Corollary 1

Under the assumptions of Theorem 1, if $Y = u(X_1, \dots, X_n)$ is an unbiased estimator of θ ($k(\theta) = \theta$), then

$$\operatorname{Var} Y \geqslant \frac{1}{nI(\theta)}.$$

Example 3

$X_1, \dots, X_n$ i.i.d $X_1 \sim b(1, \theta)$, $\frac{1}{nI(\theta)} = \frac{\theta(1-\theta)}{n}$

MLE $\hat{\theta} = \bar{X}$, $E\bar{X} = \theta$, $\operatorname{Var} \bar{X} = \frac{\theta(1-\theta)}{n}$

The variance of $\bar{X}$ attains the Cramér-Rao lower bound.

Definition 3

Under the assumptions (R0) - (R4), if $Y = u(X_1, \dots, X_n)$ is an unbiased estimator of a parameter θ, the number

$$e_Y = \frac{\frac{1}{nI(\theta)}}{\operatorname{Var} Y} = \frac{1}{nI(\theta) \operatorname{Var} Y} \qquad e_Y \in [0, 1]$$

is called the efficiency of that estimator.

If $e_Y = 1$ it is said that the estimator Y is efficient.

Example 4

$X_1, \dots, X_n$ i.i.d $X_i \sim \operatorname{Pois}(\theta)$, $\theta > 0$, MLE $\hat{\theta} = \bar{X}$, $E\bar{X} = \theta$, $\operatorname{Var} \bar{X} = \frac{\theta}{n}$

$P(X = x) = \frac{\theta^x e^{-\theta}}{x!}$

We have

$$\frac{\partial \log f(x, \theta)}{\partial \theta} = \frac{\partial}{\partial \theta}\left(x \log \theta - \theta - \log x!\right) = \frac{x}{\theta} - 1$$

$$I(\theta) = E\left[\frac{\partial \log f(X, \theta)}{\partial \theta}\right]^2 = E\left[\frac{X}{\theta} - 1\right]^2 = \frac{1}{\theta^2} E[X - \theta]^2 = \frac{1}{\theta^2} \sigma_X^2 = \frac{\theta}{\theta^2} = \frac{1}{\theta}$$

$$e_Y = \frac{1}{\left(n \frac{1}{\theta}\right) \frac{\theta}{n}} = 1$$

$Y = \bar{X}$ - efficient estimator of θ.

Example 5

$X_1, \dots, X_n$ i.i.d. $X_i \sim Beta(\theta, 1)$, $f(x, \theta) = \theta x^{\theta - 1} \mathbb{1}_{(0,1)}(x)$, $\theta > 0$

$$\log f(x, \theta) = \log \theta + (\theta - 1) \log x$$

$$\frac{\partial \log f(x, \theta)}{\partial \theta} = \frac{1}{\theta} + \log x$$

x^{θ}

$$\frac{\partial^2 \log f(x, \theta)}{\partial \theta^2} = -\frac{1}{\theta^2}$$

Thus $I(\theta) = \frac{1}{\theta^2}$.

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f(x_i, \theta) = n \log \theta + (\theta - 1) \sum_{i=1}^{n} \log x_i$$

$$\ell'(\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} \log(x_i) = 0 \Rightarrow \hat{\theta} = -\frac{n}{\sum_{i=1}^{n} \log X_i}$$

MLE of θ.

$$\ell''(\theta) = -\frac{n}{\theta^2} < 0$$

Let $Y_i = -\log X_i$, $i = 1, \dots, n$.

$$F_Y(x) = P(Y \leq x) = P(-\log X \leq x) = P(X \geq e^{-x}) =$$

$$1 - (e^{-x})^{\theta} = 1 - e^{-x\theta} \qquad Exp\left(\frac{1}{\theta}\right) \overset{d}{=} \Gamma\left(1, \frac{1}{\theta}\right)$$

$$W = \sum_{i=1}^{n} Y_i = -\sum_{i=1}^{n} \log X_i \sim \Gamma\left(n, \frac{1}{\theta}\right)$$

Fact

$$E W^k = \frac{(n+k-1)!}{\theta^k (n-1)!} \quad \text{for } k > -n.$$

Then,

$$E\hat{\theta} = n E[W^{-1}] = n \frac{(n-2)!}{\theta^{-1}(n-1)!} = \theta \frac{n}{n-1}$$

$$E\hat{\theta}^2 = n^2 E[W^{-2}] = n^2 \frac{(n-3)!}{\theta^{-2}(n-1)!} = \theta^2 \frac{n^2}{(n-1)(n-2)}$$

As a result,

$$\operatorname{Var}\hat{\theta} = \mathbb{E}\hat{\theta}^2 - (\mathbb{E}\hat{\theta})^2 = \theta^2 \frac{n^2}{(n-1)(n-2)} - \theta^2 \frac{n^2}{(n-1)^2} = \theta^2 \frac{n^2(n-1) - n^2(n-2)}{(n-1)^2(n-2)} =$$

(10)

$$\theta^2 \frac{n^2}{(n-1)^2(n-2)},$$

and

$$e_{\hat{\theta}} = \frac{1}{nI(\theta)\operatorname{Var}\hat{\theta}} = \frac{1}{n\cdot\frac{1}{\theta^2}\cdot\theta^2\frac{n^2}{(n-1)^2(n-2)}} = \frac{(n-1)^2(n-2)}{n^3} < 1$$

but $\frac{(n-1)^2(n-2)}{n^3} \to 1$

$\hat{\theta}$ is not efficient, but is asymptotically efficient.

Assumption (Additional Regularity Conditions)

(R5) The pdf $f(x,\theta)$ is three times differentiable as a function of θ. Further, for all $\theta \in \Theta$, there exists a constant c and a function $M(x)$ such that

$$\left| \frac{\partial^3 \log f(x,\theta)}{\partial\theta^3} \right| \leq M(x) \quad \text{and} \quad \mathbb{E}_{\theta_0}[M(X)] < +\infty$$

for all $\theta_0 - c < \theta < \theta_0 + c$ and all x in the support of X.

Theorem 2

Assume that $X_1, \dots, X_n$ are i.i.d with pdf $f(x,\theta_0)$, for $\theta_0 \in \Theta$ such that the regularity conditions (R0)-(R5) are satisfied. Suppose that Fisher information satisfies $0 < I(\theta_0) < +\infty$. Then any consistent sequence $\{\hat{\theta}_n\}_{n\in\mathbb{N}}$ of solutions of the equation $\frac{dL(\theta)}{d\theta} = \frac{dL(\theta, \underline{X}_n)}{d\theta} = 0$ satisfies

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N\left(0, \frac{1}{I(\theta_0)}\right).$$

Definition 4

Let $X_1, \ldots, X_n$ be i.i.d with the pdf $f(x, \theta)$. Suppose $\hat{\theta}_{1n} = \hat{\theta}_{1n}(X_1, \ldots, X_n)$ is an estimator of θ_0 such that $\sqrt{n}(\hat{\theta}_{1n} - \theta_0) \xrightarrow{\mathcal{D}} N(0, \sigma^2_{\hat{\theta}_1})$.

(i) The number

$$e(\hat{\theta}_1) = \frac{\frac{1}{I(\theta_0)}}{\sigma^2_{\hat{\theta}_1}}$$

is called the asymptotic efficiency of $\hat{\theta}_{1n}$.

(ii) If $e(\hat{\theta}_1) = 1$, it is said that $\hat{\theta}_{1n}$ is asymptotically efficient.

(iii) Suppose $\hat{\theta}_{2n} = \hat{\theta}_{2n}(X_1, \ldots, X_n)$ is an estimate of θ_0 such that $\sqrt{n}(\hat{\theta}_{2n} - \theta_0) \xrightarrow{\mathcal{D}} N(0, \sigma^2_{\hat{\theta}_2})$.

The number

$$e(\hat{\theta}_1, \hat{\theta}_2) = \frac{\sigma^2_{\hat{\theta}_2}}{\sigma^2_{\hat{\theta}_1}}$$

is called the asymptotic relative efficiency of $\hat{\theta}_{1n}$ with respect to $\hat{\theta}_{2n}$.

Example 6

$X_i = \theta + e_i$, $i = 1, \ldots, n$, $e_1, \ldots, e_n$ i.i.d

(i) $e_1 \sim$ Laplace distribution $\frac{1}{2}e^{-|x|}$

MLE of θ is $\hat{\theta}_{1n} = Me\{X_1, \ldots, X_n\} = Q_2$, $I(\theta_0) = 1$.

Also $\sqrt{n}(\hat{\theta}_{1n} - \theta_0) \xrightarrow{\mathcal{D}} N(0, 1)$.

Let $\hat{\theta}_{2n} = \bar{X}_n$

CLT implies

$$\sqrt{n}(\hat{\theta}_{2n} - \theta_0) \xrightarrow{\mathcal{D}} N(0, \sigma^2),$$

where $\sigma^2 = \mathrm{var}\, X_1 = \mathrm{Var}(e_1 + \theta) = \mathrm{Var}\, e_1 = E e_1^2 = \int_{-\infty}^{+\infty} z^2 \frac{1}{2} e^{-|z|} dz$

$$= \int_0^\infty z^2 e^{-z} dz = \Gamma(3) = 2.$$

Thus, $e(Q_2, \bar{X}) = \frac{2}{1} = 2.$

The sample median is twice as efficient as the sample mean (asymptotically).

(ii) $e_i \sim N(0,1)$.

$$\sqrt{n}(\hat{\theta}_{1n} - \theta_0) \xrightarrow{\mathcal{D}} N(0, \tfrac{\pi}{2}), \qquad \frac{\pi}{2} = \frac{1}{[2f(0)]^2}$$

$$\sqrt{n}(\hat{\theta}_{2n} - \theta_0) \sim N(0,1),$$

$$e(Me, \bar{X}) = \frac{1}{\frac{\pi}{2}} = \frac{2}{\pi} \approx 0.636 \approx \frac{1}{1.57}$$

$\bar{X}$ is 1.57 times more efficient than Q_2.

<u>Corollary 3</u>

Under the assumptions of Theorem 2, suppose $g(x)$ is a continuous function of x which is differentiable at θ_0 such that $g'(\theta_0) \neq 0$. Then

$$\sqrt{n}\left(g(\hat{\theta}_n) - g(\theta_0)\right) \xrightarrow{\mathcal{D}} N\left(0, \frac{[g'(\theta_0)]^2}{I(\theta_0)}\right).$$

3. Numerical finding of MLEs (Newton's method)

$\hat{\theta}^{(0)}$ - initial guess, $\hat{\theta}^{(1)} = \hat{\theta}^{(0)} - \frac{l'(\hat{\theta}^{(0)})}{l''(\hat{\theta}^{(0)})}$, and so on ~~etc~~

<u>Example 1</u>

$X_1, \dots, X_n$ i.i.d. $f(x, \theta) = \frac{\exp\{-(x-\theta)\}}{[1 + \exp\{-(x-\theta)\}]^2}$ $\quad x \in \mathbb{R}$, $\theta \in \mathbb{R}$

<u>No explicit form</u>