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

Then,

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

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

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

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

Therefore, we have the following inequality:

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

i.e.,

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

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

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

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

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

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

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

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:

$$\sqrt{n}(\tilde{\theta} - \theta) \longrightarrow N\left(0, \lim_{n\to\infty}\left(\frac{I(\theta)}{n}\right)^{-1}\right),$$

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

That is, when n is large, $\tilde{\theta}$ 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, $\mathrm{V}(s(X))$ 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 $X_1, X_2, \cdots, X_n$ be mutually independently distributed random variables with mean $\mathrm{E}(X_i) = \mu$ and variance $\mathrm{V}(X_i) = \sigma^2 < \infty$ for $i = 1, 2, \cdots, n$.

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

   Then, the central limit theorem is given by:

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

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

That is,

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

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

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:

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

Note that $\mathrm{E}(\overline{X}) = \mu$ and $n\mathrm{V}(\overline{X}) = \Sigma$.

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

Assume:

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

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

Then, the central limit theorem is given by:

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

i.e.,

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

Note that $\mathrm{E}(\overline{X}) = \mu$ and $n\mathrm{V}(\overline{X}) \longrightarrow \sigma^2$.

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:

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

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

Note that $\mathrm{E}(\overline{X}) = \mu$ and $n\mathrm{V}(\overline{X}) \longrightarrow \Sigma$.

[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)** $X_n \longrightarrow X$, 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 $f(X_nY_n) \longrightarrow f(ab)$.
- If $X_n \longrightarrow a$ and $Z_n \longrightarrow Z$, then $X_nZ_n \longrightarrow aZ$, 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:**

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

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** as follows:

$$\frac{\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \frac{\partial \log f(X_i; \theta)}{\partial \theta} - \mathrm{E}\Big(\frac{1}{n}\sum_{i=1}^{n} \frac{\partial \log f(X_i; \theta)}{\partial \theta}\Big)}{\sqrt{\mathrm{V}\Big(\dfrac{1}{n}\displaystyle\sum_{i=1}^{n} \frac{\partial \log f(X_i; \theta)}{\partial \theta}\Big)}} = \frac{\dfrac{1}{n}\dfrac{\partial \log L(\theta; X)}{\partial \theta} - \mathrm{E}\Big(\dfrac{1}{n}\dfrac{\partial \log L(\theta; X)}{\partial \theta}\Big)}{\sqrt{\mathrm{V}\Big(\dfrac{1}{n}\dfrac{\partial \log L(\theta; X)}{\partial \theta}\Big)}}.$$

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:

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

and

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

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

Thus, the asymptotic distribution of

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

is given by:

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

where

$$\begin{aligned}
&n\mathrm{V}\Big(\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial\theta}\Big) = \frac{1}{n}\mathrm{V}\Big(\sum_{i=1}^{n}\frac{\partial \log f(X_i;\theta)}{\partial\theta}\Big) = \frac{1}{n}\mathrm{V}\Big(\frac{\partial \log L(\theta;X)}{\partial\theta}\Big) \\
&= \frac{1}{n}I(\theta) \longrightarrow \Sigma.
\end{aligned}$$