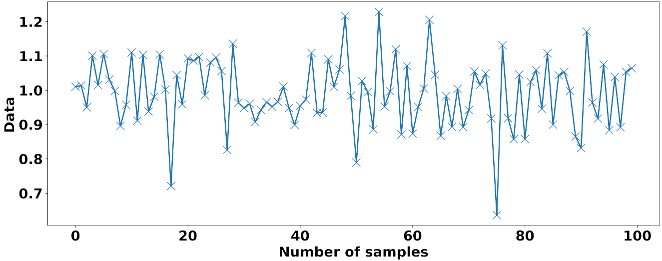
Cramér-Rao Lower Bound

Dr. Raj Thilak Rajan

Example



Consider a process e.g., a constant in noise

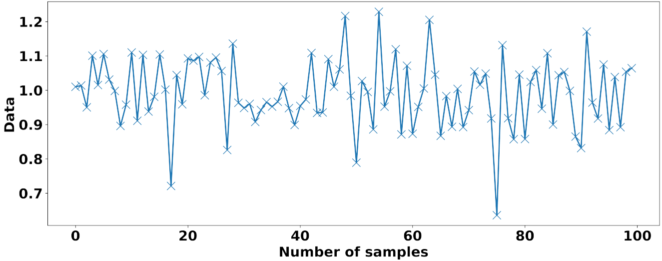
$$x[n] = A + w[n], \quad n = 0, \ldots, N-1,$$

where, we assume

- A is deterministic and *unknown*,
- $w[n]$ is a zero-mean random process with variance σ^2,
- $x[n]$ is the measured data.

How can we estimate A ?

Example



Potential estimators for A

- $\hat{A}_1 = x[0]$
- $\hat{A}_2 = \frac{1}{N}\sum_{n=0}^{N-1} x[n]$
- $\hat{A}_3 = \frac{a}{N}\sum_{n=0}^{N-1} x[n]$, for some constant a
- $\ldots$

Which estimator is *optimal* ? Estimator is also a random variable

Moments

Mean:

- $\mathbb{E}(\hat{A}_1) = \mathbb{E}(x[0]) = A$
- $\mathbb{E}(\hat{A}_2) = \mathbb{E}\left(\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right) = \frac{1}{N}\sum_{n=0}^{N-1} \mathbb{E}(x[n]) = A$
- $\mathbb{E}(\hat{A}_3) = \mathbb{E}\left(\frac{a}{N}\sum_{n=0}^{N-1} x[n]\right) = \frac{a}{N}\sum_{n=0}^{N-1} \mathbb{E}(x[n]) = aA$

Variance:

- $var(\hat{A}_1) = \sigma^2$
- $var(\hat{A}_2) = var\left(\frac{1}{N}\sum_{n=0}^{N-1} x[n]\right) = \frac{1}{N}\sum_{n=0}^{N-1} var(x[n]) = \frac{\sigma^2}{N}$
- $var(\hat{A}_3) = var\left(\frac{a}{N}\sum_{n=0}^{N-1} x[n]\right) = \frac{a^2}{N}\sum_{n=0}^{N-1} var(x[n]) = \frac{a^2\sigma^2}{N}$

Note:

- $\hat{A}_1, \hat{A}_2$ are *unbiased* estimators,
- $\hat{A}_2$ is *more efficient* than $\hat{A}_1$.

Is there an *optimal* estimator ?

Optimality criterion

Let $\hat{\theta} = g(\mathbf{x}) = [x[0], x[1], x[N-1]]$ be an estimator of θ, then

Mean square error (MSE):

$$\begin{aligned} mse(\hat{\theta}) &= \mathbb{E}\left[(\hat{\theta}-\theta)^2\right] = \mathbb{E}\left\{\left[(\hat{\theta}-(\hat{\theta})) + ((\hat{\theta})-\theta)\right]^2\right\} \\ &= \mathbb{E}\left[(\hat{\theta}-(\hat{\theta}))^2\right] + (\mathbb{E}(\hat{\theta})-\theta)^2 = var(\hat{\theta}) + (\mathbb{E}(\hat{\theta})-\theta)^2 \end{aligned}$$

Unbiased estimators: If θ is an unbiased estimator, then

$$\mathbb{E}(\hat{\theta}) = \int g(\mathbf{x})p(\mathbf{x};\theta)d\mathbf{x} = \theta \quad \text{for all } \theta,$$

where $p(\mathbf{x};\theta)$ is the probability density function. In other words, for an unbiased estimator

$$bias(\theta) = \mathbb{E}(\hat{\theta}) - \theta = 0.$$

Minimum Variance Unbiased Estimator (MVU)

- Constrain the bias of the MSE to zero, i.e., $\mathbb{E}(\hat{\theta}) = \theta$, then

$$mse(\hat{\theta}) = \mathbb{E}\left[(\hat{\theta} - \mathbb{E}(\hat{\theta}))^2\right] + (\mathbb{E}(\hat{\theta}) - \theta)^2 = \mathbb{E}\left[(\hat{\theta} - \mathbb{E}(\hat{\theta}))^2\right]$$

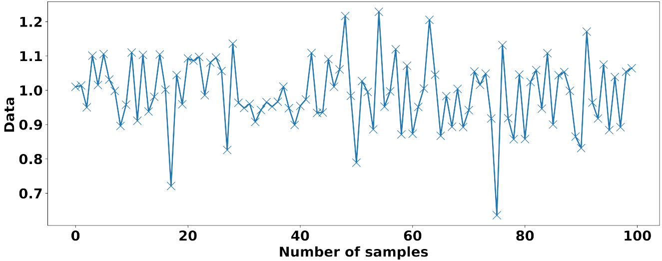
  where $\hat{\theta}$ is an unbiased estimator, and let

$$var(\hat{\theta}) \leq var(\tilde{\theta})$$

  for any other unbiased estimator $\tilde{\theta}$, then $\hat{\theta}$ is the MVU for all θ.

- *Does a MVU exist i.e., an unbiased estimator with min. variance for all θ ?*

Example



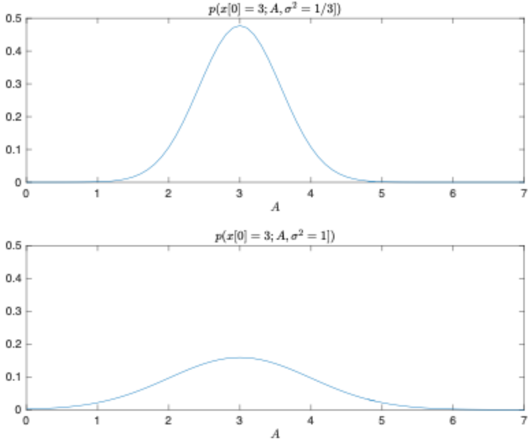
Consider the signal model

$$x[n] = A + w[n], \quad n = 0, \ldots, N-1,$$

where, we assume

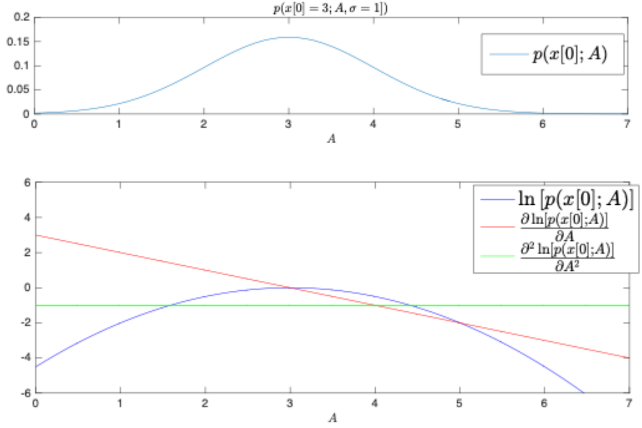
- A is deterministic and *unknown*,
- $w[n] \sim \mathcal{N}(0, \sigma^2)$,
- $x[n]$ is the measured data.

Observation (1)



- Consider a single realization: $x[0] = A + w[0]$ and the PDF $p(x[0]; A, \sigma^2)$
- Sharpness of the likelihood function determines the estimator accuracy

Observation (2)



- Measure the sharpness/curvature by $-\frac{\partial^2 \ln[p(x[0];A)]}{\partial A^2} = 1$

Score function

- The *score* function is the gradient of the log-likelihood function

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

  which indicates the steepness of the log-likelihood function.
- Mean of the score function:

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

Regularity conditions

- If $s(\mathbf{x};\theta)$ exists and is finite, and

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

  then the pdf $p(\mathbf{x};\theta)$ satisfies the following *regularity condition*

$$\mathbb{E}\left[s(\mathbf{x};\theta)\right] = \mathbb{E}\left[\frac{\partial \ln\ p(\mathbf{x};\theta)}{\partial \theta}\right] = 0, \quad \text{for all } \theta,$$

  unless the domain of the PDF for which it is nonzero depends on θ.
- If these *regularity conditions* are met, then we can estimate lower bounds on the variance of the estimator, and hopefully an MVU.

Fisher information (1)

- The variance of the score function is the *Fisher information*

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

- *Proof:* From the regularity conditions, we obtain

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

or,

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

and rearranging the terms,

$$\begin{aligned} -\int \frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2} p(\mathbf{x};\theta) d\mathbf{x} &= \int \left(\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right)^2 p(\mathbf{x};\theta) d\mathbf{x} \\ -\mathbb{E}\left[\frac{\partial^2 \ln p(\mathbf{x};\theta)}{\partial\theta^2}\right] &= \mathbb{E}\left[\left(\frac{\partial \ln p(\mathbf{x};\theta)}{\partial\theta}\right)^2\right] \end{aligned}$$

Fisher information (2)

The Fisher information is

- Non-negative, and
- Additive for *independent* observations, i.e., when

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

then

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

and for *identically* distributed observations

$$I(\theta) = Ni(\theta), \quad \text{where} \quad i(\theta) = -\mathbb{E}\left[\frac{\partial \ln p(x[n];\theta)^2)}{\partial\theta^2}\right]$$

Cramér-Rao Lower Bound theorem

- Assume the pdf $p(\mathbf{x};\theta)$ satisfies the regularity condition:

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

then the variance of any unbiased estimator $\hat{\theta}$ satisfies

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

- An estimator is *efficient* if it meets the CRLB with equality, in which case the estimator is the MVU.
- However, the converse is not necessarily true.

Finding the MVU estimator

- An unbiased estimator may be found that attains the bound for all θ iff

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

for some function g and I, then $\hat{\theta} = g(\mathbf{x})$ is an estimator with

$$\text{Mean}: \mathbb{E}(\hat{\theta}) = \theta \qquad \text{Variance}: var(\hat{\theta}) = \frac{1}{I(\theta)}.$$

Example 1(1)

$$x[n] = A + w[n] \quad n = 0, \cdots, N-1\,,$$

where $w[n] \sim \mathcal{N}(0, \sigma^2)$ is zero mean white Gaussian noise, i.e.,

$$\begin{aligned} p(\mathbf{x}; A) &= \prod_{n=0}^{N-1} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x[n]-A)^2}{2\sigma^2}\right] \\ &= \frac{1}{(2\pi\sigma^2)^{N/2}} \exp\left[-\frac{\sum_{n=0}^{N-1}(x[n]-A)^2}{2\sigma^2}\right] \end{aligned}$$

Taking the log-likelihood, we have

$$\begin{aligned} s(\mathbf{x}; \mathbf{A}) = \frac{\partial \ln \mathbf{p}(\mathbf{x}; \mathbf{A})}{\partial \mathbf{A}} &= \frac{\partial}{\partial A}\left[-\frac{1}{2\sigma^2}\sum_{n=0}^{N-1}(x[n]-A)^2\right] \\ &= \frac{1}{\sigma^2}\sum_{n=0}^{N-1}(x[n]-A) \end{aligned}$$

Example 1(2)

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

$$= \underbrace{\frac{N}{\sigma^2}}_{I(\theta)}\left(\underbrace{\frac{1}{N}\sum_{n=0}^{N-1}x[n]}_{g(\mathbf{x})} - \underbrace{A}_{\theta}\right)$$

Recollect from the CRLB theorem

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

and thus $var(\hat{A}) \geq \frac{\sigma^2}{N}$, where $\hat{A} = g(\mathbf{x})$.

CRLB for the general Gaussian model (1)

Let us assume a Gaussian distribution for the noise:

$$\mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_w) \quad \Leftrightarrow \quad p(\mathbf{w}) = \frac{1}{(2\pi)^{\frac{N}{2}} \det(\mathbf{C}_w)^{\frac{1}{2}}} \exp\left[-\frac{1}{2}\mathbf{w}^T \mathbf{C}_w^{-1} \mathbf{w}\right]$$

Then the *Gaussian model* is defined as

$$\mathbf{x} = \mathbf{h}(\theta) + \mathbf{w} \qquad \mathbf{x} \sim \mathcal{N}(\mathbf{h}(\theta), \mathbf{C}_w)$$

or,

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

CRLB for the general Gaussian model (2)

Score:

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

and

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

Fisher information:

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

CRLB:

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

CRLB for the linear Gaussian model

Consider the *linear Gaussian model*:

$$\mathbf{x} = \mathbf{h}\theta + \mathbf{w}, \qquad \mathbf{w} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}_w)$$

From CRLB for a General Gaussian model, we know

$$\frac{\partial \ln p(\mathbf{x};\theta)}{\partial \theta} = \frac{\partial \mathbf{h}^T(\theta)}{\partial \theta} \mathbf{C}_w^{-1}\big(\mathbf{x} - \mathbf{h}(\theta)\big), \qquad var(\hat{\theta}) \geq \frac{1}{\mathbf{h}^T \mathbf{C}_w^{-1} \mathbf{h}}$$

Furthermore,

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

Thus, the MVU exists and its solution reaches the CRLB:

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

Summary

Key points:

- Score function is the first derivative of the log-likelihood function w.r.t. unknown parameter
- Regularity condition are met, if the score exists, is finite and if the expectation of the score function equals zero.
- Fisher information is the covariance of the score function
- If the regularity conditions hold, then the CRLB is the inverse of the fisher information, which gives the lowest achievable bound by an unbiased estimator.
- In certain cases, the MVU can be obtained from the score function, given the CRLB.

Next session:

- Practical estimators

Assignments

Solve:

- Kay-I, Problem 3.1: Show that the regularity condition does not hold for $x[n] \sim \mathcal{U}[0, \theta]$, which are IID.
- Kay-I, Problem 3.3: Consider the data $x[n] = Ar^n + w[n]$ for $n = 0, 1, \ldots, N-1$ where $w[n]$ is WGN with variance σ. Derive the CRLB for A, and show that an efficient estimator exists and find its variance.

Derivation:

- Kay-I, 3A: Derivation of scalar Parameter CRLB
- Kay-I, 3B: Derivation of vector Parameter CRLB
- Kay-I, 3C: Derivation of general Gaussian CRLB