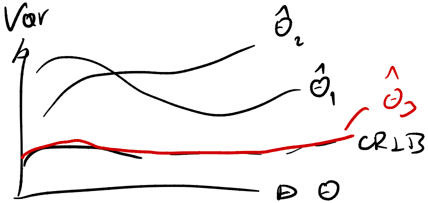
Statistical signal processing
5CTA0

Estimation theory - Cramér Rao lower bound

Cramér Rao lower bound

- Lower bound on the variance of any *unbiased* estimator



- An estimator which attains the bound is termed *efficient*
- Efficient estimators can be found by evaluating the CRLB

Cramér Rao lower bound

Cramér Rao lower bound - single parameter

The variance of any unbiased estimator $g(\mathbf{x})$ is lower bounded by

$$\mathrm{Var}[g(\mathbf{x})] \geq \frac{1}{\mathcal{I}(\theta)}, \tag{1}$$

where

$$\mathcal{I}(\theta) = \mathrm{E}\left[\left(\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta}\right)^2\right] \tag{2}$$

$$= -\mathrm{E}\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial \theta^2}\right] \tag{3}$$

is the so called Fisher information.

Regularity conditions

- Regularity condition I:

$$\frac{d}{d\theta}\int p(\mathbf{x};\theta)d\mathbf{x} = \int \frac{\partial}{\partial\theta}p(\mathbf{x};\theta)d\mathbf{x} \tag{4}$$

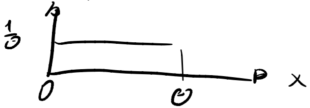
- Regularity condition II:

$$\frac{d^2}{d\theta^2}\int p(\mathbf{x};\theta)d\mathbf{x} = \int \frac{\partial^2}{\partial\theta^2}p(\mathbf{x};\theta)d\mathbf{x} \tag{5}$$

Leibniz integral rule:

$$\frac{d}{d\theta}\left(\int_{a(\theta)}^{b(\theta)} f(x,\theta)\,dx\right) = f\left(b(\theta),\theta\right)\cdot\frac{d}{d\theta}b(\theta) - f\left(a(\theta),\theta\right)\cdot\frac{d}{d\theta}a(\theta) + \int_{a(\theta)}^{b(\theta)}\frac{\partial}{\partial\theta}f(x,\theta)\,dx$$

$p(\underline{x};\theta)$



EINDHOVEN UNIVERSITY OF TECHNOLOGY

CRLB - Derivation

$$E[\hat{\theta}] = \theta \Rightarrow E[\hat{\theta}] - \theta = 0 \Rightarrow E[\hat{\theta} - \theta]$$

$$\frac{d}{d\theta}\int (\hat{\theta} - \theta)\, p(\underline{x};\theta)\, d\underline{x} = \int \frac{\partial}{\partial\theta} (\hat{\theta} - \theta)\, p(\underline{x};\theta)\, dx$$

$$= -\underbrace{\int p(\underline{x};\theta)\, dx}_{=1} + \int (\hat{\theta} - \theta) \frac{\partial}{\partial\theta} p(\underline{x};\theta)\, dx = 0$$

$$\Rightarrow \int (\hat{\theta} - \theta) \frac{\partial}{\partial\theta} p(\underline{x};\theta)\, dx = 1$$

$$\frac{d}{dx} \ln f(x) = \frac{1}{f(x)} \frac{d}{dx} f(x) \Rightarrow f(x) \frac{d}{dx} \ln f(x) = \frac{d}{dx} f(x)$$

$$\int (\hat{\theta} - \theta)\, p(\underline{x};\theta) \frac{\partial}{\partial\theta} \ln p(\underline{x};\theta)\, d\underline{x}$$

CRLB - Derivation

$$\int (\hat{\theta}-\theta)\, \underbrace{p(\underline{x};\theta)}_{\sqrt{p(\underline{x};\theta)}\sqrt{p(\underline{x};\theta)}} \frac{\partial}{\partial\theta} \ln p(\underline{x};\theta)\, d\underline{x} = 1$$

$$1 = \int \underbrace{(\hat{\theta}-\theta)\sqrt{p(\underline{x};\theta)}}_{f(\underline{x};\theta)}\ \underbrace{\frac{\partial}{\partial\theta} \ln p(\underline{x};\theta)\sqrt{p(\underline{x};\theta)}}_{g(\underline{x};\theta)}\, dx$$

$$\leq \underbrace{\int (\hat{\theta}-\overset{E[\hat{\theta}]}{\theta})^2\, p(\underline{x};\theta)\, d\underline{x}}_{Var[\hat{\theta}]} \int \left(\frac{\partial}{\partial\theta} \ln p(\underline{x};\theta)\right)^2 p(\underline{x};\theta)\, d\underline{x}$$

$$Var[\hat{\theta}] \geq E\left[\left(\frac{\partial}{\partial\theta} \ln p(\underline{x},\theta)\right)^2\right]$$

CRLB - Derivation

$$\int p(\underline{x};\theta)\,d\underline{x} = 1 \Rightarrow \frac{d}{d\theta}\int p(\underline{x},\theta)\,d\underline{x} = 0$$

$$\int \frac{\partial}{\partial\theta} p(\underline{x};\theta)\,dx = 0$$

$$\int p(\underline{x};\theta)\frac{\partial}{\partial\theta}\ln p(\underline{x};\theta)\,d\underline{x} = 0$$

$$\int \frac{\partial}{\partial\theta} p(\underline{x};\theta)\frac{\partial}{\partial\theta}\ln p(\underline{x};\theta)\,d\underline{x} + \int p(\underline{x};\theta)\frac{\partial^2}{\partial\theta^2}\ln p(\underline{x};\theta)\,d\underline{x} = 0$$

$$\int p(\underline{x},\theta)\left(\frac{\partial}{\partial\theta}\ln p(\underline{x};\theta)\right)^2 d\underline{x} = -\int p(\underline{x};\theta)\frac{\partial^2}{\partial\theta^2}\ln p(\underline{x},\theta)\,dx$$

$$E\left[\left(\frac{\partial}{\partial\theta}\ln p(\underline{x};\theta)\right)^2\right] = -E\left[\frac{\partial}{\partial^2}\ln p(\underline{x};\theta)\right]$$

Efficiency

Efficient estimator

An efficient estimator can be found if the expression $\partial \ln p(\mathbf{x};\theta)/\partial\theta$ can be expressed as

$$\frac{\partial}{\partial\theta}\ln p(\mathbf{x};\theta) = \mathcal{I}(\theta)(g(\mathbf{x})-\theta). \tag{6}$$

$$\left(\int \left[(g(\mathbf{x})-\theta)\sqrt{p(\mathbf{x};\theta)}\right]\left[\frac{\partial}{\partial\theta}\ln p(\mathbf{x};\theta)\sqrt{p(\mathbf{x};\theta)}\right] d\mathbf{x}\right)^2$$
$$\leq \int (g(\mathbf{x})-\theta)^2 p(\mathbf{x};\theta)d\mathbf{x} \int \left(\frac{\partial}{\partial\theta}\ln p(\mathbf{x};\theta)\right)^2 p(\mathbf{x};\theta)d\mathbf{x}.$$

$=$ $\quad a(\theta)\,(g(\underline{x})-\theta) = \frac{\partial}{\partial\theta}\ln p(\underline{x};\theta)$

Efficiency

$$a(\theta)(g(\underline{x})-\theta) = \frac{\partial}{\partial\theta}\ln p(\underline{x};\theta)$$

$$\frac{\partial}{\partial\theta}a(\theta)\cdot(g(\underline{x})-\theta) - a(\theta) = \frac{\partial^2}{\partial\theta^2}\ln p(\underline{x};\theta)$$

$$-\frac{\partial}{\partial\theta}a(\theta)\cdot\underbrace{(\underbrace{E[g(x)]}_{\theta}-\theta)}_{0} + a(\theta) = I(\theta)$$

$$a(\theta) = I(\theta)$$

$$\frac{\partial}{\partial\theta}\ln p(\underline{x};\theta) = I(\theta)(g(x)-\theta)$$

Efficiency - Estimation of a DC voltage

$$x_n = A + w_n, \qquad w_n \sim \mathcal{N}(0, \sigma^2)$$

$$p(\mathbf{x}; \theta) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left(-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x_n - A)^2\right)$$

$$\ln p(\mathbf{x}; A) = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x_n - A)^2$$

$$\frac{\partial}{\partial A}\ln p(\mathbf{x}; A) = \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x_n - A) = \overbrace{\frac{N}{\sigma^2}}^{I(\theta)}\Big(\underbrace{\frac{1}{N}\sum_{n=0}^{N-1} x_n}_{g(x)} - A\Big)$$

$\hat{A}_1$

$$-E\left[\frac{\partial^2}{\partial A^2}\ln p(\mathbf{x}; A)\right] = -E\left[-\frac{N}{\sigma^2}\right] = \frac{N}{\sigma^2} = I(\theta)$$

$$\mathrm{Var}[\hat{A}_1] = \frac{\sigma^2}{N}$$

CRLB for IID observations

- Joint pdf for IID observations

$$p(\mathbf{x};\theta) = \prod_{n=0}^{N-1} p(x_n;\theta) \implies \ln p(\mathbf{x};\theta) = \sum_{n=0}^{N-1} \ln p(x_n;\theta) \tag{7}$$

- Linearity of derivative and expectation

$$\begin{aligned} -\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\ln p(\mathbf{x};\theta)\right] &= -\sum_{n=0}^{N-1}\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\ln p(x_n;\theta)\right] \\ &= -N\,\mathrm{E}\left[\frac{\partial^2}{\partial\theta^2}\ln p(x_n;\theta)\right] \\ &= N i(\theta) \end{aligned} \tag{8}$$

$i(\theta)$ Fisher information of a single observation

CRLB for signals in white Gaussian noise

A

- Observation model

$$x_n = s[n;\theta] + w_n, \qquad n = 0, 1, \ldots, N-1, \tag{9}$$

- Joint pdf

$$p(\mathbf{x};\theta) = \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x_n - s[n;\theta])^2\right] \tag{10}$$

- CRLB

$$\mathrm{Var}\left[g(\mathbf{x})\right] \geq \frac{1}{\frac{1}{\sigma^2}\sum_{n=0}^{N-1}\left(\frac{\partial}{\partial\theta}s[n;\theta]\right)^2}. \tag{11}$$

$= \frac{1}{\frac{N}{\sigma^2}} = \frac{\sigma^2}{N}$

$\frac{\partial}{\partial A}$ $A = 1$

CRLB for vector parameter

Cramér Rao lower bound - vector parameter

The covariance matrix of an unbiased estimator $g(\mathbf{x})$ is lower bounded by

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} \geq \mathbf{I}^{-1}(\boldsymbol{\theta}), \qquad \underline{a} \neq 0 \qquad \underline{a}(\underline{C}_{\underline{\theta}} - \underline{I}(\theta))\underline{a} \geq 0 \tag{12}$$

where

$$\mathbf{C}_{\hat{\boldsymbol{\theta}}} = \mathrm{E}\left[(g(\mathbf{x}) - \boldsymbol{\theta})(g(\mathbf{x}) - \boldsymbol{\theta})^T\right], \tag{13}$$

and $\mathbf{I}$ is the so-called Fisher information matrix with elements

$$\begin{aligned} [\mathbf{I}(\boldsymbol{\theta})]_{i,j} &= \mathrm{E}\left[\left(\frac{\partial}{\partial\theta_i}\ln p(\mathbf{x};\boldsymbol{\theta})\right)\left(\frac{\partial}{\partial\theta_j}p(\mathbf{x};\boldsymbol{\theta})\right)\right] \\ &= -\mathrm{E}\left[\frac{\partial^2}{\partial\theta_i\partial\theta_j}\ln p(\mathbf{x};\boldsymbol{\theta})\right]. \end{aligned} \tag{14}$$