UCLA STAT 110 A

Applied Probability & Statistics for Engineers

•**Instructor**: **Ivo Dinov,**

Asst. Prof. In Statistics and Neurology

•**Teaching Assistant:** **Christopher Barr**

University of California, Los Angeles, Fall 2004

http://www.stat.ucla.edu/~dinov/

Chapter 6

Point Estimation

6.1

General Concepts of Point Estimation

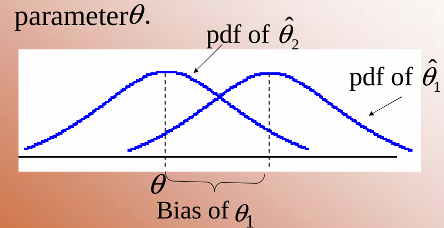
Point Estimator

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

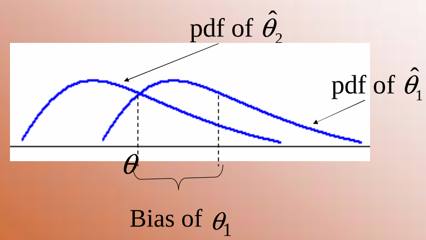
Unbiased Estimator

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

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



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



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.

Principle of Unbiased Estimation

When choosing among several different estimators of θ, select one that is unbiased.

Unbiased Estimator

Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with mean μ and variance σ^2. Then the estimator

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

is an unbiased estimator.

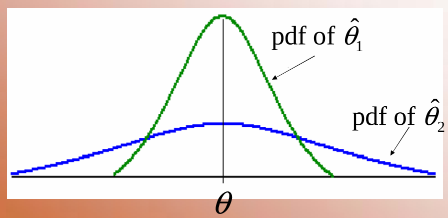
Unbiased Estimator

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

Principle of Minimum Variance Unbiased Estimation

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

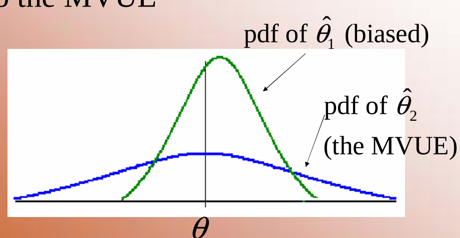
Graphs of the pdf's of two different unbiased estimators



MVUE for a Normal Distribution

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

A biased estimator that is preferable to the MVUE



The Estimator for μ

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

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

The Estimator for μ

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

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

The Estimator for μ

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

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

The Estimator for μ

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

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

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 $s_{\hat{\theta}}$.

6.2

Methods of Point Estimation

Moments

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

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

Moment Estimators

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

Likelihood Function

Let $X_1, X_2, \ldots, X_n$ have joint pmf or pdf

$$f(x_1, \ldots, x_n; \theta_1, \ldots, \theta_m)$$

where parameters $\theta_1, \ldots, \theta_m$ have unknown values. When $x_1, \ldots, x_n$ are the observed sample values and *f* is regarded as a function of $\theta_1, \ldots, \theta_m$, it is called the *likelihood function*.

Maximum Likelihood Estimators

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

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

for all $\theta_1, \ldots, \theta_m$

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

The Invariance Principle

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

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 θ is approx. unbiased $[E(\hat{\theta}) \approx \theta]$ and has variance that is nearly as small as can be achieved by any estimator.

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

(Log)Likelihood Function

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

(Log)Likelihood Function

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

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

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

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

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

(Log)Likelihood Function

- Suppose $\{X_1, \ldots, X_n\}$ IID Poisson(λ), λ is unknown. Estimate λ by: $MLE(\lambda) = \lambda\text{^} = \text{ArgMax}_\lambda L(\lambda \mid (\{X_1, \ldots, X_n\})$

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

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

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