If the following holds,

$$\lim_{n\to\infty} P(|Z_n - \theta| < \epsilon) = 1,$$

for any positive ϵ, then we say that Z_n converges to θ in probability.

θ is called a **probability limit (確率極限)** of Z_n.

$$\text{plim } Z_n = \theta.$$

(b) Let $\hat{\theta}_n$ be an estimator of parameter θ.

If $\hat{\theta}_n$ converges to θ in probability, we say that $\hat{\theta}_n$ is a consistent estimator of θ.

3. A General Case of **Chebyshev's Inequality:**

For $g(X) \geq 0$,

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

where k is a positive constant.

4. **Example:** For a random variable X, set $g(X) = (X - \mu)'(X - \mu)$, $\mathrm{E}(X) = \mu$ and $\mathrm{V}(X) = \Sigma$.

   Then, we have the following inequality:

$$P((X - \mu)'(X - \mu) \geq k) \leq \frac{\mathrm{tr}(\Sigma)}{k}.$$

   Note as follows:

$$\begin{aligned}\mathrm{E}((X - \mu)'(X - \mu)) &= \mathrm{E}\Big(\mathrm{tr}((X - \mu)'(X - \mu))\Big) = \mathrm{E}\Big(\mathrm{tr}((X - \mu)(X - \mu)')\Big) \\ &= \mathrm{tr}\Big(\mathrm{E}((X - \mu)(X - \mu)')\Big) = \mathrm{tr}(\Sigma).\end{aligned}$$

5. **Example 1 (Univariate Case):**

Suppose that $X_i \sim (\mu, \sigma^2)$, $i = 1, 2, \cdots, n$.

Then, the sample average $\overline{X}$ is a consistent estimator of μ.

Proof:

Note that $g(\overline{X}) = (\overline{X} - \mu)^2$, $\epsilon^2 = k$, $\mathrm{E}(g(\overline{X})) = \mathrm{V}(\overline{X}) = \dfrac{\sigma^2}{n}$.

Use Chebyshev's inequality.

If $n \longrightarrow \infty$,

$$P(|\overline{X} - \mu| \geq \epsilon) \leq \frac{\sigma^2}{n\epsilon^2} \longrightarrow 0, \qquad \text{for any } \epsilon.$$

That is. for any ϵ,

$$\lim_{n\to\infty} P(|\overline{X} - \mu| < \epsilon) = 1.$$

$\Longrightarrow$ **Chebyshev's inequality**

6. **Example 2 (Multivariate Case):**

Suppose that $X_i \sim (\mu, \Sigma)$, $i = 1, 2, \cdots, n$.

Then, the sample average $\overline{X}$ is a consistent estimator of μ.

Proof:

Note that $g(\overline{X}) = (\overline{X} - \mu)'(\overline{X} - \mu)$, $\epsilon^2 = k$, $\mathrm{E}(g(\overline{X})) = \mathrm{tr}\Big(\mathrm{V}(\overline{X})\Big) = \mathrm{tr}\Big(\frac{1}{n}\Sigma\Big)$.

Use Chebyshev's inequality.

If $n \longrightarrow \infty$,

$$P((\overline{X} - \mu)'(\overline{X} - \mu) \geq k) = P(|\overline{X} - \mu| \geq \epsilon) \leq \frac{\mathrm{tr}(\Sigma)}{n\epsilon^2} \longrightarrow 0, \quad \text{for any positive } \epsilon.$$

That is. for any positive ϵ, $\lim_{n\to\infty} P((\overline{X} - \mu)'(\overline{X} - \mu) < k) = 1$.

Note that $|\overline{X} - \mu| = \sqrt{(\overline{X} - \mu)'(\overline{X} - \mu)}$, which is the distance between X and μ.

$\Longrightarrow$ **Chebyshev's inequality**

7. **Some Formulas:**

Let X_n and Y_n be the random variables which satisfy $\text{plim}\, X_n = c$ and $\text{plim}\, Y_n = d$. Then,

(a) $\text{plim}\,(X_n + Y_n) = c + d$

(b) $\text{plim}\, X_n Y_n = cd$

(c) $\text{plim}\, X_n/Y_n = c/d$ for $d \neq 0$

(d) $\text{plim}\, g(X_n) = g(c)$ for a function $g(\cdot)$

$\Longrightarrow$ **Slutsky's Theorem (スルツキー定理)**

8. **Central Limit Theorem (中心極限定理)**

Univariate Case: $X_1, X_2, \cdots, X_n$ are mutually independently and identically distributed as $X_i \sim (\mu,\ \sigma^2)$.

Then,

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

which implies

$$\sqrt{n}(\overline{X} - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \longrightarrow N(0, \sigma^2).$$

Multivariate Case: $X_1, X_2, \cdots, X_n$ are mutually independently and identically distributed as $X_i \sim (\mu,\ \Sigma)$.

Then,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \longrightarrow N(0, \Sigma)$$

9. **Central Limit Theorem (Generalization)**

$X_1, X_2, \cdots, X_n$ are mutually independently and identically distributed as $X_i \sim (\mu,\ \Sigma_i)$.

Then,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu) \longrightarrow N(0, \Sigma),$$

where

$$\Sigma = \lim_{n\to\infty} \left(\frac{1}{n} \sum_{i=1}^{n} \Sigma_i \right).$$

10. **Definition:** Let $\hat{\theta}_n$ be a consistent estimator of θ.

Suppose that $\sqrt{n}(\hat{\theta}_n - \theta)$ converges to $N(0, \Sigma)$ in distribution.

Then, we say that $\hat{\theta}_n$ has an **asymptotic distribution (漸近分布)**: $N(\theta, \Sigma/n)$.

10.1 MLE: Asymptotic Properties

1. $X_1, X_2, \cdots, X_n$ are random variables with density function $f(x; \theta)$.

   Let $\hat{\theta}_n$ be a maximum likelihood estimator of θ.

   Then, under some **regularity conditions**. $\hat{\theta}_n$ is a consistent estimator of θ and the asymptotic distribution of $\sqrt{n}(\hat{\theta} - \theta)$ is given by: $N\left(0, \lim\left(\dfrac{I(\theta)}{n}\right)^{-1}\right)$.

2. **Regularity Conditions:**

   (a) The domain of X_i does not depend on θ.

(b) There exists at least third-order derivative of $f(x;\theta)$ with respect to θ, and their derivatives are finite.

3. Thus, MLE is

(i) consistent,

(ii) asymptotically normal, and

(iii) asymptotically efficient.

Proof: The log-likelihood function is given by:

$$\log L(\theta) = \log \prod_{i=1}^{n} f(X_i;\theta) = \sum_{i=1}^{n} \log f(X_i;\theta)$$

Note that the MLE $\tilde{\theta}$ satisfies:

$$\frac{\partial \log L(\tilde{\theta})}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial \log f(X_i;\tilde{\theta})}{\partial \theta} = 0.$$

X_i is a random variable.

On the other hand, the integration of $L(\theta)$ with respect to $x = (x_1, x_2, \cdots, x_n)$ is one, because $L(\theta)$ is a joint distribution of x_1, x_2, $\cdots$, x_n. Therefore, we have:

$$\int L(\theta)\mathrm{d}x = 1.$$

Taking the first-derivative of the above equation on both sides with respect to θ, we obtain:

$$\int \frac{\partial L(\theta)}{\partial \theta}\mathrm{d}x = 0,$$

which is rewritten as:

$$\int \frac{\partial L(\theta)}{\partial \theta}\mathrm{d}x = \int \frac{\partial \log L(\theta)}{\partial \theta} L(\theta)\mathrm{d}x = \mathrm{E}\Big(\frac{\partial \log L(\theta)}{\partial \theta}\Big) = 0.$$

Taking the derivative with respective θ, again (the second-derivative of $\int L(\theta)\mathrm{d}x = 1$

on both sides with respect to θ), we have:

$$\int \frac{\partial^2 \log L(\theta)}{\partial\theta^2} L(\theta)\mathrm{d}x + \int \frac{\partial \log L(\theta)}{\partial\theta} \frac{\partial \log L(\theta)}{\partial\theta'} L(\theta)\mathrm{d}x = 0,$$

which is rewritten as follows:

$$-\int \frac{\partial^2 \log L(\theta)}{\partial\theta^2} L(\theta)\mathrm{d}x = \int \frac{\partial \log L(\theta)}{\partial\theta} \frac{\partial \log L(\theta)}{\partial\theta'} L(\theta)\mathrm{d}x.$$

That is, we can derive the following:

$$-\mathrm{E}\Big(\frac{\partial^2 \log L(\theta)}{\partial\theta\partial\theta'}\Big) = \mathrm{E}\Big(\frac{\partial \log L(\theta)}{\partial\theta} \frac{\partial \log L(\theta)}{\partial\theta'}\Big) = \mathrm{V}\Big(\frac{\partial \log L(\theta)}{\partial\theta}\Big) \equiv I(\theta),$$

where the second equality holds because of $\mathrm{E}\Big(\frac{\partial \log L(\theta)}{\partial\theta}\Big) = 0$.

$I(\theta)$ is called Fisher's information matrix (or simply, information matrix).

Thus, the first-derivative of $L(\theta)$ is distributed as mean zero and variance $I(\theta)$, i.e.,

$$\frac{\partial \log L(\theta)}{\partial\theta} = \sum_{i=1}^{n} \frac{\partial \log f(X_i;\theta)}{\partial\theta} \sim (0, I(\theta)).$$

Note that we do not know the distribution of the first-derivative of $L(\theta)$, because we do not specify functional form of $f(\cdot)$

Using the central limit theorem (generalization) shown above, asymptotically we obtain the following distribution:

$$\frac{1}{\sqrt{n}} \frac{\partial \log L(\theta)}{\partial \theta} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{\partial \log f(X_i; \theta)}{\partial \theta} \longrightarrow N(0, \Sigma),$$

where $\Sigma = \lim_{n\to\infty} \Big(\frac{1}{n} I(\theta)\Big)$.

Let $\tilde{\theta}$ be the maximum likelihood estimator.

Linearizing $\dfrac{\partial \log L(\tilde{\theta})}{\partial \theta}$ around $\tilde{\theta} = \theta$, we obtain:

$$0 = \frac{1}{\sqrt{n}} \frac{\partial \log L(\tilde{\theta})}{\partial \theta} \approx \frac{1}{\sqrt{n}} \frac{\partial \log L(\theta)}{\partial \theta} + \frac{1}{\sqrt{n}} \frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'} (\tilde{\theta} - \theta),$$

where the rest of terms (i.e., the second-order term, the third-order term, ...) are ignored, which implies that the distribution of $\dfrac{1}{\sqrt{n}}\dfrac{\partial \log L(\theta)}{\partial \theta}$ is asymptotically equivalent to that of $\dfrac{1}{\sqrt{n}}\dfrac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'}(\tilde{\theta} - \theta)$.

We have already known the distribution of $\dfrac{1}{\sqrt{n}}\dfrac{\partial \log L(\theta)}{\partial \theta}$ as follows:

$$\frac{1}{\sqrt{n}}\frac{\partial \log L(\theta)}{\partial \theta} \approx -\frac{1}{\sqrt{n}}\frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'}(\tilde{\theta} - \theta) = \left(-\frac{1}{n}\frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'}\right)\sqrt{n}(\tilde{\theta} - \theta) \longrightarrow N(0, \Sigma).$$

Note as follows:

$$-\frac{1}{n}\frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'} \longrightarrow \lim_{n\to\infty}\left(\frac{1}{n}\mathrm{E}\Big(-\frac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'}\Big)\right) = \lim_{n\to\infty}\Big(\frac{1}{n}I(\theta)\Big) = \Sigma.$$

Thus, $\left(-\dfrac{1}{n}\dfrac{\partial^2 \log L(\theta)}{\partial \theta \partial \theta'}\right)\sqrt{n}(\tilde{\theta}-\theta)$ asymptotically has the same distribution as $\Sigma\sqrt{n}(\tilde{\theta}-\theta)$.

Therefore,

$$\mathrm{V}(\Sigma\sqrt{n}(\widehat{\theta} - \theta)) = \Sigma\mathrm{V}(\sqrt{n}(\widehat{\theta} - \theta))\Sigma' \longrightarrow \Sigma.$$

Note that $\Sigma = \Sigma'$. Thus, we have the asymptotic variance of $\sqrt{n}(\widehat{\theta} - \theta)$ as follows:

$$\mathrm{V}(\sqrt{n}(\widehat{\theta} - \theta)) \longrightarrow \Sigma^{-1}\Sigma\Sigma^{-1} = \Sigma^{-1}.$$

Finally, we obtain:

$$\sqrt{n}(\widehat{\theta} - \theta) \longrightarrow N(0, \Sigma^{-1}).$$