Cramer Rao theorem

Minimum Variance Unbiased Estimator (MVUE)

$\hat{\theta}$ is an estimator of θ.

$$\text{MSE}= E(\theta-\hat{\theta})^2 = b^2(\hat{\theta})+\text{var}(\hat{\theta})$$

For a good estimator $b^2(\hat{\theta})$ should be '0' (i.e. unbiased) and $\text{var}(\hat{\theta})$ should be as small as possible. Such an estimator is called minimum variance unbiased estimator. $\hat{\theta}$ is an MVUE if

$$E(\hat{\theta}) = \theta$$

and $\quad Var(\hat{\theta}) \le Var(\hat{\theta}^*)$

where $\hat{\theta}^*$ is any other unbiased estimator of θ.

Can we reduce the variance of an unbiased estimator indefinitely? The answer is given by the Cramer Rao theorem.

Suppose $\hat{\theta}$ is an unbiased estimator of random sequence $X_1,.......X_N$ which depends on a parameter θ . Let us denote the sequence by the vector

$$\mathbf{X} = \begin{bmatrix} X_1 \\ X_2 \\ \vdots \\ X_N \end{bmatrix}$$

Let $f_{\mathbf{X}}(x_1,......,x_N/\theta)$ be the joint density function which characterises $\mathbf{X}$.. This function is also called likelihood function. θ may also be random. In that case likelihood function will represent conditional joint density function.

$L(\mathbf{x}/\theta) = \ln f_{\mathbf{X}}(x_1......x_N/\theta)$ is called log likelihood function.

Statement of the Cramer Rao theorem:

If $\hat{\theta}$ is an unbiased estimator of θ, then

$$Var(\hat{\theta}) \ge \frac{1}{I(\theta)}$$

where $I(\theta) = E(\frac{\partial L}{\partial \theta})^2$ and $I(\theta)$ is a measure of average information in the random sequence and is called fisher information statistic.

The equality of CR bound holds if $\frac{\partial L}{\partial \theta} = c(\hat{\theta}-\theta)$ where c is a constant.

Proof: $\hat{\theta}$ is an unbiased estimator of θ

$$\therefore E(\hat{\theta}-\theta) = 0\,.$$

$$\Rightarrow \int_{-\infty}^{\infty} (\hat{\theta}-\theta) f_{\mathbf{X}}(\mathbf{x}/\theta) d\mathbf{x} = 0.$$

Differentiate w.r.t. θ, we get

$$\int_{-\infty}^{\infty} \frac{\partial}{\partial \theta}\{(\hat{\theta}-\theta) f_{\mathbf{X}}(\mathbf{x}/\theta)\} d\mathbf{x} = 0.$$

(since line of integration are not function of θ.)

$$= \int_{-\infty}^{\infty} (\hat{\theta}-\theta)\frac{\partial}{\partial\theta} f_{\mathbf{Y}}(\mathbf{x}/\theta)\}dy - \int_{-\infty}^{\infty} f_{\mathbf{X}}(\mathbf{x}/\theta)d\mathbf{x} = 0.$$

$$=> \quad \int_{-\infty}^{\infty} (\hat{\theta}-\theta)\frac{\partial}{\partial\theta} f_{\mathbf{X}}(\mathbf{x}/\theta)\}d\mathbf{x} = 1 \qquad (1)$$

Note that $\frac{\partial}{\partial\theta} f_{\mathbf{X}}(\mathbf{x}/\theta) = \frac{\partial}{\partial\theta}\{\ln f_{\mathbf{X}}(\mathbf{x}/\theta)\} f_{\mathbf{X}}(\mathbf{x}/\theta)$

$$=(\frac{\partial L}{\partial\theta}) f_{\mathbf{X}}(\mathbf{x}/\theta)$$

$\therefore$ from (1) $\int_{-\infty}^{\infty} (\hat{\theta}-\theta)\{\frac{\partial}{\partial\theta} L(\mathbf{x}/\theta)\} f_{\mathbf{X}}(\mathbf{x}/\theta)\}d\mathbf{x} = 1.$

$$\Rightarrow \left\{ \int_{-\infty}^{\infty} \underbrace{(\hat{\theta}-\theta)\sqrt{\mathrm{f}_{\mathbf{X}}(\mathbf{x}/\theta)}}_{a} \underbrace{\frac{\partial}{\partial\theta} L(\mathbf{x}/\theta)\sqrt{f_{\mathbf{X}}(\mathbf{x}/\theta)}}_{b} d\mathbf{x} \right\}^2 = 1. \qquad (2)$$

since $f_{\mathbf{X}}(\mathbf{x}/\theta)$ is ≥ 0.

Recall the Cauchy Schawarz Ineaquality

$$|<\mathbf{a},\mathbf{b}>|^2 \leq \|\mathbf{a}\|^2 \; \|\mathbf{b}\|^2$$

where the equality holds when $\mathbf{a} = c\mathbf{b}$ (where c is any scalar).
Applying this inequality to the L.H.S. of equation (2) we get

$$\int_{-\infty}^{\infty} (\hat{\theta}-\theta)^2 \, \mathrm{f}_{\mathbf{X}}(\mathbf{x}/\theta)\, \mathrm{dy} \quad \int_{-\infty}^{\infty} \left((\frac{\partial}{\partial\theta} L(\mathbf{x}/\theta) \right)^2 \mathrm{f}_{\mathbf{X}}(\mathbf{x}/\theta)\, \mathrm{d}\mathbf{x}$$

$$= \mathrm{var}(\hat{\theta}) \;\; \mathrm{I}(\theta)$$

$$\therefore L.H.S \leq \mathrm{var}(\hat{\theta})\, \mathrm{I}(\theta)$$

But R.H.S. = 1

$$\mathrm{var}(\hat{\theta})\, \mathrm{I}(\theta) \geq 1.$$

$$=> \mathrm{var}(\hat{\theta}) \geq \frac{1}{I(\theta)}$$

The equality will hold when

$$c(\hat{\theta}-\theta)\sqrt{f_{\mathbf{X}}(\mathbf{x}/\theta)} = \frac{\partial}{\partial\theta}\{L(\mathbf{x}/\theta)\sqrt{f_{\mathbf{X}}(\mathbf{x}/\theta)}\}$$

$$\boxed{=> \frac{\partial L(\mathbf{x}/\theta)}{\partial\theta} = c\,(\hat{\theta}-\theta)}$$

Taking the double partial derivative of the log-likelihood function get that

$$\mathrm{E}\left(\frac{\partial \mathrm{L}}{\partial\theta}\right)^2 = -\,\mathrm{E}\,\frac{\partial^2 \mathrm{L}}{\partial\theta^2}$$

If $\hat{\theta}$ satisfies CR -bound with equality, then $\hat{\theta}$ is called an <u>efficient estimator.</u>

Example

Let $X_1 ....X_N$ are iid Gaussian random sequence with known variance σ^2 and unknown mean μ.

Suppose $\hat{\mu} = \frac{1}{N}\sum_{i=1}^{N} X_i$ which is unbiased.

Find CR bound and hence show that $\hat{\mu}$ is an efficient estimator.

Likelihood function

$f_{\mathbf{X}}(x_1, x_2, ...x_N / \theta)$ will be product of individual densities (since iid)

$$\therefore f_{\mathbf{X}}(x_1, x_2, ...x_N / \theta) = \frac{1}{(\sqrt{(2\pi)^N}\ \sigma^{\mathrm{N}}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2}$$

so that $L(\mathbf{X}/\mu) = -\ln(\sqrt{2\pi})^N\ \sigma^{\mathrm{N}}\ - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i-\mu)^2$

$$\text{Now}\quad \frac{\partial L}{\partial \mu} = 0 - \frac{1}{2\sigma^2}(-2\sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{X_i} - \mu) = 0$$

$$\therefore \frac{\partial^2 L}{\partial \mu^2} = -\frac{\mathrm{N}}{\sigma^2}$$

$$\text{So that}\ \ \mathrm{E}\frac{\partial^2 L}{\partial \mu^2} = -\ \frac{\mathrm{N}}{\sigma^2}$$

$$\therefore \text{ CR Bound} = \frac{1}{I(\theta)} = \frac{1}{-\mathrm{E}\frac{\partial^2 L}{\partial \mu^2}} = \frac{1}{\frac{\mathrm{N}}{\sigma^2}} = \frac{\sigma^2}{\mathrm{N}}$$

$$\frac{\partial L}{\partial \theta} = \frac{1}{2\sigma^2}\ \sum_{\mathrm{i}=1}^{\mathrm{N}}(\mathrm{X_i} - \mu) = \frac{\mathrm{N}}{\sigma^2}\left(\sum_i \frac{X_i}{N} - \mu\right)$$

$$= \frac{N}{\sigma^2}(\hat{\mu} - \mu)$$

$$\text{Hence}\ \ -\frac{\partial L}{\partial \theta} = \mathrm{c}\,(\hat{\theta} - \theta)$$

and $\hat{\mu}$ is an efficient estimator.

Criteria for Estimation

- Maximum Likelihood
- Minimum Mean Square Error.
- Baye’s Method.
- Maximum Entropy Method.

Maximum Likelihood Estimator (MLE)

Most popular and simple.

Given a random sequence $X_1 ....X_N$ and unknown non random parameter θ upon which the joint density of $X_1 ....X_N$ depends.

$f_{\mathbf{X}}(x_1, \mathrm{x}_2, ...x_N / \theta)$ is called the likelihood function (for continuous function …., for discrete it will be joint probability fn).
$L(\mathbf{x}/\theta) = \ln f_{\mathbf{X}}(x_1, x_2, ..., x_N / \theta)$ is called log likelihood function.

The maximum likelihood estimator $\widehat{\theta}_{MLE}$ is such an estimator that

$$f_{\mathbf{X}}(x_1, x_2 ...., x_N / \hat{\theta}_{MLE}) \geq f_{\mathbf{X}}(x_1, x_2, ..., x_N / \theta), \forall \theta$$

If the likelihood function is differentiable w.r.t. θ, the $\widehat{\theta}_{MLE}$ is given by

$$\frac{\partial}{\partial \theta} f_{\mathbf{X}}\,(x_1, \ldots x_N / \theta)\Big|_{\hat{\theta}_{\mathrm{MLE}}} = 0$$

or $\dfrac{\partial \mathrm{L}(\mathbf{x} \mid \theta)}{\partial \theta}\Big|_{\hat{\theta}_{\mathrm{MLE}}} = 0$

If we have a number of unknown parameters given by $\boldsymbol{\theta} = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_N \end{bmatrix}$

Then MLE is given by a set of conditions.

.

$$\left.\frac{\partial \mathrm{L}}{\partial \theta_1}\right]_{\theta_1 = \hat{\theta}_{1\mathrm{MLE}}} = \left.\frac{\partial \mathrm{L}}{\partial \theta_2}\right]_{\theta_2 = \hat{\theta}_{2\mathrm{MLE}}} = ;.... = \left.\frac{\partial \mathrm{L}}{\partial \theta_{\mathrm{M}}}\right]_{\theta_{\mathrm{M}} = \hat{\theta}_{\mathrm{MMLE}}} = 0$$

<u>Example</u> Let $X_1 ....X_N$ are independent identically distributed sequence of $N(\mu, \sigma^2)$ distributed random variables. Find MLE for μ, σ^2.

$$f_{\mathbf{X}}(x_1, x_2, ..., x_N / \mu, \sigma^2) = \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left(\frac{x_i - \mu}{\sigma}\right)^2}$$

$$L(X / \mu, \sigma^2) = \ln f_{\mathbf{X}}(X_1, ..., X_N / \mu, \sigma^2)$$

$$= \text{- N ln } - N \ln \sqrt{2\pi} - \mathrm{N} \ln \sigma - \frac{1}{2} \sum_{\mathrm{i=1}}^{\mathrm{N}} \left(\frac{\mathrm{x_i} - \mu}{\sigma}\right)^2$$

$$\frac{\partial L}{\partial \mu} = 0 \Rightarrow \sum_{\mathrm{i=1}}^{\mathrm{N}} \left(\frac{\mathrm{x_i} - \mu}{\sigma}\right)^2 = \sum_{\mathrm{i=1}}^{\mathrm{N}} (\mathrm{x_i} - \hat{\mu}_{\mathrm{MLE}}) = 0$$

$$\frac{\partial \mathrm{L}}{\partial \sigma} = 0 = -\frac{\mathrm{N}}{\hat{\sigma}_{\mathrm{MLE}}} + \frac{\sum (\mathrm{x_i} - \hat{\mu}_{\mathrm{MLE}})^2}{\hat{\sigma}_{\mathrm{MLE}}} = 0$$

Solving we get

$$\hat{\mu}_{\mathrm{MLE}} = \frac{1}{\mathrm{N}} \sum_{\mathrm{i=1}}^{\mathrm{N}} \mathrm{x_i} \quad \text{and}$$

$$\hat{\sigma}_{\mathrm{MLE}}{}^2 = \frac{1}{\mathrm{N}} \sum_{\mathrm{i=1}}^{\mathrm{N}} \left(\mathrm{x_i} - \hat{\mu}_{\mathrm{MLE}}\right)^2$$