Lecture 08: Cramér-Rao Lower Bound

Mathematical Statistics II, MATH 60062/70062

Tuesday February 15, 2022

Reference: Casella & Berger, 7.3.2

Recap: Uniformly minimum-variance unbiased estimator

An estimator $W^* = W^*(\boldsymbol{X})$ is a **uniformly minimum-variance unbiased estimator (UMVUE)** of $\tau(\theta)$ if

1. $E_\theta(W^*) = \tau(\theta)$ for all $\theta \in \Theta$.
2. $\mathrm{Var}_\theta(W^*) \leq \mathrm{Var}_\theta(W)$ for all $\theta \in \Theta$, where W is any other unbiased estimator of $\tau(\theta)$.

We will discuss two approaches to find UMVUEs:

1. (**Cramér-Rao Inequality**) Determine a **lower bound** on the variance of *any* unbiased estimator of $\tau(\theta)$. If we can find an unbiased estimator whose variance attains this lower bound, we have found the UMVUE.
2. (**Rao-Blackwell Theorem**) Relate the property of UMVUEs with the notation of **sufficiency** and **completeness**.

Cramér-Rao Inequality

Suppose $\boldsymbol{X} \sim f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)$. Let $W(\boldsymbol{X})$ be any estimator satisfying

$$\frac{d}{d\theta} E_\theta[W(\boldsymbol{X})] = \int_{\mathcal{X}} \frac{\partial}{\partial\theta}[W(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)] d\boldsymbol{x}$$

and

$$\mathrm{Var}_\theta(W(\boldsymbol{X})) < \infty.$$

Then

$$\mathrm{Var}_\theta(W(\boldsymbol{X})) \geq \frac{(\frac{d}{d\theta} E_\theta[W(\boldsymbol{X})])^2}{E_\theta\left[(\frac{\partial}{\partial\theta} \log f_{\boldsymbol{X}}(\boldsymbol{X} \mid \theta))^2\right]}$$

Cramér-Rao Inequality

Suppose $\boldsymbol{X} \sim f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)$. Let $W(\boldsymbol{X})$ be any estimator satisfying

$$\frac{d}{d\theta} E_\theta[W(\boldsymbol{X})] = \int_{\mathcal{X}} \frac{\partial}{\partial \theta} [W(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)] d\boldsymbol{x}$$

and

$$\mathrm{Var}_\theta(W(\boldsymbol{X})) < \infty.$$

Then

$$\mathrm{Var}_\theta(W(\boldsymbol{X})) \geq \frac{(\frac{d}{d\theta} E_\theta[W(\boldsymbol{X})])^2}{E_\theta\left[(\frac{\partial}{\partial\theta} \log f_{\boldsymbol{X}}(\boldsymbol{X} \mid \theta))^2\right]}$$

The quantity on the RHS is called the **Cramér-Rao Lower Bound (CRLB)** on the variance of the estimator $W(\boldsymbol{X})$. Note that this lower bound is not restricted to unbiased estimators.

Cramér-Rao Lower Bound (CRLB)

The CRLB is given by

$$\frac{(\frac{d}{d\theta}E_\theta[W(\boldsymbol{X})])^2}{E_\theta\left[(\frac{\partial}{\partial\theta}\log f_{\boldsymbol{X}}(\boldsymbol{X} \mid \theta))^2\right]}.$$

- If $W(\boldsymbol{X})$ is an unbiased estimator of $\tau(\theta)$, then the numerator becomes $[\tau'(\theta)]^2$.
- If $W(\boldsymbol{X})$ is an unbiased estimator of θ, then the numerator becomes 1.

The denominator is called the **information number** or **Fisher information** of the sample $\boldsymbol{X}$.

- The number does not depend on $W(\boldsymbol{X})$. It is a property of the distribution $f_{\boldsymbol{X}}(\boldsymbol{X} \mid \theta)$.
- The larger the number, the more information the sample has about θ and the smaller the bound on the variance.

Cramér-Rao Lower Bound, iid case

If the assumptions for the Cramér-Rao Lower Bound are satisfied and, additionally, if the sample $\boldsymbol{X}$ consists of $X_1, \ldots, X_n$ which are iid from $f_X(x \mid \theta)$, then the denominator of the CRLB

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

Denoting by $I_n(\theta)$ and $I_1(\theta)$ the Fisher information based on the sample $\boldsymbol{X}$ and that based on one observation X, respectively,

$$I_n(\theta) = nI_1(\theta).$$

Cramér-Rao Lower Bound, Exponential family

If the assumptions for the Cramér-Rao Lower Bound are satisfied and $f_X(x \mid \theta)$ satisfies

$$\frac{d}{d\theta} E_\theta \left[\frac{\partial}{\partial \theta} \log f_X(X \mid \theta) \right] = \int \frac{\partial}{\partial \theta} \left[\left(\frac{\partial}{\partial \theta} \log f_X(x \mid \theta) \right) f_X(x \mid \theta) \right] dx,$$

which is true for an Exponential family, then

$$I_1(\theta) = E_\theta \left[\left(\frac{\partial}{\partial \theta} \log f_X(X \mid \theta) \right)^2 \right] = -E_\theta \left[\frac{\partial^2}{\partial \theta^2} \log f_X(X \mid \theta) \right]$$

Poisson CRLB

Suppose that $X_1, \ldots, X_n$ are iid $\mathrm{Pois}(\theta)$, where $\theta > 0$. Find the CRLB on the variance of unbiased estimators of $\tau(\theta) = \theta$.

Poisson CRLB

Suppose that $X_1, \ldots, X_n$ are iid $\mathrm{Pois}(\theta)$, where $\theta > 0$. Find the CRLB on the variance of unbiased estimators of $\tau(\theta) = \theta$.

The CRLB is

$$\frac{1}{I_n(\theta)} = \frac{1}{nI_1(\theta)},$$

where

$$I_1(\theta) = E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_X(X \mid \theta)\right)^2\right] = -E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f_X(X \mid \theta)\right].$$

Poisson CRLB

Suppose that $X_1, \ldots, X_n$ are iid $\mathrm{Pois}(\theta)$, where $\theta > 0$. Find the CRLB on the variance of unbiased estimators of $\tau(\theta) = \theta$.

The CRLB is

$$\frac{1}{I_n(\theta)} = \frac{1}{nI_1(\theta)},$$

where

$$I_1(\theta) = E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_X(X \mid \theta)\right)^2\right] = -E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f_X(X \mid \theta)\right].$$

The log of Poisson PMF is

$$\log f_X(x \mid \theta) = \log\left(\frac{e^{-\theta}\theta^x}{x!}\right) = -\theta + x\log\theta - \log x!$$

for $x = 0, 1, 2, \ldots$.

Therefore,

$$\frac{\partial^2}{\partial\theta^2}\log f_X(x \mid \theta) = -\frac{x}{\theta^2},$$

and the Fisher information based on one observation is

$$I_1(\theta) = -E_\theta\left(-\frac{X}{\theta^2}\right) = \frac{1}{\theta}.$$

The CRLB is

$$\frac{1}{nI_1(\theta)} = \frac{\theta}{n}.$$

Therefore,

$$\frac{\partial^2}{\partial\theta^2} \log f_X(x \mid \theta) = -\frac{x}{\theta^2},$$

and the Fisher information based on one observation is

$$I_1(\theta) = -E_\theta \left(-\frac{X}{\theta^2} \right) = \frac{1}{\theta}.$$

The CRLB is

$$\frac{1}{nI_1(\theta)} = \frac{\theta}{n}.$$

Note: The sample mean $W(\boldsymbol{X}) = \bar{X}$ is an unbiased estimator of θ for the $\mathrm{Pois}(\theta)$ model, and $\mathrm{Var}_\theta(\bar{X}) = \theta/n$. Therefore, $W(\boldsymbol{X}) = \bar{X}$ is the UMVUE for θ.

Gamma CRLB

Suppose that $X_1, \ldots, X_n$ are iid $\mathrm{Gamma}(\alpha_0, \beta)$, where α_0 is known and $\beta > 0$. The PDF is given by

$$f_X(x \mid \beta) = \frac{1}{\Gamma(\alpha_0)\beta^{\alpha_0}} x^{\alpha_0 - 1} e^{-x/\beta}.$$

Gamma CRLB

Suppose that $X_1, \ldots, X_n$ are iid $\mathrm{Gamma}(\alpha_0, \beta)$, where α_0 is known and $\beta > 0$. The PDF is given by

$$f_X(x \mid \beta) = \frac{1}{\Gamma(\alpha_0)\beta^{\alpha_0}} x^{\alpha_0 - 1} e^{-x/\beta}.$$

The CRLB is

$$\frac{1}{I_n(\beta)} = \frac{1}{nI_1(\beta)},$$

where

$$I_1(\beta) = E_\beta \left[\left(\frac{\partial}{\partial \beta} \log f_X(X \mid \beta) \right)^2 \right] = -E_\beta \left[\frac{\partial^2}{\partial \beta^2} \log f_X(X \mid \beta) \right].$$

Gamma CRLB

Suppose that $X_1, \ldots, X_n$ are iid $\mathrm{Gamma}(\alpha_0, \beta)$, where α_0 is known and $\beta > 0$. The PDF is given by

$$f_X(x \mid \beta) = \frac{1}{\Gamma(\alpha_0)\beta^{\alpha_0}} x^{\alpha_0 - 1} e^{-x/\beta}.$$

The CRLB is

$$\frac{1}{I_n(\beta)} = \frac{1}{nI_1(\beta)},$$

where

$$I_1(\beta) = E_\beta\left[\left(\frac{\partial}{\partial\beta}\log f_X(X \mid \beta)\right)^2\right] = -E_\beta\left[\frac{\partial^2}{\partial\beta^2}\log f_X(X \mid \beta)\right].$$

The log of Gamma PDF is

$$\log f_X(x \mid \beta) = -\Gamma(\alpha_0) - \alpha_0 \log\beta + (\alpha_0 - 1)\log x - \frac{x}{\beta}$$

for $x > 0$.

Therefore,

$$\frac{\partial^2}{\partial \beta^2} \log f_X(x \mid \beta) = \frac{\alpha_0}{\beta^2} - \frac{2x}{\beta^3},$$

and the Fisher information based on one observation is

$$I_1(\beta) = -E_\beta\left(\frac{\alpha_0}{\beta^2} - \frac{2X}{\beta^3}\right) = \frac{\alpha_0}{\beta^2}.$$

The CRLB is

$$\frac{1}{nI_1(\beta)} = \frac{\beta^2}{n\alpha_0}.$$

Therefore,

$$\frac{\partial^2}{\partial \beta^2} \log f_X(x \mid \beta) = \frac{\alpha_0}{\beta^2} - \frac{2x}{\beta^3},$$

and the Fisher information based on one observation is

$$I_1(\beta) = -E_\beta \left(\frac{\alpha_0}{\beta^2} - \frac{2X}{\beta^3} \right) = \frac{\alpha_0}{\beta^2}.$$

The CRLB is

$$\frac{1}{nI_1(\beta)} = \frac{\beta^2}{n\alpha_0}.$$

Consider the estimator $W(\boldsymbol{X}) = \bar{X}/\alpha_0$. Note that

$$E_\beta[W(\boldsymbol{X})] = \beta$$

and

$$\mathrm{Var}_\beta[W(\boldsymbol{X})] = \frac{\alpha_0 \beta^2}{n\alpha_0^2} = \frac{\beta^2}{n\alpha_0}.$$

Therefore, $W(\boldsymbol{X}) = \bar{X}/\alpha_0$ is the UMVUE for β.

Attainment of the CRLB

Let $X_1, \ldots, X_n$ be iid $f_X(x \mid \theta)$, where $f_X(x \mid \theta)$ satisfies the conditions stated for the Cramér-Rao Inequality. Let $L(\theta \mid \boldsymbol{x}) = \prod_{i=1}^n f(x_i \mid \theta)$ denote the likelihood function. If $W(\boldsymbol{X}) = W(X_1, \ldots, X_n)$ is any unbiased estimator of $\tau(\theta)$, then $W(\boldsymbol{X})$ attains the Cramér-Rao Lower Bound if and only if

$$\frac{\partial}{\partial\theta} \log L(\theta \mid \boldsymbol{x}) = a(\theta)[W(\boldsymbol{x}) - \tau(\theta)]$$

for some function $a(\theta)$.

Gamma CRLB

Suppose that $X_1, \ldots, X_n$ are iid $\mathrm{Gamma}(\alpha_0, \beta)$, where α_0 is known and $\beta > 0$. The likelihood function is

$$\begin{aligned} L(\beta \mid \boldsymbol{x}) &= \prod_{i=1}^{n} \frac{1}{\Gamma(\alpha_0)\beta^{\alpha_0}} x_i^{\alpha_0 - 1} e^{-x_i/\beta} \\ &= \left(\frac{1}{\Gamma(\alpha_0)\beta^{\alpha_0}}\right)^n \left(\prod_{i=1}^{n} x_i\right)^{\alpha_0 - 1} e^{-\sum_{i=1}^{n} x_i/\beta}. \end{aligned}$$

The log-likelihood function is

$$\log L(\beta \mid \boldsymbol{x}) = -n \log \Gamma(\alpha_0) - n\alpha_0 \log \beta + (\alpha_0 - 1) \sum_{i=1}^{n} \log x_i - \frac{\sum_{i=1}^{n} x_i}{\beta}$$

The score function

$$\begin{aligned}\frac{\partial}{\partial\beta}\log L(\beta \mid \boldsymbol{x}) &= -\frac{n\alpha_0}{\beta} + \frac{\sum_{i=1}^n x_i}{\beta^2} \\ &= \frac{n\alpha_0}{\beta^2}\left(\frac{\sum_{i=1}^n x_i}{n\alpha_0} - \beta\right) \\ &= a(\beta)[W(\boldsymbol{x}) - \tau(\beta)],\end{aligned}$$

where $W(\boldsymbol{x}) = \bar{x}/\alpha_0$. Because $W(\boldsymbol{X}) = \bar{X}/\alpha_0$ is an unbiased estimator of $\tau(\beta) = \beta$, the variance of $W(\boldsymbol{X})$ attains the CRLB.

Unresolved issues with the CRLB

- Not all distributions are sufficiently smooth to satisfy the regularity conditions for the CRLB.
  - In general, if the range of the PDF/PMF depends on the parameter, the theorem will not be applicable.
- The CRLB may be *unattainable*.
  - Even if we can write

    $$\frac{\partial}{\partial\theta}\log L(\theta \mid \boldsymbol{x}) = a(\theta)[W(\boldsymbol{x}) - \tau(\theta)],$$

    the function $\tau(\theta)$ may not be what want to estimate.

Appendix: Proof of Cramér-Rao Inequality

Suppose $\boldsymbol{X} \sim f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)$. Let $W(\boldsymbol{X})$ be any estimator satisfying

$$\frac{d}{d\theta} E_\theta[W(\boldsymbol{X})] = \int_{\mathcal{X}} \frac{\partial}{\partial\theta}[W(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta)] d\boldsymbol{x}$$

and

$$\mathrm{Var}_\theta(W(\boldsymbol{X})) < \infty.$$

Then

$$\mathrm{Var}_\theta(W(\boldsymbol{X})) \geq \frac{(\frac{d}{d\theta} E_\theta[W(\boldsymbol{X})])^2}{E_\theta\left[(\frac{\partial}{\partial\theta} \log f_{\boldsymbol{X}}(\boldsymbol{X} \mid \theta))^2\right]}$$

Cauchy-Schwarz Inequality. For any two random variables X and Y,

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

where the covariance can be calculated as

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

Upon rearrangement, we get a lower bound on the variance of X,

$$\mathrm{Var}(X) \geq \frac{[\mathrm{Cov}(X, Y)]^2}{\mathrm{Var}(Y)}.$$

The Cramér-Rao Inequality follows from choosing X to be the estimator $W(\boldsymbol{X})$ and Y to be the quantity $\frac{\partial}{\partial\theta} \log f_{\boldsymbol{X}}(\boldsymbol{X} \mid \theta)$ and applying the Cauchy-Schwarz Inequality.

First, note that

$$
\begin{aligned}
E_\theta\left[\frac{\partial}{\partial\theta}\log f_{\boldsymbol{X}}(\boldsymbol{X}\mid\theta)\right] &= \int_{\mathcal{X}} \frac{\partial}{\partial\theta}\log f_{\boldsymbol{X}}(\boldsymbol{x}\mid\theta) f_{\boldsymbol{X}}(\boldsymbol{x}\mid\theta)d\boldsymbol{x} \\
&= \int_{\mathcal{X}} \frac{\frac{\partial}{\partial\theta} f_{\boldsymbol{X}}(\boldsymbol{x}\mid\theta)}{f_{\boldsymbol{X}}(\boldsymbol{x}\mid\theta)} f_{\boldsymbol{X}}(\boldsymbol{x}\mid\theta)d\boldsymbol{x} \\
&= \int_{\mathcal{X}} \frac{\partial}{\partial\theta} f_{\boldsymbol{X}}(\boldsymbol{x}\mid\theta)d\boldsymbol{x} \\
&= \frac{d}{d\theta}\underbrace{\int_{\mathcal{X}} f_{\boldsymbol{X}}(\boldsymbol{x}\mid\theta)d\boldsymbol{x}}_{=1} = 0
\end{aligned}
$$

\partial{\theta} g(f(x, \theta)) = \partial{\theta} f(x, \theta)g'(f(x, \theta))

The interchange of derivative and integral above is justified based on the assumptions.

Next, consider

$$
\begin{aligned}
\mathrm{Cov}_\theta\left(W(\boldsymbol{X}), \frac{\partial}{\partial\theta}\log f_{\boldsymbol{X}}(\boldsymbol{X}\mid\theta)\right) &= E_\theta\left[W(\boldsymbol{X})\frac{\partial}{\partial\theta}\log f_{\boldsymbol{X}}(\boldsymbol{X}\mid\theta)\right] \\
&= \int_{\mathcal{X}} W(\boldsymbol{x})\frac{\partial}{\partial\theta}\log f_{\boldsymbol{X}}(\boldsymbol{x}\mid\theta) f_{\boldsymbol{X}}(\boldsymbol{x}\mid\theta)d\boldsymbol{x} \\
&= \int_{\mathcal{X}} W(\boldsymbol{x})\frac{\frac{\partial}{\partial\theta} f_{\boldsymbol{X}}(\boldsymbol{x}\mid\theta)}{f_{\boldsymbol{X}}(\boldsymbol{x}\mid\theta)} f_{\boldsymbol{X}}(\boldsymbol{x}\mid\theta)d\boldsymbol{x} \\
&= \int_{\mathcal{X}} W(\boldsymbol{x})\frac{\partial}{\partial\theta} f_{\boldsymbol{X}}(\boldsymbol{x}\mid\theta)d\boldsymbol{x} \\
&= \frac{d}{d\theta}\int_{\mathcal{X}} W(\boldsymbol{x}) f_{\boldsymbol{X}}(\boldsymbol{x}\mid\theta)d\boldsymbol{x} \\
&= \frac{d}{d\theta}E_\theta\left[W(\boldsymbol{X})\right]
\end{aligned}
$$

Applying the Cauchy-Schwarz Inequality

$$
\begin{aligned}
\mathrm{Var}_\theta(W(\boldsymbol{X})) &\geq \frac{\left[\mathrm{Cov}_\theta\left(W(\boldsymbol{X}), \frac{\partial}{\partial\theta}\log f_{\boldsymbol{X}}(\boldsymbol{X}\mid\theta)\right)\right]^2}{\mathrm{Var}_\theta\left(\frac{\partial}{\partial\theta}\log f_{\boldsymbol{X}}(\boldsymbol{X}\mid\theta)\right)} \\
&= \frac{(\frac{d}{d\theta}E_\theta[W(\boldsymbol{X})])^2}{E_\theta\left[(\frac{\partial}{\partial\theta}\log f_{\boldsymbol{X}}(\boldsymbol{X}\mid\theta))^2\right]}.
\end{aligned}
$$

Appendix: Proof of Cramér-Rao Lower Bound, iid case

If the assumptions for the Cramér-Rao Lower Bound are satisfied and, additionally, if the sample $\boldsymbol{X}$ consists of $X_1, \ldots, X_n$ which are iid from $f_X(x \mid \theta)$, then the denominator of the CRLB

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

Because $X_1, \ldots, X_n$ are iid,

$$\begin{aligned}
E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_{\boldsymbol{X}}(\boldsymbol{X} \mid \theta)\right)^2\right] &= E_\theta\left[\left(\frac{\partial}{\partial\theta}\log\prod_{i=1}^n f_X(X_i \mid \theta)\right)^2\right] \\
&= E_\theta\left[\left(\frac{\partial}{\partial\theta}\sum_{i=1}^n \log f_X(X_i \mid \theta)\right)^2\right] \\
&= E_\theta\left[\left(\sum_{i=1}^n \frac{\partial}{\partial\theta}\log f_X(X_i \mid \theta)\right)^2\right] \\
&= \sum_{i=1}^n E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_X(X_i \mid \theta)\right)^2\right] \\
&= nE_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_X(X \mid \theta)\right)^2\right].
\end{aligned}$$

Appendix: Proof of information equality

Under regularity conditions,

$$\frac{d}{d\theta}E_\theta\left[\frac{\partial}{\partial\theta}\log f_X(X \mid \theta)\right] = \int \frac{\partial}{\partial\theta}\left[\left(\frac{\partial}{\partial\theta}\log f_X(x \mid \theta)\right) f_X(x \mid \theta)\right] dx,$$

which is true for an Exponential family, then

$$I_1(\theta) = E_\theta\left[\left(\frac{\partial}{\partial\theta}\log f_X(X \mid \theta)\right)^2\right] = -E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f_X(X \mid \theta)\right]$$

By definition

$$
\begin{aligned}
E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f_X(X\mid\theta)\right] &= \int_{\mathcal{X}}\frac{\partial^2}{\partial\theta^2}\log f_X(x\mid\theta) f_X(x\mid\theta)dx \\
&= \int_{\mathcal{X}}\frac{\partial}{\partial\theta}\left[\frac{\frac{\partial}{\partial\theta}f_X(x\mid\theta)}{f_X(x\mid\theta)}\right] f_X(x\mid\theta)dx,
\end{aligned}
$$

where the derivative

$$
\frac{\partial}{\partial\theta}\left[\frac{\frac{\partial}{\partial\theta}f_X(x\mid\theta)}{f_X(x\mid\theta)}\right] = \frac{\frac{\partial^2}{\partial\theta^2}f_X(x\mid\theta)}{f_X(x\mid\theta)} - \frac{\left[\frac{\partial}{\partial\theta}f_X(x\mid\theta)\right]^2}{\left[f_X(x\mid\theta)\right]^2}.
$$

Therefore,

$$
E_\theta\left[\frac{\partial^2}{\partial\theta^2}\log f_X(X\mid\theta)\right] = \int_{\mathcal{X}}\frac{\partial^2}{\partial\theta^2}f_X(x\mid\theta)dx - \int_{\mathcal{X}}\frac{\left[\frac{\partial}{\partial\theta}f_X(x\mid\theta)\right]^2}{f_X(x\mid\theta)}dx,
$$

Consider the first term

$$\begin{aligned}\int_{\mathcal{X}} \frac{\partial^2}{\partial\theta^2} f_X(x \mid \theta) dx &= \frac{d}{d\theta} \int_{\mathcal{X}} \frac{\partial}{\partial\theta} f_X(x \mid \theta) dx \\ &= \frac{d}{d\theta} E_\theta \left[\frac{\partial}{\partial\theta} \log f_X(X \mid \theta) \right] = 0,\end{aligned}$$

and the second term

$$\begin{aligned}\int_{\mathcal{X}} \frac{\left[\frac{\partial}{\partial\theta} f_X(x \mid \theta)\right]^2}{f_X(x \mid \theta)} dx &= \int_{\mathcal{X}} \left[\frac{\partial}{\partial\theta} \log f_X(x \mid \theta) \right]^2 f_X(x \mid \theta) dx \\ &= E_\theta \left[\left(\frac{\partial}{\partial\theta} \log f_X(X \mid \theta) \right)^2 \right].\end{aligned}$$

Therefore,

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

Appendix: Proof of CRLB attainment

Let $X_1, \ldots, X_n$ be iid $f_X(x \mid \theta)$, where $f_X(x \mid \theta)$ satisfies the conditions stated for the Cramér-Rao Inequality. Let $L(\theta \mid \boldsymbol{x}) = \prod_{i=1}^{n} f(x_i \mid \theta)$ denote the likelihood function. If $W(\boldsymbol{X}) = W(X_1, \ldots, X_n)$ is any unbiased estimator of $\tau(\theta)$, then $W(\boldsymbol{X})$ attains the Cramér-Rao Lower Bound if and only if

$$\frac{\partial}{\partial\theta} \log L(\theta \mid \boldsymbol{x}) = a(\theta)[W(\boldsymbol{x}) - \tau(\theta)]$$

for some function $a(\theta)$.

In the proof of Cramér-Rao Inequality

$$\begin{aligned}\mathrm{Var}_\theta(W(\boldsymbol{X})) &\geq \frac{\left[\mathrm{Cov}_\theta\left(W(\boldsymbol{X}), \frac{\partial}{\partial\theta}\log f_{\boldsymbol{X}}(\boldsymbol{X} \mid \theta)\right)\right]^2}{\mathrm{Var}_\theta\left(\frac{\partial}{\partial\theta}\log f_{\boldsymbol{X}}(\boldsymbol{X} \mid \theta)\right)} \\ &= \frac{(\frac{d}{d\theta}E_\theta[W(\boldsymbol{X})])^2}{E_\theta\left[(\frac{\partial}{\partial\theta}\log f_{\boldsymbol{X}}(\boldsymbol{X} \mid \theta))^2\right]},\end{aligned}$$

the Cauchy-Schwarz Inequality is used where

- X is chosen to be $W(\boldsymbol{X})$ whose mean is $\tau(\theta)$.
- Y is chosen to be $\frac{\partial}{\partial\theta}\log f_{\boldsymbol{X}}(\boldsymbol{X} \mid \theta)$ whose mean is 0.

We have *equality* (attainment of CRLB) when the correlation of $W(\boldsymbol{X})$ and $\frac{\partial}{\partial\theta}\log f_{\boldsymbol{X}}(\boldsymbol{X} \mid \theta)$ equals ± 1, i.e.,

$$\begin{aligned}c[W(\boldsymbol{x}) - \tau(\theta)] &= \frac{\partial}{\partial\theta}\log f_{\boldsymbol{X}}(\boldsymbol{x} \mid \theta) - 0 \\ &= \frac{\partial}{\partial\theta}\log L(\theta \mid \boldsymbol{x}).\end{aligned}$$