CHAPTER 7

ESTIMATION

In Chap. 6, we found that once the family to which a random variable belongs is determined, the problem of approximating or *estimating* the numerical value of pertinent parameters remains. Even though we were able to define sample statistics that allow us to estimate the mean, variance, and standard deviation of a random variable in a logical manner, we were unable to assess their effectiveness. In this chapter we consider the mathematical properties of these statistics. We also present a brief introduction to the theory of estimation. The ideas developed here will be used extensively throughout the remainder of the text.

7.1 POINT ESTIMATION

In an estimation problem there is at least one parameter θ whose value is to be approximated on the basis of a sample. The approximation is done by using an appropriate statistic. A statistic used to approximate or estimate a population parameter θ is called a *point estimator* for θ and is denoted by $\hat{\theta}$ (the symbol is called a "hat"); the numerical value assumed by this statistic when evaluated for a given sample is called a *point estimate* for θ. For example, in estimating the mean coal consumption by electric utilities for a given year (see Example 6.3.1), the statistic $\bar{X}$ was used. Thus $\bar{X}$ is a point estimator for μ and we write $\hat{\mu} = \bar{X}$. In Example 6.3.1 we evaluated this statistic for a particular sample and obtained the value 408.3 million tons. This number is called a *point estimate* for μ. Note that there is a difference in the terms "estimator" and "estimate." The estimator is the statistic used to generate the estimate; it is a random variable. An estimate is a number.

Once a logical point estimator for a parameter θ has been developed, the natural question to ask is, "How good is this estimator?" Obviously, we want the estimator to generate estimates that can be expected to be close in value to θ. This can be expected to occur if the estimator $\hat{\theta}$ possesses two properties.

Desirable Properties of a Point Estimator

1. $\hat{\theta}$ to be *unbiased* for θ.
2. $\hat{\theta}$ to have a small variance for large sample sizes.

The word "unbiased" was explained graphically in Chap. 6. Basically, it means "centered at the right spot," where the right spot is the parameter being estimated. The term "unbiased" is a technical term. To be able to prove analytically that an estimator $\hat{\theta}$ is an unbiased estimator for a parameter θ, we need a formal definition for the term. This definition is given here.

Definition 7.1.1 (Unbiased). An estimator $\hat{\theta}$ is an unbiased estimator for a parameter θ if and only if $E[\hat{\theta}] = \theta$.

Recall that $\hat{\theta}$ is a statistic; therefore it is also a random variable and, as such, has a mean, or expected, value. To say that $\hat{\theta}$ is unbiased for θ implies that the mean of the estimator $\hat{\theta}$ is equal to the parameter θ that it is estimating. Thus an estimator $\hat{\mu}$ is an unbiased estimator for μ if and only if $E[\hat{\mu}] = \mu$; an estimator $\hat{\sigma}^2$ is unbiased for σ^2 if and only if $E[\hat{\sigma}^2] = \sigma^2$; an estimator $\hat{\sigma}$ is unbiased for σ if and only if $E[\hat{\sigma}] = \sigma$. Let us reexamine the estimators $\bar{X}$, S^2, and S developed in Chap. 6 in light of this new definition.

Theorem 7.1.1. Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample of size n from a distribution with mean μ. The sample mean, $\bar{X}$, is an unbiased estimator for μ.

Proof. By Definition 6.3.1,

$$E[\bar{X}] = E[1/n(X_1 + X_2 + X_3 + \cdots + X_n)]$$

By the Rules for Expectation (Theorem 3.3.1),

$$E[\bar{X}] = 1/n(E[X_1] + E[X_2] + E[X_3] + \cdots + E[X_n])$$

Since $X_1, X_2, X_3, \ldots, X_n$ constitutes a random sample from a distribution with mean μ, each of these random variables has mean μ. Therefore

$$E[\bar{X}] = 1/n\Big(\underbrace{\mu + \mu + \mu + \cdots + \mu}_{n \text{ terms}}\Big) = 1/n(n\mu) = \mu$$

and the proof is complete.

It is important to realize that since $\hat{\theta}$ is a statistic, in repeated sampling the estimates generated will vary from sample to sample. To say that $\hat{\theta}$ is unbiased for θ implies that these estimates vary about θ; it also implies that the *average* value of these estimates can be expected to lie reasonably close to θ. For example, since $\bar{X}$ is unbiased for μ, for k repetitions of an experiment the observed sample means $\bar{x}_1$, $\bar{x}_2, \bar{x}_3, \ldots, \bar{x}_k$ will vary about μ and the *average* value of these k estimates should lie reasonably close to μ.

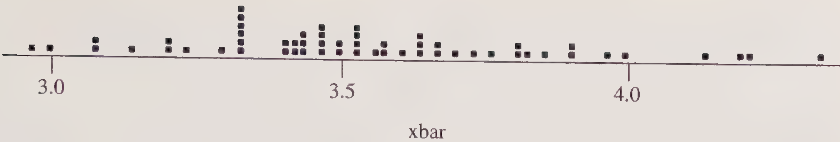

FIGURE 7.1
Plot of experimental x values of Example 7.1.1.

Example 7.1.1. Consider the experiment of rolling a single fair die. Let X denote the number obtained. X is discrete, with density given by

$$f(x) = 1/6 \qquad x = 1, 2, 3, 4, 5, 6$$

The average value of X is

$$\mu = E[X] = \Sigma\, x f(x) = 3.5$$

Now consider tossing a single die 30 times and recording the average toss, $\bar{x}$. If this process is repeated many times the $\bar{x}$ values will vary from sample to sample. Since $\bar{X}$ is an unbiased estimator for the true average value, μ, the observed $\bar{x}$ values are expected to vary around the value 3.5. This experiment was conducted in class 56 times. The results were as follows:

3.43	3.33	3.60	2.97	3.50	4.20	3.07	3.33	3.86	3.80	3.00
3.30	3.40	4.00	3.83	3.43	3.90	3.07	3.47	3.23	3.20	3.76
3.50	3.70	4.13	3.97	3.42	3.57	3.47	3.80	3.53	3.20	3.13
3.63	3.33	3.33	3.56	3.47	4.33	3.53	3.33	3.47	3.73	3.90
3.32	4.21	3.63	3.67	3.53	3.43	3.40	3.53	3.63	3.42	3.67
3.57										

Figure 7.1 shows a dot plot of these data. Notice that, as expected, the $\bar{x}$ values vary and the value 3.5 is close to the center of the data points. The average of the 56 $\bar{x}$ values is 3.548, a little higher than the ideal theoretical value of 3.5.

It is equally important to understand what the term "unbiased" does *not* imply. It does not imply that any *one* estimate will be close in value to the parameter being estimated. In reference to Example 6.3.1, the estimated mean coal consumption by electric utilities was $\hat{\mu} = \bar{x} \doteq 408.3$ million tons. This estimate is unbiased in the sense that it was generated by means of the unbiased estimator $\bar{X}$. This *alone* does not guarantee that the actual mean coal consumption by electric utilities across the country is anywhere close to 408.3 million tons. This is unfortunate. Usually, statistical studies are not repeated over and over so that the estimates obtained can be averaged. In general, only one sample is drawn; one estimate is obtained. To have some assurance that this estimate is close in value to θ, the parameter being estimated, ideally the estimator used not only should be unbiased, but also it should have a small variance for large sample sizes. In this way, even though the estimated values fluctuate about θ, the variability is small. Each estimate produced can be expected to be fairly close in value to θ. Theorem 7.1.2 shows that $\bar{X}$ has this property.

Theorem 7.1.2. Let $\bar{X}$ be the sample mean based on a random sample of size n from a distribution with mean μ and variance σ^2. Then

$$\operatorname{Var} \bar{X} = \frac{\sigma^2}{n}$$

The proof of this theorem is based on the Rules for Variance (Theorem 3.3.3) and is similar to that of Theorem 7.1.1. Note that since σ^2 is constant, as the sample size n increases, the variance of $\bar{X}$, σ^2/n, decreases and can be made as small as we wish by choosing n sufficiently large. This implies that a sample mean based on a large sample can be expected to lie reasonably close to μ; one based on a small sample may vary widely from the actual population mean. This points out the advantages of working with a large sample and the danger of placing too much emphasis on conclusions drawn from small samples. Keep in mind that many of the examples and exercises presented in this text are based on small samples. This is done for illustrative purposes only. We do *not* mean to imply that samples this small are common in research.

Since the standard deviation of any random variable is the square root of its variance, the standard deviation of the sample mean is the square root of the variance of $\bar{X}$. Thus the standard deviation of $\bar{X}$ is $\sqrt{\sigma^2/n} = \sigma/\sqrt{n}$. This standard deviation plays a vital role in the development of techniques used in making inferences on the true value of μ based on information concerning the observed value of $\bar{x}$. The name given to this special standard deviation is *standard error of the mean.*

Definition 7.1.2 (Standard error of the mean). Let $\bar{X}$ denote the sample mean based on a sample of size n drawn from a distribution with standard deviation σ. The standard deviation of $\bar{X}$ is given by $\sigma/\sqrt{n}$ and is called the standard error of the mean.

In Chap. 6 we defined the sample variance S^2 by dividing $\sum_{i=1}^{n}(X_i - \bar{X})^2$ by $n - 1$. This was done so that the resulting estimator would be unbiased for σ^2. This result is stated formally in Theorem 7.1.3. The proof of this theorem is found in Appendix C.

Theorem 7.1.3. Let S^2 be the sample variance based on a random sample of size n from a distribution with mean μ and variance σ^2. S^2 is an unbiased estimator for σ^2.

It should be noted that even though S^2 is an unbiased estimator for σ^2, it can be shown that S is not unbiased for σ (see Exercise 8). This emphasizes the fact that unbiasedness is desirable in an estimator but not essential.

7.2 THE METHOD OF MOMENTS AND MAXIMUM LIKELIHOOD

In this section we consider two methods for deriving point estimators for distribution parameters. The first, called the *method of moments,* is a simple method that was first proposed by Karl Pearson in 1894. The second, called the *method of maximum likelihood,* is more complex. It was used by C. F. Gauss to solve isolated problems over 170 years ago. In the early 1900s the method was formalized by R. A. Fisher and has been used extensively since that time.

To begin, recall that terms of the form $E[X^k]$ ($k = 1, 2, 3, \ldots$) are called the *kth moments* for X. Since an expectation is a theoretical average, logic implies that the moments for X can be estimated via an arithmetic average. That is, an estimator M_k for $E[X^k]$ based on a random sample of size n is

$$M_k = \sum_{i=1}^{n} \frac{X_i^k}{n}$$

For example,

$$M_1 = \sum_{i=1}^{n} (X_i/n) = \bar{X}$$

$$M_2 = \sum_{i=1}^{n} (X_i^2/n)$$

$$M_3 = \sum_{i=1}^{n} (X_i^3/n)$$

and so forth.

The method of moments exploits the fact that in many cases the moments for X can be expressed as a function of θ, the parameter to be estimated. We can often obtain a reasonable estimator for θ by replacing the theoretical moments by their estimators and solving the resulting equation for $\hat{\theta}$.

You have already used the technique quite naturally in solving some of the problems in the last section! We now formalize the idea. The technique is illustrated by finding the method of moments estimator for the parameter p of a binomial random variable.

Example 7.2.1. A forester plants five rows of 20 pine seedlings, each row to serve as an eventual windbreak. The soil and wind conditions to which the seedlings are subjected are identical. The variable being studied is X, the number of seedlings per row that survive the first winter. We are dealing with a random sample of size $m = 5$ from a binomial distribution with parameters $n = 20$ and p unknown. We want to use the method of moments to derive an estimator for p. To do so, note that since X is binomial,

$$E[X] = np = 20p$$

We now replace the first moment of X, $E[X]$, by its estimator $M_1 = (\Sigma_{i=1}^{5} X_i)/5 = \bar{X}$ to obtain the equation

$$\bar{X} = 20\hat{p}$$

This equation is solved for $\hat{p}$ to obtain the estimator

$$\hat{p} = \overline{X}/20$$

When the experiment is conducted, these data result:

$$x_1 = 18 \qquad x_3 = 15 \qquad x_5 = 20$$
$$x_2 = 17 \qquad x_4 = 19$$

For these data $\bar{x} = (\sum_{i=1}^{5} x_i)/5 = 17.8$. The method of moments estimate for p, the probability that a seedling will survive the first winter, is

$$\hat{p} = \bar{x}/20 = 17.8/20 = .89$$

Occasionally there are two parameters, θ_1 and θ_2, to be estimated from a single sample. To use the method of moments in this case, we must obtain two equations relating the moments of the distribution to these parameters. We then replace the theoretical moments by their estimators and solve the resulting equations simultaneously for $\hat{\theta}_1$ and $\hat{\theta}_2$. This idea is illustrated by finding estimators for α and β, the parameters that identify the gamma distribution.

Example 7.2.2. Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a gamma distribution with parameters α and β. From Theorem 4.3.2 we know that $E[X] = \alpha\beta$ and $\operatorname{Var} X = \alpha\beta^2$. Recall that since $\operatorname{Var} X = E[X^2] - (E[X])^2$, the first two moments of X are functions of α and β. The equations relating the moments to these unknown parameters are

$$E[X] = \alpha\beta$$
$$E[X^2] - (E[X])^2 = \alpha\beta^2$$

We now replace $E[X]$ and $E[X^2]$ by their estimators, M_1 and M_2, respectively, to obtain

$$M_1 = \hat{\alpha}\hat{\beta}$$
$$M_2 - M_1^2 = \hat{\alpha}\hat{\beta}^2$$

Solving this set of equations simultaneously, we see that

$$M_2 - M_1^2 = M_1\hat{\beta}$$

This implies that

$$\hat{\beta} = (M_2 - M_1^2)/M_1$$

and

$$\hat{\alpha} = M_1/\hat{\beta} = M_1^2/(M_2 - M_1^2)$$

Maximum Likelihood Estimators

The maximum likelihood method for deriving estimators is more complex than the method of moments. However, it is based on an appealing notion. Recall that the density f for a random variable X usually has at least one parameter θ associated with it. Assume that we have a random sample $x_1, x_2, x_3, \ldots, x_n$ available. The method of maximum likelihood in a sense picks out of all the possible values of θ the one most likely to have produced these observations. Before formalizing the method, let us demonstrate the idea in a simple context.

Example 7.2.3. Water samples of a specific size are taken from a river suspected of having been polluted by improper treatment procedures at an upstream sewage disposal plant. Let X denote the number of coliform organism found per sample, and assume that X is a Poisson random variable with parameter k. Let $x_1, x_2, x_3, \ldots, x_n$ be a random sample from the distribution of X. We want to determine the value of k that gives the highest probability of observing this sample. Since random sampling implies independence,

$$\begin{aligned} &P[X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n] \\ &\quad = P[X_1 = x_1]P[X_2 = x_2] \cdots P[X_n = x_n] \\ &\quad = \prod_{i=1}^{n} P[X_i = x_i] \end{aligned}$$

Recall that the density for X is given by

$$P[X = x] = f(x) = \frac{e^{-k}k^x}{x!} \qquad x = 0, 1, 2, \ldots$$

Therefore the probability of obtaining the given sample is

$$\prod_{i=1}^{n} P[X_i = x_i] = \prod_{i=1}^{n} f(x_i) = \prod_{i=1}^{n} \frac{e^{-k}k^{x_i}}{x_i!}$$

Note that this probability is a function of k, which we denote by $L(k)$. Using the laws of exponents,

$$L(k) = \frac{e^{-nk}k^{\sum_{i=1}^{n} x_i}}{\prod_{i=1}^{n} x_i!}$$

This function is called the "likelihood function." It gives us the probability of observing the values $x_1, x_2, \ldots, x_n$ as a function of the parameter k. We want to find the value of k that maximizes this probability. That is, of all the possible values for k, we want to find the one that gives us the highest probability of observing the values that we did observe. To find this value of k, we use elementary calculus to maximize the likelihood function. This can be done directly. However, to simplify the process, we first take the natural logarithm of $L(k)$ and use the laws of logarithms to simplify the resulting expression

$$\ln L(k) = -nk + \sum_{i=1}^{n} x_i \ln k - \ln \prod_{i=1}^{n} x_i!$$

The value of k that maximizes $\ln L(k)$ also maximizes $L(k)$. Therefore, to complete the derivation, we differentiate $\ln L(k)$ with respect to k, set the derivative equal to 0, and solve for k:

$$\frac{d \ln L(k)}{dk} = -n + \left(\sum_{i=1}^{n} x_i \right) \Big/ k$$

$$0 = -n + \left(\sum_{i=1}^{n} x_i \right) \Big/ k$$

$$k = \left(\sum_{i=1}^{n} x_i \right) \Big/ n = x$$

Since this procedure does not give us the exact value of k but rather provides a logical method for estimating k, we write $\hat{k} = \overline{X}$. That is, the sample mean is the "maximum likelihood estimator" for the parameter k of a Poisson random variable.

Suppose that a random sample of size 4 yields these data:

$$x_1 = 12 \qquad x_2 = 15 \qquad x_3 = 16 \qquad x_4 = 17$$

Since the value of k that is most likely to have produced this sample is $x = 15$, it is natural to take this value as our estimate for k.

Although our example involves a discrete random variable, the same general method is used in the continuous case. This method is summarized as follows:

Method of Moments Technique for Estimating θ

1. Obtain a random sample $x_1, x_2, x_3, \ldots, x_n$ from the distribution of a random variable X with density f and associated parameter θ.
2. Define a function $L(\theta)$ by

$$L(\theta) = \prod_{i=1}^{n} f(x_i)$$

   This function is called the *likelihood function* for the sample.
3. Find the expression for θ that maximizes the likelihood function. This can be done directly or by maximizing $\ln L(\theta)$.
4. Replace θ by $\hat{\theta}$ to obtain an expression for the maximum likelihood estimator for θ.
5. Find the observed value of this estimator for a given sample.

As with the method of moments, the maximum likelihood procedure can be applied when the density for X is characterized by two parameters. We illustrate the technique by finding the maximum likelihood estimators for μ and σ^2, the mean and variance of a normal random variable.

Example 7.2.4. Let $x_1, x_2, x_3, \ldots, x_n$ be a random sample from a normal distribution with mean μ and variance σ^2. The density for X is

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(1/2)[(x-\mu)/\sigma]^2}$$

The likelihood function for the sample is a function of both μ and σ. In particular,

$$
\begin{aligned}
L(\mu, \sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} e^{-(1/2)[(x_i-\mu)/\sigma]^2} \\
&= \left(\frac{1}{\sqrt{2\pi}}\right)^n \left(\frac{1}{\sigma}\right)^n e^{-(1/2\sigma^2)\sum_{i=1}^{n}(x_i-\mu)^2}
\end{aligned}
$$

The logarithm of the likelihood function is

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

To maximize this function, we take the partial derivatives with respect to μ and σ, set these derivatives equal to 0, and solve the equations simultaneously for μ and σ:

$$
\begin{cases}
\dfrac{\partial \ln L(\mu, \sigma)}{\partial \mu} = \dfrac{1}{\sigma^2} \displaystyle\sum_{i=1}^{n} (x_i - \mu) \\[2ex]
\dfrac{\partial \ln L(\mu, \sigma)}{\partial \sigma} = \dfrac{-n}{\sigma} + \displaystyle\sum_{i=1}^{n} \dfrac{(x_i - \mu)^2}{\sigma^3} = \dfrac{-n\sigma^2 + \sum_{i=1}^{n} (x_i - \mu)^2}{\sigma^3}
\end{cases}
$$

$$
\begin{cases}
\dfrac{1}{\sigma^2} \displaystyle\sum_{i=1}^{n} (x_i - \mu) = 0 \\[2ex]
\dfrac{-n\sigma^2 + \sum_{i=1}^{n} (x_i - \mu)^2}{\sigma^3} = 0
\end{cases}
$$

$$
\begin{cases}
\displaystyle\sum_{i=1}^{n} x_i - n\mu = 0 & \text{or} \quad \mu = \left(\displaystyle\sum_{i=1}^{n} x_i\right) \Big/ n = x \\[2ex]
-n\sigma^2 + \displaystyle\sum_{i=1}^{n} (x_i - \mu)^2 = 0 & \text{or} \quad \sigma^2 = \left[\displaystyle\sum_{i=1}^{n} (x_i - \mu)^2\right] \Big/ n
\end{cases}
$$

Realizing that these are not the true values of μ and σ^2 but are only estimates, we see that the maximum likelihood estimators for these parameters are

$$
\hat{\mu} = \bar{X}
$$

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

The method of moments estimator for a parameter and the maximum likelihood estimator often agree. However, if they do not, the maximum likelihood estimator is usually preferred.

7.3 FUNCTIONS OF RANDOM VARIABLES—DISTRIBUTION OF $\bar{X}$

There is one drawback to point estimation. It yields a single value for the unknown parameter θ. Is there any assurance that this estimate is even close in value to θ? The best answer is that in most cases the point estimators used are logical. To get an idea not only of the value of the parameter being estimated, but also of the accuracy of the estimate, researchers turn to the method of *interval estimation* or *confidence intervals.* An interval estimator is what the name implies. It is a random interval, an interval whose endpoints L_1 and L_2 are each statistics. It is used to determine a numerical interval based on a sample. It is hoped that the numerical interval obtained will contain the population parameter being estimated. By expanding from a point to an interval, we create a little room for error and in so doing gain the ability, based on probability theory, to report the confidence that we have in the estimate.

In later chapters we shall derive confidence intervals for many important parameters. To do so, we must know the distribution of some key random variables. In this section we consider a technique for identifying the distribution of a random variable from its moment generating function. This technique depends on the result given in Theorem 7.3.1.

Theorem 7.3.1. Let X and Y be random variables with moment generating functions $m_X(t)$ and $m_Y(t)$, respectively. If $m_X(t) = m_Y(t)$ for all t in some open interval about 0, then X and Y have the same distribution.

The proof of this theorem is based on transform theory and is beyond the scope of this text. The theorem implies that the moment generating function, when it exists, serves as a "fingerprint" for the random variable. We illustrate this idea by proving Theorem 4.4.3, the "standardization" theorem for normal random variables.

Example 7.3.1. (Proof of the standardization theorem) Let X be a normal random variable with mean μ and variance σ^2. Recall from Theorem 4.4.1 that the moment generating function for X is

$$m_X(t) = E[e^{tX}] = e^{\mu t + \sigma^2 t^2/2}$$

The moment generating function for a standard normal random variable Z is

$$m_Z(t) = e^{0t + (1)^2 t^2/2} = e^{t^2/2}$$

Let $Y = (X - \mu)/\sigma = (1/\sigma)X - \mu/\sigma$. The moment generating function for Y is given by

$$\begin{aligned} m_Y(t) = E[e^{tY}] &= E[e^{(t/\sigma)X - (\mu/\sigma)t}] \\ &= E[e^{(t/\sigma)X} e^{(-\mu/\sigma)t}] \\ &= e^{(-\mu/\sigma)t} E[e^{(t/\sigma)X}] \end{aligned}$$

Note that $E[e^{(t/\sigma)X}] = m_X(t/\sigma) = e^{\mu(t/\sigma) + \sigma^2 t^2/2\sigma^2}$. Substituting, we obtain

$$m_Y(t) = e^{(-\mu/\sigma)t} e^{(\mu/\sigma)t + t^2/2} = e^{t^2/2} = m_Z(t)$$

We have shown that Y and Z have the same moment generating function. By Theorem 7.3.1 these variables have the same distribution. In particular, they are both *standard normal* random variables.

Many of the statistics used in data analysis entail summing a collection of random variables. The following theorem together with Theorem 7.3.1 will help to determine the distribution of such statistics.

Theorem 7.3.2. Let X_1 and X_2 be independent random variables with moment generating functions $m_{X_1}(t)$ and $m_{X_2}(t)$, respectively. Let $Y = X_1 + X_2$. The moment generating function for Y is given by

$$m_Y(t) = m_{X_1}(t)m_{X_2}(t)$$

Proof. By definition

$$m_Y(t) = E[e^{tY}] = E[e^{tX_1+tX_2}] = E[e^{tX_1}e^{tX_2}]$$

Since X_1 and X_2 are independent, e^{tX_1} and e^{tX_2} are also independent. By Theorem 5.2.2

$$m_Y(t) = E[e^{tX_1}e^{tX_2}] = E[e^{tX_1}]E[e^{tX_2}] = m_{X_1}(t)m_{X_2}(t)$$

This theorem can be extended easily to include a sum of more than two random variables. That is, we can say that the moment generating function for the sum of a finite number of *independent* random variables is the product of the moment generating functions of the individual variables. The requirement that the random variables be independent is not restrictive, since in most cases the sum of interest is a function of the elements of a random sample. The term "random sample" implies independence. (See Definition 6.1.1.) Theorem 7.3.2 is illustrated by showing that the sum of a collection of independent normal random variables is normal.

Example 7.3.2. (Distribution of the sum of independent normally distributed random variables) Let $X_1, X_2, X_3, \ldots, X_n$ be independent normal random variables with means $\mu_1, \mu_2, \mu_3, \ldots, \mu_n$ and variances $\sigma_1^2, \sigma_2^2, \sigma_3^2, \ldots, \sigma_n^2$, respectively. Let $Y = X_1 + X_2 + X_3 + \cdots + X_n$. Note that the moment generating function for X_i is given by

$$m_{X_i}(t) = e^{\mu_i t+\sigma_i^2 t^2/2} \qquad i = 1, 2, 3, \ldots, n$$

and the moment generating function for Y is

$$m_Y(t) = \prod_{i=1}^{n} m_{X_i}(t) = \exp\left[\left(\sum_{i=1}^{n} \mu_i\right)t + \left(\sum_{i=1}^{n} \sigma_i^2\right)t^2\Big/2\right]$$

The function on the right is the moment generating function for a normal random variable with mean $\mu = \Sigma_{i=1}^n \mu_i$ and variance $\sigma^2 = \Sigma_{i=1}^n \sigma_i^2$.

Distribution of $\bar{X}$

One of the more useful statistics that we have studied is $\bar{X}$, the sample mean. Since $\bar{X}$ is a statistic, it is also a random variable. It makes sense to ask, "What is the distribution of $\bar{X}$?" We have already seen that the center of location for $\bar{X}$ is μ, the mean of the population from which the sample is drawn. We have also seen that its variance is σ^2/n, the original population variance divided by the sample size. We have not yet mentioned the type of distribution possessed by the statistic. Does $\bar{X}$ follow some distribution such as the gamma, uniform, or normal distributions that we have already studied, or must we introduce a new distribution now? The next theorem, whose derivation is outlined in Exercise 38, will help us to answer this question.

Theorem 7.3.3. Let X be a random variable with moment generating function $m_X(t)$. Let $Y = \alpha + \beta X$. The moment generating function for Y is

$$m_Y(t) = e^{\alpha t} m_X(\beta t)$$

We illustrate the use of this theorem in a numerical context.

Example 7.3.3. Let X denote the maximum wind speed per day recorded at the weather station of a particular locality. Assume that X is normally distributed with mean 10 miles per hour (mph) and standard deviation 4 mph. Engineers are constructing a bridge over a deep canyon in the area. They suspect that the maximum wind speed at the bridge site is given by $Y = 2X - 5$. What is the distribution of Y? To answer this question we first note that the moment generating function for X is

$$\begin{aligned} m_X(t) &= e^{\mu t + \sigma^2 t^2/2} \\ &= e^{10t + 16t^2/2} \end{aligned}$$

We next apply Theorem 7.3.3 with $\alpha = -5$ and $\beta = 2$ to see that the moment generating function for Y is

$$\begin{aligned} m_Y(t) &= e^{-5t} e^{10(2t) + 16(2t)^2/2} \\ &= e^{15t + 64t^2/2} \end{aligned}$$

This is the moment generating function for a normal random variable with mean 15 mph and variance 64. Since the moment generating function for a random variable is its fingerprint, we know that the maximum speed at the bridge site is normally distributed with an average speed of 15 mph and a standard deviation of 8 mph.

Theorem 7.3.3 is interesting in its own right, but its primary purpose at this time is to help us derive the next very important theorem. This theorem answers the question posed earlier concerning the distribution of $\bar{X}$. In particular, it assures us that when sampling from a *normal* distribution the random variable $\bar{X}$ will itself be *normally* distributed.

Theorem 7.3.4 (Distribution of $\bar{X}$—normal population). Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. Then $\bar{X}$ is normally distributed with mean μ and variance σ^2/n.

The derivation of this theorem is not hard. It is outlined in Exercises 38 to 42. We feel that by working through the derivation for yourself you will have a better understanding of the point being made. The other exercises presented are also important. They contain some results that will have major practical consequences later. Be sure to give them all a try!

7.4 INTERVAL ESTIMATION AND THE CENTRAL LIMIT THEOREM

As mentioned previously, point estimation does not give us the ability to report the accuracy of our estimate. To do this, we must turn to the method of interval estimation. The statistics used to extend a point estimate for a parameter θ to an interval of values that should contain the true value of θ vary from parameter to parameter. However, the method for deriving these statistics is basically the same in each case. In this section we illustrate the method by deriving a "confidence interval" for the mean of a normal random variable when its variance is assumed to be known. In later chapters we apply the general technique illustrated here to find confidence intervals for other important parameters.

The term "confidence interval" is a technical term that we now define.

Definition 7.4.1 (Confidence interval). A $100(1 - \alpha)\%$ confidence interval for a parameter θ is a random interval $[L_1, L_2]$ such that

$$P[L_1 \leq \theta \leq L_2] \doteq 1 - \alpha$$

regardless of the value of θ.

One general statement will guide in the construction of most of the confidence intervals presented in this text:

> To construct a $100(1 - \alpha)\%$ confidence interval for a parameter θ, we shall find a random variable whose expression involves θ and whose probability distribution is known at least approximately.

Confidence Interval on the Mean: Variance Known

To use this guideline to find a $100(1 - \alpha)\%$ confidence interval for the mean of a normal random variable whose variance is known, we must find a random variable whose expression involves μ and whose distribution is known. This is easy to do. Note that in Theorem 7.3.4, we showed that under the given conditions the sample mean, $\bar{X}$, is normally distributed with mean μ and variance σ^2/n. This implies that the random variable

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

is *standard normal.* Note that this random variable involves the parameter μ and its distribution is known. We illustrate how this random variable can be used to generate a 95% confidence interval for μ. The technique used can be generalized easily to obtain any desired degree of confidence.

Example 7.4.1. Acute myeloblastic leukemia is among the most deadly of cancers. Past experience indicates that the time in months that a patient survives after initial

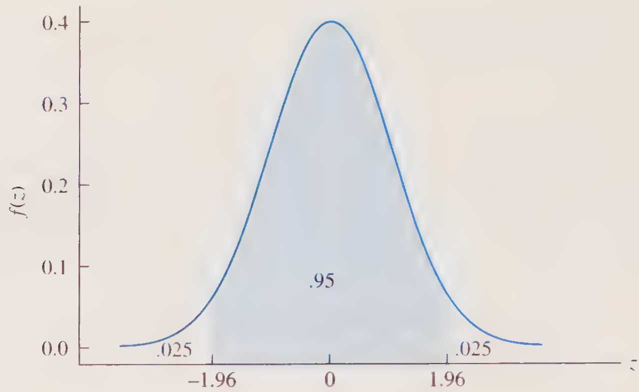

FIGURE 7.2
Partition of Z needed to obtain a 95% confidence interval for μ.

diagnosis of the disease is normally distributed with a mean of 13 months and a standard deviation of 3 months. A new treatment is being investigated which should prolong the average survival time without affecting variability. Let $X_1, X_2, X_3, \ldots, X_n$ denote a random sample from the distribution of X, the survival time under the new treatment. We are assuming that X is normally distributed with $\sigma^2 = 9$ and μ unknown. We want to find statistics L_1 and L_2 so that $P[L_1 \le \mu \le L_2] \doteq .95$. To do so, consider the partition of the standard normal curve shown in Fig. 7.2. It can be seen that

$$P[-1.96 \le Z \le 1.96] = .95$$

In this case $Z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$, and hence we may conclude that

$$P\left[-1.96 \le \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \le 1.96\right] = .95$$

To find L_1 and L_2, we algebraically isolate μ in the center of the preceding inequality as follows:

$$P[-1.96\sigma/\sqrt{n} \le \bar{X} - \mu \le 1.96\sigma/\sqrt{n}] = .95$$
$$P[-\bar{X} - 1.96\sigma/\sqrt{n} \le -\mu \le -\bar{X} + 1.96\sigma/\sqrt{n}] = .95$$
$$P[\bar{X} - 1.96\sigma/\sqrt{n} \le \mu \le \bar{X} + 1.96\sigma/\sqrt{n}] = .95$$

From this we see that the lower and upper bounds for a 95% confidence interval are

$$L_1 = \bar{X} - 1.96\sigma/\sqrt{n} \qquad L_2 = \bar{X} + 1.96\sigma/\sqrt{n}$$

These statistics have the property that in repeated sampling from the population, 95% of the numerical intervals generated are expected to contain μ; by chance, 5% will not. This idea is illustrated in Fig. 7.3.

Note that since we are assuming that σ^2 is known, the confidence bounds, $\bar{X} \pm 1.96\sigma/\sqrt{n}$, just derived are *statistics*. Given a particular set of observations on X, their numerical values can be determined easily.

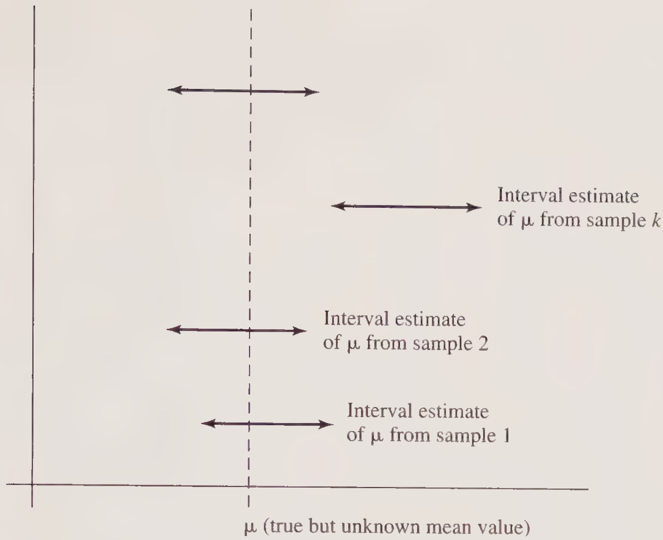

FIGURE 7.3
Of the intervals constructed by using $[L_1, L_2]$, 95% are expected to contain μ, the true but unknown population mean.

Example 7.4.2. In Example 7.1.1 fifty-six samples, each of size 30, were generated. Each sample was obtained by tossing of a single fair die 30 times. The sample mean was found for each sample. For the single die experiment, it can be shown that $E[X^2] = 15.167$, and hence the variance of X is given by

$$\sigma^2 = \mathrm{Var}(X) = E[X^2] - E[X]^2 = 15.167 - (3.5)^2 = 2.92$$

The standard deviation of X is $\sqrt{2.92} = 1.7088$. The standard error of the mean, $\sigma/\sqrt{30}$, has the value .3119. Thus the formula for a 95% confidence interval on μ in this case is

$$\bar{X} \pm 1.96(\sigma/\sqrt{n}) \qquad \text{or} \qquad \bar{X} \pm 1.96(.3119)$$

Each of the $\bar{x}$ values found in Example 7.1.1 is substituted into the above formula. We thus generate fifty six 95% confidence intervals on μ. Each is trying to trap the true mean value of 3.5. Some will succeed, and others will fail. Theoretically, 95% or about 53 will succeed and 5% or about 3 will fail. How well did the experiment work? Figure 7.4 gives the results of this exercise. The first column gives the value of $\bar{x}$, the second gives the lower 95% confidence limit, and the third shows the upper 95% confidence limit. The fourth column, result, is coded so that its value is 1 if the true mean of 3.5 falls between the lower and upper confidence limits. The last column states whether or not the interval in question actually trapped or missed the true mean. In this case, the results of our experiment agree extremely well with those predicted by theory even though X is not normal. You will soon see why.

Example 7.4.3. When the experiment of Example 7.4.1 is conducted, the following observations on X, the survival time under the new treatment, result:

	xbar	lower	upper	result	caught
1	3.43	2.81868	4.04132	1	trapped
2	3.33	2.71868	3.94132	1	trapped
3	3.60	2.98868	4.21132	1	trapped
4	2.97	2.35868	3.58132	1	trapped
5	3.50	2.88868	4.11132	1	trapped
6	4.20	3.58868	4.81132	0	missed
7	3.07	2.45868	3.68132	1	trapped
8	3.33	2.71868	3.94132	1	trapped
9	3.86	3.24868	4.47132	1	trapped
10	3.80	3.18868	4.41132	1	trapped
11	3.00	2.38868	3.61132	1	trapped
12	3.30	2.68868	3.91132	1	trapped
13	3.40	2.78868	4.01132	1	trapped
14	4.00	3.38868	4.61132	1	trapped
15	3.83	3.21868	4.44132	1	trapped
16	3.43	2.81868	4.04132	1	trapped
17	3.90	3.28868	4.51132	1	trapped
18	3.07	2.45868	3.68132	1	trapped
19	3.47	2.85868	4.08132	1	trapped
20	3.23	2.61868	3.84132	1	trapped
21	3.20	2.58868	3.81132	1	trapped
22	3.76	3.14868	4.37132	1	trapped
23	3.50	2.88868	4.11132	1	trapped
24	3.70	3.08868	4.31132	1	trapped
25	4.13	3.51868	4.74132	0	missed
26	3.97	3.35868	4.58132	1	trapped
27	3.42	2.80868	4.03132	1	trapped
28	3.57	2.95868	4.18132	1	trapped

	xbar	lower	upper	result	caught
29	3.47	2.85868	4.08132	1	trapped
30	3.80	3.18868	4.41132	1	trapped
31	3.53	2.91868	4.14132	1	trapped
32	3.20	2.58868	3.81132	1	trapped
33	3.13	2.51868	3.74132	1	trapped
34	3.63	3.01868	4.24132	1	trapped
35	3.33	2.71868	3.94132	1	trapped
36	3.33	2.71868	3.94132	1	trapped
37	3.56	2.94868	4.17132	1	trapped
38	3.47	2.85868	4.08132	1	trapped
39	4.33	3.71868	4.94132	0	missed
40	3.53	2.91868	4.14132	1	trapped
41	3.33	2.71868	3.94132	1	trapped
42	3.47	2.85868	4.08132	1	trapped
43	3.73	3.11868	4.34132	1	trapped
44	3.90	3.28868	4.51132	1	trapped
45	3.32	2.70868	3.93132	1	trapped
46	4.21	3.59868	4.82132	0	missed
47	3.63	3.01868	4.24132	1	trapped
48	3.67	3.05868	4.28132	1	trapped
49	3.53	2.91868	4.14132	1	trapped
50	3.43	2.81868	4.04132	1	trapped
51	3.40	2.78868	4.01132	1	trapped
52	3.53	2.91868	4.14132	1	trapped
53	3.63	3.01868	4.24132	1	trapped
54	3.42	2.80868	4.03132	1	trapped
55	3.67	3.05868	4.28132	1	trapped
56	3.57	2.95868	4.18132	1	trapped

FIGURE 7.4
Results of the experiment of Example 7.4.2.

8.0	13.6	13.2	13.6
12.5	14.2	14.9	14.5
13.4	8.6	11.5	16.0
14.2	19.0	17.9	17.0

Based on these data, $\hat{\mu} = \bar{x} = 13.88$ months. This point estimate is extended to a 95% confidence interval by evaluating the statistics L_1 and L_2. In particular,

$$\begin{aligned} L_1 = \bar{x} - 1.96\sigma/\sqrt{n} &= 13.88 - 1.96(3/\sqrt{16}) \\ &= 13.88 - 1.47 \\ &= 12.41 \text{ months} \\ L_2 = \bar{x} + 1.96\sigma/\sqrt{n} &= 13.88 + 1.47 \\ &= 15.35 \text{ months} \end{aligned}$$

Based on these data, the interval estimate for μ is [12.41, 15.35]. Does the true mean survival time for patients receiving the new treatment really lie between 12.41 and 15.35 months? Unfortunately, there is no way of knowing. The interval [12.41, 15.35]

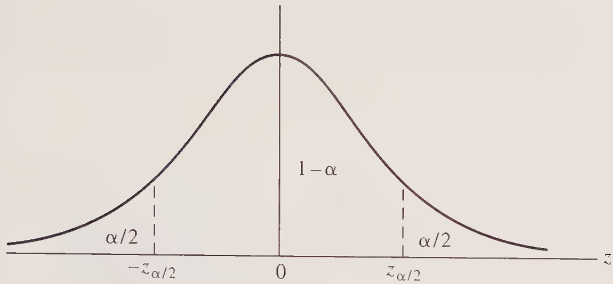

FIGURE 7.5
Partition of Z to obtain a $100(1 - \alpha)\%$ confidence interval for μ.

is a 95% confidence interval. This means that the procedure used is expected to trap μ 95% of the time. We hope that the interval obtained from our particular sample does so.

To obtain the general formula for a $100(1 - \alpha)\%$ confidence interval on the mean of a normal random variable whose variance is known, we need only to partition the standard normal curve as shown in Fig. 7.5. The algebraic argument of Example 7.4.1 goes through exactly as presented with the point $x_{.025} = 1.96$ being replaced by $z_{\alpha/2}$. This change results in the general formula given in Theorem 7.4.1.

> **Theorem 7.4.1 [$100(1 - \alpha)\%$ Confidence interval on μ when σ^2 is known].** Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample of size n from a normal distribution with mean μ and variance σ^2. A $100(1 - \alpha)\%$ confidence interval on μ is given by
>
> $$\bar{X} \pm z_{\alpha/2}\sigma/\sqrt{n}$$

Let us point out that the preceding confidence interval is very idealistic. It is usable only in settings in which the population standard deviation, σ, is known. In practice, this is seldom the case. In most real life problems both μ and σ^2 must be estimated from available data. When this occurs, the previous confidence interval is *not* appropriate. In Sec. 8.2 we shall show how to overcome this problem. Meanwhile, view this interval as a prototype for confidence intervals in general. It is useful as an aid for understanding how confidence intervals are derived and interpreted.

There are several things to notice concerning the preceding formula. First, every confidence interval on μ is centered at $\bar{x}$, the unbiased point estimate for μ. Second, the length of the confidence interval is dependent on three factors. These are the desired confidence, the amount of variability in X, and the sample size (n). The desired confidence determines the value of the z point used. The higher the confidence desired, the larger this value becomes. When a random variable displays a high degree of variability, it is hard to predict its behavior. Thus the larger σ becomes, the longer the confidence interval must become. Sample size works in reverse. With all other factors held constant, as n increases, the length of the confidence interval decreases. We can say that the length of a confidence interval on μ is directly proportional to σ and to the confidence desired and inversely proportional to the sample size.

Central Limit Theorem

There is one further point to be made. Theorem 7.4.1 does require that the base variable X be normal. If this condition is not satisfied, then the confidence bounds given can be used as long as the sample is not too small. Empirical studies have shown that for samples as small as 25, the above bounds are usually satisfactory even though approximate. This is due to a remarkable theorem, first formulated in the early nineteenth century by Laplace and Gauss. This theorem, known as the *Central Limit Theorem,* gives the distribution of $\bar{X}$ when sampling from a distribution that is not necessarily normal.

Theorem 7.4.2 (Central Limit Theorem). Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution with mean μ and variance σ^2. Then for large n, $\bar{X}$ is approximately normal with mean μ and variance σ^2/n. Furthermore, for large n, the random variable $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ is approximately standard normal.

Example 7.4.4 illustrates the Central Limit Theorem graphically.

Example 7.4.4. Consider a single die toss. We have tossed a single die 30 times and have repeated the experiment 56 times to obtain 56 $\bar{x}$ values. According to the Central Limit Theorem, a histogram of these data is expected to exhibit an approximate bell shape. The center of the bell is expected to lie close to 3.5, the true value of μ; the variance of the data should be close in value to .0973, the true value of σ^2/n; and the standard deviation of the data should approximate well the true value of the standard error of the mean, .3119. Figure 7.6 shows the histogram for the data of Example 7.1.1. Notice that the bell shape is not perfect. There is a slight right skew due to the fact that there were a few relatively large $\bar{x}$ values obtained via the experimentation. The mean for these data is 3.548, a little higher than the true mean of 3.5; the sample variance is .0911, a little smaller than the theoretical value of .0973; the estimated value of the standard error of the mean based on these data is .3019, a little smaller than the theoretical value of .3119. As the size of the sample upon which each $\bar{x}$ value is based increases, the histogram is expected to exhibit a more pronounced bell shape and the estimates for the mean, variance, and standard deviation of $\bar{X}$ are expected to agree more closely with those predicted by theory.

Please note the differences between the Central Limit Theorem and Theorem 7.3.4. The former does not require that sampling be from a normal distribution, whereas normality is assumed in the latter; the former claims that $\bar{X}$ will be approximately normally distributed for large sample sizes, whereas the latter claims that $\bar{X}$ will be exactly normally distributed regardless of the sample size involved.

The Central Limit Theorem is important to us for two reasons. First, it allows us to make inferences on the mean of a distribution based on relatively large samples

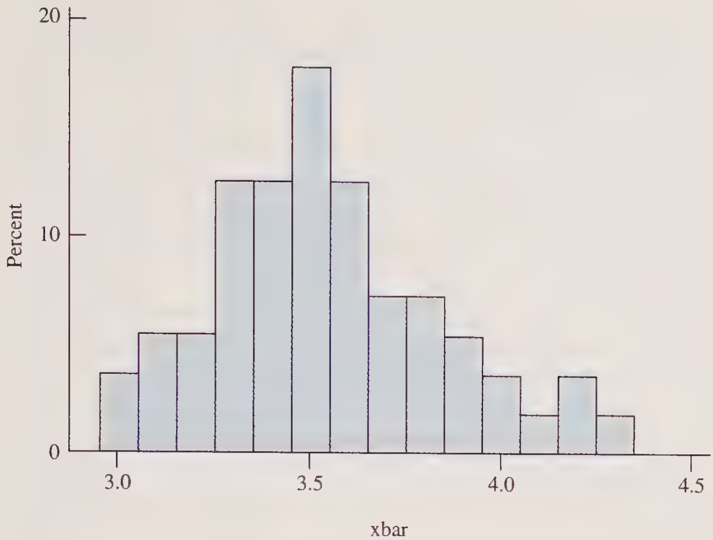

FIGURE 7.6 Histogram of the 56 $\bar{x}$ values given in Example 7.1.1.

without having to be overly concerned as to whether or not we are sampling from a normal distribution. Second, it allows us to justify analytically the normal approximations to the binomial distribution.

Example 7.4.5. (Normal Approximation to the Binomial Distribution) Let $X_1, X_2, \ldots, X_n$ be a random sample drawn from a point binomial distribution (see Exercise 45, Chap. 3). Recall that each of these random variables is binomial with parameters 1 and p. Each has mean p, variance $p(1-p)$, and moment generating function of the form $q + pe^t$. Let $X = \Sigma_{i=1}^{n} X_i$. Since $X_1, X_2, \ldots, X_n$ are independent, the moment generating function for X is given by

$$m_X(t) = \prod_{i=1}^{n} (q + pe^t) = (q + pe^t)^n$$

This is the moment generating function for a binomial random variable with parameters n and p. By the Central Limit Theorem $\bar{X} = (\Sigma_{i=1}^{n} X_i)/n = X/n$ is approximately normal with mean p and variance $p(1-p)/n$. Now consider the binomial random variable $n(X/n) = X$. Since X is a linear function of the approximately normal random variable X/n, we can apply Exercise 41 with $a_1 = n$ and $a_i = 0$, $i \neq 1$, to conclude that X is approximately normal with mean np and variance $[n^2p(1-p)]/n = np(1-p)$.

Exercises 49, 50, 55, 56, 58, 61, and 62 will give you practice in the application of the Central Limit Theorem.

CHAPTER SUMMARY

In this chapter we considered the ideas of point and interval estimation. We introduced three types of point estimators. These are unbiased estimators, method of moments estimators, and maximum likelihood estimators. Unbiased estimators are estimators whose mean value is equal to the parameter being estimated. We showed that $\bar{X}$ is unbiased for μ, that S^2 is unbiased for σ^2, but that S is not unbiased for σ. Method of moments estimators are derived by noting that the parameters that characterize a distribution are often functions of the kth moments of the distribution. Maximum likelihood estimators are found by choosing the value of the parameter θ that maximizes the likelihood function. In this way in some sense we pick out of all possible values of θ the one that is most likely to have produced the observed data.

In order to develop the idea of interval estimation, we introduced some theorems that help us to determine the distribution of a random variable. In particular, we noted that the moment generating function for a random variable is its "fingerprint." To determine its distribution we look at its moment generating function. This technique was used to verify the standardization theorem used in earlier chapters. It was also used to show that a linear function of independent normal random variables is normal, that a sum of independent chi-squared random variables is chi-squared, and that $\bar{X}$ is normally distributed when sampling from a normal distribution.

We introduced the general concept of a $100(1 - \alpha)\%$ confidence interval on a parameter θ. This is a random interval, an interval of the form $[L_1, L_2]$, where L_1 and L_2 are statistics with the property that a priori θ will be trapped between L_1 and L_2 with probability $1 - \alpha$. We used information just developed on the distribution of $\bar{X}$ to develop specific formulas for constructing a $100(1 - \alpha)\%$ confidence interval on the mean of a normal distribution. Finally, we considered the Central Limit Theorem. This theorem concerns the approximate distribution of $\bar{X}$ when sampling from a nonnormal distribution. It allows us to make inferences on the mean of any distribution when relatively large samples are available. It also allows us to justify some of the approximation techniques presented earlier in the text.

We introduced and defined important terms that you should know. These are:

Point estimator	Point estimate
Unbiased	Weighted mean
kth moments	Likelihood function
Confidence interval or interval estimator	Interval estimate
Methods of moments estimator	Sample standard error
Standard error of the mean	Maximum likelihood estimator
Central Limit Theorem	

EXERCISES

Section 7.1

1. Let $X_1, X_2, X_3, \ldots, X_{20}$ be a random sample from a distribution with mean 8 and variance 5. Find the mean and variance of $\bar{X}$.
2. Let $X_1, X_2, X_3, \ldots, X_{15}$ be a random sample from a Poisson distribution with parameter λs. Give an unbiased estimator for this parameter.

3. Let X denote the number of paint defects found in a square yard section of a car body painted by a robot. These data are obtained:

8	5	0	10
0	3	1	12
2	7	9	6

Assume that X has a Poisson distribution with parameter λs.
(*a*) Find an unbiased estimate for λs.
(*b*) Find an unbiased estimate for the average number of flaws per square yard.
(*c*) Find an unbiased estimate for the average number of flaws per square foot.

4. An interactive computer system is available at a large installation. Let X denote the number of requests for this system received per hour. Assume that X has a Poisson distribution with parameter λs. These data are obtained:

25	20	20
30	24	15
10	23	4

(*a*) Find an unbiased estimate for λs.
(*b*) Find an unbiased estimate for the average number of requests received per hour.
(*c*) Find an unbiased estimate for the average number of requests received per quarter hour.

5. Let X_1, X_2, X_3, X_4, X_5 be a random sample from a binomial distribution with $n = 10$ and p unknown.
(*a*) Show that $\overline{X}/10$ is an unbiased estimator for p.
(*b*) Estimate p based on these data: 3, 4, 4, 5, 6.

6. An experiment is conducted to study the effect of a power surge on data stored in a digital computer. A "word" is a sequence of 8 bits. Each bit is either "on" (activated) or "off" (not activated) at any given time. Twenty 8-bit words are stored, and a power surge is induced. Let X denote the number of bit reversals that result per word. Assume that X is binomially distributed with $n = 8$ and p, the probability of a bit reversal, unknown. These data result:

1	0	0	0
0	0	1	1
0	1	2	1
1	0	1	0
2	2	3	0

(*a*) Find an unbiased estimate for p.
(*b*) Based on the estimate for p just found, approximate the probability that in another 8-bit word a similar power surge will result in no bit reversals.
(*c*) A data line utilizes 64 bits. Based on the estimate for p just found, approximate the probability that at most one bit reversal will occur.

7. Stress tests are conducted on fiberglass rods used in communications networks. The random variable studied is X, the distance in inches from the anchored end of the rod to the crack location when the rod is subjected to extreme stress. Assume that X is uniformly distributed over the interval $(0, b)$. These data are obtained on 10 test rods:

10	7	11	12	8
8	9	10	9	13

(*a*) Find an unbiased estimate for the average distance from the anchored end of the rod to the crack.
(*b*) Find an unbiased estimate for the variance of X.
(*c*) Find an unbiased estimate for b.
(*d*) Find an estimate for σ, the standard deviation of X. Is this estimate unbiased?

8. Note that S is a statistic, and unless X is constant, its value will vary from sample to sample. Therefore Var $S > 0$. To show that S is not unbiased for σ, use proof by contradiction. That is, assume that $E[S] = \sigma$ and obtain a contradiction. *Hint:* Use Theorem 3.3.2.

9. (*Weighted means.*) Assume that one has k independent random samples of sizes $n_1, n_2, n_3, \ldots, n_k$ from the same distribution. These samples generate k unbiased estimators for the mean, namely, $\bar{X}_1, \bar{X}_2, \bar{X}_3, \ldots, \bar{X}_k$.
(*a*) Show that the arithmetic average of these estimators, $(\bar{X}_1 + \bar{X}_2 + \bar{X}_3 + \cdots + \bar{X}_k)/k$, is also unbiased for μ.
(*b*) Certain mineral elements required by plants are classed as macronutrients. Macronutrients are measured in terms of their percentage of the dry weight of the plant. Proportions of each element vary in different species and in the same species grown under differing conditions. One macronutrient is sulfur. In a study of winter cress, a member of the mustard family, these data, based on three independent random samples, are obtained:

$$\bar{x}_1 = .8 \qquad \bar{x}_2 = .95 \qquad \bar{x}_3 = .7$$
$$n_1 = 9 \qquad n_2 = 3 \qquad n_3 = 200$$

Use the result of part (*a*) to obtain an unbiased estimate for μ, the mean proportion of sulfur by dry weight in winter cress. By averaging the three values .8, .95, and .7 to obtain the estimate for μ, each sample is being given equal importance or "weight." Does this seem reasonable in this problem? Explain.
(*c*) To take sample sizes into account, a "weighted" mean is used. This estimator, $\hat{\mu}_W$, is given by

$$\hat{\mu}_W = \frac{n_1\bar{X}_1 + n_2\bar{X}_2 + \cdots + n_k\bar{X}_k}{n_1 + n_2 + \cdots + n_k}$$

Show that $\hat{\mu}_W$ is an unbiased estimator for μ.
(*d*) Use the data of part (*b*) to find the weighted estimate for the mean proportion of sulfur by dry weight in winter cress. Compare your answer to the estimate found in part (*b*).

10. Let X denote the number of heads obtained when a fair coin is tossed 4 times.
(*a*) What is $E[X]$ and Var X?
(*b*) Perform the experiment of tossing a fair coin 4 times and recording the number of heads obtained 10 times. You thus obtain a random sample of size 10 from a binomial distribution with $n = 4$ and $p = 1/2$.

(c) Based on your 10 observations, estimate the mean and variance of X. Compare your answers to those of your classmates. Do the observed values of $\bar{X}$ fluctuate about the theoretical mean of 2? Do the observed values of S^2 fluctuate about the theoretical variance of 1?

(d) Average the values of $\bar{X}$ that you have available. Is the average value close to 2? Average the values of S^2 that you have available. Is the average value of S^2 close to 1?

11. Let X denote the number of heads obtained when a fair coin is tossed 4 times. Perform this experiment 3 times, and record the value of X for each set of four tosses. In this way you obtain a single sample of size 3 from a binomial distribution with $n = 4$ and $p = 1/2$.

(a) Find the numerical value of $\bar{X}$ for your sample.

(b) Repeat the experiment 9 more times, recording the value of $\bar{X}$ each time.

(c) What is $E[\bar{X}]$? Average your 10 values of $\bar{X}$. Is the average value close to the theoretical mean of 2?

(d) What is Var $\bar{X}$? Find the value of S^2 for the 10 observations on $\bar{X}$. Does this value lie close to the theoretical value of 1/3?

12. Consider the experiment of rolling a pair of fair dice until a sum of 7 is obtained. Let X denote the number of trials needed to obtain a sum of 7.

(a) Notice that X is discrete. What is the distribution of X?

(b) What is the theoretical average value of X? That is, what is μ?

(c) What is the theoretical variance of X? That is, what is σ^2?

(d) Perform the experiment described 25 times, and thus obtain a sample of size $n = 25$ observations on X. Plot a stem-and-leaf diagram for your data. Does the distribution appear to be symmetric? Use your data to obtain unbiased estimates for μ and σ^2. Compare your answers to the true values of these parameters found in parts (b) and (c), respectively.

(e) Consider the random variable $\bar{X}$, the average number of trials needed to roll a sum of 7 based on 25 trials. What is $E[\bar{X}]$? What is Var $\bar{X}$?

(f) Pool the class observations on $\bar{X}$. Plot these values on a number line. Do they fluctuate about μ as expected? Find the average value of these observed $\bar{X}$ values. Is it close to μ as expected? Find the variance of the $\bar{X}$ values. Is this sample variance close in value to $\sigma^2/25$ as expected?

13. Ozone levels around Los Angeles have been measured as high as 220 parts per billion (ppb). Concentrations this high can cause the eyes to burn and are a hazard to both plant and animal life. These data were obtained on the ozone level in a forested area near Seattle, Washington (based on information found in "Twigs," *Americans Forests,* April 1990, p. 71):

160	176	160	180	167	164
165	163	162	168	173	179
170	196	185	163	162	163
172	162	167	161	169	178
161					

(a) Construct a double stem-and-leaf diagram for these data. Do these data appear to be skewed? If so, in which direction?

(*b*) Construct a boxplot for these data, and identify the potential outlier that is flagged by this technique. Assume that the point in question is a legitimate data point. In this case, do you believe that it is truly an outlier or probably simply a natural consequence of the distribution involved? Explain.

(*c*) Use these data to estimate the mean and variance of the ozone level in this area.

14. In this exercise you will show that the most logical estimator for σ^2, namely, $\sum_{i=1}^{n}(X_i - \bar{X})^2/n$, is a biased estimator for σ^2 and tends to underestimate the true variance. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution with mean μ and variance σ^2.

(*a*) Show that $\sum_{i=1}^{n}(X_i - \bar{X})^2/n = (n-1)S^2/n$.

(*b*) Verify that $E[\sum_{i=1}^{n}(X_i - \bar{X})^2/n] \neq \sigma^2$, thus showing that this estimator is not an unbiased estimator for σ^2. Argue that it tends to underestimate σ^2.

(*c*) Consider the theoretical setting described in Exercise 12. Based on samples of size $n = 25$, what is $E[S^2]$? What is $E[\sum_{i=1}^{n}(X_i - \bar{X})^2/25]$?

Section 7.2

15. Suppose that when the experiment described in Example 7.2.1 is conducted, these data result:

$$x_1 = 13 \qquad x_3 = 15 \qquad x_5 = 17$$
$$x_2 = 12 \qquad x_4 = 10$$

Use the method of moments to estimate p, the probability that a seedling will survive the first winter.

16. Let $X_1, X_2, \ldots, X_m$ be a random sample of size m from a binomial distribution with parameters n, assumed to be known, and p. Show that the method of moments estimator for p is $\hat{p} = \bar{X}/n$.

17. Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with parameter λs. Find the method of moments estimator for λs. Find the method of moments estimator for λ, the parameter underlying the Poisson process under observation.

18. In the study of traffic flow at an intersection a Poisson process with parameter λ is assumed. The basic unit of time assumed is 1 minute. These data are obtained on X, the number of vehicles arriving at the intersection during a 2-minute period:

2	5	1	3	2
3	0	8	2	5

Use these data to estimate λs, the average number of vehicles arriving during a two-minute period, and λ, the average number arriving per minute. (Use the results of Exercise 17.)

19. Use the information obtained in Example 7.2.2 to find an estimator for σ^2, the variance of a gamma random variable. Is the estimator obtained unbiased for σ^2? *Hint:* Express M_1 and M_2 as arithmetic averages, and compare your result to that of Theorem 6.3.1.

20. An acid solution made by mixing a powder compound with water is used to etch aluminum. The pH of the solution, X, will vary due to slight variations in the amount of water used, the potency of the dry compound, and the pH of the water itself. Assume that X is gamma distributed with α and β unknown. From these data, estimate α, β, μ, and σ^2 using the method of moments:

1.2	2.0	1.6	1.8	1.1
2.5	2.1	2.6	2.2	1.7
1.5	1.7	2.0	3.0	1.8

21. Assume that the data of Exercise 13 are drawn from an exponential distribution with parameter β. Find the method of moments estimate for β. Use this to find the method of moments estimate for σ^2. Is this the same estimate as that obtained in Exercise 13(*c*)?

22. Assume that the burn time of a sparkler as described in Exercise 32 of Chap. 6 follows a gamma distribution with parameters α and β. Use the data of Exercise 32 to

(*a*) find an unbiased estimate for the average burn time.
(*b*) find the method of moments estimate for the average burn time.
(*c*) find an unbiased estimate for the variance in burn time.
(*d*) find the method of moments estimate for the variance in burn time.

23. Find the method of moments estimator for the parameter p of a geometric distribution.

24. Use the results of Exercise 23 and your data from Exercise 12(*d*) to find the method of moments estimate for p, the probability of rolling a sum of 7 on a single roll of a pair of fair dice. Compare your estimate to the true probability of 1/6.

25. Using the method of moments estimator for p found in Exercise 23, find an estimator for σ^2 for the geometric distribution. Use this estimator to estimate σ^2 for your data from Exercise 12(*d*). Does this estimate differ from that found in Exercise 12(*d*)? If so, which estimate is closest to the true value of σ^2?

26. Let X be normal with mean μ and variance σ^2, both of which are unknown. Find the method of moments estimators for these parameters. Are the estimators obtained unbiased for their respective parameters? Explain.

27. Carbon dioxide is an odorless, colorless gas that constitutes about .035% by volume of the atmosphere. It affects the heat balance by acting as a one-way screen. It lets in the sun's heat to warm the oceans and the land but blocks some of the infrared heat that is radiated from the earth. This reflected heat is absorbed into the lower atmosphere, producing a greenhouse effect which causes the earth's surface to become warmer than it would be otherwise. Systematic measurements of CO_2 began in 1957 with Charles D. Keeling monitoring at Mauna Loa in Hawaii.

(*a*) Assume that these CO_2 readings (in ppm) are obtained:

319	338	337	339	328
325	340	331	341	336
330	330	321	327	337
320	343	350	322	334
326	349	341	338	332
339	335	338	333	334

Construct a stem-and-leaf diagram for these data using 31, 32, 32, 33, 33, 34, 34, 35 as stems. Graph leaves 0–4 on the first of each repeated stem and leaves 5–9 on the other. Is it reasonable to assume that the CO_2 level in the atmosphere is normally distributed? Explain.

(*b*) Estimate μ and σ^2 using the method of moments estimators.

(c) Find an unbiased estimate for σ^2.

28. Based on the data of Exercise 18, what is the maximum likelihood estimate for λ, the average number of vehicles arriving at an intersection per minute?

29. Based on the data of Exercise 27, what are the maximum likelihood estimates for the mean and variance of the atmospheric CO_2 level?

30. Let $X_1, X_2, X_3, \ldots, X_m$ be a random sample of size m from a binominal distribution with parameters n, assumed to be known, and p. Find the maximum likelihood estimator for p. Does it differ from the method of moments estimator found in Exercise 16?

31. Let W be an exponential random variable with parameter β unknown. Find the maximum likelihood estimator for β based on a sample of size n. Does it differ from the method of moments estimator?

32. A computer center employs consultants to answer users' questions. The center is open from 9 a.m. to 5 p.m. each weekday. Assume that calls arriving at the center constitute a Poisson process with unknown parameter λ calls per hour. To estimate λ, these observations were obtained on X, the number of calls arriving per hour:

8	6	12	15	12
4	9	7	20	10

(*a*) Find the maximum likelihood estimate for λ.

(*b*) Estimate the average time of arrival of the first call of the day. *Hint:* Consider Theorem 4.3.3.

33. A study of the noise level on takeoff of jets at a particular airport is studied. The random variable is X, the noise level in decibels of the jet as it passes over the first residential area adjacent to the airport. This random variable is assumed to have a gamma distribution with $\alpha = 2$ and β unknown.

(*a*) Find the maximum likelihood estimate for β based on a sample of size n.

(*b*) Use $\hat{\beta}$ to find an estimate for the mean value of X. Is this estimator unbiased for μ?

(*c*) Find the maximum likelihood estimate for β based on these data:

55	65	60	73	80
64	57	75	62	86
69	100	70	82	65
72	67	61	95	52

(*d*) Estimate the average decibel reading of these jets.

34. Computer terminals have a battery pack that maintains the configuration of the terminal. These packs must be replaced occasionally. Let X denote the life span in years of such a battery. Assume that X is exponentially distributed with unknown parameter β. Find the maximum likelihood estimate for β based on these data:

1.7	4.0	1.9	2.0	1.7
2.1	2.7	4.2	1.8	2.2
3.1	1.5	2.4	6.2	7.0
3.6	1.4	5.0	3.8	1.6

35. To estimate the proportion of defective microprocessor chips being produced by a particular maker, samples of five chips are selected at 10 randomly selected times during the day. These chips are inspected, and X, the number of defective chips in each batch of size 5, is recorded. Assume that X is binomially distributed with $n = 5$ and p unknown. Use these data to find the maximum likelihood estimate for p:

1	0	1	2	0
0	0	0	1	0

36. A new material is being tested for possible use in the brake shoes of automobiles. These shoes are expected to last for at least 75,000 miles. Fifteen sets of four of these experimental shoes are subjected to accelerated life testing. The random variable X, the number of shoes in each group of four that fail early, is assumed to be binomially distributed with $n = 4$ and p unknown. Find the maximum likelihood estimate for p based on these data:

1	0	1	0	2	1	0	1
0	1	0	0	0	1	0	

If an early failure rate in excess of 10% is unacceptable from a business point of view, would you have some doubts concerning the use of this new material? Explain.

Section 7.3

37. In each part the moment generating function for a random variable X is given. Identify the family to which the random variable belongs, and give the numerical values of pertinent distribution parameters.

(*a*) $m_X(t) = e^{2t+9t^2/2}$
(*b*) $m_X(t) = e^{8t^2}$
(*c*) $m_X(t) = .25e^t/(1 - .75e^t)$
(*d*) $m_X(t) = (.5 + .5e^t)^5$
(*e*) $m_X(t) = e^{6(e^t-1)}$
(*f*) $m_X(t) = (1 - 3t)^{-5}$
(*g*) $m_X(t) = (1 - 2t)^{-8}$
(*h*) $m_X(t) = (1 - .5t)^{-1}$

38. *(Distribution of a linear function of X.)*

(*a*) Let X be a random variable with moment generating function $m_X(t)$. Let $Y = \alpha + \beta X$. Show that $m_Y(t) = e^{\alpha t} m_X(\beta t)$. *Hint:* $m_Y(t) = E[e^{tY}] = E[e^{(\alpha + \beta X)t}]$.

(*b*) Let X be a normal random variable with mean 10 and variance 4. Find the moment generating function for the random variable $Y = 8 + 3X$. What is the distribution of Y?

39. *(Distribution of a sum of independent random variables.)* Let $X_1, X_2, X_3, \ldots, X_n$ be a collection of independent random variables with moment generating functions $m_{X_i}(t)$ ($i = 1, 2, 3, \ldots, n$, respectively). Let $a_0, a_1, a_2, \ldots, a_n$ be real numbers, and let

$$Y = a_0 + a_1X_1 + a_2X_2 + a_3X_3 + \cdots + a_nX_n$$

Show that the moment generating function for Y is given by

$$m_Y(t) = e^{a_0t}\prod_{i=1}^{n} m_{X_i}(a_it)$$

Note that this extends the result of Exercise 38(*a*) to more than one variable.

40. Let X_1 and X_2 be independent normal random variables with means 2 and 5 and variances 9 and 1, respectively. Let $Y = 3X_1 + 6X_2 - 8$. Use Exercise 39 to find the moment generating function for Y. What is the distribution of Y?

41. *(Distribution of a linear combination of independent normally distributed random variables.)* In this exercise you will prove that any linear combination of independent normally distributed random variables is also normally distributed. Let $X_1, X_2, X_3, \ldots, X_n$ be independent normal random variables with means μ_i and σ_i^2 ($i = 1, 2, 3, \ldots, n$, respectively). Let $a_0, a_1, a_2, \ldots, a_n$ be real numbers, and let

$$Y = a_0 + a_1X_1 + a_2X_2 + \cdots + a_nX_n$$

Use Exercise 39 to show that Y is normal with mean $\mu = a_0 + \sum_{i=1}^{n} a_i\mu_i$ and variance $\sigma^2 = \sum_{i=1}^{n} a_i^2\sigma_i^2$.

42. In this exercise you will prove that when sampling from a normal distribution, $\bar{X}$ is normally distributed. Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2. Use Exercise 41 to show that $\bar{X}$ is normal with mean μ and variance σ^2/n.

43. Let X_1 and X_2 be independent chi-squared random variables with 5 and 10 degrees of freedom, respectively. Show that $X_1 + X_2$ is a chi-squared random variable with 15 degrees of freedom.

44. *(Distribution of a sum of independent chi-squared random variables.)* In this exercise you will prove that the sum of a collection of independent chi-squared random variables also has a chi-squared distribution. Let $X_1, X_2, X_3, \ldots, X_n$ be independent chi-squared random variables with $\gamma_1, \gamma_2, \gamma_3, \ldots, \gamma_n$ degrees of freedom, respectively. Let

$$Y = X_1 + X_2 + X_3 + \cdots + X_n$$

Show that Y is a chi-squared random variable with γ degrees of freedom where $\gamma = \sum_{i=1}^{n} \gamma_i$.

45. *(Distribution of Z^2.)* It can be shown that the square of a standard normal random variable has a chi-squared distribution with $\gamma = 1$. That is, the random variable Z^2 follows a chi-squared distribution with 1 degree of freedom. Let $X_1, X_2, X_3, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance σ^2. Use Exercise 44 to show that

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

has a chi-squared distribution with n degrees of freedom.

46. Let X denote the time required to do a computation using an algorithm written in programming language A, and let Y denote the time required to do the same calculation using an algorithm written in programming language B. Assume that X is normally distributed with mean 10 seconds and standard deviation 3 seconds and that Y is normally distributed with mean 9 seconds and standard deviation 4 seconds.

(*a*) What is the distribution of the random variable $X - Y$?

(*b*) Find the probability that a given calculation will run faster using A than when using B.

Section 7.4

47. As heat is added to a material its temperature rises. The heat capacity is a quantitative statement of the increase in temperature for a specified addition of heat. These data are obtained on X, the measured heat capacity of liquid ethylene glycol at constant pressure and 80° C. Measurements are in calories per gram degree Celsius:

.645	.654	.640	.627	.626
.649	.629	.631	.643	.633
.646	.630	.634	.631	.651
.659	.638	.645	.655	.624
.658	.658	.658	.647	.665

Past experience indicates that $\sigma = .01$.

(*a*) Evaluate $\bar{X}$ for these data, thereby obtaining an unbiased point estimate for μ.

(*b*) Assume that X is normally distributed. Find a 95% confidence interval for μ.

(*c*) Would you expect a 90% confidence interval for μ based on these data to be longer or shorter than the interval of part (*b*)? Explain. Verify your answer by finding a 90% confidence interval on μ. *Hint:* Begin by sketching a curve similar to that shown in Fig. 7.3 with $1 - \alpha = .90$ and $\alpha/2 = .05$.

(*d*) Would you expect a 99% confidence interval for μ based on these data to be longer or shorter than the interval of part (*b*)? Explain. Verify your answer by finding a 99% confidence interval on μ.

48. The late manifestation of an injury following exposure to a sufficient dose of radiation is common. These data are obtained on the variable X, the time in days that elapses between the exposure to radiation and the appearance of peak erythema (skin redness):

16	12	14	16	13	9	15	7
20	19	11	14	9	13	11	3
8	21	16	16	12	16	14	20
7	14	18	14	18	13	11	16
18	16	11	13	14	16	15	15

(*a*) Even though the time at which the peak redness appears is recorded to the nearest day, time is actually a continuous random variable. Sketch a stem-and-leaf diagram for these data. Does the diagram lend support to the assumption that X is normally distributed?

(*b*) Evaluate $\bar{X}$ for these data.
(*c*) Assume that $\sigma = 4$ and find a 95% confidence interval on the mean time to the appearance of peak redness. Would you be surprised to hear a claim that $\mu = 17$ days? Explain, based on the confidence interval.

49. When fission occurs, many of the nuclear fragments formed have too many neutrons for stability. Some of these neutrons are expelled almost instantaneously. These observations are obtained on X, the number of neutrons released during fission of plutonium-239:

3	2	2	2	2	3	3	3
3	3	3	3	4	3	2	3
3	2	3	3	3	3	3	1
3	3	3	3	3	3	3	3
3	3	2	3	3	3	3	3

(*a*) Is X normally distributed? Explain.
(*b*) Estimate the mean number of neutrons expelled during fission of plutonium-239.
(*c*) Assume that $\sigma = .5$. Find a 99% confidence interval on μ. What theorem justifies the procedure you used to construct this interval?
(*d*) The reported value of μ is 3.0. Do these data refute this value? Explain.

50. *(Central Limit Theorem.)* Consider an infinite population with 25% of the elements having the value 1, 25% the value 2, 25% the value 3, and 25% the value 4. If X is the value of a randomly selected item, then X is a discrete random variable whose possible values are 1, 2, 3, and 4.
(*a*) Find the population mean μ and population variance σ^2 for the random variable X.
(*b*) List all 16 possible distinguishable samples of size 2, and for each calculate the value of the sample mean. Represent the value of the sample mean $\bar{X}$ using a probability histogram (use one bar for each of the possible values for $\bar{X}$). Note that although this is a very small sample, the distribution of $\bar{X}$ does not look like the population distribution and has the general shape of the normal distribution.
(*c*) Calculate the mean and variance of the distribution of $\bar{X}$ and show that, as expected, they are equal to μ and σ^2/n, respectively.

REVIEW EXERCISES

51. Consider the random variable X with density given by

$$f(x) = (1 + \theta)x^\theta \qquad 0 < x < 1 \qquad \theta > -1$$

(*a*) Show that $\int_0^1 f(x)\,dx = 1$ regardless of the specific value chosen for θ.
(*b*) Find $E[X]$.
(*c*) Find the method of moments estimator for θ.
(*d*) Find the method of moments estimate for θ based on these data:

.5 .3 .1 .1 .2

(*e*) Find the maximum likelihood estimator for θ.
(*f*) Find the maximum likelihood estimate for θ based on these data of part (*d*). Does this value agree with the method of moments estimate?

52. Consider the random variable X with density given by

$$f(x) = 1/\theta \qquad 0 < x < \theta$$

(*a*) Find $E[X]$.
(*b*) Find the method of moments estimator for θ. Is this estimator unbiased for θ?
(*c*) Find the method of moments estimate for θ based on these data:

1	.5	1.4	2.0	.25

53. Studies have shown that the random variable X, the processing time required to do a multiplication on a new 3-D computer, is normally distributed with mean μ and standard deviation 2 microseconds. A random sample of 16 observations is to be taken.
(*a*) What is the distribution of $\bar{X}$?
(*b*) These data are obtained:

42.65	45.15	39.32	44.44
41.63	41.54	41.59	45.68
46.50	41.35	44.37	40.27
43.87	43.79	43.28	40.70

Based on these data, find an unbiased estimate for μ.
(*c*) Find a 95% confidence interval for μ. Would you be surprised to read that the average time required to process a multiplication on this system is 42.2 microseconds? Explain, based on the confidence interval.

54. Let X denote the unit price of a 3.5-inch floppy diskette. These observations are obtained from a random sample of 10 suppliers:

\$3.83	3.54	3.44	3.89	3.65
3.70	3.59	3.37	4.04	3.93

(*a*) Find an unbiased estimate for the mean price of these diskettes.
(*b*) Find an unbiased estimate for the variance in the price of these diskettes.
(*c*) Find the sample standard deviation. Is this an unbiased estimate for σ?
(*d*) Assume that X is normally distributed. Find the maximum likelihood estimate for σ^2. Does this agree with your answer to (*b*)?

55. *(Central Limit Theorem.)* In an attempt to approximate the proportion p of improperly sealed packages produced on an assembly line, a random sample of 100 packages is selected and inspected. Let

$$X_i = \begin{cases} 1 & \text{if the } i\text{th package selected is improperly sealed} \\ 0 & \text{otherwise} \end{cases}$$

(*a*) What is the distribution of X_i?
(*b*) Based on the Central Limit Theorem, what is the approximate distribution of $\bar{X}$?

(*c*) When the experiment is conducted, we observe five improperly sealed packages. Find a point estimate for the proportion of improperly sealed packages being produced on this assembly line.

56. (*Central Limit Theorem.*) In a study of the size of various computer systems the random variable X, the number of files stored, is considered. Past experience indicates that $\sigma = 5$. These data are obtained:

7	8	4	5	9	9
4	12	8	1	8	7
3	13	2	1	17	7
12	5	6	2	1	13
14	10	2	4	9	11
3	5	12	6	10	7

(*a*) Find an unbiased estimate for μ, the mean number of files per system.
(*b*) Based on the Central Limit Theorem, what is the approximate distribution of $\bar{X}$?
(*c*) Find an approximate 98% confidence interval on μ.
(*d*) In describing the size of such systems, an executive states that the average number of files exceeds 10. Does this statement surprise you? Explain.

57. Let X denote the time expended by a terminal user in a computing session (time from log on to log off). Assume that X is normally distributed with $\mu_X = 15$ minutes and $\sigma_X = 4$ minutes. Let Y denote the time required to access the system. Assume that Y is normally distributed with mean 1.5 minutes and $\sigma_Y = .5$ minutes. Assume that X and Y are independent.
(*a*) Find $m_X(t)$ and $m_Y(t)$.
(*b*) The random variable $T = X + Y$ denotes the total time required by the user to run a job. Find the moment generating function for T.
(*c*) What is the distribution of T?
(*d*) Find the probability that the total time required exceeds 20 minutes.

58. Let $X_1, X_2, \ldots, X_{100}$ be a random sample of size 100 from a gamma distribution with $\alpha = 5$ and $\beta = 3$.
(*a*) Find the moment generating function for $Y = \Sigma_{i=1}^{100} X_i$.
(*b*) What is the distribution of Y?
(*c*) Find the moment generating function for $\bar{X} = Y/n$.
(*d*) What is the distribution of $\bar{X}$?
(*e*) Use the Central Limit Theorem to approximate the probability that $\bar{X}$ is at most 14.

59. Consider the random variable X with density given by

$$f(x) = (1/\theta^2)xe^{-x/\theta} \qquad x > 0 \qquad \theta > 0$$

(*a*) What is the distribution of X?
(*b*) What is $E[X]$?
(*c*) Find the method of moments estimator for θ.
(*d*) Find the maximum likelihood estimator for θ based on a random sample of size n. Does this estimator differ from that found in part (*c*)?
(*e*) Estimate θ based on these data:

3	5	2	3	4
1	4	3	3	3

(*f*) Are the estimators found in parts (*c*) and (*d*) unbiased estimators for θ?

60. Let X be normally distributed with mean 2 and variance 25.

(*a*) What is the distribution of the random variable $(X - 2)/5$?

(*b*) What is the distribution of the random variable $[(X - 2)/5]^2$?

(*c*) Let $X_1, X_2, X_3, \ldots, X_{10}$ represent a random sample from the distribution of X. What is the distribution of the random variable

$$\sum_{i=1}^{10} \left(\frac{X_i - 2}{5} \right)^2$$

61. *(Central Limit Theorem.)* In this problem you will use the Central Limit Theorem to justify the normal approximation to the Poisson distribution given earlier. That is, you will show that a Poisson random variable X with parameter λs can be approximated using a normal random variable with mean and variance λs. To do so, let $Y_1, Y_2, Y_3, \ldots, Y_n$ be a random sample of size n from a Poisson distribution with parameter $\lambda s/n$.

(*a*) Use moment generating function techniques to show that

$$X = \sum_{i=1}^{n} Y_i$$

has a Poisson distribution with parameter λs.

(*b*) Use the Central Limit Theorem to find the approximate distribution of $\bar{Y}$.

(*c*) Note that $n\bar{Y} = X$. Use this observation to argue that X is approximately normally distributed with mean λs and variance λs.

62. *(Central Limit Theorem.)* Consider the experiment of tossing a fair die once. Let X denote the number that occurs. Theoretically, X follows a discrete uniform distribution.

(*a*) Find the theoretical density, mean, and variance for X.

(*b*) Now consider an experiment in which the die is tossed 20 times and the results averaged. By the Central Limit Theorem, what is the theoretical mean and variance for the random variable $\bar{X}$?

(*c*) Perform the experiment of part (*b*) 25 times and record the value of $\bar{X}$ each time. (You will toss the die 500 times and obtain a data set that consists of 25 averages.) What shape should the stem-and-leaf diagram for these data assume? Explain. Construct a stem-and-leaf diagram for your data. Did the diagram take the shape that you expected?

(*d*) Approximately what value would you expect to obtain if you averaged the data of part (*c*)? Average your 25 observations on $\bar{X}$. Did the result come out as expected?

(*e*) Approximately what value would you expect to obtain if you found the sample variance for the data of part (*c*)? Explain. Find s^2 for your 25 observations on $\bar{X}$. Did the result come out as expected?

(*f*) If you were to construct 95% confidence intervals on μ based on each of the values of $\bar{X}$ found in part (*c*), approximately how many of them would you expect to contain the true value of μ? From your data, can you find an example of a confidence interval that does contain μ? of a confidence interval that does not contain μ?

63. Consider Example 6.2.1. Assume that X follows the exponential distribution with parameter β.

(*a*) Find the method of moments estimate for β.

(*b*) Find the maximum likelihood estimate for β.

(*c*) Are the answers to parts (*a*) and (*b*) the same?

(*d*) Use the estimated value of β *to* approximate the probability that a battery of this type will last at least 1000 hours.

64. Assume that a single fair die is tossed 30 times. Let X denote the number obtained per toss. Suppose that $\bar{x}$ assumes the value 2.83 for these 30 tosses.

(*a*) Find a 95% confidence interval for the mean value of X. Did the interval you constructed trap the true mean of 3.5?

(*b*) If we construct a 90% confidence interval on μ, will the interval have a chance of trapping μ? Explain based on what you learned in part (*a*).

(*c*) If we construct a 99% confidence interval on μ, will the interval have a chance of trapping the true mean? Explain.

(*d*) Construct a 99% confidence interval on μ. Did this interval trap the mean?