- **Convergence in Distribution (分布収束)** $X_n \longrightarrow X$, i.e., X converges in distribution to X. The distribution of X_n converges to the distribution of X as n goes to infinity.

Some Formulas

X_n and Y_n : Convergence in Probability

Z_n : Convergence in Distribution

- If $X_n \longrightarrow a$, then $f(X_n) \longrightarrow f(a)$.
- If $X_n \longrightarrow a$ and $Y_n \longrightarrow b$, then $f(X_n Y_n) \longrightarrow f(ab)$.
- If $X_n \longrightarrow a$ and $Z_n \longrightarrow Z$, then $X_n Z_n \longrightarrow aZ$, i.e., aZ is distributed with mean $\mathrm{E}(aZ) = a\mathrm{E}(Z)$ and variance $\mathrm{V}(aZ) = a^2\mathrm{V}(Z)$.

[End of Review]

8. **Weak Law of Large Numbers (大数の弱法則) — Review:**

たいすう

n random variables $X_1, X_2, \cdots, X_n$ are assumed to be mutually independently and identically distributed, where $\mathrm{E}(X_i) = \mu$ and $\mathrm{V}(X_i) = \sigma^2 < \infty$.

Then, $\overline{X} \longrightarrow \mu$ as $n \longrightarrow \infty$, which is called the **weak law of large numbers**.

$\longrightarrow$ Convergence in probability

$\longrightarrow$ Proved by Chebyshev's inequality

9. **Some Formulas of Expectaion and Variance in Multivariate Cases — Review:**

A vector of randam variavle X: $\mathrm{E}(X) = \mu$ and $\mathrm{V}(X)((X - \mu)(X - \mu)') = \Sigma$

Then, $\mathrm{E}(AX) = A\mu$ and $\mathrm{V}(AX) = A\Sigma A'$.

Proof:

$\mathrm{E}(AX) = A\mathrm{E}(X) = A\mu$

$\mathrm{V}(AX) = \mathrm{E}((AX - A\mu)(AX - A\mu)') = \mathrm{E}(A(X-\mu)(A(X-\mu))')$

$\quad = \mathrm{E}(A(X-\mu)(X-\mu)'A') = A\mathrm{E}((X-\mu)(X-\mu)')A' = A\mathrm{V}(X)A' = A\Sigma A'$

10. **Asymptotic Normality of MLE — Proof:**

The density (or probability) function of X_i is given by $f(x_i;\theta)$.

The likelihood function is: $L(\theta;x) \equiv f(x;\theta) = \prod_{i=1}^n f(x_i;\theta)$,

where $x = (x_1, x_2, \cdots, x_n)$.

MLE of θ results in the following maximization problem:

$$\max_{\theta} \ \log L(\theta;x).$$

A solution of the above problem is given by MLE of θ, denoted by $\tilde{\theta}$.

That is, $\tilde{\theta}$ is given by the θ which satisfies the following equation:

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

Replacing x_i by the underlying random variable X_i, $\dfrac{\partial \log f(X_i; \theta)}{\partial \theta}$ is taken as the ith random variable, i.e., X_i in the **Central Limit Theorem II**.

Consider applying **Central Limit Theorem II**.

In this case, we need the following expectation and variance:

$$\mathrm{E}\Big(\frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log f(X_i; \theta)}{\partial \theta}\Big) \qquad \text{and} \qquad \mathrm{V}\Big(\frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log f(X_i; \theta)}{\partial \theta}\Big).$$

Defining the variance:

$$\mathrm{V}\Big(\frac{\partial \log f(X_i; \theta)}{\partial \theta}\Big) = \Sigma_i,$$

we can rewrite the information matrix as follows:

$$I(\theta) = \mathrm{V}\Big(\frac{\partial \log L(\theta; X)}{\partial \theta}\Big) = \mathrm{V}\Big(\sum_{i=1}^{n} \frac{\partial \log f(X_i; \theta)}{\partial \theta}\Big)$$
$$= \sum_{i=1}^{n} \mathrm{V}\Big(\frac{\partial \log f(X_i; \theta)}{\partial \theta}\Big) = \sum_{i=1}^{n} \Sigma_i$$

The third equality holds when $X_1, X_2, \cdots, X_n$ are mutually independent.

Note that $\mathrm{E}\Big(\frac{\partial \log L(\theta; X)}{\partial \theta}\Big) = 0$ and $\mathrm{V}\Big(\frac{\partial \log L(\theta; X)}{\partial \theta}\Big) = I(\theta)$.

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

$$\sqrt{n}\left(\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \log f(X_i; \theta)}{\partial \theta} - \mathrm{E}\Big(\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \log f(X_i; \theta)}{\partial \theta}\Big)\right) \longrightarrow N(0, \Sigma),$$

where

$$nV\Big(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial\theta}\Big) = \frac{1}{n}V\Big(\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial\theta}\Big) = \frac{1}{n}V\Big(\frac{\partial \log L(\theta;X)}{\partial\theta}\Big)$$

$$= \frac{1}{n}I(\theta) \longrightarrow \Sigma.$$

That is,

$$\frac{1}{\sqrt{n}}\frac{\partial \log L(\theta;X)}{\partial\theta} \longrightarrow N(0,\Sigma),$$

where $X = (X_1, X_2, \cdots, X_n)$.

Now, consider replacing θ by $\tilde{\theta}$, i.e.,

$$\frac{1}{\sqrt{n}}\frac{\partial \log L(\tilde{\theta};X)}{\partial\theta},$$

which is expanded around $\tilde{\theta} = \theta$ as follows:

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

Therefore,

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

The left-hand side is rewritten as:

$$\frac{1}{\sqrt{n}}\frac{\partial^2 \log L(\theta; X)}{\partial\theta\partial\theta'} = \sqrt{n}\frac{1}{n}\frac{\partial^2 \log L(\theta; X)}{\partial\theta\partial\theta'}(\tilde{\theta} - \theta).$$

Then,

$$\sqrt{n}(\tilde{\theta} - \theta) \approx -\Big(\frac{1}{n}\frac{\partial^2 \log L(\theta; X)}{\partial\theta\partial\theta'}\Big)^{-1}\Big(\frac{1}{\sqrt{n}}\frac{\partial \log L(\theta; X)}{\partial\theta}\Big)$$

$$\longrightarrow N(0, \Sigma^{-1}\Sigma\Sigma^{-1}) = N(0, \Sigma^{-1}).$$

Note that

$$\frac{1}{n}\frac{\partial^2 \log L(\theta;X)}{\partial\theta\partial\theta'} \longrightarrow \lim_{n\to\infty}\frac{1}{n}\mathrm{E}\Big(\frac{\partial^2 \log L(\theta;X)}{\partial\theta\partial\theta'}\Big) = \Sigma,$$

and $\Big(\frac{1}{n}\frac{\partial^2 \log L(\theta;X)}{\partial\theta\partial\theta'}\Big)^{-1}\Big(\frac{1}{\sqrt{n}}\frac{\partial \log L(\theta;X)}{\partial\theta}\Big)$ has the same asymptotic distribution as $\Sigma^{-1}\Big(\frac{1}{\sqrt{n}}\frac{\partial \log L(\theta;X)}{\partial\theta}\Big)$.