Math 362: Mathematical Statistics II

Le Chen
le.chen@emory.edu
chenle02@gmail.com

Emory University
Atlanta, GA

Last updated on Spring 2021
Last compiled on January 15, 2023

2021 Spring

Chapter 5. Estimation

Chapter 5. Estimation

§ 5.5 MVE: The Cramér-Rao Lower Bound

Question: Can one identify the unbiased estimator having the *smallest* variance?

Short answer: In many cases, yes!

We are going to develop the theory to answer this question in details!

Regular Estimation/Condition: The set of y (resp. k) values, where $f_Y(y;\theta) \neq 0$ (resp. $p_X(k;\theta) \neq 0$), does not depend on θ.

i.e., the domain of the pdf does not depend on the parameter (so that one can differentiate under integration).

Definition. The **Fisher's Information** of a continous (resp. discrete) random variable Y (resp. X) with pdf $f_Y(y;\theta)$ (resp. $p_X(k;\theta)$) is defined as

$$I(\theta) = \mathbb{E}\left[\left(\frac{\partial \ln f_Y(Y;\theta)}{\partial\theta}\right)^2\right] \qquad \left(\text{resp.}\quad \mathbb{E}\left[\left(\frac{\partial \ln p_X(X;\theta)}{\partial\theta}\right)^2\right]\right).$$

Lemma. Under regular condition, let $Y_1, \cdots, Y_n$ be a random sample of size n from the continuous population pdf $f_Y(y;\theta)$. Then the Fisher Information in the random sample $Y_1, \cdots, Y_n$ equals n times the Fisher information in X:

$$\mathbb{E}\left[\left(\frac{\partial \ln f_{Y_1,\cdots,Y_n}(Y_1,\cdots,Y_n;\theta)}{\partial\theta}\right)^2\right] = n\,\mathbb{E}\left[\left(\frac{\partial \ln f_Y(Y;\theta)}{\partial\theta}\right)^2\right] = n\,I(\theta). \quad (1)$$

(A similar statement holds for the discrete case $p_X(k;\theta)$).

Proof. Based on two observations:

$$LHS = \mathbb{E}\left[\left(\sum_{i=1}^{n} \frac{\partial}{\partial\theta} \ln f_{Y_i}(Y_i;\theta)\right)^2\right]$$

$$\mathbb{E}\left(\frac{\partial}{\partial\theta}\ln f_{Y_i}(Y_i;\theta)\right) = \int_{\mathbb{R}} \frac{\frac{\partial}{\partial\theta} f_Y(y;\theta)}{f_Y(y;\theta)} f_Y(y;\theta)\mathrm{d}y = \int_{\mathbb{R}} \frac{\partial}{\partial\theta} f_Y(y;\theta)\mathrm{d}y$$

$$\overset{\text{R.C.}}{=} \frac{\partial}{\partial\theta}\int_{\mathbb{R}} f_Y(y;\theta)\mathrm{d}y = \frac{\partial}{\partial\theta}1 = 0.$$

□

Lemma. Under regular condition, if $\ln f_Y(y;\theta)$ is twice differentiable in θ, then

$$I(\theta) = -\mathbb{E}\left[\frac{\partial^2}{\partial\theta^2}\ln f_Y(Y;\theta)\right]. \tag{2}$$

(A similar statement holds for the discrete case $p_X(k;\theta)$).

Proof. This is due to the two facts:

$$\frac{\partial^2}{\partial\theta^2}\ln f_Y(Y;\theta) = \frac{\frac{\partial^2}{\partial\theta^2} f_Y(Y;\theta)}{f_Y(Y;\theta)} - \underbrace{\left(\frac{\frac{\partial}{\partial\theta} f_Y(Y;\theta)}{f_Y(Y;\theta)}\right)^2}_{=\left(\frac{\partial}{\partial\theta}\ln f_Y(Y;\theta)\right)^2}$$

$$\mathbb{E}\left(\frac{\frac{\partial^2}{\partial\theta^2} f_Y(Y;\theta)}{f_Y(Y;\theta)}\right) = \int_{\mathbb{R}} \frac{\frac{\partial^2}{\partial\theta^2} f_Y(y;\theta)}{f_Y(y;\theta)} f_Y(y;\theta)\mathrm{d}y = \int_{\mathbb{R}} \frac{\partial^2}{\partial\theta^2} f_Y(y;\theta)\mathrm{d}y.$$

$$\overset{R.C.}{=} \frac{\partial^2}{\partial\theta^2}\int_{\mathbb{R}} f_Y(y;\theta)\mathrm{d}y = \frac{\partial^2}{\partial\theta^2}1 = 0.$$

□

Theorem (Cramér-Rao Inequality) Under regular condition, let $Y_1, \cdots, Y_n$ be a random sample of size n from the continuous population pdf $f_Y(y;\theta)$. Let $\widehat{\theta} = \widehat{\theta}(Y_1, \cdots, Y_n)$ be any unbiased estimator for θ. Then

$$\mathrm{Var}(\widehat{\theta}) \geq \frac{1}{n\, I(\theta)}.$$

(A similar statement holds for the discrete case $p_X(k;\theta)$).

Proof. If $n = 1$, then by Cauchy-Schwartz inequality,

$$\mathbb{E}\left[(\widehat{\theta}-\theta)\frac{\partial}{\partial\theta}\ln f_Y(Y;\theta)\right] \leq \sqrt{\mathrm{Var}(\widehat{\theta}) \times I(\theta)}$$

On the other hand,

$$\begin{aligned}
\mathbb{E}\left[(\widehat{\theta}-\theta)\frac{\partial}{\partial\theta}\ln f_Y(Y;\theta)\right] &= \int_{\mathbb{R}} (\widehat{\theta}-\theta)\frac{\frac{\partial}{\partial\theta} f_Y(y;\theta)}{f_Y(y;\theta)} f_Y(y;\theta)\mathrm{d}y \\
&= \int_{\mathbb{R}} (\widehat{\theta}-\theta)\frac{\partial}{\partial\theta} f_Y(y;\theta)\mathrm{d}y \\
&= \frac{\partial}{\partial\theta}\underbrace{\int_{\mathbb{R}} (\widehat{\theta}-\theta) f_Y(y;\theta)\mathrm{d}y}_{=\mathbb{E}(\widehat{\theta}-\theta)=0} + 1 = 1.
\end{aligned}$$

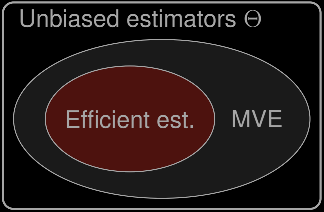
For general n, apply for (1). $\square$.

Definition. Let Θ be the set of all estimators $\widehat{\theta}$ that are unbiased for the parameter θ. We say that $\widehat{\theta}^*$ is a **best** or **minimum-variance** esimator (MVE) if $\widehat{\theta}^* \in \Theta$ and

$$\mathrm{Var}(\widehat{\theta}^*) \leq \mathrm{Var}(\widehat{\theta}) \qquad \text{for all } \widehat{\theta} \in \Theta.$$

Definition. An unbiased estimator $\widehat{\theta}$ is **efficient** if $\mathrm{Var}(\widehat{\theta})$ is equal to the Cramér-Rao lower bound, i.e., $\mathrm{Var}\widehat{\theta} = (n\, I(\theta))^{-1}$.

The **efficiency** of an unbiased estimator $\widehat{\theta}$ is defined to be $\left(nI(\theta)\mathrm{Var}(\widehat{\theta})\right)^{-1}$.



E.g. 1. $X \sim$Bernoulli(p). Check whether $\widehat{p} = \overline{X}$ is efficient?

Step 1. Compute Fisher's Information:

$$p_X(k;p) = p^k(1-p)^{1-k}.$$

$$\ln p_X(k;p) = k \ln p + (1-k)\ln(1-p)$$

$$\frac{\partial}{\partial p} \ln p_X(k;p) = \frac{k}{p} - \frac{1-k}{1-p}$$

$$-\frac{\partial^2}{\partial^2 p} \ln p_X(k;p) = \frac{k}{p^2} + \frac{1-k}{(1-p)^2}$$

$$-\mathbb{E}\left[\frac{\partial^2}{\partial^2 p} \ln p_X(X;p)\right] = \mathbb{E}\left[\frac{X}{p^2} + \frac{1-X}{(1-p)^2}\right] = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{pq}.$$

$$\boxed{I(p) = \frac{1}{pq}, \quad q = 1-p.}$$

Step 2. Compute Var($\widehat{p}$).

$$\text{Var}(\widehat{p}) = \frac{1}{n^2}\text{Var}\left(\sum_{i=1}^n X_i\right) = \frac{1}{n^2} npq = \frac{pq}{n}$$

Conclusion Because $\widehat{p}$ is unbiased and Var($\widehat{p}$) $= (nI(p))^{-1}$, $\widehat{p}$ is efficient.

E.g. 2. Exponential distr.: $f_Y(y;\lambda) = \lambda e^{-\lambda y}$ for $y \geq 0$. Is $\widehat{\lambda} = 1/\overline{Y}$ efficent?

Answer No, because $\widehat{\lambda}$ is biased. Nevertheless, we can still compute Fisher's Information as follows

Fisher's Inf.

$$\ln f_Y(y;\lambda) = \ln\lambda - \lambda y$$

$$\frac{\partial}{\partial\lambda}\ln f_Y(y;\lambda) = \frac{1}{\lambda} - y$$

$$-\frac{\partial^2}{\partial^2\lambda}\ln f_Y(y;\lambda) = \frac{1}{\lambda^2}$$

$$-\mathbb{E}\left[\frac{\partial^2}{\partial^2\lambda}\ln f_Y(Y;\lambda)\right] = \mathbb{E}\left[\frac{1}{\lambda^2}\right] = \frac{1}{\lambda^2}.$$

$$\boxed{I(\lambda) = \lambda^{-2}}$$

Try: $\widehat{\lambda}^* := \frac{n-1}{n}\frac{1}{\overline{Y}}$. It is unbiased. Is it efficient?

E.g. 2'. Exponential distr.: $f_Y(y;\theta) = \theta^{-1} e^{-y/\theta}$ for $y \geq 0$. $\widehat{\theta} = \overline{Y}$ efficent?

Step. 1. Compute Fisher's Information:

$$\ln f_Y(y;\theta) = -\ln\theta - \frac{y}{\theta}$$

$$\frac{\partial}{\partial\theta}\ln f_Y(y;\theta) = -\frac{1}{\theta} + \frac{y}{\theta^2}$$

$$-\frac{\partial^2}{\partial^2\theta}\ln f_Y(y;\theta) = -\frac{1}{\theta^2} + \frac{2y}{\theta^3}$$

$$-\mathbb{E}\left[\frac{\partial^2}{\partial^2\theta}\ln f_Y(Y;\theta)\right] = \mathbb{E}\left[-\frac{1}{\theta^2} + \frac{2Y}{\theta^3}\right] = -\frac{1}{\theta^2} + \frac{2\theta}{\theta^3} = \theta^{-2}.$$

$$\boxed{I(\theta) = \theta^{-2}}$$

Step 2. Compute $\mathrm{Var}(\widehat{\theta})$:

$$\mathrm{Var}(\overline{Y}) = \frac{1}{n^2}\sum_{i=1}^{n}\mathrm{Var}(Y_i) = \frac{1}{n^2}n\theta^2 = \frac{\theta^2}{n}.$$

Conclusion. Because $\widehat{\theta}$ is unbiased and $\mathrm{Var}(\widehat{p}) = (nI(p))^{-1}$, $\widehat{\theta}$ is efficient.

E.g. 3. $f_Y(y;\theta) = 2y/\theta^2$ for $y \in [0,\theta]$. $\widehat{\theta} = \frac{3}{2}\overline{Y}$ efficent?

Step. 1. Compute Fisher's Information:

$$\ln f_Y(y;\theta) = \ln(2y) - 2\ln\theta$$

$$\frac{\partial}{\partial\theta}\ln f_Y(y;\theta) = -\frac{2}{\theta}$$

By the definition of Fisher's information,

$$I(\theta) = \mathbb{E}\left[\left(\frac{\partial}{\partial\theta}\ln f_Y(y;\theta)\right)^2\right] = \mathbb{E}\left[\left(-\frac{2}{\theta}\right)^2\right] = \frac{4}{\theta^2}.$$

However, if we compute

$$-\frac{\partial^2}{\partial^2\theta}\ln f_Y(y;\theta) = -\frac{2}{\theta^2}$$

$$-\mathbb{E}\left[\frac{\partial^2}{\partial^2\theta}\ln f_Y(Y;\theta)\right] = \mathbb{E}\left[-\frac{2}{\theta^2}\right] = -\frac{2}{\theta^2} \neq \frac{4}{\theta^2} = I(\theta). \qquad (\dagger)$$

Step 2. Compute $\mathrm{Var}(\widehat{\theta})$:

$$\mathrm{Var}(\widehat{\theta}) = \frac{9}{4n}\mathrm{Var}(Y) = \frac{9}{4n}\frac{\theta^2}{18} = \frac{\theta^2}{8n}.$$

Discussion. Even though $\widehat{\theta}$ is unbiased, we have two discripencies: (†) and

$$\mathrm{Var}(\widehat{\theta}) = \frac{\theta^2}{8n} \leq \frac{\theta^2}{4n} = \frac{1}{nI(\theta)}$$

This is because this is not a regular estimation!