THE CRAMER-RAO LOWER BOUND

The CRLB is a lower bound on the variance of any unbiased estimator of a parameter $\underline{\theta}$.

It is useful in many ways:

1. If $\hat{\underline{\theta}}$ achieves the CRLB for all $\underline{\theta} \in \Theta$, the $\hat{\underline{\theta}}$ is a MVUE.

2. The CRLB provides a benchmark against which we can compare the performance of any unbiased estimator. We're doing well if our estimator is "close" to the CRLB

3. The CRLB allows us to rule out impossible estimators. We know it is impossible to find an estimator that beats the CRLB. This is useful in feasibility studies.

4. The theory behind the CRLB tells us precisely when the bound is achievable.

$\boxed{\text{CRLB: Scalar Parameter}}$

Theorem | Consider $\underline{X} \sim f_\theta(\underline{x}) = f(\underline{x};\theta)$ where θ is fixed but unknown. Assume $f(x;\theta)$ satisfies

$$E\left\{ \frac{\partial \log f(\underline{\mathbf{X}};\theta)}{\partial\theta} \right\} = 0$$

where the expectation is with respect to $f(\underline{x};\theta)$.

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

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

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

$$I(\theta) := E\left\{ \left(\frac{\partial \log f(\underline{X};\theta)}{\partial\theta} \right)^2 \right\}$$

Here $I(\theta)$ is evaluated at the true value of the unknown parameter, and the expectation is w.r.t $f(\underline{x};\theta)$

Furthermore, the bound holds with equality iff

$$\frac{\partial \log f(\underline{x};\theta)}{\partial\theta} = I(\theta)\cdot(\hat{\theta}(x)-\theta) \qquad \forall x \in \mathcal{X}$$

In this case we say $\hat{\theta}$ is efficient.

Remarks 1. $\underline{X}$ can be continuous or discrete; we only need differentiability w.r.t. θ.

2. When viewed as a function of θ, $f(\underline{x};\theta)$ is called the <u>likelihood</u> of θ, and $\log f(\underline{x};\theta)$ the <u>log-likelihood</u>.

3. The function

$$\frac{\partial \log f(\underline{x};\theta)}{\partial \theta}$$

is called the <u>score function</u>.

4. The condition

$$E\left\{ \frac{\partial \log f(\underline{X};\theta)}{\partial \theta} \right\} = 0$$

is called a <u>regularity condition</u>.

5. Using integration by parts, the Fisher information can be rewritten

a $$I(\theta) =$$

6. An efficient estimator does not always exist.

7. If

$$\frac{\partial \log f(\underline{x};\theta)}{\partial \theta} = I(\theta)\left(\hat{\theta}(\underline{x}) - \theta\right) \quad \forall \theta \ \forall \underline{x}$$

then $\hat{\theta}$ is a MVUE.

Example | Suppose $\underline{x} = [x_1, \ldots, x_N]^T$ where

$$X_i \sim N(\mu, \sigma^2), \quad i = 1, \ldots, N$$

with $\theta = \mu$ (assume σ^2 is known)

Let's compute the CRLB for θ.

First, we need to check the condition

$$E\left\{ \frac{\partial \log f(\underline{x} ; \theta)}{\partial \theta} \right\} = 0$$

- $\log f(\underline{x} ; \mu) = -\frac{N}{2} \log (2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum (x_i - \mu)^2$
- $\frac{\partial \log f(\underline{x} ; \mu)}{\partial \mu} = \frac{1}{\sigma^2} \sum (x_i - \mu)$
- $E\left\{ \frac{1}{\sigma^2} \sum (X_i - \mu) \right\} = \frac{1}{\sigma^2} \sum (\underbrace{E X_i}_{\mu} - \mu) = 0$

Now let's compute the CRLB

- $-\frac{\partial^2}{\partial \mu^2} \log f(\underline{x} ; \mu) = -\frac{\partial}{\partial \mu} \left\{ \frac{1}{\sigma^2} \sum (x_i - \mu) \right\} = \frac{N}{\sigma^2}$

$\Rightarrow I(\mu) = E\left\{ \frac{N}{\sigma^2} \right\} = \frac{N}{\sigma^2}.$

$\Rightarrow$ If $\hat{\mu}$ is any unbiased estimator of μ,
then $\text{Var}\{\hat{\mu}\} \geq 1/I(\mu) = \frac{\sigma^2}{N}.$ ▨

Recall the sample mean, $\hat{\mu} = \bar{X} = \frac{1}{N} \Sigma X_i$. Previously we saw $E\hat{\mu} = \mu$ and $Var\{\hat{\mu}\} = \frac{\sigma^2}{N}$

$\Longrightarrow$ $\bar{X}$ is efficient and a MVUE.

Regularity Condition

$$E\left\{ \frac{\partial \log f(\underline{X} ; \theta)}{\partial \theta} \right\} = \int \frac{\partial \log f(\underline{x} ; \theta)}{\partial \theta} \cdot f(\underline{x} ; \theta)\, d\underline{x}$$

$$= \int \frac{\left(\frac{\partial f(\underline{x} ; \theta)}{\partial \theta} \right)}{f(\underline{x} ; \theta)} \cdot f(\underline{x} ; \theta)\, d\underline{x}$$

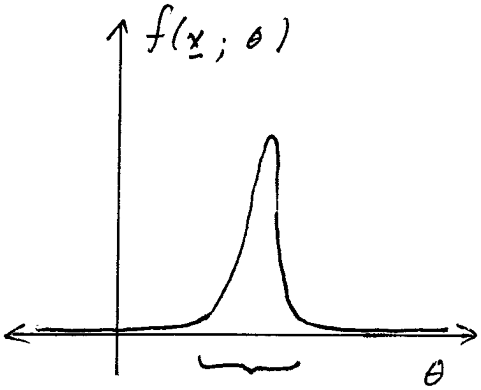
$$= \int \frac{\partial f(\underline{x} ; \theta)}{\partial \theta}\, d\underline{x}$$

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

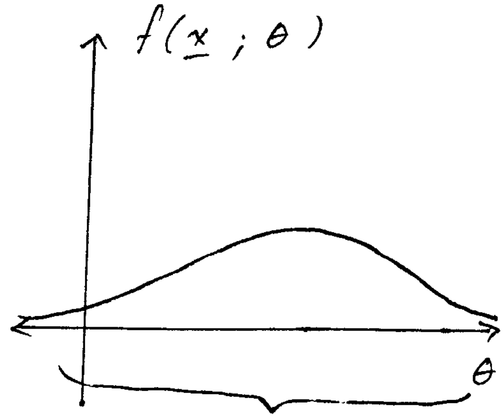
$$= \frac{\partial}{\partial \theta} \{ 1 \} = 0$$

So the regularity condition holds provided we can interchange $\frac{\partial}{\partial \theta}$ and $\int$ (or Σ for discrete $\underline{X}$). This is true for many distributions. A case where it is not true is when the <u>support</u> of $\underline{X}$ depends on θ. For example, $X \sim unif(0, \theta)$.

Fisher Information and Average Curvature

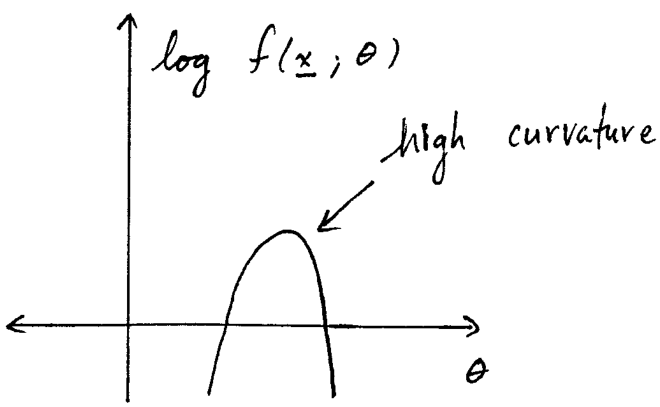


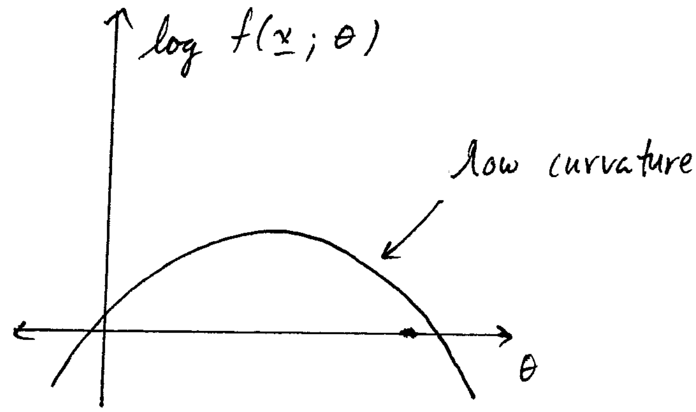
highly likely that θ is in this range



harder to pinpoint θ

The operator $-\dfrac{\partial^2}{\partial\theta^2}$ measures curvature





So $I(\theta)$ reflects the average curvature of the log-likelihood $\log f(\underline{x};\theta)$

<u>Conclusion</u>: θ is easy to estimate

$\Longleftrightarrow$ $f(\underline{x};\theta)$ is "peaky" near θ (on average)

$\Longleftrightarrow$ $\log f(\underline{x};\theta)$ has high curvature at θ (on average)

$\Longleftrightarrow$ $I(\theta) = -E\left\{\dfrac{\partial^2}{\partial\theta^2}\log f(\underline{X};\theta)\right\}$ is large

$\Longleftrightarrow$ CRLB is small

Exercise Consider the correlated bivariate Gaussian

$$\underline{X} = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix} \sim N\left(\begin{bmatrix} \mu \\ \mu \end{bmatrix}, \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \right)$$

where ρ is known. Find the CRLB. Hint:

$$\log f(\underline{x} ; \mu) = \frac{-1}{1+\rho} \left(\mu^2 - \mu (x_1 + x_2) \right) + C$$

Does an efficient estimator exist?

Solution | Let's check the regularity condition first.

$$\frac{\partial \log f(\underline{x};\mu)}{\partial \mu} = -\frac{(2\mu-(x_1+x_2))}{1+\rho}$$

$$E\left\{-\frac{(2\mu-(X_1+X_2))}{(1+\rho)}\right\} = -\frac{(2\mu-(\mu+\mu))}{1+\rho} = 0$$

Because $EX_1 = EX_2 = \mu$. Now

$$-\frac{\partial^2}{\partial\mu^2}\log f(\underline{x};\mu) = \frac{2}{1+\rho}$$

$$\Rightarrow \quad I(\mu) = E\left\{\frac{2}{1+\rho}\right\} = \frac{2}{1+\rho}.$$

$\Rightarrow$ For any unbiased $\hat{\mu}$, $Var(\hat{\mu}) \geq \frac{1+\rho}{2} = CRLB.$

Notice that

$$\frac{\partial \log f(\underline{x};\mu)}{\partial\mu} = \frac{2}{1+\rho}\left(\frac{x_1+x_2}{2}-\mu\right)$$

$$= I(\theta)\cdot(\hat{\theta}(\underline{x})-\theta)$$

$\Rightarrow$ $\hat{\mu} = \frac{x_1+x_2}{2}$ is efficient for all μ,

and hence an MVUE.

Proof of CRLB: Scalar Case

Let $\hat{\theta}$ be an unbiased estimator.

Consider the random variables

$$Y = \hat{\theta}(\underline{X}) - \theta$$
$$Z = \frac{\partial \log f(\underline{X}\,;\,\theta)}{\partial\theta}$$

Note that both have zero mean.

We will show

$$E\{Y \cdot Z\} = 1.$$

The CRLB then follows by application of the Cauchy-Schwarz inequality:

$$\begin{aligned} 1 = E\{Y\cdot Z\} &= (E\{Y\cdot Z\})^2 \\ &\leq E\{Y^2\}\cdot E\{Z^2\} \\ &= E\{(\hat{\theta}(\underline{X})-\theta)^2\}\cdot E\left\{\left(\frac{\partial \log f(\underline{X}\,;\theta)}{\partial\theta}\right)^2\right\} \\ &= \mathrm{Var}_\theta(\hat{\theta})\cdot I(\theta) \end{aligned}$$

$$\Longrightarrow \quad \mathrm{Var}(\hat{\theta}) \geq \frac{1}{I(\theta)}.$$

We now show $E\{YZ\} = 1$. Since $\hat{\theta}$ is unbiased

(b) $\theta =$

Differentiating both sides w.r.t. θ

$$1 = \frac{\partial}{\partial \theta} \int \hat{\theta}(\underline{x}) f(\underline{x} ; \theta) \, d\underline{x}$$

$$= \int \hat{\theta}(\underline{x}) \frac{\partial}{\partial \theta} f(\underline{x} ; \theta) \, d\underline{x} \qquad (1)$$

$$= \int \hat{\theta}(\underline{x}) \frac{\left(\frac{\partial f(\underline{x};\theta)}{\partial \theta}\right)}{f(\underline{x};\theta)} f(\underline{x};\theta) \, dx$$

$$= \int \hat{\theta}(\underline{x}) \frac{\partial \log f(\underline{x};\theta)}{\partial \theta} f(\underline{x};\theta) \, d\underline{x}$$

$$= \int (\hat{\theta}(\underline{x}) - \theta) \frac{\partial \log f(\underline{x};\theta)}{\partial \theta} f(\underline{x};\theta) \, d\underline{x} \qquad (2)$$

$$= E\{Y \cdot Z\}.$$

Remarks|

- Technically, (1) requires an additional assumption on it being valid to exchange $\frac{\partial}{\partial \theta}$ and $\int$ here
- (2) follows from the regularity condition

Equality holds in the Cauchy-Schwarz inequality iff $\exists\, k(\theta)$ (a constant not depending on $\underline{x}$) such that

$$\frac{\partial \log f(\underline{x};\theta)}{\partial\theta} = k(\theta)(\hat{\theta}(\underline{x})-\theta) \qquad \forall\, \underline{x}\in\mathcal{X}$$

Taking the derivative w.r.t θ of both sides

$$\frac{\partial^2}{\partial\theta^2}\log f(\underline{x};\theta) = -k(\theta) + k'(\theta)\cdot(\hat{\theta}(\underline{x})-\theta),$$

and taking $-E\{\cdot\}$ we get

$$I(\theta) = -E\left\{\frac{\partial^2}{\partial\theta^2}\log f(\underline{x};\theta)\right\}$$

$$= k(\theta) - k'(\theta)\underbrace{E\{\hat{\theta}(\underline{X})-\theta\}}_{=0}$$

$$= k(\theta). \qquad \blacksquare$$

CRLB: Vector Parameter

The vector CRLB has the form

$$\text{Cov}_{\underline{\theta}}(\hat{\underline{\theta}}) \geq I(\underline{\theta})^{-1}$$

where

$$\text{Cov}_{\underline{\theta}}(\hat{\underline{\theta}}) = E\left\{(\hat{\underline{\theta}}-\underline{\theta})(\hat{\underline{\theta}}-\underline{\theta})^T\right\}$$

$$= \begin{bmatrix} \text{Var}(\hat{\theta}_1) & \text{Cov}(\hat{\theta}_1, \theta_2) & \cdots & \text{Cov}(\hat{\theta}_1, \hat{\theta}_p) \\ \text{Cov}(\hat{\theta}_2, \hat{\theta}_1) & \text{Var}(\hat{\theta}_2) & & \\ \vdots & \vdots & & \vdots \\ \text{Cov}(\hat{\theta}_p, \hat{\theta}_1) & & & \text{Var}(\hat{\theta}_p) \end{bmatrix}$$

is the <u>covariance matrix</u> of $\hat{\underline{\theta}}$ and

$$I(\underline{\theta}) = E\left\{\left(\frac{\partial}{\partial\underline{\theta}} \log f(\underline{X};\underline{\theta})\right)\left(\frac{\partial}{\partial\underline{\theta}} \log f(\underline{X};\underline{\theta})\right)^T\right\}$$

is the <u>Fisher information matrix</u> of $\underline{\theta}$, and

$$\text{Cov}_{\underline{\theta}}(\hat{\underline{\theta}}) \geq I(\underline{\theta})^{-1}$$

means

$$\text{Cov}_{\underline{\theta}}(\hat{\underline{\theta}}) - I(\underline{\theta})^{-1} \text{ is positive semi-definite.}$$

Recall that if $\phi : \mathbb{R}^p \longrightarrow \mathbb{R}$ then

$$\frac{\partial \phi}{\partial \underline{\theta}} = \left[\frac{\partial \phi}{\partial \theta_1} \quad \cdots \quad \frac{\partial \phi}{\partial \theta_p} \right]^T =: \nabla_{\underline{\theta}}\, \phi \longleftarrow \boxed{\text{alternate notation}}$$

Analogous to the scalar case, it can be shown that

$$I(\underline{\theta}) = E\left\{ \left(\frac{\partial}{\partial \underline{\theta}} \log f(\underline{X}; \theta) \right) \left(\frac{\partial}{\partial \underline{\theta}} \log f(\underline{X}; \underline{\theta}) \right)^T \right\}$$

$$= -E\left\{ \frac{\partial^2}{\partial \underline{\theta}\, \partial \underline{\theta}^T} \log f(\underline{X}; \underline{\theta}) \right\}$$

where

$$\frac{\partial \phi}{\partial \underline{\theta}^T} = \left[\frac{\partial \phi}{\partial \theta_1} \quad \cdots \quad \frac{\partial \phi}{\partial \theta_p} \right] = \left(\frac{\partial \phi}{\partial \underline{\theta}} \right)^T$$

and

$$\frac{\partial^2 \phi}{\partial \underline{\theta}\, \partial \underline{\theta}^T} = \begin{bmatrix} \frac{\partial^2 \phi}{\partial \theta_1^2} & \frac{\partial^2 \phi}{\partial \theta_1 \partial \theta_2} & \cdots & \frac{\partial^2 \phi}{\partial \theta_1 \partial \theta_p} \\ \frac{\partial^2 \phi}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 \phi}{\partial \theta_2^2} & & \vdots \\ \vdots & & \ddots & \\ \frac{\partial^2 \phi}{\partial \theta_p \partial \theta_1} & & \cdots & \frac{\partial^2 \phi}{\partial \theta_p^2} \end{bmatrix} =: \nabla_{\underline{\theta}}^2\, \phi$$

Theorem (Vector CRLB)

Let $\underline{X} \sim f(\underline{x}\,;\underline{\theta})$ where $\underline{\theta} \in \Theta \subseteq \mathbb{R}^p$. Assume

① Θ is an open subset of $\mathbb{R}^p$

② $f(\underline{x}\,;\underline{\theta})$ is differentiable in $\underline{\theta}$

③ The following regularity condition holds:

$$E\left\{\frac{\partial}{\partial \underline{\theta}} \log f(\underline{X}\,;\underline{\theta})\right\} = \underline{0} \qquad \forall\, \underline{\theta} \in \Theta$$

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

$$\mathrm{Cov}_{\underline{\theta}}(\hat{\underline{\theta}}) \geq I(\underline{\theta})^{-1}$$

with equality iff

$$\frac{\partial}{\partial \underline{\theta}} \log f(\underline{x}\,;\underline{\theta}) = I(\underline{\theta}) \cdot (\hat{\underline{\theta}}(\underline{x}) - \underline{\theta}) \qquad \forall\, \underline{x} \in \mathcal{X}.$$

Proof For the most part, the proof generalizes the proof of the scalar case, although some new techniques are necessary. See the book for details.

If A and B are symmetric matrices and $A \geq B$, then $a_{ii} \geq b_{ii}$ $\forall i$. This follows by taking $\underline{z}_i = [0 \cdots 0\ 1\ 0 \cdots 0]^T$ and noting
$\uparrow$ ith position

$$0 \leq \underline{z}_i^T (A - B) \underline{z}$$

$$= a_{ii} - b_{ii}.$$

Therefore we have the following

Corollary | Under the assumptions of the CRLB, if $\underline{\hat{\theta}}$ is unbiased then

$$Var(\hat{\theta}_i) \geq [I(\underline{\theta})]^{-1}_{ii}$$

Thus, the vector CRLB implies scalar lower bounds on each component of $\underline{\hat{\theta}}$.

Furthermore, if $Cov_{\underline{\theta}}(\underline{\hat{\theta}}) = I(\underline{\theta})^{-1}$ $\forall \underline{\theta}$, then $\underline{\hat{\theta}}$ is a MVUE because

$$Var_{\underline{\theta}}(\underline{\hat{\theta}}) = E\{(\underline{\hat{\theta}} - \underline{\theta})^T(\underline{\hat{\theta}} - \underline{\theta})\}$$

$$= E\{\sum_{i=1}^{N} (\hat{\theta}_i - \theta_i)^2\} = \sum_{i=1}^{N} E\{(\hat{\theta}_i - \theta_i)^2\}$$

$$= \sum_{i=1}^{N} Var(\hat{\theta}_i)$$

is minimized.

<u>Exercise</u> Suppose $\underline{x} = [x_1 \cdots x_N]^T$ where

$$X_i \overset{iid}{\sim} N(\mu, \sigma^2).$$

Find the CRLB for $\underline{\theta} = [\mu \ \ \sigma^2]^T$.

Note: $$\log f(\underline{x}; \underline{\theta}) = -\frac{N}{2}\log(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{N}(x_i - \mu)^2$$

Solution

$$\frac{\partial \log f(\underline{x}\,;\,\underline{\theta})}{\partial \mu} = \frac{1}{\sigma^2}\sum_{i=1}^{N}(x_i-\mu)$$

$$\frac{\partial \log f(\underline{x}\,;\,\underline{\theta})}{\partial(\sigma^2)} = \frac{-N}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{N}(x_i-\mu)^2$$

$$\frac{\partial^2 \log f(\underline{x}\,;\,\underline{\theta})}{\partial \mu^2} = -\frac{N}{\sigma^2}$$

$$\frac{\partial^2 \log f(\underline{x}\,;\,\underline{\theta})}{\partial(\sigma^2)^2} = \frac{N}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{N}(x_i-\mu)^2$$

$$\frac{\partial^2 \log f(\underline{x}\,;\,\underline{\theta})}{\partial \mu\,\partial(\sigma^2)} = \frac{-1}{\sigma^4}\sum_{i=1}^{N}(x_i-\mu)$$

$$I(\underline{\theta}) = -E\begin{bmatrix} -\frac{N}{\sigma^2} & -\frac{1}{\sigma^4}\sum_{i=1}^{N}(X_i-\mu) \\ -\frac{1}{\sigma^4}\sum_{i=1}^{N}(X_i-\mu) & \frac{N}{2\sigma^4} - \frac{1}{\sigma^6}\sum_{i=1}^{N}(X_i-\mu)^2 \end{bmatrix}$$

$$= \begin{bmatrix} \frac{N}{\sigma^2} & 0 \\ 0 & \frac{N}{2\sigma^4} \end{bmatrix}$$

$$\Longrightarrow \quad \mathrm{Var}(\hat{\mu}) \geq \frac{\sigma^2}{N}$$

$$\mathrm{Var}(\widehat{\sigma^2}) \geq \frac{2\sigma^4}{N}$$

Summary

- CRLB = lower bound on variance of any unbiased estimator
- Bound given by Fisher Information (matrix)
- score function = $\frac{\partial}{\partial\theta} \log f(\underline{x}\,;\underline{\theta})$
  - determines regularity condition and condition for equality
- Proof: application of Cauchy-Schwarz

Key

a. $-E\left\{\dfrac{\partial^2 \log f(\underline{X}\,;\theta)}{\partial\theta^2}\right\}$

b. $\int \hat{\theta}(\underline{x})\, f(\underline{x}\,;\,\theta)\, dx$