Mathematical Statistics II

Ch.6 Estimation

Jungsoon Choi

jungsoonchoi@hanyang.ac.kr

Table of Contents

Ch6.7 Sufficient Statistics

Sufficient Statistics

Def 6.7-1 (Factorization Thm)

Let $X_1, \ldots, X_n \sim f(x_1, \ldots, x_n|\theta)$. The statistic $Y(\mathbf{X}) = u(X_1, \ldots, X_n)$ is a *sufficient statistic* (S.S.) for θ iff

$$\underbrace{f(x_1, \ldots, x_n|\theta)}_{\text{joint dist}^{n}} = g(y|\theta)\underbrace{h(x_1, \ldots, x_n)}_{\text{not contain } \theta},$$

where $g(y|\theta)$ is the pdf or pmf of $Y(\mathbf{X}) = u(X_1, \ldots, X_n)$ and h is only a function of $(x_1, \ldots, x_n)$.

Note

If $Y(\mathbf{X})$ is a S.S. for θ and $v(Y)$ is a 1-1 function, then $v(Y)$ is also a S.S.

pop

$f(x|\theta)$ $\longrightarrow$ $X_1, \cdots, X_n \longrightarrow g(X_1, \cdots, X_n) = Y$ a r.v statistic sufficient

①

$\Downarrow$ $L(\theta)$: f^{tn} of θ

$\underline{\hat{\theta}} = g(X_1, \cdots, X_n)$: statistic

estimator

②

$$L(\theta) = f(x_1, \cdots, x_n | \theta) = f_Y(y|\theta)$$

pmf (pdf) of Y

* Sufficient Statistic (S.S) for θ is a f^{tn} of $X_1, \cdots, X_n$ and is enough statistic to estimate θ.

Thus, SS for θ instead of all $X_1, \cdots, X_n$ is enough to estimate θ.

$Y \xrightarrow{1-1 f^{tn}} v(Y)$: S.S.

$\sum X_i$: S.S for θ

$\hookrightarrow$ $\frac{\sum X_i}{n} = \bar{X}$: SS for θ

$\bar{X}^3$: "

$\vdots$

Example 6.7-1

Let $X_1, \ldots, X_n \sim \mathrm{Poi}(\lambda)$ (iid). Find the S.S. for λ.

Example 6.7-2

Let $X_1, \ldots, X_n \sim \mathrm{N}(\mu, 1)$ (iid). Find the S.S. for μ.

$$f(x_1,\cdots,x_n|\theta) = g(y|\theta)\cdot h(x_1,\cdots,x_n)$$

$$\Rightarrow Y = u(X_1,\cdots,X_n) \ : \ \text{S.S for } \theta$$

$$f(x_1,\cdots,x_n|\lambda) = \prod_{i=1}^{n} f(x_i|\lambda) = \prod_{i=1}^{n} \frac{\lambda^{x_i}\cdot e^{-\lambda}}{x_i!}$$

$$= \frac{\lambda^{\Sigma x_i} e^{-n\lambda}}{\prod_{i=1}^{n}(x_i!)} = \left\{\lambda^{n\bar{x}} e^{-n\lambda}\right\}\left\{\frac{1}{\prod_{i=1}^{n} x_i!}\right\}$$

$$= \underbrace{\left\{\frac{(n\lambda)^{\Sigma x_i} e^{-n\lambda}}{(\Sigma x_i)!}\right\}}_{\Sigma X_i \sim \text{pmf of } Poi(n\lambda)}\left\{\frac{(\Sigma x_i)!}{\prod_{i=1}^{n} x_i! \cdot n^{\Sigma x_i}}\right\}$$

$\therefore \sum_{i=1}^{n} X_i$ is a S.S for λ

2). $X_i \sim N(\mu, 1)$

$$f(x_1,\cdots,x_n|\mu) = \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}} e^{-\frac{(x_i-\mu)^2}{2}} = (2\pi)^{-\frac{n}{2}}\cdot e^{-\frac{\Sigma(x_i-\mu)^2}{2}}$$

$$= (2\pi)^{-\frac{n}{2}}\cdot \exp\left\{-\frac{\Sigma(x_i-\bar{x}+\bar{x}-\mu)^2}{2}\right\}$$

$$= (2\pi)^{-\frac{n}{2}} \exp\left\{-\frac{1}{2}\sum_{i=1}^{n}(x_i-\bar{x})^2 - \frac{1}{2}\underbrace{\sum_{i=1}^{n}(\bar{x}-\mu)^2}_{= n(\bar{x}-\mu)^2}\right\}$$

$$= \underbrace{\exp\left\{-\frac{(\bar{x}-\mu)^2}{2/n}\right\}\frac{1}{\sqrt{2\pi\frac{1}{n}}}}_{\bar{X} \sim N(\mu, \frac{1}{n})} \times \exp\left\{-\frac{\Sigma(x_i-\bar{x})^2}{2}\right\}(2\pi)^{-\frac{n}{2}+\frac{1}{2}}\frac{1}{\sqrt{n}}$$

$$-\Sigma(x_i-\bar{x})(\bar{x}-\mu)$$

$$-(\bar{x}-\mu)\Sigma(x_i-\bar{x}) = 0.$$

$\therefore \bar{X}$ is S.S for μ

ΣX_i & $\bar{X}^3$ are also S.S for μ

But $\bar{X}^2$ is not S.S for μ

pop $f(x|\theta)$ $\Rightarrow$ $X_1, \cdots, X_n \Rightarrow Y = u(X_1, \cdots, X_n)$: S.S for θ

$\quad\quad\quad\quad \| \quad\quad$ if v is a 1-1 f^{tn},

$\quad\quad\quad\quad \hat{\theta}$ $\leftarrow$ $\quad\quad$ $v(Y)$ is also a ss for θ

How to find a S.S for θ ?

1) Factorization Thm

: $f(x_1, \cdots, x_n|\theta) = g(y|\theta)h(x_1, \cdots, x_n)$, $Y = u(X_1, \cdots, X_n)$: S.S for θ

2) Conditional Dist^n — S.S check.

Given a statistic, $Y = u(X_1, \cdots, X_n)$, $\dfrac{f(x_1, \cdots, x_n|\theta)}{g(y|\theta)}$ $h(x_1, \cdots, x_n)$. does not depend on θ (just f^{tn} of x_i).

$\Rightarrow Y = u(X_1, \cdots, X_n)$: S.S for θ

only f^{tn} of $x_1, \cdots, x_n$

$\Rightarrow Y$: S.S for θ.

e.g). $X_i \sim N(\mu, 1)$.

$f(x_1, \cdots, x_n|\mu)$

$$= \underbrace{\exp\left(-\frac{(\bar{x}-\mu)^2}{2/n}\right)\frac{1}{\sqrt{2\pi\frac{1}{n}}}}_{\text{pdf of } \bar{X} \sim N(\mu, \frac{1}{n})} \times \underbrace{\exp\left\{-\frac{\Sigma(x_i-\bar{x})^2}{2}\right\}(2\pi)^{-\frac{n}{2}+\frac{1}{2}}\frac{1}{\sqrt{n}}}_{f^{tn} \text{ of } x_i}$$

$\therefore \bar{X}$ is a S.S for μ.

$\bar{X} \sim N(\mu, \frac{1}{n})$

3) In exponential family,

$\sum_{i=1}^{n} K(X_i)$: S.S for θ.

Note

Let $X_1, \ldots, X_n \sim f(x_1, \ldots, x_n|\theta)$ and $Y(\mathbf{X}) = u(X_1, \ldots, X_n)$ be a statistic with the pdf or pmf $g(y|\theta)$. Then $Y(\mathbf{X})$ is a S.S. for θ iff

$$P(X_1 = x_1, \ldots, X_n = x_n | Y = y) = \frac{f(x_1, \ldots, x_n|\theta)}{g(y|\theta)} = h(x_1, \ldots, x_n)$$

does not depend on θ.

Example 6.7-3

Let $X_1, \ldots, X_n \sim \mathrm{Ber}(p)$ (iid). Then $Y = \sum_{i=1}^{n} X_i \sim \mathrm{B}(n, p)$ is a S.S. for p.

$X_1, \cdots, X_n \sim Ber(p)$.

$f_{X_i}(x) = p^{x_i}(1-p)^{1-x_i}$

$f(x_1, \dots, x_n | p) = p^{\Sigma x_i}(1-p)^{n-\Sigma x_i}$

$Y = \sum_{i=1}^{n} X_i$

$g(y|\theta) = \binom{n}{y} p^y (1-p)^{n-y}$

$$\frac{p^{\Sigma x_i}(1-p)^{n-\Sigma x_i}}{\binom{n}{y} \cdot p^{\Sigma x_i}(1-p)^{n-\Sigma x_i}}$$

$$\frac{f(x_1, \cdots, x_n | p)}{g(y|p)} = \frac{1}{\binom{n}{y}}$$

: does not depend on p.

$\therefore \Sigma x_i$ s.s for p.

Sufficient Statistics in Exponential family

Definition of Exponential family

A family of pdfs or pmfs is called an *exponential family* if it can be expressed as

$$f(x|\theta) = \exp\left[K(x)p(\theta) + S(x) + q(\theta)\right],$$

where $K(x)$ and $S(x)$ are functions of observation x and $p(\theta)$ and $q(\theta)$ are functions of parameter θ. The range of x does not depend on θ.

In uniform distribution, range of x depend on parameter

Examples

- $X \sim \mathrm{Poi}(\lambda)$

K(X)

$$f(x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!} = \exp\left[x \ln\lambda - \ln x! - \lambda\right]$$

- $X \sim \mathrm{N}(\mu, 1)$

$$f(x|\mu) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2}\right]$$
$$= \exp\left[x\mu - \frac{x^2}{2} - \frac{\mu^2}{2} - 0.5\ln(2\pi)\right]$$

K(X) ⇒ ΣX_i is S.S for μ

예로 $\bar{X} \to \Sigma X_i$

- $X \sim \mathrm{Ber}(p)$

$$f(x|p) = p^x(1-p)^{1-x} = \exp\left[x \ln\left(\frac{p}{1-p}\right) + \ln(1-p)\right]$$

ΣX_i

Thm 6.7-1

If $X_1, \ldots, X_n$ are random variables with a pdf or pmf of the exponential family such as

$$f(x|\theta) = \exp\left[K(x)p(\theta) + S(x) + q(\theta)\right],$$

then $Y = \sum_{i=1}^{n} K(X_i)$ is a S.S for θ.

$\exp^{K(x)p(\theta)} \cdot \exp^{S(x)} \cdot \exp^{q(\theta)}$

$\exp^{K(x)p(\theta) + q(\theta)} \cdot \exp^{S(x)}$.

(pf) The joint pdf(or pmf) of $X_1, \ldots, X_n$ is given

$$
\begin{aligned}
f(x_1, \ldots, x_n|\theta) &= \Pi_{i=1}^{n} f(x_i|\theta) \\
&= \exp\left[p(\theta)\sum_{i=1}^{n} K(x_i) + \sum_{i=1}^{n} S(x_i) + nq(\theta)\right] \\
&= \exp\left[p(\theta)\sum_{i=1}^{n} K(x_i) + nq(\theta)\right] \exp\left[\sum_{i=1}^{n} S(x_i)\right].
\end{aligned}
$$

Thus, $\sum_{i=1}^{n} K(x_i)$ is a S.S for θ.

$$f(x|\theta) = \exp\left(x\log\left(\frac{\theta}{1-\theta}\right) + \log(1-\theta)\right) = e^{x\log\frac{\theta}{1-\theta}} \cdot e^{\log(1-\theta)}$$

$$\prod_{i=1}^{n} f(x|\theta) = \exp\left(\log\left(\frac{\theta}{1-\theta}\right)\Sigma x + n\log(1-\theta)\right)$$

$$e^{\Sigma x\log\frac{\theta}{1-\theta}} \cdot e^{n\log(1-\theta)}$$

$Poi(\lambda)$

$$\frac{e^{-\lambda}\lambda^{x}}{x!} \qquad \frac{e^{-\lambda n}\cdot\lambda^{\Sigma x_i}}{\prod_{i=1}^{n} x_i!} \Rightarrow \frac{(n\lambda)^{\Sigma x_i}\cdot e^{-\lambda n}}{(\Sigma x_i)!} \cdot \frac{(\Sigma x_i)!}{\prod_{i=1}^{n}(x_i!)\, n^{\Sigma x_i}}$$

$$\Downarrow$$

$$\exp[\ln e^{-\lambda} + \ln\lambda^{x} - \ln x!]$$

$$\parallel$$

$$\exp[x\ln\lambda - \lambda - \ln x!] \Rightarrow \exp[\ln\lambda\,\Sigma x - n\lambda - \ln\prod_{i=1}^{n} x_i!]$$

Example 6.7-4

Let $X_1, \ldots, X_n \sim \mathrm{Exp}(\theta)$ with the pdf

$$f(x|\theta) = \frac{1}{\theta} e^{-x/\theta} = \exp\left[x \left(-\frac{1}{\theta} \right) - \ln \theta \right].$$

Find the S.S. for θ.

Note

If there is a unique MLE for the parameter θ, then the MLE $(\hat{\theta})$ is a function of S.S.

Let $Y = u(x_1, \ldots, x_n)$ be a S.S for θ.

$$L(\theta) = f(x_1, \ldots, x_n|\theta) = g(u(x_1, \ldots, x_n)|\theta)h(x_1, \ldots, x_n)$$
$$\frac{d \log [L(\theta)]}{d\theta} = \frac{d \log [g(u(x_1, \ldots, x_n)|\theta)]}{d\theta} = 0$$

Thus, the unique MLE of θ is the function of $u(x_1, \ldots, x_n)$.

$$L(\theta) = f(x_1, \cdots, x_n | \theta) = g(y|\theta) h(x_1, \cdots, x_n)$$

$$\ell(\theta) = \ln L(\theta) = \ln g(y|\theta) + \ln h(x_1, \cdots, x_n).$$

$$\frac{\partial \ell(\theta)}{\partial \theta} = \frac{\partial \ln g(y|\theta)}{\partial \theta} = 0$$

$\hat{\theta}_{MLE}$

$X_i \sim Ber(p) \quad i = 1, \cdots, n \quad iid.$

$$f(x_i|p) = p^{x_i}(1-p)^{1-x_i}$$

$$L(p) = f(x_1, \cdots, x_n | p) = p^{\sum_{i=1}^{n} x_i}(1-p)^{n - \sum_{i=1}^{n} x_i}$$

$$\ell(p) = \Sigma x_i \ln p + (n - \Sigma x_i) \ln(1-p).$$

$$\frac{\partial \ln(p)}{\partial p} = \frac{\Sigma x_i}{p} + \frac{-(n - \Sigma x_i)}{1-p} = 0$$

$$\Sigma x_i - p\Sigma x_i - pn + p\Sigma x_i = 0$$

$$\therefore \hat{p} = \frac{\Sigma x_i}{n} \Rightarrow \hat{p}_{MLE} = \frac{\Sigma X_i}{n}$$

$Y = \sum_{i=1}^{n} X_i \quad$: s.s for p.

$\sim Bin(n, p).$

$$f_Y(y) = \binom{n}{y} p^y (1-p)^{n-y}$$

$$\ln f_Y(y) = y \ln p + (n-y) \ln(1-p).$$

$$\frac{\partial \ln f_Y(y)}{\partial p} = \frac{y}{p} - \frac{n-y}{1-p} = 0$$

$$y - yp = np - py$$

$$\therefore p = \frac{y}{n} \Rightarrow \hat{p}_{MLE} = \frac{Y}{n} = \frac{\Sigma X_i}{n}$$

$X_i \sim N(\mu, 1)$.

$\sum X_i$: s.s for μ

$$Y = \sum_{i=1}^{n} X_i \sim N(n\mu, n).$$

$$\bar{X} = \frac{\sum X_i}{n} \sim N\left(\mu, \frac{1}{n}\right).$$

Joint Sufficient Statistics

Definition

Let $X_1, \ldots, X_n \sim f(x_1, \ldots, x_n|\theta_1, \theta_2)$. The statistics $Y_1 = u_1(X_1, \ldots, X_n)$ and $Y_2 = u_2(X_1, \ldots, X_n)$ are the *joint sufficient statistics* for θ_1 and θ_2 iff

$$f(x_1, \ldots, x_n|\theta_1, \theta_2) = g(y_1, y_2|\theta_1, \theta_2)h(x_1, \ldots, x_n).$$

* For the multiple parameters $\underset{\sim}{\theta}$ (vector of parameter)
the exponential family is

$$f(x \mid \underset{\sim}{\theta}) = \exp\left[\sum_{j=1}^{k} w_j(\underset{\sim}{\theta}) K_j(x) + q(\underset{\sim}{\theta}) + S(x)\right]$$

where # of parameters = k (in general $\leq k$)

$\Rightarrow \sum_{i=1}^{n} K_1(x_i), \sum_{i=1}^{n} K_2(x_i), \cdots, \sum_{i=1}^{n} K_k(x_i)$ are joint S.S for $\underset{\sim}{\theta}$

Example 6.7-5

Let $X_1, \ldots, X_n \sim \mathrm{N}(\theta_1 = \mu, \theta_2 = \sigma^2)$ (iid). Find the joint S.S. for $\theta_1 = \mu$ and $\theta_2 = \sigma^2$.

$X_1, \cdots, X_n \sim N(\mu, \sigma^2)$.

$$f(x|\underset{\sim}{\theta}) = \exp\left[\left(-\frac{1}{2\theta_2}\right)x^2 + \left(\frac{\theta_1}{\theta_2}\right)x - \frac{\theta_1^2}{2\theta_2} + \quad\right]$$

$\Rightarrow \left(\sum_{i=1}^{n} X_i, \sum_{i=1}^{n} X_i^2\right)$ are joint S.S for θ_1 & θ_2.

$\left(\bar{X}, S^2 = \frac{\Sigma X_i^2 - (\Sigma X_i)^2/n}{n-1}\right)$ are "

$$f(x|\underset{\sim}{\theta}) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$= \exp\left[-\frac{1}{2}\ln 2\pi - \frac{1}{2}\ln\sigma^2 - \frac{(x-\mu)^2}{2\sigma^2}\right]$$

$$- \frac{x^2}{2\sigma^2} + \frac{\mu}{\sigma^2}x - \frac{\mu^2}{2\sigma^2}$$

$$= \exp\left[\left(-\frac{1}{2\theta_2}\right)x^2 + \frac{\theta_1}{\theta_2}x - \frac{\theta_1^2}{2\theta_2} - \frac{1}{2}\ln\theta_2 - \frac{1}{2}\ln 2\pi\right]$$

$\Rightarrow \left(\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2\right)$ S.S for $\underset{\sim}{\theta}$.

$\Rightarrow \left(\bar{X}, \frac{\Sigma X_i^2 - (\Sigma X_i)^2/n}{n-1} = S^2\right)$. "

$$S^2 = \frac{\Sigma(X_i - \bar{X})^2}{n-1} = \frac{\Sigma X_i^2 - (\Sigma X_i)^2/n^2}{n-1}$$

$$-2\bar{X}\,\Sigma X_i \qquad \Sigma\bar{X}^2$$

$$-\cancel{2}n\bar{X}^2 \qquad n\bar{X}^2$$

$$\frac{\Sigma X_i^2 - n\bar{X}^2}{n-1}$$

$$\Sigma X_i^2 - \frac{(\Sigma X_i)^2}{n}$$

MSE (mean squared error)

MSE (mean squared error)

$$
\begin{aligned}
\mathrm{MSE}(\hat{\theta}) &= E[(\hat{\theta} - \theta)^2] \\
&= Var(\hat{\theta}) + \{Bias(\hat{\theta})\}^2
\end{aligned}
$$

: Given the estimator $\hat{\theta}$ for θ, evaluate the performance of the estimator $\hat{\theta}$.

* MSE

: Given the estimator $\hat{\theta}$ for θ, we evaluate the performance of the estimator $\hat{\theta}$

$$
\begin{aligned}
MSE(\hat{\theta}) &= E[(\hat{\theta}-\theta)^2] = E[\{\hat{\theta} - E[\hat{\theta}] + E[\hat{\theta}] - \theta\}^2] \\
&= E[\{\hat{\theta} - E[\hat{\theta}]\}^2] + E[\{\underbrace{E[\hat{\theta}] - \theta}_{\text{constant}}\}^2] + 2E[\{\underbrace{\hat{\theta} - E[\hat{\theta}]}_{f^{tn} \text{ of } X_1, \cdots, X_n}\}\{\underbrace{E[\hat{\theta}] - \theta}_{\text{constant} \Rightarrow f^{tn} \text{ of } \theta}\}] \\
&= Var[\hat{\theta}] + \{E[\hat{\theta}] - \theta\}^2 + 2\underbrace{E[\hat{\theta} - E[\hat{\theta}]]}_{E[\hat{\theta}] - E[\hat{\theta}] = 0}\{E[\hat{\theta}] - \theta\} \\
&= Var[\hat{\theta}] + \{Bias(\hat{\theta})\}^2
\end{aligned}
$$

cf). $E[\hat{\theta}] = \theta \Rightarrow \hat{\theta}$: UE of θ

$Bias(\hat{\theta}) = E[\hat{\theta}] - \theta$

To minimize MSE,

Var ↓ , Bias ↓

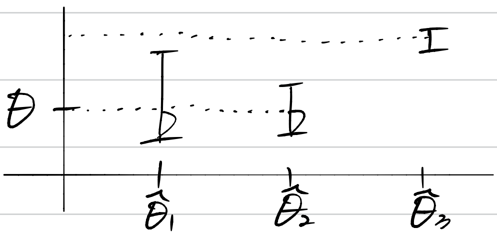
But Var ↓ ⇒ Bias ↑

Bias ↓ ⇒ Var ↑

* Idea.

① Find the UE of θ (Bias = 0).

② $\hat{\theta}$ among UEs whose variance ↓



$Bias(\hat{\theta})$ term is more important than $Var[\hat{\theta}]$

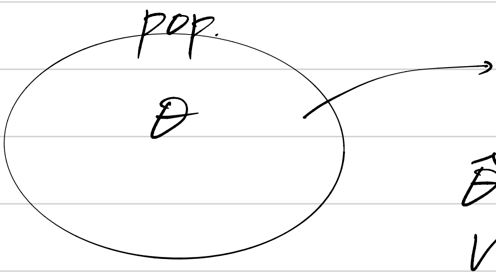
* MSE

$MSE(\hat{\theta}) = Var(\hat{\theta}) + \{Bias(\hat{\theta})\}^2$

$\Rightarrow$ UMVUE for θ

(Uniformly minimum variance unbiased estimator).

Best estimator for θ



$X_1, \cdots, X_n$

$\hat{\theta}_{1,UE}$

$Var[\hat{\theta}_{1,UE}] = MSE(\hat{\theta}_{1,UE})$.

better U.E for θ.

$MSE(\hat{\theta}_{2,UE}) = Var(\hat{\theta}_{2,UE})$

$\hat{\theta}_{2,UE}$: better UE for θ rather than $\hat{\theta}_{1,UE}$

Y_1 : S.S for θ $\quad E[\hat{\theta}_{1,UE} \mid Y_1]$

Rao-Blackwell Thm

Thm 6.7-2 (Rao-Blackwell Thm)

Let $X_1, \ldots, X_n \sim f(x|\theta)$ be random variables. Suppose that Y_1 is a S.S for θ and Y_2 is an unbiased estimator of θ. Define $u(y_1) = E[Y_2|y_1]$. Then, existing UE

- $u(Y_1)$ is an unbiased estimator of θ.
- $\mathrm{Var}[u(Y_1)] \leq \mathrm{Var}(Y_2)$.

: Key Thm. to improve U.E Y_2 of θ.

$X_i \sim f(x|\theta)$

① $Y_1 = u_1(X_1, \cdots, X_n)$: S.S for θ

② $Y_2 = u_2(X_1, \cdots, X_n)$: UE of θ ($\Leftrightarrow E[Y_2] = \theta$).

$\Rightarrow$ Define $u(Y_1) = E[Y_2 | Y_1]$: U.E of θ

pf). Define $(Y_1, Y_2) \sim g(y_1, y_2 | \theta)$.

$Y_1 \sim g_1(y_1|\theta)$.

Since Y_1 is S.S for θ, the dist^n of $Y_2 | Y_1 = y_1$ is

$$\frac{g(y_1, y_2|\theta)}{g_1(y_1|\theta)} = h(y_2|y_1) \text{ : Not depend on "}\theta\text{"}$$

① $E[u(Y_1)] = E[E[Y_2|Y_1]] = E[Y_2] = \theta$.

② $Var[u(Y_1)] = E[\{u(Y_1) - \theta\}^2]$ by ①

$$= E[\{\underbrace{E[Y_2|Y_1] - \theta}_{*}\}^2] \underset{\text{since } Y_1 \text{ : S.S for } \theta.}{=} E[\{E[\underbrace{Y_2 - \theta}_{=: Z.}|Y_1]\}^2]$$

where $* = \int y_2 \underbrace{h(y_2|y_1)}_{\text{Not depend on } \theta.} dy_2 - \theta$ $\qquad \theta \underbrace{\int h(y_2|y_1) dy_2}_{=1}$

$$= \int y_2 h(y_2|y_1) dy_2 - \int \theta h(y_2|y_1) dy_2$$

$$= \int (y_2 - \theta) h(y_2|y_1) dy_2 = E[Y_2 - \theta | Y_1]$$

Let $Z = Y_2 - \theta \Rightarrow Var[u(Y_1)] = E[\{E[Z|Y_1]\}^2]$

$\Rightarrow Var(Z|Y_1) = E[Z^2|Y_1] - \{E[Z|Y_1]\}^2 \geq 0$.

$\therefore \{E[Z|Y_1]\}^2 \leq E[Z^2|Y_1]$

$\{E[Y_2 - \theta|Y_1]\}^2 \leq E[(Y_2-\theta)^2|Y_1]$ … $**$

by

$\rightarrow Var[u(Y_1)] = E[\{E[Y_2 - \theta|Y_1]\}^2] \leq E[E\{(Y_2-\theta)^2|Y_1\}] = E[(Y_2-\theta)^2]$

$= Var(Y_2)$

cf). $E[E[X|Y]] = E[X]$

Complete Statistic

$Z = u(X_1, \cdots, X_n)$ is complete statistic iff

$E[h(Z)|\theta] \equiv 0$ for all $\theta \in \Omega$

$\Rightarrow h(Z) = 0$ for any Z.

Basu's Thm

Let Y be a complete sufficient statistic for θ. Assume that a statistic Z is free of θ. Then, Y and Z are independent.

Example

Assume that X_i follows a normal distribution with mean μ and variance σ^2, $i = 1, \cdots, n$. Show that $\overline{X}$ and S^2 are independent.

* Basu's Thm

$Y = u(X_1, \cdots, X_n)$: C.S.S for θ.

$Z = h(X_1, \cdots, X_n)$: free of θ

$\Rightarrow Y \,\&\, Z$: indep.

pf). For discrete distⁿ,

Since Z is free of θ, $P(Z=z)$ does not depend on θ.

Since Y is S.S for θ, $P(Z=z \mid Y=y)$ does not depend on θ.

To show : $Y \,\&\, Z$ are indep. $\Leftrightarrow P(Z=z \mid Y=y) = P(Z=z)$. for $\forall y$.

$\underline{P(Z=z)} = \sum_{\forall y} P(Z=z, Y=y) = \underline{\sum_{\forall y} P(Z=z \mid Y=y) \cdot P_\theta(Y=y)}$. *

Since $\sum_{\forall y} P_\theta(Y=y) = 1$,

$P(Z=z)$ is constant at that time

$P(Z=z) = P(Z=z) \cdot \sum_{\forall y} P_\theta(Y=y) = \sum_{\forall y} P(Z=z) P_\theta(Y=y)$. **

Define the statistic

$g(Y) = \underbrace{P(Z=z \mid Y=y)}_{\text{free of } \theta} - \underbrace{P(Z=z)}_{\text{free of } \theta}$ ftn of Y : free of θ.

$$\begin{aligned} E[g(Y)] &= \sum_{\forall y} g(y) P(Y=y) \\ &= \sum_{\forall y} \{ P(Z=z \mid Y=y) - P(Z=z) \} P(Y=y). \\ &= \sum_{\forall y} P(Z=z \mid Y=y) P(Y=y) - \sum_{\forall y} P(Z=z) P(Y=y) \\ &= \underbrace{P(Z=z)}_{\text{by } *} - \underbrace{P(Z=z)}_{\text{by } **} = 0 \quad \text{for } \forall \theta \end{aligned}$$

$E[g(Y)] = 0$ for $\forall \theta$.

Again $E[g(Y)] = 0$ for $\forall \theta$

Since Y is complete, $g(Y) = 0$ for $\forall y$

$\therefore P(Z=z \mid Y=y) = P(Z=z)$

$\therefore$ Y & Z are indep.

(eg). $X_i \sim N(\mu, \sigma^2)$. $i = 1, \cdots, n$ iid.

$\Rightarrow \bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

we also know that $\bar{X}$ is a S.S for μ.

$$E[h(\bar{X}) \mid \mu] = \int_{-\infty}^{\infty} h(y) \cdot \cancel{\frac{1}{\sqrt{2\pi \frac{\sigma^2}{n}}}} \exp\left\{ -\frac{n(y-\mu)^2}{2\sigma^2} \right\} dy \equiv 0 \text{ for } \forall \mu.$$

$$\Leftrightarrow \int_{-\infty}^{\infty} h(y) \underbrace{\exp\left\{ -\frac{n(y-\mu)^2}{2\sigma^2} \right\}}_{>0} dy \equiv 0 \text{ for } \forall \mu$$

$\therefore h(y) \equiv 0$ for $\forall y$.

$\therefore$ $\bar{X}$ is complete S.S for μ. Using Basu's Thm.

$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$: free of μ.

$\Rightarrow \bar{X}$ & S^2 are indep.

* Basu's Thm

Y : CSS for θ
Z : free of θ $\Big\} \Rightarrow$ Y & Z : indep.

eg). $X_i \sim N(\mu, \sigma^2)$.

$\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.

$E[h(\bar{X}) \mid \mu] = \int_{-\infty}^{\infty} h(y) \frac{1}{\sqrt{2\pi \cdot \frac{\sigma^2}{n}}} \exp\left\{-\frac{n(y-\mu)^2}{2\sigma^2}\right\} dy \equiv 0$ for $\forall \mu$.

$\Leftrightarrow \int_{-\infty}^{\infty} h(y) \cdot \underbrace{\exp\left\{-\frac{n(y-\mu)^2}{2\sigma^2}\right\}}_{>0} dy = 0$ for $\forall \mu$

$h(y) \equiv 0$. for $\forall y$.

1). $\bar{X}$: CSS for μ.

2). $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$.

S^2 : free of μ.

By Basu's Thm.

$\bar{X}$ & S^2 indep.

Fisher Information

Fisher Information

Let $X \sim f(x|\theta)$. If X is a continuous r.v., then,

$$\int f(x|\theta)dx = 1, \quad \text{for any } \theta$$

$$\frac{d}{d\theta}\int f(x|\theta)dx = \int \frac{df(x|\theta)}{d\theta}dx = 0$$

$$\int \frac{df(x|\theta)/d\theta}{f(x|\theta)} f(x|\theta)dx = 0$$

$$\int \frac{d}{d\theta}[\log f(x|\theta)] f(x|\theta)dx = 0$$

$$\frac{\partial \log f(x|\theta)}{\partial \theta} = \ell'(\theta)$$

$$E_X\left[\frac{d}{d\theta}[\log f(x|\theta)]\right] = 0$$

$E[\text{score function}] = 0$

where $\frac{d}{d\theta}[\log f(x|\theta)]$ is called the score function.

$$\int \frac{d}{d\theta}\left[\log f(x|\theta)\right] f(x|\theta)dx = 0$$

$$\int \left[\frac{d^2}{d\theta^2}\left[\log f(x|\theta)\right]\right] f(x|\theta)dx + \int \frac{d}{d\theta}\left[\log f(x|\theta)\right] \frac{d}{d\theta} f(x|\theta)dx = 0$$

$$\int \left[\frac{d^2 \log f(x|\theta)}{d\theta^2}\right] f(x|\theta)dx + \int \left[\frac{d\left[\log f(x|\theta)\right]}{d\theta}\right]^2 f(x|\theta)dx = 0$$

$\ell''(\theta)$

$$E\left[l^{''}(\theta)\right] + E\left[(l^{'}(\theta))^2\right] = 0$$

$$E\left[(l^{'}(\theta))^2\right] = -E\left[l^{''}(\theta)\right]$$

ℓ ($\mathcal{I}x$).

where $l(\theta) = \log L(\theta)$ is a log-likelihood function.

$\ell(\theta) = \log f(x|\theta)$.

Fisher Information

The Fisher Information, $\mathrm{I}(\theta)$, is defined as

$$\mathrm{I}(\theta) = \int \left[\frac{d\left[\log f(x|\theta)\right]}{d\theta} \right]^2 f(x|\theta)dx$$
$$= E\left[(l'(\theta))^2 \right] = -E\left[l''(\theta) \right]$$

where $l(\theta) = \log L(\theta)$ is a log-likelihood function.

$E[\{l'(\theta)\}^2] = -E[l''(\theta)]$
$= Var[l'(\theta)]$

$E[(l'(\theta))^2] =$ Fisher

$l'(\theta)$: score function.

$$l'(\theta) = \frac{\partial \log f(x|\theta)}{\partial \theta} \quad : \text{score function.}$$

$E[\text{score } f^{tn}] = 0 \iff E[l'(\theta)] = 0.$

$l(\theta) = \log f(x|\theta).$

$$\int l''(\theta) f(x|\theta)\,dx + \int \{l'(\theta)\}^2 f(x|\theta)\,dx = 0.$$
$$E[l''(\theta)] + E[\{l'(\theta)\}^2] = 0.$$
$$E[l''(\theta)] = -E[\{l'(\theta)\}^2].$$

Fisher information.

$$I(\theta) = E[\{l'(\theta)\}^2]$$
$$= -E[l''(\theta)] = \underline{Var[l'(\theta)]}\ ?$$

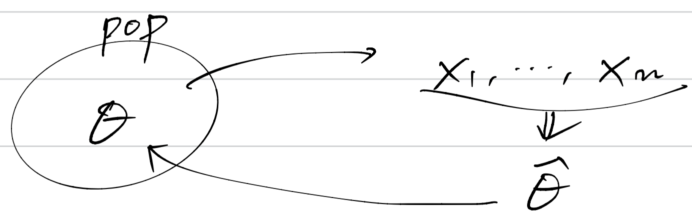
$$Var[l'(\theta)] = E[\{l'(\theta)\}^2] - \{E[l'(\theta)]\}^2$$

* Fisher information.

$$I(\theta) = E[\{l'(\theta)\}^2] = -E[l''(\theta)]$$

measuring the amount of information for θ, given X.

How much information can a simple of data provide about the parameter, θ. ?



$$L'(\theta_2) = 0$$
$$L''(\theta_2) < 0$$
$$-L''(\theta_2) > 0$$

$$L'(\theta_1) = 0$$
$$L''(\theta_1) < 0$$
$$\Downarrow$$
$$-L''(\theta_1) > 0$$

$$-L''(\theta_1) < -L''(\theta_2).$$

$$-\ell''(\theta_1) < -\ell''(\theta_2).$$

Example 1

Find the Fisher Information, $\mathrm{I}(\theta)$, for each case.

- $X \sim N(\theta, \sigma^2)$ where σ^2 is known.
- $X \sim \mathrm{Ber}(\theta)$.

$X \sim N(\theta, \sigma^2)$. σ^2 known.

$$I(\theta) = E[\{l'(\theta)\}^2] = -E[l''(\theta)]$$

$$f(x|\theta) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\theta)^2}{2\sigma^2}\right\} = L(\theta).$$

$$l(\theta) = \log L(\theta) = -\frac{1}{2}\log(2\pi\sigma^2) - \frac{(x-\theta)^2}{2\sigma^2}$$

$$l'(\theta) = \frac{\partial l(\theta)}{\partial\theta} = \frac{2(x-\theta)}{2\sigma^2} = \frac{x-\theta}{\sigma^2}$$

$$l''(\theta) = \frac{\partial^2 l(\theta)}{\partial\theta^2} = \frac{-1}{\sigma^2}$$

$$\therefore I(\theta) = -E[l''(\theta)] = \frac{1}{\sigma^2} = \frac{1}{Var(X)}$$

$X \sim Ber(\theta)$. $E[X] = \sum_{0}^{1} x \cdot \theta^x (1-\theta)^{1-x} = \theta$

$f(x|\theta) = \theta^x (1-\theta)^{1-x}$ $x = 0, 1$ $Var(X) = \theta - \theta^2$

$l(\theta) = x\log\theta + (1-x)\log(1-\theta)$ $= \theta(1-\theta)$

$$l'(\theta) = \frac{x}{\theta} + \frac{x-1}{1-\theta}$$

$$l''(\theta) = -\frac{x}{\theta^2} + \frac{1-x}{(1-\theta)^2}$$

$$-E[l''(\theta)] = -E\left[-\frac{x}{\theta^2} + \frac{1-x}{(1-\theta)^2}\right]$$

$$= E\left[\frac{x}{\theta^2}\right] - E\left[\frac{1-x}{(1-\theta)^2}\right]$$

$$= \frac{1}{\theta} - \left(\frac{1-\theta}{(1-\theta)^2}\right) = \frac{1}{\theta} - \frac{1}{1-\theta} = \frac{1-\theta-\theta}{\theta(1-\theta)} = \frac{1-2\theta}{\theta(1-\theta)}$$

Example 2

Let $X_1, \ldots, X_n \sim f(x|\theta)$ (iid) The likelihood function of $X_1, \ldots, X_n$ is $L_n(\theta) = \prod_{i=1}^{n} f(x_i|\theta)$. Then, the log-likelihood function is

$$l_n(\theta) = \log L_n(\theta) = \sum_{i=1}^{n} l(\theta|x_i)$$

The Fisher information in the random sample is

$$\begin{aligned} \mathrm{I}_n(\theta) &= E\left[(dl_n(\theta)/d\theta)^2\right] = E\left[\sum_{i=1}^{n} dl(\theta|x_i)/d\theta\right]^2 \\ &= -E\left[\sum_{i=1}^{n} d^2 l(\theta|x_i)/d\theta^2\right] \\ &= \sum_{i=1}^{n} \mathrm{I}_i(\theta) = n\mathrm{I}(\theta) \end{aligned}$$

$$l_n(\theta) = \log L_n(\theta) = \sum_{i=1}^{n} l(\theta|x_i).$$

$$I_n(\theta) = E[\{l_n'(\theta)\}^2] = E[\{\sum_{i=1}^{n} l'(\theta)\}^2]$$
$$= -E[l_n''(\theta)] = -E[\sum_{i=1}^{n} l''(\theta)] = \sum_{i=1}^{n} \underbrace{-E[l''(\theta)]}_{I_i(\theta).}$$

$$l_n'(\theta) = \frac{\partial l_n(\theta)}{\partial \theta} = \sum_{i=1}^{n} \frac{\partial l(\theta)}{\partial \theta} = \sum_{i=1}^{n} l'(\theta)$$

$$l_n''(\theta) = \frac{\partial^2 l_n(\theta)}{\partial \theta^2} = \sum_{i=1}^{n} l''(\theta)$$

$$I_n(\theta) = n \cdot I(\theta).$$

$$L_n(\theta) = \prod_{i=1}^{n} f(x_i|\theta)$$
$$l_n(\theta) = \log L_n(\theta) = \sum_{i=1}^{n} l(\theta|x_i)$$

$$I_n(\theta) = E[\{\sum_{i=1}^{n} \frac{d l_n(\theta|x_i)}{d\theta}\}^2] = -E[\sum_{i=1}^{n} \frac{d^2 l_n(\theta|x_i)}{d\theta^2}]$$
$$= -E[n \cdot l''(\theta)] = -nE[l''(\theta)]$$
$$= n \cdot I(\theta).$$

Cramer-Rao Inequality

Cramer-Rao Inequality

Cramer-Rao Inequality

Let $Y = u(X_1, \ldots, X_n)$ be an U.E of θ. Then

$$Var(Y) \geq \frac{1}{\mathrm{I}_n(\theta)} \quad \text{for any } \theta$$

where $\frac{1}{\mathrm{I}_n(\theta)}$ is called the Cramer-Rao lower bound.

$\hat{\theta}$: estimator of θ

$MSE(\hat{\theta}) = Var(\hat{\theta}) + \{Bias(\hat{\theta})\}^2$

1) $Bias(\hat{\theta}) = 0$. UE of θ

2) Smallest $Var(\hat{\theta}_{UE})$

$\Rightarrow$ $\hat{\theta}_{UE}$ is the best U.E of θ

$$Var(\hat{\theta}_{UE}) = MSE(\hat{\theta}_{UE}) \geq \frac{1}{I_n(\theta)}$$

Best UE $\hat{\theta}_{UE}$ for θ s.t $\underline{Var(\hat{\theta}_{UE})} = \frac{1}{I_n(\theta)}$

$\qquad \qquad \qquad \qquad \qquad \qquad \qquad \overset{\shortparallel}{MSE}(\hat{\theta}_{UE})$.

$Y = u(X_1, \cdots, X_n)$: UE of θ

Since Y is UE of θ, $E[Y] = E[u(X_1, \cdots, X_n)] = \theta$

Assume X_i $i = 1, \cdots, n$ are continuous r.Vs.

$$\int \cdots \int u(x_1, \cdots, x_n) f(x_1, \cdots, x_n | \theta)\, dx_1 \cdots dx_n = \theta$$

Take a derivate w.r.t θ.

$$\int \cdots \int \left[u(x_1, \cdots, x_n) \frac{\partial \log f(x_1, \cdots, x_n | \theta)}{\partial \theta} \right] f(x_1, \cdots, x_n | \theta)\, dx_1 \cdots dx_n = 1$$

$=: Z$ (score f^{tn}).

$E[Z] = 0$, $E[YZ] = 1$, $E[Y] = \theta$

Also, $E[Z^2] = Var[Z] + \{E[Z]\}^2 = Var[Z] = I_n(\theta)$.

$\Rightarrow$ $Cov(Y, Z) = 1$. $\qquad E[(Y-\theta)Z] = E[YZ - \theta Z]$

$\qquad \qquad \qquad \qquad = 1 - \theta E(Z) = 1$.

$$\{Corr(Y, Z)\}^2 = \frac{\{Cov(Y, Z)\}^2}{Var(Y)\, Var(Z)} \leq 1.$$

$$\frac{1}{Var(Y)\, I_n(\theta)} \leq 1$$

$$\therefore Var(Y) \geq \frac{1}{I_n(\theta)}$$

Efficient Estimator (UMVUE)

Let $Y = u(X_1, \ldots, X_n)$ be an U.E of θ. If

$$Var(Y) = \frac{1}{\mathrm{I}_n(\theta)} \quad \text{for any } \theta$$

then, Y is called an efficient estimator of θ. Y is also called a best unbiased estimator or a uniformly minimum variance unbiased estimator (UMVUE) of θ.

Efficiency

Let $Y = u(X_1, \ldots, X_n)$ be an U.E of θ. The efficiency of Y is given by

$$\frac{1/\mathrm{I}_n(\theta)}{Var(Y)} \leq 1$$

$Y = u(X_1, \ldots, X_n)$: UE of θ

vs.

Best UE of θ

Best

$$\frac{MSE(\hat{\theta}_{UE})}{MSE(Y)} = \frac{1/I_n(\theta)}{Var(Y)} \leq 1.$$

Example 3

Let $X_1, \ldots, X_n \sim N(\mu, \theta)$ (iid) and μ is known.

- Find the U.E of θ.
- Find the Cramer-Rao lower bound.
- Compute the efficiency of the U.E of θ.

$X_1, \cdots, X_n \sim N(\mu, \theta)$, μ : known.

$$f(x_1, \cdots, x_n \mid \theta) = \left(\frac{1}{\sqrt{2\pi\theta}}\right)^n \exp\left\{-\frac{\Sigma(x_i-\mu)^2}{2\theta}\right\} = L(\theta).$$

① $$l(\theta) = \log L(\theta) = -\frac{n}{2}\log(2\pi\theta) - \frac{\Sigma(x_i-\mu)^2}{2\theta}$$

$$\frac{\partial l(\theta)}{\partial\theta} = -\frac{n}{2\theta} + \frac{\Sigma(x_i-\mu)^2}{2}\left(\frac{1}{\theta^2}\right) = 0$$

$$\therefore \hat{\theta}_{MLE} = \frac{\Sigma(X_i-\mu)^2}{n} : \text{MLE of } \theta$$

we know that $\frac{\Sigma(X_i-\bar{X})^2}{\theta} \sim \chi^2(n-1)$

$$E\left[\frac{\Sigma(X_i-\bar{X})^2}{\theta}\right] = n-1 \Rightarrow E\left[\frac{\Sigma(X_i-\bar{X})^2}{n-1}\right] = \theta.$$

$$Var\left[\frac{\Sigma(X_i-\bar{X})^2}{\theta}\right] = 2(n-1).$$

$$\hat{\theta}_{UE} = \frac{\Sigma(X_i-\bar{X})^2}{n-1} : \text{UE of } \theta$$

② $$\frac{\partial^2 l(\theta)}{\partial\theta^2} = \frac{n}{2\theta^2} - \frac{\Sigma(x_i-\mu)^2}{\theta^3}$$

$$I_n(\theta) = E\left[\frac{\Sigma(X_i-\mu)^2}{\theta^3} - \frac{n}{2\theta^2}\right] = \frac{n\theta}{\theta^3} - \frac{n}{2\theta^2} = \frac{n}{2\theta^2}$$

$\Rightarrow$ C-R low bound : $\frac{1}{I_n(\theta)} = \frac{2\theta^2}{n}$

③ $$\frac{1/I_n(\theta)}{Var(\hat{\theta}_{UE})} = \frac{2\theta^2/n}{2\theta^2/(n-1)} = \frac{n-1}{n} \leq 1$$

* Since $var\left(\frac{\Sigma(X_i-\bar{X})^2}{\theta}\right) = 2(n-1)$,

$$Var(\hat{\theta}_{UE}) = Var\left(\frac{\Sigma(X_i-\bar{X})^2}{n-1}\right) = \frac{2\theta^2}{n-1}$$

Limiting Distributions of MLE

Limiting Distributions of MLE

1) CLT
2) convergence

기말 시험에는 X.

Limiting Distributions of MLE

Let $X_1, \ldots, X_n \sim f(x|\theta)$ (iid) and $\hat{\theta}$ be the MLE of θ.

$$\hat{\theta} \sim \mathrm{N}\left(\theta, \frac{1}{n\mathrm{I}(\theta)}\right)$$

Thus, MLE $\hat{\theta}$ is asymptotically efficient.

MLE $\hat{\theta}$ s.t $\frac{\partial l_n(\hat{\theta})}{\partial \theta} = l_n'(\hat{\theta}) = 0$ where $l_n(\theta) = \log L_n(\theta)$.

By Taylor's series expanded about θ,

$$\frac{\partial l_n(\theta)}{\partial \theta} + (\hat{\theta} - \theta) \frac{\partial^2 l_n(\theta)}{\partial \theta^2} \approx 0$$

$$\Rightarrow \hat{\theta} - \theta = \frac{l_n'(\theta)}{-l_n''(\theta)} = \frac{\sum Z_i}{\sum Z_i^*}$$

Let $Z_i = \frac{\partial l(\theta|X_i)}{\partial \theta} = l'(\theta)$ & $Z_i^* = \frac{-\partial^2 l(\theta|X_i)}{\partial \theta^2}$

$\Rightarrow E[Z_i] = 0$, $Var[Z_i] = I(\theta)$ $E[Z_i^*] = I(\theta)$.

$E[\sum Z_i] = 0$, $Var[\sum Z_i] = n \cdot I(\theta)$

By CLT, $\frac{\sum Z_i - 0}{\sqrt{n \cdot I(\theta)}} = \frac{\sum Z_i}{\sqrt{n I(\theta)}} \xrightarrow{n \to \infty} N(0,1)$

By convergency thm, $\frac{1}{n} \sum Z_i^* \xrightarrow{P} E[Z_i^*] = I(\theta)$.

$$\frac{\hat{\theta} - \theta}{\sqrt{\frac{1}{n \cdot I(\theta)}}} = \sqrt{n I(\theta)} \left[\hat{\theta} - \theta\right] = \sqrt{n I(\theta)} \frac{\sum Z_i}{\sum Z_i^*}$$

$$= \frac{\sum Z_i}{\sqrt{n \cdot I(\theta)}} \times \frac{n I(\theta)}{\sum Z_i^*}$$

$$= \frac{\sum Z_i - 0}{\sqrt{n I(\theta)}} \times \frac{I(\theta)}{\frac{1}{n} \sum Z_i^*} \longrightarrow N(0,1)$$

by CLT

$\xrightarrow{P} 1$

$$\frac{\hat{\theta}_{MLE} - \theta}{\sqrt{\frac{1}{n I(\theta)}}} \xrightarrow{n \to \infty} N(0,1).$$