Imperial College London

Lecture 03: The Cramér-Rao Lower Bound

Statistical Modelling I

Dr. Riccardo Passeggeri

Outline

1. The Cramér-Rao Lower Bound and Fisher Information

2. Example

3. Proof: CRLB

4. Proof: Information Identity

The Cramér-Rao Lower Bound and Fisher Information

Can we identify optimal estimators?

Is there an estimator T of θ such that $MSE_\theta(T) \leq MSE_\theta(S)$ for all estimators S?

Theorem: Cramér-Rao Lower Bound

Suppose $T = T(X)$ is an unbiased estimator for $\theta \in \Theta \subset \mathbb{R}$ based on $X = (X_1, \ldots, X_n)$ with joint pdf $f_\theta(x)$. Under mild regularity conditions,

$$Var_\theta(T) \geq \frac{1}{I(\theta)},$$

where

$$I(\theta) = E_\theta\left[\left\{\frac{\partial}{\partial\theta}\log f_\theta(X)\right\}^2\right]$$

is the *Fisher information* of the sample.

Remark: Fisher Information Identity

The Fisher information can also be written as

$$I(\theta) = -E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f_\theta(X)\right].$$

Corollary: Information from a Random Sample

Suppose $X_1, \ldots, X_n$ are a random sample. Then

$$f_\theta(x) = \prod_{i=1}^{n} f_\theta^{(1)}(x_i),$$

where $x = (x_1, \ldots, x_n)$ and $f_\theta^{(1)}$ is the pdf/pmf of a single observation. This implies

$$I_f(\theta) = -E_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f_\theta(X)\right) = \sum_{i=1}^{n} -E_\theta\left(\frac{\partial^2}{\partial\theta^2}\log f_\theta^{(1)}(X_i)\right) = nI_{f^{(1)}}(\theta).$$

Thus for a random sample, the Fisher information is proportional to the sample size.

Example

Example: find $I_f(\theta)$ when $X_1, \ldots, X_n \sim \text{Bernoulli}(\theta)$ iid

Summary: $X_1, \ldots, X_n \sim \text{Bernoulli}(\theta)$

For any *unbiased* estimator T, $Var_\theta(T) \geq \theta(1-\theta)/n = Var(\bar{X})$

Proof: CRLB

Proof Outline

We will show that if $T = T(X)$ is unbiased estimator for θ based on X with joint pdf $f_\theta(x)$

$$Var_\theta(T) \geq \frac{1}{I(\theta)}.$$

Proof outline

- Step 1: Cauchy-Schwarz inequality
- Step 2: Simplify lower bound*

*Regularity conditions must include/imply

(R1) The set $A = \{x \in \mathbb{R}^n : f_\theta(x) > 0\}$ does not depend on θ, Θ is an open interval in $\mathbb{R}$. For all $\theta \in \Theta$ there exists $\frac{\partial}{\partial\theta} f_\theta(x)$.

(R2) Exchanging of differentiation and integration is allowed.

Step 1: Cauchy-Schwarz

Use the Cauchy-Schwarz inequality:

$$[E(YZ)]^2 \leq E(Y^2)E(Z^2)$$

$$\begin{aligned} Var_\theta(T)I_f(\theta) &= E_\theta[(T - E_\theta T)^2]E_\theta[(\frac{\partial}{\partial\theta}\log f_\theta(X))^2] \\ &\geq \left(E_\theta\left[(T - E_\theta T)\frac{\partial}{\partial\theta}\log f_\theta(X)\right]\right)^2 \end{aligned}$$

Step 2: Simplifying the Bound

$$
\begin{aligned}
E_\theta\left[(T - E_\theta T)\frac{\partial}{\partial\theta}\log f_\theta(X)\right] &= E_\theta\left[(T - E_\theta T)\frac{\frac{\partial}{\partial\theta} f_\theta(X)}{f_\theta(X)}\right] \\
&= \int (T(x) - E_\theta T))\,\frac{\frac{\partial}{\partial\theta} f_\theta(x)}{f_\theta(x)} f_\theta(x)dx \\
&= \int T(x)\frac{\partial}{\partial\theta} f_\theta(x)dx - \int E_\theta T\frac{\partial}{\partial\theta} f_\theta(x)dx \\
&= \frac{\partial}{\partial\theta}\int T(x) f_\theta(x)dx - E_\theta T\frac{\partial}{\partial\theta}\int f_\theta(x)dx \\
&= \frac{\partial}{\partial\theta}E_\theta(T) - 0 \\
&= \frac{\partial}{\partial\theta}\theta = 1
\end{aligned}
$$

Summary

Using steps 1 and 2, we have shown $Var_\theta(T)I_f(\theta) \geq 1$. Rearranging completes the proof.

Proof: Information Identity

Proof

We want to show $E_\theta[(\frac{\partial}{\partial\theta}\log f_\theta(X))^2] = -E_\theta[\left(\frac{\partial}{\partial\theta}\right)^2 \log f_\theta(X)]$

Letting $f_\theta' = \frac{\partial}{\partial\theta} f_\theta$ and $f_\theta'' = \frac{\partial}{\partial\theta} f_\theta'$,

$$\begin{aligned} E_\theta[\left(\frac{\partial}{\partial\theta}\right)^2 \log f_\theta(X)] =& E_\theta[\frac{\partial}{\partial\theta}\frac{f_\theta'(X)}{f_\theta(X)}] = E_\theta[-\frac{f_\theta'(X)}{f_\theta^2(X)} f_\theta'(X) + \frac{f_\theta''(X)}{f_\theta(X)}] \\ =& E_\theta[-(\frac{\partial}{\partial\theta}\log f_\theta(X))^2] + E_\theta[\frac{f_\theta''(X)}{f_\theta(X)}]. \end{aligned}$$

Furthermore,

$$E_\theta[\frac{f_\theta''(X)}{f_\theta(X)}] = \int \frac{f_\theta''(x)}{f_\theta(x)} f_\theta(x)dx = \int f_\theta''(x)dx = (\frac{\partial}{\partial\theta})^2 \underbrace{\int f_\theta(x)dx}_{=1} = 0.$$

Summary

We have seen that the Fisher information and CRLB allow us to study the optimal unbiased estimator (in terms of MSE) for fixed n.

Recall: unbiased estimator may not exist, so this bound is not achieved exactly
$\Rightarrow$ What can be said about estimators when n is large?