theta}_n)}{\sqrt{n}}\Big).$$

Table 1.1: Type I and Type II Errors

	H_0 is true.	H_0 is false.
Acceptance of H_0	Correct judgment	Type II Error (Probability β)
Rejection of H_0	Type I Error (Probability α = Significance Level)	Correct judgment ($1-\beta$ = Power)

1.8 Testing Hypothesis

1.8.1 Basic Concepts in Testing Hypothesis

Given the population distribution $f(x;\theta)$, we want to judge from the observed values $x_1, x_2, \cdots, x_n$ whether the hypothesis on the parameter θ, e.g. $\theta = \theta_0$, is correct or not. The hypothesis that we want to test is called the **null hypothesis**, which is denoted by $H_0 : \theta = \theta_0$. The hypothesis against the null hypothesis, e.g. $\theta \neq \theta_0$, is called the **alternative hypothesis**, which is denoted by $H_1 : \theta \neq \theta_0$.

Type I and Type II Errors: When we test the null hypothesis H_0, as shown in Table 1.1 we have four cases, i.e., (i) we accept H_0 when H_0 is true, (ii) we reject H_0 when H_0 is true, (iii) we accept H_0 when H_0 is false, and (iv) we reject H_0 when H_0 is false. (i) and (iv) are correct judgments, while (ii) and (iii) are not correct. (ii) is called a **type I error** and (iii) is called a **type II error**. The probability which a type I error occurs is called the **significance level**, which is denoted by α, and the probability of committing a type II error is denoted by β. Probability of (iv) is called the **power** or the **power function**, because it is a function of the parameter θ.

Testing Procedures: The testing procedure is summarized as follows.

1. Construct the null hypothesis (H_0) on the parameter.
2. Consider an appropriate statistic, which is called a **test statistic**. Derive a distribution function of the test statistic when H_0 is true.
3. From the observed data, compute the observed value of the test statistic.
4. Compare the distribution and the observed value of the test statistic. When the observed value of the test statistic is in the tails of the distribution, we consider that H_0 is not likely to occur and we reject H_0.

The region that H_0 is unlikely to occur and accordingly H_0 is rejected is called the **rejection region** or the **critical region**, denoted by R. Conversely, the region that

H_0 is likely to occur and accordingly H_0 is accepted is called the **acceptance region**, denoted by A.

Using the rejection region R and the acceptance region A, the type I and II errors and the power are formulated as follows. Suppose that the test statistic is give by $T = T(X_1, X_2, \cdots, X_n)$. The probability of committing a type I error, i.e., the significance level α, is given by:

$$P(T(X_1, X_2, \cdots, X_n) \in R | H_0 \text{ is true}) = \alpha,$$

which is the probability that rejects H_0 when H_0 is true. Conventionally, the significance level $\alpha = 0.1, 0.05, 0.01$ is chosen in practice. The probability of committing a type II error, i.e., β, is represented as:

$$P(T(X_1, X_2, \cdots, X_n) \in A | H_0 \text{ is not true}) = \beta,$$

which corresponds to the probability that accepts H_0 when H_0 is not true. The power is defined as $1 - \beta$, i.e.,

$$P(T(X_1, X_2, \cdots, X_n) \in R | H_0 \text{ is not true}) = 1 - \beta,$$

which is the probability that rejects H_0 when H_0 is not true.

1.8.2 Power Function

Let X_1, X_2, $\cdots$, X_n be mutually independently, identically and normally distributed with mean μ and variance σ^2. Assume that σ^2 is known.

In Figure 1.3, we consider the hypothesis on the population mean μ, i.e., the null hypothesis $H_0 : \mu = \mu_0$ against the alternative hypothesis $H_1 : \mu = \mu_1$, where $\mu_1 > \mu_0$ is taken. The dark shadow area corresponds to the probability of committing a type I error, i.e., the significance level, while the light shadow area indicates the probability of committing a type II error. The probability of the right-hand side of f^* in the distribution under H_1 represents the power of the test, i.e., $1 - \beta$.

In the case of normal population, the distribution of sample mean $\overline{X}$ is given by:

$$\overline{X} \sim N\Big(\mu, \frac{\sigma^2}{n}\Big).$$

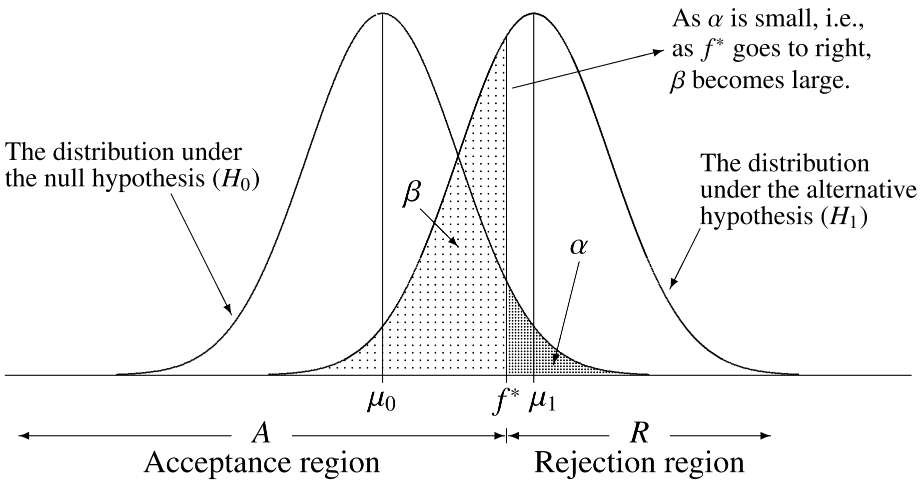
For the distribution of $\overline{X}$, see the moment-generating function of $\overline{X}$ in Theorem on p.29. By normalization, we have:

$$\frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1).$$

Therefore, under the null hypothesis $H_0 : \mu = \mu_0$, we obtain:

$$\frac{\overline{X} - \mu_0}{\sigma / \sqrt{n}} \sim N(0, 1),$$

Figure 1.3: Type I Error (α) and Type II Error (β)



where μ is replaced by μ_0. Since the significance level α is the probability which rejects H_0 when H_0 is true, it is given by:

$$\alpha = P\Big(\overline{X} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}\Big),$$

where z_α denotes $100\times\alpha$ percent point of the standard normal density function. Therefore, the rejection region is given by: $\overline{X} > \mu_0 + z_\alpha \sigma/\sqrt{n}$.

Since the power $1 - \beta$ is the probability which rejects H_0 when H_1 is true, it is given by:

$$\begin{aligned}
1 - \beta &= P\Big(\overline{X} > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}\Big) = P\Big(\frac{\overline{X} - \mu_1}{\sigma/\sqrt{n}} > \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z_\alpha\Big) \\
&= 1 - P\Big(\frac{\overline{X} - \mu_1}{\sigma/\sqrt{n}} < \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z_\alpha\Big) = 1 - F\Big(\frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z_\alpha\Big),
\end{aligned}$$

where $F(\cdot)$ represents the standard normal cumulative distribution function, which is given by $F(x) = \int_{-\infty}^{x} (2\pi)^{-1/2} \exp(-\frac{1}{2}t^2)\,\mathrm{d}t$. The power function is a function of μ_1, given μ_0 and α.

1.8.3 Testing Hypothesis on Population Mean

Let $X_1, X_2, \cdots, X_n$ be mutually independently, identically and normally distributed with mean μ and variance σ^2. Assume that σ^2 is known.

Consider testing the null hypothesis $H_0: \mu = \mu_0$. When the null hypothesis H_0 is true, the distribution of $\overline{X}$ is given by:

$$\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Therefore, the test statistic is given by: $\sqrt{n}(\overline{X}-\mu_0)/\sigma$, while the test statistic value is: $\sqrt{n}(\overline{x}-\mu_0)/\sigma$, where the sample mean $\overline{X}$ is replaced by the observed value $\overline{x}$.

Depending on the alternative hypothesis, we have the following three cases.

1. **The alternative hypothesis $H_1 : \mu < \mu_0$** (one-sided test):

   We have: $P\Big(\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} < -z_\alpha\Big) = \alpha$. Therefore, when $\dfrac{\overline{x}-\mu_0}{\sigma/\sqrt{n}} < -z_\alpha$, we reject the null hypothesis $H_0 : \mu = \mu_0$ at the significance level α.

2. **The alternative hypothesis $H_1 : \mu > \mu_0$** (one-sided test):

   We have: $P\Big(\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha\Big) = \alpha$. Therefore, when $\dfrac{\overline{x}-\mu_0}{\sigma/\sqrt{n}} > z_\alpha$, we reject the null hypothesis $H_0 : \mu = \mu_0$ at the significance level α.

3. **The alternative hypothesis $H_1 : \mu \neq \mu_0$** (two-sided test):

   We have: $P\Big(\Big|\dfrac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}\Big| > z_{\alpha/2}\Big) = \alpha$. Therefore, when $\Big|\dfrac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}\Big| > z_{\alpha/2}$, we reject the null hypothesis $H_0 : \mu = \mu_0$ at the significance level α.

When the sample size n is large enough, the testing procedure above can be applied to the cases: (i) the distribution of X_i is not known and (ii) σ^2 is replaced by its estimator S^2 (in the case where σ^2 is not known).

1.8.4 Wald Test

From (1.16), under the null hypothesis $H_0 : \theta = \theta_0$, as $n \longrightarrow \infty$, the maximum likelihood estimator $\hat{\theta}_n$ is distributed as follows:

$$\frac{\hat{\theta}_n - \theta_0}{\sigma(\hat{\theta}_n)/\sqrt{n}} \sim N(0,1).$$

For $H_0 : \theta = \theta_0$ and $H_1 : \theta \neq \theta_0$, replacing $X_1, X_2, \cdots, X_n$ in $\hat{\theta}_n$ by the observed values $x_1, x_2, \cdots, x_n$, we obtain the following testing procedure:

1. If we have:

   $$\left|\frac{\hat{\theta}_n - \theta_0}{\sigma(\hat{\theta}_n)/\sqrt{n}}\right| > z_{\alpha/2},$$

   we reject the null hypothesis H_0 at the significance level α, because the probability which H_0 occurs is small enough.

2. As for $H_0 : \theta = \theta_0$ and $H_1 : \theta > \theta_0$, if we have:

   $$\frac{\hat{\theta}_n - \theta_0}{\sigma(\hat{\theta}_n)/\sqrt{n}} > z_\alpha,$$

   we reject H_0 at the significance level α.

3. For $H_0: \theta = \theta_0$ and $H_1: \theta < \theta_0$, when we have the following:

$$\frac{\hat{\theta}_n - \theta_0}{\sigma(\hat{\theta}_n)/\sqrt{n}} < -z_\alpha,$$

we reject H_0 at the significance level α.

The testing procedure introduced here is called the **Wald test**.

Example 1.18: $X_1, X_2, \cdots, X_n$ are mutually independently, identically and exponentially distributed. Consider the following exponential probability density function:

$$f(x;\gamma) = \gamma e^{-\gamma x},$$

for $0 < x < \infty$.

Using the Wald test, we want to test the null hypothesis $H_0: \gamma = \gamma_0$ against the alternative hypothesis $H_1: \gamma \neq \gamma_0$.

Generally, as $n \longrightarrow \infty$, the distribution of the maximum likelihood estimator of the parameter γ, $\hat{\gamma}_n$, is asymptotically represented as:

$$\frac{\hat{\gamma}_n - \gamma}{\sigma(\hat{\gamma}_n)/\sqrt{n}} \sim N(0,1),$$

where

$$\sigma^2(\gamma) = \left(\mathrm{E}\left(\left(\frac{\mathrm{d}\log f(X;\gamma)}{\mathrm{d}\gamma}\right)^2\right)\right)^{-1} = -\left(\mathrm{E}\left(\frac{\mathrm{d}^2\log f(X;\gamma)}{\mathrm{d}\gamma^2}\right)\right)^{-1}.$$

See (1.14) and (1.16) for the above properties on the maximum likelihood estimator.

Therefore, under the null hypothesis $H_0: \gamma = \gamma_0$, when n is large enough, we have the following distribution:

$$\frac{\hat{\gamma}_n - \gamma_0}{\sigma(\hat{\gamma}_n)/\sqrt{n}} \sim N(0,1).$$

As for the null hypothesis $H_0: \gamma = \gamma_0$ against the alternative hypothesis $H_1: \gamma \neq \gamma_0$, if we have:

$$\left|\frac{\hat{\gamma}_n - \gamma_0}{\sigma(\hat{\gamma}_n)/\sqrt{n}}\right| > z_{\alpha/2},$$

we can reject H_0 at the significance level α.

We need to derive $\sigma^2(\gamma)$ and $\hat{\gamma}_n$ to perform the testing procedure. First, $\sigma^2(\gamma)$ is given by:

$$\sigma^2(\gamma) = -\left(\mathrm{E}\left(\frac{\mathrm{d}^2\log f(X;\gamma)}{\mathrm{d}\gamma^2}\right)\right)^{-1} = \gamma^2.$$

Note that the first- and the second-derivatives of $\log f(X;\gamma)$ with respect to γ are given by:

$$\frac{\mathrm{d}\log f(X;\gamma)}{\mathrm{d}\gamma} = \frac{1}{\gamma} - X, \qquad \frac{\mathrm{d}^2\log f(X;\gamma)}{\mathrm{d}\gamma^2} = -\frac{1}{\gamma^2}.$$

Next, the maximum likelihood estimator of γ, i.e., $\hat{\gamma}_n$, is obtained as follows. Since X_1, X_2 $\cdots$, X_n are mutually independently and identically distributed, the likelihood function $l(\gamma)$ is given by:

$$l(\gamma) = \prod_{i=1}^{n} f(x_i;\gamma) = \prod_{i=1}^{n} \gamma e^{-\gamma x_i} = \gamma^n e^{-\gamma \sum x_i}.$$

Therefore, the log-likelihood function is written as:

$$\log l(\gamma) = n\log(\gamma) - \gamma\sum_{i=1}^{n} x_i.$$

We obtain the value of γ which maximizes $\log l(\gamma)$. Solving the following equation:

$$\frac{\mathrm{d}\log l(\gamma)}{\mathrm{d}\gamma} = \frac{n}{\gamma} - \sum_{i=1}^{n} x_i = 0,$$

the maximum likelihood estimator of γ, i.e., $\hat{\gamma}_n$ is represented as:

$$\hat{\gamma}_n = \frac{n}{\sum_{i=1}^{n} X_i} = \frac{1}{\overline{X}}.$$

Then, we have the following:

$$\frac{\hat{\gamma}_n - \gamma}{\sigma(\hat{\gamma}_n)/\sqrt{n}} = \frac{\hat{\gamma}_n - \gamma}{\hat{\gamma}_n/\sqrt{n}} \longrightarrow N(0,1),$$

where $\hat{\gamma}_n$ is given by $1/\overline{X}$.

For $H_0:\ \gamma = \gamma_0$ and $H_1:\ \gamma \neq \gamma_0$, if we have:

$$\left|\frac{\hat{\gamma}_n - \gamma_0}{\hat{\gamma}_n/\sqrt{n}}\right| > z_{\alpha/2},$$

we reject H_0 at the significance level α.

1.8.5 Likelihood Ratio Test

Suppose that the population distribution is given by $f(x;\theta)$, where $\theta = (\theta_1, \theta_2)$. Consider testing the null hypothesis $\theta_1 = \theta_1^*$ against the alternative hypothesis $H_1 : \theta_1 \neq \theta_1^*$, using the observed values $(x_1, x_2, \cdots, x_n)$ corresponding to the random sample $(X_1, X_2, \cdots, X_n)$.

Let θ_1 and θ_2 be $1 \times k_1$ and $1 \times k_2$ vectors, respectively. Therefore, $\theta = (\theta_1, \theta_2)$ denotes a $1 \times (k_1 + k_2)$ vector. Since we take the null hypothesis as $H_0: \theta_1 = \theta_1^*$, the number of restrictions is given by k_1, which is equal to the dimension of θ_1.

The likelihood function is written as:

$$l(\theta_1, \theta_2) = \prod_{i=1}^{n} f(x_i; \theta_1, \theta_2).$$

Let $(\widetilde{\theta}_1, \widetilde{\theta}_2)$ be the maximum likelihood estimator of (θ_1, θ_2). That is, $(\widetilde{\theta}_1, \widetilde{\theta}_2)$ indicates the solution of (θ_1, θ_2), obtained from the following equations:

$$\frac{\partial l(\theta_1, \theta_2)}{\partial \theta_1} = 0, \qquad \frac{\partial l(\theta_1, \theta_2)}{\partial \theta_2} = 0.$$

The solution $(\widetilde{\theta}_1, \widetilde{\theta}_2)$ is called the **unconstrained maximum likelihood estimator**, because the null hypothesis $H_0 : \theta_1 = \theta_1^*$ is not taken into account.

Let $\hat{\theta}_2$ be the maximum likelihood estimator of θ_2 under the null hypothesis $H_0 : \theta_1 = \theta_1^*$. That is, $\hat{\theta}_2$ is a solution of the following equation:

$$\frac{\partial l(\theta_1^*, \theta_2)}{\partial \theta_2} = 0.$$

The solution $\hat{\theta}_2$ is called the **constrained maximum likelihood estimator** of θ_2, because the likelihood function is maximized with respect to θ_2 subject to the constraint $\theta_1 = \theta_1^*$.

Define λ as follows:

$$\lambda = \frac{l(\theta_1^*, \hat{\theta}_2)}{l(\widetilde{\theta}_1, \widetilde{\theta}_2)},$$

which is called the **likelihood ratio**.

As n goes to infinity, it is known that we have:

$$-2\log(\lambda) \sim \chi^2(k_1),$$

where k_1 denotes the number of the constraints.

Let $\chi^2_\alpha(k_1)$ be the $100 \times \alpha$ percent point from the chi-square distribution with k_1 degrees of freedom. When $-2\log(\lambda) > \chi^2_\alpha(k_1)$, we reject the null hypothesis $H_0 : \theta_1 = \theta_1^*$ at the significance level α. If $-2\log(\lambda)$ is close to zero, we accept the null hypothesis. When $(\theta_1^*, \hat{\theta}_2)$ is close to $(\widetilde{\theta}_1, \widetilde{\theta}_2)$, $-2\log(\lambda)$ approaches zero.

The likelihood ratio test is useful in the case where it is not easy to derive the distribution of $(\widetilde{\theta}_1, \widetilde{\theta}_2)$.

Example 1.19: $X_1, X_2, \cdots, X_n$ are mutually independently, identically and exponentially distributed. Consider the following exponential probability density function:

$$f(x;\gamma) = \gamma e^{-\gamma x},$$

for $0 < x < \infty$.

Using the likelihood ratio test, we want to test the null hypothesis $H_0 : \ \gamma = \gamma_0$ against the alternative hypothesis $H_1 : \ \gamma \neq \gamma_0$. Remember that in Example 1.18 we test the hypothesis with the Wald test.

In this case, the likelihood ratio is given by:

$$\lambda = \frac{l(\gamma_0)}{l(\hat{\gamma}_n)},$$

where $\hat{\gamma}_n$ is derived in Example 1.18, i.e.,

$$\hat{\gamma}_n = \frac{n}{\sum_{i=1}^n X_i} = \frac{1}{\overline{X}}.$$

Since the number of the constraint is equal to one, as the sample size n goes to infinity we have the following asymptotic distribution:

$$-2\log\lambda \longrightarrow \chi^2(1).$$

The likelihood ratio is computed as follows:

$$\lambda = \frac{l(\gamma_0)}{l(\hat{\gamma}_n)} = \frac{\gamma_0^n e^{-\gamma_0 \sum X_i}}{\hat{\gamma}_n^n e^{-n}}.$$

If $-2\log\lambda > \chi^2_\alpha(1)$, we reject the null hypothesis $H_0 : \ \mu = \mu_0$ at the significance level α. Note that $\chi^2_\alpha(1)$ denotes the $100 \times \alpha$ percent point from the chi-square distribution with one degree of freedom.

Example 1.20: Suppose that $X_1, X_2, \cdots, X_n$ are mutually independently, identically and normally distributed with mean zero and variance σ^2.

The normal probability density function with mean μ and variance σ^2 is given by:

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

By the likelihood ratio test, we want to test the null hypothesis $H_0 : \ \mu = \mu_0$ against the alternative hypothesis $H_1 : \mu \neq \mu_0$.

The likelihood ratio is given by:

$$\lambda = \frac{l(\mu_0, \widetilde{\sigma}^2)}{l(\hat{\mu}, \hat{\sigma}^2)},$$

where $\widetilde{\sigma}^2$ is the constrained maximum likelihood estimator with the constraint $\mu = \mu_0$, while $(\hat{\mu}, \hat{\sigma}^2)$ denotes the unconstrained maximum likelihood estimator. In this case, since the number of the constraint is one, the asymptotic distribution is as follows:

$$-2\log\lambda \longrightarrow \chi^2(1).$$

Now, we derive $l(\mu_0, \widetilde{\sigma}^2)$ and $l(\hat{\mu}, \hat{\sigma}^2)$. $l(\mu, \sigma^2)$ is written as:

$$\begin{aligned}
l(\mu, \sigma^2) = f(x_1, x_2, \cdots, x_n; \mu, \sigma^2) &= \prod_{i=1}^{n} f(x_i; \mu, \sigma^2) \\
&= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{1}{2\sigma^2}(x_i - \mu)^2\Big) \\
&= (2\pi\sigma^2)^{-n/2} \exp\Big(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i - \mu)^2\Big).
\end{aligned}$$

The log-likelihood function $\log l(\mu, \sigma^2)$ is represented as:

$$\log l(\mu, \sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2.$$

For the numerator of the likelihood ratio, under the constraint $\mu = \mu_0$, maximize $\log l(\mu_0, \sigma^2)$ with respect to σ^2. Since we obtain the first-derivative:

$$\frac{\partial \log l(\mu_0, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i - \mu_0)^2 = 0,$$

the constrained maximum likelihood estimator $\widetilde{\sigma}^2$ is given by:

$$\widetilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_0)^2.$$

Therefore, replacing σ^2 by $\widetilde{\sigma}^2$, $l(\mu_0, \widetilde{\sigma}^2)$ is written as:

$$l(\mu_0, \widetilde{\sigma}^2) = (2\pi\widetilde{\sigma}^2)^{-n/2} \exp\Big(-\frac{1}{2\widetilde{\sigma}^2}\sum_{i=1}^{n}(x_i - \mu_0)^2\Big) = (2\pi\widetilde{\sigma}^2)^{-n/2}\exp\Big(-\frac{n}{2}\Big).$$

For the denominator of the likelihood ratio, because the unconstrained maximum likelihood estimators are obtained as:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2,$$

$l(\hat{\mu}, \hat{\sigma}^2)$ is written as:

$$l(\hat{\mu}, \hat{\sigma}^2) = (2\pi\hat{\sigma}^2)^{-n/2} \exp\Big(-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}(x_i - \hat{\mu})^2\Big) = (2\pi\hat{\sigma}^2)^{-n/2}\exp\Big(-\frac{n}{2}\Big).$$

Thus, the likelihood ratio is given by:

$$\lambda = \frac{l(\mu_0, \widetilde{\sigma}^2)}{l(\hat{\mu}, \hat{\sigma}^2)} = \frac{(2\pi\widetilde{\sigma}^2)^{-n/2} \exp\Big(-\dfrac{n}{2}\Big)}{(2\pi\hat{\sigma}^2)^{-n/2} \exp\Big(-\dfrac{n}{2}\Big)} = \Big(\frac{\widetilde{\sigma}^2}{\hat{\sigma}^2}\Big)^{-n/2}.$$

Asymptotically, we have:

$$-2\log\lambda = n(\log\widetilde{\sigma}^2 - \log\hat{\sigma}^2) \sim \chi^2(1).$$

When $-2\log\lambda > \chi^2_\alpha(1)$, we reject the null hypothesis $H_0:\ \mu = \mu_0$ at the significance level α.

1.9 Regression Analysis

1.9.1 Setup of the Model

When (X_1, Y_1), (X_2, Y_2), $\cdots$, (X_n, Y_n) are available, suppose that there is a linear relationship between Y and X, i.e.,

$$Y_i = \beta_1 + \beta_2 X_i + u_i, \tag{1.21}$$

for $i = 1, 2, \cdots, n$.

X_i and Y_i denote the ith observations. Y_i is called the **dependent variable** or the **unexplanatory variable**, while X_i is known as the **independent variable** or the **explanatory variable**. β_1 and β_2 are unknown **parameter**s to be estimated. u_i is the unobserved **error term** assumed to be a random variable with mean zero and variance σ^2. β_1 and β_2 are called the **regression coefficient**s.

X_i is assumed to be nonstochastic, but Y_i is stochastic because Y_i depends on the error u_i. The error terms u_1, u_2, $\cdots$, u_n are assumed to be mutually independently and identically distributed. It is assumed that u_i has a distribution with mean zero, i.e., $\mathrm{E}(u_i) = 0$ is assumed. Taking the expectation on both sides of equation (1.21), the expectation of Y_i is represented as:

$$\begin{aligned}\mathrm{E}(Y_i) &= \mathrm{E}(\beta_1 + \beta_2 X_i + u_i) = \beta_1 + \beta_2 X_i + \mathrm{E}(u_i) \\ &= \beta_1 + \beta_2 X_i,\end{aligned} \tag{1.22}$$

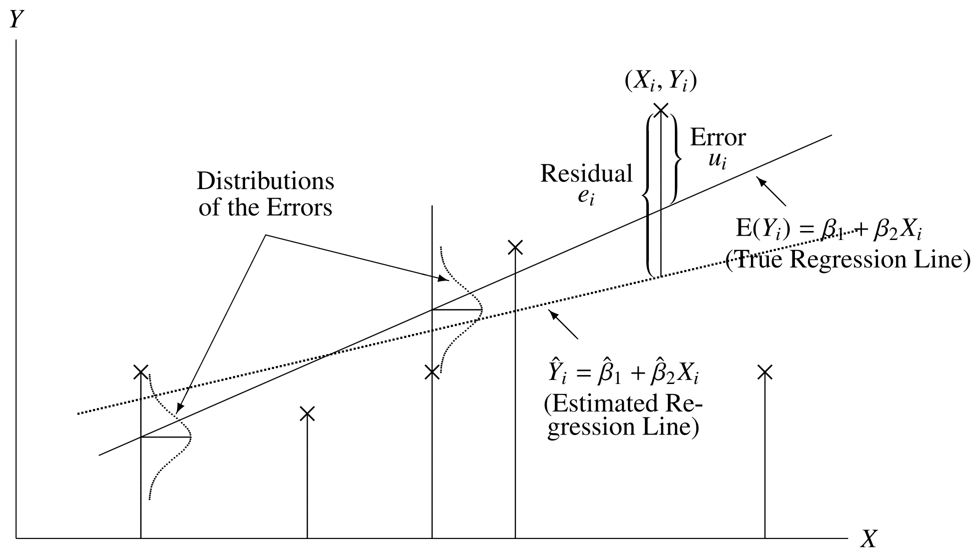
for $i = 1, 2, \cdots, n$. Using $\mathrm{E}(Y_i)$ we can rewrite (1.21) as $Y_i = \mathrm{E}(Y_i) + u_i$. Equation (1.22) represents the true regression line.

Let $\hat{\beta}_1$ and $\hat{\beta}_2$ be estimators of β_1 and β_2. Replacing (β_1, β_2) by $(\hat{\beta}_1, \hat{\beta}_2)$, equation (1.21) turns out to be:

$$Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + e_i, \tag{1.23}$$

for $i = 1, 2, \cdots, n$, where e_i is called the **residual**. The residual e_i is taken as the experimental value of u_i.

Figure 1.4: True and Estimated Regression Lines



We define $\hat{Y}_i$ as follows:

$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i, \tag{1.24}$$

for $i = 1, 2, \cdots, n$, which is interpreted as the predicted value of Y_i. Equation (1.24) indicates the estimated regression line, which is different from equation (1.22). Moreover, using $\hat{Y}_i$ we can rewrite (1.23) as $Y_i = \hat{Y}_i + e_i$.

Equations (1.22) and (1.24) are displayed in Figure 1.4. Consider the case of $n = 6$ for simplicity. $\times$ indicates the observed data series. The true regression line (1.22) is represented by the solid line, while the estimated regression line (1.24) is drawn with the dotted line. Based on the observed data, β_1 and β_2 are estimated as: $\hat{\beta}_1$ and $\hat{\beta}_2$.

In the next section, we consider how to obtain the estimates of β_1 and β_2, i.e., $\hat{\beta}_1$ and $\hat{\beta}_2$.

1.9.2 Ordinary Least Squares Estimation

Suppose that (X_1, Y_1), (X_2, Y_2), $\cdots$, (X_n, Y_n) are available. For the regression model (1.21), we consider estimating β_1 and β_2. Replacing β_1 and β_2 by their estimates $\hat{\beta}_1$ and $\hat{\beta}_2$, remember that the residual e_i is given by:

$$e_i = Y_i - \hat{Y}_i = Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i.$$

The sum of squared residuals is defined as follows:

$$S(\hat{\beta}_1, \hat{\beta}_2) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2.$$

It might be plausible to choose the $\hat{\beta}_1$ and $\hat{\beta}_2$ which minimize the sum of squared residuals, i.e., $S(\hat{\beta}_1, \hat{\beta}_2)$. This method is called the **ordinary least squares (OLS) estimation**. To minimize $S(\hat{\beta}_1, \hat{\beta}_2)$ with respect to $\hat{\beta}_1$ and $\hat{\beta}_2$, we set the partial derivatives equal to zero:

$$\frac{\partial S(\hat{\beta}_1, \hat{\beta}_2)}{\partial \hat{\beta}_1} = -2\sum_{i=1}^{n}(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) = 0,$$
$$\frac{\partial S(\hat{\beta}_1, \hat{\beta}_2)}{\partial \hat{\beta}_2} = -2\sum_{i=1}^{n} X_i(Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i) = 0,$$

which yields the following two equations:

$$\overline{Y} = \hat{\beta}_1 + \hat{\beta}_2 \overline{X}, \tag{1.25}$$
$$\sum_{i=1}^{n} X_i Y_i = n\overline{X}\hat{\beta}_1 + \hat{\beta}_2 \sum_{i=1}^{n} X_i^2, \tag{1.26}$$

where $\overline{Y} = (1/n)\sum_{i=1}^{n} Y_i$ and $\overline{X} = (1/n)\sum_{i=1}^{n} X_i$. Multiplying (1.25) by $n\overline{X}$ and subtracting (1.26), we can derive $\hat{\beta}_2$ as follows:

$$\hat{\beta}_2 = \frac{\sum_{i=1}^{n} X_i Y_i - n\overline{X}\overline{Y}}{\sum_{i=1}^{n} X_i^2 - n\overline{X}^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2}. \tag{1.27}$$

From equation (1.25), $\hat{\beta}_1$ is directly obtained as follows:

$$\hat{\beta}_1 = \overline{Y} - \hat{\beta}_2 \overline{X}. \tag{1.28}$$

When the observed values are taken for Y_i and X_i for $i = 1, 2, \cdots, n$, we say that $\hat{\beta}_1$ and $\hat{\beta}_2$ are called the **ordinary least squares estimate**s (or simply the **least squares estimate**s) of β_1 and β_2. When Y_i for $i = 1, 2, \cdots, n$ are regarded as the random sample, we say that $\hat{\beta}_1$ and $\hat{\beta}_2$ are called the **ordinary least squares estimator**s (or the **least squares estimator**s) of β_1 and β_2.

1.9.3 Properties of Least Squares Estimator

Equation (1.27) is rewritten as:

$$\begin{aligned}\hat{\beta}_2 &= \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} = \frac{\sum_{i=1}^{n}(X_i - \overline{X})Y_i}{\sum_{i=1}^{n}(X_i - \overline{X})^2} - \frac{\overline{Y}\sum_{i=1}^{n}(X_i - \overline{X})}{\sum_{i=1}^{n}(X_i - \overline{X})^2} \\ &= \sum_{i=1}^{n} \frac{X_i - \overline{X}}{\sum_{i=1}^{n}(X_i - \overline{X})^2} Y_i = \sum_{i=1}^{n} \omega_i Y_i. \end{aligned} \tag{1.29}$$

In the third equality, $\sum_{i=1}^n (X_i - \overline{X}) = 0$ is utilized because of $\overline{X} = (1/n)\sum_{i=1}^n X_i$. In the fourth equality, ω_i is defined as:

$$\omega_i = \frac{X_i - \overline{X}}{\sum_{i=1}^n (X_i - \overline{X})^2}.$$

ω_i is nonstochastic because X_i is assumed to be nonstochastic. ω_i has the following properties:

$$\sum_{i=1}^n \omega_i = \sum_{i=1}^n \frac{X_i - \overline{X}}{\sum_{i=1}^n (X_i - \overline{X})^2} = \frac{\sum_{i=1}^n (X_i - \overline{X})}{\sum_{i=1}^n (X_i - \overline{X})^2} = 0, \tag{1.30}$$

$$\sum_{i=1}^n \omega_i X_i = \sum_{i=1}^n \omega_i (X_i - \overline{X}) = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{\sum_{i=1}^n (X_i - \overline{X})^2} = 1, \tag{1.31}$$

$$\begin{aligned}\sum_{i=1}^n \omega_i^2 &= \sum_{i=1}^n \left(\frac{X_i - \overline{X}}{\sum_{i=1}^n (X_i - \overline{X})^2}\right)^2 = \frac{\sum_{i=1}^n (X_i - \overline{X})^2}{\left(\sum_{i=1}^n (X_i - \overline{X})^2\right)^2} \\ &= \frac{1}{\sum_{i=1}^n (X_i - \overline{X})^2}.\end{aligned} \tag{1.32}$$

The first equality of equation (1.31) comes from equation (1.30).

From now on, we focus only on $\hat{\beta}_2$, because usually β_2 is more important than β_1 in the regression model (1.21). In order to obtain the properties of the least squares estimator $\hat{\beta}_2$, we rewrite equation (1.29) as:

$$\begin{aligned}\hat{\beta}_2 &= \sum_{i=1}^n \omega_i Y_i = \sum_{i=1}^n \omega_i(\beta_1 + \beta_2 X_i + u_i) \\ &= \beta_1 \sum_{i=1}^n \omega_i + \beta_2 \sum_{i=1}^n \omega_i X_i + \sum_{i=1}^n \omega_i u_i \\ &= \beta_2 + \sum_{i=1}^n \omega_i u_i.\end{aligned} \tag{1.33}$$

In the fourth equality of (1.33), equations (1.30) and (1.31) are utilized.

Mean and Variance of $\hat{\beta}_2$: $u_1, u_2, \cdots, u_n$ are assumed to be mutually independently and identically distributed with mean zero and variance σ^2, but they are not necessarily normal. Remember that we do not need normality assumption to obtain mean and variance but the normality assumption is required to test a hypothesis.

From equation (1.33), the expectation of $\hat{\beta}_2$ is derived as follows:

$$\begin{aligned}\mathrm{E}(\hat{\beta}_2) &= \mathrm{E}(\beta_2 + \sum_{i=1}^n \omega_i u_i) = \beta_2 + \mathrm{E}(\sum_{i=1}^n \omega_i u_i) \\ &= \beta_2 + \sum_{i=1}^n \omega_i \mathrm{E}(u_i) = \beta_2.\end{aligned} \tag{1.34}$$

It is shown from (1.34) that the ordinary least squares estimator $\hat{\beta}_2$ is an unbiased estimator of β_2.

From (1.33), the variance of $\hat{\beta}_2$ is computed as:

$$\begin{aligned}
\mathrm{V}(\hat{\beta}_2) &= \mathrm{V}(\beta_2 + \sum_{i=1}^{n} \omega_i u_i) = \mathrm{V}(\sum_{i=1}^{n} \omega_i u_i) = \sum_{i=1}^{n} \mathrm{V}(\omega_i u_i) = \sum_{i=1}^{n} \omega_i^2 \mathrm{V}(u_i) \\
&= \sigma^2 \sum_{i=1}^{n} \omega_i^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i - \overline{X})^2}. \qquad (1.35)
\end{aligned}$$

From Theorem on p.15, the second and the fourth equalities hold. The third equality holds because $u_1, u_2, \cdots, u_n$ are mutually independent (see the theorem on p.20). The last equality comes from equation (1.32).

Thus, $\mathrm{E}(\hat{\beta}_2)$ and $\mathrm{V}(\hat{\beta}_2)$ are given by (1.34) and (1.35).

Gauss-Markov Theorem: It has been discussed above that $\hat{\beta}_2$ is represented as (1.29), which implies that $\hat{\beta}_2$ is a linear estimator, i.e., linear in Y_i. In addition, (1.34) indicates that $\hat{\beta}_2$ is an unbiased estimator. Therefore, summarizing these two facts, it is shown that $\hat{\beta}_2$ is a **linear unbiased estimator**. Furthermore, here we show that $\hat{\beta}_2$ has minimum variance within a class of the linear unbiased estimators.

Consider the alternative linear unbiased estimator $\widetilde{\beta}_2$ as follows:

$$\widetilde{\beta}_2 = \sum_{i=1}^{n} c_i Y_i = \sum_{i=1}^{n} (\omega_i + d_i) Y_i,$$

where $c_i = \omega_i + d_i$ is defined and d_i is nonstochastic. Then, $\widetilde{\beta}_2$ is transformed into:

$$\begin{aligned}
\widetilde{\beta}_2 &= \sum_{i=1}^{n} c_i Y_i = \sum_{i=1}^{n} (\omega_i + d_i)(\beta_1 + \beta_2 X_i + u_i) \\
&= \beta_1 \sum_{i=1}^{n} \omega_i + \beta_2 \sum_{i=1}^{n} \omega_i X_i + \sum_{i=1}^{n} \omega_i u_i + \beta_1 \sum_{i=1}^{n} d_i + \beta_2 \sum_{i=1}^{n} d_i X_i + \sum_{i=1}^{n} d_i u_i \\
&= \beta_2 + \beta_1 \sum_{i=1}^{n} d_i + \beta_2 \sum_{i=1}^{n} d_i X_i + \sum_{i=1}^{n} \omega_i u_i + \sum_{i=1}^{n} d_i u_i.
\end{aligned}$$

Equations (1.30) and (1.31) are used in the forth equality. Taking the expectation on both sides of the above equation, we obtain:

$$\begin{aligned}
\mathrm{E}(\widetilde{\beta}_2) &= \beta_2 + \beta_1 \sum_{i=1}^{n} d_i + \beta_2 \sum_{i=1}^{n} d_i X_i + \sum_{i=1}^{n} \omega_i \mathrm{E}(u_i) + \sum_{i=1}^{n} d_i \mathrm{E}(u_i) \\
&= \beta_2 + \beta_1 \sum_{i=1}^{n} d_i + \beta_2 \sum_{i=1}^{n} d_i X_i.
\end{aligned}$$

Note that d_i is not a random variable and that $\mathrm{E}(u_i) = 0$. Since $\widetilde{\beta}_2$ is assumed to be unbiased, we need the following conditions:

$$\sum_{i=1}^{n} d_i = 0, \qquad \sum_{i=1}^{n} d_i X_i = 0.$$

When these conditions hold, we can rewrite $\widetilde{\beta}_2$ as:

$$\widetilde{\beta}_2 = \beta_2 + \sum_{i=1}^{n} (\omega_i + d_i) u_i.$$

The variance of $\widetilde{\beta}_2$ is derived as:

$$\begin{aligned}
\mathrm{V}(\widetilde{\beta}_2) &= \mathrm{V}\Big(\beta_2 + \sum_{i=1}^{n} (\omega_i + d_i) u_i\Big) = \mathrm{V}\Big(\sum_{i=1}^{n} (\omega_i + d_i) u_i\Big) = \sum_{i=1}^{n} \mathrm{V}\Big((\omega_i + d_i) u_i\Big) \\
&= \sum_{i=1}^{n} (\omega_i + d_i)^2 \mathrm{V}(u_i) = \sigma^2 (\sum_{i=1}^{n} \omega_i^2 + 2 \sum_{i=1}^{n} \omega_i d_i + \sum_{i=1}^{n} d_i^2) \\
&= \sigma^2 (\sum_{i=1}^{n} \omega_i^2 + \sum_{i=1}^{n} d_i^2).
\end{aligned}$$

From unbiasedness of $\widetilde{\beta}_2$, using $\sum_{i=1}^{n} d_i = 0$ and $\sum_{i=1}^{n} d_i X_i = 0$, we obtain:

$$\sum_{i=1}^{n} \omega_i d_i = \frac{\sum_{i=1}^{n} (X_i - \overline{X}) d_i}{\sum_{i=1}^{n} (X_i - \overline{X})^2} = \frac{\sum_{i=1}^{n} X_i d_i - \overline{X} \sum_{i=1}^{n} d_i}{\sum_{i=1}^{n} (X_i - \overline{X})^2} = 0,$$

which is utilized to obtain the variance of $\widetilde{\beta}_2$ in the third line of the above equation. From (1.35), the variance of $\hat{\beta}_2$ is given by: $\mathrm{V}(\hat{\beta}_2) = \sigma^2 \sum_{i=1}^{n} \omega_i^2$. Therefore, we have:

$$\mathrm{V}(\widetilde{\beta}_2) \geq \mathrm{V}(\hat{\beta}_2),$$

because of $\sum_{i=1}^{n} d_i^2 \geq 0$. When $\sum_{i=1}^{n} d_i^2 = 0$, i.e., when $d_1 = d_2 = \cdots = d_n = 0$, we have the equality: $\mathrm{V}(\widetilde{\beta}_2) = \mathrm{V}(\hat{\beta}_2)$. Thus, in the case of $d_1 = d_2 = \cdots = d_n = 0$, $\hat{\beta}_2$ is equivalent to $\widetilde{\beta}_2$.

As shown above, the least squares estimator $\hat{\beta}_2$ gives us the **linear unbiased minimum variance estimator**, or equivalently the **best linear unbiased estimator (BLUE)**, which is called the **Gauss-Markov theorem**.

Asymptotic Properties of $\hat{\beta}_2$: We assume that as n goes to infinity we have the following:

$$\frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2 \longrightarrow M < \infty,$$

where M is a constant value. From (1.32), we obtain:

$$n\sum_{i=1}^{n}\omega_i^2 = \frac{1}{(1/n)\sum_{i=1}^{n}(X_i-\overline{X})} \longrightarrow \frac{1}{M}.$$

Note that $f(x_n) \longrightarrow f(m)$ when $x_n \longrightarrow m$, where m is a constant value and $f(\cdot)$ is a function.

Here, we show both consistency of $\hat{\beta}_2$ and asymptotic normality of $\sqrt{n}(\hat{\beta}_2-\beta_2)$. First, we prove that $\hat{\beta}_2$ is a consistent estimator of β_2. As in (1.10), Chebyshev's inequality is given by:

$$P(|X-\mu|>\epsilon) \leq \frac{\sigma^2}{\epsilon^2},$$

where $\mu = \mathrm{E}(X)$ and $\sigma^2 = \mathrm{V}(X)$. Here, we replace X, $\mathrm{E}(X)$ and $\mathrm{V}(X)$ by $\hat{\beta}_2$,

$$\mathrm{E}(\hat{\beta}_2)=\beta_2, \quad \mathrm{V}(\hat{\beta}_2)=\sigma^2\sum_{i=1}^{n}\omega_i^2 = \frac{\sigma^2}{\sum_{i=1}^{n}(X_i-\overline{X})},$$

respectively. Then, when $n \longrightarrow \infty$, we obtain the following result:

$$P(|\hat{\beta}_2-\beta_2|>\epsilon) \leq \frac{\sigma^2\sum_{i=1}^{n}\omega_i^2}{\epsilon^2} = \frac{\sigma^2}{\epsilon^2\sum_{i=1}^{n}(X_i-\overline{X})} \longrightarrow 0,$$

where $\sum_{i=1}^{n}\omega_i^2 \longrightarrow 0$ because $n\sum_{i=1}^{n}\omega_i^2 \longrightarrow 1/M$ from the assumption. Thus, we obtain the result that $\hat{\beta}_2 \longrightarrow \beta_2$ as $n \longrightarrow \infty$. Therefore, we can conclude that $\hat{\beta}_2$ is a consistent estimator of β_2.

Next, we want to show that $\sqrt{n}(\hat{\beta}_2-\beta_2)$ is asymptotically normal. Noting that $\hat{\beta}_2 = \beta_2 + \sum_{i=1}^{n}\omega_i u_i$ as in (1.33) from Corollary 2 on p.35 (central limit theorem), asymptotic normality is shown as follows:

$$\frac{\sum_{i=1}^{n}\omega_i u_i - \mathrm{E}(\sum_{i=1}^{n}\omega_i u_i)}{\sqrt{\mathrm{V}(\sum_{i=1}^{n}\omega_i u_i)}} = \frac{\sum_{i=1}^{n}\omega_i u_i}{\sigma\sqrt{\sum_{i=1}^{n}\omega_i^2}} = \frac{\hat{\beta}_2-\beta_2}{\sigma/\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}} \longrightarrow N(0,1),$$

where $\mathrm{E}(\sum_{i=1}^{n}\omega_i u_i)=0$, $\mathrm{V}(\sum_{i=1}^{n}\omega_i u_i)=\sigma^2\sum_{i=1}^{n}\omega_i^2$ and $\sum_{i=1}^{n}\omega_i u_i = \hat{\beta}_2-\beta_2$ are substituted in the second equality. Moreover, we can rewrite as follows:

$$\frac{\hat{\beta}_2-\beta_2}{\sigma/\sqrt{\sum_{i=1}^{n}(X_i-\overline{X})^2}} = \frac{\sqrt{n}(\hat{\beta}_2-\beta_2)}{\sigma/\sqrt{(1/n)\sum_{i=1}^{n}(X_i-\overline{X})^2}} \longrightarrow \frac{\sqrt{n}(\hat{\beta}_2-\beta_2)}{\sigma/\sqrt{M}} \longrightarrow N(0,1),$$

or equivalently,

$$\sqrt{n}(\hat{\beta}_2-\beta_2) \longrightarrow N(0,\frac{\sigma^2}{M}).$$

Thus, asymptotic normality of $\sqrt{n}(\hat{\beta}_2-\beta_2)$ is shown.

Finally, replacing σ^2 by its consistent estimator s^2, it is known as follows:

$$\frac{\hat{\beta}_2 - \beta_2}{s\sqrt{\sum_{i=1}^n (X_i - \overline{X})^2}} \longrightarrow N(0,1), \tag{1.36}$$

where s^2 is defined as:

$$s^2 = \frac{1}{n-2}\sum_{i=1}^n e_i^2 = \frac{1}{n-2}\sum_{i=1}^n (Y_i - \hat{\beta}_1 - \hat{\beta}_2 X_i)^2, \tag{1.37}$$

which is a consistent and unbiased estimator of σ^2.

Thus, using (1.36), in large sample we can construct the confidence interval discussed in Section 1.7.6 and test the hypothesis discussed in Section 1.8.

Exact Distribution of $\hat{\beta}_2$: We have shown asymptotic normality of $\sqrt{n}(\hat{\beta}_2 - \beta_2)$, which is one of the large sample properties. Now, we discuss the small sample properties of $\hat{\beta}_2$. In order to obtain the distribution of $\hat{\beta}_2$ in small sample, the distribution of the error term has to be assumed. Therefore, the extra assumption is that $u_i \sim N(0, \sigma^2)$. Writing equation (1.33), again, $\hat{\beta}_2$ is represented as:

$$\hat{\beta}_2 = \beta_2 + \sum_{i=1}^n \omega_i u_i.$$

First, we obtain the distribution of the second term in the above equation. From Theorem on p.29, $\sum_{i=1}^n \omega_i u_i$ is distributed as:

$$\sum_{i=1}^n \omega_i u_i \sim N(0, \sigma^2 \sum_{i=1}^n \omega_i^2),$$

which is easily shown using the moment-generating function. Therefore, from Example 1.9 on p.23, $\hat{\beta}_2$ is distributed as:

$$\hat{\beta}_2 = \beta_2 + \sum_{i=1}^n \omega_i u_i \sim N(\beta_2, \sigma^2 \sum_{i=1}^n \omega_i^2),$$

or equivalently,

$$\frac{\hat{\beta}_2 - \beta_2}{\sigma\sqrt{\sum_{i=1}^n \omega_i^2}} = \frac{\hat{\beta}_2 - \beta_2}{\sigma/\sqrt{\sum_{i=1}^n (X_i - \overline{X})^2}} \sim N(0,1),$$

for any n.

Moreover, replacing σ^2 by its estimator s^2 defined in (1.37), it is known that we have:

$$\frac{\hat{\beta}_2 - \beta_2}{s/\sqrt{\sum_{i=1}^n (X_i - \overline{X})^2}} \sim t(n-2),$$

where $t(n-2)$ denotes t distribution with $n-2$ degrees of freedom. See Section 2.2.10 for derivation of the t distribution. Thus, under normality assumption on the error term u_i, the $t(n-2)$ distribution is used for the confidence interval and the testing hypothesis in small sample.

1.9.4 Multiple Regression Model

In Sections 1.9.1 – 1.9.3, only one independent variable, i.e., X_i, is taken into the regression model. In this section, we extend it to more independent variables, which is called the **multiple regression**. We consider the following regression model:

$$\begin{aligned} Y_i &= \beta_1 X_{i,1} + \beta_2 X_{i,2} + \cdots + \beta_k X_{i,k} + u_i \\ &= X_i\beta + u_i, \end{aligned}$$

for $i = 1, 2, \cdots, n$, where X_i and β denote a $1 \times k$ vector of the independent variables and a $k \times 1$ vector of the unknown parameters to be estimated, which are represented as:

$$X_i = (X_{i,1}, X_{i,2}, \cdots, X_{i,k}), \qquad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}.$$

$X_{i,j}$ denotes the ith observation of the jth independent variable. The case of $k = 2$ and $X_{i,1} = 1$ for all i is exactly equivalent to (1.21). Therefore, the matrix form above is a generalization of (1.21). Writing all the equations for $i = 1, 2, \cdots, n$, we have:

$$\begin{aligned} Y_1 &= \beta_1 X_{1,1} + \beta_2 X_{1,2} + \cdots + \beta_k X_{1,k} + u_1, \\ Y_2 &= \beta_1 X_{2,1} + \beta_2 X_{2,2} + \cdots + \beta_k X_{2,k} + u_2, \\ &\vdots \\ Y_n &= \beta_1 X_{n,1} + \beta_2 X_{n,2} + \cdots + \beta_k X_{n,k} + u_n, \end{aligned}$$

which is rewritten as:

$$\begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix} = \begin{pmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,k} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n,1} & X_{n,2} & \cdots & X_{n,k} \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix} + \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_k \end{pmatrix}.$$

Again, the above equation is compactly rewritten as:

$$Y = X\beta + u. \tag{1.38}$$

where Y, X and u are denoted by:

$$Y = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \qquad X = \begin{pmatrix} X_{1,1} & X_{1,2} & \cdots & X_{1,k} \\ X_{2,1} & X_{2,2} & \cdots & X_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ X_{n,1} & X_{n,2} & \cdots & X_{n,k} \end{pmatrix}, \qquad u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_k \end{pmatrix}.$$

Utilizing the matrix form (1.38), we derive the ordinary least squares estimator of β, denoted by $\hat{\beta}$. In equation (1.38), replacing β by $\hat{\beta}$, we have the following equation:

$$Y = X\hat{\beta} + e,$$

where e denotes a $1 \times n$ vector of the residuals. The ith element of e is given by e_i. The sum of squared residuals is written as follows:

$$\begin{aligned} S(\hat{\beta}) &= \sum_{i=1}^{n} e_i^2 = e'e = (Y - X\hat{\beta})'(Y - X\hat{\beta}) = (Y' - \hat{\beta}'X')(Y - X\hat{\beta}) \\ &= Y'Y - Y'X\hat{\beta} - \hat{\beta}'X'Y + \hat{\beta}'X'X\hat{\beta} = Y'Y - 2Y'X\hat{\beta} + \hat{\beta}'X'X\hat{\beta}. \end{aligned}$$

See Appendix 1.5 for the transpose in the fourth equality. In the last equality, note that $\hat{\beta}'X'Y = Y'X\hat{\beta}$ because both are scalars. To minimize $S(\hat{\beta})$ with respect to $\hat{\beta}$, we set the first derivative of $S(\hat{\beta})$ equal to zero, i.e.,

$$\frac{\partial S(\hat{\beta})}{\partial \hat{\beta}} = -2X'Y + 2X'X\hat{\beta} = 0.$$

See Appendix 1.5 for the derivatives of matrices. Solving the equation above with respect to $\hat{\beta}$, the ordinary least squares estimator of β is given by:

$$\hat{\beta} = (X'X)^{-1}X'Y. \tag{1.39}$$

See Appendix 1.5 for the inverse of the matrix. Thus, the ordinary least squares estimator is derived in the matrix form.

Now, in order to obtain the properties of $\hat{\beta}$ such as mean, variance, distribution and so on, (1.39) is rewritten as follows:

$$\begin{aligned} \hat{\beta} &= (X'X)^{-1}X'Y = (X'X)^{-1}X'(X\beta + u) = (X'X)^{-1}X'X\beta + (X'X)^{-1}X'u \\ &= \beta + (X'X)^{-1}X'u. \end{aligned} \tag{1.40}$$

Taking the expectation on both sides of equation (1.40), we have the following:

$$\mathrm{E}(\hat{\beta}) = \mathrm{E}(\beta + (X'X)^{-1}X'u) = \beta + (X'X)^{-1}X'\mathrm{E}(u) = \beta,$$

because of $\mathrm{E}(u) = 0$ by the assumption of the error term u_i. Thus, unbiasedness of $\hat{\beta}$ is shown.

The variance of $\hat{\beta}$ is obtained as:

$$\begin{aligned} \mathrm{V}(\hat{\beta}) &= \mathrm{E}((\hat{\beta} - \beta)(\hat{\beta} - \beta)') = \mathrm{E}\Big((X'X)^{-1}X'u((X'X)^{-1}X'u)'\Big) \\ &= \mathrm{E}((X'X)^{-1}X'uu'X(X'X)^{-1}) = (X'X)^{-1}X'\mathrm{E}(uu')X(X'X)^{-1} \\ &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} = \sigma^2(X'X)^{-1}. \end{aligned}$$

The first equality is the definition of variance in the case of vector. In the fifth equality, $\mathrm{E}(uu') = \sigma^2 I_n$ is used, which implies that $\mathrm{E}(u_i^2) = \sigma^2$ for all i and $\mathrm{E}(u_i u_j) = 0$ for

$i \neq j$. Remember that $u_1, u_2, \cdots, u_n$ are assumed to be mutually independently and identically distributed with mean zero and variance σ^2.

Under normality assumption on the error term u, it is known that the distribution of $\hat{\beta}$ is given by:

$$\hat{\beta} \sim N(\beta, \sigma^2 (X'X)^{-1}).$$

Taking the jth element of $\hat{\beta}$, its distribution is given by:

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 a_{jj}), \qquad \text{i.e.,} \qquad \frac{\hat{\beta}_j - \beta_j}{\sigma \sqrt{a_{jj}}} \sim N(0, 1),$$

where a_{jj} denotes the jth diagonal element of $(X'X)^{-1}$.

Replacing σ^2 by its estimator s^2, we have the following t distribution:

$$\frac{\hat{\beta}_j - \beta_j}{s \sqrt{a_{jj}}} \sim t(n-k),$$

where $t(n-k)$ denotes the t distribution with $n-k$ degrees of freedom. s^2 is taken as follows:

$$s^2 = \frac{1}{n-k} \sum_{i=1}^{n} e_i^2 = \frac{1}{n-k} e'e = \frac{1}{n-k} (Y - X\hat{\beta})'(Y - X\hat{\beta}),$$

which leads to an unbiased estimator of σ^2.

Using the central limit theorem, without normality assumption we can show that as $n \longrightarrow \infty$, under the condition of $(1/n)X'X \longrightarrow M$ we have the following result:

$$\frac{\hat{\beta}_j - \beta_j}{s \sqrt{a_{jj}}} \longrightarrow N(0, 1),$$

where M denotes a $k \times k$ constant matrix.

Thus, we can construct the confidence interval and the testing procedure, using the t distribution under the normality assumption or the normal distribution without the normality assumption.

Appendix 1.1: Integration by Substitution

Univariate Case: For a function of x, $f(x)$, we perform integration by substitution, using $x = \psi(y)$. Then, it is easy to obtain the following formula:

$$\int f(x)\, \mathrm{d}x = \int \psi'(y) f(\psi(y))\, \mathrm{d}y,$$

which formula is called the **integration by substitution**.

Proof:
Let $F(x)$ be the integration of $f(x)$, i.e.,

$$F(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t,$$

which implies that $F'(x) = f(x)$.
Differentiating $F(x) = F(\psi(y))$ with respect to y, we have:

$$f(x) \equiv \frac{\mathrm{d}F(\psi(y))}{\mathrm{d}y} = \frac{\mathrm{d}F(x)}{\mathrm{d}x}\frac{\mathrm{d}x}{\mathrm{d}y} = f(x)\psi'(y) = f(\psi(y))\psi'(y).$$

Bivariate Case: For $f(x, y)$, define $x = \psi_1(u, v)$ and $y = \psi_2(u, v)$.

$$\iint f(x, y)\,\mathrm{d}x\,\mathrm{d}y = \iint J f(\psi_1(u, v), \psi_2(u, v))\,\mathrm{d}u\,\mathrm{d}v,$$

where J is called the **Jacobian**, which represents the following determinant:

$$J = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} = \frac{\partial x}{\partial u}\frac{\partial y}{\partial v} - \frac{\partial x}{\partial v}\frac{\partial y}{\partial u}.$$

Appendix 1.2: Integration by Parts

Let $h(x)$ and $g(x)$ be functions of x. Then, we have the following formula:

$$\int h(x)g'(x)\,\mathrm{d}x = h(x)g(x) - \int h'(x)g(x)\,\mathrm{d}x,$$

which is called the **integration by parts**.

Proof:
Consider the derivative of $f(x)g(x)$ with respect to x, i.e.,

$$\big(f(x)g(x)\big)' = f'(x)g(x) + f(x)g'(x).$$

Integrating the above equation on both sides, we have:

$$\int \big(f(x)g(x)\big)'\,\mathrm{d}x = \int f'(x)g(x)\,\mathrm{d}x + \int f(x)g'(x)\,\mathrm{d}x.$$

Therefore, we obtain:

$$f(x)g(x) = \int f'(x)g(x)\,\mathrm{d}x + \int f(x)g'(x)\,\mathrm{d}x.$$

Thus, the following result is derived.

$$\int f(x)g'(x)\,\mathrm{d}x = f(x)g(x) - \int f'(x)g(x)\,\mathrm{d}x.$$

When we want to integrate $f(x)g'(x)$ within the range between a and b for $a < b$, the above formula is modified as:

$$\int_a^b f(x)g'(x)\,\mathrm{d}x = \Big[f(x)g(x)\Big]_a^b - \int_a^b f'(x)g(x)\,\mathrm{d}x.$$

Appendix 1.3: Taylor Series Expansion

Consider approximating $f(x)$ around $x = x_0$ by the **Taylor series expansion**.. Then, $f(x)$ is approximated as follows:

$$\begin{aligned} f(x) &= f(x_0) + f'(x_0)(x - x_0) + \frac{1}{2!}f''(x_0)(x - x_0)^2 + \frac{1}{3!}f'''(x_0)(x - x_0)^3 + \cdots \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} f^{(n)}(x_0)(x - x_0)^n, \end{aligned}$$

where $f^{(n)}(x_0)$ denotes the nth derivative of $f(x)$ evaluated at $x = x_0$. Note that $f^{(0)}(x_0) = f(x_0)$ and $0! = 1$.

In addition, the following approximation is called the ***k*th order Taylor series expansion**:

$$f(x) \approx \sum_{n=0}^{k} \frac{1}{n!} f^{(n)}(x_0)(x - x_0)^n.$$

Appendix 1.4: Cramer-Rao Inequality

As seen in (1.13) and (1.14), the Cramer-Rao inequality is given by:

$$\mathrm{V}(\hat{\theta}_n) \geq \frac{\sigma^2(\theta)}{n},$$

where

$$\sigma^2(\theta) = \frac{1}{\mathrm{E}\Big(\big(\frac{\partial \log f(X;\theta)}{\partial\theta}\big)^2\Big)} = \frac{1}{\mathrm{V}\Big(\big(\frac{\partial \log f(X;\theta)}{\partial\theta}\big)\Big)} = -\frac{1}{\mathrm{E}\Big(\frac{\partial^2 \log f(X;\theta)}{\partial\theta^2}\Big)}.$$

Proof:
We prove the above inequality and the equalities in $\sigma^2(\theta)$. The likelihood function $l(\theta; x) = l(\theta; x_1, x_2, \cdots, x_n)$ is a joint density of $X_1, X_2, \cdots, X_n$. Therefore, the integration of $l(\theta; x_1, x_2, \cdots, x_n)$ with respect to $x_1, x_2, \cdots, x_n$ is equal to one. See Section 1.7.5 for the likelihood function. That is, we have the following equation:

$$1 = \int l(\theta; x)\,\mathrm{d}x, \tag{1.41}$$

where the likelihood function $l(\theta; x)$ is given by $l(\theta; x) = \prod_{i=1}^{n} f(x_i; \theta)$ and $\int \cdots\,\mathrm{d}x$ implies n-tuple integral.

Differentiating both sides of equation (1.41) with respect to θ, we obtain the following equation:

$$\begin{aligned} 0 &= \int \frac{\partial l(\theta; x)}{\partial\theta}\,\mathrm{d}x = \int \frac{1}{l(\theta; x)}\frac{\partial l(\theta; x)}{\partial\theta} l(\theta; x)\,\mathrm{d}x \\ &= \int \frac{\partial \log l(\theta; x)}{\partial\theta} l(\theta; x)\,\mathrm{d}x = \mathrm{E}\Big(\frac{\partial \log l(\theta; X)}{\partial\theta}\Big), \end{aligned} \tag{1.42}$$

which implies that the expectation of $\partial \log l(\theta; X)/\partial\theta$ is equal to zero. In the third equality, note that $\mathrm{d}\log x/\,\mathrm{d}x = 1/x$.

Now, let $\hat\theta_n$ be an estimator of θ. The definition of the mathematical expectation of the estimator $\hat\theta_n$ is represented as:

$$\mathrm{E}(\hat\theta_n) = \int \hat\theta_n l(\theta; x)\,\mathrm{d}x. \tag{1.43}$$

Differentiating equation (1.43) with respect to θ on both sides, we can rewrite as follows:

$$\begin{aligned} \frac{\partial \mathrm{E}(\hat\theta_n)}{\partial\theta} &= \int \hat\theta_n \frac{\partial l(\theta; x)}{\partial\theta}\,\mathrm{d}x = \int \hat\theta_n \frac{\partial \log l(\theta; x)}{\partial\theta} l(\theta; x)\,\mathrm{d}x \\ &= \int \Big(\hat\theta_n - \mathrm{E}(\hat\theta_n)\Big)\Big(\frac{\partial \log l(\theta; x)}{\partial\theta} - \mathrm{E}(\frac{\partial \log l(\theta; x)}{\partial\theta})\Big) l(\theta; x)\,\mathrm{d}x \\ &= \mathrm{Cov}\Big(\hat\theta_n, \frac{\partial \log l(\theta; X)}{\partial\theta}\Big). \end{aligned} \tag{1.44}$$

In the second equality, $\mathrm{d}\log x/\,\mathrm{d}x = 1/x$ is utilized. The third equality holds because of $\mathrm{E}(\partial \log l(\theta; X)/\partial\theta) = 0$ from equation (1.42).

For simplicity of discussion, suppose that θ is a scalar. Taking the square on both sides of equation (1.44), we obtain the following expression:

$$\begin{aligned} \Big(\frac{\partial \mathrm{E}(\hat\theta_n)}{\partial\theta}\Big)^2 &= \left(\mathrm{Cov}\Big(\hat\theta_n, \frac{\partial \log l(\theta; X)}{\partial\theta}\Big)\right)^2 = \rho^2 \mathrm{V}(\hat\theta_n)\mathrm{V}\Big(\frac{\partial \log l(\theta; X)}{\partial\theta}\Big) \\ &\le \mathrm{V}(\hat\theta_n)\mathrm{V}\left(\frac{\partial \log l(\theta; X)}{\partial\theta}\right), \end{aligned} \tag{1.45}$$

where ρ denotes the correlation coefficient between $\hat{\theta}_n$ and $\partial \log l(\theta; X)/\partial\theta$. Note that we have the definition of ρ is given by:

$$\rho = \frac{\mathrm{Cov}\Big(\hat{\theta}_n, \dfrac{\partial \log l(\theta; X)}{\partial\theta}\Big)}{\sqrt{\mathrm{V}(\hat{\theta}_n)}\sqrt{\mathrm{V}\Big(\dfrac{\partial \log l(\theta; X)}{\partial\theta}\Big)}}.$$

Moreover, we have $-1 \le \rho \le 1$ (i.e., $\rho^2 \le 1$). Then, the inequality (1.45) is obtained, which is rewritten as:

$$\mathrm{V}(\hat{\theta}_n) \ge \frac{\Big(\dfrac{\partial \mathrm{E}(\hat{\theta}_n)}{\partial\theta}\Big)^2}{\mathrm{V}\Big(\dfrac{\partial \log l(\theta; X)}{\partial\theta}\Big)}. \tag{1.46}$$

When $\mathrm{E}(\hat{\theta}_n) = \theta$, i.e., when $\hat{\theta}_n$ is an unbiased estimator of θ, the numerator in the right-hand side of equation (1.46) is equal to one. Therefore, we have the following result:

$$\mathrm{V}(\hat{\theta}_n) \ge \frac{1}{\mathrm{V}\Big(\dfrac{\partial \log l(\theta; X)}{\partial\theta}\Big)} = \frac{1}{\mathrm{E}\Big(\Big(\dfrac{\partial \log l(\theta; X)}{\partial\theta}\Big)^2\Big)}.$$

Note that we have $\mathrm{V}(\partial \log l(\theta; X)/\partial\theta) = \mathrm{E}((\partial \log l(\theta; X)/\partial\theta)^2)$ in the equality above, because of $\mathrm{E}(\partial \log l(\theta; X)/\partial\theta) = 0$.

Moreover, the denominator in the right-hand side of the above inequality is rewritten as follows:

$$\begin{aligned}
\mathrm{E}\Big(\Big(\frac{\partial \log l(\theta; X)}{\partial\theta}\Big)^2\Big) &= \mathrm{E}\Big(\Big(\sum_{i=1}^{n} \frac{\partial \log f(X_i; \theta)}{\partial\theta}\Big)^2\Big) = \sum_{i=1}^{n} \mathrm{E}\Big(\Big(\frac{\partial \log f(X_i; \theta)}{\partial\theta}\Big)^2\Big) \\
&= n\mathrm{E}\Big(\Big(\frac{\partial \log f(X; \theta)}{\partial\theta}\Big)^2\Big) = n\int_{-\infty}^{\infty} \Big(\frac{\partial \log f(x; \theta)}{\partial\theta}\Big)^2 f(x; \theta)\,\mathrm{d}x.
\end{aligned}$$

In the first equality, $\log l(\theta; X) = \sum_{i=1}^{n} \log f(X_i; \theta)$ is utilized. Since X_i, $i = 1, 2, \cdots, n$, are mutually independent, the second equality holds. The third equality holds because $X_1, X_2, \cdots, X_n$ are identically distributed.

Therefore, we obtain the following inequality:

$$\mathrm{V}(\hat{\theta}_n) \ge \frac{1}{\mathrm{E}\Big(\Big(\dfrac{\partial \log l(\theta; X)}{\partial\theta}\Big)^2\Big)} = \frac{1}{n\mathrm{E}\Big(\Big(\dfrac{\partial \log f(X; \theta)}{\partial\theta}\Big)^2\Big)} = \frac{\sigma^2(\theta)}{n},$$

which is equivalent to (1.13).

Next, we prove the equalities in (1.14), i.e.,

$$-\mathrm{E}\Big(\frac{\partial^2 \log f(X; \theta)}{\partial\theta^2}\Big) = \mathrm{E}\Big(\Big(\frac{\partial \log f(X; \theta)}{\partial\theta}\Big)^2\Big) = \mathrm{V}\Big(\frac{\partial \log f(X; \theta)}{\partial\theta}\Big).$$

Differentiating $\int f(x;\theta)\,\mathrm{d}x = 1$ with respect to θ, we obtain as follows:

$$\int \frac{\partial f(x;\theta)}{\partial\theta}\,\mathrm{d}x = 0.$$

We assume that the range of x does not depend on the parameter θ and that $\partial f(x;\theta)/\partial\theta$ exists. The above equation is rewritten as:

$$\int \frac{\partial \log f(x;\theta)}{\partial\theta} f(x;\theta)\,\mathrm{d}x = 0, \tag{1.47}$$

or equivalently,

$$\mathrm{E}\Big(\frac{\partial \log f(X;\theta)}{\partial\theta}\Big) = 0. \tag{1.48}$$

Again, differentiating equation (1.47) with respect to θ,

$$\int \frac{\partial^2 \log f(x;\theta)}{\partial\theta^2} f(x;\theta)\,\mathrm{d}x + \int \frac{\partial \log f(x;\theta)}{\partial\theta}\frac{\partial f(x;\theta)}{\partial\theta}\,\mathrm{d}x = 0,$$

i.e.,

$$\int \frac{\partial^2 \log f(x;\theta)}{\partial\theta^2} f(x;\theta)\,\mathrm{d}x + \int \Big(\frac{\partial \log f(x;\theta)}{\partial\theta}\Big)^2 f(x;\theta)\,\mathrm{d}x = 0,$$

i.e.,

$$\mathrm{E}\Big(\frac{\partial^2 \log f(x;\theta)}{\partial\theta^2}\Big) + \mathrm{E}\Big(\Big(\frac{\partial \log f(x;\theta)}{\partial\theta}\Big)^2\Big) = 0.$$

Thus, we obtain:

$$-\mathrm{E}\Big(\frac{\partial^2 \log f(x;\theta)}{\partial\theta^2}\Big) = \mathrm{E}\Big(\Big(\frac{\partial \log f(x;\theta)}{\partial\theta}\Big)^2\Big).$$

Moreover, from equation (1.48), the following equation is derived.

$$\mathrm{E}\Big(\Big(\frac{\partial \log f(x;\theta)}{\partial\theta}\Big)^2\Big) = \mathrm{V}\Big(\frac{\partial \log f(x;\theta)}{\partial\theta}\Big).$$

Therefore, we have:

$$-\,\mathrm{E}\Big(\frac{\partial^2 \log f(X;\theta)}{\partial\theta^2}\Big) = \mathrm{E}\Big(\Big(\frac{\partial \log f(X;\theta)}{\partial\theta}\Big)^2\Big) = \mathrm{V}\Big(\frac{\partial \log f(X;\theta)}{\partial\theta}\Big).$$

Thus, the Cramer-Rao inequality is derived as:

$$\mathrm{V}(\hat{\theta}_n) \geq \frac{\sigma^2(\theta)}{n},$$

where

$$\sigma^2(\theta) = \frac{1}{\mathrm{E}\Big(\Big(\frac{\partial \log f(X;\theta)}{\partial\theta}\Big)^2\Big)} = \frac{1}{\mathrm{V}\Big(\Big(\frac{\partial \log f(X;\theta)}{\partial\theta}\Big)\Big)} = -\frac{1}{\mathrm{E}\Big(\frac{\partial^2 \log f(X;\theta)}{\partial\theta^2}\Big)}.$$

Appendix 1.5: Some Formulas of Matrix Algebra

1. Let $A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{l1} & a_{l2} & \cdots & a_{lk} \end{pmatrix} = [a_{ij}]$, which is a $l \times k$ matrix, where a_{ij} denotes ith row and jth column of A. The **transpose** of A, denoted by A', is defined as:

   $$A' = \begin{pmatrix} a_{11} & a_{21} & \cdots & a_{l1} \\ a_{12} & a_{22} & \cdots & a_{l2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1k} & a_{2k} & \cdots & a_{lk} \end{pmatrix} = [a_{ji}],$$

   where the ith row of A' is the ith column of A.

2. $(Ax)' = x'A'$,

   where A and x are a $l \times k$ matrix and a $k \times 1$ vector, respectively.

3. $a' = a$,

   where a denotes a scalar.

4. $\dfrac{\partial a'x}{\partial x} = a$,

   where a and x are $k \times 1$ vectors.

5. $\dfrac{\partial x'Ax}{\partial x} = (A + A')x$,

   where A and x are a $k \times k$ matrix and a $k \times 1$ vector, respectively.

   Especially, when A is symmetric,

   $\dfrac{\partial x'Ax}{\partial x} = 2Ax$.

6. Let A and B be $k \times k$ matrices, and I_k be a $k \times k$ **identity matrix** (one in the diagonal elements and zero in the other elements).

   When $AB = I_k$, B is called the **inverse** of A, denoted by $B = A^{-1}$.

   That is, $AA^{-1} = A^{-1}A = I_k$.

7. Let A be a $k \times k$ matrix and x be a $k \times 1$ vector.

   If A is a **positive definite matrix**, for any x we have:

   $$x'Ax > 0.$$

   If A is a **positive semidefinite matrix**, for any x we have:

   $$x'Ax \geq 0.$$

If A is a **negative definite matrix**, for any x we have:

$$x'Ax < 0.$$

If A is a **negative semidefinite matrix**, for any x we have:

$$x'Ax \leq 0.$$

References

Greene, W.H., 1993, *Econometric Analysis* (Second Edition), Prentice Hall.

Greene, W.H., 1997, *Econometric Analysis* (Third Edition), Prentice-Hall.

Greene, W.H., 2000, *Econometric Analysis* (Fourth Edition), Prentice-Hall.

Hogg, R.V. and Craig, A.T., 1995, *Introduction to Mathematical Statistics* (Fifth Edition), Prentice Hall.

Judge, G., Hill, C., Griffiths, W. and Lee, T., 1980, *The Theory and Practice of Econometrics*, John Wiley & Sons.

Mood, A.M., Graybill, F.A. and Boes, D.C., 1974, *Introduction to the Theory of Statistics* (Third Edition), McGraw-Hill.

Stuart, A. and Ord, J.K., 1991, *Kendall's Advanced Theory of Statistics, Vol.2* (Fifth Edition), Edward Arnold.

Stuart, A. and Ord, J.K., 1994, *Kendall's Advanced Theory of Statistics, Vol.1* (Sixth Edition), Edward Arnold.

Exercises and Answers to Chapter 1

$\boxed{1}$ The continuous type of random variable X has the following density function:

$$f(x) = \begin{cases} a - x, & \text{if } 0 < x < a, \\ 0, & \text{otherwise.} \end{cases}$$

Answer the following questions.

(1) Find a.

(2) Obtain mean and variance of X.

(3) When $Y = X^2$, derive the density function of Y.

[Answer]

(1) From the property of the density function, i.e., $\int f(x)\,\mathrm{d}x = 1$, we need to have:

$$\int f(x)\,\mathrm{d}x = \int_0^a (a-x)\,\mathrm{d}x = \left[ax - \frac{1}{2}x^2\right]_0^a = \frac{1}{2}a^2 = 1.$$

Therefore, $a = \sqrt{2}$ is obtained, taking into account $a > 0$.

(2) The definitions of mean and variance are given by: $\mathrm{E}(X) = \int x f(x)\,\mathrm{d}x$ and $\mathrm{V}(X) = \int (x-\mu)^2 f(x)\,\mathrm{d}x$, where $\mu = \mathrm{E}(X)$. Therefore, mean of X is:

$$\begin{aligned}\mathrm{E}(X) &= \int x f(x)\,\mathrm{d}x = \int_0^a x(a-x)\,\mathrm{d}x = \left[\frac{1}{2}ax^2 - \frac{1}{3}x^3\right]_0^a = \frac{1}{6}a^3 \\ &= \frac{\sqrt{2}}{3} \qquad \longleftarrow \qquad a = \sqrt{2} \text{ is substituted.}\end{aligned}$$

Variance of X is:

$$\begin{aligned}\mathrm{V}(X) &= \int (x-\mu)^2 f(x)\,\mathrm{d}x = \int x^2 f(x)\,\mathrm{d}x - \mu^2 = \int_0^a x^2(a-x)\,\mathrm{d}x - \mu^2 \\ &= \left[\frac{1}{3}ax^3 - \frac{1}{4}x^4\right]_0^a - \mu^2 = \frac{1}{12}a^4 - \mu^2 = \frac{1}{3} - \left(\frac{\sqrt{2}}{3}\right)^2 = \frac{1}{9}.\end{aligned}$$

(3) Let $f(x)$ be the density function of X and $F(x)$ be the distribution function of X. And let $g(y)$ be the density function of Y and $G(y)$ be the distribution function of Y. Using $Y = X^2$, we obtain:

$$\begin{aligned}G(y) &= P(Y < y) = P(X^2 < y) = P(-\sqrt{y} < X < \sqrt{y}) = F(\sqrt{y}) - F(-\sqrt{y}) \\ &= F(\sqrt{y}) \qquad \longleftarrow \qquad F(-\sqrt{y}) = 0.\end{aligned}$$

Moreover, from the relationship between the density and the distribution functions, we obtain the following:

$$\begin{aligned}g(y) &= \frac{\mathrm{d}G(y)}{\mathrm{d}y} = \frac{\mathrm{d}F(\sqrt{y})}{\mathrm{d}y} = \frac{\mathrm{d}F(x)}{\mathrm{d}x}\frac{\mathrm{d}\sqrt{y}}{\mathrm{d}y} \qquad \longleftarrow \qquad x = \sqrt{y} \\ &= F'(x)\frac{1}{2\sqrt{y}} = f(x)\frac{1}{2\sqrt{y}} = f(\sqrt{y})\frac{1}{2\sqrt{y}} \\ &= (\sqrt{2} - \sqrt{y})\frac{1}{2\sqrt{y}}, \qquad \text{for } 0 < y < 2.\end{aligned}$$

The range of y is obtained as: $0 < x < \sqrt{2} \Longrightarrow 0 < x^2 < 2 \Longrightarrow 0 < y < 2$.

$\boxed{2}$ The continuous type of random variable X has the following density function:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

Answer the following questions.

(1) Compute mean and variance of X.
(2) When $Y = X^2$, compute mean and variance of Y.
(3) When $Z = e^X$, obtain mean and variance of Z.

[Answer]

(1) The definitions of mean and variance are: $\mathrm{E}(X) = \displaystyle\int x f(x)\,\mathrm{d}x$ and $\mathrm{V}(X) = \displaystyle\int (x-\mu)^2 f(x)\,\mathrm{d}x$, where $\mu = \mathrm{E}(X)$. Therefore, mean of X is:

$$\mathrm{E}(X) = \int x f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,\mathrm{d}x = -\frac{1}{\sqrt{2\pi}} \Big[e^{-\frac{1}{2}x^2} \Big]_{-\infty}^{\infty} = 0.$$

In the third equality, we utilize: $\dfrac{\mathrm{d}e^{-\frac{1}{2}x^2}}{\mathrm{d}x} = -xe^{-\frac{1}{2}x^2}$.

Variance of X is:

$$\begin{aligned}\mathrm{V}(X) &= \int (x-\mu)^2 f(x)\,\mathrm{d}x = \int x^2 f(x)\,\mathrm{d}x - \mu^2 = \int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,\mathrm{d}x - \mu^2 \\ &= \Big[-x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \Big]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,\mathrm{d}x - \mu^2 = 1.\end{aligned}$$

In the fourth equality, the following formula is used.

$$\int_a^b h'(x)g(x)\,\mathrm{d}x = \Big[h(x)g(x) \Big]_a^b - \int_a^b h(x)g'(x)\,\mathrm{d}x,$$

where $g(x) = x$ and $h'(x) = x\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$ are set.

And in the first term of the fourth equality, we use:

$$\lim_{x\to\pm\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} = 0.$$

In the second term of the fourth equality, we utilize the property that the integration of the density function is equal to one.

(2) When $Y = X^2$, mean of Y is:

$$\mathrm{E}(Y) = \mathrm{E}(X^2) = \mathrm{V}(X) - \mu_x^2 = 1$$

From (1), note that $\mathrm{V}(X) = 1$ and $\mu_x = \mathrm{E}(X) = 0$.

Variance of Y is:

$$\begin{aligned}
\mathrm{V}(Y) &= \mathrm{E}(Y - \mu_y)^2 \qquad \longleftarrow \qquad \mu_y = \mathrm{E}(Y) = 1 \\
&= \mathrm{E}(Y^2) - \mu_y^2 = \mathrm{E}(X^4) - \mu_y^2 = \int_{-\infty}^{\infty} x^4 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,\mathrm{d}x - \mu_y^2 \\
&= \int_{-\infty}^{\infty} x^3 \cdot x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,\mathrm{d}x - \mu_y^2 \\
&= \left[-x^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \right]_{-\infty}^{\infty} + 3\int_{-\infty}^{\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,\mathrm{d}x - \mu_y^2 \\
&= 3\mathrm{E}(X^2) - \mu_y^2 \qquad \longleftarrow \qquad \mathrm{E}(X^2) = 1,\ \mu_y = 1 \\
&= 2
\end{aligned}$$

In the sixth equality, the following formula on integration is utilized.

$$\int_a^b h'(x)g(x)\,\mathrm{d}x = \Big[h(x)g(x)\Big]_a^b - \int_a^b h(x)g'(x)\,\mathrm{d}x,$$

where $g(x) = x^3$ and $h'(x) = x\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}x^2}$ are set.

In the first term of the sixth equality, we use:

$$\lim_{x\to\pm\infty} x^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} = 0.$$

(3) For $Z = e^X$, mean of Z is:

$$\begin{aligned}
\mathrm{E}(Z) = \mathrm{E}(e^X) &= \int_{-\infty}^{\infty} e^x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,\mathrm{d}x = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x^2-2x)}\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-1)^2+\frac{1}{2}}\,\mathrm{d}x = e^{\frac{1}{2}} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-1)^2}\,\mathrm{d}x = e^{\frac{1}{2}}.
\end{aligned}$$

In the sixth equality, $\dfrac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-1)^2}$ is a normal distribution with mean one and variance one, and accordingly its integration is equal to one.

Variance of Z is:

$$\begin{aligned}
\mathrm{V}(Z) &= \mathrm{E}(Z - \mu_z)^2 \qquad \longleftarrow \qquad \mu_z = \mathrm{E}(Z) = e^{\frac{1}{2}} \\
&= \mathrm{E}(Z^2) - \mu_z^2 = \mathrm{E}(e^{2X}) - \mu_z^2 = \int_{-\infty}^{\infty} e^{2x} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}\,\mathrm{d}x - \mu_z^2 \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x^2-4x)}\,\mathrm{d}x - \mu_z^2 = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-2)^2+2}\,\mathrm{d}x - \mu_z^2 \\
&= e^2 \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}(x-2)^2}\,\mathrm{d}x - \mu_z^2 = e^2 - e.
\end{aligned}$$

The eighth equality comes from the facts that $\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(x-2)^2}$ is a normal distribution with mean two and variance one and that its integration is equal to one.

$\boxed{3}$ The continuous type of random variable X has the following density function:

$$f(x) = \begin{cases} \frac{1}{\lambda}e^{-\frac{x}{\lambda}}, & \text{if } 0 < x, \\ 0, & \text{otherwise.} \end{cases}$$

Answer the following questions.

(1) Compute mean and variance of X.

(2) Derive the moment-generating function of X.

(3) Let $X_1, X_2, \cdots, X_n$ be the random variables, which are mutually independently distributed and have the density function shown above. Prove that the density function of $Y = X_1 + X_2 + \cdots + X_n$ is given by the chi-square distribution with $2n$ degrees of freedom when $\lambda = 2$. Note that the chi-square distribution with m degrees of freedom is given by:

$$f(x) = \begin{cases} \frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})}x^{\frac{m}{2}-1}e^{-\frac{x}{2}}, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

[Answer]

(1) Mean of X is:

$$\begin{aligned} \mathrm{E}(X) &= \int xf(x)\,\mathrm{d}x = \int_0^\infty x\frac{1}{\lambda}e^{-\frac{x}{\lambda}}\,\mathrm{d}x \\ &= \Big[-xe^{-\frac{x}{\lambda}}\Big]_0^\infty + \int_0^\infty e^{-\frac{x}{\lambda}}\,\mathrm{d}x = \Big[-\lambda e^{-\frac{x}{\lambda}}\Big]_0^\infty = \lambda. \end{aligned}$$

In the third equality, the following formula is used:

$$\int_a^b h'(x)g(x)\,\mathrm{d}x = \Big[h(x)g(x)\Big]_a^b - \int_a^b h(x)g'(x)\,\mathrm{d}x.$$

where $g(x) = x$ and $h'(x) = \frac{1}{\lambda}e^{-\frac{x}{\lambda}}$ are set.

And we utilize:

$$\lim_{x\to\infty} xe^{-\frac{x}{\lambda}} = 0, \qquad \lim_{x\to\infty} e^{-\frac{x}{\lambda}} = 0.$$

Variance of X is:

$$\begin{aligned}
\mathrm{V}(X) &= \int (x-\mu)^2 f(x)\,\mathrm{d}x = \int x^2 f(x)\,\mathrm{d}x - \mu^2 \quad \longleftarrow \quad \mu = \mathrm{E}(X) = \lambda \\
&= \int_0^\infty x^2 \frac{1}{\lambda} e^{-\frac{x}{\lambda}}\,\mathrm{d}x - \mu^2 = \Big[-x^2 e^{-\frac{x}{\lambda}}\Big]_0^\infty + 2\int_0^\infty x e^{-\frac{x}{\lambda}}\,\mathrm{d}x - \mu^2 \\
&= \Big[-x^2 e^{-\frac{x}{\lambda}}\Big]_0^\infty + 2\lambda \int_0^\infty x \frac{1}{\lambda} e^{-\frac{x}{\lambda}}\,\mathrm{d}x - \mu^2 \\
&= 2\lambda \mathrm{E}(X) - \mu^2 \quad \longleftarrow \quad \mu = \mathrm{E}(X) = \lambda \\
&= 2\lambda^2 - \lambda^2 = \lambda^2.
\end{aligned}$$

In the third equality, we utilize:

$$\int_a^b h'(x)g(x)\,\mathrm{d}x = \Big[h(x)g(x)\Big]_a^b - \int_a^b h(x)g'(x)\,\mathrm{d}x,$$

where $g(x) = x^2$ and $h'(x) = \dfrac{1}{\lambda} e^{-\frac{x}{\lambda}}$.

In the sixth equality, the following formulas are used:

$$\lim_{x\to\infty} x^2 e^{-\frac{x}{\lambda}} = 0, \qquad \mu = \mathrm{E}(X) = \int_0^\infty x e^{-\frac{x}{\lambda}}\,\mathrm{d}x.$$

(2) The moment-generating function of X is:

$$\begin{aligned}
\phi(\theta) &= \mathrm{E}(e^{\theta X}) = \int e^{\theta x} f(x)\,\mathrm{d}x = \int_0^\infty e^{\theta x} \frac{1}{\lambda} e^{-\frac{x}{\lambda}}\,\mathrm{d}x = \int_0^\infty \frac{1}{\lambda} e^{-(\frac{1}{\lambda}-\theta)x}\,\mathrm{d}x \\
&= \frac{1/\lambda}{1/\lambda - \theta} \int_0^\infty (\frac{1}{\lambda} - \theta) e^{-(\frac{1}{\lambda}-\theta)x}\,\mathrm{d}x = \frac{1}{1-\lambda\theta}.
\end{aligned}$$

In the last equality, since $(\dfrac{1}{\lambda} - \theta)e^{-(\frac{1}{\lambda}-\theta)x}$ is a density function, its integration is one. λ in $f(x)$ is replaced by $\dfrac{1}{\lambda} - \theta$.

(3) We want to show that the moment-generating function of Y is equivalent to that of a chi-square distribution with $2n$ degrees of freedom.

Because X_1, X_2, $\cdots$, X_n are mutually independently distributed, the moment-generating function of X_i, $\phi_i(\theta)$, is:

$$\phi_i(\theta) = \frac{1}{1-2\theta} = \phi(\theta),$$

which corresponds to the case $\lambda = 2$ of (2).

For $\lambda = 2$, the moment-generating function of $Y = X_1 + X_2 + \cdots + X_n$, $\phi_y(\theta)$, is:

$$\begin{aligned}
\phi_y(\theta) &= \mathrm{E}(e^{\theta Y}) = \mathrm{E}(e^{\theta(X_1+X_2+\cdots+X_n)}) = \mathrm{E}(e^{\theta X_1})\mathrm{E}(e^{\theta X_2})\cdots\mathrm{E}(e^{\theta X_n}) \\
&= \phi_1(\theta)\phi_2(\theta)\cdots\phi_n(\theta) = \Big(\phi(\theta)\Big)^n = \Big(\frac{1}{1-2\theta}\Big)^n = \Big(\frac{1}{1-2\theta}\Big)^{\frac{2n}{2}}.
\end{aligned}$$

Therefore, the moment-generating function of Y is:

$$\phi_y(\theta) = \Big(\frac{1}{1-2\theta}\Big)^{\frac{2n}{2}}.$$

A chi-square distribution with m degrees of freedom is given by:

$$f(x) = \frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})} x^{\frac{m}{2}-1} e^{-\frac{x}{2}}, \quad \text{for } x > 0.$$

The moment-generating function of the above density function, $\phi_{\chi^2}(\theta)$, is:

$$\begin{aligned}
\phi_{\chi^2}(\theta) = \mathrm{E}(e^{\theta X}) &= \int_0^\infty e^{\theta x} \frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})} x^{\frac{m}{2}-1} e^{-\frac{x}{2}}\, \mathrm{d}x \\
&= \int_0^\infty \frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})} x^{\frac{m}{2}-1} e^{-\frac{1}{2}(1-2\theta)x}\, \mathrm{d}x \\
&= \int_0^\infty \frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})} \Big(\frac{y}{1-2\theta}\Big)^{\frac{m}{2}-1} e^{-\frac{1}{2}y} \frac{1}{1-2\theta}\, \mathrm{d}x \\
&= \Big(\frac{1}{1-2\theta}\Big)^{\frac{m}{2}-1} \frac{1}{1-2\theta} \int_0^\infty \frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})} y^{\frac{m}{2}-1} e^{-\frac{1}{2}y}\, \mathrm{d}x = \Big(\frac{1}{1-2\theta}\Big)^{\frac{m}{2}}.
\end{aligned}$$

In the fourth equality, use $y = (1-2\theta)x$. In the sixth equality, since the function in the integration corresponds to the chi-square distribution with m degrees of freedom, the integration is one. Thus, $\phi_y(\theta)$ is equivalent to $\phi_{\chi^2}(\theta)$ for $m = 2n$. That is, $\phi_y(\theta)$ is the moment-generating function of a chi square distribution with $2n$ degrees of freedom. Therefore, $Y \sim \chi^2(2n)$.

$\boxed{4}$ The continuous type of random variable X has the following density function:

$$f(x) = \begin{cases} 1, & \text{if } 0 < x < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Answer the following questions.

(1) Compute mean and variance of X.

(2) When $Y = -2\log X$, derive the moment-generating function of Y. Note that the log represents the natural logarithm (i.e., $y = -2\log x$ is equivalent to $x = e^{-\frac{1}{2}y}$).

(3) Let Y_1 and Y_2 be the random variables which have the density function obtained in (2). Suppose that Y_1 is independent of Y_2. When $Z = Y_1 + Y_2$, compute the density function of Z.

[Answer]

(1) Mean of X is:

$$\mathrm{E}(X) = \int x f(x)\,\mathrm{d}x = \int_0^1 x\,\mathrm{d}x = \Big[\frac{1}{2}x^2\Big]_0^1 = \frac{1}{2}.$$

Variance of X is:

$$\begin{aligned}\mathrm{V}(X) &= \int (x-\mu)^2 f(x)\,\mathrm{d}x = \int x^2 f(x)\,\mathrm{d}x - \mu^2 \quad \longleftarrow \quad \mu = \mathrm{E}(X) = \frac{1}{2}\\ &= \int_0^1 x^2\,\mathrm{d}x - \mu^2 = \Big[\frac{1}{3}x^3\Big]_0^1 - \mu^2 = \frac{1}{3} - \Big(\frac{1}{2}\Big)^2 = \frac{1}{12}.\end{aligned}$$

(2) For $Y = -2\log X$, we obtain the moment-generating function of Y, $\phi_y(\theta)$.

$$\begin{aligned}\phi_y(\theta) &= \mathrm{E}(e^{\theta Y}) = \mathrm{E}(e^{-2\theta\log X}) = \mathrm{E}(X^{-2\theta}) = \int x^{-2\theta} f(x)\,\mathrm{d}x\\ &= \int_0^1 x^{-2\theta}\,\mathrm{d}x = \Big[\frac{1}{1-2\theta}x^{1-2\theta}\Big]_0^1 = \frac{1}{1-2\theta}.\end{aligned}$$

(3) Let Y_1 and Y_2 be the random variables which have the density function obtained from (2). And, assume that Y_1 is independent of Y_2. For $Z = Y_1 + Y_2$, we want to have the density function of Z.

The moment-generating function of Z, $\phi_z(\theta)$, is:

$$\begin{aligned}\phi_z(\theta) &= \mathrm{E}(e^{\theta Z}) = \mathrm{E}(e^{\theta(Y_1+Y_2)}) = \mathrm{E}(e^{\theta Y_1})\mathrm{E}(e^{\theta Y_2}) = \big(\phi_y(\theta)\big)^2\\ &= \Big(\frac{1}{1-2\theta}\Big)^2 = \Big(\frac{1}{1-2\theta}\Big)^{\frac{4}{2}},\end{aligned}$$

which is equivalent to the moment-generating function of the chi square distribution with 4 degrees of freedom. Therefore, $Z \sim \chi^2(4)$. Note that the chi-square density function with n degrees of freedom is given by:

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & \text{for } x > 0,\\ 0, & \text{otherwise.}\end{cases}$$

The moment-generating function $\phi(\theta)$ is:

$$\phi(\theta) = \Big(\frac{1}{1-2\theta}\Big)^{\frac{n}{2}}.$$

$\boxed{5}$ The continuous type of random variable X has the following density function:

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & \text{if } x > 0,\\ 0, & \text{otherwise.}\end{cases}$$

Answer the following questions. $\Gamma(a)$ is called the gamma function, defined as:

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x}\,\mathrm{d}x.$$

(1) What are mean and variance of X?

(2) Compute the moment-generating function of X.

[Answer]

(1) For mean:

$$\begin{aligned}
\mathrm{E}(X) &= \int_{-\infty}^{\infty} x f(x)\,\mathrm{d}x = \int_0^\infty x \frac{1}{\Gamma(\frac{n}{2})} 2^{-\frac{n}{2}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}\,\mathrm{d}x \\
&= \frac{2^{-\frac{n}{2}}}{2^{-\frac{n+2}{2}}} \frac{\Gamma(\frac{n+2}{2})}{\Gamma(\frac{n}{2})} \int_0^\infty \frac{1}{\Gamma(\frac{n+2}{2})} 2^{-\frac{n+2}{2}} x^{\frac{n+2}{2}-1} e^{-\frac{x}{2}}\,\mathrm{d}x \\
&= 2\frac{n}{2} \int_0^\infty \frac{1}{\Gamma(\frac{n'}{2})} 2^{-\frac{n'}{2}} x^{\frac{n'}{2}-1} e^{-\frac{x}{2}}\,\mathrm{d}x = n.
\end{aligned}$$

Note that $\Gamma(s+1) = s\Gamma(s)$, $\Gamma(1) = 1$, and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. Using $n' = n+2$, from the property of the density function, we have:

$$\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = \int_0^\infty \frac{1}{\Gamma(\frac{n'}{2})} 2^{-\frac{n'}{2}} x^{\frac{n'}{2}-1} e^{-\frac{x}{2}}\,\mathrm{d}x = 1,$$

which is utilized in the fifth equality.

For variance, from $\mathrm{V}(X) = \mathrm{E}(X^2) - \mu^2$ we compute $\mathrm{E}(X^2)$ as follows:

$$\begin{aligned}
\mathrm{E}(X^2) &= \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x = \int_0^\infty x^2 \frac{1}{\Gamma(\frac{n}{2})} 2^{-\frac{n}{2}} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}\,\mathrm{d}x \\
&= \int_0^\infty \frac{1}{\Gamma(\frac{n}{2})} 2^{-\frac{n}{2}} x^{\frac{n+4}{2}-1} e^{-\frac{x}{2}}\,\mathrm{d}x \\
&= \frac{2^{-\frac{n}{2}}}{2^{-\frac{n+4}{2}}} \frac{\Gamma(\frac{n+4}{2})}{\Gamma(\frac{n}{2})} \int_0^\infty \frac{1}{\Gamma(\frac{n+4}{2})} 2^{-\frac{n+4}{2}} x^{\frac{n+4}{2}-1} e^{-\frac{x}{2}}\,\mathrm{d}x \\
&= 4(\frac{n+2}{2}\frac{n}{2}) \int_0^\infty \frac{1}{\Gamma(\frac{n'}{2})} 2^{-\frac{n'}{2}} x^{\frac{n'}{2}-1} e^{-\frac{x}{2}}\,\mathrm{d}x = n(n+2),
\end{aligned}$$

where $n' = n+4$ is set. Therefore, $\mathrm{V}(X) = n(n+2) - n^2 = 2n$ is obtained.

(2) The moment-generating function of X is:

$$\phi(\theta) = \mathrm{E}(e^{\theta X}) = \int_{-\infty}^{\infty} e^{\theta x} f(x)\,\mathrm{d}x = \int_0^\infty e^{\theta x} \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} \exp(-\frac{x}{2})\,\mathrm{d}x$$

$$
\begin{aligned}
&= \int_0^\infty \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} \exp\left(-\frac{1}{2}(1-2\theta)x\right) \mathrm{d}x \\
&= \int_0^\infty \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} \left(\frac{y}{1-2\theta}\right)^{\frac{n}{2}-1} \exp(-\frac{1}{2}y) \frac{1}{1-2\theta} \mathrm{d}y \\
&= \left(\frac{1}{1-2\theta}\right)^{\frac{n}{2}} \int_0^\infty \frac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} y^{\frac{n}{2}-1} \exp(-\frac{1}{2}y)\, \mathrm{d}y = \left(\frac{1}{1-2\theta}\right)^{\frac{n}{2}}.
\end{aligned}
$$

Use $y = (1-2\theta)x$ in the fifth equality. Note that $\dfrac{\mathrm{d}x}{\mathrm{d}y} = (1-2\theta)^{-1}$. In the seventh equality, the integration corresponds to the chi-square distribution with n degrees of freedom.

$\boxed{6}$ The continuous type of random variables X and Y are mutually independent and assumed to be $X \sim N(0,1)$ and $Y \sim N(0,1)$. Define $U = X/Y$. Answer the following questions. When $X \sim N(0,1)$, the density function of X is represented as:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}.$$

(1) Derive the density function of U.

(2) Prove that the first moment of U does not exist.

[Answer]

(1) The density of U is obtained as follows. The densities of X and Y are:

$$
\begin{aligned}
f(x) &= \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2), \quad -\infty < x < \infty, \\
g(y) &= \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}y^2), \quad -\infty < y < \infty.
\end{aligned}
$$

Since X is independent of Y, the joint density of X and Y is:

$$
\begin{aligned}
h(x,y) = f(x)g(y) &= \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2) \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}y^2) \\
&= \frac{1}{2\pi} \exp(-\frac{1}{2}(x^2+y^2)).
\end{aligned}
$$

Using $u = \dfrac{x}{y}$ and $v = y$, the transformation of the variables is performed. For $x = uv$ and $y = v$, we have the Jacobian:

$$J = \begin{vmatrix} \dfrac{\partial x}{\partial u} & \dfrac{\partial x}{\partial v} \\ \dfrac{\partial y}{\partial u} & \dfrac{\partial y}{\partial v} \end{vmatrix} = \begin{vmatrix} v & u \\ 0 & 1 \end{vmatrix}.$$

Using transformation of variables, the joint density of U and V, $s(u, v)$ is given by:

$$s(u, v) = h(uv, v)|J| = \frac{1}{2\pi}\exp(-\frac{1}{2}v^2(1+u^2))|v|.$$

The marginal density of U is:

$$\begin{aligned} p(u) &= \int s(u, v)\,\mathrm{d}v = \frac{1}{2\pi}\int_{-\infty}^{\infty} |v|\exp(-\frac{1}{2}v^2(1+u^2))\,\mathrm{d}v \\ &= \frac{1}{\pi}\int_0^{\infty} v\exp(-\frac{1}{2}v^2(1+u^2))\,\mathrm{d}v \\ &= \frac{1}{\pi}\left[-\frac{1}{1+u^2}\exp(-\frac{1}{2}v^2(1+u^2))\right]_{v=0}^{\infty} = \frac{1}{\pi(1+u^2)}, \end{aligned}$$

which corresponds to Cauchy distribution.

(2) We prove that the first moment of U is infinity, i.e.,

$$\begin{aligned} \mathrm{E}(U) &= \int uf(u)\,\mathrm{d}u = \int_{-\infty}^{\infty} u\frac{1}{\pi(1+u^2)}\,\mathrm{d}u \\ &= \int_1^{\infty} \frac{1}{2\pi}\frac{1}{x}\,\mathrm{d}x \quad \longleftarrow \quad x = 1+u^2 \text{ is used.} \\ &= \left[\frac{1}{2\pi}\log x\right]_1^{\infty} \quad \longleftarrow \quad \frac{\mathrm{d}\log x}{\mathrm{d}x} = \frac{1}{x} \\ &= \infty. \end{aligned}$$

For $-\infty < u < \infty$, the range of $x = 1+u^2$ is give by $1 < x < \infty$.

$\boxed{7}$ The continuous type of random variables has the following joint density function:

$$f(x, y) = \begin{cases} x+y, & \text{if } 0 < x < 1 \text{ and } 0 < y < 1, \\ 0, & \text{otherwise.} \end{cases}$$

Answer the following questions.

(1) Compute the expectation of XY.

(2) Obtain the correlation coefficient between X and Y.

(3) What is the marginal density function of X?

[Answer]

(1) The expectation of XY is:

$$\begin{aligned} \mathrm{E}(XY) &= \int_0^1\int_0^1 xyf(x, y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^1\int_0^1 xy(x+y)\,\mathrm{d}x\,\mathrm{d}y \\ &= \int_0^1 \left[\frac{1}{3}yx^3 + \frac{1}{2}y^2x^2\right]_0^1 \mathrm{d}y = \int_0^1 (\frac{1}{3}y + \frac{1}{2}y^2)\,\mathrm{d}y \\ &= \left[\frac{1}{6}y^2 + \frac{1}{6}y^3\right]_0^1 = \frac{1}{3}. \end{aligned}$$

(2) We want to obtain the correlation coefficient between X and Y, which is represented as: $\rho = \mathrm{Cov}(X, Y)/\sqrt{\mathrm{V}(X)\mathrm{V}(Y)}$. Therefore, $\mathrm{E}(X)$, $\mathrm{E}(Y)$, $\mathrm{V}(X)$, $\mathrm{V}(Y)$ and $\mathrm{Cov}(X, Y)$ have to be computed.

$\mathrm{E}(X)$ is:

$$\begin{aligned}
\mathrm{E}(X) &= \int_0^1 \int_0^1 x f(x,y)\,\mathrm{d}x\,\mathrm{d}y = \int_0^1 \int_0^1 x(x+y)\,\mathrm{d}x\,\mathrm{d}y \\
&= \int_0^1 \left[\frac{1}{3}x^3 + \frac{1}{2}yx^2\right]_0^1 \mathrm{d}y = \int_0^1 (\frac{1}{3} + \frac{1}{2}y)\,\mathrm{d}y \\
&= \left[\frac{1}{3}y + \frac{1}{4}y^2\right]_0^1 = \frac{7}{12}.
\end{aligned}$$

In the case where x and y are exchangeable, the functional form of $f(x,y)$ is unchanged. Therefore, $\mathrm{E}(Y)$ is:

$$\mathrm{E}(Y) = \mathrm{E}(X) = \frac{7}{12}.$$

For $\mathrm{V}(X)$,

$$\begin{aligned}
\mathrm{V}(X) &= \mathrm{E}\Big((X-\mu)^2\Big) \quad \longleftarrow \quad \mu = \mathrm{E}(X) = \frac{7}{12} \\
&= \mathrm{E}(X^2) - \mu^2 = \int_0^1 \int_0^1 x^2 f(x,y)\,\mathrm{d}x\,\mathrm{d}y - \mu^2 \\
&= \int_0^1 \int_0^1 x^2(x+y)\,\mathrm{d}x\,\mathrm{d}y - \mu^2 = \int_0^1 \left[\frac{1}{4}x^4 + \frac{1}{3}yx^3\right]_0^1 \mathrm{d}y - \mu^2 \\
&= \int_0^1 (\frac{1}{4} + \frac{1}{3}y)\,\mathrm{d}y - \mu^2 = \left[\frac{1}{4}y + \frac{1}{6}y^2\right]_0^1 - \mu^2 \\
&= \frac{5}{12} - \Big(\frac{7}{12}\Big)^2 = \frac{11}{144}.
\end{aligned}$$

For $\mathrm{V}(Y)$,

$$\mathrm{V}(Y) = \mathrm{V}(X) = \frac{11}{144}.$$

For $\mathrm{Cov}(X, Y)$,

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathrm{E}\Big((X-\mu_x)(Y-\mu_y)\Big) = \mathrm{E}(XY) - \mu_x\mu_y \\
&= \frac{1}{3} - \frac{7}{12}\frac{7}{12} = -\frac{1}{144},
\end{aligned}$$

where

$$\mu_x = \mathrm{E}(X) = \frac{7}{12}, \quad \mu_y = \mathrm{E}(Y) = \frac{7}{12}.$$

Therefore, ρ is:

$$\rho = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{V}(X)\mathrm{V}(Y)}} = \frac{-1/144}{\sqrt{(11/144)(11/144)}} = -\frac{1}{11}.$$

(3) The marginal density function of X, $f_x(x)$, is:

$$f_x(x) = \int f(x, y)\,\mathrm{d}y = \int_0^1 (x + y)\,\mathrm{d}y = \Big[xy + \frac{1}{2}y^2\Big]_{y=0}^{1} = x + \frac{1}{2},$$

for $0 < x < 1$.

$\boxed{8}$ The discrete type of random variable X has the following density function:

$$f(x) = \frac{e^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, \cdots.$$

Answer the following questions.

(1) Prove $\sum_{x=0}^{\infty} f(x) = 1$.

(2) Compute the moment-generating function of X.

(3) From the moment-generating function, obtain mean and variance of X.

[Answer]

(1) We can show $\sum_{x=0}^{\infty} f(x) = 1$ as:

$$\sum_{x=0}^{\infty} f(x) = \sum_{x=0}^{\infty} e^{-\lambda}\frac{\lambda^x}{x!} = e^{-\lambda}\sum_{x=0}^{\infty}\frac{\lambda^x}{x!} = e^{-\lambda}e^{\lambda} = 1.$$

Note that $e^x = \sum_{k=0}^{\infty}\frac{x^k}{k!}$, because we have $f^{(k)}(x) = e^x$ for $f(x) = e^x$. As shown in Appendix 1.3, the formula of Taylor series expansion is:

$$f(x) = \sum_{k=0}^{\infty}\frac{1}{k!}f^{(k)}(x_0)(x - x_0)^k.$$

The Taylor series expansion around $x = 0$ is:

$$f(x) = \sum_{k=0}^{\infty}\frac{1}{k!}f^{(k)}(0)x^k = \sum_{k=0}^{\infty}\frac{1}{k!}x^k = \sum_{k=0}^{\infty}\frac{x^k}{k!}.$$

Here, replace x by λ and k by x.

(2) The moment-generating function of X is:

$$\begin{aligned}\phi(\theta) &= \mathrm{E}(e^{\theta X}) = \sum_{x=0}^{\infty} e^{\theta x} f(x) = \sum_{x=0}^{\infty} e^{\theta x} e^{-\lambda} \frac{\lambda^x}{x!} = \sum_{x=0}^{\infty} e^{-\lambda} \frac{(e^{\theta}\lambda)^x}{x!} \\ &= e^{-\lambda} \exp(e^{\theta}\lambda) \sum_{x=0}^{\infty} \exp(-e^{\theta}\lambda) \frac{(e^{\theta}\lambda)^x}{x!} = e^{-\lambda} \exp(e^{\theta}\lambda) \sum_{x=0}^{\infty} e^{-\lambda'} \frac{\lambda'^x}{x!} \\ &= \exp(-\lambda)\exp(e^{\theta}\lambda) = \exp\big(\lambda(e^{\theta}-1)\big).\end{aligned}$$

Note that $\lambda' = \exp(e^{\theta}\lambda)$.

(3) Based on the moment-generating function, we obtain mean and variance of X.

For mean, because of $\phi(\theta) = \exp\big(\lambda(e^{\theta}-1)\big)$, $\phi'(\theta) = \lambda e^{\theta} \exp\big(\lambda(e^{\theta}-1)\big)$ and $\mathrm{E}(X) = \phi'(0)$, we obtain:

$$\mathrm{E}(X) = \phi'(0) = \lambda.$$

For variance, from $\mathrm{V}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2$, we obtain $\mathrm{E}(X^2)$. Note that $\mathrm{E}(X^2) = \phi''(0)$ and $\phi''(\theta) = (1+\lambda e^{\theta})\lambda e^{\theta} \exp\big(\lambda(e^{\theta}-1)\big)$. Therefore,

$$\mathrm{V}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = \phi''(0) - (\phi'(0))^2 = (1+\lambda)\lambda - \lambda^2 = \lambda.$$

$\boxed{9}$ $X_1, X_2, \cdots, X_n$ are mutually independently and normally distributed with mean μ and variance σ^2, where the density function is given by:

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

Then, answer the following questions.

(1) Obtain the maximum likelihood estimators of mean μ and variance σ^2.

(2) Check whether the maximum likelihood estimator of σ^2 is unbiased. If it is not unbiased, obtain an unbiased estimator of σ^2. (Hint: use the maximum likelihood estimator.)

(3) We want to test the null hypothesis $H_0: \mu = \mu_0$ by the likelihood ratio test. Obtain the test statistic and explain the testing procedure.

[Answer]

(1) The joint density is:

$$\begin{aligned}f(x_1, x_2, \cdots, x_n; \mu, \sigma^2) &= \prod_{i=1}^{n} f(x_i; \mu, \sigma^2) \\ &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu)^2\right) \\ &= (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right) = l(\mu, \sigma^2).\end{aligned}$$

Taking the logarithm, we have:

$$\log l(\mu,\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

The derivatives of the log-likelihood function $\log l(\mu, \sigma^2)$ with respect to μ and σ^2 are set to be zero.

$$\frac{\partial \log l(\mu,\sigma^2)}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(x_i-\mu) = 0,$$

$$\frac{\partial \log l(\mu,\sigma^2)}{\partial\sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu)^2 = 0.$$

Solving the two equations, we have the solution of (μ, σ^2), denoted by $(\hat{\mu}, \hat{\sigma}^2)$:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x},$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\mu)^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\overline{x})^2.$$

Therefore, the maximum likelihood estimators of μ and σ^2, $(\hat{\mu}, \hat{\sigma}^2)$, are as follows:

$$\overline{X}, \quad S^{**2} = \frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2.$$

(2) Take the expectation to check whether S^{**2} is unbiased.

$$\begin{aligned}
\mathrm{E}(S^{**2}) &= \mathrm{E}\Big(\frac{1}{n}\sum_{i=1}^{n}(X_i-\overline{X})^2\Big) = \frac{1}{n}\mathrm{E}\Big(\sum_{i=1}^{n}(X_i-\overline{X})^2\Big) \\
&= \frac{1}{n}\mathrm{E}\Big(\sum_{i=1}^{n}((X_i-\mu)-(\overline{X}-\mu))^2\Big) \\
&= \frac{1}{n}\mathrm{E}\Big(\sum_{i=1}^{n}((X_i-\mu)^2 - 2(X_i-\mu)(\overline{X}-\mu) + (\overline{X}-\mu)^2)\Big) \\
&= \frac{1}{n}\mathrm{E}\Big(\sum_{i=1}^{n}(X_i-\mu)^2 - 2(\overline{X}-\mu)\sum_{i=1}^{n}(X_i-\mu) + n(\overline{X}-\mu)^2\Big) \\
&= \frac{1}{n}\mathrm{E}\Big(\sum_{i=1}^{n}(X_i-\mu)^2 - 2n(\overline{X}-\mu)^2 + n(\overline{X}-\mu)^2\Big) \\
&= \frac{1}{n}\mathrm{E}\Big(\sum_{i=1}^{n}(X_i-\mu)^2 - n(\overline{X}-\mu)^2\Big)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{n}\mathrm{E}\Big(\sum_{i=1}^{n}(X_i-\mu)^2\Big) - \frac{1}{n}\mathrm{E}\Big(n(\overline{X}-\mu)^2\Big) \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}\Big((X_i-\mu)^2\Big) - \mathrm{E}\Big((\overline{X}-\mu)^2\Big) \\
&= \frac{1}{n}\sum_{i=1}^{n}\mathrm{V}(X_i) - \mathrm{V}(\overline{X}) = \frac{1}{n}\sum_{i=1}^{n}\sigma^2 - \frac{\sigma^2}{n} \\
&= \sigma^2 - \frac{1}{n}\sigma^2 = \frac{n-1}{n}\sigma^2 \neq \sigma^2.
\end{aligned}
$$

Therefore, S^{**2} is not unbiased. Based on S^{**2}, we obtain the unbiased estimator of σ^2. Multiplying $n/(n-1)$ on both sides of $\mathrm{E}(S^{**2}) = \sigma^2(n-1)/n$, we obtain:

$$\frac{n}{n-1}\mathrm{E}(S^{**2}) = \sigma^2.$$

Therefore, the unbiased estimator of σ^2 is:

$$\frac{n}{n-1}S^{**2} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i-\overline{X})^2 = S^2.$$

(3) The likelihood ratio is defined as:

$$\lambda = \frac{\max_{\sigma^2} l(\mu_0,\sigma^2)}{\max_{\mu,\sigma^2} l(\mu,\sigma^2)} = \frac{l(\mu_0,\widetilde{\sigma}^2)}{l(\hat{\mu},\hat{\sigma}^2)}.$$

Since the number of restriction is one, we have:

$$-2\log\lambda \longrightarrow \chi^2(1).$$

$l(\mu,\sigma^2)$ is given by:

$$l(\mu,\sigma^2) = (2\pi\sigma^2)^{-n/2}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right).$$

Taking the logarithm, $\log l(\mu,\sigma^2)$ is:

$$\log l(\mu,\sigma^2) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2.$$

On the numerator, under the restriction $\mu = \mu_0$, $\log l(\mu_0,\sigma^2)$ is maximized with respect to σ^2 as follows:

$$\frac{\partial \log l(\mu_0,\sigma^2)}{\partial\sigma^2} = -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(x_i-\mu_0)^2 = 0.$$

This solution of σ^2 is $\widetilde{\sigma}^2$, which is represented as:

$$\widetilde{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_0)^2.$$

Then, $l(\mu_0, \widetilde{\sigma}^2)$ is:

$$l(\mu_0, \widetilde{\sigma}^2) = (2\pi\widetilde{\sigma}^2)^{-n/2} \exp\left(-\frac{1}{2\widetilde{\sigma}^2}\sum_{i=1}^{n}(x_i - \mu_0)^2\right) = (2\pi\widetilde{\sigma}^2)^{-n/2} \exp\left(-\frac{n}{2}\right).$$

On the denominator, from the question (1), we have:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} x_i, \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \hat{\mu})^2.$$

Therefore, $l(\hat{\mu}, \hat{\sigma}^2)$ is:

$$l(\hat{\mu}, \hat{\sigma}^2) = (2\pi\hat{\sigma}^2)^{-n/2} \exp\left(-\frac{1}{2\hat{\sigma}^2}\sum_{i=1}^{n}(x_i - \hat{\mu})^2\right) = (2\pi\hat{\sigma}^2)^{-n/2} \exp\left(-\frac{n}{2}\right).$$

The likelihood ratio is:

$$\lambda = \frac{\max_{\sigma^2} l(\mu_0, \sigma^2)}{\max_{\mu,\sigma^2} l(\mu, \sigma^2)} = \frac{l(\mu_0, \widetilde{\sigma}^2)}{l(\hat{\mu}, \hat{\sigma}^2)} = \frac{(2\pi\widetilde{\sigma}^2)^{-n/2}\exp(-n/2)}{(2\pi\hat{\sigma}^2)^{-n/2}\exp(-n/2)} = \left(\frac{\widetilde{\sigma}^2}{\hat{\sigma}^2}\right)^{-n/2}.$$

As n goes to infinity, we obtain:

$$-2\log\lambda = n(\log\widetilde{\sigma}^2 - \log\hat{\sigma}^2) \sim \chi^2(1).$$

When $-2\log\lambda > \chi^2_\alpha(1)$, the null hypothesis $H_0 : \mu = \mu_0$ is rejected by the significance level α, where $\chi^2_\alpha(1)$ denotes the $100 \times \alpha$ percent point of the Chi-square distribution with one degree of freedom.

$\boxed{10}$ Answer the following questions.

(1) The discrete type of random variable X is assumed to be Bernoulli. The Bernoulli distribution is given by:

$$f(x) = p^x(1-p)^{1-x}, \qquad x = 0, 1.$$

Let X_1, X_2, $\cdots$,X_n be random variables drawn from the Bernoulli trials. Compute the maximum likelihood estimator of p.

(2) Let Y be a random variable from a binomial distribution, denoted by $f(y)$, which is represented as:

$$f(y) = {}_nC_y p^y (1-p)^{n-y}, \qquad y = 0, 1, 2, \cdots, n.$$

Then, prove that Y/n goes to p as n is large.

(3) For the random variable Y in the question (2), Let us define:

$$Z_n \equiv \frac{Y - np}{\sqrt{np(1-p)}}.$$

Then, Z_n goes to a standard normal distribution as n is large.

(4) The continuous type of random variable X has the following density function:

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{n}{2}}\Gamma(\frac{n}{2})} x^{\frac{n}{2}-1} e^{-\frac{x}{2}}, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

where $\Gamma(a)$ denotes the Gamma function, i.e.,

$$\Gamma(a) = \int_0^\infty x^{a-1} e^{-x}\,\mathrm{d}x.$$

Then, show that X/n approaches one when $n \longrightarrow \infty$.

[Answer]

(1) When X is a Bernoulli random variable, the probability function of X is given by:

$$f(x; p) = p^x (1-p)^{1-x}, \qquad x = 0, 1.$$

The joint probability function of X_1, X_2, $\cdots$,X_n is:

$$\begin{aligned} f(x_1, x_2, \cdots, x_n; p) &= \prod_{i=1}^n f(x_i; p) = \prod_{i=1}^n p^{x_i}(1-p)^{1-x_i} \\ &= p^{\sum_i x_i}(1-p)^{n-\sum_i x_i} = l(p). \end{aligned}$$

Take the logarithm of $l(p)$.

$$\log l(p) = (\sum_i x_i)\log(p) + (n - \sum_i x_i)\log(1-p).$$

The derivative of the log-likelihood function $\log l(p)$ with respect to p is set to be zero.

$$\frac{\mathrm{d}\log l(p)}{\mathrm{d}p} = \frac{\sum_i x_i}{p} - \frac{n - \sum_i x_i}{1-p} = \frac{\sum_i x_i - np}{p(1-p)} = 0.$$

Solving the above equation, we have:

$$p = \frac{1}{n}\sum_{i=1}^{n} x_i = \overline{x}.$$

Therefore, the maximum likelihood estimator of p is:

$$\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i = \overline{X}.$$

(2) Mean and variance of Y are:

$$\mathrm{E}(Y) = np, \qquad \mathrm{V}(Y) = np(1-p).$$

Therefore, we have:

$$\mathrm{E}(\frac{Y}{n}) = \frac{1}{n}\mathrm{E}(Y) = p, \qquad \mathrm{V}(\frac{Y}{n}) = \frac{1}{n^2}\mathrm{V}(Y) = \frac{p(1-p)}{n}.$$

Chebyshev's inequality indicates that for a random variable X and $g(x) \geq 0$ we have:

$$P(g(X) \geq k) \leq \frac{\mathrm{E}(g(X))}{k},$$

where $k > 0$.

Here, when $g(X) = (X - \mathrm{E}(X))^2$ and $k = \epsilon^2$ are set, we can rewrite as:

$$P(|X - \mathrm{E}(X)| \geq \epsilon) \leq \frac{\mathrm{V}(X)}{\epsilon^2},$$

where $\epsilon > 0$.

Replacing X by $\frac{Y}{n}$, we apply Chebyshev's inequality.

$$P(|\frac{Y}{n} - \mathrm{E}(\frac{Y}{n})| \geq \epsilon) \leq \frac{\mathrm{V}(\frac{Y}{n})}{\epsilon^2}.$$

That is, as $n \longrightarrow \infty$,

$$P(|\frac{Y}{n} - p| \geq \epsilon) \leq \frac{p(1-p)}{n\epsilon^2} \longrightarrow 0.$$

Therefore, we obtain:

$$\frac{Y}{n} \longrightarrow p.$$

(3) Let X_1, X_2, $\cdots$, X_n be Bernoulli random variables, where $P(X_i = x) = p^x(1-p)^{1-x}$ for $x = 0, 1$. Define $Y = X_1 + X_2 + \cdots + X_n$. Because Y has a binomial distribution, Y/n is taken as the sample mean from X_1, X_2, $\cdots$, X_n, i.e., $Y/n = (1/n)\sum_{i=1}^n X_i$. Therefore, using $\mathrm{E}(Y/n) = p$ and $\mathrm{V}(Y/n) = p(1-p)/n$, by the central limit theorem, as $n \longrightarrow \infty$, we have:

$$\frac{Y/n - p}{\sqrt{p(1-p)/n}} \longrightarrow N(0,1).$$

Moreover,

$$Z_n \equiv \frac{Y - np}{\sqrt{np(1-p)}} = \frac{Y/n - p}{\sqrt{p(1-p)/n}}.$$

Therefore,

$$Z_n \longrightarrow N(0,1).$$

(4) When $X \sim \chi^2(n)$, we have $\mathrm{E}(X) = n$ and $\mathrm{V}(X) = 2n$. Therefore, $\mathrm{E}(X/n) = 1$ and $\mathrm{V}(X/n) = 2/n$.

Apply Chebyshev's inequality. Then, we have:

$$P(|\frac{X}{n} - \mathrm{E}(\frac{X}{n})| \geq \epsilon) \leq \frac{\mathrm{V}(\frac{X}{n})}{\epsilon^2},$$

where $\epsilon > 0$. That is, as $n \longrightarrow \infty$, we have:

$$P(|\frac{X}{n} - 1| \geq \epsilon) \leq \frac{2}{n\epsilon^2} \longrightarrow 0.$$

Therefore,

$$\frac{X}{n} \longrightarrow 1.$$

$\boxed{11}$ Consider n random variables X_1, X_2, $\cdots$, X_n, which are mutually independently and exponentially distributed. Note that the exponential distribution is given by:

$$f(x) = \lambda e^{-\lambda x}, \qquad x > 0.$$

Then, answer the following questions.

(1) Let $\hat{\lambda}$ be the maximum likelihood estimator of λ. Obtain $\hat{\lambda}$.

(2) When n is large enough, obtain mean and variance of $\hat{\lambda}$.

[Answer]

(1) Since $X_1, \cdots, X_n$ are mutually independently and exponentially distributed, the likelihood function $l(\lambda)$ is written as:

$$l(\lambda) = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda \sum x_i}.$$

The log-likelihood function is:

$$\log l(\lambda) = n \log(\lambda) - \lambda \sum_{i=1}^{n} x_i.$$

We want the λ which maximizes $\log l(\lambda)$. Solving the following equation:

$$\frac{\mathrm{d} \log l(\lambda)}{\mathrm{d}\lambda} = \frac{n}{\lambda} - \sum_{i=1}^{n} x_i = 0,$$

and replacing x_i by X_i, the maximum likelihood estimator of λ, denoted by $\hat{\lambda}$, is:

$$\hat{\lambda} = \frac{n}{\sum_{i=1}^{n} X_i}.$$

(2) $X_1, X_2, \cdots, X_n$ are mutually independent. Let $f(x_i; \lambda)$ be the density function of X_i. For the maximum likelihood estimator of λ, i.e., $\hat{\lambda}_n$, as $n \longrightarrow \infty$, we have the following property:

$$\sqrt{n}(\hat{\lambda}_n - \lambda) \longrightarrow N\Big(0, \sigma^2(\lambda)\Big),$$

where

$$\sigma^2(\lambda) = \frac{1}{\mathrm{E}\left[\left(\dfrac{\mathrm{d} \log f(X; \lambda)}{\mathrm{d}\lambda}\right)^2\right]}.$$

Therefore, we obtain $\sigma^2(\lambda)$. The expectation in $\sigma^2(\hat{\lambda}_n)$ is:

$$\begin{aligned}\mathrm{E}\left[\left(\frac{\mathrm{d} \log f(X; \lambda)}{\mathrm{d}\lambda}\right)^2\right] &= \mathrm{E}\left[\left(\frac{1}{\lambda} - X\right)^2\right] = \mathrm{E}\left(\frac{1}{\lambda^2} - \frac{2}{\lambda}X + X^2\right) \\ &= \frac{1}{\lambda^2} - \frac{2}{\lambda}\mathrm{E}(X) + \mathrm{E}(X^2) = \frac{1}{\lambda^2},\end{aligned}$$

where $\mathrm{E}(X)$ and $\mathrm{E}(X^2)$ are:

$$\mathrm{E}(X) = \frac{1}{\lambda}, \qquad \mathrm{E}(X^2) = \frac{2}{\lambda^2}.$$

Therefore, we have:

$$\sigma^2(\lambda) = \frac{1}{\mathrm{E}\left[\left(\dfrac{\mathrm{d} \log f(X; \lambda)}{\mathrm{d}\lambda}\right)^2\right]} = \lambda^2.$$

As n is large, $\hat{\lambda}_n$ approximately has the following distribution:

$$\hat{\lambda}_n \sim N(\lambda, \frac{\lambda^2}{n}).$$

Thus, as n goes to infinity, mean and variance are given by λ and λ^2/n.

$\boxed{12}$ The n random variables $X_1, X_2, \cdots, X_n$ are mutually independently distributed with mean μ and variance σ^2. Consider the following two estimators of μ:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad \widetilde{X} = \frac{1}{2}(X_1 + X_n).$$

Then, answer the following questions.

(1) Is $\overline{X}$ unbiased? How about $\widetilde{X}$?
(2) Which is more efficient, $\overline{X}$ or $\widetilde{X}$?
(3) Examine whether $\overline{X}$ and $\widetilde{X}$ are consistent.

[Answer]

(1) We check whether $\overline{X}$ and $\widetilde{X}$ are unbiased.

$$\mathrm{E}(\overline{X}) = \mathrm{E}(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n}\mathrm{E}(\sum_{i=1}^{n} X_i) = \frac{1}{n}\sum_{i=1}^{n}\mathrm{E}(X_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu,$$
$$\mathrm{E}(\widetilde{X}) = \frac{1}{2}\Big(\mathrm{E}(X_1) + \mathrm{E}(X_n)\Big) = \frac{1}{2}(\mu + \mu) = \mu.$$

Thus, both are unbiased.

(2) We examine which is more efficient, $\overline{X}$ or $\widetilde{X}$.

$$\mathrm{V}(\overline{X}) = \mathrm{V}(\frac{1}{n}\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\mathrm{V}(\sum_{i=1}^{n} X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{V}(X_i) = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{\sigma^2}{n},$$
$$\mathrm{V}(\widetilde{X}) = \frac{1}{4}\Big(\mathrm{V}(X_1) + \mathrm{V}(X_n)\Big) = \frac{1}{4}(\sigma^2 + \sigma^2) = \frac{\sigma^2}{2}.$$

Therefore, because of $\mathrm{V}(\overline{X}) < \mathrm{V}(\widetilde{X})$, $\overline{X}$ is more efficient than $\widetilde{X}$ when $n > 2$.

(3) We check if $\overline{X}$ and $\widetilde{X}$ are consistent. Apply Chebyshev's inequality. For $\overline{X}$,

$$P(|\overline{X} - \mathrm{E}(\overline{X})| \geq \epsilon) \leq \frac{\mathrm{V}(\overline{X})}{\epsilon^2},$$

where $\epsilon > 0$. That is, when $n \longrightarrow \infty$, we have:

$$P(|\overline{X} - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \longrightarrow 0.$$

Therefore, we obtain:

$$\overline{X} \longrightarrow \mu.$$

Next, for $\widetilde{X}$, we have:

$$P(|\widetilde{X} - \mathrm{E}(\widetilde{X})| \geq \epsilon) \leq \frac{\mathrm{V}(\widetilde{X})}{\epsilon^2},$$

where $\epsilon > 0$. That is, when $n \longrightarrow \infty$, the following equation is obtained:

$$P(|\widetilde{X} - \mu| \geq \epsilon) \leq \frac{\sigma^2}{2\epsilon^2} \not\longrightarrow 0.$$

$\overline{X}$ is a consistent estimator of μ, but $\widetilde{X}$ is not consistent.

$\boxed{13}$ The 9 random samples:

21 23 32 20 36 27 26 28 30

which are obtained from the normal population $N(\mu, \sigma^2)$. Then, answer the following questions.

(1) Obtain the unbiased estimates of μ and σ^2.

(2) Obtain both 90 and 95 percent confidence intervals for μ.

(3) Test the null hypothesis $H_0: \mu = 24$ and the alternative hypothesis $H_1: \mu > 24$ by the significance level 0.10. How about 0.05?

[Answer]

(1) The unbiased estimators of μ and σ^2, denoted by $\overline{X}$ and S^2, are given by:

$$\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2.$$

The unbiased estimates of μ and σ^2 are:

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i, \quad s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2.$$

Therefore,

$$\overline{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{9}(21 + 23 + 32 + 20 + 36 + 27 + 26 + 28 + 30) = 27,$$

$$\begin{aligned} s^2 &= \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2 \\ &= \frac{1}{8}\Big((21-27)^2 + (23-27)^2 + (32-27)^2 + (20-27)^2 \\ &\quad +(36-27)^2 + (27-27)^2 + (26-27)^2 + (28-27)^2 + (30-27)^2\Big) \\ &= \frac{1}{8}(36 + 16 + 25 + 49 + 81 + 0 + 1 + 1 + 9) = 27.25. \end{aligned}$$

(2) We obtain the confidence intervals of μ. The following sample distribution is utilized:

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1).$$

Therefore,

$$P\Big(\Big|\frac{\overline{X}-\mu}{S/\sqrt{n}}\Big| < t_{\alpha/2}(n-1)\Big) = 1-\alpha,$$

where $t_{\alpha/2}(n-1)$ denotes the $100\times\alpha/2$ percent point of the t distribution, which is obtained given probability α and $n-1$ degrees of freedom. Therefore, we have:

$$P\Big(\overline{X} - t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2}(n-1)\frac{S}{\sqrt{n}}\Big) = 1-\alpha.$$

Replacing $\overline{X}$ and S^2 by $\overline{x}$ and s^2, the $100\times(1-\alpha)$ percent confidence interval of μ is:

$$\Big(\overline{x} - t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}, \overline{x} + t_{\alpha/2}(n-1)\frac{s}{\sqrt{n}}\Big).$$

Since $\overline{x} = 27$, $s^2 = 27.25$, $n = 9$, $t_{0.05}(8) = 1.860$ and $t_{0.025}(8) = 2.306$, the 90 percent confidence interval of μ is:

$$(27 - 1.860\sqrt{\frac{27.25}{9}}, 27 + 1.860\sqrt{\frac{27.25}{9}}) = (23.76, 30.24),$$

and the 95 percent confidence interval of μ is:

$$(27 - 2.306\sqrt{\frac{27.25}{9}}, 27 + 2.306\sqrt{\frac{27.25}{9}}) = (22.99, 31.01).$$

(3) We test the null hypothesis $H_0: \mu = 24$ and the alternative hypothesis $H_1: \mu > 24$ by the significance levels 0.10 and 0.05. The distribution of $\overline{X}$ is:

$$\frac{\overline{X}-\mu}{S/\sqrt{n}} \sim t(n-1).$$

Therefore, under the null hypothesis $H_0: \mu = \mu_0$, we obtain

$$\frac{\overline{X}-\mu_0}{S/\sqrt{n}} \sim t(n-1).$$

Note that μ is replaced by μ_0. For the alternative hypothesis $H_1: \mu > \mu_0$, since we have:

$$P\Big(\frac{\overline{X}-\mu_0}{S/\sqrt{n}} > t_\alpha(n-1)\Big) = \alpha,$$

we reject the null hypothesis $H_0 : \mu = \mu_0$ by the significance level α when we have:

$$\frac{\overline{x} - \mu_0}{s/\sqrt{n}} > t_\alpha(n-1).$$

Substitute $\overline{x} = 27$, $s^2 = 27.25$, $\mu_0 = 24$, $n = 9$, $t_{0.10}(8) = 1.397$ and $t_{0.05}(8) = 1.860$ into the above formula. Then, we obtain:

$$\frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{27-24}{\sqrt{27.25/9}} = 1.724 > t_{0.10}(8) = 1.397.$$

Therefore, we reject the null hypothesis $H_0 : \mu = 24$ by the significance level $\alpha = 0.10$. And we obtain:

$$\frac{\overline{x} - \mu_0}{s/\sqrt{n}} = \frac{27-24}{\sqrt{27.25/9}} = 1.724 < t_{0.05}(8) = 1.860.$$

Therefore, the null hypothesis $H_0 : \mu = 24$ is accepted by the significance level $\alpha = 0.05$.

$\boxed{14}$ The 16 samples $X_1, X_2, \cdots, X_{16}$ are randomly drawn from the normal population with mean μ and known variance $\sigma^2 = 2^2$. The sample average is given by $\overline{x} = 36$. Then, answer the following questions.

(1) Obtain the 95 percent confidence interval for μ.

(2) Test the null hypothesis $H_0 : \mu = 35$ and the alternative hypothesis $H_1 : \mu = 36.5$ by the significance level 0.05.

(3) Compute the power of the test in the above question (2).

[Answer]

(1) We obtain the 95 percent confidence interval of μ. The distribution of $\overline{X}$ is:

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Therefore,

$$P\Big(\Big|\frac{\overline{X} - \mu}{\sigma/\sqrt{n}}\Big| < z_{\alpha/2}\Big) = 1 - \alpha,$$

where $z_{\alpha/2}$ denotes the $100\times\frac{\alpha}{2}$ percent point, which is obtained given probability α. Therefore,

$$P\Big(\overline{X} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\Big) = 1 - \alpha.$$

Replacing $\overline{X}$ by $\overline{x}$, the $100(1-\alpha)$ percent confidence interval of μ is:

$$\Big(\overline{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}}, \overline{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\Big).$$

Substituting $\overline{x} = 36$, $\sigma^2 = 2^2$, $n = 16$ and $z_{0.025} = 1.960$, the $100 \times (1-\alpha)$ percent confidence interval of μ is:

$$(36 - 1.960\frac{2}{\sqrt{16}}, 36 + 1.960\frac{2}{\sqrt{16}}) = (35.02, 36.98).$$

(2) We test the null hypothesis $H_0 : \mu = 35$ and the alternative hypothesis $H_1 : \mu = 36.5$ by the significance level 0.05. The distribution of $\overline{X}$ is:

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1).$$

Under the null hypothesis $H_0 : \mu = \mu_0$,

$$\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \sim N(0,1).$$

For the alternative hypothesis $H_1 : \mu > \mu_0$, we obtain:

$$P\Big(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha\Big) = \alpha.$$

If we have:

$$\frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha,$$

the null hypothesis $H_0 : \mu = \mu_0$ is rejected by the significance level α. Substituting $\overline{x} = 36$, $\sigma^2 = 2^2$, $n = 16$ and $z_{0.05} = 1.645$, we obtain:

$$\frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{36 - 35}{2/\sqrt{16}} = 2 > z_\alpha = 1.645.$$

The null hypothesis $H_0 : \mu = 35$ is rejected by the significance level $\alpha = 0.05$.

(3) We compute the power of the test in the question (2). The power of the test is the probability which rejects the null hypothesis under the alternative hypothesis. That is, under the null hypothesis $H_0 : \mu = \mu_0$, the region which rejects the null hypothesis is: $\overline{X} > \mu_0 + z_\alpha\sigma/\sqrt{n}$, because

$$P\Big(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} > z_\alpha\Big) = \alpha.$$

We compute the probability which rejects the null hypothesis under the alternative hypothesis $H_1 : \mu = \mu_1$. That is, under the alternative hypothesis $H_1 : \mu = \mu_1$, the following probability is known as the power of the test:

$$P\Big(\overline{X} > \mu_0 + z_\alpha \sigma/\sqrt{n}\Big).$$

Under the alternative hypothesis $H_1 : \mu = \mu_1$, we have:

$$\frac{\overline{X} - \mu_1}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Therefore, we want to compute the following probability

$$P\Big(\frac{\overline{X} - \mu_1}{\sigma/\sqrt{n}} > \frac{\mu_0 - \mu_1}{\sigma/\sqrt{n}} + z_\alpha\Big).$$

Substituting $\sigma = 2$, $n = 16$, $\mu_0 = 35$, $\mu_1 = 36.5$ and $z_\alpha = 1.645$, we obtain:

$$\begin{aligned} P\Big(\frac{\overline{X} - \mu_1}{\sigma/\sqrt{n}} > \frac{35 - 36.5}{2/\sqrt{16}} + 1.645\Big) &= P\Big(\frac{\overline{X} - \mu_1}{\sigma/\sqrt{n}} > -1.355\Big) \\ &= 1 - P\Big(\frac{\overline{X} - \mu_1}{\sigma/\sqrt{n}} > 1.355\Big) \\ &= 1 - 0.0877 = 0.9123. \end{aligned}$$

Note that $z_{0.0885} = 1.35$ and $z_{0.0869} = 1.36$.

$\boxed{15}$ $X_1, X_2, \cdots, X_n$ are assumed to be mutually independent and be distributed as a Poisson process, where the Poisson distribution is given by:

$$P(X = x) = f(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \cdots.$$

Then, answer the following questions.

(1) Obtain the maximum likelihood estimator of λ, which is denoted by $\hat{\lambda}$.

(2) Prove that $\hat{\lambda}$ is an unbiased estimator.

(3) Prove that $\hat{\lambda}$ is an efficient estimator.

(4) Prove that $\hat{\lambda}$ is an consistent estimator.

[Answer]

(1) We obtain the maximum likelihood estimator of λ, denoted by $\hat{\lambda}$. The Poisson distribution is:

$$P(X = x) = f(x; \lambda) = \frac{\lambda^x e^{-\lambda}}{x!}, \quad x = 0, 1, 2, \cdots.$$

The likelihood function is:

$$l(\lambda) = \prod_{i=1}^{n} f(x_i;\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i} e^{-\lambda}}{x_i!} = \frac{\lambda^{\sum_{i=1}^{n} x_i} e^{-n\lambda}}{\prod_{i=1}^{n} x_i!}.$$

The log-likelihood function is:

$$\log l(\lambda) = \log(\lambda) \sum_{i=1}^{n} x_i - n\lambda - \log(\prod_{i=1}^{n} x_i!).$$

The derivative of the log-likelihood function with respect to λ is:

$$\frac{\partial \log l(\lambda)}{\partial \lambda} = \frac{1}{\lambda} \sum_{i=1}^{n} x_i - n = 0.$$

Solving the above equation, the maximum likelihood estimator $\hat{\lambda}$ is:

$$\hat{\lambda} = \frac{1}{n} \sum_{i=1}^{n} X_i = \overline{X}.$$

(2) We prove that $\hat{\lambda}$ is an unbiased estimator of λ.

$$\mathrm{E}(\hat{\lambda}) = \mathrm{E}(\frac{1}{n} \sum_{i=1}^{n} X_i) = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}(X_i) = \frac{1}{n} \sum_{i=1}^{n} \lambda = \lambda.$$

(3) We prove that $\hat{\lambda}$ is an efficient estimator of λ, where we show that the equality holds in the Cramer-Rao inequality. First, we obtain $\mathrm{V}(\hat{\lambda})$ as:

$$\mathrm{V}(\hat{\lambda}) = \mathrm{V}(\frac{1}{n} \sum_{i=1}^{n} X_i) = \frac{1}{n^2} \sum_{i=1}^{n} \mathrm{V}(X_i) = \frac{1}{n^2} \sum_{i=1}^{n} \lambda = \frac{\lambda}{n}.$$

The Cramer-Rao lower bound is given by:

$$\begin{aligned} \frac{1}{n\mathrm{E}\left[\left(\dfrac{\partial \log f(X;\lambda)}{\partial \lambda}\right)^2\right]} &= \frac{1}{n\mathrm{E}\left[\left(\dfrac{\partial (X \log \lambda - \lambda - \log X!)}{\partial \lambda}\right)^2\right]} \\ &= \frac{1}{n\mathrm{E}\left[\left(\dfrac{X}{\lambda} - 1\right)^2\right]} = \frac{\lambda^2}{n\mathrm{E}[(X-\lambda)^2]} \\ &= \frac{\lambda^2}{n\mathrm{V}(X)} = \frac{\lambda^2}{n\lambda} = \frac{\lambda}{n}. \end{aligned}$$

Therefore,

$$\mathrm{V}(\hat{\lambda}) = \frac{1}{n\mathrm{E}\left[\left(\dfrac{\partial \log f(X;\lambda)}{\partial\lambda}\right)^2\right]}.$$

That is, $\mathrm{V}(\hat{\lambda})$ is equal to the lower bound of the Cramer-Rao inequality. Therefore, $\hat{\lambda}$ is efficient.

(4) We show that $\hat{\lambda}$ is a consistent estimator of λ. Note as follows:

$$\mathrm{E}(\hat{\lambda}) = \lambda, \qquad \mathrm{V}(\hat{\lambda}) = \frac{\lambda}{n}.$$

In Chebyshev's inequality:

$$P(|\hat{\lambda} - \mathrm{E}(\hat{\lambda})| \geq \epsilon) \leq \frac{\mathrm{V}(\hat{\lambda})}{\epsilon^2},$$

$\mathrm{E}(\hat{\lambda})$ and $\mathrm{V}(\hat{\lambda})$ are substituted. Then, we have:

$$P(|\hat{\lambda} - \lambda| > \epsilon) < \frac{\lambda}{n\epsilon^2} \longrightarrow 0,$$

which implies that $\hat{\lambda}$ is consistent.

$\boxed{16}$ $X_1, X_2, \cdots, X_n$ are mutually independently distributed as normal random variables. Note that the normal density is:

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

Then, answer the following questions.

(1) Prove that the sample mean $\overline{X} = (1/n)\sum_{i=1}^n X_i$ is normally distributed with mean μ and variance σ^2/n.

(2) Define:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}.$$

Show that Z is normally distributed with mean zero and variance one.

(3) Consider the sample unbiased variance:

$$S^2 = \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X})^2.$$

The distribution of $(n-1)S^2/\sigma^2$ is known as a Chi-square distribution with $n-1$ degrees of freedom. Obtain mean and variance of S^2. Note that a Chi-square distribution with m degrees of freedom is:

$$f(x) = \begin{cases} \dfrac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})} x^{\frac{m}{2}-1} e^{-\frac{x}{2}}, & \text{if } x > 0, \\ 0, & \text{otherwise.} \end{cases}$$

(4) Prove that S^2 is an consistent estimator of σ^2.

[Answer]

(1) The distribution of the sample mean $\overline{X} = (1/n)\sum_{i=1}^n X_i$ is derived using the moment-generating function. Note that for $X \sim N(\mu, \sigma^2)$ the moment-generating function $\phi(\theta)$ is:

$$\begin{aligned}
\phi(\theta) \equiv \mathrm{E}(e^{\theta X}) &= \int_{-\infty}^{\infty} e^{\theta x} f(x)\,\mathrm{d}x = \int_{-\infty}^{\infty} e^{\theta x} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}(x-\mu)^2+\theta x}\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\left(x^2-2(\mu+\sigma^2\theta)x+\mu^2\right)}\,\mathrm{d}x \\
&= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\left(x-(\mu+\sigma^2\theta)\right)^2+(\mu\theta+\frac{1}{2}\sigma^2\theta^2)}\,\mathrm{d}x \\
&= e^{\mu\theta+\frac{1}{2}\sigma^2\theta^2} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2\sigma^2}\left(x-(\mu+\sigma^2\theta)\right)^2}\,\mathrm{d}x = \exp\Big(\mu\theta + \frac{1}{2}\sigma^2\theta^2\Big).
\end{aligned}$$

In the integration above, $N(\mu + \sigma^2\theta, \sigma^2)$ is utilized. Therefore, we have:

$$\phi_i(\theta) = \exp\Big(\mu\theta + \frac{1}{2}\sigma^2\theta^2\Big).$$

Now, consider the moment-generating function of $\overline{X}$, denoted by $\phi_{\overline{x}}(\theta)$:

$$\begin{aligned}
\phi_{\overline{x}}(\theta) \equiv \mathrm{E}(e^{\theta\overline{X}}) &= \mathrm{E}(e^{\theta\frac{1}{n}\sum_{i=1}^n X_i}) = \mathrm{E}\Big(\prod_{i=1}^n e^{\frac{\theta}{n}X_i}\Big) = \prod_{i=1}^n \mathrm{E}(e^{\frac{\theta}{n}X_i}) = \prod_{i=1}^n \phi_i\Big(\frac{\theta}{n}\Big) \\
&= \prod_{i=1}^n \exp\Big(\mu\frac{\theta}{n} + \frac{1}{2}\sigma^2\Big(\frac{\theta}{n}\Big)^2\Big) = \exp\Big(\mu\theta + \frac{1}{2}\sigma^2\frac{\theta^2}{n}\Big) = \exp\Big(\mu\theta + \frac{1}{2}\frac{\sigma^2}{n}\theta^2\Big),
\end{aligned}$$

which is equivalent to the moment-generating function of the normal distribution with mean μ and variance σ^2/n.

(2) We derive the distribution of Z, which is shown as:

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}}.$$

From the question (1), the moment-generating function of $\overline{X}$, denoted by $\phi_{\overline{x}}(\theta)$, is:

$$\phi_{\overline{x}}(\theta) \equiv \mathrm{E}(e^{\theta\overline{X}}) = \exp\Big(\mu\theta + \frac{1}{2}\frac{\sigma^2}{n}\theta^2\Big).$$

The moment-generating function of Z, denoted by $\phi_z(\theta)$:

$$\begin{aligned}
\phi_z(\theta) &\equiv \mathrm{E}(e^{\theta Z}) = \mathrm{E}\Big(\exp(\theta\frac{\overline{X}-\mu}{\sigma/\sqrt{n}})\Big) \\
&= \exp\Big(-\theta\frac{\mu}{\sigma/\sqrt{n}}\Big)\mathrm{E}\Big(\exp(\frac{\theta}{\sigma/\sqrt{n}}\overline{X})\Big) \\
&= \exp\Big(-\theta\frac{\mu}{\sigma/\sqrt{n}}\Big)\phi_{\overline{x}}(\frac{\theta}{\sigma/\sqrt{n}}) \\
&= \exp\Big(-\theta\frac{\mu}{\sigma/\sqrt{n}}\Big)\exp\Big(\mu\frac{\theta}{\sigma/\sqrt{n}}+\frac{1}{2}\frac{\sigma^2}{n}\Big(\frac{\theta}{\sigma/\sqrt{n}}\Big)^2\Big) = \exp(\frac{1}{2}\theta^2),
\end{aligned}$$

which is the moment-generating function of $N(0,1)$.

(3) First, as preliminaries, we derive mean and variance of the chi-square distribution with m degrees of freedom. The chi-square distribution with m degrees of freedom is:

$$f(x) = \frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})}x^{\frac{m}{2}-1}e^{-\frac{x}{2}}, \quad \text{if } x > 0.$$

Therefore, the moment-generating function $\phi_{\chi^2}(\theta)$ is:

$$\begin{aligned}
\phi_{\chi^2}(\theta) &= \mathrm{E}(e^{\theta X}) = \int_0^\infty e^{\theta x}\frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})}x^{\frac{m}{2}-1}e^{-\frac{x}{2}}\,\mathrm{d}x \\
&= \int_0^\infty \frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})}x^{\frac{m}{2}-1}e^{-\frac{1}{2}(1-2\theta)x}\,\mathrm{d}x \\
&= \int_0^\infty \frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})}\Big(\frac{y}{1-2\theta}\Big)^{\frac{m}{2}-1}e^{-\frac{1}{2}y}\frac{1}{1-2\theta}\,\mathrm{d}x \\
&= \Big(\frac{1}{1-2\theta}\Big)^{\frac{m}{2}-1}\frac{1}{1-2\theta}\int_0^\infty \frac{1}{2^{\frac{m}{2}}\Gamma(\frac{m}{2})}y^{\frac{m}{2}-1}e^{-\frac{1}{2}y}\,\mathrm{d}x = (1-2\theta)^{-\frac{m}{2}}.
\end{aligned}$$

In the fourth equality, use $y = (1-2\theta)x$. The first and second derivatives of the moment-generating function is:

$$\phi'_{\chi^2}(\theta) = m(1-2\theta)^{-\frac{m}{2}-1}, \qquad \phi''_{\chi^2}(\theta) = m(m+2)(1-2\theta)^{-\frac{m}{2}-2}.$$

Therefore, we obtain:

$$\mathrm{E}(X) = \phi'_{\chi^2}(0) = m, \qquad \mathrm{E}(X^2) = \phi''_{\chi^2}(0) = m(m+2).$$

Thus, for the chi-square distribution with m degrees of freedom, mean is given by m and variance is:

$$\mathrm{V}(X) = \mathrm{E}(X^2) - (\mathrm{E}(X))^2 = m(m+2) - m^2 = 2m.$$

Therefore, using $(n-1)S^2/\sigma^2 \sim \chi^2(n-1)$, we have:

$$\mathrm{E}(\frac{(n-1)S^2}{\sigma^2}) = n-1, \qquad \mathrm{V}(\frac{(n-1)S^2}{\sigma^2}) = 2(n-1),$$

which implies

$$\frac{n-1}{\sigma^2}\mathrm{E}(S^2) = n-1, \qquad (\frac{n-1}{\sigma^2})^2\mathrm{V}(S^2) = 2(n-1).$$

Finally, mean and variance of S^2 are:

$$\mathrm{E}(S^2) = \sigma^2, \qquad \mathrm{V}(S^2) = \frac{2\sigma^4}{n-1}.$$

(4) We show that S^2 is a consistent estimator of σ^2. Chebyshev's inequality is utilized, which is:

$$P(|S^2 - \mathrm{E}(S^2)| \geq \epsilon) \leq \frac{\mathrm{V}(S^2)}{\epsilon^2}.$$

Substituting $\mathrm{E}(S^2)$ and $\mathrm{V}(S^2)$, we obtain:

$$P(|S^2 - \sigma^2| \geq \epsilon) \leq \frac{2\sigma^4}{(n-1)\epsilon^2} \longrightarrow 0.$$

Therefore, S^2 is consistent.

Statistical Tables

Table 1.2: Standard Normal Distribution — $Z \sim N(0, 1)$

$$\alpha = P(Z > z_\alpha) = \int_{z_\alpha}^{\infty} \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}x^2)\,\mathrm{d}x$$

z_α	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.50000	.49601	.49202	.48803	.48405	.48006	.47608	.47210	.46812	.46414
0.1	.46017	.45620	.45224	.44828	.44433	.44038	.43644	.43251	.42858	.42465
0.2	.42074	.41683	.41294	.40905	.40517	.40129	.39743	.39358	.38974	.38591
0.3	.38209	.37828	.37448	.37070	.36693	.36317	.35942	.35569	.35197	.34827
0.4	.34458	.34090	.33724	.33360	.32997	.32636	.32276	.31918	.31561	.31207
0.5	.30854	.30503	.30153	.29806	.29460	.29116	.28774	.28434	.28096	.27760
0.6	.27425	.27093	.26763	.26435	.26109	.25785	.25463	.25143	.24825	.24510
0.7	.24196	.23885	.23576	.23270	.22965	.22663	.22363	.22065	.21770	.21476
0.8	.21186	.20897	.20611	.20327	.20045	.19766	.19489	.19215	.18943	.18673
0.9	.18406	.18141	.17879	.17619	.17361	.17106	.16853	.16602	.16354	.16109
1.0	.15866	.15625	.15386	.15151	.14917	.14686	.14457	.14231	.14007	.13786
1.1	.13567	.13350	.13136	.12924	.12714	.12507	.12302	.12100	.11900	.11702
1.2	.11507	.11314	.11123	.10935	.10749	.10565	.10383	.10204	.10027	.09853
1.3	.09680	.09510	.09342	.09176	.09012	.08851	.08692	.08534	.08379	.08226
1.4	.08076	.07927	.07780	.07636	.07493	.07353	.07215	.07078	.06944	.06811
1.5	.06681	.06552	.06426	.06301	.06178	.06057	.05938	.05821	.05705	.05592
1.6	.05480	.05370	.05262	.05155	.05050	.04947	.04846	.04746	.04648	.04551
1.7	.04457	.04363	.04272	.04182	.04093	.04006	.03920	.03836	.03754	.03673
1.8	.03593	.03515	.03438	.03362	.03288	.03216	.03144	.03074	.03005	.02938
1.9	.02872	.02807	.02743	.02680	.02619	.02559	.02500	.02442	.02385	.02330
2.0	.02275	.02222	.02169	.02118	.02068	.02018	.01970	.01923	.01876	.01831
2.1	.01786	.01743	.01700	.01659	.01618	.01578	.01539	.01500	.01463	.01426
2.2	.01390	.01355	.01321	.01287	.01255	.01222	.01191	.01160	.01130	.01101
2.3	.01072	.01044	.01017	.00990	.00964	.00939	.00914	.00889	.00866	.00842
2.4	.00820	.00798	.00776	.00755	.00734	.00714	.00695	.00676	.00657	.00639
2.5	.00621	.00604	.00587	.00570	.00554	.00539	.00523	.00508	.00494	.00480
2.6	.00466	.00453	.00440	.00427	.00415	.00402	.00391	.00379	.00368	.00357
2.7	.00347	.00336	.00326	.00317	.00307	.00298	.00289	.00280	.00272	.00264
2.8	.00256	.00248	.00240	.00233	.00226	.00219	.00212	.00205	.00199	.00193
2.9	.00187	.00181	.00175	.00169	.00164	.00159	.00154	.00149	.00144	.00139
3.0	.00135	.00131	.00126	.00122	.00118	.00114	.00111	.00107	.00104	.00100
3.1	.00097	.00094	.00090	.00087	.00084	.00082	.00079	.00076	.00074	.00071
3.2	.00069	.00066	.00064	.00062	.00060	.00058	.00056	.00054	.00052	.00050
3.3	.00048	.00047	.00045	.00043	.00042	.00040	.00039	.00038	.00036	.00035
3.4	.00034	.00032	.00031	.00030	.00029	.00028	.00027	.00026	.00025	.00024
3.5	.00023	.00022	.00022	.00021	.00020	.00019	.00019	.00018	.00017	.00017
3.6	.00016	.00015	.00015	.00014	.00014	.00013	.00013	.00012	.00012	.00011
3.7	.00011	.00010	.00010	.00010	.00009	.00009	.00008	.00008	.00008	.00008
3.8	.00007	.00007	.00007	.00006	.00006	.00006	.00006	.00005	.00005	.00005
3.9	.00005	.00005	.00004	.00004	.00004	.00004	.00004	.00004	.00003	.00003

Table 1.3: Chi-Square Distribution — $X \sim \chi^2(m)$

$$\alpha = P(X > \chi^2_\alpha) = \int_{\chi^2_\alpha}^{\infty} f(\chi^2)\,\mathrm{d}\chi^2$$

m \ α	.995	.99	.975	.95	.90	.10	.05	.025	.010	.005
1	.0000393	.000157	.000982	.00393	.0158	2.71	3.84	5.02	6.63	7.88
2	.0100	.0201	.0506	.103	.211	4.61	5.99	7.38	9.21	10.60
3	.0717	.115	.216	.352	.584	6.25	7.81	9.35	11.34	12.84
4	.207	.297	.484	.711	1.06	7.78	9.49	11.14	13.28	14.86
5	.412	.554	.831	1.15	1.61	9.24	11.07	12.83	15.09	16.75
6	.676	.872	1.24	1.64	2.20	10.64	12.59	14.45	16.81	18.55
7	.989	1.24	1.69	2.17	2.83	12.02	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	3.49	13.36	15.51	17.53	20.09	21.95
9	1.73	2.09	2.70	3.33	4.17	14.68	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	4.87	15.99	18.31	20.48	23.21	25.19
11	2.60	3.05	3.82	4.57	5.58	17.28	19.68	21.92	24.73	26.76
12	3.07	3.57	4.40	5.23	6.30	18.55	21.03	23.34	26.22	28.30
13	3.57	4.11	5.01	5.89	7.04	19.81	22.36	24.74	27.69	29.82
14	4.07	4.66	5.63	6.57	7.79	21.06	23.68	26.12	29.14	31.32
15	4.60	5.23	6.26	7.26	8.55	22.31	25.00	27.49	30.58	32.80
16	5.14	5.81	6.91	7.96	9.31	23.54	26.30	28.85	32.00	34.27
17	5.70	6.41	7.56	8.67	10.09	24.77	27.59	30.19	33.41	35.72
18	6.26	7.01	8.23	9.39	10.86	25.99	28.87	31.53	34.81	37.16
19	6.84	7.63	8.91	10.12	11.65	27.20	30.14	32.85	36.19	38.58
20	7.43	8.26	9.59	10.85	12.44	28.41	31.41	34.17	37.57	40.00
21	8.03	8.90	10.28	11.59	13.24	29.62	32.67	35.48	38.93	41.40
22	8.64	9.54	10.98	12.34	14.04	30.81	33.92	36.78	40.29	42.80
23	9.26	10.20	11.69	13.09	14.85	32.01	35.17	38.08	41.64	44.18
24	9.89	10.86	12.40	13.85	15.66	33.20	36.42	39.36	42.98	45.56
25	10.52	11.52	13.12	14.61	16.47	34.38	37.65	40.65	44.31	46.93
26	11.16	12.20	13.84	15.38	17.29	35.56	38.89	41.92	45.64	48.29
27	11.81	12.88	14.57	16.15	18.11	36.74	40.11	43.19	46.96	49.65
28	12.46	13.56	15.31	16.93	18.94	37.92	41.34	44.46	48.28	50.99
29	13.12	14.26	16.05	17.71	19.77	39.09	42.56	45.72	49.59	52.34
30	13.79	14.95	16.79	18.49	20.60	40.26	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	29.05	51.81	55.76	59.34	63.69	66.77
50	27.99	29.71	32.36	34.76	37.69	63.17	67.50	71.42	76.15	79.49
60	35.53	37.48	40.48	43.19	46.46	74.40	79.08	83.30	88.38	91.95
70	43.28	45.44	48.76	51.74	55.33	85.53	90.53	95.02	100.43	104.21
80	51.17	53.54	57.15	60.39	64.28	96.58	101.88	106.63	112.33	116.32
90	59.20	61.75	65.65	69.13	73.29	107.57	113.15	118.14	124.12	128.30
100	67.33	70.06	74.22	77.93	82.36	118.50	124.34	129.56	135.81	140.17

Table 1.4: F Distribution — $F \sim F(m_1, m_2)$

$$\alpha = \mathrm{P}(F > F_\alpha) = \int_{F_\alpha}^{\infty} f(F)\,\mathrm{d}F$$

m_1 =Degree of freedom in the numerator
m_2 =Degree of freedom in the denominator

m_2 \ m_1	α	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	.050	161.	200.	216.	225.	230.	234.	237.	239.	241.	242.	243.	244.	245.	245.
	.025	648.	800.	864.	900.	922.	937.	948.	957.	963.	969.	973.	977.	980.	983.
	.010	4052	5000	5403	5625	5764	5859	5928	5981	6022	6056	6083	6106	6126	6143
	.005	16211	20000	21615	22500	23056	23437	23715	23925	24091	24224	24334	24426	24505	24572
2	.050	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.4	19.4
	.025	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4	39.4	39.4	39.4
	.010	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.4	99.4
	.005	199.	199.	199.	199.	199.	199.	199.	199.	199.	199.	199.	199.	199.	199.
3	.050	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.76	8.74	8.73	8.71
	.025	17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5	14.4	14.4	14.3	14.3	14.3
	.010	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	27.1	27.1	27.0	26.9
	.005	55.6	49.8	47.5	46.2	45.4	44.8	44.4	44.1	43.9	43.7	43.5	43.4	43.3	43.2
4	.050	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.94	5.91	5.89	5.87
	.025	12.2	10.6	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.79	8.75	8.72	8.68
	.010	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.5	14.4	14.3	14.2
	.005	31.3	26.3	24.3	23.2	22.5	22.0	21.6	21.4	21.1	21.0	20.8	20.7	20.6	20.5
5	.050	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.70	4.68	4.66	4.64
	.025	10.0	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.57	6.52	6.49	6.46
	.010	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.96	9.89	9.82	9.77
	.005	22.8	18.3	16.5	15.6	14.9	14.5	14.2	14.0	13.8	13.6	13.5	13.4	13.3	13.2
6	.050	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.98	3.96
	.025	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.41	5.37	5.33	5.30
	.010	13.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.79	7.72	7.66	7.60
	.005	18.6	14.5	12.9	12.0	11.5	11.1	10.8	10.6	10.4	10.3	10.1	10.0	9.95	9.88
7	.050	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.60	3.57	3.55	3.53
	.025	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.71	4.67	4.63	4.60
	.010	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.54	6.47	6.41	6.36
	.005	16.2	12.4	10.9	10.1	9.52	9.16	8.89	8.68	8.51	8.38	8.27	8.18	8.10	8.03
8	.050	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.31	3.28	3.26	3.24
	.025	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.24	4.20	4.16	4.13
	.010	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.73	5.67	5.61	5.56
	.005	14.7	11.0	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	7.10	7.01	6.94	6.87
9	.050	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.10	3.07	3.05	3.03
	.025	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.03	3.96	3.91	3.87	3.83	3.80
	.010	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.18	5.11	5.05	5.01
	.005	13.6	10.1	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.31	6.23	6.15	6.09
10	.050	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.94	2.91	2.89	2.86
	.025	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.66	3.62	3.58	3.55
	.010	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.77	4.71	4.65	4.60
	.005	12.8	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.75	5.66	5.59	5.53
11	.050	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.82	2.79	2.76	2.74
	.025	6.72	5.26	4.63	4.28	4.04	3.88	3.76	3.66	3.59	3.53	3.47	3.43	3.39	3.36
	.010	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.46	4.40	4.34	4.29
	.005	12.2	8.91	7.60	6.88	6.42	6.10	5.86	5.68	5.54	5.42	5.32	5.24	5.16	5.10
12	.050	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.72	2.69	2.66	2.64
	.025	6.55	5.10	4.47	4.12	3.89	3.73	3.61	3.51	3.44	3.37	3.32	3.28	3.24	3.21
	.010	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.22	4.16	4.10	4.05
	.005	11.8	8.51	7.23	6.52	6.07	5.76	5.52	5.35	5.20	5.09	4.99	4.91	4.84	4.77
13	.050	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.63	2.60	2.58	2.55
	.025	6.41	4.97	4.35	4.00	3.77	3.60	3.48	3.39	3.31	3.25	3.20	3.15	3.12	3.08
	.010	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.91	3.86
	.005	11.4	8.19	6.93	6.23	5.79	5.48	5.25	5.08	4.94	4.82	4.72	4.64	4.57	4.51

Table 1.4: F Distribution — $F \sim F(m_1, m_2)$: < Continued >

$$\alpha = \mathrm{P}(F > F_\alpha) = \int_{F_\alpha}^{\infty} f(F)\,\mathrm{d}F$$

m_1 =Degree of freedom in the numerator
m_2 =Degree of freedom in the denominator

m_2 \ m_1	α	15	16	17	18	19	20	25	30	40	50	60	80	100	200
1	.050	246.	246.	247.	247.	248.	248.	249.	250.	251.	252.	252.	253.	253.	254.
	.025	985.	987.	989.	990.	992.	993.	998.	1001	1006	1008	1010	1012	1013	1016
	.010	6157	6170	6181	6192	6201	6209	6240	6261	6287	6303	6313	6326	6334	6350
	.005	24630	24681	24727	24767	24803	24836	24960	25044	25148	25211	25253	25306	25337	25401
2	.050	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5	19.5	19.5	19.5	19.5
	.025	39.4	39.4	39.4	39.4	39.4	39.4	39.5	39.5	39.5	39.5	39.5	39.5	39.5	39.5
	.010	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5	99.5	99.5	99.5	99.5
	.005	199.	199.	199.	199.	199.	199.	199.	199.	199.	199.	199.	199.	199.	199.
3	.050	8.70	8.69	8.68	8.67	8.67	8.66	8.63	8.62	8.59	8.58	8.57	8.56	8.55	8.54
	.025	14.3	14.2	14.2	14.2	14.2	14.2	14.1	14.1	14.0	14.0	14.0	14.0	14.0	13.9
	.010	26.9	26.8	26.8	26.8	26.7	26.7	26.6	26.5	26.4	26.4	26.3	26.3	26.2	26.2
	.005	43.1	43.0	42.9	42.9	42.8	42.8	42.6	42.5	42.3	42.2	42.1	42.1	42.0	41.9
4	.050	5.86	5.84	5.83	5.82	5.81	5.80	5.77	5.75	5.72	5.70	5.69	5.67	5.66	5.65
	.025	8.66	8.63	8.61	8.59	8.58	8.56	8.50	8.46	8.41	8.38	8.36	8.33	8.32	8.29
	.010	14.2	14.2	14.1	14.1	14.0	14.0	13.9	13.8	13.7	13.7	13.7	13.6	13.6	13.5
	.005	20.4	20.4	20.3	20.3	20.2	20.2	20.0	19.9	19.8	19.7	19.6	19.5	19.5	19.4
5	.050	4.62	4.60	4.59	4.58	4.57	4.56	4.52	4.50	4.46	4.44	4.43	4.42	4.41	4.39
	.025	6.43	6.40	6.38	6.36	6.34	6.33	6.27	6.23	6.18	6.14	6.12	6.10	6.08	6.05
	.010	9.72	9.68	9.64	9.61	9.58	9.55	9.45	9.38	9.29	9.24	9.20	9.16	9.13	9.08
	.005	13.1	13.1	13.0	13.0	12.9	12.9	12.8	12.7	12.5	12.5	12.4	12.3	12.3	12.2
6	.050	3.94	3.92	3.91	3.90	3.88	3.87	3.83	3.81	3.77	3.75	3.74	3.72	3.71	3.69
	.025	5.27	5.24	5.22	5.20	5.18	5.17	5.11	5.07	5.01	4.98	4.96	4.93	4.92	4.88
	.010	7.56	7.52	7.48	7.45	7.42	7.40	7.30	7.23	7.14	7.09	7.06	7.01	6.99	6.93
	.005	9.81	9.76	9.71	9.66	9.62	9.59	9.45	9.36	9.24	9.17	9.12	9.06	9.03	8.95
7	.050	3.51	3.49	3.48	3.47	3.46	3.44	3.40	3.38	3.34	3.32	3.30	3.29	3.27	3.25
	.025	4.57	4.54	4.52	4.50	4.48	4.47	4.40	4.36	4.31	4.28	4.25	4.23	4.21	4.18
	.010	6.31	6.28	6.24	6.21	6.18	6.16	6.06	5.99	5.91	5.86	5.82	5.78	5.75	5.70
	.005	7.97	7.91	7.87	7.83	7.79	7.75	7.62	7.53	7.42	7.35	7.31	7.25	7.22	7.15
8	.050	3.22	3.20	3.19	3.17	3.16	3.15	3.11	3.08	3.04	3.02	3.01	2.99	2.97	2.95
	.025	4.10	4.08	4.05	4.03	4.02	4.00	3.94	3.89	3.84	3.81	3.78	3.76	3.74	3.71
	.010	5.52	5.48	5.44	5.41	5.38	5.36	5.26	5.20	5.12	5.07	5.03	4.99	4.96	4.91
	.005	6.81	6.76	6.72	6.68	6.64	6.61	6.48	6.40	6.29	6.22	6.18	6.12	6.09	6.02
9	.050	3.01	2.99	2.97	2.96	2.95	2.94	2.89	2.86	2.83	2.80	2.79	2.77	2.76	2.73
	.025	3.77	3.74	3.72	3.70	3.68	3.67	3.60	3.56	3.51	3.47	3.45	3.42	3.40	3.37
	.010	4.96	4.92	4.89	4.86	4.83	4.81	4.71	4.65	4.57	4.52	4.48	4.44	4.42	4.36
	.005	6.03	5.98	5.94	5.90	5.86	5.83	5.71	5.62	5.52	5.45	5.41	5.36	5.32	5.26
10	.050	2.85	2.83	2.81	2.80	2.79	2.77	2.73	2.70	2.66	2.64	2.62	2.60	2.59	2.56
	.025	3.52	3.50	3.47	3.45	3.44	3.42	3.35	3.31	3.26	3.22	3.20	3.17	3.15	3.12
	.010	4.56	4.52	4.49	4.46	4.43	4.41	4.31	4.25	4.17	4.12	4.08	4.04	4.01	3.96
	.005	5.47	5.42	5.38	5.34	5.31	5.27	5.15	5.07	4.97	4.90	4.86	4.81	4.77	4.71
11	.050	2.72	2.70	2.69	2.67	2.66	2.65	2.60	2.57	2.53	2.51	2.49	2.47	2.46	2.43
	.025	3.33	3.30	3.28	3.26	3.24	3.23	3.16	3.12	3.06	3.03	3.00	2.97	2.96	2.92
	.010	4.25	4.21	4.18	4.15	4.12	4.10	4.01	3.94	3.86	3.81	3.78	3.73	3.71	3.66
	.005	5.05	5.00	4.96	4.92	4.89	4.86	4.74	4.65	4.55	4.49	4.45	4.39	4.36	4.29
12	.050	2.62	2.60	2.58	2.57	2.56	2.54	2.50	2.47	2.43	2.40	2.38	2.36	2.35	2.32
	.025	3.18	3.15	3.13	3.11	3.09	3.07	3.01	2.96	2.91	2.87	2.85	2.82	2.80	2.76
	.010	4.01	3.97	3.94	3.91	3.88	3.86	3.76	3.70	3.62	3.57	3.54	3.49	3.47	3.41
	.005	4.72	4.67	4.63	4.59	4.56	4.53	4.41	4.33	4.23	4.17	4.12	4.07	4.04	3.97
13	.050	2.53	2.51	2.50	2.48	2.47	2.46	2.41	2.38	2.34	2.31	2.30	2.27	2.26	2.23
	.025	3.05	3.03	3.00	2.98	2.96	2.95	2.88	2.84	2.78	2.74	2.72	2.69	2.67	2.63
	.010	3.82	3.78	3.75	3.72	3.69	3.66	3.57	3.51	3.43	3.38	3.34	3.30	3.27	3.22
	.005	4.46	4.41	4.37	4.33	4.30	4.27	4.15	4.07	3.97	3.91	3.87	3.81	3.78	3.71

Table 1.4: F Distribution — $F \sim F(m_1, m_2)$: < Continued >

$$\alpha = \mathrm{P}(F > F_\alpha) = \int_{F_\alpha}^{\infty} f(F)\,\mathrm{d}F$$

m_1 =Degree of freedom in the numerator
m_2 =Degree of freedom in the denominator

m_2 \ m_1	α	1	2	3	4	5	6	7	8	9	10	11	12	13	14
	.050	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.57	2.53	2.51	2.48
14	.025	6.30	4.86	4.24	3.89	3.66	3.50	3.38	3.29	3.21	3.15	3.09	3.05	3.01	2.98
	.010	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.86	3.80	3.75	3.70
	.005	11.1	7.92	6.68	6.00	5.56	5.26	5.03	4.86	4.72	4.60	4.51	4.43	4.36	4.30
	.050	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.51	2.48	2.45	2.42
15	.025	6.20	4.77	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	3.01	2.96	2.92	2.89
	.010	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.73	3.67	3.61	3.56
	.005	10.8	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.33	4.25	4.18	4.12
	.050	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.46	2.42	2.40	2.37
16	.025	6.12	4.69	4.08	3.73	3.50	3.34	3.22	3.12	3.05	2.99	2.93	2.89	2.85	2.82
	.010	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.62	3.55	3.50	3.45
	.005	10.6	7.51	6.30	5.64	5.21	4.91	4.69	4.52	4.38	4.27	4.18	4.10	4.03	3.97
	.050	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.41	2.38	2.35	2.33
17	.025	6.04	4.62	4.01	3.66	3.44	3.28	3.16	3.06	2.98	2.92	2.87	2.82	2.79	2.75
	.010	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.52	3.46	3.40	3.35
	.005	10.4	7.35	6.16	5.50	5.07	4.78	4.56	4.39	4.25	4.14	4.05	3.97	3.90	3.84
	.050	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.37	2.34	2.31	2.29
18	.025	5.98	4.56	3.95	3.61	3.38	3.22	3.10	3.01	2.93	2.87	2.81	2.77	2.73	2.70
	.010	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.43	3.37	3.32	3.27
	.005	10.2	7.21	6.03	5.37	4.96	4.66	4.44	4.28	4.14	4.03	3.94	3.86	3.79	3.73
	.050	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.34	2.31	2.28	2.26
19	.025	5.92	4.51	3.90	3.56	3.33	3.17	3.05	2.96	2.88	2.82	2.76	2.72	2.68	2.65
	.010	8.18	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.36	3.30	3.24	3.19
	.005	10.1	7.09	5.92	5.27	4.85	4.56	4.34	4.18	4.04	3.93	3.84	3.76	3.70	3.64
	.050	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.31	2.28	2.25	2.23
20	.025	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.72	2.68	2.64	2.60
	.010	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.29	3.23	3.18	3.13
	.005	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.76	3.68	3.61	3.55
	.050	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.28	2.25	2.22	2.20
21	.025	5.83	4.42	3.82	3.48	3.25	3.09	2.97	2.87	2.80	2.73	2.68	2.64	2.60	2.56
	.010	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.24	3.17	3.12	3.07
	.005	9.83	6.89	5.73	5.09	4.68	4.39	4.18	4.01	3.88	3.77	3.68	3.60	3.54	3.48
	.050	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.26	2.23	2.20	2.17
22	.025	5.79	4.38	3.78	3.44	3.22	3.05	2.93	2.84	2.76	2.70	2.65	2.60	2.56	2.53
	.010	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.18	3.12	3.07	3.02
	.005	9.73	6.81	5.65	5.02	4.61	4.32	4.11	3.94	3.81	3.70	3.61	3.54	3.47	3.41
	.050	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.24	2.20	2.18	2.15
23	.025	5.75	4.35	3.75	3.41	3.18	3.02	2.90	2.81	2.73	2.67	2.62	2.57	2.53	2.50
	.010	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.14	3.07	3.02	2.97
	.005	9.63	6.73	5.58	4.95	4.54	4.26	4.05	3.88	3.75	3.64	3.55	3.47	3.41	3.35
	.050	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.22	2.18	2.15	2.13
24	.025	5.72	4.32	3.72	3.38	3.15	2.99	2.87	2.78	2.70	2.64	2.59	2.54	2.50	2.47
	.010	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.09	3.03	2.98	2.93
	.005	9.55	6.66	5.52	4.89	4.49	4.20	3.99	3.83	3.69	3.59	3.50	3.42	3.35	3.30
	.050	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.20	2.16	2.14	2.11
25	.025	5.69	4.29	3.69	3.35	3.13	2.97	2.85	2.75	2.68	2.61	2.56	2.51	2.48	2.44
	.010	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	3.06	2.99	2.94	2.89
	.005	9.48	6.60	5.46	4.84	4.43	4.15	3.94	3.78	3.64	3.54	3.45	3.37	3.30	3.25
	.050	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.13	2.09	2.06	2.04
30	.025	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.46	2.41	2.37	2.34
	.010	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.91	2.84	2.79	2.74
	.005	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.25	3.18	3.11	3.06

Table 1.4: F Distribution — $F \sim F(m_1, m_2)$: < Continued >

$$\alpha = \mathrm{P}(F > F_\alpha) = \int_{F_\alpha}^{\infty} f(F)\,\mathrm{d}F$$

m_1 =Degree of freedom in the numerator
m_2 =Degree of freedom in the denominator

m_2 \ m_1	α	15	16	17	18	19	20	25	30	40	50	60	80	100	200
	.050	2.46	2.44	2.43	2.41	2.40	2.39	2.34	2.31	2.27	2.24	2.22	2.20	2.19	2.16
14	.025	2.95	2.92	2.90	2.88	2.86	2.84	2.78	2.73	2.67	2.64	2.61	2.58	2.56	2.53
	.010	3.66	3.62	3.59	3.56	3.53	3.51	3.41	3.35	3.27	3.22	3.18	3.14	3.11	3.06
	.005	4.25	4.20	4.16	4.12	4.09	4.06	3.94	3.86	3.76	3.70	3.66	3.60	3.57	3.50
	.050	2.40	2.38	2.37	2.35	2.34	2.33	2.28	2.25	2.20	2.18	2.16	2.14	2.12	2.10
15	.025	2.86	2.84	2.81	2.79	2.77	2.76	2.69	2.64	2.59	2.55	2.52	2.49	2.47	2.44
	.010	3.52	3.49	3.45	3.42	3.40	3.37	3.28	3.21	3.13	3.08	3.05	3.00	2.98	2.92
	.005	4.07	4.02	3.98	3.95	3.91	3.88	3.77	3.69	3.59	3.52	3.48	3.43	3.39	3.33
	.050	2.35	2.33	2.32	2.30	2.29	2.28	2.23	2.19	2.15	2.12	2.11	2.08	2.07	2.04
16	.025	2.79	2.76	2.74	2.72	2.70	2.68	2.61	2.57	2.51	2.47	2.45	2.42	2.40	2.36
	.010	3.41	3.37	3.34	3.31	3.28	3.26	3.17	3.10	3.02	2.97	2.93	2.89	2.86	2.81
	.005	3.92	3.87	3.83	3.80	3.76	3.73	3.62	3.54	3.44	3.37	3.33	3.28	3.25	3.18
	.050	2.31	2.29	2.27	2.26	2.24	2.23	2.18	2.15	2.10	2.08	2.06	2.03	2.02	1.99
17	.025	2.72	2.70	2.67	2.65	2.63	2.62	2.55	2.50	2.44	2.41	2.38	2.35	2.33	2.29
	.010	3.31	3.27	3.24	3.21	3.19	3.16	3.07	3.00	2.92	2.87	2.83	2.79	2.76	2.71
	.005	3.79	3.75	3.71	3.67	3.64	3.61	3.49	3.41	3.31	3.25	3.21	3.15	3.12	3.05
	.050	2.27	2.25	2.23	2.22	2.20	2.19	2.14	2.11	2.06	2.04	2.02	1.99	1.98	1.95
18	.025	2.67	2.64	2.62	2.60	2.58	2.56	2.49	2.44	2.38	2.35	2.32	2.29	2.27	2.23
	.010	3.23	3.19	3.16	3.13	3.10	3.08	2.98	2.92	2.84	2.78	2.75	2.71	2.68	2.62
	.005	3.68	3.64	3.60	3.56	3.53	3.50	3.38	3.30	3.20	3.14	3.10	3.04	3.01	2.94
	.050	2.23	2.21	2.20	2.18	2.17	2.16	2.11	2.07	2.03	2.00	1.98	1.96	1.94	1.91
19	.025	2.62	2.59	2.57	2.55	2.53	2.51	2.44	2.39	2.33	2.30	2.27	2.24	2.22	2.18
	.010	3.15	3.12	3.08	3.05	3.03	3.00	2.91	2.84	2.76	2.71	2.67	2.63	2.60	2.55
	.005	3.59	3.54	3.50	3.46	3.43	3.40	3.29	3.21	3.11	3.04	3.00	2.95	2.91	2.85
	.050	2.20	2.18	2.17	2.15	2.14	2.12	2.07	2.04	1.99	1.97	1.95	1.92	1.91	1.88
20	.025	2.57	2.55	2.52	2.50	2.48	2.46	2.40	2.35	2.29	2.25	2.22	2.19	2.17	2.13
	.010	3.09	3.05	3.02	2.99	2.96	2.94	2.84	2.78	2.69	2.64	2.61	2.56	2.54	2.48
	.005	3.50	3.46	3.42	3.38	3.35	3.32	3.20	3.12	3.02	2.96	2.92	2.86	2.83	2.76
	.050	2.18	2.16	2.14	2.12	2.11	2.10	2.05	2.01	1.96	1.94	1.92	1.89	1.88	1.84
21	.025	2.53	2.51	2.48	2.46	2.44	2.42	2.36	2.31	2.25	2.21	2.18	2.15	2.13	2.09
	.010	3.03	2.99	2.96	2.93	2.90	2.88	2.79	2.72	2.64	2.58	2.55	2.50	2.48	2.42
	.005	3.43	3.38	3.34	3.31	3.27	3.24	3.13	3.05	2.95	2.88	2.84	2.79	2.75	2.68
	.050	2.15	2.13	2.11	2.10	2.08	2.07	2.02	1.98	1.94	1.91	1.89	1.86	1.85	1.82
22	.025	2.50	2.47	2.45	2.43	2.41	2.39	2.32	2.27	2.21	2.17	2.14	2.11	2.09	2.05
	.010	2.98	2.94	2.91	2.88	2.85	2.83	2.73	2.67	2.58	2.53	2.50	2.45	2.42	2.36
	.005	3.36	3.32	3.27	3.24	3.21	3.18	3.06	2.98	2.88	2.82	2.77	2.72	2.69	2.62
	.050	2.13	2.11	2.09	2.08	2.06	2.05	2.00	1.96	1.91	1.88	1.86	1.84	1.82	1.79
23	.025	2.47	2.44	2.42	2.39	2.37	2.36	2.29	2.24	2.18	2.14	2.11	2.08	2.06	2.01
	.010	2.93	2.89	2.86	2.83	2.80	2.78	2.69	2.62	2.54	2.48	2.45	2.40	2.37	2.32
	.005	3.30	3.25	3.21	3.18	3.15	3.12	3.00	2.92	2.82	2.76	2.71	2.66	2.62	2.56
	.050	2.11	2.09	2.07	2.05	2.04	2.03	1.98	1.94	1.89	1.86	1.84	1.82	1.80	1.77
24	.025	2.44	2.41	2.39	2.36	2.35	2.33	2.26	2.21	2.15	2.11	2.08	2.05	2.02	1.98
	.010	2.89	2.85	2.82	2.79	2.76	2.74	2.64	2.58	2.49	2.44	2.40	2.36	2.33	2.27
	.005	3.25	3.20	3.16	3.12	3.09	3.06	2.95	2.87	2.77	2.70	2.66	2.60	2.57	2.50
	.050	2.09	2.07	2.05	2.04	2.02	2.01	1.96	1.92	1.87	1.84	1.82	1.80	1.78	1.75
25	.025	2.41	2.38	2.36	2.34	2.32	2.30	2.23	2.18	2.12	2.08	2.05	2.02	2.00	1.95
	.010	2.85	2.81	2.78	2.75	2.72	2.70	2.60	2.54	2.45	2.40	2.36	2.32	2.29	2.23
	.005	3.20	3.15	3.11	3.08	3.04	3.01	2.90	2.82	2.72	2.65	2.61	2.55	2.52	2.45
	.050	2.01	1.99	1.98	1.96	1.95	1.93	1.88	1.84	1.79	1.76	1.74	1.71	1.70	1.66
30	.025	2.31	2.28	2.26	2.23	2.21	2.20	2.12	2.07	2.01	1.97	1.94	1.90	1.88	1.84
	.010	2.70	2.66	2.63	2.60	2.57	2.55	2.45	2.39	2.30	2.25	2.21	2.16	2.13	2.07
	.005	3.01	2.96	2.92	2.89	2.85	2.82	2.71	2.63	2.52	2.46	2.42	2.36	2.32	2.25

Table 1.4: F Distribution — $F \sim F(m_1, m_2)$: < Continued >

$$\alpha = \mathrm{P}(F > F_\alpha) = \int_{F_\alpha}^{\infty} f(F)\, \mathrm{d}F$$

m_1 =Degree of freedom in the numerator
m_2 =Degree of freedom in the denominator

m_2 \ m_1	α	1	2	3	4	5	6	7	8	9	10	11	12	13	14
35	.050	4.12	3.27	2.87	2.64	2.49	2.37	2.29	2.22	2.16	2.11	2.08	2.04	2.01	1.99
	.025	5.48	4.11	3.52	3.18	2.96	2.80	2.68	2.58	2.50	2.44	2.39	2.34	2.30	2.27
	.010	7.42	5.27	4.40	3.91	3.59	3.37	3.20	3.07	2.96	2.88	2.80	2.74	2.69	2.64
	.005	8.98	6.19	5.09	4.48	4.09	3.81	3.61	3.45	3.32	3.21	3.12	3.05	2.98	2.93
40	.050	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.04	2.00	1.97	1.95
	.025	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.33	2.29	2.25	2.21
	.010	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.73	2.66	2.61	2.56
	.005	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	3.12	3.03	2.95	2.89	2.83
45	.050	4.06	3.20	2.81	2.58	2.42	2.31	2.22	2.15	2.10	2.05	2.01	1.97	1.94	1.92
	.025	5.38	4.01	3.42	3.09	2.86	2.70	2.58	2.49	2.41	2.35	2.29	2.25	2.21	2.17
	.010	7.23	5.11	4.25	3.77	3.45	3.23	3.07	2.94	2.83	2.74	2.67	2.61	2.55	2.51
	.005	8.71	5.97	4.89	4.29	3.91	3.64	3.43	3.28	3.15	3.04	2.96	2.88	2.82	2.76
50	.050	4.03	3.18	2.79	2.56	2.40	2.29	2.20	2.13	2.07	2.03	1.99	1.95	1.92	1.89
	.025	5.34	3.97	3.39	3.05	2.83	2.67	2.55	2.46	2.38	2.32	2.26	2.22	2.18	2.14
	.010	7.17	5.06	4.20	3.72	3.41	3.19	3.02	2.89	2.79	2.70	2.63	2.56	2.51	2.46
	.005	8.63	5.90	4.83	4.23	3.85	3.58	3.38	3.22	3.09	2.99	2.90	2.82	2.76	2.70
60	.050	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.95	1.92	1.89	1.86
	.025	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.22	2.17	2.13	2.09
	.010	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.56	2.50	2.44	2.39
	.005	8.49	5.80	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.82	2.74	2.68	2.62
70	.050	3.98	3.13	2.74	2.50	2.35	2.23	2.14	2.07	2.02	1.97	1.93	1.89	1.86	1.84
	.025	5.25	3.89	3.31	2.97	2.75	2.59	2.47	2.38	2.30	2.24	2.18	2.14	2.10	2.06
	.010	7.01	4.92	4.07	3.60	3.29	3.07	2.91	2.78	2.67	2.59	2.51	2.45	2.40	2.35
	.005	8.40	5.72	4.66	4.08	3.70	3.43	3.23	3.08	2.95	2.85	2.76	2.68	2.62	2.56
80	.050	3.96	3.11	2.72	2.49	2.33	2.21	2.13	2.06	2.00	1.95	1.91	1.88	1.84	1.82
	.025	5.22	3.86	3.28	2.95	2.73	2.57	2.45	2.35	2.28	2.21	2.16	2.11	2.07	2.03
	.010	6.96	4.88	4.04	3.56	3.26	3.04	2.87	2.74	2.64	2.55	2.48	2.42	2.36	2.31
	.005	8.33	5.67	4.61	4.03	3.65	3.39	3.19	3.03	2.91	2.80	2.72	2.64	2.58	2.52
90	.050	3.95	3.10	2.71	2.47	2.32	2.20	2.11	2.04	1.99	1.94	1.90	1.86	1.83	1.80
	.025	5.20	3.84	3.26	2.93	2.71	2.55	2.43	2.34	2.26	2.19	2.14	2.09	2.05	2.02
	.010	6.93	4.85	4.01	3.54	3.23	3.01	2.84	2.72	2.61	2.52	2.45	2.39	2.33	2.29
	.005	8.28	5.62	4.57	3.99	3.62	3.35	3.15	3.00	2.87	2.77	2.68	2.61	2.54	2.49
100	.050	3.94	3.09	2.70	2.46	2.31	2.19	2.10	2.03	1.97	1.93	1.89	1.85	1.82	1.79
	.025	5.18	3.83	3.25	2.92	2.70	2.54	2.42	2.32	2.24	2.18	2.12	2.08	2.04	2.00
	.010	6.90	4.82	3.98	3.51	3.21	2.99	2.82	2.69	2.59	2.50	2.43	2.37	2.31	2.27
	.005	8.24	5.59	4.54	3.96	3.59	3.33	3.13	2.97	2.85	2.74	2.66	2.58	2.52	2.46
150	.050	3.90	3.06	2.66	2.43	2.27	2.16	2.07	2.00	1.94	1.89	1.85	1.82	1.79	1.76
	.025	5.13	3.78	3.20	2.87	2.65	2.49	2.37	2.28	2.20	2.13	2.08	2.03	1.99	1.95
	.010	6.81	4.75	3.91	3.45	3.14	2.92	2.76	2.63	2.53	2.44	2.37	2.31	2.25	2.20
	.005	8.12	5.49	4.45	3.88	3.51	3.25	3.05	2.89	2.77	2.67	2.58	2.51	2.44	2.38
200	.050	3.89	3.04	2.65	2.42	2.26	2.14	2.06	1.98	1.93	1.88	1.84	1.80	1.77	1.74
	.025	5.10	3.76	3.18	2.85	2.63	2.47	2.35	2.26	2.18	2.11	2.06	2.01	1.97	1.93
	.010	6.76	4.71	3.88	3.41	3.11	2.89	2.73	2.60	2.50	2.41	2.34	2.27	2.22	2.17
	.005	8.06	5.44	4.41	3.84	3.47	3.21	3.01	2.86	2.73	2.63	2.54	2.47	2.40	2.35
500	.050	3.86	3.01	2.62	2.39	2.23	2.12	2.03	1.96	1.90	1.85	1.81	1.77	1.74	1.71
	.025	5.05	3.72	3.14	2.81	2.59	2.43	2.31	2.22	2.14	2.07	2.02	1.97	1.93	1.89
	.010	6.69	4.65	3.82	3.36	3.05	2.84	2.68	2.55	2.44	2.36	2.28	2.22	2.17	2.12
	.005	7.95	5.35	4.33	3.76	3.40	3.14	2.94	2.79	2.66	2.56	2.48	2.40	2.34	2.28
∞	.050	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.79	1.75	1.72	1.69
	.025	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.99	1.94	1.90	1.87
	.010	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.25	2.18	2.13	2.08
	.005	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62	2.52	2.43	2.36	2.29	2.24

Table 1.4: F Distribution — $F \sim F(m_1, m_2)$: < Continued >

$$\alpha = \mathrm{P}(F > F_\alpha) = \int_{F_\alpha}^{\infty} f(F)\,\mathrm{d}F$$

m_1 =Degree of freedom in the numerator
m_2 =Degree of freedom in the denominator

m_2 \ m_1	α	15	16	17	18	19	20	25	30	40	50	60	80	100	200
35	.050	1.96	1.94	1.92	1.91	1.89	1.88	1.82	1.79	1.74	1.70	1.68	1.65	1.63	1.60
	.025	2.24	2.21	2.18	2.16	2.14	2.12	2.05	2.00	1.93	1.89	1.86	1.82	1.80	1.75
	.010	2.60	2.56	2.53	2.50	2.47	2.44	2.35	2.28	2.19	2.14	2.10	2.05	2.02	1.96
	.005	2.88	2.83	2.79	2.76	2.72	2.69	2.58	2.50	2.39	2.33	2.28	2.22	2.19	2.11
40	.050	1.92	1.90	1.89	1.87	1.85	1.84	1.78	1.74	1.69	1.66	1.64	1.61	1.59	1.55
	.025	2.18	2.15	2.13	2.11	2.09	2.07	1.99	1.94	1.88	1.83	1.80	1.76	1.74	1.69
	.010	2.52	2.48	2.45	2.42	2.39	2.37	2.27	2.20	2.11	2.06	2.02	1.97	1.94	1.87
	.005	2.78	2.74	2.70	2.66	2.63	2.60	2.48	2.40	2.30	2.23	2.18	2.12	2.09	2.01
45	.050	1.89	1.87	1.86	1.84	1.82	1.81	1.75	1.71	1.66	1.63	1.60	1.57	1.55	1.51
	.025	2.14	2.11	2.09	2.07	2.05	2.03	1.95	1.90	1.83	1.79	1.76	1.72	1.69	1.64
	.010	2.46	2.43	2.39	2.36	2.34	2.31	2.21	2.14	2.05	2.00	1.96	1.91	1.88	1.81
	.005	2.71	2.66	2.62	2.59	2.56	2.53	2.41	2.33	2.22	2.16	2.11	2.05	2.01	1.93
50	.050	1.87	1.85	1.83	1.81	1.80	1.78	1.73	1.69	1.63	1.60	1.58	1.54	1.52	1.48
	.025	2.11	2.08	2.06	2.03	2.01	1.99	1.92	1.87	1.80	1.75	1.72	1.68	1.66	1.60
	.010	2.42	2.38	2.35	2.32	2.29	2.27	2.17	2.10	2.01	1.95	1.91	1.86	1.82	1.76
	.005	2.65	2.61	2.57	2.53	2.50	2.47	2.35	2.27	2.16	2.10	2.05	1.99	1.95	1.87
60	.050	1.84	1.82	1.80	1.78	1.76	1.75	1.69	1.65	1.59	1.56	1.53	1.50	1.48	1.44
	.025	2.06	2.03	2.01	1.98	1.96	1.94	1.87	1.82	1.74	1.70	1.67	1.63	1.60	1.54
	.010	2.35	2.31	2.28	2.25	2.22	2.20	2.10	2.03	1.94	1.88	1.84	1.78	1.75	1.68
	.005	2.57	2.53	2.49	2.45	2.42	2.39	2.27	2.19	2.08	2.01	1.96	1.90	1.86	1.78
70	.050	1.81	1.79	1.77	1.75	1.74	1.72	1.66	1.62	1.57	1.53	1.50	1.47	1.45	1.40
	.025	2.03	2.00	1.97	1.95	1.93	1.91	1.83	1.78	1.71	1.66	1.63	1.59	1.56	1.50
	.010	2.31	2.27	2.23	2.20	2.18	2.15	2.05	1.98	1.89	1.83	1.78	1.73	1.70	1.62
	.005	2.51	2.47	2.43	2.39	2.36	2.33	2.21	2.13	2.02	1.95	1.90	1.84	1.80	1.71
80	.050	1.79	1.77	1.75	1.73	1.72	1.70	1.64	1.60	1.54	1.51	1.48	1.45	1.43	1.38
	.025	2.00	1.97	1.95	1.93	1.90	1.88	1.81	1.75	1.68	1.63	1.60	1.55	1.53	1.47
	.010	2.27	2.23	2.20	2.17	2.14	2.12	2.01	1.94	1.85	1.79	1.75	1.69	1.65	1.58
	.005	2.47	2.43	2.39	2.35	2.32	2.29	2.17	2.08	1.97	1.90	1.85	1.79	1.75	1.66
90	.050	1.78	1.76	1.74	1.72	1.70	1.69	1.63	1.59	1.53	1.49	1.46	1.43	1.41	1.36
	.025	1.98	1.95	1.93	1.91	1.88	1.86	1.79	1.73	1.66	1.61	1.58	1.53	1.50	1.44
	.010	2.24	2.21	2.17	2.14	2.11	2.09	1.99	1.92	1.82	1.76	1.72	1.66	1.62	1.55
	.005	2.44	2.39	2.35	2.32	2.28	2.25	2.13	2.05	1.94	1.87	1.82	1.75	1.71	1.62
100	.050	1.77	1.75	1.73	1.71	1.69	1.68	1.62	1.57	1.52	1.48	1.45	1.41	1.39	1.34
	.025	1.97	1.94	1.91	1.89	1.87	1.85	1.77	1.71	1.64	1.59	1.56	1.51	1.48	1.42
	.010	2.22	2.19	2.15	2.12	2.09	2.07	1.97	1.89	1.80	1.74	1.69	1.63	1.60	1.52
	.005	2.41	2.37	2.33	2.29	2.26	2.23	2.11	2.02	1.91	1.84	1.79	1.72	1.68	1.59
150	.050	1.73	1.71	1.69	1.67	1.66	1.64	1.58	1.54	1.48	1.44	1.41	1.37	1.34	1.29
	.025	1.92	1.89	1.87	1.84	1.82	1.80	1.72	1.67	1.59	1.54	1.50	1.45	1.42	1.35
	.010	2.16	2.12	2.09	2.06	2.03	2.00	1.90	1.83	1.73	1.66	1.62	1.56	1.52	1.43
	.005	2.33	2.29	2.25	2.21	2.18	2.15	2.03	1.94	1.83	1.76	1.70	1.63	1.59	1.49
200	.050	1.72	1.69	1.67	1.66	1.64	1.62	1.56	1.52	1.46	1.41	1.39	1.35	1.32	1.26
	.025	1.90	1.87	1.84	1.82	1.80	1.78	1.70	1.64	1.56	1.51	1.47	1.42	1.39	1.32
	.010	2.13	2.09	2.06	2.03	2.00	1.97	1.87	1.79	1.69	1.63	1.58	1.52	1.48	1.39
	.005	2.30	2.25	2.21	2.18	2.14	2.11	1.99	1.91	1.79	1.71	1.66	1.59	1.54	1.44
500	.050	1.69	1.66	1.64	1.62	1.61	1.59	1.53	1.48	1.42	1.38	1.35	1.30	1.28	1.21
	.025	1.86	1.83	1.80	1.78	1.76	1.74	1.65	1.60	1.52	1.46	1.42	1.37	1.34	1.25
	.010	2.07	2.04	2.00	1.97	1.94	1.92	1.81	1.74	1.63	1.57	1.52	1.45	1.41	1.31
	.005	2.23	2.19	2.14	2.11	2.07	2.04	1.92	1.84	1.72	1.64	1.58	1.51	1.46	1.35
∞	.050	1.67	1.64	1.62	1.60	1.59	1.57	1.51	1.46	1.39	1.35	1.32	1.27	1.24	1.17
	.025	1.83	1.80	1.78	1.75	1.73	1.71	1.63	1.57	1.48	1.43	1.39	1.33	1.30	1.21
	.010	2.04	2.00	1.97	1.93	1.90	1.88	1.77	1.70	1.59	1.52	1.47	1.40	1.36	1.25
	.005	2.19	2.14	2.10	2.06	2.03	2.00	1.88	1.79	1.67	1.59	1.53	1.45	1.40	1.28

Table 1.5: t Distribution — $T \sim t(m)$

$$\alpha = P(T > t_\alpha) = \int_{t_\alpha}^{\infty} \frac{\Gamma(\frac{m+1}{2})}{\Gamma(\frac{m}{2})} \frac{1}{\sqrt{m\pi}} \frac{1}{(1 + x^2/m)^{(m+1)/2}} \, \mathrm{d}x$$

α / m	.10	.05	.025	.01	.005
1	3.0777	6.3137	12.7062	31.8210	63.6559
2	1.8856	2.9200	4.3027	6.9645	9.9250
3	1.6377	2.3534	3.1824	4.5407	5.8408
4	1.5332	2.1318	2.7765	3.7469	4.6041
5	1.4759	2.0150	2.5706	3.3649	4.0321
6	1.4398	1.9432	2.4469	3.1427	3.7074
7	1.4149	1.8946	2.3646	2.9979	3.4995
8	1.3968	1.8595	2.3060	2.8965	3.3554
9	1.3830	1.8331	2.2622	2.8214	3.2498
10	1.3722	1.8125	2.2281	2.7638	3.1693
11	1.3634	1.7959	2.2010	2.7181	3.1058
12	1.3562	1.7823	2.1788	2.6810	3.0545
13	1.3502	1.7709	2.1604	2.6503	3.0123
14	1.3450	1.7613	2.1448	2.6245	2.9768
15	1.3406	1.7531	2.1315	2.6025	2.9467
16	1.3368	1.7459	2.1199	2.5835	2.9208
17	1.3334	1.7396	2.1098	2.5669	2.8982
18	1.3304	1.7341	2.1009	2.5524	2.8784
19	1.3277	1.7291	2.0930	2.5395	2.8609
20	1.3253	1.7247	2.0860	2.5280	2.8453
21	1.3232	1.7207	2.0796	2.5176	2.8314
22	1.3212	1.7171	2.0739	2.5083	2.8188
23	1.3195	1.7139	2.0687	2.4999	2.8073
24	1.3178	1.7109	2.0639	2.4922	2.7970
25	1.3163	1.7081	2.0595	2.4851	2.7874
26	1.3150	1.7056	2.0555	2.4786	2.7787
27	1.3137	1.7033	2.0518	2.4727	2.7707
28	1.3125	1.7011	2.0484	2.4671	2.7633
29	1.3114	1.6991	2.0452	2.4620	2.7564
30	1.3104	1.6973	2.0423	2.4573	2.7500
31	1.3095	1.6955	2.0395	2.4528	2.7440
32	1.3086	1.6939	2.0369	2.4487	2.7385
33	1.3077	1.6924	2.0345	2.4448	2.7333
34	1.3070	1.6909	2.0322	2.4411	2.7284
35	1.3062	1.6896	2.0301	2.4377	2.7238
40	1.3031	1.6839	2.0211	2.4233	2.7045
50	1.2987	1.6759	2.0086	2.4033	2.6778
60	1.2958	1.6706	2.0003	2.3901	2.6603
70	1.2938	1.6669	1.9944	2.3808	2.6479
80	1.2922	1.6641	1.9901	2.3739	2.6387
90	1.2910	1.6620	1.9867	2.3685	2.6316
100	1.2901	1.6602	1.9840	2.3642	2.6259
∞	1.2816	1.6449	1.9600	2.3264	2.5758

Index