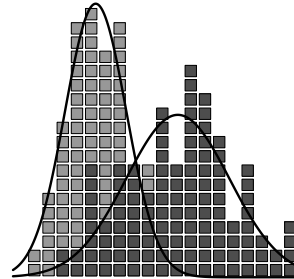# Chapter 3

Interval Estimation

One inherent weakness associated with point estimators is that they do not quantify the precision of the estimator. As an illustration, assume that the annual salaries of n people are collected, yielding the sample mean annual salary

$$\bar{x} = \$35,281.$$

You would certainly have much more faith in this point estimate of the population mean annual salary μ if $n = 500$ salaries were collected than if only $n = 5$ salaries were collected. It would be beneficial to augment a point estimator $\hat{\mu}$ with supplemental information concerning its precision. Giving the value of the sample size n, or even better $V[\hat{\mu}]$, would be helpful, but statisticians tend to use an interval (sometimes known as a statistical interval), primarily a special type of interval known as a confidence interval, to quantify the uncertainty associated with a point estimator. The four sections of this chapter consider exact confidence intervals, approximate confidence intervals, asymptotically exact confidence intervals, and other types of intervals.

3.1   Exact Confidence Intervals

Edward Russo and Paul Schoemaker designed the following ten-question quiz.

1. Martin Luther King's age at death.

2. Length of the Nile River in miles.

3. Number of countries that are members of OPEC.

4. Number of books in the Old Testament.

5. Diameter of the moon in miles.

6. Weight of an empty Boeing 747 in pounds.

7. Year in which Wolfgang Amadeus Mozart was born.

8. Gestation period, in days, of an Asian elephant.

9. Air distance in miles from London to Tokyo.

10. Deepest known point in the oceans in feet.

If you were to simply guess at an answer to each of these ten questions, you would be providing ten *point estimates* of the true answers, which was the topic of the last chapter. But this was not what Russo and Schoemaker had in mind. Instead, the authors asked for a lower bound and upper bound for the answer to each question, chosen so that you are 90 percent confident that the correct answer will fall in the interval that you specify. Provide ten intervals that are wide enough so that the probability an individual interal contains the correct answer is 0.9. Go ahead and take the quiz now before reading on; the answers are given in the preface at the beginning of the Chapter 3 reference material. Count how many of your intervals contain the correct answer.

Russo and Schoemaker had more than 1000 American and European managers take the quiz. If the managers had accurate intervals that correctly expressed their uncertainty with their knowledge, you would expect that many of them would have exactly nine of the ten intervals contain the correct answers. But that was not the case. Less than 1 percent of the managers had nine or ten of the intervals contain the correct answers. Most managers had between three and six of their intervals contain the correct answers. What happened? The managers were *overconfident* in choosing their lower and upper bounds to their intervals, and chose intervals that were too narrow.

Russo and Schoemaker's quiz provides a good illustration of the use of an interval to express the *precision* associated with how much faith we have in our point estimator. They were assessing the ability to accurately measure how well you know what you don't know. To use some of the language that will be introduced subsequently in this chapter, although the *stated* coverage of the confidence intervals was 90%, the *actual* coverage turned out to be much lower than 90%, somewhere between 30% and 60%. The confidence intervals that the managers gave were *approximate* 90% confidence intervals rather than *exact* 90% confidence intervals.

A confidence interval gives an estimate of the precision associated with a point estimate. An exact confidence interval is the preferred type of confidence interval.

Definition 3.1 A random interval of the form

$$L < \theta < U$$

is an *exact two-sided* $100(1-\alpha)\%$ *confidence interval* for the unknown parameter θ provided

$$P(L < \theta < U) = 1 - \alpha$$

for all values of θ.

Some comments regarding exact confidence intervals are listed below.

- The confidence interval $L < \theta < U$ is a random interval because the endpoints of the interval, L and U, are random variables. It is for this reason that they are set in upper case.

- The random variable L is known as the *lower bound* of the confidence interval.

- The random variable U is known as the *upper bound* of the confidence interval.

- The probability $1 - \alpha$ will be referred to as the *stated coverage*. Other terms to describe this quantity are the nominal coverage, confidence level, and the confidence coefficient.

- Although α can assume any value between 0 and 1, the most common values used by statisticians are 0.1, 0.05, and 0.01. The confidence intervals associated with these values of α are referred to as 90%, 95%, and 99% confidence intervals, respectively. Because of the increased confidence, smaller values of α result in wider confidence intervals for identical data values.

- The notation in Definition 3.1 has been radically oversimplified. The random variables L and U are typically functions of the sample size n, the stated coverage $1 - \alpha$, and the data values $X_1, X_2, \ldots, X_n$. Since the notation

$$L = g_1(n, \alpha, X_1, X_2, \ldots, X_n) \qquad \text{and} \qquad U = g_2(n, \alpha, X_1, X_2, \ldots, X_n)$$

  is cumbersome, we let just L and U suffice.

- As was the case with point estimates, a confidence interval associated with a random sample drawn from a population is not necessarily unique. There might be several exact confidence intervals from which to choose. Criteria that can be used to select among them will be introduced in this chapter.

- The confidence interval defined in Definition 3.1 is exact because $P(L < \theta < U)$ is exactly $1 - \alpha$. But not all confidence intervals are exact. Approximate confidence intervals and asymptotically exact confidence intervals will be introduced in the next two sections.

- There are three ways to report a confidence interval. The first follows Definition 3.1:

$$L < \theta < U.$$

  This can also be reported as the open interval

$$(L, U).$$

  Finally, when the confidence interval bounds are symmetric about the point estimator $\hat{\theta}$, the confidence interval can be reported as
$$\hat{\theta} \pm H,$$
  where $H = (U - L)/2$ is the confidence interval *halfwidth*. The advantage of expressing a confidence interval in this format is that the point estimate $\hat{\theta}$ and the confidence interval halfwidth H are both immediately apparent.

- Reporting a confidence interval in the form $L < \theta < U$ does not imply that θ is between L and U. For this reason, it might be clearer writing a confidence interval as simply (L, U).

- Using the experimental values $x_1, x_2, \ldots, x_n$ of the random variables $X_1, X_2, \ldots, X_n$, the confidence intervals can be written in lower case:

$$l < \theta < u, \qquad (l, u), \qquad \text{or} \qquad \hat{\theta} \pm h.$$

- The confidence interval in Definition 3.1 is a two-sided confidence interval. Some applications require a *one-sided confidence interval* when the interest is only in an upper or lower bound on a parameter A one-sided confidence interval of the form $(-\infty, U)$, which is typically reported as $\theta < U$, gives only an upper bound on the parameter θ. Conversely, a one-sided confidence interval of the form (L, ∞), which is typically reported as $\theta > L$, gives only a lower bound on the parameter θ. One-sided confidence intervals will be illustrated subsequently. The default assumption for confidence intervals is that they are two-sided.

- As was the case with point estimates, care must be taken to report the confidence interval bounds to the appropriate number of significant digits for a particular data set.

   We begin with three examples of deriving an exact two-sided confidence interval from $n = 1$ observation drawn from a continuous population. Statisticians seldom encounter a data set with just a single observation, but if data are expensive (for example, if it involves crashing a car into a wall) or difficult to collect (for example, if it involves an experiment conducted by a rover on the moon), reporting a point and interval estimate from a single observation might be appropriate.

Example 3.1 Consider a single ($n = 1$) observation X drawn from a population with probability density function

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

where θ is a positive unknown parameter.

   (a) Find the method of moments and the maximum likelihood estimators of θ.

   (b) Find an exact two-sided 90% confidence interval for θ.

   (c) Calculate the point estimators and the 90% confidence interval for the data value $x = 0.63$.

The usual notation for a random sample is $X_1, X_2, \ldots, X_n$, but since $n = 1$, the subscripts are dropped entirely to keep the notation as simple as possible. This population distribution is a special case of the beta distribution with its second parameter equal to one. The population probability density function is plotted for several values of θ in Figure 3.1. It is clear from examining these curves that it would be preferable for a point estimate $\hat{\theta}$ to be monotonically increasing with the data value X, that is, small values of X should result in small values of $\hat{\theta}$ and large values of X should result in large values of $\hat{\theta}$. The population mean of this distribution is

$$E[X] = \int_0^1 \theta x^\theta \, dx = \left[\frac{\theta x^{\theta+1}}{\theta+1} \right]_0^1 = \frac{\theta}{\theta+1}.$$

The cumulative distribution function of the population distribution on the support of X is

$$F(x) = \int_0^x \theta w^{\theta - 1} \, dw = \left[w^\theta \right]_0^x = x^\theta \qquad 0 < x < 1.$$
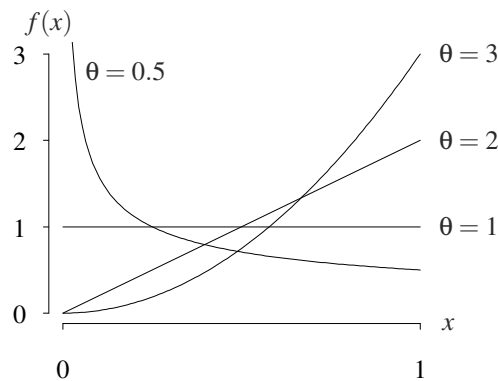

Figure 3.1: Population probability density functions.

(a) Since there is only a single positive unknown parameter θ, the population mean should be equated to the sample mean to determine the method of moments estimator. Since the sample mean for $n = 1$ is just the value of the single observation x, this equation is

$$\frac{\theta}{\theta + 1} = x.$$

Solving for θ gives the method of moments estimate

$$\hat{\theta}_1 = \frac{x}{1 - x}.$$

The subscript on $\hat{\theta}$ is needed to distinguish this estimate from the maximum likelihood estimate. To determine the maximum likelihood estimate, the likelihood function for the single observation is just the probability density function:

$$L(\theta) = f(x) = \theta x^{\theta - 1}.$$

The log likelihood function is

$$\ln L(\theta) = \ln \theta + (\theta - 1) \ln x.$$

The score is

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{1}{\theta} + \ln x.$$

Equating the score to zero and solving for θ gives the maximum likelihood estimate

$$\hat{\theta}_2 = -\frac{1}{\ln x}.$$

Figure 3.2 shows these two point estimates as functions of the data value x. There is a vertical asymptote at $x = 1$. Both estimates are monotonically increasing in x as desired. The maximum likelihood estimate $\hat{\theta}_2$ is slightly larger than the method of moments estimator $\hat{\theta}_1$ for the same value of x.
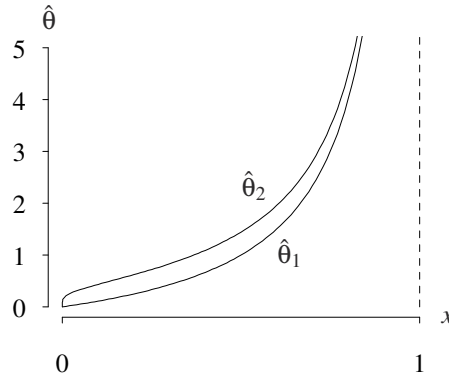

Figure 3.2: Method of moments and maximum likelihood estimators of θ.

(b) In our quest to find an exact two-sided 90% confidence interval for θ, we can begin with the fact that the probability that X lies between a and b can be calculated with

$$P(a < X < b) = \int_a^b f(x)\,dx,$$

where a and b are real constants satisfying $0 < a < b < 1$. Integration is used because the population probability distribution is continuous. Since this is to be an exact two-sided 90% confidence interval, this probability can be set to 0.9:

$$P(a < X < b) = \int_a^b f(x)\,dx = 0.9.$$

This probability statement is not yet of the form $P(L < \theta < U) = 0.9$ as in Definition 3.1, but this can be overcome. Unfortunately, there are an infinite number of pairs of a and b values satisfying this equation. One solution to the problem is to find a and b that minimize $b - a$, that is, the narrowest possible confidence interval. Another solution is to let a be the 5th percentile of the probability distribution of X and let b be the 95th percentile of the probability distribution of X. It is this second solution that is almost universally taken by statisticians. Using the cumulative distribution function of the population distribution, these percentiles can be found by solving

$$F(a) = a^\theta = 0.05$$

and

$$F(b) = b^\theta = 0.95$$

for a and b, which results in

$$a = 0.05^{1/\theta} \qquad \text{and} \qquad b = 0.95^{1/\theta}.$$

So the probability statement $P(a < X < b) = 0.9$ with this choice of a and b becomes

$$P\left(0.05^{1/\theta} < X < 0.95^{1/\theta}\right) = 0.9.$$

The next step is to perform algebraic operations on the inequality so as to place θ in the center of the inequality. Taking the natural logarithm of all parts of the inequality results in

$$P\left(\frac{1}{\theta} \ln 0.05 < \ln X < \frac{1}{\theta} \ln 0.95\right) = 0.9.$$

Finally, multiplying by θ (which is positive) and dividing by $\ln X$ (which is negative because the support of X is $0 < x < 1$),

$$P\left(\frac{\ln 0.95}{\ln X} < \theta < \frac{\ln 0.05}{\ln X}\right) = 0.9.$$

So an exact two-sided 90% confidence interval for the unknown parameter θ is

$$\frac{\ln 0.95}{\ln X} < \theta < \frac{\ln 0.05}{\ln X}.$$

This confidence interval is a random interval in that it is expressed in terms of the random variable X. For a specific data value x, it is written as

$$\frac{\ln 0.95}{\ln x} < \theta < \frac{\ln 0.05}{\ln x}.$$

(c) For the specific single observation $x = 0.63$, the two point estimators are

$$\hat{\theta}_1 = \frac{0.63}{1 - 0.63} = 1.70 \qquad \text{and} \qquad \hat{\theta}_2 = -\frac{1}{\ln 0.63} = 2.16.$$

The exact two-sided 90% confidence interval for θ is

$$\frac{\ln 0.95}{\ln 0.63} < \theta < \frac{\ln 0.05}{\ln 0.63}$$

or

$$0.11 < \theta < 6.48.$$

Since there is just one observation, the 90% confidence interval is quite wide, as one would expect. We simply cannot expect much precision on $\hat{\theta}$ from just a single observation. The point estimates and the confidence interval bounds were all reported to two digits to the right of the decimal point, which is consistent with the data value.

The initial example of an exact confidence interval considered a single observation X drawn from a population distribution with support $0 < x < 1$. The second example of an exact confidence interval considers a single observation X drawn from a population distribution with positive support $x > 0$. Population distributions of this type are used in a field known as *survival analysis*, where the interest is in a positive random variable, such as the remission time for a cancer, the duration of a strike by a labor union, or the lifetime of a coil spring.

Example 3.2 Consider a single ($n = 1$) observation X drawn from a population whose cumulative distribution function on its support is

$$F(x) = \frac{x^2}{\theta^2 + x^2} \qquad x > 0,$$

where θ is a positive unknown parameter.

(a) Find the method of moments and the maximum likelihood estimators of θ and determine which is the preferred point estimator.

(b) Find an exact two-sided 80% confidence interval for θ.

Writing the cumulative distribution function of the population distribution as

$$F(x) = \frac{(x/\theta)^2}{1 + (x/\theta)^2} \qquad x > 0$$

emphasizes the fact that θ is a scale parameter of the population probability distribution. This population distribution is a special case of a *log logistic distribution*. The probability density function can be found by differentiating the cumulative distribution function with respect to x:

$$f(x) = \frac{2\theta^2 x}{\left(\theta^2 + x^2\right)^2} \qquad x > 0.$$

Leaving out the integration details, the population mean of the population distribution is

$$E[X] = \int_0^\infty x \cdot \frac{2\theta^2 x}{\left(\theta^2 + x^2\right)^2}\, dx = \frac{\pi\theta}{2}.$$

(a) Since there is only a single unknown parameter, the method of moments estimate of θ is found by equating the population mean to the sample mean, which is just x in this case:

$$\frac{\pi\theta}{2} = x.$$

Solving for θ gives the method of moments estimate

$$\hat{\theta}_1 = \frac{2x}{\pi}.$$

Switching the notation associated with this point estimator to upper case, the expected value of the method of moments estimator is

$$E\left[\hat{\theta}_1\right] = E\left[\frac{2X}{\pi}\right] = \frac{2}{\pi}E\left[X\right] = \frac{2}{\pi} \cdot \frac{\pi\theta}{2} = \theta,$$

which indicates that the method of moments estimator is an unbiased estimator of θ. For the maximum likelihood estimator, the likelihood function is again just the probability density function because there is only a single observation:

$$L(\theta) = f(x) = \frac{2\theta^2 x}{\left(\theta^2 + x^2\right)^2}.$$

The log likelihood function is

$$\ln L(\theta) = \ln 2 + 2\ln\theta + \ln x - 2\ln\left(\theta^2 + x^2\right).$$

The score is

$$\frac{\partial \ln L(\theta)}{\partial \theta} = \frac{2}{\theta} - \frac{4\theta}{\theta^2 + x^2}.$$

Equating the score to zero and solving for θ results in the maximum likelihood estimate

$$\hat{\theta}_2 = x.$$

Switching to upper case, the expected value of the maximum likelihood estimator is

$$E\left[\hat{\theta}_2\right] = E[X] = \frac{\pi\theta}{2},$$

so the maximum likelihood estimator is a biased estimator of θ. For this reason, the method of moments estimator is the preferred point estimator. Since the maximum likelihood estimator is a constant times the method of moments estimator, that is, $\hat{\theta}_2 = (\pi/2)\hat{\theta}_1$, the method of moments estimator is equivalent to the maximum likelihood estimator with an unbiasing constant. The subscript on the point estimator will be dropped and the method of moments estimate will be referred to as just $\hat{\theta} = 2x/\pi$.

A Monte Carlo simulation experiment can be used to assess the random sampling distribution of $\hat{\theta}$ for one particular value of the population parameter θ. Random variates from the population distribution are generated by inverting the cumulative distribution function of the population:

$$F^{-1}(u) = \theta\sqrt{\frac{u}{1-u}} \qquad 0 < u < 1,$$

where u is a random number. Consider 100 replications of the point estimator for a population with θ arbitrarily set to 5. The R code below generates and plots 100 point estimators.

```
theta = 5
nrep = 100
u = runif(nrep)
x = theta * sqrt(u / (1 - u))
thetahat = 2 * x / pi
plot(1:nrep, thetahat)
```

After a call to `set.seed(20)` to initialize the random number seed, the code is executed and the resulting graph is shown in Figure 3.3, which includes a spectacular $\hat{\theta} = 18.3$ in the 12th replication. Since the method of moments estimator $\hat{\theta}$ is unbiased, the 100 point estimators hover about $\theta = 5$, which is indicated by the horizontal line. The distribution of $\hat{\theta}$ is not symmetric; it has a long right-hand tail. The sample mean of the 100 simulated point estimates is 4.2, which is slightly lower than expected.

(b) The technique for determining an exact two-sided 80% confidence interval for θ will be similar to the technique used in the previous example. We begin by choosing constants a and b such that the probability that X lies between a and b is 0.8:

$$P(a < X < b) = \int_a^b f(x)\,dx = 0.8.$$

As before, there are an infinite number of pairs of a and b satisfying this equation, but common statistical practice is to choose a to be the 10th percentile of the population distribution and b to be the 90th percentile of the population distribution. These percentiles are found by solving

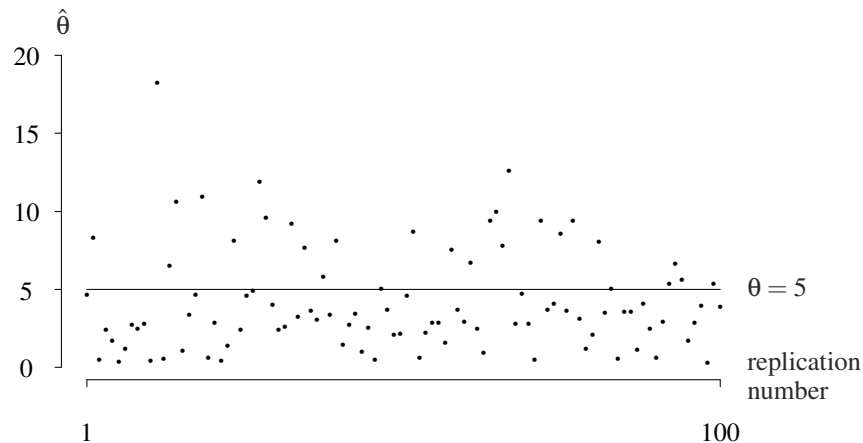$$F(a) = \frac{a^2}{\theta^2 + a^2} = 0.1.$$

Figure 3.3: Monte Carlo simulation of 100 point estimates $\hat{\theta}$.

and

$$F(b) = \frac{b^2}{\theta^2 + b^2} = 0.9$$

for *a* and *b*, which results in

$$a = \sqrt{\frac{0.1}{0.9}}\,\theta = \frac{\theta}{3} \qquad \text{and} \qquad b = \sqrt{\frac{0.9}{0.1}}\,\theta = 3\theta.$$

So the probability statement $P(a < X < b) = 0.8$ becomes

$$P\left(\frac{\theta}{3} < X < 3\theta\right) = 0.8.$$

Dividing the inequality by θ, then dividing the inequality by X (both of which are positive), then inverting, θ now appears at the center of the inequality:

$$P\left(\frac{X}{3} < \theta < 3X\right) = 0.8.$$

So we have achieved a probability statement of the form $P(L < \theta < U)$. Thus, an exact two-sided 80% confidence interval for the unknown parameter θ is

$$\frac{X}{3} < \theta < 3X.$$

For a particular observation *x*, the 80% confidence interval is

$$\frac{x}{3} < \theta < 3x.$$

Monte Carlo simulation can again be used to evaluate the confidence interval procedure. Using the same random number stream as in Figure 3.3, 100 confidence intervals are generated from a population distribution with $\theta = 5$. Confidence intervals are drawn as vertical lines that intersect the point estimators in Figure 3.4. Several conclusions can immediately be drawn from Figure 3.4: (*a*) the confidence interval bounds are highly nonsymmetric about the point estimator, (*b*) the widths of the confidence intervals increase with $\hat{\theta}$, and (*c*) twice as many confidence intervals fall below $\theta = 5$ as those that fall above $\theta = 5$ for this particular simulation. Since 79 out of the 100 simulated confidence intervals contained the population parameter $\theta = 5$, the Monte Carlo simulation supports our analytic derivation of the exact 80% confidence interval. The number of confidence intervals that cover θ will vary from one simulation experiment to the next. The number of exact two-sided 80% confidence intervals that contain θ in the Monte Carlo experiment with 100 replications is a binomial random variable with parameters $n = 100$ and $p = 0.8$. The number of exact two-sided 80% confidence intervals that miss high and the number that miss low are binomial random variables with $n = 100$ and $p = 0.1$, although the three binomial random variables are correlated (the number of confidence intervals that miss low, cover $\theta = 5$, and miss high has the trinomial distribution). Care should be taken to not draw any meaningful conclusions from a Monte Carlo simulation experiment with only 100

$\hat{\theta}$

Confidence interval summary:
7 intervals missed high
79 intervals covered $\theta = 5$
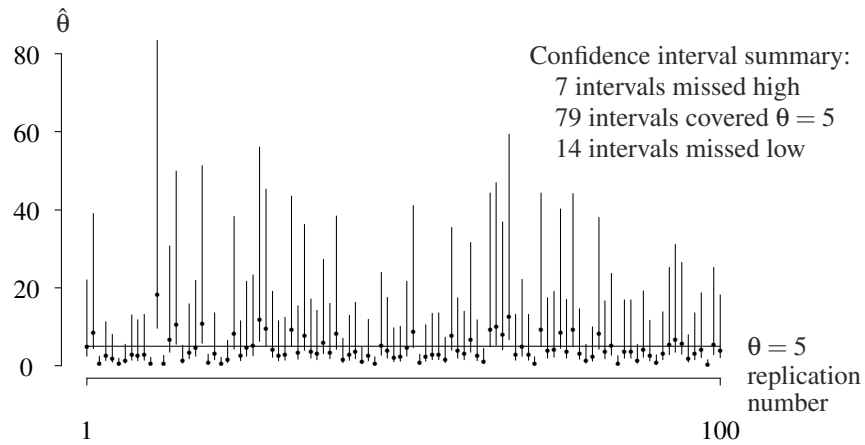14 intervals missed low

$\theta = 5$
replication
number

Figure 3.4: Monte Carlo simulation of 100 exact 80% confidence intervals.

replications. To that end, the number of replications was increased to one million, and the simulation yielded the following fractions of the confidence intervals generated that covered $\theta = 5$ for five runs:

$$0.7995 \qquad 0.7997 \qquad 0.8001 \qquad 0.8003 \qquad 0.8002.$$

These Monte Carlo simulation results provide strong evidence to support the exact coverage probability of the two-sided 80% confidence interval. In these simulation experiments, approximately 10% of the confidence intervals miss low and approximately 10% miss high. The fact that twice as many intervals missed low as those that missed high in Figure 3.4 is due to random sampling variability.

The Monte Carlo simulation in the previous example has highlighted the importance of the *width* of an exact confidence interval, which is defined next.

Definition 3.2 For the two-sided confidence interval $L < \theta < U$, the random confidence interval *width* is
$$W = U - L$$
and the random confidence interval *halfwidth* is
$$H = \frac{U - L}{2}.$$

Since the confidence interval width W and the confidence interval halfwidth H are random variables, their expected values and population variances are of interest. For any confidence interval, being exact, that is, $P(L < \theta < U) = 1 - \alpha$ for all values of θ and all sample sizes n, is a primary goal. For an exact confidence interval, one would prefer a small value of $E[W]$ or $E[H]$ as a secondary criterion because this is an indicator of increased precision. A tertiary criterion would be to have a small value of $V[W]$ or $V[H]$ because this indicates a stable width or halfwidth to the interval. The next example again considers the formulation of an exact two-sided confidence interval in the case