Cramér-Rao Lower Bound (CRLB)

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a set of iid samples from a distribution with pdf $f(x;\theta)$, or denoted by $f_\theta(x)$. Suppose $\delta(\mathbf{X})$ is an unbiased estimator of $g(\theta)$. For simplicity, assume θ is one-dimensional.

If the following holds for any function $h(\mathbf{x})$ with $\mathbb{E}|h(\mathbf{X})| < \infty$,

$$\frac{\partial}{\partial\theta}\iint h(\mathbf{x}) f(x_1,\ldots,x_n;\theta)dx_1\cdots dx_n = \iint h(\mathbf{x})\frac{\partial}{\partial\theta} f(x_1,\ldots,x_n;\theta)dx_1\cdots dx_n, \tag{1}$$

that is, we can change the order of integration and differentiation, then

$$\mathrm{Var}(\delta) \geq \frac{\big[g'(\theta)\big]^2}{n\cdot I(\theta)}. \tag{2}$$

where $I(\theta)$ in the denominator is called the **Fisher information**, which has two equivalent expressions:

$$I(\theta) = \mathrm{Var}\Big[\frac{\partial}{\partial\theta}\ln f_\theta(X)\Big] = -\mathbb{E}\Big[\frac{\partial^2}{\partial\theta^2}\ln f_\theta(X)\Big],$$

because

$$\begin{aligned}
\mathbb{E}\Big[\frac{\partial}{\partial\theta}\ln f_\theta(X)\Big] &= 0\\
\mathrm{Var}\Big[\frac{\partial}{\partial\theta}\ln f_\theta(X)\Big] &= \mathbb{E}\Big[\frac{\partial}{\partial\theta}\ln f_\theta(X)\Big]^2 = -\mathbb{E}\Big[\frac{\partial^2}{\partial\theta^2}\ln f_\theta(X)\Big];
\end{aligned}$$

$$\begin{aligned}
\mathbb{E}\Big[\frac{\partial}{\partial\theta}\ln f_\theta(X)\Big] &= \int \Big[\frac{\partial}{\partial\theta}\ln f_\theta(x)\Big] f_\theta(x)dx\\
&= \int \Big[\frac{\partial}{\partial\theta} f_\theta(x)\Big]\frac{1}{f_\theta(x)} f_\theta(x)dx\\
&= \int \Big[\frac{\partial}{\partial\theta} f_\theta(x)\Big]dx, \qquad \text{i.e., } \frac{\partial}{\partial\theta}f_\theta(x) = \Big[\frac{\partial}{\partial\theta}\ln f_\theta(x)\Big]f_\theta(x)\\
&= \frac{\partial}{\partial\theta}\int f_\theta(x)dx = \frac{\partial}{\partial\theta}1 = 0;
\end{aligned}$$

$$\begin{aligned}
\text{Since } 0 = \mathbb{E}\Big[\frac{\partial}{\partial\theta}\ln f_\theta(X)\Big], \quad 0 &= \frac{\partial}{\partial\theta}\int \Big[\frac{\partial}{\partial\theta}\ln f_\theta(x)\Big]\cdot f_\theta(x)dx\\
&= \int \Big[\frac{\partial^2}{\partial\theta^2}\ln f_\theta(x)\Big]\cdot f_\theta(x)dx + \int \Big[\frac{\partial}{\partial\theta}\ln f_\theta(x)\Big]\cdot\Big[\frac{\partial}{\partial\theta}f_\theta(x)\Big]dx\\
&= \mathbb{E}\Big[\frac{\partial^2}{\partial\theta^2}\ln f_\theta(X)\Big] + \int \Big[\frac{\partial}{\partial\theta}\ln f_\theta(x)\Big]\cdot\Big[\frac{\partial}{\partial\theta}\ln f_\theta(x)\Big]f_\theta(x)dx\\
&= \mathbb{E}\Big[\frac{\partial^2}{\partial\theta^2}\ln f_\theta(X)\Big] + \mathbb{E}\Big[\frac{\partial}{\partial\theta}\ln f_\theta(X)\Big]^2.
\end{aligned}$$

A proof for the CRLB is given in the Appendix.

Note that the MSE for an unbiased estimator is equal to its variance, so the right hand side of (2), known as the Cramér-Rao Lower Bound (CRLB), denotes the smallest MSE any unbiased estimators of $g(\theta)$ could have. If the variance of an unbiased estimator for $g(\theta)$ is the same as the CRLB, then we have found the best unbiased estimator for $g(\theta)$ (or one of the best if the best is not unique). Go through the examples in `RaoCramerans.pdf`.

Limitations of CRLB:

1. Some parametric families do not satisfy condition (1), e.g., $\mathsf{Unif}(0, \theta]$ or in general any family whose support depends on the unknown parameter θ.

2. The Cauchy-Schwartz inequality is used in the proof and sometimes the equality is impossible to achieve. That is, for some parameters and some parametric families, there does not exist an unbiased estimator that achieves the CRLB.

Asymptotic Properties of MLEs

Under some regularity conditions, the MLE $\hat{\theta}$ has the following nice asymptotic properties:

$$\text{(Consistency)} \quad : \quad \hat{\theta} \xrightarrow{P} \theta \tag{3}$$

$$\text{(Asymptotic Normality)} \quad : \quad \sqrt{n}\big(\hat{\theta} - \theta\big) \xrightarrow{D} \mathrm{N}\big(0, \frac{1}{I(\theta)}\big). \tag{4}$$

So MLEs are asymptotic efficient, achieving the CRLB.

Technically speaking, (3) and (4) are not for MLEs, but for solutions to $\ell'(\theta) = 0$.

By definition, the MLE $\hat{\theta}$ is the maximizer of the log-likelihood function

$$\ell(\theta) = \sum_{i=1}^{n} \ln f_\theta(X_i).$$

Often, we find the MLE by finding the root of $\ell'(\theta) = 0$, i.e., $\ell'(\hat{\theta}) = 0$. As we have learned before, there are some special situations, in which $\ell'(\hat{\theta}) \neq 0$ (e.g., when the maximum is achieved at a boundary point), or MLE doesn't even exist. Those special cases are ruled out by the so-called *regularity conditions*. For example, Unif$(0, \theta]$ doesn't satisfy those regularity conditions.

$\ell'(\theta) = 0$ could have multiple roots, and some of them are local maximizers not the global one. Results (3) and (4) are just for one of them (of course, the one close to the true parameter θ), but do not offer any guidance on which one we should pick – an unpleasant aspect of these results. Nevertheless, in most examples we encounter in 410, $\ell'(\theta)$ has only one root and it is the maximum, so (3) and (4) do hold for MLEs.

Why we need $\ell'(\hat{\theta}) = 0$? Let's look at how $\ell'(\hat{\theta}) = 0$ is used in the proof for asymptotic normality of $\hat{\theta}$.

- Start with a Taylor expansion of $\ell'(\hat{\theta})$ at the true parameter θ :

$$0 = \ell'(\hat{\theta}) = \ell'(\theta) + \ell''(\theta)(\hat{\theta} - \theta) + \text{Remaining Terms}.$$

  Assuming the remaining terms are small (which could be justified by the regularity conditions), we have

$$\begin{aligned} (\hat{\theta} - \theta) &\approx \frac{1}{-\ell''(\theta)} \ell'(\theta), \\ \sqrt{n}(\hat{\theta} - \theta) &\approx \frac{1}{\left[-\frac{1}{n}\ell''(\theta) \right]} \cdot \left[\sqrt{n}\frac{1}{n}\ell'(\theta) \right]. \end{aligned}$$

- Note that

$$\sqrt{n}\frac{1}{n}\ell'(\theta) = \sqrt{n} \cdot \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f_\theta(X_i) = \sqrt{n}(\bar{Z} - 0),$$

  where $Z_i = \frac{\partial}{\partial\theta} \ln f_\theta(X_i)$ are iid with

$$\mathbb{E}Z_i = 0, \quad \text{Var}(Z_i) = I(\theta).$$

  So by CLT, we have

$$\sqrt{n}\frac{1}{n}\ell'(\theta) \xrightarrow{D} \text{N}\big(0, I(\theta)\big).$$

- By WLLN,

$$\frac{1}{n}\ell''(\theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2}{\partial\theta^2} \ln f_\theta(X_i) = \bar{W} \xrightarrow{P} -I(\theta),$$

  where $W_i = \frac{\partial^2}{\partial\theta^2} \ln f_\theta(X_i)$ with $\mathbb{E}W_i = -I(\theta)$.

- The asymptotic normality (4) follows by the Slutsky's Thm.

Appendix: Proof for the CRLB

1. Start with the unbiasedness:

$$g(\theta) = \iint \delta(\mathbf{x}) f_\theta(x_1, \ldots, x_n) dx_1 \cdots dx_n.$$

   Taking derivative wrt θ on both sides of the equation, due to (1), we have

$$\begin{aligned} g'(\theta) &= \iint \delta(\mathbf{x}) \frac{\partial}{\partial \theta} f_\theta(x_1, \ldots, x_n) dx_1 \cdots dx_n \\ &= \iint \delta(\mathbf{x}) \frac{\partial}{\partial \theta} \ln f_\theta(x_1, \ldots, x_n) \cdot f_\theta(x_1, \ldots, x_n) dx_1 \cdots dx_n \\ &= \mathbb{E}\ \delta(\mathbf{X}) \cdot \frac{\partial}{\partial \theta} \ln f_\theta(\mathbf{X}) \end{aligned}$$

2. Suppose X and Y are two random variables and $\mathbb{E}Y = \mu_Y = 0$. Then

$$\mathrm{Cov}(X, Y) = \mathbb{E}XY - \mu_X \mu_Y = \mathbb{E}(XY).$$

   Since

$$\mathbb{E}\Big[\frac{\partial}{\partial \theta} \ln f_\theta(\mathbf{X})\Big] = n\mathbb{E}\Big[\frac{\partial}{\partial \theta} \ln f_\theta(X)\Big] = 0,$$

   we have

$$g'(\theta) = \mathrm{Cov}\Big(\delta(\mathbf{X}), \frac{\partial}{\partial \theta} \ln f_\theta(\mathbf{X})\Big).$$

3. Recall the Cauchy-Schwartz inequality

$$\mathrm{Cov}(X, Y)^2 \leq \mathrm{Var}(X) \cdot \mathrm{Var}(Y),$$

   and "=" holds iff X is a linear function of Y, i.e., $X = aY + b$.

   We have

$$\begin{aligned} \Big[g'(\theta)\Big]^2 &\leq \mathrm{Var}\Big[\delta(\mathbf{X})\Big] \cdot \mathrm{Var}\Big[\frac{\partial}{\partial \theta} \ln f_\theta(\mathbf{X})\Big] \\ &= \mathrm{Var}\Big[\delta(\mathbf{X})\Big] \cdot n \cdot \mathrm{Var}\Big[\frac{\partial}{\partial \theta} \ln f_\theta(X)\Big], \end{aligned}$$

   that is,

$$\mathrm{Var}\Big[\delta(\mathbf{X})\Big] \geq \frac{\Big[g'(\theta)\Big]^2}{nI(\theta)}.$$