Chapter 7

PARAMETER ESTIMATION

7.1 INTRODUCTION

Let $X_1, \ldots, X_n$ be a random sample from a distribution F_θ that is specified up to a vector of unknown parameters θ. For instance, the sample could be from a Poisson distribution whose mean value is unknown; or it could be from a normal distribution having an unknown mean and variance. Whereas in probability theory it is usual to suppose that all of the parameters of a distribution are known, the opposite is true in statistics, where a central problem is to use the observed data to make inferences about the unknown parameters.

In Section 7.2, we present the *maximum likelihood* method for determining estimators of unknown parameters. The estimates so obtained are called *point estimates*, because they specify a single quantity as an estimate of θ. In Section 7.3, we consider the problem of obtaining *interval estimates*. In this case, rather than specifying a certain value as our estimate of θ, we specify an interval in which we estimate that θ lies. Additionally, we consider the question of how much *confidence* we can attach to such an interval estimate. We illustrate by showing how to obtain an interval estimate of the unknown mean of a normal distribution whose variance is specified. We then consider a variety of interval estimation problems. In Section 7.3.1, we present an interval estimate of the mean of a normal distribution whose variance is unknown. In Section 7.3.2, we obtain an interval estimate of the variance of a normal distribution. In Section 7.4, we determine an interval estimate for the difference of two normal means, both when their variances are assumed to be known and when they are assumed to be unknown (although in the latter case we suppose that the unknown variances are equal). In Sections 7.5 and the optional Section 7.6, we present interval estimates of the mean of a Bernoulli random variable and the mean of an exponential random variable.

In the optional Section 7.7, we return to the general problem of obtaining point estimates of unknown parameters and show how to evaluate an estimator by considering its mean square error. The bias of an estimator is discussed, and its relationship to the mean square error is explored.

In the optional Section 7.8, we consider the problem of determining an estimate of an unknown parameter when there is some prior information available. This is the *Bayesian* approach, which supposes that prior to observing the data, information about θ is always available to the decision maker, and that this information can be expressed in terms of a probability distribution on θ. In such a situation, we show how to compute the *Bayes estimator*, which is the estimator whose expected squared distance from θ is minimal.

7.2 MAXIMUM LIKELIHOOD ESTIMATORS

Any statistic used to estimate the value of an unknown parameter θ is called an *estimator* of θ. The observed value of the estimator is called the *estimate*. For instance, as we shall see, the usual estimator of the mean of a normal population, based on a sample $X_1, \ldots, X_n$ from that population, is the sample mean $\overline{X} = \sum_i X_i/n$. If a sample of size 3 yields the data $X_1 = 2, X_2 = 3, X_3 = 4$, then the estimate of the population mean, resulting from the estimator $\overline{X}$, is the value 3.

Suppose that the random variables $X_1, \ldots, X_n$, whose joint distribution is assumed given except for an unknown parameter θ, are to be observed. The problem of interest is to use the observed values to estimate θ. For example, the X_i's might be independent, exponential random variables each having the same unknown mean θ. In this case, the joint density function of the random variables would be given by

$$\begin{aligned}
&f(x_1, x_2, \ldots, x_n)\\
&\quad = f_{X_1}(x_1) f_{X_2}(x_2) \cdots f_{X_n}(x_n)\\
&\quad = \frac{1}{\theta} e^{-x_1/\theta} \frac{1}{\theta} e^{-x_2/\theta} \cdots \frac{1}{\theta} e^{-x_n/\theta}, \qquad 0 < x_i < \infty, i = 1, \ldots, n\\
&\quad = \frac{1}{\theta^n} \exp\left\{-\sum_1^n x_i/\theta\right\}, \qquad 0 < x_i < \infty, i = 1, \ldots, n
\end{aligned}$$

and the objective would be to estimate θ from the observed data $X_1, X_2, \ldots, X_n$.

A particular type of estimator, known as the *maximum likelihood* estimator, is widely used in statistics. It is obtained by reasoning as follows. Let $f(x_1, \ldots, x_n|\theta)$ denote the joint probability mass function of the random variables $X_1, X_2, \ldots, X_n$ when they are discrete, and let it be their joint probability density function when they are jointly continuous random variables. Because θ is assumed unknown, we also write f as a function of θ. Now since $f(x_1, \ldots, x_n|\theta)$ represents the likelihood that the values $x_1, x_2, \ldots, x_n$ will be observed when θ is the true value of the parameter, it would seem that a reasonable estimate of θ would be that value yielding the largest likelihood of the observed values. In other words, the maximum likelihood estimate $\hat{\theta}$ is defined to be that value of θ maximizing $f(x_1, \ldots, x_n|\theta)$ where $x_1, \ldots, x_n$ are the observed values. The function $f(x_1, \ldots, x_n|\theta)$ is often referred to as the *likelihood* function of θ.

In determining the maximizing value of θ, it is often useful to use the fact that $f(x_1, \ldots, x_n|\theta)$ and $\log[f(x_1, \ldots, x_n|\theta)]$ have their maximum at the same value of θ. Hence, we may also obtain $\hat{\theta}$ by maximizing $\log[f(x_1, \ldots, x_n|\theta)]$.

EXAMPLE 7.2a (Maximum Likelihood Estimator of a Bernoulli Parameter) Suppose that n independent trials, each of which is a success with probability p, are performed. What is the maximum likelihood estimator of p?

SOLUTION The data consist of the values of $X_1, \ldots, X_n$ where

$$X_i = \begin{cases} 1 & \text{if trial } i \text{ is a success} \\ 0 & \text{otherwise} \end{cases}$$

Now

$$P\{X_i = 1\} = p = 1 - P\{X_i = 0\}$$

which can be succinctly expressed as

$$P\{X_i = x\} = p^x(1-p)^{1-x}, \quad x = 0, 1$$

Hence, by the assumed independence of the trials, the likelihood (that is, the joint probability mass function) of the data is given by

$$\begin{aligned} f(x_1, \ldots, x_n|p) &= P\{X_1 = x_1, \ldots, X_n = x_n|p\} \\ &= p^{x_1}(1-p)^{1-x_1} \cdots p^{x_n}(1-p)^{1-x_n} \\ &= p^{\Sigma_1^n x_i}(1-p)^{n-\Sigma_1^n x_i}, \quad x_i = 0, 1, \quad i = 1, \ldots, n \end{aligned}$$

To determine the value of p that maximizes the likelihood, first take logs to obtain

$$\log f(x_1, \ldots, x_n|p) = \sum_1^n x_i \log p + \left(n - \sum_1^n x_i\right) \log(1-p)$$

Differentiation yields

$$\frac{d}{dp} \log f(x_1, \ldots, x_n|p) = \frac{\sum_1^n x_i}{p} - \frac{\left(n - \sum_1^n x_i\right)}{1-p}$$

Upon equating to zero and solving, we obtain that the maximum likelihood estimate $\hat{p}$ satisfies

$$\frac{\sum_{1}^{n} x_i}{\hat{p}} = \frac{n - \sum_{1}^{n} x_i}{1 - \hat{p}}$$

or

$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Hence, the maximum likelihood estimator of the unknown mean of a Bernoulli distribution is given by

$$d(X_1, \ldots, X_n) = \frac{\sum_{i=1}^{n} X_i}{n}$$

Since $\sum_{i=1}^{n} X_i$ is the number of successful trials, we see that the maximum likelihood estimator of p is equal to the proportion of the observed trials that result in successes. For an illustration, suppose that each RAM (random access memory) chip produced by a certain manufacturer is, independently, of acceptable quality with probability p. Then if out of a sample of 1,000 tested 921 are acceptable, it follows that the maximum likelihood estimate of p is .921. ■

EXAMPLE 7.2b Two proofreaders were given the same manuscript to read. If proofreader 1 found n_1 errors, and proofreader 2 found n_2 errors, with $n_{1,2}$ of these errors being found by both proofreaders, estimate N, the total number of errors that are in the manuscript.

SOLUTION Before we can estimate N we need to make some assumptions about the underlying probability model. So let us assume that the results of the proofreaders are independent, and that each error in the manuscript is independently found by proofreader i with probability p_i, $i = 1, 2$.

To estimate N, we will start by deriving an estimator of p_1. To do so, note that each of the n_2 errors found by reader 2 will, independently, be found by proofreader 1 with probability p_i. Because proofreader 1 found $n_{1,2}$ of those n_2 errors, a reasonable estimate of p_1 is given by

$$\hat{p}_1 = \frac{n_{1,2}}{n_2}$$

However, because proofreader 1 found n_1 of the N errors in the manuscript, it is reasonable to suppose that p_1 is also approximately equal to $\frac{n_1}{N}$. Equating this to $\hat{p}_1$ gives that

$$\frac{n_{1,2}}{n_2} \approx \frac{n_1}{N}$$

or

$$N \approx \frac{n_1 n_2}{n_{1,2}}$$

Because the preceding estimate is symmetric in n_1 and n_2, it follows that it is the same no matter which proofreader is designated as proofreader 1.

An interesting application of the preceding occurred when two teams of researchers recently announced that they had decoded the human genetic code sequence. As part of their work both teams estimated that the human genome consisted of approximately 33,000 genes. Because both teams independently arrived at the same number, many scientists found this number believable. However, most scientists were quite surprised by this relatively small number of genes; by comparison it is only about twice as many as a fruit fly has. However, a closer inspection of the findings indicated that the two groups only agreed on the existence of about 17,000 genes. (That is, 17,000 genes were found by both teams.) Thus, based on our preceding estimator, we would estimate that the actual number of genes, rather than being 33,000, is

$$\frac{n_1 n_2}{n_{1,2}} = \frac{33{,}000 \times 33{,}000}{17{,}000} \approx 64{,}000$$

(Because there is some controversy about whether some of genes claimed to be found are actually genes, 64,000 should probably be taken as an upper bound on the actual number of genes.)

The estimation approach used when there are two proofreaders does not work when there are m proofreaders, when $m > 2$. Because, if for each i, we let $\hat{p}_i$ be the fraction of the errors found by at least one of the other proofreaders j, $(j \neq i)$, that are also found by i, and then set that equal to $\frac{n_i}{N}$, then the estimate of N, namely $\frac{n_i}{\hat{p}_i}$, would differ for different values of i. Moreover, with this approach it is possible that we may have that $\hat{p}_i > \hat{p}_j$ even if proofreader i finds fewer errors than does proofreader j. For instance, for $m = 3$, suppose proofreaders 1 and 2 find exactly the same set of 10 errors whereas proofreader 3 finds 20 errors with only 1 of them in common with the set of errors found by the others. Then, because proofreader 1 (and 2) found 10 of the 29 errors found by at least one of the other proofreaders, $\hat{p}_i = 10/29$, $i = 1, 2$. On the other hand, because proofreader 3 only found 1 of the 10 errors found by the others, $\hat{p}_3 = 1/10$. Therefore, although proofreader 3 found twice the number of errors as did proofreader 1, the estimate of p_3 is less than that of p_1. To obtain more reasonable estimates, we could take the preceding values of $\hat{p}_i$, $i = 1, \ldots, m$, as preliminary estimates of the p_i. Now, let n_f be the number of errors that are found by at least one proofreader. Because n_f/N is the fraction of errors that are found by at least one proofreader, this should approximately equal $1 - \prod_{i=1}^{m}(1-p_i)$, the probability that an error is found by at least one proofreader. Therefore, we have

$$\frac{n_f}{N} \approx 1 - \prod_{i=1}^{m}(1-p_i)$$

suggesting that $N \approx \hat{N}$, where

$$\hat{N} = \frac{n_f}{1 - \prod_{i=1}^{m}(1-\hat{p}_i)} \tag{7.2.1}$$

With this estimate of N, we can then reset our estimates of the p_i by using

$$\hat{p}_i = \frac{n_i}{\hat{N}}, \quad i = 1, \ldots, m \tag{7.2.2}$$

We can then reestimate N by using the new value (Equation 7.2.1). (The estimation need not stop here; each time we obtain a new estimate $\hat{N}$ of N we can use Equation 7.2.2 to obtain new estimates of the p_i, which can then be used to obtain a new estimate of N, and so on.) ■

EXAMPLE 7.2c (Maximum Likelihood Estimator of a Poisson Parameter) Suppose $X_1, \ldots, X_n$ are independent Poisson random variables each having mean λ. Determine the maximum likelihood estimator of λ.

SOLUTION The likelihood function is given by

$$\begin{aligned} f(x_1, \ldots, x_n|\lambda) &= \frac{e^{-\lambda}\lambda^{x_1}}{x_1!} \cdots \frac{e^{-\lambda}\lambda^{x_n}}{x_n!} \\ &= \frac{e^{-n\lambda}\lambda^{\Sigma_1^n x_i}}{x_1! \ldots x_n!} \end{aligned}$$

Thus,

$$\log f(x_1, \ldots, x_n|\lambda) = -n\lambda + \sum_{1}^{n} x_i \log \lambda - \log c$$

where $c = \prod_{i=1}^{n} x_i!$ does not depend on λ. Differentiation yields

$$\frac{d}{d\lambda} \log f(x_1, \ldots, x_n|\lambda) = -n + \frac{\sum_{1}^{n} x_i}{\lambda}$$

By equating to zero, we obtain that the maximum likelihood estimate $\hat{\lambda}$ equals

$$\hat{\lambda} = \frac{\sum_{1}^{n} x_i}{n}$$

and so the maximum likelihood estimator is given by

$$d(X_1, \ldots, X_n) = \frac{\sum_{i=1}^{n} X_i}{n}$$

For example, suppose that the number of people who enter a certain retail establishment in any day is a Poisson random variable having an unknown mean λ, which must be estimated. If after 20 days a total of 857 people have entered the establishment, then the maximum likelihood estimate of λ is $857/20 = 42.85$. That is, we estimate that on average, 42.85 customers will enter the establishment on a given day. ■

EXAMPLE 7.2d The number of traffic accidents in Berkeley, California, in 10 randomly chosen nonrainy days in 1998 is as follows:

$$4, 0, 6, 5, 2, 1, 2, 0, 4, 3$$

Use these data to estimate the proportion of nonrainy days that had 2 or fewer accidents that year.

SOLUTION Since there are a large number of drivers, each of whom has a small probability of being involved in an accident in a given day, it seems reasonable to assume that the daily number of traffic accidents is a Poisson random variable. Since

$$\overline{X} = \frac{1}{10} \sum_{i=1}^{10} X_i = 2.7$$

it follows that the maximum likelihood estimate of the Poisson mean is 2.7. Since the long-run proportion of nonrainy days that have 2 or fewer accidents is equal to $P\{X \leq 2\}$, where X is the random number of accidents in a day, it follows that the desired estimate is

$$e^{-2.7}(1 + 2.7 + (2.7)^2/2) = .4936$$

That is, we estimate that a little less than half of the nonrainy days had 2 or fewer accidents. ■

EXAMPLE 7.2e (Maximum Likelihood Estimator in a Normal Population) Suppose $X_1, \ldots, X_n$ are independent, normal random variables each with unknown mean μ and unknown standard deviation σ. The joint density is given by

$$\begin{aligned} f(x_1, \ldots, x_n|\mu, \sigma) &= \prod_{i=1}^{n} \frac{1}{\sqrt{2\pi}\sigma} \exp\left[\frac{-(x_i-\mu)^2}{2\sigma^2}\right] \\ &= \left(\frac{1}{2\pi}\right)^{n/2} \frac{1}{\sigma^n} \exp\left[\frac{-\sum_1^n (x_i-\mu)^2}{2\sigma^2}\right] \end{aligned}$$

The logarithm of the likelihood is thus given by

$$\log f(x_1, \ldots, x_n|\mu, \sigma) = -\frac{n}{2}\log(2\pi) - n\log\sigma - \frac{\sum_1^n (x_i-\mu)^2}{2\sigma^2}$$

In order to find the value of μ and σ maximizing the foregoing, we compute

$$\begin{aligned} \frac{\partial}{\partial\mu} \log f(x_1, \ldots, x_n|\mu, \sigma) &= \frac{\sum_{i=1}^{n}(x_i-\mu)}{\sigma^2} \\ \frac{\partial}{\partial\sigma} \log f(x_1, \ldots, x_n|\mu, \sigma) &= -\frac{n}{\sigma} + \frac{\sum_1^n (x_i-\mu)^2}{\sigma^3} \end{aligned}$$

Equating these equations to zero yields that

$$\hat{\mu} = \sum_{i=1}^{n} x_i/n$$

and

$$\hat{\sigma} = \left[\sum_{i=1}^{n}(x_i-\hat{\mu})^2/n\right]^{1/2}$$

Hence, the maximum likelihood estimators of μ and σ are given, respectively, by

$$\overline{X} \quad \text{and} \quad \left[\sum_{i=1}^{n}(X_i - \overline{X})^2/n\right]^{1/2} \tag{7.2.3}$$

It should be noted that the maximum likelihood estimator of the standard deviation σ differs from the sample standard deviation

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

in that the denominator in Equation 7.2.3 is $\sqrt{n}$ rather than $\sqrt{n-1}$. However, for n of reasonable size, these two estimators of σ will be approximately equal. ■

EXAMPLE 7.2f *Kolmogorov's law of fragmentation* states that the size of an individual particle in a large collection of particles resulting from the fragmentation of a mineral compound will have an approximate lognormal distribution, where a random variable X is said to have a *lognormal* distribution if $\log(X)$ has a normal distribution. The law, which was first noted empirically and then later given a theoretical basis by Kolmogorov, has been applied to a variety of engineering studies. For instance, it has been used in the analysis of the size of randomly chosen gold particles from a collection of gold sand. A less obvious application of the law has been to a study of the stress release in earthquake fault zones (see Lomnitz, C., "Global Tectonics and Earthquake Risk," *Developments in Geotectonics*, Elsevier, Amsterdam, 1979).

Suppose that a sample of 10 grains of metallic sand taken from a large sand pile have respective lengths (in millimeters):

$$2.2,\ 3.4,\ 1.6,\ 0.8,\ 2.7,\ 3.3,\ 1.6,\ 2.8,\ 2.5,\ 1.9$$

Estimate the percentage of sand grains in the entire pile whose length is between 2 and 3 mm.

SOLUTION Taking the natural logarithm of these 10 data values, the following transformed data set results

$$.7885,\ 1.2238,\ .4700,\ -.2231,\ .9933,\ 1.1939,\ .4700,\ 1.0296,\ .9163,\ .6419$$

Because the sample mean and sample standard deviation of these data are

$$\bar{x} = .7504, \quad s = .4351$$

it follows that the logarithm of the length of a randomly chosen grain has a normal distribution with mean approximately equal to .7504 and with standard deviation approximately equal to .4351. Hence, if X is the length of the grain, then

$$
\begin{aligned}
P\{2 < X < 3\} &= P\{\log(2) < \log(X) < \log(3)\} \\
&= P\left\{\frac{\log(2) - .7504}{.4351} < \frac{\log(X) - .7504}{.4351} < \frac{\log(3) - .7504}{.4351}\right\} \\
&= P\left\{-.1316 < \frac{\log(X) - .7504}{.4351} < .8003\right\} \\
&\approx \Phi(.8003) - \Phi(-.1316) \\
&= .3405 \quad \blacksquare
\end{aligned}
$$

The lognormal distribution is often assumed in situations where the random variable under interest can be regarded as the product of a large number of independent and identically distributed random variables. For instance, it is commonly used in finance as the distribution of the price of a security at some future time. To see why this might be reasonable, suppose that the current price of the security is s and that we are interested in $S(t)$, the price of the security after an additional time t. For a large value n, let $t_i = it/n$, and consider $S(t_1), \ldots, S(t_n)$, the prices of the security at the times $t_1, \ldots, t_n$. Now, a common assumption in finance is that the ratios $S(t_i)/S(t_{i-1})$ are approximately independent and identically distributed. Consequently, if we let $X_i = S(t_i)/S(t_{i-1})$, then writing

$$
\begin{aligned}
S(t) = S(t_n) &= S(t_0) \cdot \frac{S(t_1)}{S(t_0)} \cdot \frac{S(t_2)}{S(t_1)} \cdots \frac{S(t_n)}{S(t_{n-1})} \\
&= s \prod_{i=1}^{n} X_i
\end{aligned}
$$

we obtain, upon taking logarithms, that

$$
\log(S(t)) = \log(s) + \sum_{i=1}^{n} \log(X_i)
$$

Thus, by the central limit theorem $\log(S(t))$ will approximately have a normal distribution.

The lognormal distribution has also been shown to be a good fit for such random variables as length of patient stays in hospitals, and vehicle travel times.

In all of the foregoing examples, the maximum likelihood estimator of the population mean turned out to be the sample mean $\overline{X}$. To show that this is not always the situation, consider the following example.

EXAMPLE 7.2g (Estimating the Mean of a Uniform Distribution) Suppose $X_1, \ldots, X_n$ constitute a sample from a uniform distribution on $(0, \theta)$, where θ is unknown. Their joint density is thus

$$f(x_1, x_2, \dots, x_n|\theta) = \begin{cases} \dfrac{1}{\theta^n} & 0 < x_i < \theta, \quad i = 1, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

This density is maximized by choosing θ as small as possible. Since θ must be at least as large as all of the observed values x_i, it follows that the smallest possible choice of θ is equal to $\max(x_1, x_2, \dots, x_n)$. Hence, the maximum likelihood estimator of θ is

$$\hat{\theta} = \max(X_1, X_2, \dots, X_n)$$

It easily follows from the foregoing that the maximum likelihood estimator of $\theta/2$, the mean of the distribution, is $\max(X_1, X_2, \dots, X_n)/2$. ■

*7.2.1 Estimating Life Distributions

Let X denote the age at death of a randomly chosen child born today. That is, $X = i$ if the newborn dies in its ith year, $i \geq 1$. To estimate the probability mass function of X, let λ_i denote the probability that a newborn who has survived his or her first $i - 1$ years dies in year i. That is,

$$\lambda_i = P\{X = i|X > i - 1\} = \frac{P\{X = i\}}{P\{X > i - 1\}}$$

Also, let

$$s_i = 1 - \lambda_i = \frac{P\{X > i\}}{P\{X > i - 1\}}$$

be the probability that a newborn who survives her first $i - 1$ years also survives year i. The quantity λ_i is called the *failure rate*, and s_i is called the *survival rate*, of an individual who is entering his or her ith year. Now,

$$\begin{aligned} s_1 s_2 \cdots s_i &= P\{X > 1\} \frac{P\{X > 2\}P\{X > 3\}}{P\{X > 1\}P\{X > 2\}} \cdots \frac{P\{X > i\}}{P\{X > i - 1\}} \\ &= P\{X > i\} \end{aligned}$$

Therefore,

$$P\{X = n\} = P\{X > n - 1\}\lambda_n = s_1 \cdots s_{n-1}(1 - s_n)$$

Consequently, we can estimate the probability mass function of X by estimating the quantities s_i, $i = 1, \dots, n$. The value s_i can be estimated by looking at all individuals in the

* Optional section.

population who reached age i 1 year ago, and then letting the estimate $\hat{s}_i$ be the fraction of them who are alive today. We would then use $\hat{s}_1\hat{s}_2\cdots\hat{s}_{n-1}(1-\hat{s}_n)$ as the estimate of $P\{X=n\}$. (Note that although we are using the most recent possible data to estimate the quantities s_i, our estimate of the probability mass function of the lifetime of a newborn assumes that the survival rate of the newborn when it reaches age i will be the same as last year's survival rate of someone of age i.)

The use of the survival rate to estimate a life distribution is also of importance in health studies with partial information. For instance, consider a study in which a new drug is given to a random sample of 12 lung cancer patients. Suppose that after some time we have the following data on the number of months of survival after starting the new drug:

$$4, 7^*, 9, 11^*, 12, 3, 14^*, 1, 8, 7, 5, 3^*$$

where x means that the patient died in month x after starting the drug treatment, and x^* means that the patient has taken the drug for x months and is still alive.

Let X equal the number of months of survival after beginning the drug treatment, and let

$$s_i = P\{X > i | X > i-1\} = \frac{P\{X > i\}}{P\{X > i-1\}}$$

To estimate s_i, the probability that a patient who has survived the first $i-1$ months will also survive month i, we should take the fraction of those patients who began their ith month of drug taking and survived the month. For instance, because 11 of the 12 patients survived month 1, $\hat{s}_1 = 11/12$. Because all 11 patients who began month 2 survived, $\hat{s}_2 = 11/11$. Because 10 of the 11 patients who began month 3 survived, $\hat{s}_3 = 10/11$. Because 8 of the 9 patients who began their fourth month of taking the drug (the 9 being all but the ones labelled 1, 3, and 3*) survived month 4, $\hat{s}_4 = 8/9$. Similar reasoning holds for the others, giving the following survival rate estimates:

$$\begin{aligned}
\hat{s}_1 &= 11/12\\
\hat{s}_2 &= 11/11\\
\hat{s}_3 &= 10/11\\
\hat{s}_4 &= 8/9\\
\hat{s}_5 &= 7/8\\
\hat{s}_6 &= 7/7\\
\hat{s}_7 &= 6/7\\
\hat{s}_8 &= 4/5\\
\hat{s}_9 &= 3/4\\
\hat{s}_{10} &= 3/3\\
\hat{s}_{11} &= 3/3
\end{aligned}$$

$$\hat{s}_{12} = 1/2$$
$$\hat{s}_{13} = 1/1$$
$$\hat{s}_{14} = 1/1$$

We can now use $\prod_{i=1}^{j} \hat{s}_i$ to estimate the probability that a drug taker survives at least j time periods, $j = 1, \ldots, 14$. For instance, our estimate of $P\{X > 6\}$ is 35/54.

7.3 INTERVAL ESTIMATES

Suppose that $X_1, \ldots, X_n$ is a sample from a normal population having unknown mean μ and known variance σ^2. It has been shown that $\overline{X} = \sum_{i=1}^{n} X_i/n$ is the maximum likelihood estimator for μ. However, we don't expect that the sample mean $\overline{X}$ will exactly equal μ, but rather that it will "be close." Hence, rather than a point estimate, it is sometimes more valuable to be able to specify an interval for which we have a certain degree of confidence that μ lies within. To obtain such an interval estimator, we make use of the probability distribution of the point estimator. Let us see how it works for the preceding situation.

In the foregoing, since the point estimator $\overline{X}$ is normal with mean μ and variance σ^2/n, it follows that

$$\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} = \sqrt{n}\frac{(\overline{X} - \mu)}{\sigma}$$

has a standard normal distribution. Therefore,

$$P\left\{-1.96 < \sqrt{n}\frac{(\overline{X} - \mu)}{\sigma} < 1.96\right\} = .95$$

or, equivalently,

$$P\left\{-1.96\frac{\sigma}{\sqrt{n}} < \overline{X} - \mu < 1.96\frac{\sigma}{\sqrt{n}}\right\} = .95$$

Multiplying through by -1 yields the equivalent statement

$$P\left\{-1.96\frac{\sigma}{\sqrt{n}} < \mu - \overline{X} < 1.96\frac{\sigma}{\sqrt{n}}\right\} = .95$$

or, equivalently,

$$P\left\{\overline{X} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{X} + 1.96\frac{\sigma}{\sqrt{n}}\right\} = .95$$

That is, 95 percent of the time the value of the sample average $\bar{X}$ will be such that the distance between it and the mean μ will be less than $1.96\,\sigma/\sqrt{n}$. If we now observe the sample and it turns out that $\overline{X} = \overline{x}$, then we say that "with 95 percent confidence"

$$\overline{x} - 1.96\frac{\sigma}{\sqrt{n}} < \mu < \overline{x} + 1.96\frac{\sigma}{\sqrt{n}} \tag{7.3.1}$$

That is, "with 95 percent confidence" we assert that the true mean lies within $1.96\,\sigma/\sqrt{n}$ of the observed sample mean. The interval

$$\left(\overline{x} - 1.96\frac{\sigma}{\sqrt{n}},\ \overline{x} + 1.96\frac{\sigma}{\sqrt{n}}\right)$$

is called a *95 percent confidence interval estimate* of μ.

EXAMPLE 7.3a Suppose that when a signal having value μ is transmitted from location A the value received at location B is normally distributed with mean μ and variance 4. That is, if μ is sent, then the value received is $\mu + N$ where N, representing noise, is normal with mean 0 and variance 4. To reduce error, suppose the same value is sent 9 times. If the successive values received are 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5, let us construct a 95 percent confidence interval for μ.

Since

$$\overline{x} = \frac{81}{9} = 9$$

It follows, under the assumption that the values received are independent, that a 95 percent confidence interval for μ is

$$\left(9 - 1.96\frac{\sigma}{3},\ 9 + 1.96\frac{\sigma}{3}\right) = (7.69, 10.31)$$

Hence, we are "95 percent confident" that the true message value lies between 7.69 and 10.31. ■

The interval in Equation 7.3.1 is called a *two-sided confidence interval.* Sometimes, however, we are interested in determining a value so that we can assert with, say, 95 percent confidence, that μ is at least as large as that value.

To determine such a value, note that if Z is a standard normal random variable then

$$P\{Z < 1.645\} = .95$$

As a result,

$$P\left\{\sqrt{n}\frac{(\overline{X} - \mu)}{\sigma} < 1.645\right\} = .95$$

or

$$P\left\{\overline{X} - 1.645\frac{\sigma}{\sqrt{n}} < \mu\right\} = .95$$

Thus, a 95 percent *one-sided upper confidence interval* for μ is

$$\left(\bar{x} - 1.645\frac{\sigma}{\sqrt{n}}, \infty\right)$$

where $\bar{x}$ is the observed value of the sample mean.

A *one-sided lower confidence interval* is obtained similarly; when the observed value of the sample mean is $\bar{x}$, then the 95 percent one-sided lower confidence interval for μ is

$$\left(-\infty,\ \bar{x} + 1.645\frac{\sigma}{\sqrt{n}}\right)$$

EXAMPLE 7.3b Determine the upper and lower 95 percent confidence interval estimates of μ in Example 7.3a.

SOLUTION Since

$$1.645\frac{\sigma}{\sqrt{n}} = \frac{3.29}{3} = 1.097$$

the 95 percent upper confidence interval is

$$(9 - 1.097, \infty) = (7.903, \infty)$$

and the 95 percent lower confidence interval is

$$(-\infty, 9 + 1.097) = (-\infty, 10.097) \quad ■$$

We can also obtain confidence intervals of any specified level of confidence. To do so, recall that z_α is such that

$$P\{Z > z_\alpha\} = \alpha$$

when Z is a standard normal random variable. But this implies (see Figure 7.1) that for any α

$$P\{-z_{\alpha/2} < Z < z_{\alpha/2}\} = 1 - \alpha$$

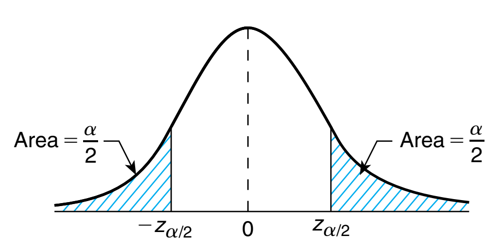


FIGURE 7.1 $P\{-z_{\alpha/2} < Z < z_{\alpha/2}\} = 1 - \alpha$.

As a result, we see that

$$P\left\{-z_{\alpha/2} < \sqrt{n}\frac{(\overline{X}-\mu)}{\sigma} < z_{\alpha/2}\right\} = 1-\alpha$$

or

$$P\left\{-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \overline{X}-\mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = 1-\alpha$$

or

$$P\left\{-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu-\overline{X} < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = 1-\alpha$$

That is,

$$P\left\{\overline{X}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \overline{X}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right\} = 1-\alpha$$

Hence, a $100(1-\alpha)$ percent two-sided confidence interval for μ is

$$\left(\overline{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\quad \overline{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

where $\overline{x}$ is the observed sample mean.

Similarly, knowing that $Z = \sqrt{n}\frac{(\overline{X}-\mu)}{\sigma}$ is a standard normal random variable, along with the identities

$$P\{Z > z_\alpha\} = \alpha$$

and

$$P\{Z < -z_\alpha\} = \alpha$$

results in one-sided confidence intervals of any desired level of confidence. Specifically, we obtain that

$$\left(\overline{x}-z_\alpha\frac{\sigma}{\sqrt{n}},\ \infty\right)$$

and

$$\left(-\infty,\ \overline{x}+z_\alpha\frac{\sigma}{\sqrt{n}}\right)$$

are, respectively, $100(1-\alpha)$ percent one-sided upper and $100(1-\alpha)$ percent one-sided lower confidence intervals for μ.

EXAMPLE 7.3c Use the data of Example 7.3a to obtain a 99 percent confidence interval estimate of μ, along with 99 percent one-sided upper and lower intervals.

SOLUTION Since $z_{.005} = 2.58$, and

$$2.58\frac{\alpha}{\sqrt{n}} = \frac{5.16}{3} = 1.72$$

it follows that a 99 percent confidence interval for μ is

$$9 \pm 1.72$$

That is, the 99 percent confidence interval estimate is (7.28, 10.72).

Also, since $z_{.01} = 2.33$, a 99 percent upper confidence interval is

$$(9 - 2.33(2/3),\ \infty) = (7.447,\ \infty)$$

Similarly, a 99 percent lower confidence interval is

$$(-\infty,\ 9 + 2.33(2/3)) = (-\infty,\ 10.553) \quad \blacksquare$$

Sometimes we are interested in a two-sided confidence interval of a certain level, say $1 - \alpha$, and the problem is to choose the sample size n so that the interval is of a certain size. For instance, suppose that we want to compute an interval of length .1 that we can assert, with 99 percent confidence, contains μ. How large need n be? To solve this, note that as $z_{.005} = 2.58$ it follows that the 99 percent confidence interval for μ from a sample of size n is

$$\left(\bar{x} - 2.58\frac{\sigma}{\sqrt{n}},\quad \bar{x} + 2.58\frac{\sigma}{\sqrt{n}}\right)$$

Hence, its length is

$$5.16\frac{\sigma}{\sqrt{n}}$$

Thus, to make the length of the interval equal to .1, we must choose

$$5.16\frac{\sigma}{\sqrt{n}} = .1$$

or

$$n = (51.6\,\sigma)^2$$

REMARK

The interpretation of "a $100(1 - \alpha)$ percent confidence interval" can be confusing. It should be noted that we are *not* asserting that the probability that $\mu \in (\bar{x} - 1.96\sigma/\sqrt{n},\ \bar{x} + 1.96\sigma/\sqrt{n})$ is .95, for there are no random variables involved in this assertion. What we are asserting is that the technique utilized to obtain this interval is such that 95 percent of the time that it is employed it will result in an interval in which μ lies. In other words, before the data are observed we can assert that with probability .95 the interval that will be obtained will contain μ, whereas after the data are obtained we can only assert that the resultant interval indeed contains μ "with confidence .95."

EXAMPLE 7.3d From past experience it is known that the weights of salmon grown at a commercial hatchery are normal with a mean that varies from season to season but with a standard deviation that remains fixed at 0.3 pounds. If we want to be 95 percent certain that our estimate of the present season's mean weight of a salmon is correct to within ± 0.1 pounds, how large a sample is needed?

SOLUTION A 95 percent confidence interval estimate for the unknown mean μ, based on a sample of size n, is

$$\mu \in \left(\bar{x} - 1.96\frac{\sigma}{\sqrt{n}},\ \bar{x} + 1.96\frac{\sigma}{\sqrt{n}} \right)$$

Because the estimate $\bar{x}$ is within $1.96(\sigma/\sqrt{n}) = .588/\sqrt{n}$ of any point in the interval, it follows that we can be 95 percent certain that $\bar{x}$ is within 0.1 of μ provided that

$$\frac{.588}{\sqrt{n}} \leq 0.1$$

That is, provided that

$$\sqrt{n} \geq 5.88$$

or

$$n \geq 34.57$$

That is, a sample size of 35 or larger will suffice. ■

7.3.1 Confidence Interval for a Normal Mean When the Variance Is Unknown

Suppose now that $X_1, \ldots, X_n$ is a sample from a normal distribution with unknown mean μ and unknown variance σ^2, and that we wish to construct a $100(1 - \alpha)$ percent confidence interval for μ. Since σ is unknown, we can no longer base our interval on the fact that $\sqrt{n}(\bar{X} - \mu)/\sigma$ is a standard normal random variable. However, by letting $S^2 = \sum_{i=1}^{n} (X_i - \bar{X})^2/(n - 1)$ denote the sample variance, then from Corollary 6.5.2 it follows that

$$\sqrt{n}\, \frac{(\bar{X} - \mu)}{S}$$

is a t-random variable with $n - 1$ degrees of freedom. Hence, from the symmetry of the t-density function (see Figure 7.2), we have that for any $\alpha \in (0, 1/2)$,

$$P\left\{ -t_{\alpha/2,n-1} < \sqrt{n}\, \frac{(\bar{X} - \mu)}{S} < t_{\alpha/2,n-1} \right\} = 1 - \alpha$$

or, equivalently,

$$P\{-t_{\alpha/2,n-1}\, \frac{S}{\sqrt{n}} < \bar{X} - \mu < t_{\alpha/2,n-1}\, \frac{S}{\sqrt{n}}\} = 1 - \alpha$$

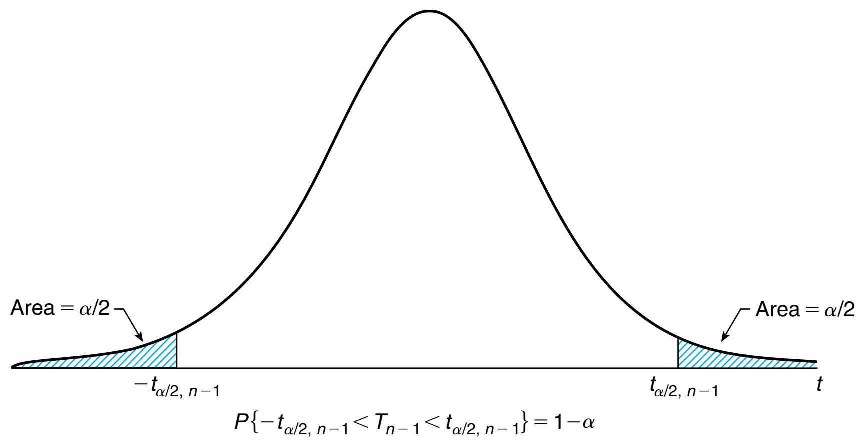


FIGURE 7.2 *t-density function.*

Multiplying all sides of the preceding by -1 and then adding $\bar{X}$ yields that

$$P\left\{\overline{X} - t_{\alpha/2,n-1}\frac{S}{\sqrt{n}} < \mu < \overline{X} + t_{\alpha/2,n-1}\frac{S}{\sqrt{n}}\right\} = 1 - \alpha$$

Thus, if it is observed that $\overline{X} = \overline{x}$ and $S = s$, then we can say that "with $100(1-\alpha)$ percent confidence"

$$\mu \in \left(\overline{x} - t_{\alpha/2,n-1}\frac{s}{\sqrt{n}},\ \overline{x} + t_{\alpha/2,n-1}\frac{s}{\sqrt{n}}\right)$$

EXAMPLE 7.3e Let us again consider Example 7.3a but let us now suppose that when the value μ is transmitted at location A then the value received at location B is normal with mean μ and variance σ^2 but with σ^2 being unknown. If 9 successive values are, as in Example 7.3a, 5, 8.5, 12, 15, 7, 9, 7.5, 6.5, 10.5, compute a 95 percent confidence interval for μ.

SOLUTION A simple calculation yields that

$$\overline{x} = 9$$

and

$$s^2 = \frac{\sum x_i^2 - 9(\overline{x})^2}{8} = 9.5$$

or

$$s = 3.082$$

Hence, as $t_{.025,8} = 2.306$, a 95 percent confidence interval for μ is

$$\left[9 - 2.306\frac{(3.082)}{3},\ 9 + 2.306\frac{(3.082)}{3}\right] = (6.63, 11.37)$$

a larger interval than obtained in Example 7.3a. The reason why the interval just obtained is larger than the one in Example 7.3a is twofold. The primary reason is that we have a larger estimated variance than in Example 7.3a. That is, in Example 7.3a we assumed that σ^2 was known to equal 4, whereas in this example we assumed it to be unknown and our estimate of it turned out to be 9.5, which resulted in a larger confidence interval. In fact, the confidence interval would have been larger than in Example 7.3a even if our estimate of σ^2 was again 4 because by having to estimate the variance we need to utilize the t-distribution, which has a greater variance and thus a larger spread than the standard normal (which can be used when σ^2 is assumed known). For instance, if it had turned out that $\bar{x} = 9$ and $s^2 = 4$, then our confidence interval would have been

$$(9 - 2.306 \cdot \tfrac{2}{3}, 9 + 2.306 \cdot \tfrac{2}{3}) = (7.46, 10.54)$$

which is larger than that obtained in Example 7.3a. ■

REMARKS

(a) The confidence interval for μ when σ is known is based on the fact that $\sqrt{n}(\overline{X} - \mu)/\sigma$ has a standard normal distribution. When σ is unknown, the foregoing approach is to estimate it by S and then use the fact that $\sqrt{n}(\overline{X} - \mu)/S$ has a t-distribution with $n - 1$ degrees of freedom.

(b) The length of a $100(1-2\alpha)$ percent confidence interval for μ is not always larger when the variance is unknown. For the length of such an interval is $2z_\alpha\sigma/\sqrt{n}$ when σ is known, whereas it is $2t_{\alpha,n-1}S/\sqrt{n}$ when σ is unknown; and it is certainly possible that the sample standard deviation S can turn out to be much smaller than σ. However, it can be shown that the mean length of the interval is longer when σ is unknown. That is, it can be shown that

$$t_{\alpha,n-1}E[S] \geq z_\alpha\sigma$$

Indeed, $E[S]$ is evaluated in Chapter 14 and it is shown, for instance, that

$$E[S] = \begin{cases} .94\,\sigma & \text{when } n = 5 \\ .97\,\sigma & \text{when } n = 9 \end{cases}$$

Since

$$z_{.025} = 1.96, \qquad t_{.025,4} = 2.78, \qquad t_{.025,8} = 2.31$$

the length of a 95 percent confidence interval from a sample of size 5 is $2 \times 1.96\sigma/\sqrt{5} = 1.75\sigma$ when σ is known, whereas its expected length is $2 \times 2.78 \times .94\sigma/\sqrt{5} = 2.34\sigma$ when σ is unknown — an increase of 33.7 percent. If the sample is of size 9, then the two values to compare are 1.31σ and 1.49σ — a gain of 13.7 percent. ■

A one-sided upper confidence interval can be obtained by noting that

$$P\left\{\sqrt{n}\frac{(\overline{X}-\mu)}{S} < t_{\alpha,n-1}\right\} = 1-\alpha$$

or

$$P\left\{\overline{X}-\mu < \frac{S}{\sqrt{n}}t_{\alpha,\,n-1}\right\} = 1-\alpha$$

or

$$P\left\{\mu > \overline{X} - \frac{S}{\sqrt{n}}t_{\alpha,\,n-1}\right\} = 1-\alpha$$

Hence, if it is observed that $\overline{X} = \overline{x}$, $S = s$, then we can assert "with $100(1-\alpha)$ percent confidence" that

$$\mu \in \left(\overline{x} - \frac{s}{\sqrt{n}}t_{\alpha,n-1},\ \infty\right)$$

Similarly, a $100(1-\alpha)$ lower confidence interval would be

$$\mu \in \left(-\infty,\ \overline{x} + \frac{s}{\sqrt{n}}t_{\alpha,\,n-1}\right)$$

Program 7.3.1 will compute both one- and two-sided confidence intervals for the mean of a normal distribution when the variance is unknown.

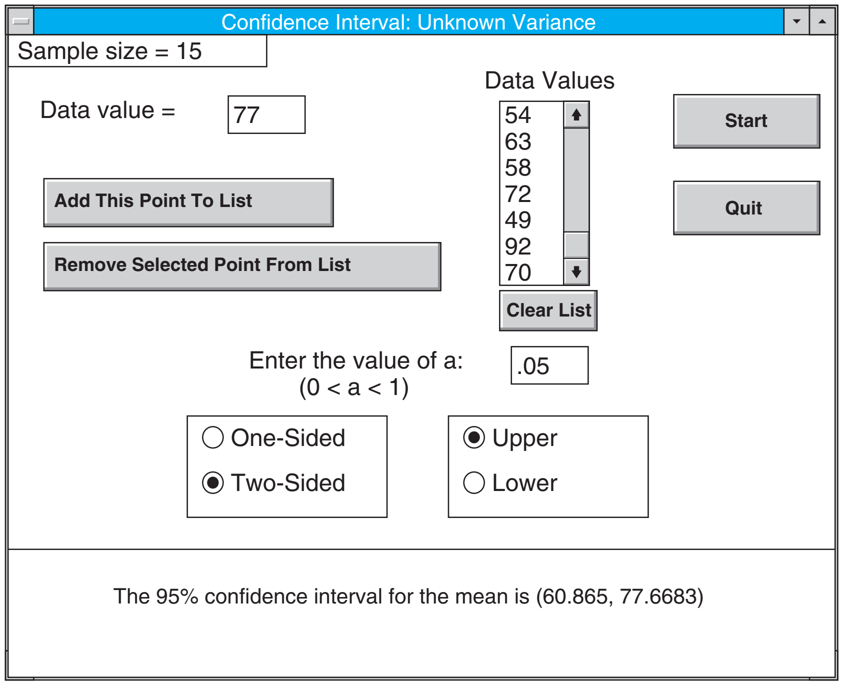
EXAMPLE 7.3f Determine a 95 percent confidence interval for the average resting pulse of the members of a health club if a random selection of 15 members of the club yielded the data 54, 63, 58, 72, 49, 92, 70, 73, 69, 104, 48, 66, 80, 64, 77. Also determine a 95 percent lower confidence interval for this mean.

SOLUTION We use Program 7.3.1 to obtain the solution (see Figure 7.3). ■

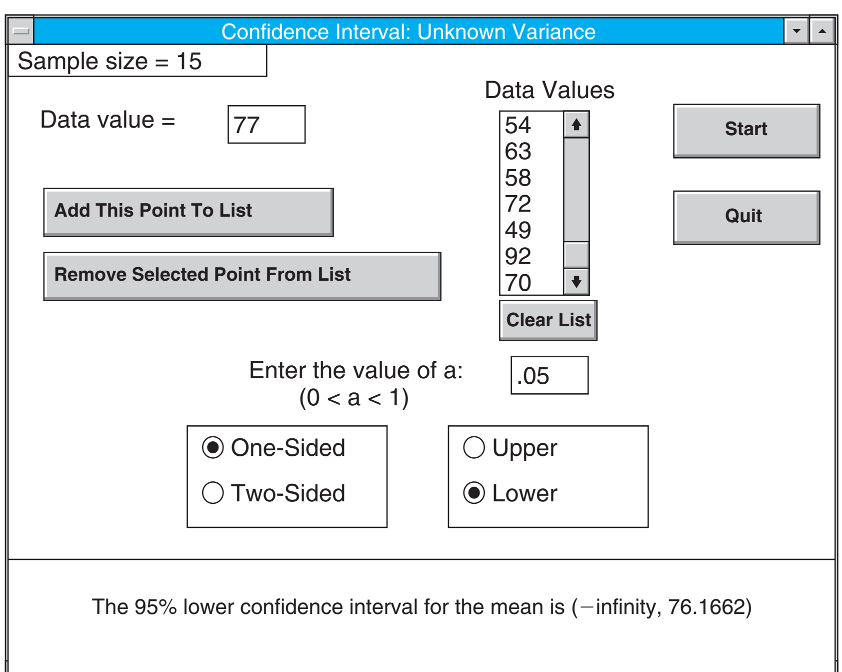
Our derivations of the $100(1-\alpha)$ percent confidence intervals for the population mean μ have assumed that the population distribution is normal. However, even when this is not the case, if the sample size is reasonably large then the intervals obtained will still be approximate $100(1-\alpha)$ percent confidence intervals for μ. This is true because, by the central limit theorem, $\sqrt{n}(\overline{X}-\mu)/\sigma$ will have approximately a normal distribution, and $\sqrt{n}(\overline{X}-\mu)/S$ will have approximately a t-distribution.

EXAMPLE 7.3g Simulation provides a powerful method for evaluating single and multi-dimensional integrals. For instance, let f be a function of an r-valued vector $(y_1, \ldots, y_r)$, and suppose that we want to estimate the quantity θ, defined by

$$\theta = \int_0^1 \int_0^1 \cdots \int_0^1 f(y_1, y_2, \ldots, y_r)\, dy_1 dy_2, \ldots, dy_r$$



(a)



(b)

FIGURE 7.3 *(a) Two-sided and (b) lower 95 percent confidence intervals for Example 7.3f.*

To accomplish this, note that if $U_1, U_2, \ldots, U_r$ are independent uniform random variables on (0, 1), then

$$\theta = E[f(U_1, U_2, \ldots, U_r)]$$

Now, the values of independent uniform (0, 1) random variables can be approximated on a computer (by so-called *pseudo random numbers*); if we generate a vector of r of them, and evaluate f at this vector, then the value obtained, call it X_1, will be a random variable with mean θ. If we now repeat this process, then we obtain another value, call it X_2, which will have the same distribution as X_1. Continuing on, we can generate a sequence $X_1, X_2, \ldots, X_n$ of independent and identically distributed random variables with mean θ; we then use their observed values to estimate θ. This method of approximating integrals is called *Monte Carlo simulation*.

For instance, suppose we wanted to estimate the one-dimensional integral

$$\theta = \int_0^1 \sqrt{1 - y^2}\, dy = E[\sqrt{1 - U^2}]$$

where U is a uniform (0, 1) random variable. To do so, let $U_1, \ldots, U_{100}$ be independent uniform (0, 1) random variables, and set

$$X_i = \sqrt{1 - U_i^2}, \qquad i = 1, \ldots, 100$$

In this way, we have generated a sample of 100 random variables having mean θ. Suppose that the computer generated values of $U_1, \ldots, U_{100}$, resulting in $X_1, \ldots, X_{100}$ having sample mean .786 and sample standard deviation .03. Consequently, since $t_{.025,99} = 1.985$, it follows that a 95 percent confidence interval for θ would be given by

$$.786 \pm 1.985(.003)$$

As a result, we could assert, with 95 percent confidence, that θ (which can be shown to equal $\pi/4$) is between .780 and .792. ■

7.3.2 Prediction Intervals

Suppose that $X_1, \ldots, X_n, X_{n+1}$ is a sample from a normal distribution with unknown mean μ and unknown variance σ^2. Suppose further that the values of $X_1, \ldots, X_n$ are to be observed and that we want to use them to predict the value of X_{n+1}. To begin, note that if the mean μ were known, then it would be the natural predictor for X_{n+1}. As it is not known, it seems natural to use its current estimator after observing $X_1, \ldots, X_n$, namely the average of these observed values, as the predicted value of X_{n+1}. That is, we should use the observed value of $\bar{X}_n = \sum_{i=1}^n X_i/n$, as the predicted value of X_{n+1}.

Suppose now that we want to determine an interval in which we predict, with a certain degree of confidence, that X_{n+1} will lie. To obtain such a prediction interval, note that as $\bar{X}_n$ is normal with mean μ and variance σ^2/n, and is independent of X_{n+1} which is normal with mean μ and variance σ^2, it follows that $X_{n+1} - \bar{X}_n$ is normal with mean 0 and variance $\sigma^2/n + \sigma^2$. Consequently,

$$\frac{X_{n+1} - \bar{X}_n}{\sigma\sqrt{1 + 1/n}} \qquad \text{is a standard normal random variable.}$$

Because this is independent of $S_n^2 = \sum_{i=1}^{n}(X_i - \bar{X}_n)^2/(n-1)$, it follows from the same argument used to establish Corollary 6.5.2, that replacing σ by its estimator S_n in the preceding expression will yield a t-random variable with $n-1$ degrees of freedom. That is,

$$\frac{X_{n+1} - \bar{X}_n}{S_n\sqrt{1 + 1/n}}$$

is a t-random variable with $n - 1$ degrees of freedom. Hence, for any $\alpha \in (0, 1/2)$,

$$P\{-t_{\alpha/2,n-1} < \frac{X_{n+1} - \bar{X}_n}{S_n\sqrt{1 + 1/n}} < t_{\alpha/2,n-1}\} = 1 - \alpha$$

which is equivalent to

$$P\{\bar{X}_n - t_{\alpha/2,n-1}\, S_n\sqrt{1 + 1/n} < X_{n+1} < \bar{X}_n + t_{\alpha/2,n-1}\, S_n\sqrt{1 + 1/n}\}$$

Hence, if the observed values of $\bar{X}_n$ and S_n are, respectively, $\bar{x}_n$ and s_n, then we can predict, with $100(1-\alpha)$ percent confidence, that X_{n+1} will lie between $\bar{x}_n - t_{\alpha/2,n-1}\, s_n\sqrt{1 + 1/n}$ and $\bar{x}_n + t_{\alpha/2,n-1}\, s_n\sqrt{1 + 1/n}$. That is, with $100(1 - \alpha)$ percent confidence, we can predict that

$$X_{n+1} \in \left(\bar{x}_n - t_{\alpha/2,n-1}\, s_n\sqrt{1 + 1/n},\ \ \bar{x}_n + t_{\alpha/2,n-1}\, s_n\sqrt{1 + 1/n}\right)$$

EXAMPLE 7.3h The following are the number of steps walked in each of the last 7 days

6822 5333 7420 7432 6252 7005 6752

Assuming that the daily number of steps can be thought of as being independent realizations from a normal distribution, give a prediction interval that, with 95 percent confidence, will contain the number of steps that will be walked tomorrow.

SOLUTION A simple calculation gives that the sample mean and sample variance of the 7 data values are

$$\bar{X}_7 = 6716.57 \qquad S_7 = 733.97$$

Because $t_{.025,6} = 2.447$, and $2.4447 \cdot 733.97\sqrt{1 + 1/7} = 1920.03$, we can predict, with 95 percent confidence, that tomorrow's number of steps will be between $6716.57 - 1920.03$ and $6716.57 + 1920.03$. That is, with 95 percent confidence, X_8 will lie in the interval $(4796.54,\ \ 8636.60)$.

7.3.3 Confidence Intervals for the Variance of a Normal Distribution

If $X_1, \ldots, X_n$ is a sample from a normal distribution having unknown parameters μ and σ^2, then we can construct a confidence interval for σ^2 by using the fact that

$$(n-1)\frac{S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Hence,

$$P\left\{\chi^2_{1-\alpha/2,n-1} \leq (n-1)\frac{S^2}{\sigma^2} \leq \chi^2_{\alpha/2,n-1}\right\} = 1-\alpha$$

or, equivalently,

$$P\left\{\frac{(n-1)S^2}{\chi^2_{\alpha/2,n-1}} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2,n-1}}\right\} = 1-\alpha$$

Hence when $S^2 = s^2$, a $100(1-\alpha)$ percent confidence interval for σ^2 is

$$\left(\frac{(n-1)s^2}{\chi^2_{\alpha/2,n-1}}, \frac{(n-1)s^2}{\chi^2_{1-\alpha/2,n-1}}\right)$$

EXAMPLE 7.3i A standardized procedure is expected to produce washers with very small deviation in their thicknesses. Suppose that 10 such washers were chosen and measured. If the thicknesses of these washers were, in inches,

.123	.133
.124	.125
.126	.128
.120	.124
.130	.126

what is a 90 percent confidence interval for the standard deviation of the thickness of a washer produced by this procedure?

SOLUTION A computation gives that

$$S^2 = 1.366 \times 10^{-5}$$

Because $\chi^2_{.05,9} = 16.917$ and $\chi^2_{.95,9} = 3.334$, and because

$$\frac{9 \times 1.366 \times 10^{-5}}{16.917} = 7.267 \times 10^{-6}, \qquad \frac{9 \times 1.366 \times 10^{-5}}{3.334} = 36.875 \times 10^{-6}$$

it follows that, with confidence .90,

$$\sigma^2 \in (7.267 \times 10^{-6}, \quad 36.875 \times 10^{-6})$$

Taking square roots yields that, with confidence .90,

$$\sigma \in (2.696 \times 10^{-3}, \quad 6.072 \times 10^{-3}) \quad \blacksquare$$

One-sided confidence intervals for σ^2 are obtained by similar reasoning and are presented in Table 7.1, which sums up the results of this section.

7.4 ESTIMATING THE DIFFERENCE IN MEANS OF TWO NORMAL POPULATIONS

Let $X_1, X_2, \ldots, X_n$ be a sample of size n from a normal population having mean μ_1 and variance σ_1^2 and let $Y_1, \ldots, Y_m$ be a sample of size m from a different normal population having mean μ_2 and variance σ_2^2 and suppose that the two samples are independent of each other. We are interested in estimating $\mu_1 - \mu_2$.

Since $\overline{X} = \sum_{i=1}^n X_i/n$ and $\overline{Y} = \sum_{i=1}^m Y_i/m$ are the maximum likelihood estimators of μ_1 and μ_2 it seems intuitive (and can be proven) that $\overline{X} - \overline{Y}$ is the maximum likelihood estimator of $\mu_1 - \mu_2$.

To obtain a confidence interval estimator, we need the distribution of $\overline{X} - \overline{Y}$. Because

$$\overline{X} \sim \mathcal{N}(\mu_1, \sigma_1^2/n)$$
$$\overline{Y} \sim \mathcal{N}(\mu_2, \sigma_2^2/m)$$

it follows from the fact that the sum of independent normal random variables is also normal, that

$$\overline{X} - \overline{Y} \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}\right)$$

TABLE 7.1 *100(1 −α) Percent Confidence Intervals*

$$X_1, \ldots, X_n \sim \mathcal{N}(\mu, \sigma^2)$$

$$\overline{X} = \sum_{i=1}^n X_i/n, \qquad S = \sqrt{\sum_{i=1}^n (X_i - \overline{X})^2/(n-1)}$$

Assumption	Parameter	Confidence Interval	Lower Interval	Upper Interval
σ^2 known	μ	$\overline{X} \pm z_{\alpha/2}\dfrac{\sigma}{\sqrt{n}}$	$\left(-\infty, \overline{X} + z_\alpha \dfrac{\sigma}{\sqrt{n}}\right)$	$\left(\overline{X} + z_\alpha \dfrac{\sigma}{\sqrt{n}}, \infty\right)$
σ^2 unknown	μ	$\overline{X} \pm t_{\alpha/2, n-1}\dfrac{S}{\sqrt{n}}$	$\left(-\infty, \overline{X} + t_{\alpha, n-1} \dfrac{S}{\sqrt{n}}\right)$	$\left(\overline{X} - t_{\alpha, n-1} \dfrac{S}{\sqrt{n}}, \infty\right)$
μ unknown	σ^2	$\left(\dfrac{(n-1)S^2}{\chi^2_{\alpha/2, n-1}}, \dfrac{(n-1)S^2}{\chi^2_{1-\alpha/2, n-1}}\right)$	$\left(0, \dfrac{(n-1)S^2}{\chi^2_{1-\alpha, n-1}}\right)$	$\left(\dfrac{(n-1)S^2}{\chi^2_{\alpha, n-1}}, \infty\right)$

Hence, assuming σ_1^2 and σ_2^2 are known, we have that

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}} \sim \mathcal{N}(0, 1) \tag{7.4.1}$$

and so

$$P\left\{-z_{\alpha/2} < \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}}} < z_{\alpha/2}\right\} = 1 - \alpha$$

or, equivalently,

$$P\left\{\overline{X} - \overline{Y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}} < \mu_1 - \mu_2 < \overline{X} - \overline{Y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right\} = 1 - \alpha$$

Hence, if $\overline{X}$ and $\overline{Y}$ are observed to equal $\bar{x}$ and $\bar{y}$, respectively, then a $100(1-\alpha)$ two-sided confidence interval estimate for $\mu_1 - \mu_2$ is

$$\mu_1 - \mu_2 \in \left(\bar{x} - \bar{y} - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}},\ \bar{x} - \bar{y} + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}\right)$$

One-sided confidence intervals for $\mu_1 - \mu_2$ are obtained in a similar fashion, and we leave it for the reader to verify that a $100(1 - \alpha)$ percent one-sided interval is given by

$$\mu_1 - \mu_2 \in \left(-\infty,\ \bar{x} - \bar{y} + z_\alpha\sqrt{\sigma_1^2/n + \sigma_2^2/m}\right)$$

Program 7.4.1 will compute both one- and two-sided confidence intervals for $\mu_1 - \mu_2$.

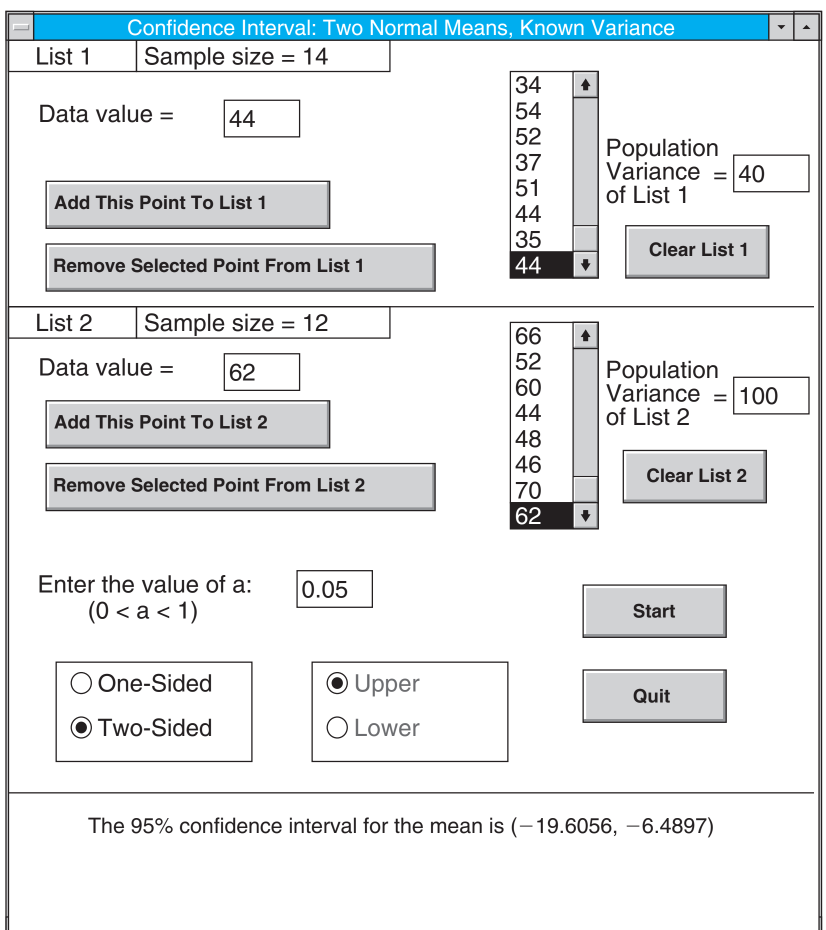
EXAMPLE 7.4a Two different types of electrical cable insulation have recently been tested to determine the voltage level at which failures tend to occur. When specimens were subjected to an increasing voltage stress in a laboratory experiment, failures for the two types of cable insulation occurred at the following voltages:

Type A		*Type B*	
36	54	52	60
44	52	64	44
41	37	38	48
53	51	68	46
38	44	66	70
36	35	52	62
34	44		

Suppose that it is known that the amount of voltage that cables having type A insulation can withstand is normally distributed with unknown mean μ_A and known variance $\sigma_A^2 = 40$, whereas the corresponding distribution for type B insulation is normal with unknown mean μ_B and known variance $\sigma_B^2 = 100$. Determine a 95 percent confidence interval for $\mu_A - \mu_B$. Determine a value that we can assert, with 95 percent confidence, exceeds $\mu_A - \mu_B$.

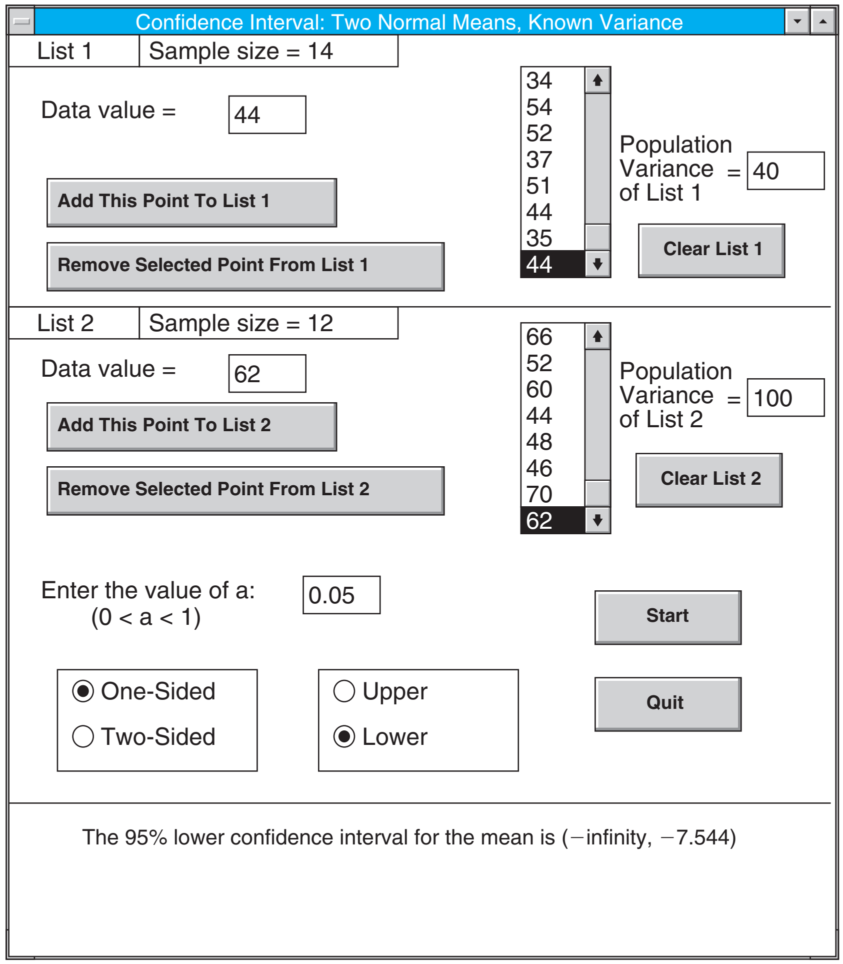
SOLUTION We run Program 7.4.1 to obtain the solution (see Figure 7.4). ■

Let us suppose now that we again desire an interval estimator of $\mu_1 - \mu_2$ but that the population variances σ_1^2 and σ_2^2 are unknown. In this case, it is natural to try to replace σ_1^2 and σ_2^2 in Equation 7.4.1 by the sample variances



(a)

FIGURE 7.4 *(a) Two-sided and (b) lower 95 percent confidence intervals for Example 7.4a.*



(b)

FIGURE 7.4 *(continued)*

$$S_1^2 = \sum_{i=1}^{n} \frac{(X_i - \overline{X})^2}{n-1}$$

$$S_2^2 = \sum_{i=1}^{m} \frac{(Y_i - \overline{Y})^2}{m-1}$$

That is, it is natural to base our interval estimate on something like

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{S_1^2/n + S_2^2/m}}$$

However, to utilize the foregoing to obtain a confidence interval, we need its distribution and it must not depend on any of the unknown parameters σ_1^2 and σ_2^2. Unfortunately, this distribution is both complicated and does indeed depend on the unknown parameters σ_1^2 and σ_2^2. In fact, it is only in the special case when $\sigma_1^2 = \sigma_2^2$ that we will be able to obtain an interval estimator. So let us suppose that the population variances, though unknown, are equal and let σ^2 denote their common value. Now, from Theorem 6.5.1 it follows that

$$(n-1)\frac{S_1^2}{\sigma^2} \sim \chi^2_{n-1}$$

and

$$(m-1)\frac{S_2^2}{\sigma^2} \sim \chi^2_{m-1}$$

Also, because the samples are independent, it follows that these two chi-square random variables are independent. Hence, from the additive property of chi-square random variables, which states that the sum of independent chi-square random variables is also chi-square with a degree of freedom equal to the sum of their degrees of freedom, it follows that

$$(n-1)\frac{S_1^2}{\sigma^2} + (m-1)\frac{S_2^2}{\sigma^2} \sim \chi^2_{n+m-2} \tag{7.4.2}$$

Also, since

$$\overline{X} - \overline{Y} \sim \mathcal{N}\left(\mu_1 - \mu_2, \frac{\sigma^2}{n} + \frac{\sigma^2}{m}\right)$$

we see that

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\dfrac{\sigma^2}{n} + \dfrac{\sigma^2}{m}}} \sim \mathcal{N}(0, 1) \tag{7.4.3}$$

Now it follows from the fundamental result that in normal sampling $\overline{X}$ and S^2 are independent (Theorem 6.5.1), that $\overline{X}_1, S_1^2, \overline{X}_2, S_2^2$ are independent random variables. Hence, using the definition of a t-random variable (as the ratio of two independent random variables, the numerator being a standard normal and the denominator being the square root of a chi-square random variable divided by its degree of freedom parameter), it follows from Equations 7.4.2 and 7.4.3 that if we let

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$$

then

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{\sigma^2(1/n + 1/m)}} \div \sqrt{S_p^2/\sigma^2} = \frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{S_p^2(1/n + 1/m)}}$$

has a t-distribution with $n+m-2$ degrees of freedom. Consequently,

$$P\left\{-t_{\alpha/2,n+m-2} \leq \frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{S_p\sqrt{1/n+1/m}} \leq t_{\alpha/2,n+m-2}\right\} = 1-\alpha$$

Therefore, when the data result in the values $\overline{X} = \overline{x}, \overline{Y} = \overline{y}, S_p = s_p$, we obtain the following $100(1-\alpha)$ percent confidence interval for $\mu_1 - \mu_2$:

$$\left(\overline{x}-\overline{y}-t_{\alpha/2,n+m-2}\, s_p\sqrt{1/n+1/m},\quad \overline{x}-\overline{y}+t_{\alpha/2,n+m-2}\, s_p\sqrt{1/n+1/m}\right) \tag{7.4.4}$$

One-sided confidence intervals are similarly obtained.

Program 7.4.2 can be used to obtain both one- and two-sided confidence intervals for the difference in means in two normal populations having unknown but equal variances.

EXAMPLE 7.4b There are two different techniques a given manufacturer can employ to produce batteries. A random selection of 12 batteries produced by technique I and of 14 produced by technique II resulted in the following capacities (in ampere hours):

Technique I		*Technique II*	
140	132	144	134
136	142	132	130
138	150	136	146
150	154	140	128
152	136	128	131
144	142	150	137
		130	135

Determine a 90 percent level two-sided confidence interval for the difference in means, assuming a common variance. Also determine a 95 percent upper confidence interval for $\mu_{\mathrm{I}} - \mu_{\mathrm{II}}$.

SOLUTION We run Program 7.4.2 to obtain the solution (see Figure 7.5). ■

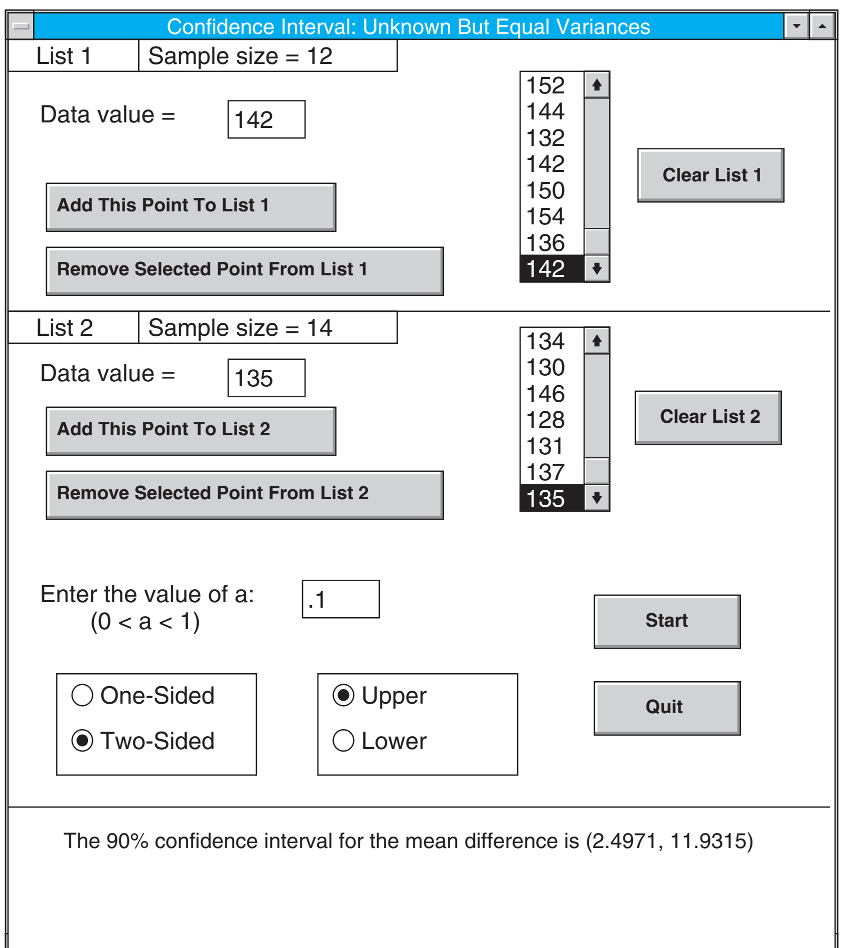
REMARK

The confidence interval given by Equation 7.4.4 was obtained under the assumption that the population variances are equal; with σ^2 as their common value, it follows that

$$\frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sqrt{\sigma^2/n+\sigma^2/m}} = \frac{\overline{X}-\overline{Y}-(\mu_1-\mu_2)}{\sigma\sqrt{1/n+1/m}}$$

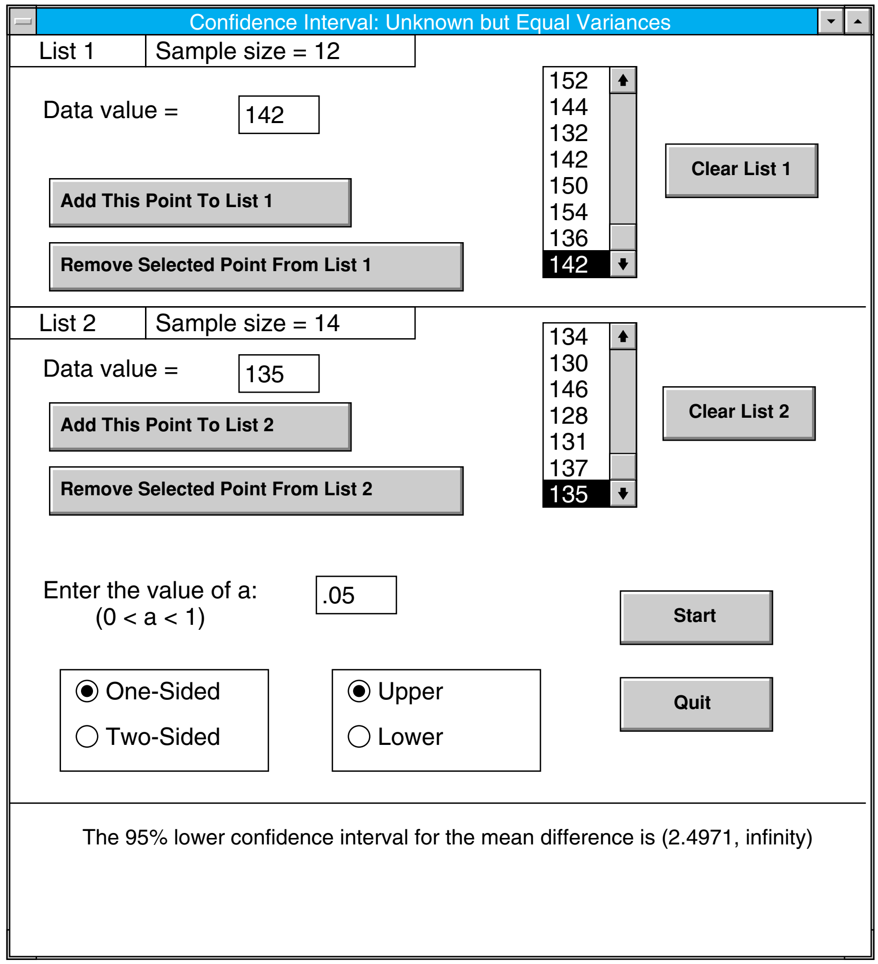
has a standard normal distribution. However, since σ^2 is unknown this result cannot be immediately applied to obtain a confidence interval; σ^2 must first be estimated. To do so, note that both sample variances are estimators of σ^2; moreover, since S_1^2 has $n-1$ degrees of freedom and S_2^2 has $m-1$, the appropriate estimator is to take a weighted average of the two sample variances, with the weights proportional to these degrees of freedom. That is, the estimator of σ^2 is the *pooled estimator*

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}$$



(a)

FIGURE 7.5 *(a) Two-sided and (b) upper 90 percent confidence intervals for Example 7.4b.*



(b)

FIGURE 7.5 *(continued)*

and the confidence interval is then based on the statistic

$$\frac{\overline{X} - \overline{Y} - (\mu_1 - \mu_2)}{\sqrt{S_p^2}\sqrt{1/n + 1/m}}$$

which, by our previous analysis, has a t-distribution with $n + m - 2$ degrees of freedom. The results of this section are summarized in Table 7.2.

TABLE 7.2 *$100(1-\sigma)$ Percent Confidence Intervals for $\mu_1 - \mu_2$*

$$X_1, \ldots, X_n \sim \mathcal{N}(\mu_1, \sigma_1^2)$$
$$Y_1, \ldots, Y_m \sim \mathcal{N}(\mu_2, \sigma_2^2)$$
$$\overline{X} = \sum_{i=1}^{n} X_i/n, \qquad S_1^2 = \sum_{i=1}^{n}(X_i - \overline{X})^2/(n-1)$$
$$\overline{Y} = \sum_{i=1}^{m} Y_i/n, \qquad S_2^2 = \sum_{i=1}^{m}(Y_i - \overline{Y})^2/(m-1)$$

Assumption	**Confidence Interval**
σ_1, σ_2 known	$\overline{X} - \overline{Y} \pm z_{\alpha/2}\sqrt{\sigma_1^2/n + \sigma_2^2/m}$
σ_1, σ_2 unknown but equal	$\overline{X} - \overline{Y} \pm t_{\alpha/2,\, n+m-2}\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}}$

Assumption	**Lower Confidence Interval**
σ_1, σ_2 known	$(-\infty, \overline{X} - \overline{Y} + z_{\alpha}\sqrt{\sigma_1^2/n + \sigma_2^2/m})$
σ_1, σ_2 unknown but equal	$\left(-\infty, \overline{X} - \overline{Y} + t_{\alpha,\, n+m-2}\sqrt{\left(\frac{1}{n} + \frac{1}{m}\right)\frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2}}\right)$

Note: Upper confidence intervals for $\mu_1 - \mu_2$ are obtained from lower confidence intervals for $\mu_2 - \mu_1$.

7.5 APPROXIMATE CONFIDENCE INTERVAL FOR THE MEAN OF A BERNOULLI RANDOM VARIABLE

Consider a population of items, each of which independently meets certain standards with some unknown probability p. If n of these items are tested to determine whether they meet the standards, how can we use the resulting data to obtain a confidence interval for p?

If we let X denote the number of the n items that meet the standards, then X is a binomial random variable with parameters n and p. Thus, when n is large, it follows by the normal approximation to the binomial that X is approximately normally distributed with mean np and variance $np(1-p)$. Hence,

$$\frac{X - np}{\sqrt{np(1-p)}} \mathrel{\dot\sim} \mathcal{N}(0, 1) \tag{7.5.1}$$

where $\dot{\sim}$ means "is approximately distributed as." Therefore, for any $\alpha \in (0, 1)$,

$$P\left\{-z_{\alpha/2} < \frac{X - np}{\sqrt{np(1-p)}} < z_{\alpha/2}\right\} \approx 1 - \alpha$$

and so if X is observed to equal x, then an approximate $100(1 - \alpha)$ percent confidence *region* for p is

$$\left\{p : -z_{\alpha/2} < \frac{x - np}{\sqrt{np(1-p)}} < z_{\alpha/2}\right\}$$

The foregoing region, however, is not an interval. To obtain a confidence *interval* for p, let $\hat{p} = X/n$ be the fraction of the items that meet the standards. From Example 7.2a, $\hat{p}$ is the maximum likelihood estimator of p, and so should be approximately equal to p. As a result, $\sqrt{n\hat{p}(1-\hat{p})}$ will be approximately equal to $\sqrt{np(1-p)}$ and so from Equation 7.5.1 we see that

$$\frac{X - np}{\sqrt{n\hat{p}(1-\hat{p})}} \dot{\sim} \mathcal{N}(0, 1)$$

Hence, for any $\alpha \in (0, 1)$ we have that

$$P\left\{-z_{\alpha/2} < \frac{X - np}{\sqrt{n\hat{p}(1-\hat{p})}} < z_{\alpha/2}\right\} \approx 1 - \alpha$$

or, equivalently,

$$P\{-z_{\alpha/2}\sqrt{n\hat{p}(1-\hat{p})} < np - X < z_{\alpha/2}\sqrt{n\hat{p}(1-\hat{p})}\} \approx 1 - \alpha$$

Dividing all sides of the preceding inequality by n, and using that $\hat{p} = X/n$, the preceding can be written as

$$P\{\hat{p} - z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n} < p < \hat{p} + z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}\} \approx 1 - \alpha$$

which yields an approximate $100(1 - \alpha)$ percent confidence interval for p.

EXAMPLE 7.5a A sample of 100 transistors is randomly chosen from a large batch and tested to determine if they meet the current standards. If 80 of them meet the standards, then an approximate 95 percent confidence interval for p, the fraction of all the transistors that meet the standards, is given by

$$(.8 - 1.96\sqrt{.8(.2)/100},\ .8 + 1.96\sqrt{.8(.2)/100}) = (.7216,\ .8784)$$

That is, with "95 percent confidence," between 72.16 and 87.84 percent of all transistors meet the standards. ■

EXAMPLE 7.5b In August 2013, the *New York Times* reported that a recent poll indicated that 52 percent of the population was in favor of the job performance of President Obama, with a margin of error of ± 4 percent. What does this mean? Can we infer how many people were questioned?

SOLUTION It has become common practice for the news media to present 95 percent confidence intervals. Since $z_{.025} = 1.96$, a 95 percent confidence interval for p, the percentage of the population that is in favor of President Obama's job performance, is given by

$$\hat{p} \pm 1.96\sqrt{\hat{p}(1-\hat{p})/n} = .52 \pm 1.96\sqrt{.52(.48)/n}$$

where n is the size of the sample. Since the "margin of error" is ± 4 percent, it follows that

$$1.96\sqrt{.52(.48)/n} = .04$$

or

$$n = \frac{(1.96)^2(.52)(.48)}{(.04)^2} = 599.29$$

That is, approximately 599 people were sampled, and 52 percent of them reported favorably on President Obama's job performance. ■

We often want to specify an approximate $100(1-\alpha)$ percent confidence interval for p that is no greater than some given length, say b. The problem is to determine the appropriate sample size n to obtain such an interval. To do so, note that the length of the approximate $100(1-\alpha)$ percent confidence interval for p from a sample of size n is

$$2z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$$

which is approximately equal to $2z_{\alpha/2}\sqrt{p(1-p)/n}$. Unfortunately, p is not known in advance, and so we cannot just set $2z_{\alpha/2}\sqrt{p(1-p)/n}$ equal to b to determine the necessary sample size n. What we can do, however, is to first take a preliminary sample to obtain a rough estimate of p, and then use this estimate to determine n. That is, we use p^*, the proportion of the preliminary sample that meets the standards, as a preliminary estimate of p; we then determine the total sample size n by solving the equation

$$2z_{\alpha/2}\sqrt{p^*(1-p^*)/n} = b$$

Squaring both sides of the preceding yields that

$$(2z_{\alpha/2})^2 p^*(1-p^*)/n = b^2$$

or

$$n = \frac{(2z_{\alpha/2})^2 p^*(1-p^*)}{b^2}$$

That is, if k items were initially sampled to obtain the preliminary estimate of p, then an additional $n-k$ (or 0 if $n \le k$) items should be sampled.

EXAMPLE 7.5c A certain manufacturer produces computer chips; each chip is independently acceptable with some unknown probability p. To obtain an approximate 99 percent confidence interval for p, whose length is approximately .05, an initial sample of 30 chips has been taken. If 26 of these chips are of acceptable quality, then the preliminary estimate of p is 26/30. Using this value, a 99 percent confidence interval of length approximately .05 would require an approximate sample of size

$$n = \frac{4(z_{.005})^2}{(.05)^2} \frac{26}{30}\left(1 - \frac{26}{30}\right) = \frac{4(2.58)^2}{(.05)^2} \frac{26}{30} \frac{4}{30} = 1{,}231$$

Hence, we should now sample an additional 1,201 chips and if, for instance, 1,040 of them are acceptable, then the final 99 percent confidence interval for p is

$$\left(\frac{1,066}{1,231} - \sqrt{1,066\left(1 - \frac{1,066}{1,231}\right)}\frac{z_{.005}}{1,231}, \frac{1,066}{1,231} + \sqrt{1,066\left(1 - \frac{1,066}{1,231}\right)}\frac{z_{.005}}{1,231}\right)$$

or

$$p \in (.84091, .89101) \quad \blacksquare$$

REMARK

As shown, a $100(1-\alpha)$ percent confidence interval for p will be of approximate length b when the sample size is

$$n = \frac{(2z_{\alpha/2})^2}{b^2} p(1-p)$$

Now it is easily shown that the function $g(p) = p(1-p)$ attains its maximum value of $\frac{1}{4}$, in the interval $0 \le p \le 1$, when $p = \frac{1}{2}$. Thus an upper bound on n is

$$n \le \frac{(z_{\alpha/2})^2}{b^2}$$

and so by choosing a sample whose size is at least as large as $(z_{\alpha/2})^2/b^2$, one can be assured of obtaining a confidence interval of length no greater than b without need of any additional sampling. $\blacksquare$

One-sided approximate confidence intervals for p are also easily obtained; Table 7.3 gives the results.

TABLE 7.3 *Approximate 100(1 −α) Percent Confidence Intervals for p*
X Is a Binomial (n, p) Random Variable
$\hat{p} = X/n$

Type of Interval	Confidence Interval
Two-sided	$\hat{p} \pm z_{\alpha/2}\sqrt{\hat{p}(1-\hat{p})/n}$
One-sided lower	$\left(-\infty,\ \hat{p} + z_{\alpha}\sqrt{\hat{p}(1-\hat{p})/n}\right)$
One-sided upper	$\left(\hat{p} - z_{\alpha}\sqrt{\hat{p}(1-\hat{p})/n},\ \infty\right)$

*7.6 CONFIDENCE INTERVAL OF THE MEAN OF THE EXPONENTIAL DISTRIBUTION

If $X_1, X_2, \ldots, X_n$ are independent exponential random variables each having mean θ, then it can be shown that the maximum likelihood estimator of θ is the sample mean $\sum_{i=1}^{n} X_i/n$. To obtain a confidence interval estimator of θ, recall from Section 5.7 that $\sum_{i=1}^{n} X_i$ has a gamma distribution with parameters n, $1/\theta$. This in turn implies (from the relationship between the gamma and chi-square distribution shown in Section 5.8.1.1) that

$$\frac{2}{\theta}\sum_{i=1}^{n} X_i \sim \chi^2_{2n}$$

Hence, for any $\alpha \in (0, 1)$

$$P\left\{\chi^2_{1-\alpha/2,\,2n} < \frac{2}{\theta}\sum_{i=1}^{n} X_i < \chi^2_{\alpha/2,\,2n}\right\} = 1 - \alpha$$

or, equivalently,

$$P\left\{\frac{2\sum_{i=1}^{n} X_i}{\chi^2_{\alpha/2,\,2n}} < \theta < \frac{2\sum_{i=1}^{n} X_i}{\chi^2_{1-\alpha/2,\,2n}}\right\} = 1 - \alpha$$

Hence, a $100(1-\alpha)$ percent confidence interval for θ is

$$\theta \in \left(\frac{2\sum_{i=1}^{n} X_i}{\chi^2_{\alpha/2,\,2n}},\ \frac{2\sum_{i=1}^{n} X_i}{\chi^2_{1-\alpha/2,\,2n}}\right)$$

* Optional section.

EXAMPLE 7.6a The successive items produced by a certain manufacturer are assumed to have useful lives that (in hours) are independent with a common density function

$$f(x) = \frac{1}{\theta}e^{-x/\theta}, \quad 0 < x < \infty$$

If the sum of the lives of the first 10 items is equal to 1,740, what is a 95 percent confidence interval for the population mean θ?

SOLUTION From Program 5.8.1b (or Table A2), we see that

$$\chi^2_{.025,20} = 34.169, \qquad \chi^2_{.975,20} = 9.661$$

and so we can conclude, with 95 percent confidence, that

$$\theta \in \left(\frac{3480}{34.169}, \frac{3480}{9.661}\right)$$

or, equivalently,

$$\theta \in (101.847, 360.211) \quad ■$$

*7.7 EVALUATING A POINT ESTIMATOR

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a sample from a population whose distribution is specified up to an unknown parameter θ, and let $d = d(\mathbf{X})$ be an estimator of θ. How are we to determine its worth as an estimator of θ? One way is to consider the square of the difference between $d(\mathbf{X})$ and θ. However, since $(d(\mathbf{X}) - \theta)^2$ is a random variable, let us agree to consider $r(d, \theta)$, the *mean square error* of the estimator d, which is defined by

$$r(d, \theta) = E[(d(\mathbf{X}) - \theta)^2]$$

as an indication of the worth of d as an estimator of θ.

It would be nice if there were a single estimator d that minimized $r(d, \theta)$ for all possible values of θ. However, except in trivial situations, this will never be the case. For example, consider the estimator d^* defined by

$$d^*(X_1, \ldots, X_n) = 4$$

That is, no matter what the outcome of the sample data, the estimator d^* chooses 4 as its estimate of θ. While this seems like a silly estimator (since it makes no use of the data), it is, however, true that when θ actually equals 4, the mean square error of this estimator is 0.

* Optional section.

Thus, the mean square error of any estimator different from d^* must, in most situations, be larger than the mean square error of d^* when $\theta = 4$.

Although minimum mean square estimators rarely exist, it is sometimes possible to find an estimator having the smallest mean square error among all estimators that satisfy a certain property. One such property is that of unbiasedness.

Definition

Let $d = d(\mathbf{X})$ be an estimator of the parameter θ. Then

$$b_\theta(d) = E[d(\mathbf{X})] - \theta$$

is called the *bias* of d as an estimator of θ. If $b_\theta(d) = 0$ for all θ, then we say that d is an *unbiased* estimator of θ. In other words, an estimator is unbiased if its expected value always equals the value of the parameter it is attempting to estimate.

EXAMPLE 7.7a Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution having unknown mean θ. Then

$$d_1(X_1, X_2, \ldots, X_n) = X_1$$

and

$$d_2(X_1, X_2, \ldots, X_n) = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

are both unbiased estimators of θ since

$$E[X_1] = E\left[\frac{X_1 + X_2 + \cdots + X_n}{n}\right] = \theta$$

More generally, $d_3(X_1, X_2, \ldots, X_n) = \sum_{i=1}^n \lambda_i X_i$ is an unbiased estimator of θ whenever $\sum_{i=1}^n \lambda_i = 1$. This follows since

$$\begin{aligned} E\left[\sum_{i=1}^n \lambda_i X_i\right] &= \sum_{i=1}^n E[\lambda_i X_i] \\ &= \sum_{i=1}^n \lambda_i E(X_i) \\ &= \theta \sum_{i=1}^n \lambda_i \\ &= \theta \quad \blacksquare \end{aligned}$$

If $d(X_1, \ldots, X_n)$ is an unbiased estimator, then its mean square error is given by

$$\begin{aligned} r(d, \theta) &= E[(d(\mathbf{X}) - \theta)^2] \\ &= E[(d(\mathbf{X}) - E[d(\mathbf{X})])^2] \qquad \text{since } d \text{ is unbiased} \\ &= \mathrm{Var}(d(\mathbf{X})) \end{aligned}$$

Thus the mean square error of an unbiased estimator is equal to its variance.

EXAMPLE 7.7b (Combining Independent Unbiased Estimators) Let d_1 and d_2 denote independent unbiased estimators of θ, having known variances σ_1^2 and σ_2^2. That is, for $i = 1, 2$,

$$E[d_i] = \theta, \qquad \mathrm{Var}(d_i) = \sigma_i^2$$

Any estimator of the form

$$d = \lambda d_1 + (1 - \lambda) d_2$$

will also be unbiased. To determine the value of λ that results in d having the smallest possible mean square error, note that

$$\begin{aligned} r(d, \theta) &= \mathrm{Var}(d) \\ &= \lambda^2 \, \mathrm{Var}(d_1) + (1 - \lambda)^2 \, \mathrm{Var}(d_2) \\ &\qquad \text{by the independence of } d_1 \text{ and } d_2 \\ &= \lambda^2 \sigma_1^2 + (1 - \lambda)^2 \sigma_2^2 \end{aligned}$$

Differentiation yields that

$$\frac{d}{d\lambda} r(d, \theta) = 2\lambda\sigma_1^2 - 2(1 - \lambda)\sigma_2^2$$

To determine the value of λ that minimizes $r(d, \theta)$ — call it $\hat{\lambda}$ — set this equal to 0 and solve for λ to obtain

$$2\hat{\lambda}\sigma_1^2 = 2(1 - \hat{\lambda})\sigma_2^2$$

or

$$\hat{\lambda} = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2} = \frac{1/\sigma_1^2}{1/\sigma_1^2 + 1/\sigma_2^2}$$

In words, the optimal weight to give an estimator is inversely proportional to its variance (when all the estimators are unbiased and independent).

For an application of the foregoing, suppose that a conservation organization wants to determine the acidity content of a certain lake. To determine this quantity, they draw some water from the lake and then send samples of this water to n different laboratories. These laboratories will then, independently, test for acidity content by using their respective titration equipment, which is of differing precision. Specifically, suppose that d_i, the result of a titration test at laboratory i, is a random variable having mean θ, the true acidity of the sample water, and variance σ_i^2, $i = 1, \ldots, n$. If the quantities σ_i^2, $i = 1, \ldots, n$ are known to the conservation organization, then they should estimate the acidity of the sampled water from the lake by

$$d = \frac{\sum_{i=1}^{n} d_i/\sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2}$$

The mean square error of d is as follows:

$$\begin{aligned}
r(d,\theta) &= \operatorname{Var}(d) \qquad \text{since } d \text{ is unbiased} \\
&= \left(\sum_{i=1}^{n} 1/\sigma_i^2\right)^{-2} \sum_{i=1}^{n} \left(\frac{1}{\sigma_i^2}\right)^2 \sigma_i^2 \\
&= \frac{1}{\sum_{i=1}^{n} 1/\sigma_i^2} \qquad \blacksquare
\end{aligned}$$

A generalization of the result that the mean square error of an unbiased estimator is equal to its variance is that the mean square error of any estimator is equal to its variance plus the square of its bias. This follows since

$$\begin{aligned}
r(d,\theta) &= E[(d(\mathbf{X}) - \theta)^2] \\
&= E[(d - E[d] + E[d] - \theta)^2] \\
&= E[(d - E[d])^2 + (E[d] - \theta)^2 + 2(E[d] - \theta)(d - E[d])] \\
&= E[(d - E[d])^2] + E[(E[d] - \theta)^2] \\
&\quad + 2E[(E[d] - \theta)(d - E[d])] \\
&= E[(d - E[d])^2] + (E[d] - \theta)^2 + 2(E[d] - \theta)E[d - E[d]] \\
&\qquad\qquad \text{since } E[d] - \theta \text{ is constant} \\
&= E[(d - E[d])^2] + (E[d] - \theta)^2
\end{aligned}$$

The last equality follows since

$$E[d - E[d]] = 0$$

Hence

$$r(d, \theta) = \text{Var}(d) + b_\theta^2(d)$$

EXAMPLE 7.7c Let $X_1, \ldots, X_n$ denote a sample from a uniform $(0, \theta)$ distribution, where θ is assumed unknown. Since

$$E[X_i] = \frac{\theta}{2}$$

a "natural" estimator to consider is the unbiased estimator

$$d_1 = d_1(\mathbf{X}) = \frac{2 \sum_{i=1}^{n} X_i}{n}$$

Since $E[d_1] = \theta$, it follows that

$$\begin{aligned}
r(d_1, \theta) &= \text{Var}(d_1) \\
&= \frac{4}{n} \text{Var}(X_i) \\
&= \frac{4}{n} \frac{\theta^2}{12} \qquad \text{since } \text{Var}(X_i) = \frac{\theta^2}{12} \\
&= \frac{\theta^2}{3n}
\end{aligned}$$

A second possible estimator of θ is the maximum likelihood estimator, which, as shown in Example 7.2d, is given by

$$d_2 = d_2(\mathbf{X}) = \max_i X_i$$

To compute the mean square error of d_2 as an estimator of θ, we need to first compute its mean (so as to determine its bias) and variance. To do so, note that the distribution function of d_2 is as follows:

$$\begin{aligned}
F_2(x) &\equiv P\{d_2(\mathbf{X}) \leq x\} \\
&= P\{\max_i X_i \leq x\} \\
&= P\{X_i \leq x \ \text{ for all } \ i = 1, \ldots, n\} \\
&= \prod_{i=1}^{n} P\{X_i \leq x\} \qquad \text{by independence} \\
&= \left(\frac{x}{\theta}\right)^n \qquad x \leq \theta
\end{aligned}$$

Hence, upon differentiating, we obtain that the density function of d_2 is

$$f_2(x) = \frac{nx^{n-1}}{\theta^n}, x \leq \theta$$

Therefore,

$$E[d_2] = \int_0^{\infty} x\, f_2(x)\, dx = \frac{n}{\theta^n} \int_0^{\theta} x^n\, dx = \frac{n}{n+1}\,\theta \tag{7.7.1}$$

Also

$$E[d_2^2] = \frac{n}{\theta^n} \int_0^{\theta} x^{n+1}\, dx = \frac{n}{n+2}\,\theta^2$$

and so

$$\begin{aligned} \operatorname{Var}(d_2) &= \frac{n}{n+2}\theta^2 - \left(\frac{n}{n+1}\theta\right)^2 \qquad (7.7.2) \\ &= n\,\theta^2 \left[\frac{1}{n+2} - \frac{n}{(n+1)^2}\right] = \frac{n\,\theta^2}{(n+2)(n+1)^2} \end{aligned}$$

Hence

$$\begin{aligned} r(d_2, \theta) &= (E(d_2) - \theta)^2 + \operatorname{Var}(d_2) \qquad (7.7.3) \\ &= \frac{\theta^2}{(n+1)^2} + \frac{n\,\theta^2}{(n+2)(n+1)^2} \\ &= \frac{\theta^2}{(n+1)^2}\left[1 + \frac{n}{n+2}\right] \\ &= \frac{2\theta^2}{(n+1)(n+2)} \end{aligned}$$

Since

$$\frac{2\theta^2}{(n+1)(n+2)} \leq \frac{\theta^2}{3n} \qquad n = 1, 2, \ldots$$

it follows that d_2 is a superior estimator of θ than is d_1.

Equation 7.7.1 suggests the use of even another estimator — namely, the unbiased estimator $(1 + 1/n)d_2(\mathbf{X}) = (1 + 1/n) \max_i X_i$. However, rather than considering this estimator directly, let us consider all estimators of the form

$$d_c(\mathbf{X}) = c \max_i X_i = c\, d_2(\mathbf{X})$$

where c is a given constant. The mean square error of this estimator is

$$\begin{aligned} r(d_c(\mathbf{X}), \theta) &= \operatorname{Var}(d_c(\mathbf{X})) + (E[d_c(\mathbf{X})] - \theta)^2 \\ &= c^2 \operatorname{Var}(d_2(\mathbf{X})) + (cE[d_2(\mathbf{X})] - \theta)^2 \\ &= \frac{c^2 n \theta^2}{(n+2)(n+1)^2} + \theta^2 \left(\frac{c\, n}{n+1} - 1 \right)^2 \\ &\qquad \text{by Equations 7.7.2 and 7.7.1} \end{aligned} \tag{7.7.4}$$

To determine the constant c resulting in minimal mean square error, we differentiate to obtain

$$\frac{d}{dc} r(d_c(\mathbf{X}), \theta) = \frac{2c\, n\theta^2}{(n+2)(n+1)^2} + \frac{2\theta^2 n}{n+1} \left(\frac{c\, n}{n+1} - 1 \right)$$

Equating this to 0 shows that the best constant c — call it c^* — is such that

$$\frac{c^*}{n+2} + c^* n - (n+1) = 0$$

or

$$c^* = \frac{(n+1)(n+2)}{n^2 + 2n + 1} = \frac{n+2}{n+1}$$

Substituting this value of c into Equation 7.7.4 yields that

$$\begin{aligned} r\left(\frac{n+2}{n+1} \max_i X_i, \theta \right) &= \frac{(n+2)n\theta^2}{(n+1)^4} + \theta^2 \left(\frac{n(n+2)}{(n+1)^2} - 1 \right)^2 \\ &= \frac{(n+2)n\theta^2}{(n+1)^4} + \frac{\theta^2}{(n+1)^4} \\ &= \frac{\theta^2}{(n+1)^2} \end{aligned}$$

A comparison with Equation 7.7.3 shows that the (biased) estimator $(n+2)/(n+1) \max_i X_i$ has about half the mean square error of the maximum likelihood estimator $\max_i X_i$. ■

*7.8 THE BAYES ESTIMATOR

In certain situations it seems reasonable to regard an unknown parameter θ as being the value of a random variable from a given probability distribution. This usually arises when, prior to the observance of the outcomes of the data $X_1, \ldots, X_n$, we have some information about the value of θ and this information is expressible in terms of a probability distribution (called appropriately the *prior* distribution of θ). For instance, suppose that from

* Optional section.

past experience we know that θ is equally likely to be near any value in the interval (0, 1). Hence, we could reasonably assume that θ is chosen from a uniform distribution on (0, 1).

Suppose now that our prior feelings about θ are that it can be regarded as being the value of a continuous random variable having probability density function $p(\theta)$; and suppose that we are about to observe the value of a sample whose distribution depends on θ. Specifically, suppose that $f(x|\theta)$ represents the likelihood — that is, it is the probability mass function in the discrete case or the probability density function in the continuous case — that a data value is equal to x when θ is the value of the parameter. If the observed data values are $X_i = x_i, i = 1, \dots, n$, then the updated, or conditional, probability density function of θ is as follows:

$$\begin{aligned} f(\theta|x_1, \dots, x_n) &= \frac{f(\theta, x_1, \dots, x_n)}{f(x_1, \dots, x_n)} \\ &= \frac{p(\theta)f(x_1, \dots, x_n|\theta)}{\int f(x_1, \dots, x_n|\theta)p(\theta)\, d\theta} \end{aligned}$$

The conditional density function $f(\theta|x_1, \dots, x_n)$ is called the *posterior* density function. (Thus, before observing the data, one's feelings about θ are expressed in terms of the prior distribution, whereas once the data are observed, this prior distribution is updated to yield the posterior distribution.)

Now we have shown that whenever we are given the probability distribution of a random variable, the best estimate of the value of that random variable, in the sense of minimizing the expected squared error, is its mean. Therefore, it follows that the best estimate of θ, given the data values $X_i = x_i$, $i = 1, \dots, n$, is the mean of the posterior distribution $f(\theta|x_1, \dots, x_n)$. This estimator, called the *Bayes estimator*, is written as $E[\theta|X_1, \dots, X_n]$. That is, if $X_i = x_i$, $i = 1, \dots, n$, then the value of the Bayes estimator is

$$E[\theta|X_1 = x_1, \dots, X_n = x_n] = \int \theta f(\theta|x_1, \dots, x_n)\, d\theta$$

EXAMPLE 7.8a Suppose that $X_1, \dots, X_n$ are independent Bernoulli random variables, each having probability mass function given by

$$f(x|\theta) = \theta^x(1-\theta)^{1-x}, \qquad x = 0, 1$$

where θ is unknown. Further, suppose that θ is chosen from a uniform distribution on (0, 1). Compute the Bayes estimator of θ.

SOLUTION We must compute $E[\theta|X_1, \dots, X_n]$. Since the prior density of θ is the uniform density

$$p(\theta) = 1, \qquad 0 < \theta < 1$$

we have that the conditional density of θ given $X_1, \ldots, X_n$ is given by

$$\begin{aligned}
f(\theta|x_1, \ldots, x_n) &= \frac{f(x_1, \ldots, x_n, \theta)}{f(x_1, \ldots, x_n)} \\
&= \frac{f(x_1, \ldots, x_n|\theta)p(\theta)}{\int_0^1 f(x_1, \ldots, x_n|\theta)p(\theta)\, d\theta} \\
&= \frac{\theta^{\Sigma_1^n x_i}(1-\theta)^{n-\Sigma_1^n x_i}}{\int_0^1 \theta^{\Sigma_1^n x_i}(1-\theta)^{n-\Sigma_1^n x_i}\, d\theta}
\end{aligned}$$

Now it can be shown that for integral values m and r

$$\int_0^1 \theta^m (1-\theta)^r\, d\theta = \frac{m!r!}{(m+r+1)!} \tag{7.8.1}$$

Hence, upon letting $x = \sum_{i=1}^n x_i$

$$f(\theta|x_1, \ldots, x_n) = \frac{(n+1)!\ \theta^x (1-\theta)^{n-x}}{x!\ (n-x)!} \tag{7.8.2}$$

Therefore,

$$\begin{aligned}
E[\theta|x_1, \ldots, x_n] &= \frac{(n+1)!}{x!(n-x)!} \int_0^1 \theta^{1+x}(1-\theta)^{n-x}\, d\theta \\
&= \frac{(n+1)!}{x!(n-x)!} \frac{(1+x)!(n-x)!}{(n+2)!} \qquad \text{from Equation 7.8.1} \\
&= \frac{x+1}{n+2}
\end{aligned}$$

Thus, the Bayes estimator is given by

$$E[\theta|X_1, \ldots, X_n] = \frac{\sum_{i=1}^n X_i + 1}{n+2}$$

As an illustration, if 10 independent trials, each of which results in a success with probability θ, result in 6 successes, then assuming a uniform (0, 1) prior distribution on θ, the Bayes estimator of θ is 7/12 (as opposed, for instance, to the maximum likelihood estimator of 6/10). ■

REMARK

The conditional distribution of θ given that $X_i = x_i$, $i = 1, \ldots, n$, whose density function is given by Equation 7.8.2, is called the beta distribution with parameters $\sum_{i=1}^{n} x_i + 1$, $n - \sum_{i=1}^{n} x_i + 1$. ■

EXAMPLE 7.8b Suppose $X_1, \ldots, X_n$ are independent normal random variables, each having unknown mean θ and known variance σ_0^2. If θ is itself selected from a normal population having known mean μ and known variance σ^2, what is the Bayes estimator of θ?

SOLUTION In order to determine $E[\theta|X_1, \ldots, X_n]$, the Bayes estimator, we need first determine the conditional density of θ given the values of $X_1, \ldots, X_n$. Now

$$f(\theta|x_1, \ldots, x_n) = \frac{f(x_1, \ldots, x_n|\theta)p(\theta)}{f(x_1, \ldots, x_n)}$$

where

$$f(x_1, \ldots, x_n|\theta) = \frac{1}{(2\pi)^{n/2}\sigma_0^n} \exp\left\{-\sum_{i=1}^{n}(x_i - \theta)^2/2\sigma_0^2\right\}$$

$$p(\theta) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\{-(\theta - \mu)^2/2\sigma^2\}$$

and

$$f(x_1, \ldots, x_n) = \int_{-\infty}^{\infty} f(x_1, \ldots, x_n|\theta)p(\theta)\, d\theta$$

With the help of a little algebra, it can now be shown that this conditional density is a *normal* density with mean

$$\begin{aligned} E[\theta|X_1, \ldots, X_n] &= \frac{n\sigma^2}{n\sigma^2 + \sigma_0^2}\overline{X} + \frac{\sigma_0^2}{n\sigma^2 + \sigma_0^2}\mu \qquad (7.8.3) \\ &= \frac{\frac{n}{\sigma_0^2}}{\frac{n}{\sigma_0^2} + \frac{1}{\sigma^2}}\overline{X} + \frac{\frac{1}{\sigma^2}}{\frac{n}{\sigma_0^2} + \frac{1}{\sigma^2}}\mu \end{aligned}$$

and variance

$$\text{Var}(\theta|X_1, \ldots, X_n) = \frac{\sigma_0^2\sigma^2}{n\sigma^2 + \sigma_0^2}$$

Writing the Bayes estimator as we did in Equation 7.8.3 is informative, for it shows that it is a weighted average of $\overline{X}$, the sample mean, and μ, the *a priori* mean. In fact, the weights given to these two quantities are in proportion to the inverses of σ_0^2/n (the conditional variance of the sample mean $\overline{X}$ given θ) and σ^2 (the variance of the prior distribution). ■

REMARK: ON CHOOSING A NORMAL PRIOR

As illustrated by Example 7.8b, it is computationally very convenient to choose a normal prior for the unknown mean θ of a normal distribution — for then the Bayes estimator is simply given by Equation 7.8.3. This raises the question of how one should go about determining whether there is a normal prior that reasonably represents one's prior feelings about the unknown mean.

To begin, it seems reasonable to determine the value — call it μ — that you *a priori* feel is most likely to be near θ. That is, we start with the mode (which equals the mean when the distribution is normal) of the prior distribution. We should then try to ascertain whether or not we believe that the prior distribution is symmetric about μ. That is, for each $a > 0$ do we believe that it is just as likely that θ will lie between $\mu - a$ and μ as it is that it will be between μ and $\mu + a$? If the answer is positive, then we accept, as a working hypothesis, that our prior feelings about θ can be expressed in terms of a prior distribution that is normal with mean μ. To determine σ, the standard deviation of the normal prior, think of an interval centered about μ that you *a priori* feel is 90 percent certain to contain θ. For instance, suppose you feel 90 percent (no more and no less) certain that θ will lie between $\mu - a$ and $\mu + a$. Then, since a normal random variable θ with mean μ and variance σ^2 is such that

$$P\left\{-1.645 < \frac{\theta - \mu}{\sigma} < 1.645\right\} = .90$$

or

$$P\{\mu - 1.645\sigma < \theta < \mu + 1.645\sigma\} = .90$$

it seems reasonable to take

$$1.645\sigma = a \quad \text{or} \quad \sigma = \frac{a}{1.645}$$

Thus, if your prior feelings can indeed be reasonably described by a normal distribution, then that distribution would have mean μ and standard deviation $\sigma = a/1.645$. As a test of whether this distribution indeed fits your prior feelings you might ask yourself such questions as whether you are 95 percent certain that θ will fall between $\mu - 1.96\sigma$ and $\mu + 1.96\sigma$, or whether you are 99 percent certain that θ will fall between $\mu - 2.58\sigma$ and $\mu + 2.58\sigma$, where these intervals are determined by the equalities

$$P\left\{-1.96 < \frac{\theta - \mu}{\sigma} < 1.96\right\} = .95$$

$$P\left\{-2.58 < \frac{\theta - \mu}{\sigma} < 2.58\right\} = .99$$

which hold when θ is normal with mean μ and variance σ^2.

EXAMPLE 7.8c Consider the likelihood function $f(x_1, \ldots, x_n|\theta)$ and suppose that θ is uniformly distributed over some interval (a, b). The posterior density of θ given $X_1, \ldots, X_n$ equals

$$\begin{aligned} f(\theta|x_1, \ldots, x_n) &= \frac{f(x_1, \ldots, x_n|\theta)p(\theta)}{\int_a^b f(x_1, \ldots, x_n|\theta)p(\theta)\, d\theta} \\ &= \frac{f(x_1, \ldots, x_n|\theta)}{\int_a^b f(x_1, \ldots, x_n|\theta)\, d\theta} \qquad a < \theta < b \end{aligned}$$

Now the *mode* of a density $f(\theta)$ was defined to be that value of θ that maximizes $f(\theta)$. By the foregoing, it follows that the mode of the density $f(\theta|x_1, \ldots, x_n)$ is that value of θ maximizing $f(x_1, \ldots, x_n|\theta)$; that is, it is just the maximum likelihood estimate of θ [when it is constrained to be in (a, b)]. In other words, the maximum likelihood estimate equals the mode of the posterior distribution when a uniform prior distribution is assumed. ■

If, rather than a point estimate, we desire an interval in which θ lies with a specified probability — say $1 - \alpha$ — we can accomplish this by choosing values a and b such that

$$\int_a^b f(\theta|x_1, \ldots, x_n)\, d\theta = 1 - \alpha$$

EXAMPLE 7.8d Suppose that if a signal of value s is sent from location A, then the signal value received at location B is normally distributed with mean s and variance 60. Suppose also that the value of a signal sent at location A is, *a priori*, known to be normally distributed with mean 50 and variance 100. If the value received at location B is equal to 40, determine an interval that will contain the actual value sent with probability .90.

SOLUTION It follows from Example 7.8b that the conditional distribution of S, the signal value sent, given that 40 is the value received, is normal with mean and variance given by

$$\begin{aligned} E[S|\text{data}] &= \frac{1/60}{1/60 + 1/100} 40 + \frac{1/100}{1/60 + 1/100} 50 = 43.75 \\ \text{Var}(S|\text{data}) &= \frac{1}{1/60 + 1/100} = 37.5 \end{aligned}$$

Hence, given that the value received is 40, $(S - 43.75)/\sqrt{37.5}$ has a standard normal distribution and so

$$P\left\{-1.645 < \frac{S - 43.75}{\sqrt{37.5}} < 1.645|\text{data}\right\} = .90$$

or

$$P\{43.75 - 1.645\sqrt{37.5} < S < 43.75 + 1.645\sqrt{37.5}|\text{data}\} = .95$$

That is, with *probability* .90, the true signal sent lies within the interval (33.68, 53.82). ■

Problems

1. Let $X_1, \ldots, X_n$ be a sample from the distribution whose density function is

$$f(x) = \begin{cases} e^{-(x-\theta)} & x \geq \theta \\ 0 & \text{otherwise} \end{cases}$$

   Determine the maximum likelihood estimator of θ.

2. Determine the maximum likelihood estimator of θ when $X_1, \ldots, X_n$ is a sample with density function

$$f(x) = \tfrac{1}{2}e^{-|x-\theta|}, \qquad -\infty < x < \infty$$

3. Let $X_1, \ldots, X_n$ be a sample from a normal μ, σ^2 population. Determine the maximum likelihood estimator of σ^2 when μ is known. What is the expected value of this estimator?

4. Determine the maximum likelihood estimates of a and λ when $X_1, \ldots, X_n$ is a sample from the Pareto density function

$$f(x) = \begin{cases} \lambda a^{\lambda} x^{-(\lambda+1)}, & \text{if } x \geq a \\ 0, & \text{if } x < a \end{cases}$$

5. Suppose that $X_1, \ldots, X_n$ are normal with mean μ_1; $Y_1, \ldots, Y_n$ are normal with mean μ_2; and $W_1, \ldots, W_n$ are normal with mean $\mu_1 + \mu_2$. Assuming that all $3n$ random variables are independent with a common variance, find the maximum likelihood estimators of μ_1 and μ_2.

6. River floods are often measured by their discharges (in units of feet cubed per second). The value v is said to be the value of a 100-year flood if

$$P\{D \geq v\} = .01$$

   where D is the discharge of the largest flood in a randomly chosen year. The following table gives the flood discharges of the largest floods of the Blackstone River in Woonsocket, Rhode Island, in each of the years from 1929 to 1965. Assuming that these discharges follow a lognormal distribution, estimate the value of a 100-year flood.

Annual Floods of the Blackstone River (1929–1965)

Year	Flood Discharge (ft^3/s)
1929	4,570
1930	1,970
1931	8,220
1932	4,530
1933	5,780
1934	6,560
1935	7,500
1936	15,000
1937	6,340
1938	15,100
1939	3,840
1940	5,860
1941	4,480
1942	5,330
1943	5,310
1944	3,830
1945	3,410
1946	3,830
1947	3,150
1948	5,810
1949	2,030
1950	3,620
1951	4,920
1952	4,090
1953	5,570
1954	9,400
1955	32,900
1956	8,710
1957	3,850
1958	4,970
1959	5,398
1960	4,780
1961	4,020
1962	5,790
1963	4,510
1964	5,520
1965	5,300

7. Recall that X is said to have a lognormal distribution with parameters μ and σ^2 if $\log(X)$ is normal with mean μ and variance σ^2. Suppose X is such a lognormal random variable.
(a) Find $E[X]$.
(b) Find $\text{Var}(X)$.

Hint: Make use of the formula for the moment generating function of a normal random variable.
(c) The following are, in minutes, travel times to work over a sequence of 10 days.

$$42, 28, 53, 57, 67, 39, 35, 50, 44, 39$$

Assuming an underlying lognormal distribution, use the data to estimate the mean travel time.

8. An electric scale gives a reading equal to the true weight plus a random error that is normally distributed with mean 0 and standard deviation $\sigma = .1$ mg. Suppose that the results of five successive weighings of the same object are as follows: 3.142, 3.163, 3.155, 3.150, 3.141.
   (a) Determine a 95 percent confidence interval estimate of the true weight.
   (b) Determine a 99 percent confidence interval estimate of the true weight.

9. The PCB concentration of a fish caught in Lake Michigan was measured by a technique that is known to result in an error of measurement that is normally distributed with a standard deviation of .08 ppm (parts per million). Suppose the results of 10 independent measurements of this fish are

$$11.2,\ 12.4,\ 10.8,\ 11.6,\ 12.5,\ 10.1,\ 11.0,\ 12.2,\ 12.4,\ 10.6$$

   (a) Give a 95 percent confidence interval for the PCB level of this fish.
   (b) Give a 95 percent lower confidence interval.
   (c) Give a 95 percent upper confidence interval.

10. The standard deviation of test scores on a certain achievement test is 11.3. If a random sample of 81 students had a sample mean score of 74.6, find a 90 percent confidence interval estimate for the average score of all students.

11. Let $X_1, \ldots, X_n, X_{n+1}$ be a sample from a normal population having an unknown mean μ and variance 1. Let $\bar{X}_n = \sum_{i=1}^{n} X_i/n$ be the average of the first n of them.
    (a) What is the distribution of $X_{n+1} - \bar{X}_n$?
    (b) If $\bar{X}_n = 4$, give an interval that, with 90 percent confidence, will contain the value of X_{n+1}.

12. If $X_1, \ldots, X_n$ is a sample from a normal population whose mean μ is unknown but whose variance σ^2 is known, show that $(-\infty, \overline{X} + z_\alpha \sigma/\sqrt{n})$ is a $100(1-\alpha)$ percent lower confidence interval for μ.

13. A sample of 20 cigarettes is tested to determine nicotine content and the average value observed was 1.2 mg. Compute a 99 percent two-sided confidence interval for the mean nicotine content of a cigarette if it is known that the standard deviation of a cigarette's nicotine content is $\sigma = .2$ mg.

14. In Problem 13, suppose that the population variance is not known in advance of the experiment. If the sample variance is .04, compute a 99 percent two-sided confidence interval for the mean nicotine content.

15. In Problem 14, compute a value c for which we can assert "with 99 percent confidence" that c is larger than the mean nicotine content of a cigarette.

16. Suppose that when sampling from a normal population having an unknown mean μ and unknown variance σ^2, we wish to determine a sample size n so as to guarantee that the resulting $100(1-\alpha)$ percent confidence interval for μ will be of size no greater than A, for given values α and A. Explain how we can approximately do this by a double sampling scheme that first takes a subsample of size 30 and then chooses the total sample size by using the results of the first subsample.

17. The following data resulted from 24 independent measurements of the melting point of lead.

330°C	322°C	345°C
328.6°C	331°C	342°C
342.4°C	340.4°C	329.7°C
334°C	326.5°C	325.8°C
337.5°C	327.3°C	322.6°C
341°C	340°C	333°C
343.3°C	331°C	341°C
329.5°C	332.3°C	340°C

Assuming that the measurements can be regarded as constituting a normal sample whose mean is the true melting point of lead, determine a 95 percent two-sided confidence interval for this value. Also determine a 99 percent two-sided confidence interval.

18. The following are scores on IQ tests of a random sample of 18 students at a large eastern university.

$$130,\ 122,\ 119,\ 142,\ 136,\ 127,\ 120,\ 152,\ 141,$$
$$132,\ 127,\ 118,\ 150,\ 141,\ 133,\ 137,\ 129,\ 142$$

(a) Construct a 95 percent confidence interval estimate of the average IQ score of all students at the university.
(b) Construct a 95 percent lower confidence interval estimate.
(c) Construct a 95 percent upper confidence interval estimate.

19. Suppose that a random sample of nine recently sold houses in a certain city resulted in a sample mean price of \$222,000, with a sample standard deviation of \$22,000. Give a 95 percent upper confidence interval for the mean price of all recently sold houses in this city.

20. A company self-insures its large fleet of cars against collisions. To determine its mean repair cost per collision, it has randomly chosen a sample of 16 accidents. If the average repair cost in these accidents is \$2,200 with a sample standard deviation of \$800, find a 90 percent confidence interval estimate of the mean cost per collision.

21. A standardized test is given annually to all sixth-grade students in the state of Washington. To determine the average score of students in her district, a school supervisor selects a random sample of 100 students. If the sample mean of these students' scores is 320 and the sample standard deviation is 16, give a 95 percent confidence interval estimate of the average score of students in that supervisor's district.

22. Each of 20 science students independently measured the melting point of lead. The sample mean and sample standard deviation of these measurements were (in degrees centigrade) 330.2 and 15.4, respectively. Construct **(a)** a 95 percent and **(b)** a 99 percent confidence interval estimate of the true melting point of lead.

23. A random sample of 300 CitiBank VISA cardholder accounts indicated a sample mean debt of \$1,220 with a sample standard deviation of \$840. Construct a 95 percent confidence interval estimate of the average debt of all cardholders.

24. In Problem 23, find the smallest value v that "with 90 percent confidence," exceeds the average debt per cardholder.

25. Verify the formula given in Table 7.1 for the $100(1-\alpha)$ percent lower confidence interval for μ when σ is unknown.

26. The following are the daily number of steps taken by a certain individual in 20 weekdays.

2,100	1,984	2,072	1,898
1,950	1,992	2,096	2,103
2,043	2,218	2,244	2,206
2,210	2,152	1,962	2,007
2,018	2,106	1,938	1,956

Assuming that the daily number of steps is normally distributed, construct **(a)** a 95 percent and **(b)** a 99 percent two-sided confidence interval for the mean number of steps. **(c)** Determine the largest value v that, "with 95 percent confidence," will be less than the mean range.

27. Studies were conducted in Los Angeles to determine the carbon monoxide concentration near freeways. The basic technique used was to capture air samples in special bags and to then determine the carbon monoxide concentration by using a spectrophotometer. The measurements in ppm (parts per million) over a sampled period during the year were 102.2, 98.4, 104.1, 101, 102.2, 100.4, 98.6, 88.2, 78.8, 83, 84.7, 94.8, 105.1, 106.2, 111.2, 108.3, 105.2, 103.2, 99, 98.8. Compute a 95 percent two-sided confidence interval for the mean carbon monoxide concentration.

28. A set of 10 determinations, by a method devised by the chemist Karl Fischer, of the percentage of water in a methanol solution yielded the following data.

$$.50, .55, .53, .56, .54,$$
$$.57, .52, .60, .55, .58$$

Assuming normality, use these data to give a 95 percent confidence interval for the actual percentage.

29. Suppose that $U_1, U_2, \ldots$ is a sequence of independent uniform (0,1) random variables, and define N by

$$N = \min\{n : U_1 + \cdots + U_n > 1\}$$

That is, N is the number of uniform (0, 1) random variables that need to be summed to exceed 1. Use random numbers to determine the value of 36 random variables having the same distribution as N, and then use these data to obtain a 95 percent confidence interval estimate of $E[N]$. Based on this interval, guess the exact value of $E[N]$.

30. An important issue for a retailer is to decide when to reorder stock from a supplier. A common policy used to make the decision is of a type called s, S: The retailer orders at the end of a period if the on-hand stock is less than s, and orders enough to bring the stock up to S. The appropriate values of s and S depend on different cost parameters, such as inventory holding costs and the profit per item sold, as well as the distribution of the demand during a period. Consequently, it is important for the retailer to collect data relating to the parameters of the demand distribution. Suppose that the following data give the numbers of a certain type of item sold in each of 30 weeks.

$$14, 8, 12, 9, 5, 22, 15, 12, 16, 7, 10, 9, 15, 15, 12,$$
$$9, 11, 16, 8, 7, 15, 13, 9, 5, 18, 14, 10, 13, 7, 11$$

Assuming that the numbers sold each week are independent random variables from a common distribution, use the data to obtain a 95 percent confidence interval for the mean number sold in a week.

31. A random sample of 16 professors at a large private university yielded a sample mean annual salary of \$90,450 with a sample standard deviation of \$9,400. Determine a 95 percent confidence interval of the average salary of all professors at that university.

32. Let $X_1, \ldots, X_{n+1}$ be a sample from a population with mean μ and variance σ^2. As noted in the text, the natural predictor of X_{n+1} based on the data values $X_1, \ldots, X_n$ is $\bar{X}_n = \sum_{i=1}^{n} X_i/n$. Determine the mean square error of this predictor. That is, find $E[(X_{n+1} - \bar{X}_n)^2]$.

33. National Safety Council data show that the number of accidental deaths due to drowning in the United States in the years from 1990 to 1993 were (in units of one thousand) 5.2, 4.6, 4.3, 4.8. Use these data to give an interval that will, with 95 percent confidence, contain the number of such deaths in 1994.

34. The daily dissolved oxygen concentration for a water stream has been recorded over 30 days. If the sample average of the 30 values is 2.5 mg/liter and the sample standard deviation is 2.12 mg/liter, determine a value which, with 90 percent confidence, exceeds the mean daily concentration.

35. Verify the formulas given in Table 7.1 for the $100(1 - \alpha)$ percent lower and upper confidence intervals for σ^2.

36. The capacities (in ampere-hours) of 10 batteries were recorded as follows:

$$140, 136, 150, 144, 148, 152, 138, 141, 143, 151$$

(a) Estimate the population variance σ^2.
(b) Compute a 99 percent two-sided confidence interval for σ^2.
(c) Compute a value v that enables us to state, with 90 percent confidence, that σ^2 is less than v.

37. Find a 95 percent two-sided confidence interval for the variance of the diameter of a rivet based on the data given here.

6.68	6.66	6.62	6.72
6.76	6.67	6.70	6.72
6.78	6.66	6.76	6.72
6.76	6.70	6.76	6.76
6.74	6.74	6.81	6.66
6.64	6.79	6.72	6.82
6.81	6.77	6.60	6.72
6.74	6.70	6.64	6.78
6.70	6.70	6.75	6.79

Assume a normal population.

38. The following are independent samples from two normal populations, both of which have the same standard deviation σ.

$$16, 17, 19, 20, 18 \qquad \text{and} \qquad 3, 4, 8$$

Use them to estimate σ.

39. The amount of beryllium in a substance is often determined by the use of a photometric filtration method. If the weight of the beryllium is μ, then the value given by the photometric filtration method is normally distributed with mean μ and standard deviation σ. A total of eight independent measurements of 3.180 mg of beryllium gave the following results.

$$3.166, 3.192, 3.175, 3.180, 3.182, 3.171, 3.184, 3.177$$

Use the preceding data to

(a) estimate σ;

(b) find a 90 percent confidence interval estimate of σ.

40. If $X_1, \ldots, X_n$ is a sample from a normal population, explain how to obtain a $100(1-\alpha)$ percent confidence interval for the population variance σ^2 when the population mean μ is known. Explain in what sense knowledge of μ improves the interval estimator compared with when it is unknown.
Repeat Problem 38 if it is known that the mean burning time is 53.6 seconds.

41. A civil engineer wishes to measure the compressive strength of two different types of concrete. A random sample of 10 specimens of the first type yielded the following data (in psi)

Type 1:	3,250,	3,268,	4,302,	3,184,	3,266
	3,297,	3,332,	3,502,	3,064,	3,116

whereas a sample of 10 specimens of the second yielded the data

Type 2:	3,094,	3,106,	3,004,	3,066,	2,984,
	3,124,	3,316,	3,212,	3,380,	3,018

If we assume that the samples are normal with a common variance, determine

(a) a 95 percent two-sided confidence interval for $\mu_1 - \mu_2$, the difference in means;

(b) a 95 percent one-sided upper confidence interval for $\mu_1 - \mu_2$;

(c) a 95 percent one-sided lower confidence interval for $\mu_1 - \mu_2$.

42. Independent random samples are taken from the output of two machines on a production line. The weight of each item is of interest. From the first machine, a sample of size 36 is taken, with sample mean weight of 120 grams and a sample variance of 4. From the second machine, a sample of size 64 is taken, with a sample mean weight of 130 grams and a sample variance of 5. It is assumed that the weights of items from the first machine are normally distributed with mean μ_1 and variance σ^2 and that the weights of items from the second machine are normally distributed with mean μ_2 and variance σ^2 (that is, the variances are assumed to be equal). Find a 99 percent confidence interval for $\mu_1 - \mu_2$, the difference in population means.

43. Do Problem 42 when it is known in advance that the population variances are 4 and 5.

44. The following are the daily numbers of company website visits resulting from advertisements on two different types of media.

Type I		*Type II*	
481	572	526	537
506	561	511	582
527	501	556	605
661	487	542	558
501	524	491	578

Find a 99 percent confidence interval for the mean difference in daily visits assuming normality with unknown but equal variances.

45. If $X_1, \ldots, X_n$ is a sample from a normal population having known mean μ_1 and unknown variance σ_1^2, and $Y_1, \ldots, Y_m$ is an independent sample from a normal population having known mean μ_2 and unknown variance σ_2^2, determine a $100(1-\alpha)$ percent confidence interval for σ_1^2/σ_2^2.

46. Two analysts took repeated readings on the hardness of city water. Assuming that the readings of analyst i constitute a sample from a normal population having variance σ_i^2, $i = 1, 2$, compute a 95 percent two-sided confidence interval for σ_1^2/σ_2^2 when the data are as follows:

Coded Measures of Hardness	
Analyst 1	**Analyst 2**
.46	.82
.62	.61
.37	.89
.40	.51
.44	.33
.58	.48
.48	.23
.53	.25
	.67
	.88

47. A problem of interest in baseball is whether a sacrifice bunt is a good strategy when there is a man on first base and no outs. Assuming that the bunter will be out but will be successful in advancing the man on base, we could compare the probability of scoring a run with a player on first base and no outs with the probability of scoring a run with a player on second base and one out.

The following data resulted from a study of randomly chosen major league baseball games played in 1959 and 1960.

(a) Give a 95 percent confidence interval estimate for the probability of scoring at least one run when there is a man on first and no outs.

(b) Give a 95 percent confidence interval estimate for the probability of scoring at least one run when there is a man on second and one out.

Base Occupied	Number of Outs	Number of Cases in Which 0 Runs Are Scored	Total Number of Cases
First	0	1,044	1,728
Second	1	401	657

48. A random sample of 1,200 engineers included 48 Hispanic Americans, 80 African Americans, and 204 females. Determine 90 percent confidence intervals for the proportion of all engineers who are

(a) female;

(b) Hispanic Americans or African Americans.

49. To estimate p, the proportion of all newborn babies that are male, the gender of 10,000 newborn babies was noted. If 5,106 of them were male, determine **(a)** a 90 percent and **(b)** a 99 percent confidence interval estimate of p.

50. An airline is interested in determining the proportion of its customers who are flying for reasons of business. If they want to be 90 percent certain that their estimate will be correct to within 2 percent, how large a random sample should they select?

51. A recent newspaper poll indicated that Candidate A is favored over Candidate B by a 53 to 47 percentage, with a margin of error of ± 4 percent. The newspaper then stated that since the 6-point gap is larger than the margin of error, its readers can be certain that Candidate A is the current choice. Is this reasoning correct?

52. A market research firm is interested in determining the proportion of households that are watching a particular sporting event. To accomplish this task, they plan on using a telephone poll of randomly chosen households. How large a sample is needed if they want to be 90 percent certain that their estimate is correct to within $\pm .02$?

53. In a recent study, 79 of 140 meteorites were observed to enter the atmosphere with a velocity of less than 25 miles per second. If we take $\hat{p} = 79/140$ as an estimate of the probability that an arbitrary meteorite that enters the atmosphere will have a speed less than 25 miles per second, what can we say, with 99 percent confidence, about the maximum error of our estimate?

54. A random sample of 100 items from a production line revealed 17 of them to be defective. Compute a 95 percent two-sided confidence interval for the probability that an item produced is defective. Determine also a 99 percent upper confidence interval for this value. What assumptions are you making?

55. Of 100 randomly detected cases of individuals having lung cancer, 67 died within 5 years of detection.

(a) Estimate the probability that a person contracting lung cancer will die within 5 years.

(b) How large an additional sample would be required to be 95 percent confident that the error in estimating the probability in part (a) is less than .02?

56. Derive $100(1-\alpha)$ percent lower and upper confidence intervals for p, when the data consist of the values of n independent Bernoulli random variables with parameter p.

57. Suppose the lifetimes of batteries are exponentially distributed with mean θ. If the average of a sample of 10 batteries is 36 hours, determine a 95 percent two-sided confidence interval for θ.

58. Determine $100(1-\alpha)$ percent one-sided upper and lower confidence intervals for θ in Problem 57.

59. Let $X_1, X_2, \ldots, X_n$ denote a sample from a population whose mean value θ is unknown. Use the results of Example 7.7b to argue that among all unbiased estimators of θ of the form $\sum_{i=1}^{n} \lambda_i X_i$, $\sum_{i=1}^{n} \lambda_i = 1$, the one with minimal mean square error has $\lambda_i \equiv 1/n$, $i = 1, \ldots, n$.

60. Consider two independent samples from normal populations having the same variance σ^2, of respective sizes n and m. That is, $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ are independent samples from normal populations each having variance σ^2. Let S_x^2 and S_y^2 denote the respective sample variances. Thus both S_x^2 and S_y^2 are unbiased estimators of σ^2. Show by using the results of Example 7.7b along with the fact that

$$\mathrm{Var}(\chi_k^2) = 2k$$

where χ_k^2 is chi-square with k degrees of freedom, that the minimum mean square estimator of σ^2 of the form $\lambda S_x^2 + (1-\lambda)S_y^2$ is

$$S_p^2 = \frac{(n-1)S_x^2 + (m-1)S_y^2}{n+m-2}$$

This is called the *pooled estimator* of σ^2.

61. Consider two estimators d_1 and d_2 of a parameter θ. If $E[d_1] = \theta$, $\text{Var}(d_1) = 6$ and $E[d_2] = \theta + 2$, $\text{Var}(d_2) = 2$, which estimator should be preferred?

62. Suppose that the number of accidents occurring daily in a certain plant has a Poisson distribution with an unknown mean λ. Based on previous experience in similar industrial plants, suppose that a statistician's initial feelings about the possible value of λ can be expressed by an exponential distribution with parameter 1. That is, the prior density is

$$p(\lambda) = e^{-\lambda}, \qquad 0 < \lambda < \infty$$

Determine the Bayes estimate of λ if there are a total of 83 accidents over the next 10 days. What is the maximum likelihood estimate?

63. The functional lifetimes in hours of computer chips produced by a certain semiconductor firm are exponentially distributed with mean $1/\lambda$. Suppose that the prior distribution on λ is the gamma distribution with density function

$$g(x) = \frac{e^{-x}x^2}{2}, \qquad 0 < x < \infty$$

If the average life of the first 20 chips tested is 4.6 hours, compute the Bayes estimate of λ.

64. Each item produced will, independently, be defective with probability p. If the prior distribution on p is uniform on (0, 1), compute the posterior probability that p is less than .2 given

(a) a total of 2 defectives out of a sample of size 10;
(b) a total of 1 defective out of a sample of size 10;
(c) a total of 10 defectives out of a sample of size 10.

65. The breaking strength of a certain type of cloth is to be measured for 10 specimens. The underlying distribution is normal with unknown mean θ but with a standard deviation equal to 3 psi. Suppose also that based on previous experience we feel that the unknown mean has a prior distribution that is normally distributed with mean 200 and standard deviation 2. If the average breaking strength of a sample of 20 specimens is 182 psi, determine a region that contains θ with probability .95.