# 7 Sampling Distributions and Point Estimation of Parameters

## CHAPTER OUTLINE

## LEARNING OBJECTIVES

After careful study of this chapter you should be able to do the following:

1. Explain the general concepts of estimating the parameters of a population or a probability distribution
2. Explain the important role of the normal distribution as a sampling distribution
3. Understand the central limit theorem
4. Explain important properties of point estimators, including bias, variance, and mean square error
5. Know how to construct point estimators using the method of moments and the method of maximum likelihood
6. Know how to compute and explain the precision with which a parameter is estimated
7. Know how to construct a point estimator using the Bayesian approach

# 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

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. We call the probability distribution of a statistic a **sampling distribution.** The notion of a sampling distribution is very important and will be discussed and illustrated later in the chapter.

## Definition

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

# 7-1 Introduction

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

- The mean $\mu$ of a single population
- The variance $\sigma^2$ (or standard deviation $\sigma$) 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 $\mu$, the estimate is $\hat{\mu} = \bar{x}$, the sample mean.
- For $\sigma^2$, 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

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 $\mu$ and variance $\sigma^2/n$, 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:

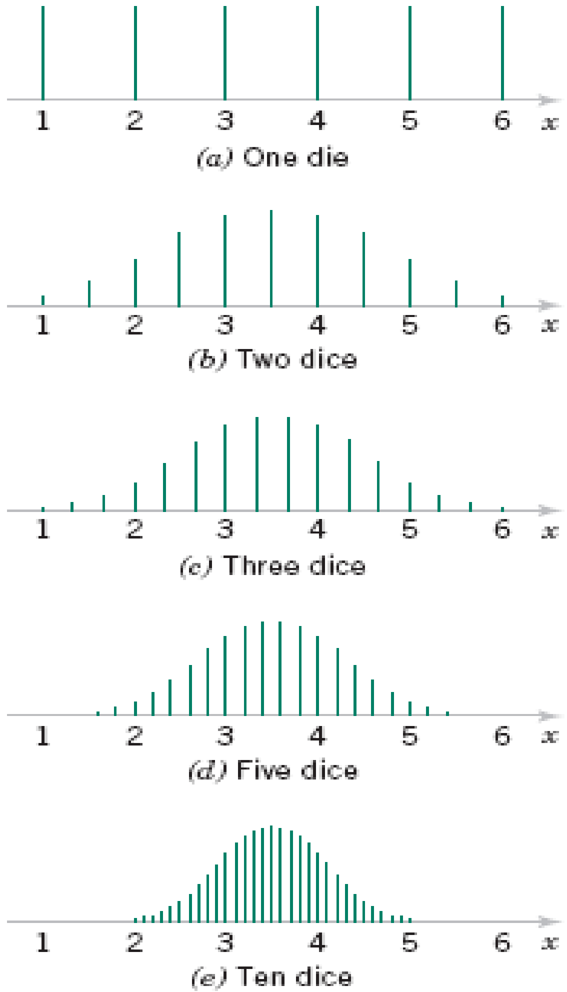
If $X_1, X_2, \ldots, X_n$ is a random sample of size $n$ taken from a population (either finite or infinite) with mean $\mu$ and finite variance $\sigma^2$, and if $\overline{X}$ is the sample mean, the limiting form of the distribution of

$$Z = \frac{\overline{X} - \mu}{\sigma/\sqrt{n}} \quad (7\text{-}1)$$

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

# 7.2 Sampling Distributions and the Central Limit Theorem

**Figure 7-1** Distributions of average scores from throwing dice. [Adapted with permission from Box, Hunter, and Hunter (1978).]



# 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,

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

# 7.2 Sampling Distributions and the Central Limit Theorem

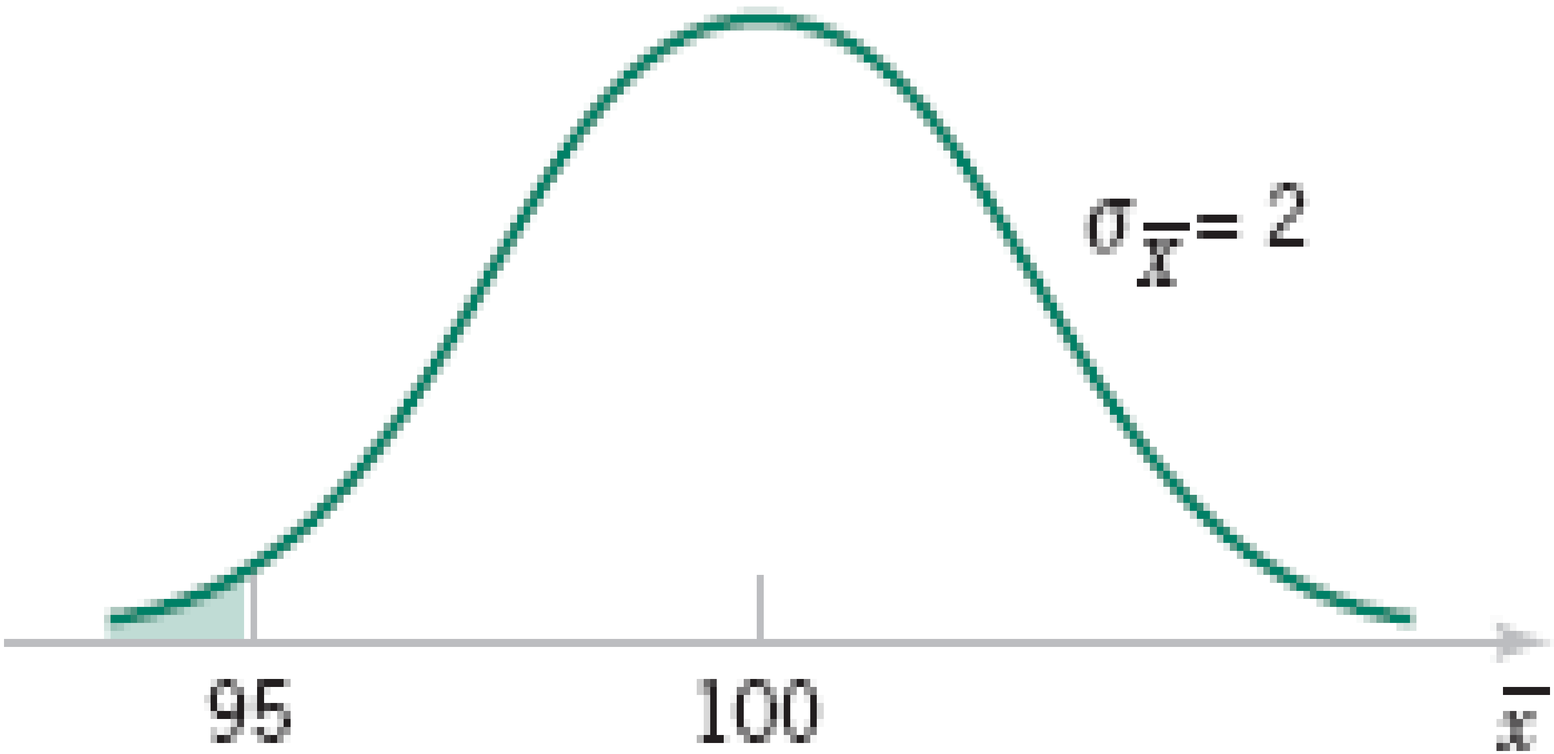



**Figure 7-2** Probability for Example 7-1

# 7.2 Sampling Distributions and the Central Limit Theorem

## Approximate Sampling Distribution of a Difference in Sample Means

If we have two independent populations with means $\mu_1$ and $\mu_2$ and variances $\sigma_1^2$ and $\sigma_2^2$ 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.

# 7-3 General Concepts of Point Estimation

## 7-3.1 Unbiased Estimators

### Definition

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

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

If the estimator is not unbiased, then the difference

$$E(\hat{\Theta}) - \theta \tag{7-6}$$

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

# 7-3 General Concepts of Point Estimation

## Example 7-1

Suppose that $X$ is a random variable with mean $\mu$ and variance $\sigma^2$. Let $X_1, X_2, \ldots, X_n$ be a random sample of size $n$ from the population represented by $X$. Show that the sample mean $\overline{X}$ and sample variance $S^2$ are unbiased estimators of $\mu$ and $\sigma^2$, respectively.

First consider the sample mean. In Equation 5.40a in Chapter 5, we showed that $E(\overline{X}) = \mu$. Therefore, the sample mean $\overline{X}$ is an unbiased estimator of the population mean $\mu$.

Now consider the sample variance. We have

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

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

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

# 7-3 General Concepts of Point Estimation

## Example 7-1 (continued)

The last equality follows from Equation 5-37 in Chapter 5. However, since $E(X_i^2) = \mu^2 + \sigma^2$ and $E(\overline{X}^2) = \mu^2 + \sigma^2/n$, we have

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

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

$$= \sigma^2$$

Therefore, the sample variance $S^2$ is an unbiased estimator of the population variance $\sigma^2$.

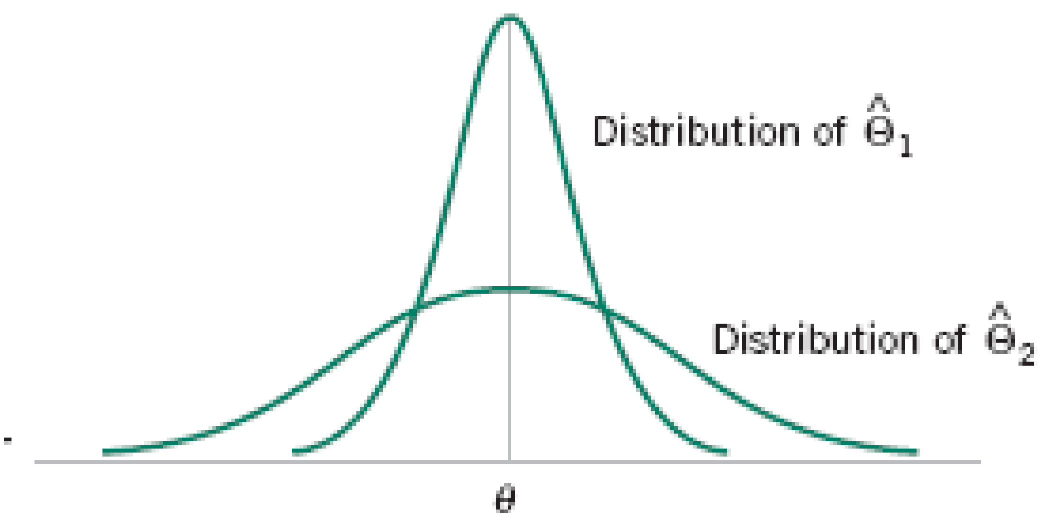
# 7-3 General Concepts of Point Estimation

## 7-3.2 Variance of a Point Estimator

**Definition**

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

**Figure 7-5** 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 $\mu$ and variance $\sigma^2$, the sample mean $\overline{X}$ is the MVUE for $\mu$.

# 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 $\mu$ and variance $\sigma^2$. Now the distribution of $\overline{X}$ is normal with mean $\mu$ and variance $\sigma^2/n$, so the standard error of $\overline{X}$ is

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

If we did not know $\sigma$ 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 $\sigma$ 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 $41.924 \pm 0.1756$, or between 41.744 and 42.104.

# 7-3 General Concepts of Point Estimation

## 7-3.4 Mean Square Error of an Estimator

**Definition**

The **mean squared error** of an estimator $\hat{\Theta}$ of the parameter $\theta$ 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

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 $\theta$, 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

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

If this relative efficiency is less than 1, we would conclude that $\hat{\Theta}_1$ is a more efficient estimator of $\theta$ 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

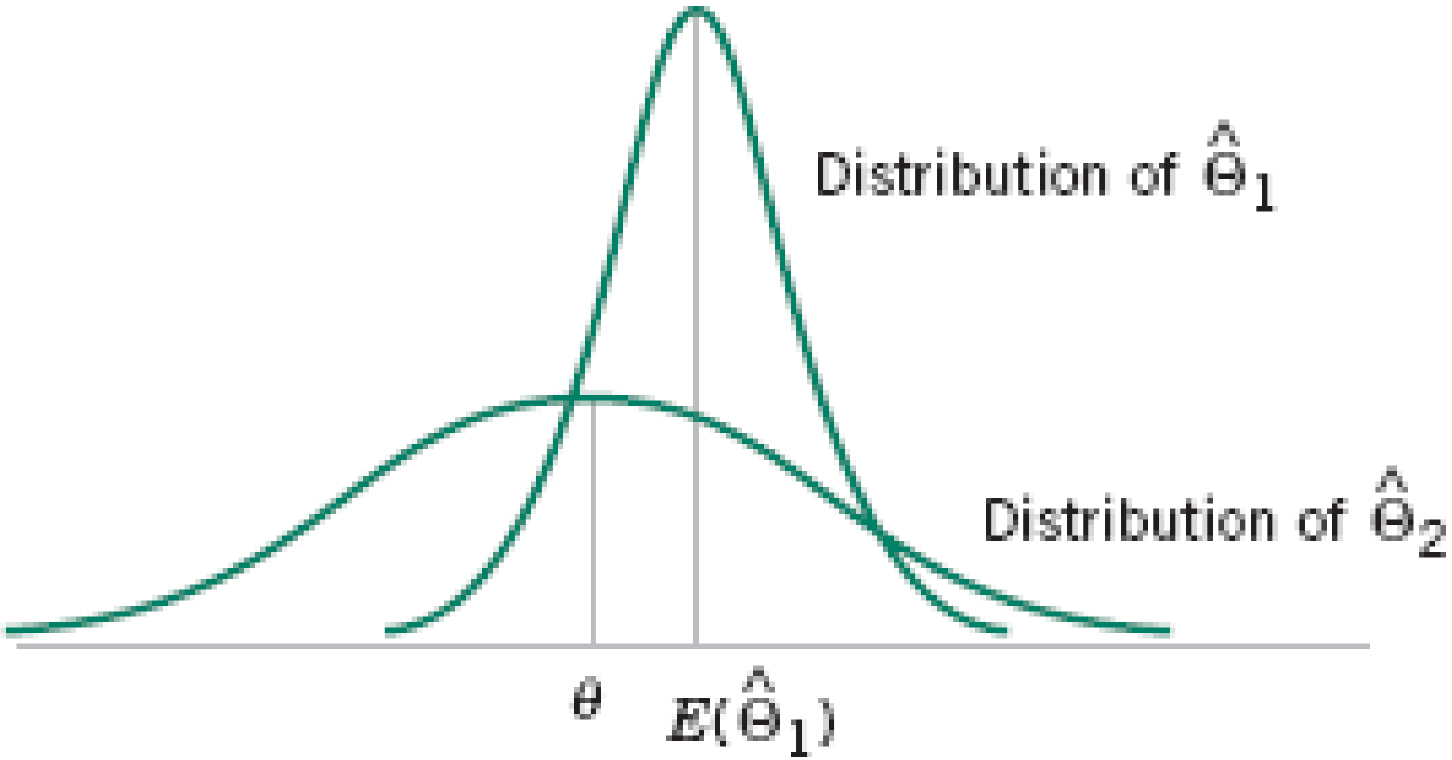



**Figure 7-6** A biased estimator $\hat{\Theta}_1$ that has smaller variance than the unbiased estimator $\hat{\Theta}_2$.

# 7-4 Methods of Point Estimation

## Definition

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 $k$th **population moment** (or **distribution moment**) is $E(X^k)$, $k = 1, 2, \ldots$. The corresponding $k$th **sample moment** is $(1/n)\sum_{i=1}^{n} X_i^k$, $k = 1, 2, \ldots$.

## Definition

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 $\mu$ and $\sigma^2$. 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

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

Solving these equations gives the moment estimators

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

Notice that the moment estimator of $\sigma^2$ is not an unbiased estimator.

# 7-4 Methods of Point Estimation

## 7-4.2 Method of Maximum Likelihood

### Definition

Suppose that $X$ is a random variable with probability distribution $f(x; \theta)$, where $\theta$ 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

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

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

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

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

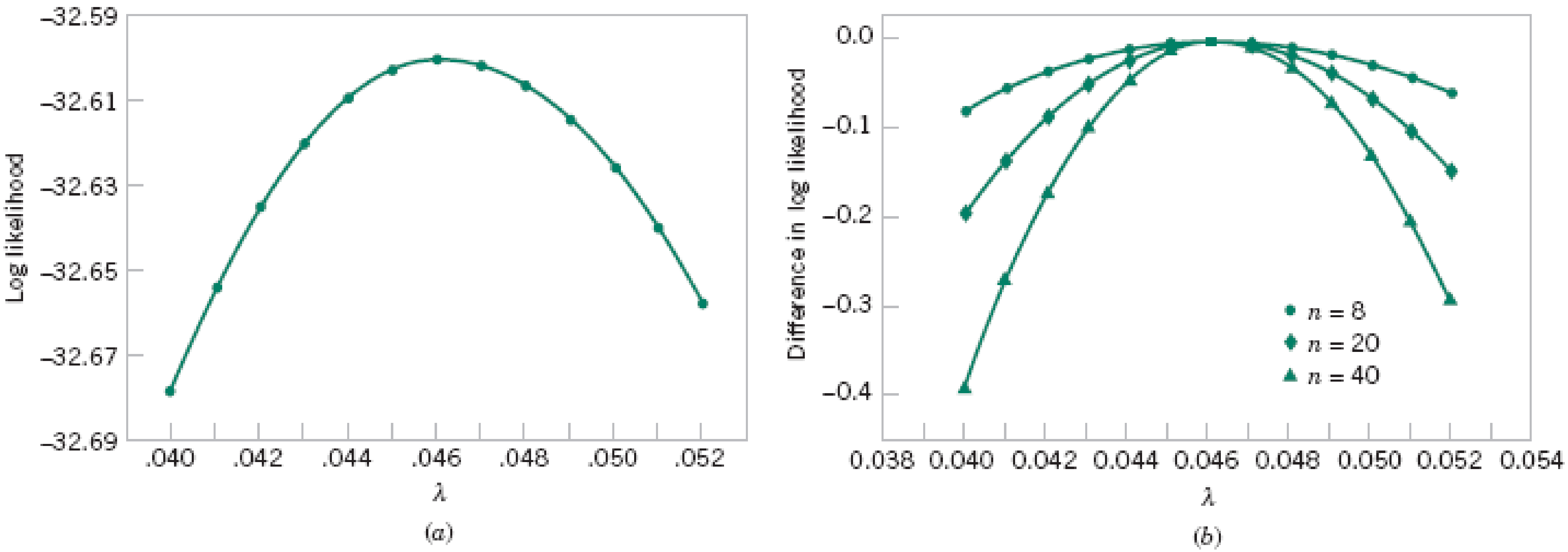



**Figure 7-7** Log likelihood for the exponential distribution, using the failure time data. (a) Log likelihood with $n = 8$ (original data). (b) Log likelihood if $n = 8$, 20, and 40.

# 7-4 Methods of Point Estimation

## Example 7-12

Let $X$ be normally distributed with mean $\mu$ and variance $\sigma^2$, where both $\mu$ and $\sigma^2$ 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

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

# 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 $\theta$,

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

(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

## 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 $\mu$ and $\sigma^2$ 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 $\sigma$ 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 $L(\theta)$.

# 7-4 Methods of Point Estimation

## Example 7-14

Let $X$ be uniformly distributed on the interval 0 to $a$. Since the density function is $f(x) = 1/a$ for $0 \le x \le a$ and zero otherwise, the likelihood function of a random sample of size $n$ is

$$L(a) = \prod_{i=1}^{n} \frac{1}{a} = \frac{1}{a^n}$$

if $0 \le x_1 \le a$, $0 \le x_2 \le a, \ldots, 0 \le x_n \le a$. Note that the slope of this function is not zero anywhere. That is, as long as $\max(x_i) \le a$, the likelihood is $1/a^n$, which is positive, but when $a < \max(x_i)$, the likelihood goes to zero, as illustrated in Fig. 7-4. Therefore, calculus methods cannot be used directly because the maximum value of the likelihood function occurs at a point of discontinuity. However, since $d/da(a^{-n}) = -n/a^{n+1}$ is less than zero for all values of $a > 0$, $a^{-n}$ is a decreasing function of $a$. This implies that the maximum of the likelihood function $L(a)$ occurs at the lower boundary point. The figure clearly shows that we could maximize $L(a)$ by setting $\hat{a}$ equal to the smallest value that it could logically take on, which is $\max(x_i)$. Clearly, $a$ cannot be smaller than the largest sample observation, so setting $\hat{a}$ equal to the largest sample value is reasonable.

# 7-4 Methods of Point Estimation

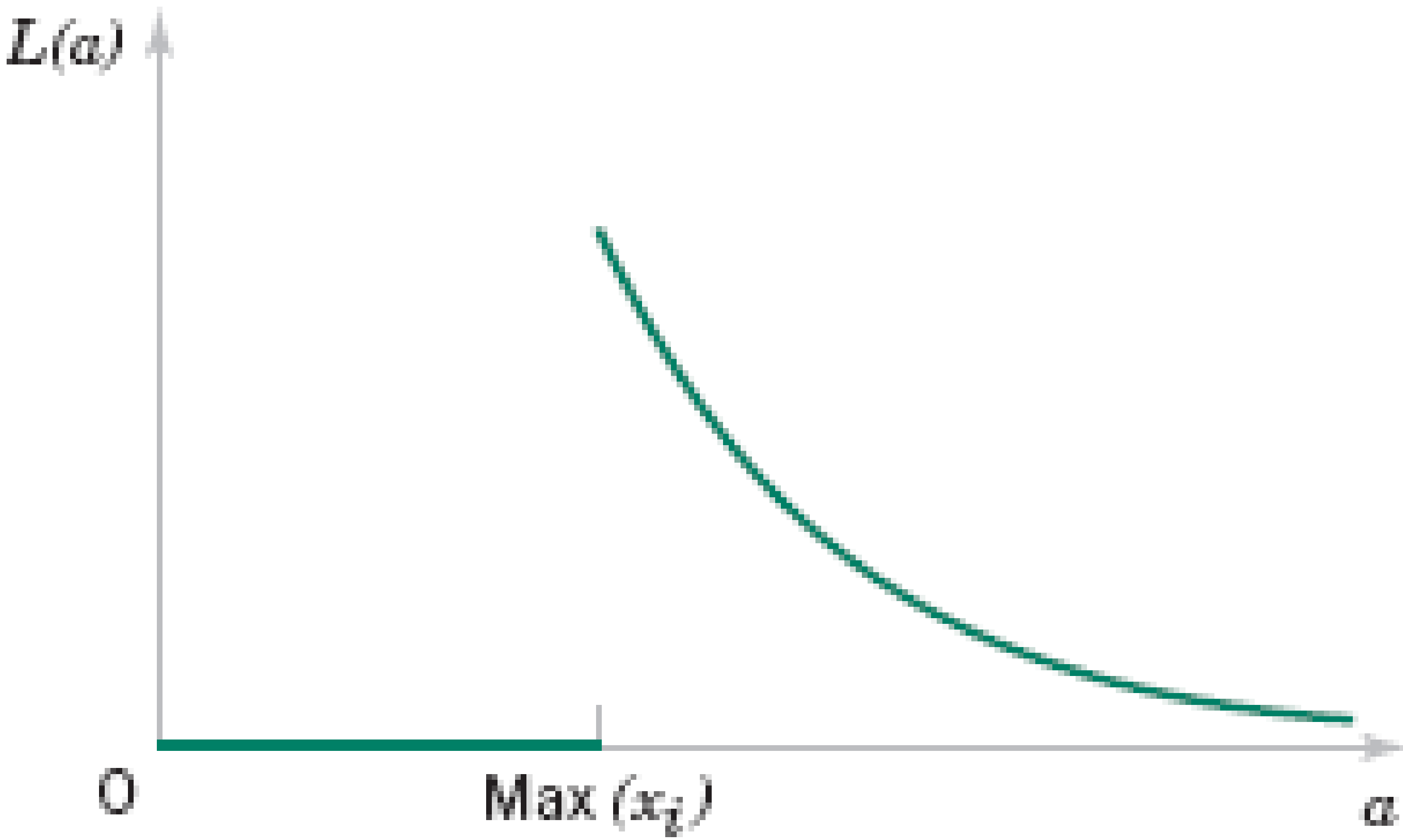



**Figure 7-8** The likelihood function for the uniform distribution in Example 7-13.

# 7-4 Methods of Point Estimation

## 7-4.3 Bayesian Estimation of Parameters

Suppose that we have some additional information about $\theta$ and that we can summarize that information in the form of a probability distribution for $\theta$, say, $f(\theta)$. This probability distribution is often called the **prior distribution** for $\theta$, and suppose that the mean of the prior is $\mu_0$ and the variance is $\sigma_0^2$. This is a very novel concept insofar as the rest of this book is concerned because we are now viewing the parameter $\theta$ as a random variable. The probabilities associated with the prior distribution are often called **subjective probabilities,** in that they usually reflect the analyst's degree of belief regarding the true value of $\theta$. The Bayesian approach to estimation uses the prior distribution for $\theta$, $f(\theta)$, and the joint probability distribution of the sample, say $f(x_1, x_2, \ldots, x_n \mid \theta)$, to find a **posterior distribution** for $\theta$, say, $f(\theta \mid x_1, x_2, \ldots, x_n)$. This posterior distribution contains information both from the sample and the prior distribution for $\theta$. In a sense, it expresses our degree of belief regarding the true value of $\theta$ after observing the sample data. It is easy conceptually to find the posterior distribution. The joint probability distribution of the sample $X_1, X_2, \ldots, X_n$ and the parameter $\theta$ (remember that $\theta$ is a random variable) is

$$f(x_1, x_2, \ldots, x_n, \theta) = f(x_1, x_2, \ldots, x_n \mid \theta) f(\theta)$$

# 7-4 Methods of Point Estimation

## 7-4.3 Bayesian Estimation of Parameters

and the marginal distribution of $X_1, X_2, \ldots, X_n$ is

$$f(x_1, x_2, \ldots, x_n) = \begin{cases} \sum_{\theta} f(x_1, x_2, \ldots, x_n, \theta), & \theta \text{ discrete} \\ \int_{-\infty}^{\infty} f(x_1, x_2, \ldots, x_n, \theta)\, d\theta, & \theta \text{ continuous} \end{cases}$$

Therefore, the desired distribution is

$$f(\theta \mid x_1, x_2, \ldots, x_n) = \frac{f(x_1, x_2, \ldots, x_n, \theta)}{f(x_1, x_2, \ldots, x_n)}$$

We define the **Bayes estimator** of $\theta$ as the value $\tilde{\theta}$ that corresponds to the mean of the posterior distribution $f(\theta \mid x_1, x_2, \ldots, x_n)$.

# 7-4 Methods of Point Estimation

## Example 7-16

Let $X_1, X_2, \ldots, X_n$ be a random sample from the normal distribution with mean $\mu$ and variance $\sigma^2$, where $\mu$ is unknown and $\sigma^2$ is known. Assume that the prior distribution for $\mu$ is normal with mean $\mu_0$ and variance $\sigma_0^2$; that is

$$f(\mu) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-(\mu-\mu_0)^2/(2\sigma_0^2)} = \frac{1}{\sqrt{2\pi\sigma_0^2}} e^{-(\mu^2-2\mu_0\mu+\mu_0^2)/(2\sigma_0^2)}$$

The joint probability distribution of the sample is

$$f(x_1, x_2, \ldots, x_n \mid \mu) = \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2\sigma^2)\sum_{i=1}^{n}(x_i-\mu)^2}$$

$$= \frac{1}{(2\pi\sigma^2)^{n/2}} e^{-(1/2\sigma^2)(\sum x_i^2 - 2\mu\sum x_i + n\mu^2)}$$

# 7-4 Methods of Point Estimation

## Example 7-16 (Continued)

Thus, the joint probability distribution of the sample and μ is

$$f(x_1, x_2, \ldots, x_n, \mu) = \frac{1}{(2\pi\sigma^2)^{n/2}\sqrt{2\pi}\sigma_0} e^{-(1/2)[(1/\sigma_0^2 + n/\sigma^2)\mu^2 - (2\mu_0/\sigma_0^2 + 2\Sigma x_i/\sigma^2)\mu + \Sigma x_i^2/\sigma^2 + \mu_0^2/\sigma_0^2]}$$

$$= e^{-(1/2)\left[\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}\right)\mu^2 - 2\left(\frac{\mu_0}{\sigma_0^2} + \frac{\bar{x}}{\sigma^2/n}\right)\mu\right]} h_1(x_1, \ldots, x_n, \sigma^2, \mu_0, \sigma_0^2)$$

Upon completing the square in the exponent

$$f(x_1, x_2, \ldots, x_n, \mu) = e^{-(1/2)\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}\right)\left[\mu^2 - \left(\frac{(\sigma^2/n)\mu_0}{\sigma_0^2 + \sigma^2/n} + \frac{\bar{x}\sigma_0^2}{\sigma_0^2 + \sigma^2/n}\right)\right]^2} h_2(x_1, \ldots, x_n, \sigma^2, \mu_0, \sigma_0^2)$$

where $h_i(x_1, \ldots, x_n, \sigma^2, \mu_0, \sigma_0^2)$ is a function of the observed values, $\sigma^2$, $\mu_0$, and $\sigma_0^2$.

# 7-4 Methods of Point Estimation

## Example 7-16 (Continued)

Now, because $f(x_1, \ldots, x_n)$ does not depend on $\mu$,

$$f(\mu \mid x_1, \ldots, x_n) = e^{-(1/2)\left(\frac{1}{\sigma_0^2}+\frac{1}{\sigma^2/n}\right)\left[\mu^2-\left(\frac{(\sigma^2/n)\mu_0+\sigma_0^2\bar{x}}{\sigma_0^2+\sigma^2/n}\right)\right]} h_3(x_1, \ldots, x_n, \sigma^2, \mu_0, \sigma_0^2)$$

This is recognized as a normal probability density function with posterior mean

$$\frac{(\sigma^2/n)\mu_0 + \sigma_0^2\bar{x}}{\sigma_0^2 + \sigma^2/n}$$

and posterior variance

$$\left(\frac{1}{\sigma_0^2} + \frac{1}{\sigma^2/n}\right)^{-1} = \frac{\sigma_0^2(\sigma^2/n)}{\sigma_0^2 + \sigma^2/n}$$

# 7-4 Methods of Point Estimation

## Example 7-16 (Continued)

Consequently, the Bayes estimate of $\mu$ is a weighted average of $\mu_0$ and $\bar{x}$. For purposes of comparison, note that the maximum likelihood estimate of $\mu$ is $\hat{\mu} = \bar{x}$.

To illustrate, suppose that we have a sample of size $n = 10$ from a normal distribution with unknown mean $\mu$ and variance $\sigma^2 = 4$. Assume that the prior distribution for $\mu$ is normal with mean $\mu_0 = 0$ and variance $\sigma_0^2 = 1$. If the sample mean is 0.75, the Bayes estimate of $\mu$ is

$$\frac{(4/10)0 + 1(0.75)}{1 + (4/10)} = \frac{0.75}{1.4} = 0.536$$

Note that the maximum likelihood estimate of $\mu$ is $\bar{x} = 0.75$.

## IMPORTANT TERMS AND CONCEPTS

- Bayes estimator
- Bias in parameter estimation
- Central limit theorem
- Estimator versus estimate
- Likelihood function
- Maximum likelihood estimator
- Mean square error of an estimator
- Minimum variance unbiased estimator
- Moment estimator
- Normal distribution as the sampling distribution of a sample mean
- Normal distribution as the sampling distribution of the difference in two sample means
- Parameter estimation
- Point estimator
- Population or distribution moments
- Posterior distribution
- Prior distribution
- Sample moments
- Sampling distribution
- Standard error and estimated standard error of an estimator
- Statistic
- Statistical inference
- Unbiased estimator