Cramér–Rao Lower Bound — From Handwritten Notes (Revised)

Converted from IMG_1173–IMG_1176

Table of contents

1 Assumptions (from IMG_1173)

We consider a regular parametric model with density $f(x \mid \theta)$ and log-likelihood $\log f(x \mid \theta)$. The assumptions are the standard "regular case of estimation":

1. Differentiability in the parameter:

$$\frac{\partial}{\partial\theta} f(x \mid \theta) \text{ exists}, \quad \theta \text{ lies in an open interval.}$$

2. The log-density is differentiable and we may treat $f(X \mid \theta)$ and $\log f(X \mid \theta)$ as random variables whose expectations can be differentiated w.r.t. θ. In particular, we can move the derivative inside the expectation:

$$\frac{\partial}{\partial\theta} \mathbb{E}_\theta[g(X, \theta)] = \mathbb{E}_\theta\left[\frac{\partial}{\partial\theta} g(X, \theta)\right],$$

   for g equal to $f(\cdot \mid \theta)$ or $\log f(\cdot \mid \theta)$, whenever the objects exist.

3. Finite Fisher's information:

$$\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X \mid \theta)\right)^2\right] < \infty.$$

We will write the score for one observation as

$$U_\theta(X) = \frac{\partial}{\partial\theta}\log f(X \mid \theta).$$

2 Theorem (from IMG_1174)

Let $T = t(X_1, \ldots, X_n)$ be an unbiased estimator of θ, so $\mathbb{E}_\theta[T] = \theta$. Under Assumptions 1–3, the variance of T satisfies the **Cramér–Rao lower bound**

$$\mathrm{Var}_\theta(T) \ge \frac{1}{n\,\mathbb{E}_\theta\left[\left(\frac{\partial}{\partial\theta}\log f(X \mid \theta)\right)^2\right]} = \frac{1}{n\,I(\theta)},$$

where $I(\theta) = \mathbb{E}_\theta[U_\theta(X)^2]$ is the Fisher information for a single observation.

3 Proof — page 1 (setup)

Because T is unbiased,

$$\mathbb{E}_\theta[T] = \theta.$$

Differentiate both sides w.r.t. θ and use the interchange of derivative and expectation together with the joint density $f(x_1, \ldots, x_n \mid \theta)$:

$$1 = \int t\,\frac{\partial}{\partial\theta} f(x_1, \ldots, x_n \mid \theta)\,dx = \int t\,\frac{\partial}{\partial\theta}\log f(x_1, \ldots, x_n \mid \theta)\,f(x_1, \ldots, x_n \mid \theta)\,dx = \mathbb{E}_\theta[T\,U_\theta(X_1, \ldots, X_n)],$$

where the joint score is the sum of the marginal scores,

$$U_\theta(X_1, \ldots, X_n) = \sum_{i=1}^{n}\frac{\partial}{\partial\theta}\log f(X_i \mid \theta).$$

4 Proof — page 2 (from IMG_1175)

Normalization of the density gives

$$1 = \int f(x \mid \theta)\, dx \quad \Rightarrow \quad 0 = \int \frac{\partial}{\partial \theta} f(x \mid \theta)\, dx = \int \frac{\partial}{\partial \theta} \log f(x \mid \theta)\, f(x \mid \theta)\, dx = \mathbb{E}_\theta[U_\theta(X)].$$

Hence,

$$\mathrm{Cov}_\theta\,(T,\ U_\theta(X_1, \dots, X_n)) = 1,$$

and

$$\mathrm{Var}_\theta(U_\theta(X_1, \dots, X_n)) = \mathbb{E}_\theta[U_\theta(X_1, \dots, X_n)^2] = n\, I(\theta)\ <\ \infty.$$

5 Proof — page 3 (from IMG_1176, corrected last step)

Start with the **correlation** between T and the joint score $U_\theta(X_1, \dots, X_n)$:

$$\rho(T, U_\theta) \equiv \frac{\mathrm{Cov}_\theta(T, U_\theta(X_1, \dots, X_n))}{\sqrt{\mathrm{Var}_\theta(T)}\,\sqrt{\mathrm{Var}_\theta(U_\theta(X_1, \dots, X_n))}}.$$

From the previous page we have $\mathrm{Cov}_\theta(T, U_\theta) = 1$, hence

$$\rho(T, U_\theta) = \frac{1}{\sqrt{\mathrm{Var}_\theta(T)}\,\sqrt{\mathrm{Var}_\theta(U_\theta(X_1, \dots, X_n))}}.$$

Because $|\rho| \le 1$,

$$1 \ge \rho(T, U_\theta)^2 = \frac{1}{\mathrm{Var}_\theta(T)\ \mathrm{Var}_\theta(U_\theta(X_1, \dots, X_n))}.$$

Invert both sides (all terms are positive) to obtain

$$\mathrm{Var}_\theta(T)\ \mathrm{Var}_\theta(U_\theta(X_1, \dots, X_n)) \ge 1.$$

Substituting $\mathrm{Var}_\theta(U_\theta(X_1, \dots, X_n)) = nI(\theta)$ yields

$$\mathrm{Var}_\theta(T) \ge \frac{1}{n\, I(\theta)}.$$

□