We present some methods for finding estimators with small variances later in this chapter. For now we wish only to point out that relative efficiency is one important criterion for comparing estimators.

Exercises

9.1 In Exercise 8.8, we considered a random sample of size 3 from an exponential distribution with density function given by

$$f(y) = \begin{cases} (1/\theta)e^{-y/\theta}, & 0 < y, \\ 0, & \text{elsewhere}, \end{cases}$$

and determined that $\hat{\theta}_1 = Y_1, \hat{\theta}_2 = (Y_1 + Y_2)/2, \hat{\theta}_3 = (Y_1 + 2Y_2)/3$, and $\hat{\theta}_5 = \overline{Y}$ are all unbiased estimators for θ. Find the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_5$, of $\hat{\theta}_2$ relative to $\hat{\theta}_5$, and of $\hat{\theta}_3$ relative to $\hat{\theta}_5$.

9.2 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a population with mean μ and variance σ^2. Consider the following three estimators for μ:

$$\hat{\mu}_1 = \frac{1}{2}(Y_1 + Y_2), \qquad \hat{\mu}_2 = \frac{1}{4}Y_1 + \frac{Y_2 + \cdots + Y_{n-1}}{2(n-2)} + \frac{1}{4}Y_n, \qquad \hat{\mu}_3 = \overline{Y}.$$

a Show that each of the three estimators is unbiased.

b Find the efficiency of $\hat{\mu}_3$ relative to $\hat{\mu}_2$ and $\hat{\mu}_1$, respectively.

9.3 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the uniform distribution on the interval $(\theta, \theta + 1)$. Let

$$\hat{\theta}_1 = \overline{Y} - \frac{1}{2} \quad \text{and} \quad \hat{\theta}_2 = Y_{(n)} - \frac{n}{n+1}.$$

a Show that both $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators of θ.

b Find the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$.

9.4 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size n from a uniform distribution on the interval $(0, \theta)$. If $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$, the result of Exercise 8.18 is that $\hat{\theta}_1 = (n+1)Y_{(1)}$ is an unbiased estimator for θ. If $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$, the results of Example 9.1 imply that $\hat{\theta}_2 = [(n+1)/n]Y_{(n)}$ is another unbiased estimator for θ. Show that the efficiency of $\hat{\theta}_1$ to $\hat{\theta}_2$ is $1/n^2$. Notice that this implies that $\hat{\theta}_2$ is a markedly superior estimator.

9.5 Suppose that $Y_1, Y_2, \ldots, Y_n$ is a random sample from a normal distribution with mean μ and variance σ^2. Two unbiased estimators of σ^2 are

$$\hat{\sigma}_1^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y})^2 \quad \text{and} \quad \hat{\sigma}_2^2 = \frac{1}{2}(Y_1 - Y_2)^2.$$

Find the efficiency of $\hat{\sigma}_1^2$ relative to $\hat{\sigma}_2^2$.

9.6 Suppose that $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size n from a Poisson distribution with mean λ. Consider $\hat{\lambda}_1 = (Y_1 + Y_2)/2$ and $\hat{\lambda}_2 = \overline{Y}$. Derive the efficiency of $\hat{\lambda}_1$ relative to $\hat{\lambda}_2$.

9.7 Suppose that $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size n from an exponential distribution with density function given by

$$f(y) = \begin{cases} (1/\theta)e^{-y/\theta}, & 0 < y, \\ 0, & \text{elsewhere}. \end{cases}$$

***8.16** Suppose that $Y_1, Y_2, \ldots, Y_n$ constitute a random sample from a normal distribution with parameters μ and σ^2.[1]

a Show that $S = \sqrt{S^2}$ is a biased estimator of σ. [*Hint:* Recall the distribution of $(n-1)S^2/\sigma^2$ and the result given in Exercise 4.112.]

b Adjust S to form an unbiased estimator of σ.

c Find an unbiased estimator of $\mu - z_\alpha \sigma$, the point that cuts off a lower-tail area of α under this normal curve.

8.17 If Y has a binomial distribution with parameters n and p, then $\hat{p}_1 = Y/n$ is an unbiased estimator of p. Another estimator of p is $\hat{p}_2 = (Y+1)/(n+2)$.

a Derive the bias of $\hat{p}_2$.

b Derive MSE($\hat{p}_1$) and MSE($\hat{p}_2$).

c For what values of p is MSE($\hat{p}_1$) $<$ MSE($\hat{p}_2$)?

8.18 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size n from a population with a uniform distribution on the interval $(0, \theta)$. Consider $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$, the smallest-order statistic. Use the methods of Section 6.7 to derive $E(Y_{(1)})$. Find a multiple of $Y_{(1)}$ that is an unbiased estimator for θ.

8.19 Suppose that $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size n from a population with an exponential distribution whose density is given by

$$f(y) = \begin{cases} (1/\theta)e^{-y/\theta}, & y > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

If $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$ denotes the smallest-order statistic, show that $\hat{\theta} = nY_{(1)}$ is an unbiased estimator for θ and find MSE($\hat{\theta}$). [*Hint:* Recall the results of Exercise 6.81.]

***8.20** Suppose that $Y_1,\ Y_2,\ Y_3,\ Y_4$ denote a random sample of size 4 from a population with an exponential distribution whose density is given by

$$f(y) \begin{cases} (1/\theta)e^{-y/\theta}, & y > 0, \\ 0, & \text{elsewhere.} \end{cases}$$

a Let $X = \sqrt{Y_1 Y_2}$. Find a multiple of X that is an unbiased estimator for θ. [*Hint:* Use your knowledge of the gamma distribution and the fact that $\Gamma(1/2) = \sqrt{\pi}$ to find $E(\sqrt{Y_1})$. Recall that the variables Y_i are independent.]

b Let $W = \sqrt{Y_1 Y_2 Y_3 Y_4}$. Find a multiple of W that is an unbiased estimator for θ^2. [Recall the hint for part (a).]

8.3 Some Common Unbiased Point Estimators

Some formal methods for deriving point estimators for target parameters are presented in Chapter 9. In this section, we focus on some estimators that merit consideration on the basis of intuition. For example, it seems natural to use the sample mean

1. Exercises preceded by an asterisk are optional.

a Estimate the mean number λ_A of nucleation sites for regime A and place a 2-standard-error bound on the error of estimation.

b Estimate the difference in the mean numbers of nucleation sites $\lambda_A - \lambda_B$ for regimes A and B. Place a 2-standard-error bound on the error of estimation. Would you say that regime B tends to produce a larger mean number of nucleation sites? Why?

8.36 If $Y_1, Y_2, \ldots, Y_n$ denote a random sample from an exponential distribution with mean θ, then $E(Y_i) = \theta$ and $V(Y_i) = \theta^2$. Thus, $E(\overline{Y}) = \theta$ and $V(\overline{Y}) = \theta^2/n$, or $\sigma_{\overline{Y}} = \theta/\sqrt{n}$. Suggest an unbiased estimator for θ and provide an estimate for the standard error of your estimator.

8.37 Refer to Exercise 8.36. An engineer observes $n = 10$ independent length-of-life measurements on a type of electronic component. The average of these 10 measurements is 1020 hours. If these lengths of life come from an exponential distribution with mean θ, estimate θ and place a 2-standard-error bound on the error of estimation.

8.38 The number of persons coming through a blood bank until the first person with type A blood is found is a random variable Y with a geometric distribution. If p denotes the probability that any one randomly selected person will possess type A blood, then $E(Y) = 1/p$ and $V(Y) = (1-p)/p^2$.

a Find a function of Y that is an unbiased estimator of $V(Y)$.

b Suggest how to form a 2-standard-error bound on the error of estimation when Y is used to estimate $1/p$.

8.5 Confidence Intervals

An *interval estimator* is a rule specifying the method for using the sample measurements to calculate two numbers that form the endpoints of the interval. Ideally, the resulting interval will have two properties: First, it will contain the target parameter θ; second, it will be relatively narrow. One or both of the endpoints of the interval, being functions of the sample measurements, will vary randomly from sample to sample. Thus, the length and location of the interval are random quantities, and we cannot be certain that the (fixed) target parameter θ will fall between the endpoints of any single interval calculated from a single sample. This being the case, our objective is to find an interval estimator capable of generating narrow intervals that have a high probability of enclosing θ.

Interval estimators are commonly called *confidence intervals*. The upper and lower endpoints of a confidence interval are called the *upper* and *lower confidence limits*, respectively. The probability that a (random) confidence interval will enclose θ (a fixed quantity) is called the *confidence coefficient*. From a practical point of view, the confidence coefficient identifies the fraction of the time, in repeated sampling, that the intervals constructed will contain the target parameter θ. If we know that the confidence coefficient associated with our estimator is high, we can be highly confident that any confidence interval, constructed by using the results from a single sample, will enclose θ.

Suppose that $\hat{\theta}_L$ and $\hat{\theta}_U$ are the (random) lower and upper confidence limits, respectively, for a parameter θ. Then, if

$$P\left(\hat{\theta}_L \le \theta \le \hat{\theta}_U\right) = 1 - \alpha,$$

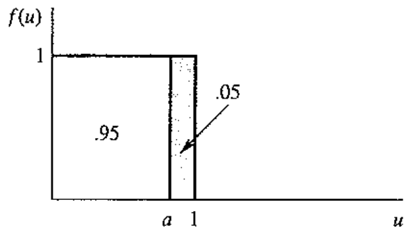

FIGURE 8.6 Density function for U, Example 8.5

Figure 8.6 contains a graph of the density function for U. Again, we see that U satisfies the requirements of a pivotal quantity. Because we seek a 95% lower confidence limit for θ, let us determine the value for a so that $P(U \leq a) = .95$. That is,

$$\int_0^a (1)\, du = .95,$$

or $a = .95$. Thus,

$$P(U \leq .95) = P\left(\frac{Y}{\theta} \leq .95\right) = P(Y \leq .95\theta) = P\left(\frac{Y}{.95} \leq \theta\right) = .95.$$

We see that $Y/.95$ is a lower confidence limit for θ, with confidence coefficient .95. Because any observed Y must be less than θ, it is intuitively reasonable to have the lower confidence limit for θ slightly larger than the observed value of Y. ■

The two preceding examples illustrate the use of the pivotal method for finding confidence limits for unknown parameters. In each instance, the interval estimates were developed on the basis of a single observation from the distribution. These examples were introduced primarily to illustrate the pivotal method. In the remaining sections of this chapter, we use this method in conjunction with the sampling distributions presented in Chapter 7 to develop some interval estimates of greater practical importance.

Exercises

8.39 Suppose that the random variable Y has a gamma distribution with parameters $\alpha = 2$ and an unknown β. In Exercise 6.46, you used the method of moment-generating functions to prove a general result implying that $2Y/\beta$ has a χ^2 distribution with 4 degrees of freedom (df). Using $2Y/\beta$ as a pivotal quantity, derive a 90% confidence interval for β.

8.40 Suppose that the random variable Y is an observation from a normal distribution with unknown mean μ and variance 1. Find a

a 95% confidence interval for μ.

b 95% upper confidence limit for μ.

c 95% lower confidence limit for μ.

8.41 Suppose that Y is normally distributed with mean 0 and unknown variance σ^2. Then Y^2/σ^2 has a χ^2 distribution with 1 df. Use the pivotal quantity Y^2/σ^2 to find a

a 95% confidence interval for σ^2.
b 95% upper confidence limit for σ^2.
c 95% lower confidence limit for σ^2.

8.42 Use the answers from Exercise 8.41 to find a

a 95% confidence interval for σ.
b 95% upper confidence limit for σ.
c 95% lower confidence limit for σ.

8.43 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size n from a population with a uniform distribution on the interval $(0, \theta)$. Let $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$ and $U = (1/\theta)Y_{(n)}$.

a Show that U has distribution function

$$F_U(u) = \begin{cases} 0, & u < 0, \\ u^n, & 0 \le u \le 1, \\ 1, & u > 1. \end{cases}$$

b Because the distribution of U does not depend on θ, U is a pivotal quantity. Find a 95% lower confidence bound for θ.

8.44 Let Y have probability density function

$$f_Y(y) = \begin{cases} \dfrac{2(\theta - y)}{\theta^2}, & 0 < y < \theta, \\ 0, & \text{elsewhere.} \end{cases}$$

a Show that Y has distribution function

$$F_Y(y) = \begin{cases} 0, & y \le 0, \\ \dfrac{2y}{\theta} - \dfrac{y^2}{\theta^2}, & 0 < y < \theta, \\ 1, & y \ge \theta. \end{cases}$$

b Show that Y/θ is a pivotal quantity.
c Use the pivotal quantity from part (b) to find a 90% lower confidence limit for θ.

8.45 Refer to Exercise 8.44.

a Use the pivotal quantity from Exercise 8.44(b) to find a 90% upper confidence limit for θ.
b If $\hat{\theta}_L$ is the lower confidence bound for θ obtained in Exercise 8.44(c) and $\hat{\theta}_U$ is the upper bound found in part (a), what is the confidence coefficient of the interval $(\hat{\theta}_L, \hat{\theta}_U)$?

8.46 Refer to Example 8.4 and suppose that Y is a single observation from an exponential distribution with mean θ.

a Use the method of moment-generating functions to show that $2Y/\theta$ is a pivotal quantity and has a χ^2 distribution with 2 df.
b Use the pivotal quantity $2Y/\theta$ to derive a 90% confidence interval for θ.
c Compare the interval you obtained in part (b) with the interval obtained in Example 8.4.

8.47 Refer to Exercise 8.46. Assume that $Y_1, Y_2, \ldots, Y_n$ is a sample of size n from an exponential distribution with mean θ.

a Use the method of moment-generating functions to show that $2\sum_{i=1}^{n} Y_i/\theta$ is a pivotal quantity and has a χ^2 distribution with $2n$ df.
b Use the pivotal quantity $2\sum_{i=1}^{n} Y_i/\theta$ to derive a 95% confidence interval for θ.

a Click the button "One Sample" a single time.

   i What is the length of the resulting interval? Is the interval longer or shorter than that obtained in Exercise 8.51(d)?

   ii Give three reasons that the interval you obtained in part (i) is shorter than the interval obtained in Exercise 8.51(d).

b Click the button "100 Samples" a few times. Each click will produce 100 intervals and provide you with the number and proportion of those 100 intervals that contain the true value of p. After each click, write down the number of intervals that captured $p = .25$.

   i How many intervals did you generate? How many of the generated intervals captured the true value of p?

   ii What percentage of all the generated intervals captured p?

8.54 **Applet Exercise** Refer to Exercises 8.51–8.53. Change the value of p to .90. Change the sample size to $n = 10$ and the confidence coefficient to 0.95. Click the button "100 Samples" a few times. After each click, write down the number of intervals that captured $p = .90$.

a When the simulation produced ten successes in ten trials, what is the resulting realized 95% confidence interval for p? What is the length of the interval? Why? How is this depicted on the display?

b How many intervals did you generate? How many of the generated intervals captured the true value of p?

c What percentage of all of the generated intervals captured p?

d Does the result of part (c) surprise you?

e Does the result in part (c) invalidate the *large-sample* confidence interval procedures presented in this section? Why?

8.55 **Applet Exercise** Refer to Exercises 8.51–8.54. Change the value of p to .90. Change the sample size to $n = 100$ and the confidence coefficient to .95. Click the button "100 Samples" a few times. After each click, write down the number of intervals that captured $p = .90$ and answer the questions posed in Exercise 8.54, parts (b)–(e).

8.56 Is America's romance with movies on the wane? In a Gallup Poll[5] of $n = 800$ randomly chosen adults, 45% indicated that movies were getting better whereas 43% indicated that movies were getting worse.

a Find a 98% confidence interval for p, the overall proportion of adults who say that movies are getting better.

b Does the interval include the value $p = .50$? Do you think that a majority of adults say that movies are getting better?

8.57 Refer to Exercise 8.29. According to the result given there, 51% of the $n = 1001$ adults polled in November 2003 claimed to be baseball fans. Construct a 99% confidence interval for the proportion of adults who professed to be baseball fans in November 2003 (after the World Series). Interpret this interval.

8.58 The administrators for a hospital wished to estimate the average number of days required for inpatient treatment of patients between the ages of 25 and 34. A random sample of 500 hospital

5. *Source:* "Movie Mania Ebbing," Gallup Poll of 800 adults, http://www.usatoday.com/snapshot/news/2001-06-14-moviemania.htm., 16–18 March 2001.

	Mean	Sample Size	Standard Deviation
Pretest: all BACC classes	13.38	372	5.59
Pretest: all traditional	14.06	368	5.45
Posttest: all BACC classes	18.50	365	8.03
Posttest: all traditional	16.50	298	6.96

a Give a 90% confidence interval for the mean posttest score for all BACC students.

b Find a 95% confidence interval for the difference in the mean posttest scores for BACC and traditionally taught students.

c Does the confidence interval in part (b) provide evidence that there is a difference in the mean posttest scores for BACC and traditionally taught students? Explain.

8.67 One suggested method for solving the electric-power shortage in a region involves constructing floating nuclear power plants a few miles offshore in the ocean. Concern about the possibility of a ship collision with the floating (but anchored) plant has raised the need for an estimate of the density of ship traffic in the area. The number of ships passing within 10 miles of the proposed power-plant location per day, recorded for $n = 60$ days during July and August, possessed a sample mean and variance of $\overline{y} = 7.2$ and $s^2 = 8.8$.

a Find a 95% confidence interval for the mean number of ships passing within 10 miles of the proposed power-plant location during a 1-day time period.

b The density of ship traffic was expected to decrease during the winter months. A sample of $n = 90$ daily recordings of ship sightings for December, January, and February yielded a mean and variance of $\overline{y} = 4.7$ and $s^2 = 4.9$. Find a 90% confidence interval for the difference in mean density of ship traffic between the summer and winter months.

c What is the population associated with your estimate in part (b)? What could be wrong with the sampling procedure for parts (a) and (b)?

***8.68** Suppose that Y_1, Y_2, Y_3, and Y_4 have a multinomial distribution with n trials and probabilities p_1, p_2, p_3, and p_4 for the four cells. Just as in the binomial case, any linear combination of Y_1, Y_2, Y_3, and Y_4 will be approximately normally distributed for large n.

a Determine the variance of $Y_1 - Y_2$. [*Hint:* Recall that the random variables Y_i are dependent.]

b A study of attitudes among residents of Florida with regard to policies for handling nuisance alligators in urban areas showed the following. Among 500 people sampled and presented with four management choices, 6% said the alligators should be completely protected, 16% said they should be destroyed by wildlife officers, 52% said they should be relocated live, and 26% said that a regulated commercial harvest should be allowed. Estimate the difference between the population proportion favoring complete protection and the population proportion favoring destruction by wildlife officers. Use a confidence coefficient of .95.

***8.69** The *Journal of Communication*, Winter 1978, reported on a study of viewing violence on TV. Samples from populations with low viewing rates (10–19 programs per week) and high viewing rates (40–49 programs per week) were divided into two age groups, and Y_i, the number of persons watching a high number of violent programs, was recorded. The data for two age groups are shown in the accompanying table, with n_i denoting the sample size for [illegible]. Y_1, Y_2, Y_3, and Y_4 have independent binomial distributions with parameters p_1, [illegible] p_4, respectively, find a 95% confidence interval for $(p_3 - p_1) - (p_4 - p_2)$. This function [illegible] p_i values represents a comparison between the change in viewing habits for young [illegible] the corresponding change for older adults, as we move from those with low view[illegible]