ECE531 Screencast 2.3: The Cramer-Rao Lower Bound for Estimating a Scalar Parameter

D. Richard Brown III

Worcester Polytechnic Institute
Theorem (The Cramer-Rao Lower Bound (scalar parameter))

Suppose that \( \hat{\theta}(y) \) is an unbiased estimator of the parameter \( \theta \) and that we have a family of densities \( \{p_Y(y; \theta); \theta \in \Lambda\} \). If

\[
E\left[ \frac{\partial \ln p(y; \theta)}{\partial \theta} \right] = 0 \text{ for all } \theta \in \Lambda
\]

then

\[
\text{var}[\hat{\theta}(Y)] \geq \frac{1}{I(\theta)}
\]

where \( I(\theta) \) is the Fisher information.

An estimator is said to be **efficient** if it achieves the CRLB. An efficient unbiased estimator is MVU. Not all MVU estimators are efficient.

The Cramer-Rao lower bound was originally described by Fisher in 1922 but was not well-known until Rao and Cramer worked on it in 1945 and 1946, respectively.

A proof of the theorem is in your textbook. The theorem only requires some calculus and the Cauchy-Schwarz inequality.
Theorem (Attainability of the CRLB (scalar parameter))

An unbiased estimator \( \hat{\theta}(y) \) can be found that attains the CRLB for all \( \theta \) if and only if

\[
\frac{\partial \ln p_Y(y; \theta)}{\partial \theta} = I(\theta)(\hat{\theta}(y) \ - \ \theta)
\]

where \( I(\theta) \) is the Fisher information.

The proof of this theorem is also in your textbook.
Example

Recall our example where we get a scalar observation of an unknown parameter $\theta \in \mathbb{R}$ in zero-mean Gaussian noise:

$$p_Y(y; \theta) = \frac{1}{\sqrt{2\pi}\sigma} \exp \left( \frac{-(y - \theta)^2}{2\sigma^2} \right)$$

We’ve already computed the Fisher information:

$$I(\theta) = \frac{1}{\sigma^2}$$

Since the conditions of the theorem are satisfied (you should check this), we can say that the variance of any unbiased estimator is lower bounded by

$$\text{var}[\hat{\theta}(Y)] \geq \sigma^2.$$ 

The estimator $\hat{\theta}(y) = y$ is unbiased and it is easy to show that it achieves this minimum variance bound. Hence $\hat{\theta}(y) = y$ is MVU.
Let’s also check the attainability condition by computing

$$
\frac{\partial \ln p_Y(y; \theta)}{\partial \theta} = \frac{\partial}{\partial \theta} \ln \left( \frac{1}{\sqrt{2\pi\sigma}} \exp \left( \frac{-(y - \theta)^2}{2\sigma^2} \right) \right)
$$

$$
= \frac{\partial}{\partial \theta} \left( \ln \left( \frac{1}{\sqrt{2\pi\sigma}} \right) + \frac{-(y - \theta)^2}{2\sigma^2} \right)
$$

$$
= 0 - \frac{\partial}{\partial \theta} \frac{(y - \theta)^2}{2\sigma^2}
$$

$$
= \frac{2(y - \theta)}{2\sigma^2}
$$

$$
= \frac{1}{\sigma^2}(y - \theta)
$$

$$
= I(\theta)(\hat{\theta}(y) - \theta)
$$
CRLB for Signals in Zero-Mean White Gaussian Noise

We assume the general system model

\[ Y_k = s_k(\theta) + W_k \quad \text{for} \quad k = 0, 1, \ldots, n - 1 \]

where \( s_k(\theta) \) is a deterministic signal with an unknown real-valued non-random scalar parameter \( \theta \) and where \( W_k \) i.i.d. \( \sim \mathcal{N}(0, \sigma^2) \).

To compute the Fisher information, we can differentiate twice to get:

\[
\frac{\partial^2}{\partial \theta^2} \ln p_Y(y; \theta) = \frac{1}{\sigma^2} \sum_{k=0}^{n-1} \left\{ [y_k - s_k(\theta)] \frac{\partial^2}{\partial \theta^2} s_k(\theta) - \left( \frac{\partial}{\partial \theta} s_k(\theta) \right)^2 \right\}
\]

We then take the expected value (over the observations) to get

\[
I(\theta) = -E \left[ \frac{\partial^2}{\partial \theta^2} \ln p_Y(Y; \theta) \right] = \frac{1}{\sigma^2} \sum_{k=0}^{n-1} \left( \frac{\partial}{\partial \theta} s_k(\theta) \right)^2
\]

and the CRLB follows immediately as \( 1/I(\theta) \). Note additive information.
Example: Sinusoidal Frequency Estimation in AWGN

Consider the case where

\[ Y_k = a \cos(\theta_k + \phi) + W_k \text{ for } k = 0, 1, \ldots, n - 1 \]

where \( a \) and \( \phi \) are known, \( \theta \in (0, \pi) \), and \( W_k \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \sigma^2) \). You can confirm that the regularity conditions of the theorems are all satisfied here.

To compute the CRLB, we can apply our prior result:

\[
\text{var}_\theta \left[ \hat{\theta}(Y) \right] \geq \frac{\sigma^2}{\sum_{k=0}^{n-1} \left( \frac{\partial}{\partial \theta} s_k(\theta) \right)^2} = \frac{\sigma^2}{a^2 \sum_{k=0}^{n-1} (k \sin(\theta_k + \phi))^2}
\]
Example: Sinusoidal Frequency Estimation in AWGN

\[ n = 10, \sigma^2/a^2 = 1, \text{ and } \phi = 0. \]