

Math 362: Mathematical Statistics II

Le Chen

le.chen@emory.edu

Emory University
Atlanta, GA

Last updated on February 6, 2020

Jan. 16th, 2020

Chapter 5: Estimation

Chapter 5. Estimation

§ 5.1 Introduction

§ 5.2 Estimating parameters: MLE and MME

§ 5.3 Interval Estimation

§ 5.4 Properties of Estimators

§ 5.5 Minimum-Variance Estimators: The Cramér-Rao Lower Bound

§ 5.4 Properties of Estimators

Question: Estimators are not in general unique (MLE or MME ...). How to select one estimator?

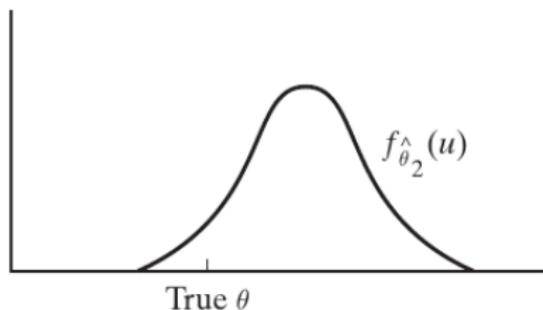
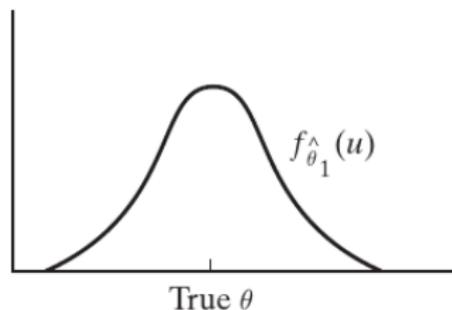
Recall: For a random sample of size n from the population with given pdf, we have X_1, \dots, X_n , which are i.i.d. r.v.'s. The estimator $\hat{\theta}$ is a function of X_i 's:

$$\hat{\theta} = \hat{\theta}(X_1, \dots, X_n).$$

Criteria:

1. Unbiased. (Mean)
2. Efficiency, the minimum-variance estimator. (Variance)
3. Sufficiency.
4. Consistency. (Asymptotic behavior)

Unbiasedness



Definition 5.4.1. Given a random sample of size n whose population distribution depends on an unknown parameter θ , let $\hat{\theta}$ be an estimator of θ .

Then $\hat{\theta}$ is called **unbiased** if $\mathbb{E}(\hat{\theta}) = \theta$;

and $\hat{\theta}$ is called **asymptotically unbiased** if $\lim_{n \rightarrow \infty} \mathbb{E}(\hat{\theta}) = \theta$.

E.g. 1. $f_Y(y; \theta) = \frac{2y}{\theta^2}$ if $y \in [0, \theta]$.

▶ $\hat{\theta}_1 = \frac{3}{2} \bar{Y}$

▶ $\hat{\theta}_2 = Y_{max}$.

▶ $\hat{\theta}_3 = \frac{2n+1}{2n} Y_{max}$.

E.g. 2. Let X_1, \dots, X_n be a random sample of size n with the unknown parameter $\theta = \mathbb{E}(X)$. Show that for any constants a_i 's,

$$\hat{\theta} = \sum_{i=1}^n a_i X_i \text{ is unbiased} \iff \sum_{i=1}^n a_i = 1.$$

E.g. 3. Let X_1, \dots, X_n be a random sample of size n with the unknown parameter $\sigma^2 = \text{Var}(X)$.

$$\blacktriangleright \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$\blacktriangleright S^2 = \text{Sample Variance} = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

$$\blacktriangleright S = \text{Sample Standard Deviation} = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2}. \quad (\text{Biased for } \sigma!)$$

E.g. 4. Exponential distr.: $f_Y(y; \lambda) = \lambda e^{-\lambda y}$ for $y \geq 0$. $\hat{\lambda} = 1/\bar{Y}$ is biased.

$n\bar{Y} = \sum_{i=1}^n Y_i \sim \text{Gamma distribution}(n, \lambda)$. Hence,

$$\begin{aligned}\mathbb{E}(\hat{\lambda}) &= \mathbb{E}(1/\bar{Y}) = n \int_0^{\infty} \frac{1}{y} \frac{\lambda^n}{\Gamma(n)} y^{n-1} e^{-\lambda y} dy \\ &= \frac{n\lambda}{n-1} \int_0^{\infty} \underbrace{\frac{\lambda^{n-1}}{\Gamma(n-1)} y^{(n-1)-1} e^{-\lambda y}}_{\text{pdf for Gamma distr. } (n-1, \lambda)} dy \\ &= \frac{n}{n-1} \lambda.\end{aligned}$$

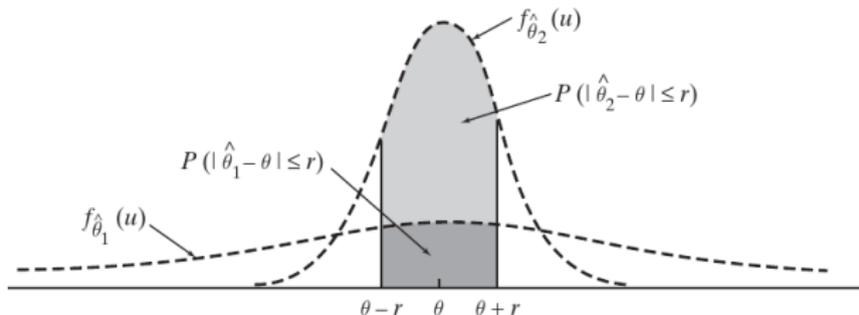
Biased! But $\mathbb{E}(\hat{\lambda}) = \frac{n}{n-1} \lambda \rightarrow \lambda$ as $n \rightarrow \infty$. (Asymptotically unbiased.)

Note: $\hat{\lambda}^* = \frac{n-1}{n\bar{Y}}$ is unbiased.

E.g. 4'. Exponential distr.: $f_Y(y; \theta) = \frac{1}{\theta} e^{-y/\theta}$ for $y \geq 0$. $\hat{\theta} = \bar{Y}$ is unbiased.

$$\mathbb{E}(\hat{\theta}) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(Y_i) = \frac{1}{n} \sum_{i=1}^n \theta = \theta.$$

Efficiency



Definition 5.4.2. Let $\hat{\theta}_1$ and $\hat{\theta}_2$ be two unbiased estimators for a parameter θ . If $\text{Var}(\hat{\theta}_1) < \text{Var}(\hat{\theta}_2)$, then we say that $\hat{\theta}_1$ is **more efficient** than $\hat{\theta}_2$. The **relative efficiency** of $\hat{\theta}_1$ w.r.t. $\hat{\theta}_2$ is the ratio $\text{Var}(\hat{\theta}_2)/\text{Var}(\hat{\theta}_1)$.

E.g. 1. $f_Y(y; \theta) = \frac{2y}{\theta^2}$ if $y \in [0, \theta]$. Which is more efficient? Find the relative efficiency of $\hat{\theta}_1$ w.r.t. $\hat{\theta}_3$.

$$\blacktriangleright \hat{\theta}_1 = \frac{3}{2} \bar{Y}$$

$$\blacktriangleright \hat{\theta}_3 = \frac{2n+1}{2n} Y_{\max}.$$

E.g. 2. Let X_1, \dots, X_n be a random sample of size n with the unknown parameter $\theta = \mathbb{E}(X)$ (suppose $\sigma^2 = \text{Var}(X) < \infty$).

Among all possible unbiased estimators $\hat{\theta} = \sum_{i=1}^n a_i X_i$ with $\sum_{i=1}^n a_i = 1$. Find the most efficient one.

Sol:

$$\text{Var}(\hat{\theta}) = \sum_{i=1}^n a_i^2 \text{Var}(X) = \sigma^2 \sum_{i=1}^n a_i^2 \geq \sigma^2 2^{n-1} \left(\sum_{i=1}^n a_i \right)^2 = 2^{n-1} \sigma^2,$$

with equality iff $a_1 = \dots = a_n = 1/n$.

Hence, the most efficient one is the sample mean $\hat{\theta} = \bar{X}$. □

Chapter 5. Estimation

§ 5.1 Introduction

§ 5.2 Estimating parameters: MLE and MME

§ 5.3 Interval Estimation

§ 5.4 Properties of Estimators

§ 5.5 Minimum-Variance Estimators: The Cramér-Rao Lower Bound

§ 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 continuous (resp. discrete) random variable Y (resp. X) with pdf $f_Y(y; \theta)$ (resp. $p_X(k; \theta)$) is defined as

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

Lemma. Under regular condition, let Y_1, \dots, 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, \dots, Y_n equals n times the Fisher information in X :

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

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

Proof. Based on two observations:

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

$$\begin{aligned} \mathbb{E} \left(\frac{\partial}{\partial \theta} \ln f_{Y_i}(Y_i; \theta) \right) &= \int_{\mathbb{R}} \frac{\frac{\partial}{\partial \theta} f_Y(y; \theta)}{f_Y(y; \theta)} f_Y(y; \theta) dy = \int_{\mathbb{R}} \frac{\partial}{\partial \theta} f_Y(y; \theta) dy \\ &\stackrel{\text{R.C.}}{=} \frac{\partial}{\partial \theta} \int_{\mathbb{R}} f_Y(y; \theta) dy = \frac{\partial}{\partial \theta} 1 = 0. \end{aligned}$$

□

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

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

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

Proof. This is due to the two facts:

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

$$\begin{aligned} \mathbb{E} \left(\frac{\frac{\partial^2}{\partial \theta^2} f_Y(Y; \theta)}{f_Y(Y; \theta)} \right) &= \int_{\mathbb{R}} \frac{\frac{\partial^2}{\partial \theta^2} f_Y(y; \theta)}{f_Y(y; \theta)} f_Y(y; \theta) dy = \int_{\mathbb{R}} \frac{\partial^2}{\partial \theta^2} f_Y(y; \theta) dy. \\ &\stackrel{R.C.}{=} \frac{\partial^2}{\partial \theta^2} \int_{\mathbb{R}} f_Y(y; \theta) dy = \frac{\partial^2}{\partial \theta^2} 1 = 0. \end{aligned}$$

□

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

$$\text{Var}(\hat{\theta}) \geq \frac{1}{n I(\theta)}.$$

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

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

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

On the other hand,

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

For general n , apply for (1).

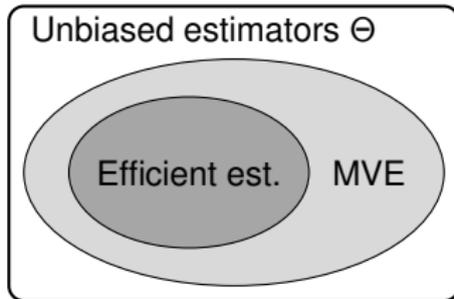
□.

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

$$\text{Var}(\hat{\theta}^*) \leq \text{Var}(\hat{\theta}) \quad \text{for all } \hat{\theta} \in \Theta.$$

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

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



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

Step 1. Compute Fisher's Information:

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

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

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

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

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

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

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

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

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

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

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

Fisher's Inf.

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

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

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

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

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

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

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

Step 1. Compute Fisher's Information:

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

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

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

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

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

Step 2. Compute $\text{Var}(\hat{\theta})$:

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

Conclusion. Because $\hat{\theta}$ is unbiased and $\text{Var}(\hat{\rho}) = (nI(\rho))^{-1}$, $\hat{\theta}$ is efficient.

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

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

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

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

Step 2. Compute $\text{Var}(\hat{\theta})$:

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

Discussion. Even though $\hat{\theta}$ is unbiased, we have two discrepancies: (\dagger) and

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

This is because this is not a regular estimation!