

**UCLA STAT 110 A**  
**Applied Probability & Statistics for Engineers**

- **Instructor:** Ivo Dinov,  
Asst. Prof. In Statistics and Neurology
- **Teaching Assistant:** Neda Farzinnia, UCLA Statistics

University of California, Los Angeles, Spring 2004  
<http://www.stat.ucla.edu/~dinov/>

Stat 110A, UCLA, Ivo Dinov Slide 1

Chapter 6  
**Point Estimation**

Slide 2 Stat 110A, UCLA, Ivo Dinov

6.1  
**General Concepts of Point Estimation**

Slide 3 Stat 110A, UCLA, Ivo Dinov

Point Estimator

A *point estimator* of a parameter  $\theta$  is a single number that can be regarded as a sensible value for  $\theta$ . A point estimator can be obtained by selecting a suitable statistic and computing its value from the given sample data.

Slide 4 Stat 110A, UCLA, Ivo Dinov

Unbiased Estimator

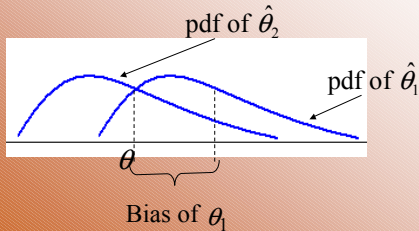
A *point estimator*  $\hat{\theta}$  is said to be an unbiased estimator of  $\theta$  if  $E(\hat{\theta}) = \theta$  for every possible value of  $\theta$ . If  $\hat{\theta}$  is not biased, the difference  $E(\hat{\theta}) - \theta$  is called the *bias* of  $\hat{\theta}$ .

Slide 5 Stat 110A, UCLA, Ivo Dinov

The pdf's of a biased estimator  $\hat{\theta}_1$  and an unbiased estimator  $\hat{\theta}_2$  for a parameter  $\theta$ .

Slide 6 Stat 110A, UCLA, Ivo Dinov

The pdf's of a biased estimator  $\hat{\theta}_1$  and an unbiased estimator  $\hat{\theta}_2$  for a parameter  $\theta$ .



Slide 7 Stat 110A, UCLA, Jon Dineen

## Unbiased Estimator

When  $X$  is a binomial rv with parameters  $n$  and  $p$ , the sample proportion  $\hat{p} = X/n$  is an unbiased estimator of  $p$ .

Slide 8 Stat 110A, UCLA, Jon Dineen

## Principle of Unbiased Estimation

When choosing among several different estimators of  $\theta$ , select one that is unbiased.

Slide 9 Stat 110A, UCLA, Jon Dineen

## Unbiased Estimator

Let  $X_1, X_2, \dots, X_n$  be a random sample from a distribution with mean  $\mu$  and variance  $\sigma^2$ . Then the estimator

$$\hat{\sigma}^2 = S^2 = \frac{\sum (X_i - \bar{X})^2}{n-1}$$

is an unbiased estimator.

Slide 10 Stat 110A, UCLA, Jon Dineen

## Unbiased Estimator

If  $X_1, X_2, \dots, X_n$  is a random sample from a distribution with mean  $\mu$ , then  $\bar{X}$  is an unbiased estimator of  $\mu$ . If in addition the distribution is continuous and symmetric, then  $\bar{X}$  and any trimmed mean are also unbiased estimators of  $\mu$ .

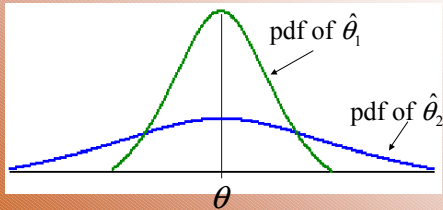
Slide 11 Stat 110A, UCLA, Jon Dineen

## Principle of Minimum Variance Unbiased Estimation

Among all estimators of  $\theta$  that are unbiased, choose the one that has the minimum variance. The resulting  $\hat{\theta}$  is called the *minimum variance unbiased estimator (MVUE)* of  $\theta$ .

Slide 12 Stat 110A, UCLA, Jon Dineen

Graphs of the pdf's of two different unbiased estimators



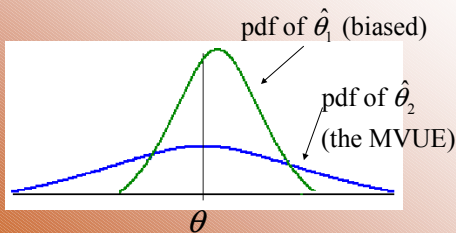
Slide 13 Stat 110A, UCLA, Leo Dineen

### MVUE for a Normal Distribution

Let  $X_1, X_2, \dots, X_n$  be a random sample from a normal distribution with parameters  $\mu$  and  $\sigma$ . Then the estimator  $\hat{\mu} = \bar{X}$  is the MVUE for  $\mu$ .

Slide 14 Stat 110A, UCLA, Leo Dineen

A biased estimator that is preferable to the MVUE



Slide 15 Stat 110A, UCLA, Leo Dineen

### The Estimator for $\mu$

$$\left( \bar{X}, \tilde{X}, \bar{X}_e, \bar{X}_{tr(10)} \right)$$

1. If the random sample comes from a normal distribution, then  $\bar{X}$  is the best estimator since it has minimum variance among all unbiased estimators.

Slide 16 Stat 110A, UCLA, Leo Dineen

### The Estimator for $\mu$

$$\left( \bar{X}, \tilde{X}, \bar{X}_e, \bar{X}_{tr(10)} \right)$$

2. If the random sample comes from a Cauchy distribution, then  $\tilde{X}$  is good (the MVUE is not known).  $\bar{X}$  and  $\bar{X}_e$  are quite bad.

Slide 17 Stat 110A, UCLA, Leo Dineen

### The Estimator for $\mu$

$$\left( \bar{X}, \tilde{X}, \bar{X}_e, \bar{X}_{tr(10)} \right)$$

3. If the underlying distribution is uniform, the best estimator is  $\bar{X}_e$  this estimator is influenced by outlying observations, but the lack of tails makes this impossible.

Slide 18 Stat 110A, UCLA, Leo Dineen

### The Estimator for $\mu$

$$\left(\bar{X}, \tilde{X}, \bar{X}_e, \bar{X}_{tr(10)}\right)$$

4. The trimmed mean  $\bar{X}_{tr(10)}$  works reasonably well in all three situations but is not the best for any of them.

Slide 19 Stat 110A, UCLA, Leo Dineen

### Standard Error

The *standard error* of an estimator  $\hat{\theta}$  is its standard deviation  $\sigma_{\hat{\theta}} = \sqrt{V(\hat{\theta})}$ . If the standard error itself involves unknown parameters whose values can be estimated, substitution into  $\sigma_{\hat{\theta}}$  yields the *estimated standard error* of the estimator, denoted  $\hat{\sigma}_{\hat{\theta}}$  or  $s_{\hat{\theta}}$ .

Slide 20 Stat 110A, UCLA, Leo Dineen

## 6.2

# Methods of Point Estimation

Slide 21 Stat 110A, UCLA, Leo Dineen

### Moments

Let  $X_1, X_2, \dots, X_n$  be a random sample from a pmf or pdf  $f(x)$ . For  $k = 1, 2, \dots$  the *kth population moment*, or *kth moment of the distribution  $f(x)$*  is  $E(X^k)$ . The *kth sample moment* is

$$\left(\frac{1}{n}\right) \sum_{i=1}^n X_i^k.$$

Slide 22 Stat 110A, UCLA, Leo Dineen

### Moment Estimators

Let  $X_1, X_2, \dots, X_n$  be a random sample from a distribution with pmf or pdf  $f(x; \theta_1, \dots, \theta_m)$ , where  $\theta_1, \dots, \theta_m$  are parameters whose values are unknown. Then the *moment estimators*  $\hat{\theta}_1, \dots, \hat{\theta}_m$  are obtained by equating the first  $m$  sample moments to the corresponding first  $m$  population moments and solving for  $\theta_1, \dots, \theta_m$ .

Slide 23 Stat 110A, UCLA, Leo Dineen

### Likelihood Function

Let  $X_1, X_2, \dots, X_n$  have joint pmf or pdf  $f(x_1, \dots, x_n; \theta_1, \dots, \theta_m)$  where parameters  $\theta_1, \dots, \theta_m$  have unknown values. When  $x_1, \dots, x_n$  are the observed sample values and  $f$  is regarded as a function of  $\theta_1, \dots, \theta_m$ , it is called the *likelihood function*.

Slide 24 Stat 110A, UCLA, Leo Dineen

## Maximum Likelihood Estimators

The maximum likelihood estimates (mle's)  $\hat{\theta}_1, \dots, \hat{\theta}_m$  are those values of the  $\theta_i$ 's that maximize the likelihood function so that

$$f(x_1, \dots, x_n; \hat{\theta}_1, \dots, \hat{\theta}_m) \geq f(x_1, \dots, x_n; \theta_1, \dots, \theta_m)$$

for all  $\theta_1, \dots, \theta_m$

When the  $X_i$ 's are substituted in the place of the  $x_i$ 's, the *maximum likelihood estimators* result.

Slide 25 Stat 110A, UCLA, Leo Dineen

## The Invariance Principle

Let  $\hat{\theta}_1, \dots, \hat{\theta}_m$  be the mle's of the parameters  $\theta_1, \dots, \theta_m$ . Then the mle of any function  $h(\theta_1, \dots, \theta_m)$  of these parameters is the function  $h(\hat{\theta}_1, \dots, \hat{\theta}_m)$  of the mle's.

Slide 26 Stat 110A, UCLA, Leo Dineen

## Desirable Property of the Maximum Likelihood Estimate

Under very general conditions on the joint distribution of the sample, when the sample size  $n$  is large, the maximum likelihood estimator of any parameter  $\theta$  is approx. unbiased [ $E(\hat{\theta}) \approx \theta$ ] and has variance that is nearly as small as can be achieved by any estimator.

$$\text{mle } \hat{\theta} \approx \text{MVUE of } \theta$$

Slide 27 Stat 110A, UCLA, Leo Dineen

## (Log)Likelihood Function

- Suppose we have a sample  $\{X_1, \dots, X_n\}$  IID  $D(\theta)$  with probability density function  $p = p(X | \theta)$ . Then the joint density  $p(\{X_1, \dots, X_n\} | \theta)$  is a function of the (unknown) parameter  $\theta$ .
- Likelihood function  $l(\theta | \{X_1, \dots, X_n\}) = p(\{X_1, \dots, X_n\} | \theta)$
- Log-likelihood  $L(\theta | \{X_1, \dots, X_n\}) = \text{Log}_e l(\theta | \{X_1, \dots, X_n\})$
- Maximum-likelihood estimation (MLE):
- Suppose  $\{X_1, \dots, X_n\}$  IID  $N(\mu, \sigma^2)$ ,  $\mu$  is unknown. We estimate it by:  $\text{MLE}(\mu) = \hat{\mu} = \text{ArgMax}_{\mu} L(\mu | \{X_1, \dots, X_n\})$

Slide 28 Stat 110A, UCLA, Leo Dineen

## (Log)Likelihood Function

- Suppose  $\{X_1, \dots, X_n\}$  IID  $N(\mu, \sigma^2)$ ,  $\mu$  is unknown. We estimate it by:  $\text{MLE}(\mu) = \hat{\mu} = \text{ArgMax}_{\mu} L(\mu | \{X_1, \dots, X_n\})$

$$\text{MLE}(\mu) = \text{Log} \left( \prod_{i=1}^n \frac{e^{-(x_i - \mu)^2 / 2\sigma^2}}{\sqrt{2\pi\sigma^2}} \right) = L(\mu)$$

$$0 = L'(\hat{\mu}) = \frac{1}{(2\pi\sigma^2)^{n/2}} \left( e^{-\sum_{i=1}^n (x_i - \hat{\mu})^2 / 2\sigma^2} \right) \frac{\sum_{i=1}^n 2(x_i - \hat{\mu})}{2\sigma^2}$$

$$\Leftrightarrow 0 = 2 \sum_{i=1}^n (x_i - \hat{\mu}) \Leftrightarrow \hat{\mu} = \frac{\sum_{i=1}^n x_i}{n}$$

Similarly can show that:  $\text{MLE}(\sigma) = \hat{\sigma} = \frac{\sum_{i=1}^n (x_i - \mu)^2}{n-1}$

Slide 29 Stat 110A, UCLA, Leo Dineen

## (Log)Likelihood Function

- Suppose  $\{X_1, \dots, X_n\}$  IID  $\text{Poisson}(\lambda)$ ,  $\lambda$  is unknown. Estimate  $\lambda$  by:  $\text{MLE}(\lambda) = \hat{\lambda} = \text{ArgMax}_{\lambda} L(\lambda | \{X_1, \dots, X_n\})$

$$\text{MLE}(\lambda) = \text{Log} \left( \prod_{i=1}^n \frac{e^{-\lambda} \lambda^{x_i}}{(x_i)!} \right) = L(\lambda)$$

$$0 = L'(\hat{\lambda}) = \frac{\partial}{\partial \lambda} \text{Log} \left( \frac{e^{-n\lambda} \lambda^{\sum_{i=1}^n x_i}}{\prod_{i=1}^n (x_i)!} \right) =$$

$$= \frac{\partial}{\partial \lambda} \left( -n\lambda + \text{Log}(\lambda)^{\sum_{i=1}^n x_i} \right) = -n + \frac{1}{\lambda} \sum_{i=1}^n x_i \Leftrightarrow \hat{\lambda} = \frac{\sum_{i=1}^n x_i}{n}$$

Slide 30 Stat 110A, UCLA, Leo Dineen