CH.7 Sampling Distributions and Point Estimation of Parameters

- Introduction
  - Parameter estimation, sampling distribution, statistic, point estimator, point estimate
- Sampling distribution and the Central Limit Theorem
- General Concepts of Point Estimation
  - Unbiased estimators
  - Variance of a point estimator
  - Standard error of an estimator
  - Mean squared error of an estimator
- Methods of point estimation
  - Method of moments
  - Method of maximum likelihood

7-1 Introduction

- The field of *statistical inference* consists of those methods used to make decisions or to draw conclusions about a **population**.
- These methods utilize the information contained in a **sample** from the population in drawing conclusions.
- Statistical inference may be divided into two major areas:
    - Parameter estimation
    - Hypothesis testing

7-1 Introduction

Parameter estimation *examples*

•Estimate the mean fill volume of soft drink cans: Soft drink cans are filled by automated filling machines. The fill volume may vary because of differences in soft drinks, automated machines, and measurement procedures.

• Estimate the mean diameter of bottle openings of a specific bottle manufactured in a plant: The diameter of bottle openings may vary because of machines, molds and measurements.

7-1 Introduction

Hypothesis testing *example*

•*2* machine types are used for filling soft drink cans: *m1* and *m2*

• You have a hypothesis that *m1* results in larger fill volume of soft drink cans than does *m2*.

•Construct the statistical hypothesis as follows:

- the mean fill volume using machime *m1* is larger than the mean fill volume using machine *m2*.

• Try to draw conclusions about a stated hypothesis.

7-1 Introduction

Suppose that we want to obtain a point estimate of a population parameter. We know that before the data is collected, the observations are considered to be random variables, say $X_1, X_2, \ldots, X_n$. Therefore, any function of the observation, or any **statistic,** is also a random variable. For example, the sample mean $\overline{X}$ and the sample variance S^2 are statistics and they are also random variables.

- Since a statistic is a random variable, it has a probability distribution.
- The probability distribution of a statistic is called a **sampling distribution**

Definition

A point estimate of some population parameter θ is a single numerical value $\hat{\theta}$ of a statistic $\hat{\Theta}$. The statistic $\hat{\Theta}$ is called the point estimator.

7-1 Introduction

μ

point estimate $\overline{x}$

A **point estimate** of some population parameter θ is a single numerical value $\hat{\theta}$ of a statistic $\hat{\Theta}$. The statistic $\hat{\Theta}$ is called the **point estimator**.

$\overline{X}$ point estimator

$x_1=25$, $x_2=30$, $x_3=29$, $x_4=31$

$$\overline{x} = \frac{25+30+29+31}{4} = 28.75$$

Point estimate of μ

7-1 Introduction

σ^2 point estimate s^2

A **point estimate** of some population parameter θ is a single numerical value $\hat{\theta}$ of a statistic $\hat{\Theta}$. The statistic $\hat{\Theta}$ is called the **point estimator**.

S^2 point estimator

$x_1=25$, $x_2=30$, $x_3=29$, $x_4=31$

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \overline{x})^2}{n-1} = 6.9$$

⇨ Point estimate of σ^2

7-1 Introduction

Estimation problems occur frequently in engineering. We often need to estimate

- The mean μ of a single population
- The variance σ^2 (or standard deviation σ) of a single population
- The proportion p of items in a population that belong to a class of interest
- The difference in means of two populations, $\mu_1 - \mu_2$
- The difference in two population proportions, $p_1 - p_2$

7-1 Introduction

Reasonable point estimates of these parameters are as follows:

- For μ, the estimate is $\hat{\mu} = \bar{x}$, the sample mean.
- For σ^2, the estimate is $\hat{\sigma}^2 = s^2$, the sample variance.
- For p, the estimate is $\hat{p} = x/n$, the sample proportion, where x is the number of items in a random sample of size n that belong to the class of interest.
- For $\mu_1 - \mu_2$, the estimate is $\hat{\mu}_1 - \hat{\mu}_2 = \bar{x}_1 - \bar{x}_2$, the difference between the sample means of two independent random samples.
- For $p_1 - p_2$, the estimate is $\hat{p}_1 - \hat{p}_2$, the difference between two sample proportions computed from two independent random samples.

7.2 Sampling Distributions and the Central Limit Theorem

Statistical inference is concerned with making decisions about a population based on the information contained in a random sample from that population.

Definitions:

The random variables $X_1, X_2, \ldots, X_n$ are a **random sample** of size n if (a) the X_i's are independent random variables, and (b) every X_i has the same probability distribution.

A **statistic** is any function of the observations in a random sample.

The probability distribution of a statistic is called a **sampling distribution.**

7.2 Sampling Distributions and the Central Limit Theorem

- The probability distribution of $\bar{X}$ is called the sampling distribution of mean.
- Suppose that a random sample of size n is taken from a **normal population** with mean μ and variance σ^2.
- Each observation $X_1, X_2, \ldots, X_n$ is normally and independently distributed with mean μ and variance σ^2
- Linear functions of independent normally distributed random variables are also normally distributed. **(Reproductive property of Normal Distr.)**
- The sample mean $\bar{X} = \dfrac{X_1 + X_2 + \ldots + X_n}{n}$ has a normal distribution
- with mean $\mu_{\bar{X}} = \dfrac{\mu + \mu + \ldots + \mu}{n} = \mu$
- and variance $\sigma^2_{\bar{X}} = \dfrac{\sigma^2 + \sigma^2 + \ldots + \sigma^2}{n^2} = \dfrac{\sigma^2}{n}$

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

7.2 Sampling Distributions and the Central Limit Theorem

If we are sampling from a population that has an unknown probability distribution, the sampling distribution of the sample mean will still be approximately normal with mean μ and variance σ^2/n, if the sample size n is large. This is one of the most useful theorems in statistics, called the **central limit theorem.** The statement is as follows:

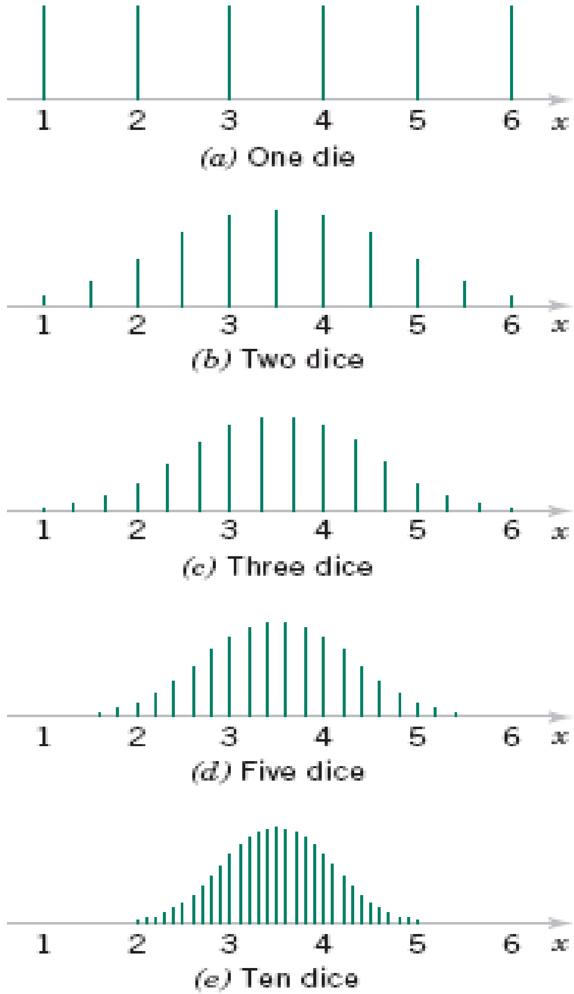
If $X_1, X_2, \ldots, X_n$ is a random sample of size n taken from a population (either finite or infinite) with mean μ and finite variance σ^2, and if $\overline{X}$ is the sample mean, the limiting form of the distribution of

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \tag{7-1}$$

as $n \rightarrow \infty$, is the standard normal distribution.

7.2 Sampling Distributions and the Central Limit Theorem

Distributions of average scores from throwing dice.



Practically

if **n ≥30**, the normal approximation will be satisfactory regardless of the shape of the population.

İf **n<30**, the central limit theorem will work if the distribution of the population is not severely nonnormal.

(continuous, unimodal, symmetric)

7.2 Sampling Distributions and the Central Limit Theorem

Example 7-1

An electronics company manufactures resistors that have a mean resistance of 100 ohms and a standard deviation of 10 ohms. The distribution of resistance is normal. Find the probability that a random sample of $n = 25$ resistors will have an average resistance less than 95 ohms.

Note that the sampling distribution of $\overline{X}$ is normal, with mean $\mu_{\overline{X}} = 100$ ohms and a standard deviation of

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}} = \frac{10}{\sqrt{25}} = 2$$

Therefore, the desired probability corresponds to the shaded area in Fig. 7-1. Standardizing the point $\overline{X} = 95$ in Fig. 7-2, we find that

$$z = \frac{95 - 100}{2} = -2.5$$

and therefore,

$$\begin{aligned} P(\overline{X} < 95) &= P(Z < -2.5) \\ &= 0.0062 \end{aligned}$$

7.2 Sampling Distributions and the Central Limit Theorem

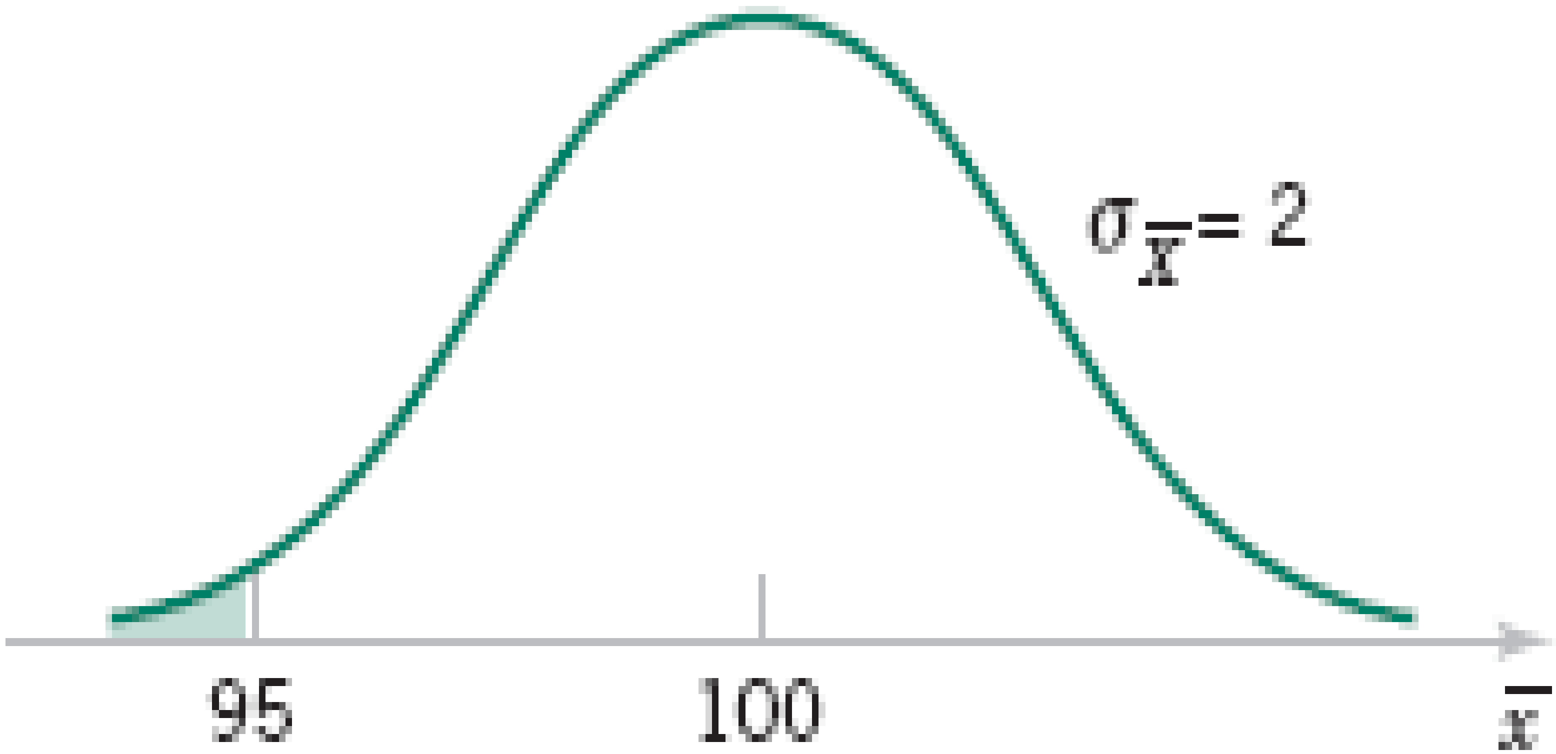


Figure 7-2 Probability for Example 7-1

7.2 Sampling Distributions and the Central Limit Theorem

Example 7-2

X has a continuous distribution:

Find the distributon of the sample mean of a random sample of size n=40?

$$f(x) = \begin{cases} 1/2, & 4 \le x \le 6 \\ 0, & otherwise \end{cases}$$

$$\mu = 5 \qquad \sigma^2 = (6-4)^2/12 = 1/3$$

By C.L.T, $\bar{X}$ is approximately normally distributed with

$$\mu_{\bar{X}} = 5 \qquad \sigma^2_{\bar{X}} = 1/[3(40)] = 1/120$$

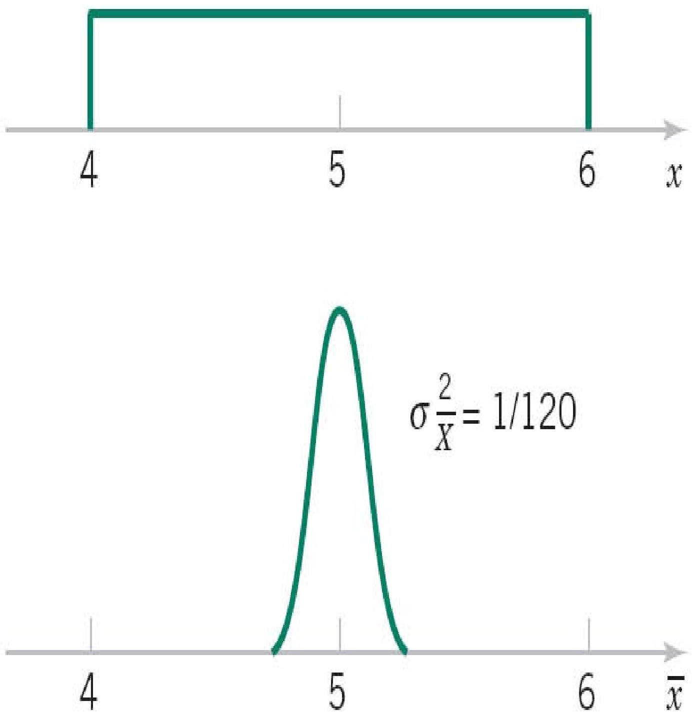


Figure 7-3 The distributions of X and $\bar{X}$ for Example 7-2.

7.2 Sampling Distributions and the Central Limit Theorem

- Two independent populations
  - 1st population has mean μ_1 and variance σ_1^2
  - 2nd population has mean μ_2 and variance σ_2^2

- The statistic $\bar{X}_1 - \bar{X}_2$ has the following mean and variance :

- mean $\mu_{\bar{X}_1 - \bar{X}_2} = \mu_{\bar{X}_1} - \mu_{\bar{X}_2} = \mu_1 - \mu_2$

- and variance $\sigma^2_{\bar{X}_1 - \bar{X}_2} = \sigma^2_{\bar{X}_1} + \sigma^2_{\bar{X}_2} = \dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}$

7.2 Sampling Distributions and the Central Limit Theorem

Approximate Sampling Distribution of a Difference in Sample Means

If we have two independent populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2 and if $\overline{X}_1$ and $\overline{X}_2$ are the sample means of two independent random samples of sizes n_1 and n_2 from these populations, then the sampling distribution of

$$Z = \frac{\overline{X}_1 - \overline{X}_2 - (\mu_1 - \mu_2)}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}} \qquad (7\text{-}4)$$

is approximately standard normal, if the conditions of the central limit theorem apply. If the two populations are normal, the sampling distribution of Z is exactly standard normal.

7.2 Sampling Distributions and the Central Limit Theorem *Ex.7-3*

- Two independent and approximately normal populations
  - X_1: life of a component with old process $\mu_1 = 5000$ $\sigma_1 = 40$
  - X_2: life of a component with improved process $\mu_2 = 5050$ $\sigma_2 = 30$
- $n_1 = 16$, $n_2 = 25$
- What is the probability that the difference in the two sample means $\bar{X}_2 - \bar{X}_1$

is at least 25 hours?

$$\mu_{\bar{X}_1} = \mu_1 = 5000 \qquad \sigma^2_{\bar{X}_1} = \frac{\sigma_1^2}{n_1} = \frac{40^2}{16} = 100$$

$$\mu_{\bar{X}_2} = \mu_2 = 5050 \qquad \sigma^2_{\bar{X}_2} = \frac{\sigma_2^2}{n_2} = \frac{30^2}{25} = 36$$

$$\mu_{\bar{X}_2 - \bar{X}_1} = \mu_{\bar{X}_2} - \mu_{\bar{X}_1} = 50$$

$$\sigma^2_{\bar{X}_2 - \bar{X}_1} = \sigma^2_{\bar{X}_2} + \sigma^2_{\bar{X}_1} = 136$$

$$P(\bar{X}_2 - \bar{X}_1 \geq 25) = P\left(Z \geq \frac{25 - \mu_{\bar{X}_2 - \bar{X}_1}}{\sigma_{\bar{X}_2 - \bar{X}_1}}\right) = P\left(Z \geq \frac{25 - 50}{\sqrt{136}}\right) = P\left(Z \geq -2.14\right) = 0.9838$$

7-3 General Concepts of Point Estimation

- We may have several different choices for the point estimator of a parameter. Ex: to estimate the mean of a population
  - Sample mean
  - Sample median
  - The average of the smallest and largest observations in the sample
- Which point estimator is the best one?
- Need to examine their statistical properties and develop some criteria for comparing estimators
- For instance, an estimator should be close to the true value of the unknown parameter

7-3 General Concepts of Point Estimation

7-3.1 Unbiased Estimators

Definition

The point estimator $\hat{\Theta}$ is an **unbiased estimator** for the parameter θ if

$$E(\hat{\Theta}) = \theta \qquad (7\text{-}5)$$

If the estimator is not unbiased, then the difference

$$\textbf{bias} = \boxed{E(\hat{\Theta}) - \theta} \qquad (7\text{-}6)$$

is called the **bias** of the estimator $\hat{\Theta}$.

➔ When an estimator is unbiased, the bias is zero.

7-3 General Concepts of Point Estimation

Are $\bar{X}$ and S^2 unbiased estimators of μ and σ^2?

$$E(\bar{X}) = E\left(\frac{1}{n}X_1 + \frac{1}{n}X_2 + \ldots + \frac{1}{n}X_n\right) = n\frac{1}{n}\mu \stackrel{\checkmark}{=} \mu$$

Unbiased !

$$E(S^2) = E\left(\frac{\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}{n-1}\right) = \frac{1}{n-1}E\left(\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2\right)$$

$$= \frac{1}{n-1}E\left(\sum_{i=1}^{n}\left(X_i^2 - 2\bar{X}X_i + \bar{X}^2\right)\right) = \frac{1}{n-1}E\left(\sum_{i=1}^{n}X_i^2 - \sum_{i=1}^{n}2\bar{X}X_i + \sum_{i=1}^{n}\bar{X}^2\right)$$

$$= \frac{1}{n-1}E\left(\sum_{i=1}^{n}X_i^2 - 2n\bar{X}^2 + n\bar{X}^2\right) = \frac{1}{n-1}E\left(\sum_{i=1}^{n}X_i^2 - n\bar{X}^2\right)$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n}E\left(X_i^2\right) - nE\left(\bar{X}^2\right)\right)$$

7-3 General Concepts of Point Estimation

Example 7-1 (continued)

$$V(X_i^2) = \sigma^2 = E(X_i^2) - \mu^2 \Rightarrow \boxed{E(X_i^2) = \sigma^2 + \mu^2}$$

$$V(\bar{X}) = \frac{\sigma^2}{n} = E(\bar{X}^2) - \left(E(\bar{X})\right)^2 = E(\bar{X}^2) - \mu^2 \Rightarrow \boxed{E(\bar{X}^2) = \frac{\sigma^2}{n} + \mu^2}$$

$$E(S^2) = \frac{1}{n-1}\left[\sum_{i=1}^{n}(\mu^2 + \sigma^2) - n(\mu^2 + \sigma^2/n)\right]$$

$$= \frac{1}{n-1}(n\mu^2 + n\sigma^2 - n\mu^2 - \sigma^2)$$

$$= \sigma^2$$

Unbiased !

7-3 General Concepts of Point Estimation

- There is not a unique unbiased estimator.
- n=10 data 12.8 9.4 8.7 11.6 13.1 9.8 14.1 8.5 12.1 10.3
- There are several unbiased estimators of μ
  - Sample mean (11.04)
  - Sample median (10.95)
  - The average of the smallest and largest observations in the sample (11.3)
  - A single observation from the population (12.8)
- Cannot rely on the property of unbiasedness alone to select the estimator.
- Need a method to select among unbiased estimators.

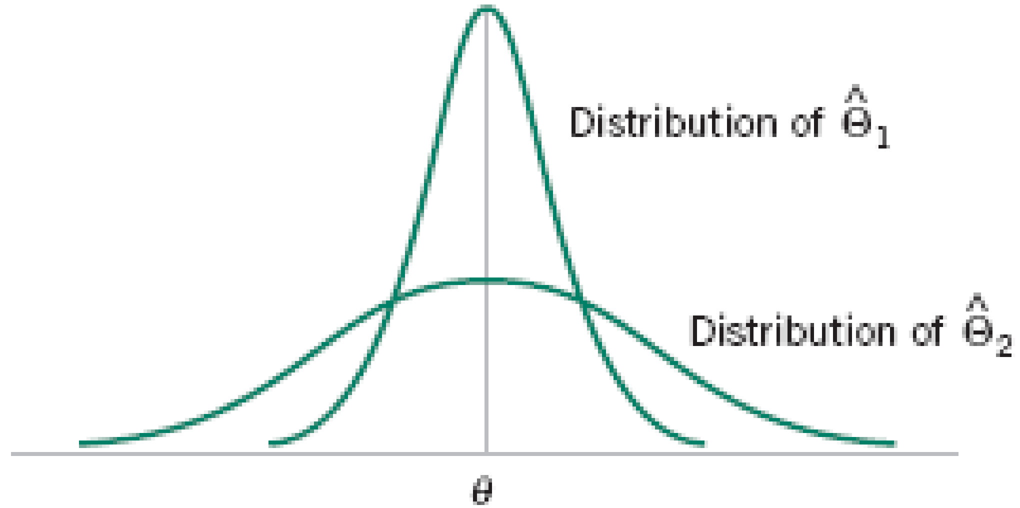
7-3 General Concepts of Point Estimation

7-3.2 Variance of a Point Estimator

Definition

If we consider all unbiased estimators of θ, the one with the smallest variance is called the **minimum variance unbiased estimator** (MVUE).

The sampling distributions of two unbiased estimators $\hat{\Theta}_1$ and $\hat{\Theta}_2$.



7-3 General Concepts of Point Estimation

7-3.2 Variance of a Point Estimator

If $X_1, X_2, \ldots, X_n$ is a random sample of size n from a normal distribution with mean μ and variance σ^2, the sample mean $\overline{X}$ is the MVUE for μ.

$$V(X_i) = \sigma^2$$

$$V(\overline{X}) = \frac{\sigma^2}{n}$$

$$V(\overline{X}) < V(X_i) \quad for\ n \geq 2$$

The sample mean is better estimator of μ than a single observation X_i

7-3 General Concepts of Point Estimation

7-3.3 Standard Error: Reporting a Point Estimate

Definition

The **standard error** of an estimator $\hat{\Theta}$ is its standard deviation, given by $\sigma_{\hat{\Theta}} = \sqrt{V(\hat{\Theta})}$. If the standard error involves unknown parameters that can be estimated, substitution of those values into $\sigma_{\hat{\Theta}}$ produces an **estimated standard error**, denoted by $\hat{\sigma}_{\hat{\Theta}}$.

7-3 General Concepts of Point Estimation

7-3.3 Standard Error: Reporting a Point Estimate

Suppose we are sampling from a normal distribution with mean μ and variance σ^2. Now the distribution of $\overline{X}$ is normal with mean μ and variance σ^2/n, so the standard error of $\overline{X}$ is

$$\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}$$

If we did not know σ but substituted the sample standard deviation S into the above equation, the estimated standard error of $\overline{X}$ would be

$$\hat{\sigma}_{\overline{X}} = \frac{S}{\sqrt{n}}$$

7-3 General Concepts of Point Estimation

Example 7-5

An article in the *Journal of Heat Transfer* (Trans. ASME, Sec. C, 96, 1974, p. 59) described a new method of measuring the thermal conductivity of Armco iron. Using a temperature of 100°F and a power input of 550 watts, the following 10 measurements of thermal conductivity (in Btu/hr-ft-°F) were obtained:

41.60, 41.48, 42.34, 41.95, 41.86,
42.18, 41.72, 42.26, 41.81, 42.04

A point estimate of the mean thermal conductivity at 100°F and 550 watts is the sample mean or

$$\bar{x} = 41.924 \text{ Btu/hr-ft-°F}$$

7-3 General Concepts of Point Estimation

Example 7-5 (continued)

The standard error of the sample mean is $\sigma_{\overline{X}} = \sigma/\sqrt{n}$, and since σ is unknown, we may replace it by the sample standard deviation $s = 0.284$ to obtain the estimated standard error of $\overline{X}$ as

$$\hat{\sigma}_{\overline{X}} = \frac{s}{\sqrt{n}} = \frac{0.284}{\sqrt{10}} = 0.0898$$

Notice that the standard error is about 0.2 percent of the sample mean, implying that we have obtained a relatively precise point estimate of thermal conductivity. If we can assume that thermal conductivity is normally distributed, 2 times the standard error is $2\hat{\sigma}_{\overline{X}} = 2(0.0898) = 0.1796$, and we are highly confident that the true mean thermal conductivity is with the interval 41.924 ± 0.1756, or between 41.744 and 42.104.

Probability that true mean is within $\overline{X} \pm 2\hat{\sigma}_{\overline{X}}$ **is 0.9545**

7-3 General Concepts of Point Estimation

7-3.4 Mean Square Error of an Estimator

There may be cases where we may need to use a biased estimator.

So we need another comparison measure:

Definition

The **mean squared error** of an estimator $\hat{\Theta}$ of the parameter θ is defined as

$$\text{MSE}(\hat{\Theta}) = E(\hat{\Theta} - \theta)^2 \qquad (7\text{-}7)$$

7-3 General Concepts of Point Estimation

7-3.4 Mean Square Error of an Estimator

$$MSE(\hat{\Theta}) = E(\hat{\Theta} - \theta)^2$$

$$= E\left[\hat{\Theta} - E(\hat{\Theta})\right]^2 + \left[\theta - E(\hat{\Theta})\right]^2$$

$$= V(\hat{\Theta}) \qquad\qquad + (bias)^2$$

The MSE of $\hat{\Theta}$ is equal to the variance of the estimator plus the squared bias.

If $\hat{\Theta}$ is an unbiased estimator of θ, MSE of $\hat{\Theta}$ is equal to the variance of $\hat{\Theta}$

7-3 General Concepts of Point Estimation

7-3.4 Mean Square Error of an Estimator

$$V(\hat{\Theta}) + (bias)^2$$

$$E\left[\hat{\Theta} - E(\hat{\Theta})\right]^2 + \left[\theta - E(\hat{\Theta})\right]^2$$

$$= E\left[\hat{\Theta}^2 - 2\hat{\Theta}E(\hat{\Theta}) + E^2(\hat{\Theta})\right] + \left[\theta^2 - 2\theta E(\hat{\Theta}) + E^2(\hat{\Theta})\right]$$

$$= E(\hat{\Theta}^2) - 2E(\hat{\Theta})E(\hat{\Theta}) + E^2(\hat{\Theta}) + \theta^2 - 2\theta E(\hat{\Theta}) + E^2(\hat{\Theta})$$

$$= E(\hat{\Theta}^2) - E^2(\hat{\Theta}) + \theta^2 - 2\theta E(\hat{\Theta}) + E^2(\hat{\Theta})$$

$$= E(\hat{\Theta}^2) + \theta^2 - 2\theta E(\hat{\Theta})$$

$$\overset{\checkmark}{=} E(\hat{\Theta} - \theta)^2$$

$$MSE(\hat{\Theta})$$

7-3 General Concepts of Point Estimation

7-3.4 Mean Square Error of an Estimator

The mean squared error is an important criterion for comparing two estimators. Let $\hat{\Theta}_1$ and $\hat{\Theta}_2$ be two estimators of the parameter θ, and let MSE ($\hat{\Theta}_1$) and MSE ($\hat{\Theta}_2$) be the mean squared errors of $\hat{\Theta}_1$ and $\hat{\Theta}_2$. Then the **relative efficiency of $\hat{\Theta}_2$ to $\hat{\Theta}_1$** is defined as

$$\frac{MSE(\hat{\Theta}_1)}{MSE(\hat{\Theta}_2)} \tag{7-8}$$

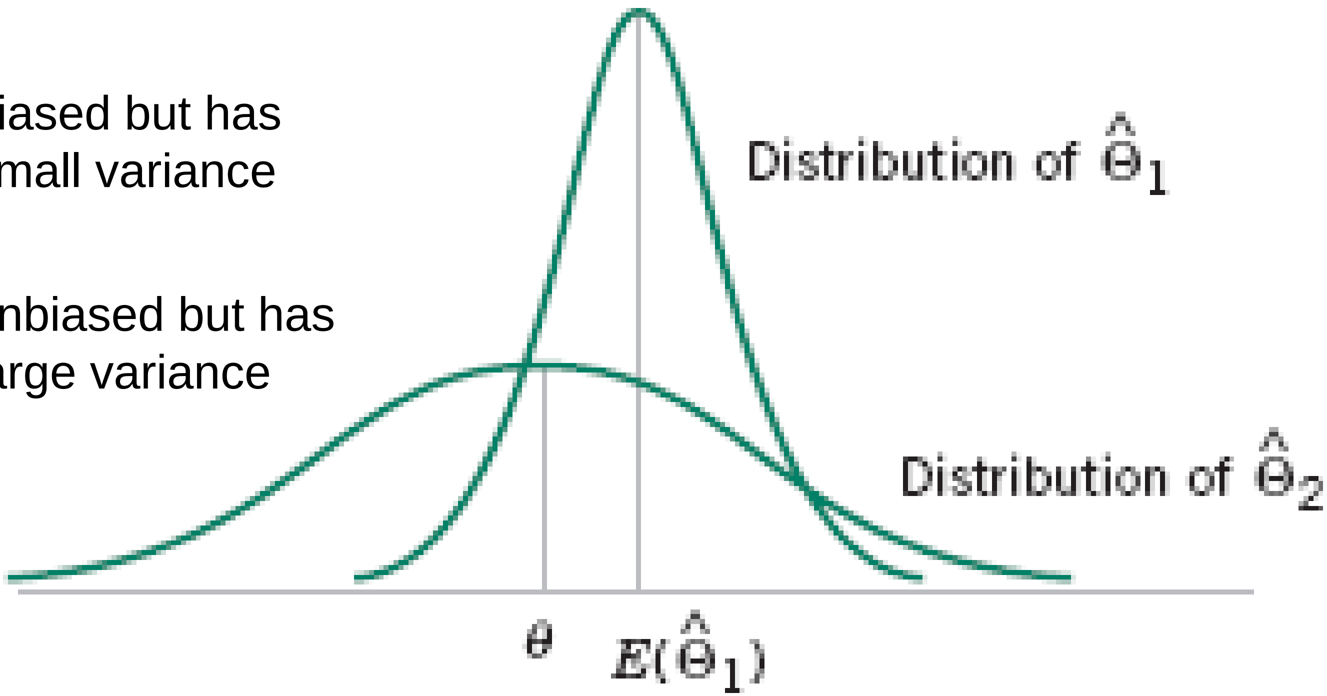
If this relative efficiency is less than 1, we would conclude that $\hat{\Theta}_1$ is a more efficient estimator of θ than $\hat{\Theta}_2$, in the sense that it has a smaller mean square error.

7-3 General Concepts of Point Estimation

7-3.4 Mean Square Error of an Estimator

$\hat{\Theta}_1$ biased but has small variance

$\hat{\Theta}_2$ unbiased but has large variance



An estimate based on $\hat{\Theta}_1$ would more likely be close to the true value of θ than would an estimate based on $\hat{\Theta}_2$. .

7-3 General Concepts of Point Estimation

7-3.4 Mean Square Error of an Estimator

Exercise: Calculate the MSE of the following estimators.

$$\hat{\Theta}_1 = \bar{X}$$

$$\hat{\Theta}_2 = \tilde{X}$$

7-4 Methods of Point Estimation

How can good estimators be obtained?

7-4.1 Method of Moments

Let $X_1, X_2, \ldots, X_n$ be a random sample from the probability distribution $f(x)$, where $f(x)$ can be a discrete probability mass function or a continuous probability density function. The kth **population moment** (or **distribution moment**) is $E(X^k)$, $k = 1, 2, \ldots$. The corresponding kth **sample moment** is $(1/n)\sum_{i=1}^{n} X_i^k$, $k = 1, 2, \ldots$.

Let $X_1, X_2, \ldots, X_n$ be a random sample from either a probability mass function or probability density function with m unknown parameters $\theta_1, \theta_2, \ldots, \theta_m$. The **moment estimators** $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_m$ are found by equating the first m population moments to the first m sample moments and solving the resulting equations for the unknown parameters.

7-4 Methods of Point Estimation

Example 7-7

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a normal distribution with parameters μ and σ^2. For the normal distribution $E(X) = \mu$ and $E(X^2) = \mu^2 + \sigma^2$. Equating $E(X)$ to $\overline{X}$ and $E(X^2)$ to $\frac{1}{n}\sum_{i=1}^{n} X_i^2$ gives

$$E(X) = \mu = \overline{X}, \qquad \mu^2 + \sigma^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2$$

Solving these equations gives the moment estimators

$$\hat{\mu} = \overline{X}, \qquad \hat{\sigma}^2 = \frac{\sum_{i=1}^{n} X_i^2 - n\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right)^2}{n} = \frac{\sum_{i=1}^{n} (X_i - \overline{X})^2}{n}$$

Notice that the moment estimator of σ^2 is not an unbiased estimator.

7-4 Methods of Point Estimation

7-4.2 Method of Maximum Likelihood

Suppose that X is a random variable with probability distribution $f(x; \theta)$, where θ is a single unknown parameter. Let $x_1, x_2, \ldots, x_n$ be the observed values in a random sample of size n. Then the **likelihood function** of the sample is

$$L(\theta) = f(x_1; \theta) \cdot f(x_2; \theta) \cdot \cdots \cdot f(x_n; \theta) \quad (7\text{-}9)$$

Note that the likelihood function is now a function of only the unknown parameter θ. The **maximum likelihood estimator (MLE)** of θ is the value of θ that maximizes the likelihood function $L(\theta)$.

7-4 Methods of Point Estimation

Example 7-9

Let X be a Bernoulli random variable. The probability mass function is

$$f(x;p) = \begin{cases} p^x(1-p)^{1-x}, & x = 0, 1 \\ 0, & \text{otherwise} \end{cases}$$

where p is the parameter to be estimated. The likelihood function of a random sample of size n is

$$\begin{aligned} L(p) &= p^{x_1}(1-p)^{1-x_1} p^{x_2}(1-p)^{1-x_2} \cdots p^{x_n}(1-p)^{1-x_n} \\ &= \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^{\sum_{i=1}^{n} x_i}(1-p)^{n-\sum_{i=1}^{n} x_i} \end{aligned}$$

7-4 Methods of Point Estimation

Example 7-9 (continued)

We observe that if $\hat{p}$ maximizes $L(p)$, $\hat{p}$ also maximizes $\ln L(p)$. Therefore,

$$\ln L(p) = \left(\sum_{i=1}^{n} x_i\right)\ln p + \left(n - \sum_{i=1}^{n} x_i\right)\ln(1-p)$$

$$y = \ln f(x)$$

$$\frac{dy}{dx} = \frac{f'(x)}{f(x)}$$

Now

$$\frac{d \ln L(p)}{dp} = \frac{\sum_{i=1}^{n} x_i}{p} - \frac{\left(n - \sum_{i=1}^{n} x_i\right)}{1-p}$$

Equating this to zero and solving for p yields $\hat{p} = (1/n)\sum_{i=1}^{n} x_i$. Therefore, the maximum likelihood estimator of p is

$$\hat{P} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

7-4 Methods of Point Estimation

Example 7-12

Let X be normally distributed with mean μ and variance σ^2, where both μ and σ^2 are unknown. The likelihood function for a random sample of size n is

$$L(\mu, \sigma^2) = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{-(x_i-\mu)^2/(2\sigma^2)} = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2\sigma^2)\sum_{i=1}^{n}(x_i-\mu)^2}$$

and

$$\ln L(\mu, \sigma^2) = -\frac{n}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2$$

$$y = (ax+b)^m$$

$$\frac{dy}{dx} = ma(ax+b)^{m-1}$$

7-4 Methods of Point Estimation

Example 7-12 (continued)

Now

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = \frac{1}{\sigma^2} \sum_{i=1}^{n} (x_i - \mu) = 0$$

$$\frac{\partial \ln L(\mu, \sigma^2)}{\partial (\sigma^2)} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^2 = 0$$

The solutions to the above equation yield the maximum likelihood estimators

$$\hat{\mu} = \overline{X} \qquad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} (X_i - \overline{X})^2$$

Once again, the maximum likelihood estimators are equal to the moment estimators.

7-4 Methods of Point Estimation

Properties of the Maximum Likelihood Estimator

Under very general and not restrictive conditions, when the sample size n is large and if $\hat{\Theta}$ is the maximum likelihood estimator of the parameter θ,

(1) $\hat{\Theta}$ is an approximately unbiased estimator for θ $[E(\hat{\Theta}) \simeq \theta]$,

(2) the variance of $\hat{\Theta}$ is nearly as small as the variance that could be obtained with any other estimator, and

(3) $\hat{\Theta}$ has an approximate normal distribution.

7-4 Methods of Point Estimation

Properties of the Maximum Likelihood Estimator

$$MLE \ of \ \sigma^2 \ is$$

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

$$E(\hat{\sigma}^2) = \frac{n-1}{n}\sigma^2$$

$$bias = E(\hat{\sigma}^2) - \sigma^2 = \frac{-\sigma^2}{n}$$

bias is negative. MLE for σ^2 tends to underestimate σ^2

⇒ The bias approaches zero as n increases.

MLE for σ^2 is an asymptotically unbiased estimator for σ^2

7-4 Methods of Point Estimation

The Invariance Property

Let $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k$ be the maximum likelihood estimators of the parameters $\theta_1, \theta_2, \ldots, \theta_k$. Then the maximum likelihood estimator of any function $h(\theta_1, \theta_2, \ldots, \theta_k)$ of these parameters is the same function $h(\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k)$ of the estimators $\hat{\Theta}_1, \hat{\Theta}_2, \ldots, \hat{\Theta}_k$.

7-4 Methods of Point Estimation

Example 7-13

In the normal distribution case, the maximum likelihood estimators of μ and σ^2 were $\hat{\mu} = \overline{X}$ and $\hat{\sigma}^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/n$. To obtain the maximum likelihood estimator of the function $h(\mu, \sigma^2) = \sqrt{\sigma^2} = \sigma$, substitute the estimators $\hat{\mu}$ and $\hat{\sigma}^2$ into the function h, which yields

$$\hat{\sigma} = \sqrt{\hat{\sigma}^2} = \left[\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2\right]^{1/2}$$

Thus, the maximum likelihood estimator of the standard deviation σ is *not* the sample standard deviation S.

7-4 Methods of Point Estimation

Complications in Using Maximum Likelihood Estimation

- It is not always easy to maximize the likelihood function because the equation(s) obtained from $dL(\theta)/d\theta = 0$ may be difficult to solve.
- It may not always be possible to use calculus methods directly to determine the maximum of $L(\theta)$.