Cramér-Rao lower bound contd.

Tuesday, 9 March 2021   15:38

θ : unknown and deterministic

$\mathbf{x} = [x[0], x[1], \dots x[N-1]]^T \in \mathbb{R}^N$

$p(\mathbf{x};\theta)$: likelihood function

$x[n] = A + w[n]$
$\mathbf{x} \sim \mathcal{N}(A, \sigma^2 I)$

$\hat{\theta} = g(\mathbf{x})$

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

CRLB:

Assume that the PDF satisfies the regularity condition

(1) $E\left[\frac{\partial}{\partial\theta}\ln p(\mathbf{x};\theta)\right] = 0 \quad \forall\,\theta$

Then, variance of any unbiased estimator $\hat{\theta}$

(2) $\mathrm{var}(\hat{\theta}) \geq \dfrac{1}{-E\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right]}$ ←

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

(average curvature) $= I(\theta)$ "Fisher information"

(4) $\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta} = I(\theta)[\hat{\theta} - \theta]$ [Efficient]

$g(\mathbf{x}) = \hat{\theta}$

then $\hat{\theta}$ would be the MVU

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

Estimators that achieve the CRLB are called "efficient" estimators.

Proof (3):

$$\frac{\partial}{\partial\theta} E\left[\frac{\partial}{\partial\theta}\ln p(\mathbf{x};\theta)\right] = 0$$

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

$$\int\left[\frac{\partial^2}{\partial\theta^2}\ln p(\mathbf{x};\theta)\, p(\mathbf{x};\theta) + \frac{\partial}{\partial\theta}\ln p(\mathbf{x};\theta)\frac{\partial}{\partial\theta}p(\mathbf{x};\theta)\right]d\mathbf{x} = 0$$

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

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

$= I(\theta) \rightarrow$ scalar

"Fisher Information"

Properties:

(1) Non-negative

(2) Additive for independent observations:

$$\ln p(\mathbf{x};\theta) = \ln\prod_{n=0}^{N-1} p(x[n];\theta) = \sum_{n=0}^{N-1}\ln p(x[n];\theta)$$

$x[0] = A + w[0] \rightarrow \sigma_1^2$
$x[1] = A + w[1] \rightarrow \sigma_2^2$
$\sigma_1^2 = \sigma_2^2 = \sigma^2$

$$-E\left[\frac{\partial^2}{\partial\theta^2}\ln p(\mathbf{x};\theta)\right] = \sum_{n=0}^{N-1} -E\left[\frac{\partial^2}{\partial\theta^2}\ln p(x[n];\theta)\right]$$

IID observations: $I(\theta) = N\, i(\theta)$ ↳ Fisher information from one sample

Efficient Estimators:

$E(\hat{\theta}) = \theta \qquad \mathrm{var}(\hat{\theta}) = \frac{1}{I(\theta)}$

(*) $\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta} = I(\theta)[\hat{\theta} - \theta]$

$$E\left[\frac{\partial}{\partial\theta}\ln p(\mathbf{x};\theta)\right] = I(\theta)\left[E(\hat{\theta}) - \theta\right] = 0$$

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

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

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

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

(*) $\frac{\partial^2}{\partial\theta^2}\ln p(\mathbf{x};\theta) = \frac{\partial I(\theta)}{\partial\theta}(\hat{\theta}-\theta) - I(\theta)$

$$-E\left[\frac{\partial^2}{\partial\theta^2}\ln p(\mathbf{x};\theta)\right] = -\frac{\partial I(\theta)}{\partial\theta}\left[\underbrace{E(\hat{\theta}) - \theta}_{=0}\right] + I(\theta)$$

$$= I(\theta)$$

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

Example:

$\mathbf{x} = \mathbf{h}(\theta) + \mathbf{w}$

$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, C) \qquad C: N \times N$

$\mathbf{x} \sim \mathcal{N}(\mathbf{h}(\theta), C)$

$$p(\mathbf{x};\theta) = \frac{1}{(2\pi)^{N/2}|C|^{1/2}}\exp\left[-\frac{1}{2}(\mathbf{x}-\mathbf{h}(\theta))^T C^{-1}(\mathbf{x}-\mathbf{h}(\theta))\right]$$

$$\ln p(\mathbf{x};\theta) = k - \frac{1}{2}[\mathbf{x}-\mathbf{h}(\theta)]^T C^{-1}[\mathbf{x}-\mathbf{h}(\theta)]$$

$$\frac{\partial}{\partial\theta}\ln p(\mathbf{x};\theta) = \cancel{+\frac{1}{2}}\cdot\cancel{2}\,\underbrace{[\mathbf{x}-\mathbf{h}(\theta)]^T C^{-1}}\frac{\partial \mathbf{h}(\theta)}{\partial\theta}$$

$$\frac{\partial^2}{\partial\theta^2}\ln p(\mathbf{x};\theta) = [\mathbf{x}-\mathbf{h}(\theta)]^T C^{-1}\frac{\partial^2\mathbf{h}(\theta)}{\partial\theta^2} - \frac{\partial\mathbf{h}^T(\theta)}{\partial\theta}C^{-1}\frac{\partial\mathbf{h}(\theta)}{\partial\theta}$$

$$-E\left[\frac{\partial^2}{\partial\theta^2}\ln p(\mathbf{x};\theta)\right] = \frac{\partial\mathbf{h}^T(\theta)}{\partial\theta}C^{-1}\frac{\partial\mathbf{h}(\theta)}{\partial\theta}$$

$$\mathrm{var}(\hat{\theta}) \geq \frac{1}{\frac{\partial\mathbf{h}^T(\theta)}{\partial\theta}C^{-1}\frac{\partial\mathbf{h}(\theta)}{\partial\theta}}$$

→ The more $\mathbf{h}(\theta)$ depends on θ; smaller will be the CRLB

→ $\mathbf{h}(\theta) = $ const. $\Rightarrow$ CRLB $= \infty$

→ $\mathbf{h}(\theta) = \mathbf{h}\theta$ [linear data model]

$$\mathrm{var}(\hat{\theta}) \geq \frac{1}{\underbrace{\mathbf{h}^T C^{-1}\mathbf{h}}}$$

→ not a function of θ

→ non-linear models; CRLB depends on θ

Specialization:

$C = \sigma^2 I$

$\mathbf{h}\theta = \mathbf{1}_N A$

$$\mathrm{Var}(\hat{A}) \geq \frac{1}{\mathbf{1}^T(\sigma^{-2}I)\mathbf{1}} = \frac{\sigma^2}{N}$$

Linear model: $\mathbf{x} = \mathbf{h}\theta + \mathbf{w}$

$$\frac{\partial}{\partial\theta}\ln p(\mathbf{x};\theta) = [\mathbf{x}-\mathbf{h}\theta]^T C^{-1}\mathbf{h}$$

$$= \mathbf{h}^T C^{-1}\mathbf{x} - \theta\,\mathbf{h}^T C^{-1}\mathbf{h}$$

$$\frac{\partial}{\partial\theta}\ln p(\mathbf{x};\theta) = \underbrace{\mathbf{h}^T C^{-1}\mathbf{h}}_{I(\theta)}\left[\underbrace{(\mathbf{h}^T C^{-1}\mathbf{h})^{-1}\mathbf{h}^T C^{-1}\mathbf{x}}_{g(\mathbf{x}) = \hat{\theta}} - \theta\right]$$

$$\hat{\theta}_{MVU} = (\mathbf{h}^T C^{-1}\mathbf{h})^{-1}\mathbf{h}^T C^{-1}\mathbf{x}$$

IID: $C = \sigma^2 I$, $\mathbf{h} = \mathbf{1}$ } $\hat{\theta}_{MVU} = (\mathbf{1}^T(\sigma^{-2}I)\mathbf{1})^{-1}\mathbf{1}^T(\sigma^{-2}I)\mathbf{x}$

$$= \frac{1}{N\cancel{\sigma^{-2}}}\cdot\cancel{\sigma^{-2}}\,\mathbf{1}^T\mathbf{x}$$

$$= \frac{1}{N}\sum_{n=0}^{N-1}x[n] = \bar{x}$$

Transformation of parameters:

$$x[n] = A + w[n]$$

- Estimator of A^2 ?
- Knowing the CRLB for A, how to compute the CRLB for A^2

$$\alpha = g(\theta)$$

$$\mathrm{var}(\hat{\alpha}) \geq \frac{\left(\frac{\partial g(\theta)}{\partial\theta}\right)^2}{-E\left[\frac{\partial^2}{\partial\theta^2}\ln p(\mathbf{x};\theta)\right]}$$

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

$$\int\hat{\alpha}\, p(\mathbf{x};\theta)\, d\mathbf{x} = g(\theta)$$

$$\int\hat{\alpha}\frac{\partial}{\partial\theta}p(\mathbf{x};\theta)\, d\mathbf{x} = \frac{\partial}{\partial\theta}g(\theta)$$

$$\int\hat{\alpha}\frac{\partial}{\partial\theta}\ln p(\mathbf{x};\theta)\cdot p(\mathbf{x};\theta)\, d\mathbf{x} = \frac{\partial g(\theta)}{\partial\theta}$$

Using regularity condition: $\alpha\, E\left[\frac{\partial}{\partial\theta}\ln p(\mathbf{x};\theta)\right] = 0$

$$\int(\hat{\alpha}-\alpha)\frac{\partial}{\partial\theta}\ln p(\mathbf{x};\theta)\, p(\mathbf{x};\theta)\, d\mathbf{x} = \frac{\partial g(\theta)}{\partial\theta}$$

$$\int(\hat{\alpha}-\alpha)\sqrt{p(\mathbf{x};\theta)}\,\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\sqrt{p(\mathbf{x};\theta)}\, d\mathbf{x} = \frac{\partial g(\theta)}{\partial\theta}$$

$$\boxed{\mathrm{Var}(\hat{\alpha}) \geq \frac{\left(\frac{\partial}{\partial\theta}g(\theta)\right)^2}{E\left[\frac{\partial}{\partial\theta}\ln p(\mathbf{x};\theta)\right]^2}}$$