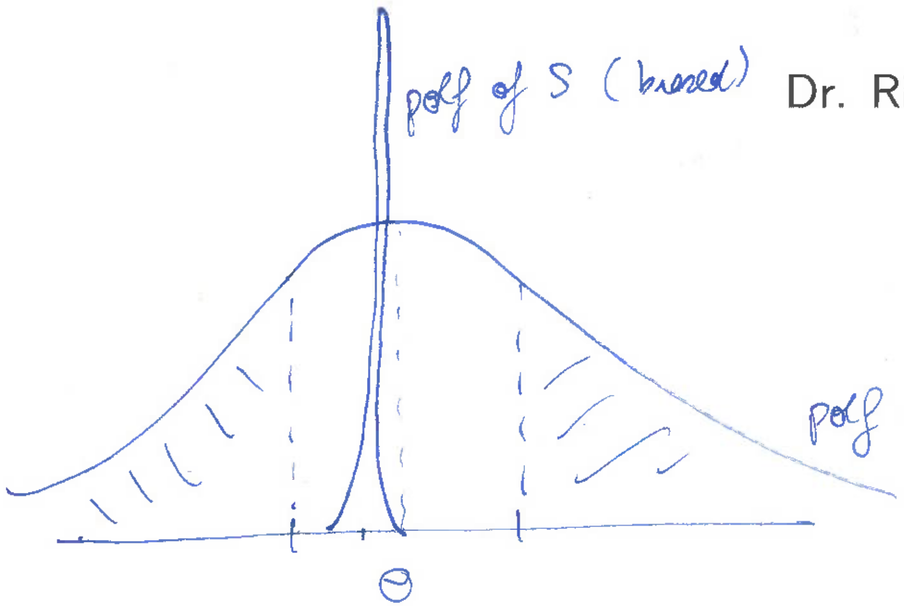
Imperial College London

Lecture 03: The Cramér-Rao Lower Bound

Statistical Modelling I



Dr. Riccardo Passeggeri

Outline

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?

IN GENERAL NO, BUT YES WHEN WE FOCUS ON UBIASED ESTIMATORS

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,

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

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 $X_1, \dots, X_n$ are a random sample. Then

$$f_\theta(x) = \prod_{i=1}^{n} f_\theta^{(1)}(x_i), \quad \cancel{\neq f_\theta^{(2)}}$$

where $x = (x_1, \dots, x_n)$ 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.

THE MORE DATA YOU COLLECT, THE MORE PRECISE YOUR ESTIMATES ARE

Example

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

$$I_f(\theta) = n\, I_{f^{(1)}}(\theta) \qquad f^{(1)}(x) = \theta^x (1-\theta)^{1-x}$$

$$\frac{\partial}{\partial\theta} \log f^{(1)}(x) = \frac{x}{\theta} - \frac{1-x}{1-\theta} = \frac{x-\theta}{\theta(1-\theta)}$$

$$I_{f^{(1)}}(\theta) = E\left[\left(\frac{x-\theta}{\theta(1-\theta)}\right)^2\right] = \frac{1}{\theta^2(1-\theta)^2} \operatorname{Var}(X_1) = \frac{\theta(1-\theta)}{\theta^2(1-\theta)^2} = \frac{1}{\theta(1-\theta)}$$

$$\Rightarrow I_f(\theta) = n \frac{1}{\theta(1-\theta)} \Rightarrow$$

T IS AN UNBIASED ESTIMATOR

$$\operatorname{VAR}(T) \geq \frac{\theta(1-\theta)}{n}$$

WHEN T IS THE SAMPLE MEAN THE $\operatorname{VAR}(T) = \frac{\theta(1-\theta)}{n}$

$\Rightarrow$ T IS OPTIMAL

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

GIVEN TWO VECTORS

a, b

$|a \cdot b| \leq \|a\|\|b\|$

Use the Cauchy-Schwarz inequality:

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

$$Var_\theta(T)I_f(\theta) = E_\theta[(T - E_\theta T)^2]E_\theta[(\frac{\partial}{\partial\theta}\log f_\theta(X))^2]$$
$$\geq \Big(\underbrace{E_\theta\Big[(T - E_\theta T)\frac{\partial}{\partial\theta}\log f_\theta(X)\Big]}_{=1}\Big)^2$$

C-S INEQ.

$\Rightarrow VAR(T) \geq \frac{1}{I_f(\theta)}$

Step 2: Simplifying the Bound

CHAIN RULE

$$E_\theta\left[(T - E_\theta T)\frac{\partial}{\partial\theta}\log f_\theta(X)\right] \overset{?}{=} E_\theta\left[(T - E_\theta T)\frac{\frac{\partial}{\partial\theta}f_\theta(X)}{f_\theta(X)}\right]$$

DEF OF E_θ

$$= \int (T(x) - E_\theta T))\frac{\frac{\partial}{\partial\theta}f_\theta(x)}{f_\theta(x)}f_\theta(x)dx$$

ADDITIVITY

$$= \int T(x)\frac{\partial}{\partial\theta}f_\theta(x)dx - \int \underline{E_\theta} T\frac{\partial}{\partial\theta}f_\theta(x)dx$$

ASSUMPTION (R2)

$$= \frac{\partial}{\partial\theta}\int T(x)f_\theta(x)dx - E_\theta T\frac{\partial}{\partial\theta}\int f_\theta(x)dx$$

DEF OF E_θ AND $\int f_\theta(x)dx = 1$ SO $\frac{\partial}{\partial\theta}1 = 0$

$$= \frac{\partial}{\partial\theta}E_\theta(T) - 0$$

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

$E_\theta(T) = \theta$ BECAUSE T IS UNBIASED

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

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

$$E_\theta\left[\left(\frac{\partial}{\partial\theta}\right)^2 \log f_\theta(X)\right] \overset{\text{CHAIN RULE}}{=} E_\theta\left[\frac{\partial}{\partial\theta}\frac{f'_\theta(X)}{f_\theta(X)}\right] \overset{\text{PRODUCT RULE}}{=} E_\theta\left[-\frac{f'_\theta(X)}{f^2_\theta(X)} f'_\theta(X) + \frac{f''_\theta(X)}{f_\theta(X)}\right]$$

$$= E_\theta\left[-\left(\frac{\partial}{\partial\theta}\log f_\theta(X)\right)^2\right] + \underbrace{E_\theta\left[\frac{f''_\theta(X)}{f_\theta(X)}\right]}_{=0}.$$

Furthermore,

$$E_\theta\left[\frac{f''_\theta(X)}{f_\theta(X)}\right] \underset{\text{DEF OF } E_\theta}{=} \int \frac{f''_\theta(x)}{\cancel{f_\theta(x)}} \cancel{f_\theta(x)}\, dx = \int f''_\theta(x)\, dx \underset{\substack{\text{ASSUMPTION} \\ \text{(R2)}}}{=} \left(\frac{\partial}{\partial\theta}\right)^2 \underbrace{\int f_\theta(x)\, dx}_{=1} = 0.$$

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?

JENSEN INEQUALITY

GIVEN A CONVEX FUNCTION f AND A R.V. X

THEN
$$f(E[X]) \leq E[f(X)]$$

$$0 \leq VAR(X) = E[X^2] - E[X]^2 \Rightarrow E[X]^2 \leq E[X^2] \quad (f(x) = x^2)$$