

# STAT 512 su 2021 Lec 10 slides

## First principles of estimation: sufficiency and Rao-Blackwell theorem

Karl B. Gregory

University of South Carolina

These slides are an instructional aid; their sole purpose is to display, during the lecture, definitions, plots, results, etc. which take too much time to write by hand on the blackboard. They are not intended to explain or expound on any material.

We want to estimate  $\theta$  based on  $(X_1, \dots, X_n) \sim f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta)$ .

How do we build an estimator of  $\theta$ ?

A good estimator will use all the information about  $\theta$  contained in  $X_1, \dots, X_n$ .

A *sufficient statistic*  $T(X_1, \dots, X_n)$  carries all the info. about  $\theta$  from  $X_1, \dots, X_n$ .



## Sufficient statistic

A statistic  $T(X_1, \dots, X_n)$  is *sufficient for*  $\theta$  if the joint pdf/pmf of  $X_1, \dots, X_n$ , conditional on the value of  $T(X_1, \dots, X_n)$ , is free of  $\theta$ .

Affect of  $\theta$  on the distribution of  $X_1, \dots, X_n$  is expressed fully in  $T(X_1, \dots, X_n)$ .

Easier to remember: “ $T(X_1, \dots, X_n)$  contains all info about  $\theta$ ”.

**Example:** Let  $X_1, X_2, X_3 \stackrel{\text{ind}}{\sim} \text{Poisson}(\lambda)$ , estimator for  $\lambda$  wanted.

We find that the sample outcomes

$$(X_1, X_2, X_3) = (1, 3, 5) \quad \text{and} \quad (X_1, X_2, X_3) = (1, 1, 7)$$

have the same thing to say about  $\lambda$ . Why??

## Theorem (Checking for sufficiency in the discrete case)

For discrete

$$(X_1, \dots, X_n) \sim p_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta),$$

$T(X_1, \dots, X_n)$  is a sufficient statistic for  $\theta$  if

$$\frac{p_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta)}{p_T(T(x_1, \dots, x_n); \theta)}$$

is free of  $\theta$  for all  $(x_1, \dots, x_n)$  in the support of  $(X_1, \dots, X_n)$ , where

$$T(X_1, \dots, X_n) \sim p_T(\cdot; \theta).$$

Note that for a random sample  $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} p_X(\cdot; \theta)$ , the joint pmf is

$$p_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) = \prod_{i=1}^n p_X(x_i; \theta)$$

**Exercise:** Let  $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Poisson}(\lambda)$ .

- 1 Check whether  $T(X_1, \dots, X_n) = \sum_{i=1}^n X_i$  is a sufficient statistic for  $\lambda \dots$
- 2 Interpret.

**Instructor:** Prove previous theorem. Illustrate with  $X_1, X_2, X_3 \stackrel{\text{ind}}{\sim} \text{Bernoulli}(p)$ .

**Exercise:** Let  $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Bernoulli}(p)$ .

Check whether  $T(X_1, \dots, X_n) = \sum_{i=1}^n X_i$  is a sufficient statistic for  $p$ .



## Theorem (Checking for sufficiency in the continuous case)

For continuous

$$(X_1, \dots, X_n) \sim f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta),$$

$T(X_1, \dots, X_n)$  is a sufficient statistic for  $\theta$  if

$$\frac{f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta)}{f_T(T(x_1, \dots, x_n); \theta)}$$

is free of  $\theta$  for all  $(x_1, \dots, x_n)$  in the support of  $(X_1, \dots, X_n)$ , where

$$T(X_1, \dots, X_n) \sim f_T(\cdot; \theta).$$

Note that for a random sample  $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} f_X(\cdot; \theta)$ , the joint pdf is

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f_X(x_i; \theta)$$

**Exercise:** Let  $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Uniform}(0, \theta)$ .

Check whether  $T(X_1, \dots, X_n) = X_{(n)}$  is a sufficient statistic for  $\theta$ .

**Exercise:** Let  $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Exponential}(\lambda)$ .

Check whether  $T(X_1, \dots, X_n) = \sum_{i=1}^n X_i$  is a sufficient statistic for  $\lambda$ .

## Theorem (Easier check for sufficiency by factorization)

For

$$(X_1, \dots, X_n) \sim f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta),$$

$T(X_1, \dots, X_n)$  is a sufficient statistic for  $\theta$  if

$$f_{X_1, \dots, X_n}(x_1, \dots, x_n; \theta) = g(T(x_1, \dots, x_n); \theta) \cdot h(x_1, \dots, x_n)$$

for all  $\theta \in \Theta$  and all  $(x_1, \dots, x_n)$  in the support of  $(X_1, \dots, X_n)$  for some functions

$$\underbrace{g(t; \theta)}_{\text{function of } t \text{ and } \theta} \quad \text{and} \quad \underbrace{h(x_1, \dots, x_n)}_{\text{does not involve } \theta}$$

Also works for the discrete case.

Can be used to *find* a sufficient statistic.

**Exercise:** Let  $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Beta}(1, \theta)$ .

Check whether  $T(X_1, \dots, X_n) = \prod_{i=1}^n (1 - X_i)$  is a sufficient statistic for  $\theta$ .

**Exercise:** Let  $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} f_X(x; \mu) = \pi^{-1}[1 + (x - \mu)^2]^{-1}$ .

Check whether  $T(X_1, \dots, X_n) = \sum_{i=1}^n X_i$  is a sufficient statistic for  $\mu$ .

A statistic may consist of multiple functions of the data:

$$T(X_1, \dots, X_n) = (T_1(X_1, \dots, X_n), \dots, T_K(X_1, \dots, X_n)), \quad K \geq 1.$$

A parameter  $\theta$  may consist of several values:

$$\theta = (\theta_1, \dots, \theta_d), \quad d \geq 1.$$

We can still establish sufficiency using the factorization theorem.

**Exercise:** Let  $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$ .

Check whether  $T(X_1, \dots, X_n) = (\bar{X}_n, S_n^2)$  is a sufficient statistic for  $(\mu, \sigma^2)$ .

**Exercise:** Let  $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Gamma}(\alpha, \beta)$ .

Find a sufficient statistic for  $(\alpha, \beta)$  by factorization.



**Exercise:** Let  $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} f_X(x; \theta)$ . Show that

$$T(X_1, \dots, X_n) = (X_{(1)}, \dots, X_{(n)})$$

is sufficient for  $\theta$ .

We can use sufficient statistics to find “good” estimators.

Specifically, we can find the unbiased estimator with the smallest possible variance.



## Minimum variance unbiased estimator (MVUE)

Let  $\hat{\theta}$  be an unbiased estimator of  $\theta \in \Theta \subset \mathbb{R}$ . If

$$\text{Var } \hat{\theta} \leq \text{Var } \tilde{\theta}$$

for every unbiased estimator  $\tilde{\theta}$  of  $\theta$ , then  $\hat{\theta}$  is a *minimum-variance unbiased estimator (MVUE)* for  $\theta$ .

## Theorem (Rao-Blackwell)

Let  $\tilde{\theta}$  be an unbiased estimator of  $\theta \in \Theta \subset \mathbb{R}$  and  $T$  be a sufficient statistic for  $\theta$ . Moreover define

$$\hat{\theta} = \mathbb{E}[\tilde{\theta} | T].$$

Then  $\hat{\theta}$  is an estimator of  $\theta$  such that

$$\mathbb{E}\hat{\theta} = \theta \quad \text{and} \quad \text{Var } \hat{\theta} \leq \text{Var } \tilde{\theta}.$$

Says in a complicated way that MVUEs are always functions of sufficient statistics.

Are *you* trying to find an MVUE? Use a sufficient statistic!

**Instructor:** Show proof of the Rao-Blackwell Theorem.

## Theorem (Uniqueness of the MVUE)

*If  $\hat{\theta}$  is a MVUE for  $\theta$ , it is unique, i.e. there is no other unbiased estimator with variance as small.*

Now a recipe for *the* MVUE:

- 1 Find a sufficient statistic  $T(X_1, \dots, X_n)$  for  $\theta$ .
- 2 Find a function of  $T(X_1, \dots, X_n)$  which is unbiased for  $\theta$ .  
This is the MVUE.



Works also for finding the MVUE of a function  $\tau = \tau(\theta)$  of the parameter:  
Just find a function of  $T(X_1, \dots, X_n)$  which is unbiased for  $\tau$ .

**Exercise:** Let  $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Exponential}(\lambda)$ .

- 1 Find the MVUE for  $\lambda$ .
- 2 Find the MVUE for  $\gamma = \lambda^2$ .

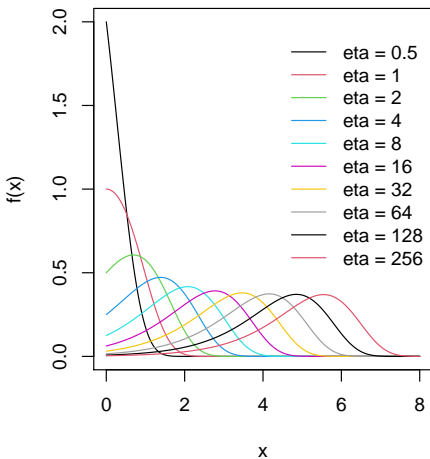
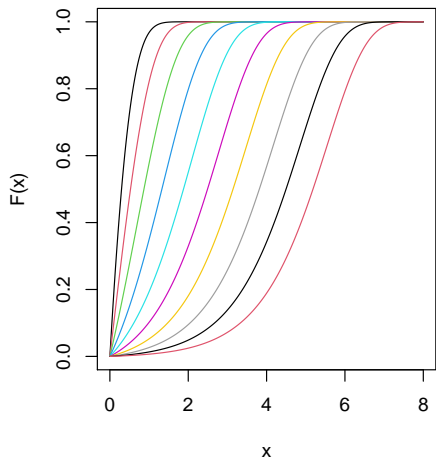
**Exercise:** Let  $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Bernoulli}(p)$ . Find the MVUE for  $p$ .

**Exercise:** Let  $X_1, \dots, X_n$  be a rs from a distribution with cdf given by

$$F_X(x; \eta) = \begin{cases} 1 - e^{-(e^x - 1)/\eta} & x \geq 0 \\ 0 & x < 0. \end{cases}$$

- ① Show that  $\sum_{i=1}^n e^{X_i}$  is a sufficient statistic for  $\eta$ .
- ② Argue that  $\hat{\eta} = n^{-1} \sum_{i=1}^n (e^{X_i} - 1)$  is the MVUE for  $\eta$ .

*Hint: Let  $Y_i = e^{X_i} - 1$ ,  $i = 1, \dots, n$ .*





**Exercise:** Let  $X_1, \dots, X_n \stackrel{\text{ind}}{\sim} \text{Normal}(\mu, \sigma^2)$ .

Argue whether  $\bar{X}_n$  and  $S_n^2$  the MVUEs for  $\mu$  and  $\sigma^2$ , respectively.