$$I(\theta) = -\mathbb{E}\left[-\frac{X}{\theta^2} - \frac{1-X}{(1-\theta)^2}\right] = \frac{\theta}{\theta^2} + \frac{1-\theta}{(1-\theta)^2} = \frac{1}{\theta} + \frac{1}{1-\theta} = \frac{1}{\theta(1-\theta)}$$

Example 2.

$X_1, \ldots, X_n$ i.i.d. such that

$$X_i = \theta + e_i, \quad i = 1, \ldots, n, \qquad \text{(location model)}$$

where $e_1, \ldots, e_n$ are i.i.d ~~with~~ $e_i \sim f(x)$.

Then $X_i \sim f_X(x, \theta) = f(x-\theta)$

Assume that f satisfies the regularity conditions. Then

$$I(\theta) = \int_{-\infty}^{+\infty} \left(\frac{f'(x-\theta)}{f(x-\theta)}\right)^2 f(x-\theta)\,dx = \left\{\begin{matrix} z = x-\theta \\ dz = dx \end{matrix}\right\} = \int_{-\infty}^{+\infty} \left(\frac{f'(z)}{f(z)}\right)^2 f(z)\,dz.$$

Hence, in the location model, the Fisher information does not depend on θ.

Suppose that X_i has the Laplace distribution, i.e., $f(x, \theta) = \frac{1}{2} e^{-|x_i - \theta|}$

Since

$$X_i = \theta + e_i,$$

$$e_i \sim f(z_i) = \tfrac{1}{2} e^{-|z_i|}.$$

Furthermore, $f'(z) = -\frac{1}{2} e^{-|z|} \operatorname{sgn}(z)$.

Therefore,

$$I(\theta) = \int_{-\infty}^{+\infty} \left(\frac{f'(z)}{f(z)}\right)^2 f(z)\,dz = \int_{-\infty}^{+\infty} f(z)\,dz = 1$$

Remark 2.

If $X_1, \ldots, X_n$ are i.i.d, $X_i \sim f(x, \theta)$ and $I(\theta)$ is the Fisher information of X_1, then $nI(\theta)$ is the Fisher information of the sample.

Proof

$$\operatorname{Var}\left(\frac{\partial \log L(\theta, \underline{X})}{\partial\theta}\right) = \operatorname{Var}\left(\sum_{i=1}^{n} \frac{\partial \log f(X_i, \theta)}{\partial\theta}\right) = \sum_{i=1}^{n} \operatorname{Var}\left(\frac{\partial \log f(X_i,\theta)}{\partial\theta}\right) = nI(\theta)$$

Theorem 1 (Cramér - Rao inequality)

Let $X_1, \dots, X_n$ be i.i.d. with pdf $f(x,\theta)$, $\theta \in \Theta$.

Assume that the regularity conditions (R0)-(R4) hold.

Let $Y = u(X_1, \dots, X_n)$ be a statistic with mean $\mathbb{E}Y = \mathbb{E}[u(X_1,\dots,X_n)] = k(\theta)$. Then

$$\operatorname{Var} Y \geq \frac{[k'(\theta)]^2}{nI(\theta)}$$

Proof

I) (Consider the) continuous case

1. We have $k(\theta) = \mathbb{E}Y = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} u(x_1, \dots, x_n) f(x_1, \theta) \cdot \dots \cdot f(x_n, \theta)\, dx_1 \dots dx_n$

2. By the above

$$k'(\theta) = \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} u(x_1,\dots,x_n) \left[\sum_{i=1}^{n} \frac{1}{f(x_i,\theta)} \frac{\partial f(x_i,\theta)}{\partial\theta}\right] f(x_1,\theta) \cdot \dots \cdot f(x_n,\theta)\, dx_1 \dots dx_n$$

$$= \int_{-\infty}^{+\infty} \dots \int_{-\infty}^{+\infty} u(x_1,\dots,x_n) \left[\sum_{i=1}^{n} \frac{\partial \log f(x_i;\theta)}{\partial\theta}\right] f(x_1,\theta) \dots f(x_n,\theta)\, dx_1 \dots dx_n$$

3. Define random variable $Z = \sum_{i=1}^{n} \frac{\partial \log f(X_i, \theta)}{\partial\theta}$

4. Then $\mathbb{E}Z = 0$, $\operatorname{Var} Z = nI(\theta)$.

5. Moreover, $k'(\theta) = \mathbb{E}[Y \cdot Z] = \mathbb{E}Y \cdot \mathbb{E}Z + \rho\sigma_Y\sigma_Z \overset{(4)}{=} \rho\sigma_Y\sqrt{nI(\theta)}$,

where $\rho = \operatorname{corr}(Y, Z)$

6. Thus

$$\rho = \frac{k'(\theta)}{\sigma_Y\sqrt{nI(\theta)}}.$$

7. Since $\rho^2 \leq 1$, we have

$$\frac{[k'(\theta)]^2}{\sigma_Y^2\, nI(\theta)} \leq 1 \iff \operatorname{Var} Y \geq \frac{[k'(\theta)]^2}{nI(\theta)}$$

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

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

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

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

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

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

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

Example 5!

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

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

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

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

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

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

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

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

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

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

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

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

Fact

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

Then,

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

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

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 $X_1, \ldots, X_n$ 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

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