Law of Large Numbers 2: 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 $\overline{X}_n = \dfrac{1}{n}\sum_{i=1}^{n} X_i$.
We say that $\overline{X}_n$ converges in probability to μ.

Proof:
Using (10), Chebyshev's inequality is represented as follows:

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

where X in (10) is replaced by $\overline{X}_n$.
We know $\mathrm{E}(\overline{X}_n) = \mu$ and $\mathrm{V}(\overline{X}_n) = \dfrac{\sigma^2}{n}$, which are substituted into the above inequality.
Then, we obtain:

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

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

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

That is, $\overline{X}_n \longrightarrow \mu$ is obtained as $n \longrightarrow \infty$, which is written as: $\mathrm{plim}\, \overline{X}_n = \mu$.
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 (確率収束)**.
In this case, we say that $\overline{X}_n$ converges in probability to μ.

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

$$m_n = \mathrm{E}(\sum_{i=1}^{n} X_i), \qquad V_n = \mathrm{V}(\sum_{i=1}^{n} X_i),$$

and assume that

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

Then, we obtain the following result:

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

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

Proof:
Remember Chebyshev's inequality:

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

Replace X, $\mathrm{E}(X)$ and $\mathrm{V}(X)$
by $\overline{X}_n$, $\mathrm{E}(\overline{X}_n) = \dfrac{m_n}{n}$ and $\mathrm{V}(\overline{X}_n) = \dfrac{V_n}{n^2}$.
Then, we obtain:

$$P(|\overline{X}_n - \frac{m_n}{n}| > \epsilon) \le \frac{V_n}{n^2\epsilon^2}.$$

As n goes to infinity,

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

Therefore, $\overline{X}_n \longrightarrow \lim_{n\to\infty} \frac{m_n}{n}$ as $n \longrightarrow \infty$.

6.3 Central Limit Theorem (中心極限定理) and Convergence in Distribution (分布収束)

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 $\theta = 0$.

(*) Remark:

$O(x)$ 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:

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

Note that $x \longrightarrow 0$ when $n \longrightarrow \infty$ and that $\lim_{x\to 0}(1+x)^{1/x} = e$ as in Section 2.3 (p.35).
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.110 in Section 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,

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

We say that $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}$ converges in distribution to $N(0,1)$.
$\Longrightarrow$ **Convergence in distribution (分布収束)**

The following expression is also possible:

$$\sqrt{n}(\overline{X}_n - \mu) \longrightarrow N(0,\sigma^2). \qquad (11)$$

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

$$\lim_{n\to\infty} n\mathrm{V}(\overline{X}_n) = \sigma^2 < \infty, \qquad \text{where } \overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i.$$

Then, when $n \longrightarrow \infty$, we have:

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

Summary: Let $X_1, X_2, \cdots, X_n, \cdots$ be a sequence of random variables. Let X be a random variable. Let F_n be the distribution function of X_n and F be that of X.

- X_n converges in probability to X if $\lim_{n\to\infty} P(|X_n - X| \geq \epsilon) = 0$ or $\lim_{n\to\infty} P(|X_n - X| < \epsilon) = 1$ for all $\epsilon > 0$.
Equivalently, we write $X_n \xrightarrow{P} X$.

- X_n converges in distribution to X (or F) if $\lim_{n\to\infty} F_n(x) = F(x)$ for all x.
Equivalently, we write $X_n \xrightarrow{D} X$ or $X_n \xrightarrow{D} F$.

7 Statistical Inference

7.1 Point Estimation (点推定)

Suppose that the underlying distribution 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, $x_2,\ \cdots,\ x_n$.

$\hat{\theta}_n(x_1,\ x_2,\ \cdots,\ x_n)$ is constructed to estimate the parameter θ.

$\hat{\theta}_n(x_1,\ x_2,\ \cdots,\ x_n)$ takes a certain value given the n observed data.

$\hat{\theta}_n(x_1,\ x_2,\ \cdots,\ x_n)$ 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:

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

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

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

7.2 Statistic, Estimate and Estimator (統計量，推定値，推定量)

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, $X_2,\ \cdots,\ X_n$ take 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.

$\hat{\theta}_n(x_1, x_2, \cdots, x_n)$ 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 (推定量)** of θ, which should be distinguished from the **estimate (推定値)** 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 $\overline{X} = \dfrac{1}{n}\sum_{i=1}^{n} X_i$, while the estimate of μ is $\overline{x} = \dfrac{1}{n}\sum_{i=1}^{n} x_i$.

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

There are numerous estimators and estimates of θ.

All of $\dfrac{1}{n}\sum_{i=1}^{n} X_i$, $\dfrac{X_1 + X_n}{2}$, 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.

Both $S^2 = \dfrac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2$ and $S^{*2} = \dfrac{1}{n}\sum_{i=1}^{2}(X_i - \overline{X})^2$ 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 7.4 and 7.5.

In addition, note as follows.

A function of random variables is called a **statistic (統計量)**. The statistic for estimation of the parameter is called an estimator.

Therefore, an estimator is a family of a statistic.

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

Properties of $\overline{X}$: From Theorem on p.138, mean and variance of $\overline{X}$ are obtained as follows:

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

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

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