

4. **Cramer-Rao Lower Bound** (クラメル・ラオの下限): $(I(\theta))^{-1}$

Suppose that an unbiased estimator of θ is given by $s(X)$.

Then, we have the following:

$$V(s(X)) \geq (I(\theta))^{-1}$$

Proof:

The expectation of $s(X)$ is:

$$E(s(X)) = \int s(x)L(\theta; x)dx.$$

Differentiating the above with respect to θ ,

$$\begin{aligned} \frac{\partial E(s(X))}{\partial \theta} &= \int s(x) \frac{\partial L(\theta; x)}{\partial \theta} dx = \int s(x) \frac{\partial \log L(\theta; x)}{\partial \theta} L(\theta; x) dx \\ &= \text{Cov} \left(s(X), \frac{\partial \log L(\theta; X)}{\partial \theta} \right) \end{aligned}$$

For simplicity, let $s(X)$ and θ be scalars.

Then,

$$\begin{aligned} \left(\frac{\partial \mathbb{E}(s(X))}{\partial \theta} \right)^2 &= \left(\text{Cov} \left(s(X), \frac{\partial \log L(\theta; X)}{\partial \theta} \right) \right)^2 = \rho^2 \text{V}(s(X)) \text{V} \left(\frac{\partial \log L(\theta; X)}{\partial \theta} \right) \\ &\leq \text{V}(s(X)) \text{V} \left(\frac{\partial \log L(\theta; X)}{\partial \theta} \right), \end{aligned}$$

where ρ denotes the correlation coefficient between $s(X)$ and $\frac{\partial \log L(\theta; X)}{\partial \theta}$, i.e.,

$$\rho = \frac{\text{Cov} \left(s(X), \frac{\partial \log L(\theta; X)}{\partial \theta} \right)}{\sqrt{\text{V}(s(X))} \sqrt{\text{V} \left(\frac{\partial \log L(\theta; X)}{\partial \theta} \right)}}.$$

Note that $|\rho| \leq 1$.

Therefore, we have the following inequality:

$$\left(\frac{\partial \mathbf{E}(s(X))}{\partial \theta}\right)^2 \leq \mathbf{V}(s(X)) \mathbf{V}\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right),$$

i.e.,

$$\mathbf{V}(s(X)) \geq \frac{\left(\frac{\partial \mathbf{E}(s(X))}{\partial \theta}\right)^2}{\mathbf{V}\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right)}$$

Especially, when $\mathbf{E}(s(X)) = \theta$,

$$\mathbf{V}(s(X)) \geq \frac{1}{-\mathbf{E}\left(\frac{\partial^2 \log L(\theta; X)}{\partial \theta^2}\right)} = (I(\theta))^{-1}.$$

Even in the case where $s(X)$ is a vector, the following inequality holds.

$$\mathbf{V}(s(X)) \geq (I(\theta))^{-1},$$

where $I(\theta)$ is defined as:

$$\begin{aligned} I(\theta) &= -\mathbb{E}\left(\frac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'}\right) \\ &= \mathbb{E}\left(\frac{\partial \log L(\theta; X)}{\partial \theta} \frac{\partial \log L(\theta; X)}{\partial \theta'}\right) = \mathbb{V}\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right). \end{aligned}$$

The variance of any unbiased estimator of θ is larger than or equal to $(I(\theta))^{-1}$.

5. Asymptotic Normality of MLE:

Let $\tilde{\theta}$ be MLE of θ .

As n goes to infinity, we have the following result:

$$\sqrt{n}(\tilde{\theta} - \theta) \longrightarrow N\left(0, \left(\lim_{n \rightarrow \infty} \left(\frac{I(\theta)}{n}\right)\right)^{-1}\right) \quad \text{or} \quad N(0, \Sigma^{-1}),$$

where it is assumed that $\lim_{n \rightarrow \infty} \left(\frac{I(\theta)}{n}\right) \equiv \Sigma$.

That is, when n is large, we have the following distribution:

$$\sqrt{n}(\tilde{\theta} - \theta) \sim N\left(0, \left(\frac{I(\theta)}{n}\right)^{-1}\right).$$

That is, $\tilde{\theta}$ is approximately distributed as follows:

$$\tilde{\theta} \sim N\left(\theta, (I(\theta))^{-1}\right).$$

Suppose that $s(X) = \tilde{\theta}$.

When n is large, $V(s(X))$ is approximately equal to $(I(\theta))^{-1}$.

Practically, we utilize the following approximated distribution:

$$\tilde{\theta} \sim N\left(\theta, (I(\tilde{\theta}))^{-1}\right).$$

Then, we can obtain the significance test and the confidence interval for θ

6. **Central Limit Theorem:** Let X_1, X_2, \dots, X_n be mutually independently distributed random variables with mean $E(X_i) = \mu$ and variance $V(X_i) = \sigma^2 < \infty$ for $i = 1, 2, \dots, n$.

Define $\bar{X} = (1/n) \sum_{i=1}^n X_i$.

Then, the central limit theorem is given by:

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

Note that $E(\bar{X}) = \mu$ and $V(\bar{X}) = \sigma^2/n$.

That is,

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

Note that $E(\bar{X}) = \mu$ and $nV(\bar{X}) = \sigma^2$.

In the case where X_i is a vector of random variable with mean μ and variance $\Sigma < \infty$, the central limit theorem is given by:

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

Note that $E(\bar{X}) = \mu$ and $nV(\bar{X}) = \Sigma$.

7. **Central Limit Theorem II:** Let X_1, X_2, \dots, X_n be mutually independently distributed random variables with mean $E(X_i) = \mu$ and variance $V(X_i) = \sigma_i^2$ for $i = 1, 2, \dots, n$.

Assume:

$$\sigma^2 = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sigma_i^2 < \infty.$$

Define $\bar{X} = (1/n) \sum_{i=1}^n X_i$.

Then, the central limit theorem is given by:

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

i.e.,

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

Note that $E(\bar{X}) = \mu$ and $nV(\bar{X}) \longrightarrow \sigma^2$.

In the case where X_i is a vector of random variable with mean μ and variance Σ_i , the central limit theorem is given by:

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

where $\Sigma = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Sigma_i < \infty$.

Note that $E(\bar{X}) = \mu$ and $nV(\bar{X}) \longrightarrow \Sigma$.

[Review of Asymptotic Theories]

- **Convergence in Probability (確率収束)** $X_n \longrightarrow a$, i.e., X_n converges in probability to a , where a is a fixed number.

- **Convergence in Distribution** (分布収束) $X_n \rightarrow X$, 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 $X_n \rightarrow a$, then $f(X_n) \rightarrow f(a)$.
- If $X_n \rightarrow a$ and $Y_n \rightarrow b$, then $f(X_n Y_n) \rightarrow f(ab)$.
- If $X_n \rightarrow a$ and $Z_n \rightarrow Z$, then $X_n Z_n \rightarrow aZ$, i.e., aZ is distributed with mean $E(aZ) = aE(Z)$ and variance $V(aZ) = a^2V(Z)$.

[End of Review]

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

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

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

→ Convergence in probability

→ Proved by Chebyshev's inequality

9. **Some Formulas of Expectation and Variance in Multivariate Cases**

— **Review:**

A vector of random variable X : $E(X) = \mu$ and $V(X) \equiv E((X - \mu)(X - \mu)') = \Sigma$

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

Proof:

$$E(AX) = AE(X) = A\mu$$

$$\begin{aligned} V(AX) &= E((AX - A\mu)(AX - A\mu)') = E(A(X - \mu)(A(X - \mu))') \\ &= E(A(X - \mu)(X - \mu)'A') = AE((X - \mu)(X - \mu)')A' = AV(X)A' = A\Sigma A' \end{aligned}$$

10. Asymptotic Normality of MLE — Proof:

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

For simplicity, consider the case where θ is a scalar.

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

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

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 , $\frac{\partial \log f(X_i; \theta)}{\partial \theta}$ is taken as the i th random variable, i.e., X_i in the **Central Limit Theorem II**.

Consider applying **Central Limit Theorem II** as follows:

$$\frac{\frac{1}{n} \sum_{i=1}^n \frac{\partial \log f(X_i; \theta)}{\partial \theta} - \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n \frac{\partial \log f(X_i; \theta)}{\partial \theta}\right)}{\sqrt{\mathbb{V}\left(\frac{1}{n} \sum_{i=1}^n \frac{\partial \log f(X_i; \theta)}{\partial \theta}\right)}} = \frac{\frac{1}{n} \frac{\partial \log L(\theta; X)}{\partial \theta} - \mathbb{E}\left(\frac{1}{n} \frac{\partial \log L(\theta; X)}{\partial \theta}\right)}{\sqrt{\mathbb{V}\left(\frac{1}{n} \frac{\partial \log L(\theta; X)}{\partial \theta}\right)}}.$$

Note that

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

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

$$\mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n \frac{\partial \log f(X_i; \theta)}{\partial \theta}\right) = \mathbb{E}\left(\frac{1}{n} \frac{\partial \log L(\theta; X)}{\partial \theta}\right) = 0,$$

and

$$\mathbb{V}\left(\frac{1}{n} \sum_{i=1}^n \frac{\partial \log f(X_i; \theta)}{\partial \theta}\right) = \mathbb{V}\left(\frac{1}{n} \frac{\partial \log L(\theta; X)}{\partial \theta}\right) = \frac{1}{n^2} I(\theta).$$

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

$$\begin{aligned}
& \frac{\frac{1}{n} \sum_{i=1}^n \frac{\partial \log f(X_i; \theta)}{\partial \theta} - \mathbb{E}\left(\frac{1}{n} \sum_{i=1}^n \frac{\partial \log f(X_i; \theta)}{\partial \theta}\right)}{\sqrt{\mathbb{V}\left(\frac{1}{n} \sum_{i=1}^n \frac{\partial \log f(X_i; \theta)}{\partial \theta}\right)}} \\
&= \frac{\frac{1}{n} \frac{\partial \log L(\theta; X)}{\partial \theta} - \mathbb{E}\left(\frac{1}{n} \frac{\partial \log L(\theta; X)}{\partial \theta}\right)}{\sqrt{\mathbb{V}\left(\frac{1}{n} \frac{\partial \log L(\theta; X)}{\partial \theta}\right)}} = \frac{\frac{1}{n} \frac{\partial \log L(\theta; X)}{\partial \theta} - \frac{1}{n} \mathbb{E}\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right)}{\frac{1}{\sqrt{n}} \sqrt{\frac{1}{n} \mathbb{V}\left(\frac{\partial \log L(\theta; X)}{\partial \theta}\right)}} \\
&= \frac{\frac{1}{n} \frac{\partial \log L(\theta; X)}{\partial \theta}}{\frac{1}{\sqrt{n}} \sqrt{\frac{1}{n} I(\theta)}} = \frac{\frac{1}{\sqrt{n}} \frac{\partial \log L(\theta; X)}{\partial \theta}}{\sqrt{\frac{1}{n} I(\theta)}} \rightarrow N(0, 1)
\end{aligned}$$

We have assumed that $\lim_{n \rightarrow \infty} \frac{1}{n} I(\theta) = \sigma^2$.

Therefore,

$$\frac{1}{\sqrt{n}} \frac{\partial \log L(\theta; X)}{\partial \theta} \longrightarrow N(0, \sigma^2)$$

In the case where θ is a vector, we have:

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

where $\Sigma = \lim_{n \rightarrow \infty} \frac{1}{n} I(\theta)$, which is a $k \times k$ matrix when θ is a $k \times 1$ vector, and $X = (X_1, X_2, \dots, X_n)$.

Now, replacing θ by $\tilde{\theta}$, consider the asymptotic distribution of

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

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:

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

Then,

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

Using the law of large number, note that

$$\begin{aligned} -\frac{1}{n} \frac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'} &\longrightarrow \lim_{n \rightarrow \infty} \frac{1}{n} \left(-\mathbb{E} \left(\frac{\partial^2 \log L(\theta; X)}{\partial \theta \partial \theta'} \right) \right) \\ &= \lim_{n \rightarrow \infty} \frac{1}{n} \left(\mathbb{V} \left(\frac{\partial \log L(\theta; X)}{\partial \theta} \right) \right) = \lim_{n \rightarrow \infty} \frac{1}{n} I(\theta) = \Sigma, \end{aligned}$$