Chapter 9

Properties of Point Estimators and Methods of Estimation

9.1 Introduction

In Chapter 8 we presented some intuitive estimators for parameters often of interest in practical problems. An estimator $\hat{\theta}$ for a target parameter θ is a function of the random variables observed in a sample and, therefore, is itself a random variable. Consequently, an estimator has a probability distribution, which we call the *sampling distribution* of the estimator. We noted in Section 8.2 that, if $E(\hat{\theta}) = \theta$, then the estimator has the (sometimes) desirable property of being unbiased.

In this chapter we undertake a more formal and detailed examination of some of the mathematical properties of point estimators—particularly the notions of efficiency, consistency, and sufficiency. We present a result, the Rao–Blackwell theorem, that provides a link between sufficient statistics and unbiased estimators for parameters. Generally speaking, an unbiased estimator with small variance is, or can be made to be, a function of a sufficient statistic. We also demonstrate a method that can sometimes be used to find minimum-variance unbiased estimators for parameters of interest. We then offer two other useful methods for deriving estimators: the method of moments and the method of maximum likelihood. Some properties of estimators derived by these methods are discussed as well.

9.2 Relative Efficiency

It usually is possible to obtain more than one unbiased estimator for the same target parameter θ. In Section 8.2 (Figure 8.3), we mentioned that, if $\hat{\theta}_1$ and $\hat{\theta}_2$ denote two unbiased estimators for the same parameter θ, we prefer to use the estimator with the smaller variance. That is, if both estimators are unbiased, $\hat{\theta}_1$ is *relatively more efficient* than $\hat{\theta}_2$ if $V(\hat{\theta}_2) > V(\hat{\theta}_1)$. In fact, we use the ratio $V(\hat{\theta}_2)/V(\hat{\theta}_1)$ to define the *relative efficiency* of two unbiased estimators.

Definition 9.1

Given two unbiased estimators, $\hat{\theta}_1$ and $\hat{\theta}_2$, of a parameter θ, with variances $V(\hat{\theta}_1)$ and $V(\hat{\theta}_2)$, respectively, then the *efficiency* of $\hat{\theta}_1$ relative to $\hat{\theta}_2$, denoted $\text{eff}(\hat{\theta}_1, \hat{\theta}_2)$, is defined to be the ratio

$$\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)}.$$

If $\hat{\theta}_1$ and $\hat{\theta}_2$ are unbiased estimators for θ, the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$, $\text{eff}(\hat{\theta}_1, \hat{\theta}_2)$, is greater than 1 only if $V(\hat{\theta}_2) > V(\hat{\theta}_1)$. In this case $\hat{\theta}_1$ is a better unbiased estimator than $\hat{\theta}_2$. For example, if $\text{eff}(\hat{\theta}_1, \hat{\theta}_2) = 1.8$, then $V(\hat{\theta}_2) = (1.8)V(\hat{\theta}_1)$ and $\hat{\theta}_1$ is preferred to $\hat{\theta}_2$. Similarly, if $\text{eff}(\hat{\theta}_1, \hat{\theta}_2)$ is less than 1, say, .73, then $V(\hat{\theta}_2) = (.73)V(\hat{\theta}_1)$ and $\hat{\theta}_2$ is preferred to $\hat{\theta}_1$. Let us consider an example involving two different estimators for a population mean. Suppose that we wish to estimate the mean of a normal population. Let $\hat{\theta}_1$ be the sample *median*, the middle observation when the sample measurements are ordered according to magnitude (n odd) or the average of the two middle observations (n even). Let $\hat{\theta}_2$ be the sample mean. Although proof is omitted, it can be shown that the variance of the sample median, for large n, is $V(\hat{\theta}_1) = (1.2533)^2(\sigma^2/n)$. Then the efficiency of the sample median relative to the sample mean is

$$\text{eff}\,(\hat{\theta}_1,\,\hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)} = \frac{\sigma^2/n}{(1.2533)^2\sigma^2/n} = \frac{1}{(1.2533)^2} = .6366.$$

Thus we see that the amount of variability associated with the sample mean is approximately 63% of the variability associated with the sample median. Therefore, we would prefer to use the sample mean as the estimator for the population mean.

EXAMPLE 9.1 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the uniform distribution on the interval $(0, \theta)$. Two unbiased estimators for θ are

$$\hat{\theta}_1 = 2\overline{Y} \qquad \text{and} \qquad \hat{\theta}_2 = \left(\frac{n+1}{n}\right) Y_{(n)}$$

where $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$. Find the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$.

Solution Because each Y_i has a uniform distribution on the interval $(0, \theta)$, $\mu = E(Y_i) = \theta/2$ and $\sigma^2 = V(Y_i) = \theta^2/12$. Therefore,

$$E(\hat{\theta}_1) = E(2\overline{Y}) = 2E(\overline{Y}) = 2(\mu) = 2\left(\frac{\theta}{2}\right) = \theta$$

and $\hat{\theta}_1$ is unbiased, as claimed. Further,

$$V(\hat{\theta}_1) = V(2\overline{Y}) = 4V(\overline{Y}) = 4\left[\frac{V(Y_i)}{n}\right] = \left(\frac{4}{n}\right)\left(\frac{\theta^2}{12}\right) = \frac{\theta^2}{3n}.$$

To find the mean and variance of $\hat{\theta}_2$, recall (see Section 6.7) that the density function of $Y_{(n)}$ is given by

$$g_{(n)}(y) = n[F_Y(y)]^{n-1} f_Y(y) = \begin{cases} n\left(\dfrac{y}{\theta}\right)^{n-1}\left(\dfrac{1}{\theta}\right), & 0 \le y \le \theta \\ 0, & \text{elsewhere} \end{cases}$$

because

$$P(Y_i \le y) = F_Y(y) = \begin{cases} 0, & y < 0 \\ \dfrac{y}{\theta}, & 0 \le y \le \theta \\ 1, & y > \theta. \end{cases}$$

Thus

$$E(Y_{(n)}) = \frac{n}{\theta^n}\int_0^\theta y^n dy = \left(\frac{n}{n+1}\right)\theta$$

and it follows that $E\{[(n+1)/n]Y_{(n)}\} = \theta$; that is, $\hat{\theta}_2$ is an unbiased estimator for θ.

Because

$$E(Y_{(n)}^2) = \frac{n}{\theta^n}\int_0^\theta y^{n+1} dy = \left(\frac{n}{n+2}\right)\theta^2$$

we obtain

$$V(Y_{(n)}) = E(Y_{(n)}^2) - [E(Y_{(n)})]^2 = \left[\frac{n}{n+2} - \left(\frac{n}{n+1}\right)^2\right]\theta^2$$

and

$$\begin{aligned} V(\hat{\theta}_2) = V\left[\left(\frac{n+1}{n}\right)Y_{(n)}\right] &= \left(\frac{n+1}{n}\right)^2 V(Y_{(n)}) \\ = \left[\frac{(n+1)^2}{n(n+2)} - 1\right]\theta^2 &= \frac{\theta^2}{n(n+2)}. \end{aligned}$$

Therefore, the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$ is given by

$$\text{eff}\,(\hat{\theta}_1,\,\hat{\theta}_2) = \frac{V(\hat{\theta}_2)}{V(\hat{\theta}_1)} = \frac{\theta^2/[n(n+2)]}{\theta^2/3n} = \frac{3}{n+2}.$$

This efficiency is less than one if $n > 1$. That is, if $n > 1$, $\hat{\theta}_2$ has a smaller variance than $\hat{\theta}_1$, and therefore $\hat{\theta}_2$ is generally preferable to $\hat{\theta}_1$ as an estimator of θ.

We will 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 8.4 we considered a random sample of size three from an exponential distribution with a 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

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

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 8.14 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

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

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

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

In Exercise 8.15, we determined that $\hat{\theta}_1 = nY_{(1)}$ is an unbiased estimator of θ with $\text{MSE}(\hat{\theta}_1) = \theta^2$. Consider the estimator $\hat{\theta}_2 = \overline{Y}$, and find the efficiency of $\hat{\theta}_1$ relative to $\hat{\theta}_2$.

***9.8** Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a probability density function $f(y)$, which has unknown parameter θ. If $\hat{\theta}$ is an unbiased estimator of θ, then under very general conditions

$$V(\hat{\theta}) \geq I(\theta) \qquad \text{where} \qquad I(\theta) = 1 \Big/ nE\left[-\frac{\partial^2 \ln f(Y)}{\partial \theta^2}\right].$$

(This is known as the Cramer–Rao inequality.) If $V(\hat{\theta}) = I(\theta)$, the estimator $\hat{\theta}$ is said to be *efficient.*[1]

a. Suppose that $f(y)$ is the normal density with mean μ and variance σ^2. Show that $\overline{Y}$ is an efficient estimator of μ.

b. This inequality also holds for discrete probability functions $p(y)$. Suppose that $p(y)$ is the Poisson probability function with mean λ. Show that $\overline{Y}$ is an efficient estimator of λ.

[1]Exercises preceded by an asterisk are optional.

9.3 Consistency

Suppose that a coin, which has probability p of resulting in heads, is tossed n times. If the tosses are independent, then Y, the number of heads among the n tosses, has a binomial distribution. If the true value of p is unknown, the sample proportion Y/n is an estimator of p. What happens to this sample proportion as the number of tosses n increases? Our intuition leads us to believe that, as n gets larger, Y/n should get closer to the true value of p. That is, our estimator should get closer to the quantity being estimated as the amount of information in the sample increases.

Because Y/n is a random variable, we may express this "closeness" to p in probabilistic terms. In particular, let us examine the probability that the distance between the estimator and the target parameter, $|(Y/n) - p|$, will be less than some arbitrary positive real number, ε. If our intuition is correct and n is large, this probability,

$$P\left(\left|\frac{Y}{n} - p\right| \leq \varepsilon\right)$$

should be close to 1. If this probability, in fact, does tend to 1 as $n \to \infty$, we then say that (Y/n) is a *consistent estimator* of p or that (Y/n) "converges in probability to p."

Definition 9.2

The estimator $\hat{\theta}_n$ is said to be a *consistent estimator* of θ if, for any positive number ε,

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \leq \varepsilon) = 1$$

or equivalently,

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| > \varepsilon) = 0.$$

The notation $\hat{\theta}_n$ expresses that the estimator for θ is calculated by using a sample of size n. For example, $\overline{Y}_2$ is the average of two observations, whereas $\overline{Y}_{100}$ is the average of the 100 observations contained in a sample of size $n = 100$. If $\hat{\theta}_n$ is an unbiased estimator, the following theorem can often be used to prove that the estimator is consistent.

Theorem 9.1 *An unbiased estimator $\hat{\theta}_n$ for θ is a consistent estimator of θ if*

$$\lim_{n\to\infty} V(\hat{\theta}_n) = 0.$$

Proof

If Y is any random variable with $E(Y) = \mu$ and $V(Y) = \sigma^2 < \infty$, and if k is any nonnegative constant, Tchebysheff's theorem (see Theorem 4.13) implies that

$$P(|Y - \mu| > k\sigma) \leq \frac{1}{k^2}.$$

Because $\hat{\theta}_n$ is an unbiased estimator for θ, it follows that $E(\hat{\theta}_n) = \theta$. Let $\sigma_{\hat{\theta}_n} = \sqrt{V(\hat{\theta}_n)}$ denote the standard error of the estimator $\hat{\theta}_n$. If we apply Tchebysheff's theorem for the random variable $\hat{\theta}_n$, we obtain

$$P\left(\left|\hat{\theta}_n - \theta\right| > k\sigma_{\hat{\theta}_n}\right) \leq \frac{1}{k^2}.$$

Let n be any fixed sample size. For any positive number ε,

$$k = \frac{\varepsilon}{\sigma_{\hat{\theta}_n}}$$

is a positive number. Application of Tchebysheff's theorem for this fixed n and this choice of k shows that

$$P\left(\left|\hat{\theta}_n - \theta\right| > \varepsilon\right) = P\left(\left|\hat{\theta}_n - \theta\right| > \left[\frac{\varepsilon}{\sigma_{\hat{\theta}_n}}\right]\sigma_{\hat{\theta}_n}\right) \leq \frac{1}{\left(\varepsilon/\sigma_{\hat{\theta}_n}\right)^2} = \frac{V(\hat{\theta}_n)}{\varepsilon^2}$$

or

$$P\left(\left|\hat{\theta}_n - \theta\right| > \varepsilon\right) \leq \frac{V(\hat{\theta}_n)}{\varepsilon^2}.$$

Now we need to take the limit as $n \to \infty$ of the preceding sequence of probabilities. Notice that, for any fixed n,

$$0 \leq P\left(\left|\hat{\theta}_n - \theta\right| > \varepsilon\right) \leq \frac{V(\hat{\theta}_n)}{\varepsilon^2}.$$

If $\lim_{n\to\infty} V(\hat{\theta}_n) = 0$, then

$$\lim_{n\to\infty}(0) \leq \lim_{n\to\infty} P\left(\left|\hat{\theta}_n - \theta\right| > \varepsilon\right) \leq \lim_{n\to\infty} \frac{V(\hat{\theta}_n)}{\varepsilon^2} = 0.$$

Thus $\hat{\theta}_n$ is a consistent estimator for θ.

The consistency property given in Definition 9.2 and discussed in Theorem 9.1 involves a particular type of convergence of $\hat{\theta}_n$ to θ. For this reason the statement "$\hat{\theta}_n$ is a consistent estimator for θ" is sometimes replaced by the equivalent statement "$\hat{\theta}_n$ converges in probability to θ."

EXAMPLE 9.2 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a distribution with mean μ and variance $\sigma^2 < \infty$. Show that $\overline{Y}_n = \frac{1}{n}\sum_{i=1}^{n} Y_i$ is a consistent estimator of μ. (Note : we use the notation $\overline{Y}_n$ to explicitly indicate that $\overline{Y}$ is calculated by using a sample of size n.)

Solution We know from earlier chapters that $E(\overline{Y}_n) = \mu$ and $V(\overline{Y}_n) = \sigma^2/n$. Because $\overline{Y}_n$ is unbiased for μ and $V(\overline{Y}_n) \to 0$ as $n \to \infty$, Theorem 9.1 establishes that $\overline{Y}_n$ is a consistent estimator of μ. Equivalently, we may say that $\overline{Y}_n$ converges in probability to μ.

The fact that $\overline{Y}_n$ is consistent for μ, or converges in probability to μ, is sometimes referred to as the *law of large numbers*. It provides the theoretical justification for the averaging process employed by many experimenters to obtain precision in measurements. For example, an experimenter may take the average of the weights of many animals to obtain a more precise estimate of the average weight of animals of this species. His feeling, a feeling confirmed by Theorem 9.1, is that the average of many independently selected weights should be quite close to the true mean weight, with high probability.

In Section 8.3 we considered an intuitive estimator for $\mu_1 - \mu_2$, the difference in the means of two populations. The estimator discussed at that time was $\overline{Y}_1 - \overline{Y}_2$, the difference in the means of independent random samples selected from two populations. The results of Theorem 9.2 will be very useful in establishing the consistency of such estimators.

Theorem 9.2 *Suppose that $\hat{\theta}_n$ converges in probability to θ and that $\hat{\theta}'_n$ converges in probability to θ'.*

(a) $\hat{\theta}_n + \hat{\theta}'_n$ *converges in probability to* $\theta + \theta'$.

(b) $\hat{\theta}_n \times \hat{\theta}'_n$ *converges in probability to* $\theta \times \theta'$.

(c) $\hat{\theta}_n / \hat{\theta}'_n$ *converges in probability to* θ/θ', *provided that* $\theta' \neq 0$.

(d) If $g(\cdot)$ *is a real-valued function that is continuous at* θ, *then* $g(\hat{\theta}_n)$ *converges in probability to* $g(\theta)$.

The proof of Theorem 9.2 closely resembles the corresponding proof in the case where $\{a_n\}$ and $\{b_n\}$ are sequences of real numbers converging to real limits a and b, respectively. For example, if $a_n \to a$ and $b_n \to b$ then

$$a_n + b_n \to a + b.$$

EXAMPLE 9.3 Suppose that $Y_1, Y_2, \ldots, Y_n$ represent a random sample such that $E(Y_i) = \mu$, $E(Y_i^2) = \mu'_2$, and $E(Y_i^4) = \mu'_4$ are all finite. Show that

$$S_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (Y_i - \overline{Y}_n)^2$$

is a consistent estimator of $\sigma^2 = V(Y_i)$. (Note : we use subscript n on both S^2 and $\overline{Y}$ to explicitly convey their dependence on the value of the sample size, n.)

Solution We have seen in earlier chapters that S^2, now written as S_n^2, is

$$S_n^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} Y_i^2 - n\overline{Y}_n^2\right) = \left(\frac{n}{n-1}\right)\left(\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}_n^2\right).$$

The statistic $(1/n)\sum_{i=1}^{n} Y_i^2$ is the average of n independent and identically distributed random variables, with $E(Y_i^2) = \mu_2'$ and $V(Y_i^2) = \mu_4' - (\mu_2')^2 < \infty$. By the law of large numbers (Example 9.2), we know that $(1/n)\sum_{i=1}^{n} Y_i^2$ converges in probability to μ_2'.

Example 9.2 also implies that $\overline{Y}_n$ converges in probability to μ. Since the function $g(x) = x^2$ is continuous for all finite values of x, Theorem 9.2(d) implies that $\overline{Y}_n^2$ converges in probability to μ^2. It then follows from Theorem 9.2(a) that

$$\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}_n^2$$

converges in probability to $\mu_2' - \mu^2 = \sigma^2$. Because $n/(n-1)$ is a sequence of constants converging to one as $n \to \infty$, we can conclude that S_n^2 converges in probability to σ^2. Equivalently, S_n^2, the sample variance, is a consistent estimator for σ^2, the population variance.

In Section 8.6 we considered large-sample confidence intervals for some parameters of practical interest. In particular, if $Y_1, Y_2, \ldots, Y_n$ is a random sample from any distribution with mean μ and variance σ^2, we established that

$$\overline{Y} \pm z_{\alpha/2}\left(\frac{\sigma}{\sqrt{n}}\right)$$

is a valid large-sample confidence interval with confidence coefficient approximately equal to $(1-\alpha)$. If σ^2 is known, this interval can and should be calculated. However, if σ^2 is not known but the sample size is large, we recommended substituting S for

σ in the calculation, since this entails no significant loss of accuracy. The following theorem provides the theoretical justification for these claims.

Theorem 9.3 *Suppose that U_n has a distribution function that converges to a standard normal distribution function as $n \to \infty$. If W_n converges in probability to 1, then the distribution function of U_n/W_n converges to a standard normal distribution function.*

This result follows from a general result known as *Slutsky's Theorem* (see Serfling 1980). The proof of this result is beyond the scope of this text. However, the usefulness of the result is illustrated in the following example.

EXAMPLE 9.4 Suppose that $Y_1, Y_2, \ldots, Y_n$ is a random sample of size n from a distribution with $E(Y_i) = \mu$ and $V(Y_i) = \sigma^2$. Define S_n^2 as

$$S_n^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y}_n)^2.$$

Show that the distribution function of

$$\sqrt{n}\left(\frac{\overline{Y}_n - \mu}{S_n}\right)$$

converges to a standard normal distribution function.

Solution In Example 9.3 we showed that S_n^2 converges in probability to σ^2. Notice that $g(x) = +\sqrt{x/c}$ is a continuous function of x if both x and c are positive. Hence it follows from Theorem 9.2(d) that $S_n/\sigma = +\sqrt{S_n^2/\sigma^2}$ converges in probability to 1. We also know from the central limit theorem (Theorem 7.4) that the distribution function of

$$U_n = \sqrt{n}\left(\frac{\overline{Y}_n - \mu}{\sigma}\right)$$

converges to a standard normal distribution function. Therefore, Theorem 9.3 implies that the distribution function of

$$\sqrt{n}\left(\frac{\overline{Y}_n - \mu}{\sigma}\right)\Bigg/(S_n/\sigma) = \sqrt{n}\left(\frac{\overline{Y}_n - \mu}{S_n}\right)$$

converges to a standard normal distribution function.

The result of Example 9.4 tells us that, when n is large, $\sqrt{n}(\overline{Y}_n - \mu)/S_n$ has approximately a standard normal distribution *whatever* is the form of the distribution from which the sample is taken. If the sample is taken from a *normal distribution*, the results of Chapter 7 imply that $t = \sqrt{n}(\overline{Y}_n - \mu)/S_n$ has a t distribution with $n - 1$ degrees of freedom. Combining this information, we see that, if a large sample is taken from a normal distribution, the distribution function of $t = \sqrt{n}(\overline{Y}_n - \mu)/S_n$ can be approximated by a standard normal distribution function. That is, as n gets large, and hence as the number of degrees of freedom gets large, the t distribution function converges to the standard normal distribution function.

If we obtain a large sample from any distribution, we know from Example 9.4 that $\sqrt{n}(\overline{Y}_n - \mu)/S_n$ has approximately a standard normal distribution. Therefore, it follows that

$$P\left[-z_{\alpha/2} \leq \sqrt{n}\left(\frac{\overline{Y}_n - \mu}{S_n}\right) \leq z_{\alpha/2}\right] \approx 1 - \alpha.$$

If we manipulate the inequalities in the probability statement to isolate μ in the middle, we obtain

$$P\left[\overline{Y}_n - z_{\alpha/2}\left(\frac{S_n}{\sqrt{n}}\right) \leq \mu \leq \overline{Y}_n + z_{\alpha/2}\left(\frac{S_n}{\sqrt{n}}\right)\right] \approx 1 - \alpha.$$

Thus, $\overline{Y}_n \pm z_{\alpha/2}(S_n/\sqrt{n})$ forms a valid large-sample confidence interval for μ, with confidence coefficient approximately equal to $1 - \alpha$. Similarly, Theorem 9.3 can be applied to show that

$$\hat{p}_n \pm z_{\alpha/2}\sqrt{\frac{\hat{p}_n \hat{q}_n}{n}}$$

is a valid large-sample confidence interval for p with confidence coefficient approximately equal to $1 - \alpha$.

In this section we have seen that the property of consistency tells us something about the distance between an estimator and the quantity being estimated. We have seen that, when the sample size is large, $\overline{Y}_n$ is close to μ and S_n^2 is close to σ^2, with high probability. We will see other examples of consistent estimators in the exercises and later in the chapter.

In this section, we have used the notation $\overline{Y}_n$, S_n^2, $\hat{p}_n$, and in general, $\hat{\theta}_n$ to explicitly convey the dependence of the estimators on the sample size, n. We needed to do so because we were interested in computing

$$\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \leq \varepsilon).$$

If the above limit is 1, then $\hat{\theta}_n$ is a "consistent" estimator for θ (more precisely, $\hat{\theta}_n$ a consistent *sequence of estimators* for θ). Unfortunately, this notation makes our estimators look overly complicated. Henceforth, we will revert to the notation $\hat{\theta}$ as our estimator for θ and not explicitly display the dependence of the estimator on n. The dependence of $\hat{\theta}$ on the sample size, n, is always implicit and should be used whenever the consistency of the estimator is considered.

EXERCISES

9.9 Refer to Exercise 9.3. Show that both $\hat{\theta}_1$ and $\hat{\theta}_2$ are consistent estimators for θ.

9.10 Refer to Exercise 9.5. Is $\hat{\sigma}_2^2$ a consistent estimator of σ^2?

9.11 Suppose that $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ are independent random samples from populations with means μ_1 and μ_2 and variances σ_1^2 and σ_2^2, respectively. Show that $\overline{X} - \overline{Y}$ is a consistent estimator of $\mu_1 - \mu_2$.

9.12 In Exercise 9.11, suppose that the populations are normally distributed with

$\sigma_1^2 = \sigma_2^2 = \sigma^2$. Show that

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

is a consistent estimator of σ^2.

9.13 Let Y_1, Y_2, ..., Y_n denote a random sample from the probability density function

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

where $\theta > 0$. Show that $\overline{Y}$ is a consistent estimator of $\theta/(\theta+1)$.

9.14 If Y has a binomial distribution with n trials and success probability p, show that Y/n is a consistent estimator of p.

9.15 Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of size n from a normal population with mean μ and variance σ^2. Assuming $n = 2k$ for some integer k, one possible estimator for σ^2 is given by:

$$\hat{\sigma}^2 = \frac{1}{2k}\sum_{i=1}^{k}(Y_{2i} - Y_{2i-1})^2.$$

a. Show that $\hat{\sigma}^2$ an unbiased estimator for σ^2.

b. Show that $\hat{\sigma}^2$ is a consistent estimator for σ^2.

9.16 Refer to Exercise 9.15. Suppose that $Y_1, Y_2, \ldots, Y_n$ is a random sample of size n

from a Poisson distributed population with mean λ. Again, assume that $n = 2k$ for some integer k. Consider

$$\hat{\lambda} = \frac{1}{2k}\sum_{i=1}^{k}(Y_{2i} - Y_{2i-1})^2.$$

a. Show that $\hat{\lambda}$ an unbiased estimator for λ.

b. Show that $\hat{\lambda}$ is a consistent estimator for λ.

9.17 Refer to Exercise 9.15. Suppose that $Y_1, Y_2, \ldots, Y_n$ is a random sample of size n from a population for which the first four moments are finite. That is, $m_1' = E(Y_1) < \infty$, $m_2' = E(Y_1^2) < \infty$, $m_3' = E(Y_1^3) < \infty$, and $m_4' = E(Y_1^4) < \infty$. (Note : this assumption is valid for the normal and Poisson distributions in Exercise 9.15 and 9.16, respectively.) Again, assume that $n = 2k$ for some integer k. Consider

$$\hat{\sigma}^2 = \frac{1}{2k}\sum_{i=1}^{k}(Y_{2i} - Y_{2i-1})^2.$$

a. Show that $\hat{\sigma}^2$ an unbiased estimator for σ^2.

b. Show that $\hat{\sigma}^2$ is a consistent estimator for σ^2.

c. Why did you need the assumption that $m_4' = E(Y_1^4) < \infty$?

9.18 Let $Y_1,\ Y_2,\ Y_3,\ \cdots$ be independent standard normal random variables.

a. What is the distribution of $\sum_{i=1}^{n} Y_i^2$?

b. Let $W_n = \dfrac{1}{n}\sum_{i=1}^{n} Y_i^2$. Does W_n *converge in probability* to some constant? If so, what is the value of the constant?

9.19 Suppose that $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size n from a normal distribution with mean μ and variance 1. Consider the first observation Y_1 as an estimator for μ.

a. Show that Y_1 is an unbiased estimator for μ.

b. Find $P(|Y_1 - \mu| \leq 1)$.

c. Look at the basic definition of consistency given in Definition 9.2. Based on the result of (b), is Y_1 a consistent estimator for μ?

***9.20** It is sometimes relatively easy to establish consistency or lack of consistency by appealing directly to Definition 9.2, evaluating $P(|\hat{\theta}_n - \theta| \leq \varepsilon)$ directly, and then showing that $\lim_{n\to\infty} P(|\hat{\theta}_n - \theta| \leq \varepsilon) = 1$. 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 $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$, we showed in Exercise 6.59 that the probability distribution function of $Y_{(n)}$ is given by

$$F_{(n)}(y) = \begin{cases} 0, & y < 0 \\ (y/\theta)^n, & 0 \leq y \leq \theta \\ 1, & y > \theta. \end{cases}$$

a. For each $n \geq 1$ and every $\varepsilon > 0$, it follows that $P(|Y_{(n)} - \theta| \leq \varepsilon) = P(\theta - \varepsilon \leq$

$Y_{(n)} \leq \theta + \varepsilon)$. If $\varepsilon > \theta$, verify that $P(\theta - \varepsilon \leq Y_{(n)} \leq \theta + \varepsilon) = 1$ and that, for every positive $\varepsilon < \theta$, we obtain $P(\theta - \varepsilon \leq Y_{(n)} \leq \theta + \varepsilon) = 1 - [(\theta - \varepsilon)/\theta]^n$.

b. Using the result from (a), show that $Y_{(n)}$ is a consistent estimator for θ by showing that, for every $\varepsilon > 0$, we have $\lim_{n \to \infty} P(|Y_{(n)} - \theta| \leq \varepsilon) = 1$.

***9.21** Use the method described in Exercise 9.20 to show that, if $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$ when $Y_1, Y_2, \ldots, Y_n$ are independent uniform random variables on the interval $(0, \theta)$, then $Y_{(1)}$ is *not* a consistent estimator for θ. Hint: Based on the methods of Section 6.7, $Y_{(1)}$ has the distribution function

$$F_{(1)}(y) = \begin{cases} 0, & y < 0 \\ 1 - (1 - y/\theta)^n, & 0 \leq y \leq \theta \\ 1, & y > \theta. \end{cases}$$

***9.22** Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size n from a Pareto distribution (see Exercise 6.14). Then the methods of Section 6.7 imply that $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$ has the distribution function given by

$$F_{(1)}(y) = \begin{cases} 0, & y \leq \beta \\ 1 - (\beta/y)^{\alpha n}, & y > \beta. \end{cases}$$

Use the method described in Exercise 9.20 to show that $Y_{(1)}$ is a consistent estimator of β.

***9.23** Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size n from a power family distribution (see Exercise 6.13). Then the methods of Section 6.7 imply that $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$ has the distribution function given by

$$F_{(n)}(y) = \begin{cases} 0, & y < 0 \\ (y/\theta)^{\alpha n}, & 0 \le y \le \theta \\ 1, & y > \theta. \end{cases}$$

Use the method described in Exercise 9.20 to show that $Y_{(n)}$ is a consistent estimator of θ.

9.24 Let $Y_1, Y_2, \ldots, Y_n$ be independent random variables, each with probability density function

$$f(y) = \begin{cases} 3y^2, & 0 \le y \le 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Show that $\overline{Y}$ converges in probability to some constant, and find the constant.

9.25 If $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a gamma distribution with parameters α and β, show that $\overline{Y}$ converges in probability to some constant and find the constant.

9.26 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the probability density function

$$f(y) = \begin{cases} \dfrac{2}{y^2}, & y \ge 2 \\ 0, & \text{elsewhere.} \end{cases}$$

Does the law of large numbers apply to $\overline{Y}$ in this case? Why or why not?

9.27 An experimenter wishes to compare the numbers of bacteria of types A and B in samples of water. A total of n independent water samples are taken, and counts are made for each sample. Let X_i denote the number of type A bacteria, and let Y_i denote the number of type B bacteria for sample i. Assume that the two bacteria types are sparsely distributed within a water sample, so that $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ can be considered independent random samples from Poisson distributions with means λ_1 and λ_2, respectively. Suggest an estimator of $\lambda_1/(\lambda_1 + \lambda_2)$. What properties does your estimator have?

9.28 Suppose that Y has a binomial distribution based on n trials and success probability p. Then $\hat{p}_n = Y/n$ is an unbiased estimator of p. Use Theorem 9.3 to prove that the distribution of $(\hat{p}_n - p)/\sqrt{\hat{p}_n\hat{q}_n/n}$ converges to a standard normal distribution. [Hint: Write Y as we did in Section 7.5.]

9.4 Sufficiency

Up to this point we have chosen estimators on the basis of intuition. Thus we chose $\overline{Y}$ and S^2 as the estimators of the mean and variance, respectively, of the normal distribution. (It *seems* as though these estimators should be good estimators of the population parameters.) We have seen that it is sometimes desirable to use estimators that are unbiased. Indeed, $\overline{Y}$ and S^2 have been shown to be unbiased estimators of the population mean μ and variance σ^2, respectively. Notice that we have used the information in a sample of size n to calculate the value of two statistics that function as estimators for the parameters of interest. At this stage the actual sample values are no longer important; rather, we summarize the information in the sample that relates to the parameters of interest by using the statistics $\overline{Y}$ and S^2. Has this process of summarizing or reducing the data to the two statistics, $\overline{Y}$ and S^2, retained all the information about μ and σ^2 in the original set of n sample observations? Or has some information about these parameters been lost or obscured through the process of reducing the data? In this section we will present methods for finding statistics that, in a sense, summarize *all* of the information in a sample about a target parameter. Such statistics are said to have the property of *sufficiency*; or more simply, they are called *sufficient statistics*. As we will see in the next section, "good" estimators are (or can be made to be) functions of any sufficient statistic. Indeed, sufficient statistics often can be used to develop estimators that have the minimum variance among all unbiased estimators.

To illustrate the notion of a sufficient statistic, let us consider the outcomes of n trials of a binomial experiment, $X_1, X_2, \ldots, X_n$, where

$$X_i = \begin{cases} 1, & \text{if the } i\text{th trial is a success} \\ 0, & \text{if the } i\text{th trial is a failure.} \end{cases}$$

If p is the probability of success on any trial then, for $i = 1,\ 2, \ldots,\ n$,

$$X_i = \begin{cases} 1, & \text{with probability } p \\ 0, & \text{with probability } q = 1 - p. \end{cases}$$

Suppose that we are given a value of $Y = \sum_{i=1}^{n} X_i$, the number of successes among the n trials. If we know the value of Y, can we gain any further information about p by looking at other functions of $X_1,\ X_2,\ \ldots,\ X_n$? One way to answer this question is to look at the conditional distribution of $X_1,\ X_2,\ \ldots,\ X_n$, given Y:

$$P(X_1 = x_1, \ldots, X_n = x_n | Y = y) = \frac{P(X_1 = x_1, \ldots, X_n = x_n, Y = y)}{P(Y = y)}.$$

The numerator on the right side of the expression above is 0 if $\sum_{i=1}^{n} x_i \neq y$, and it is the probability of an independent sequence of 0's and 1's with a total of y 1's and $(n - y)$ 0's if $\sum_{i=1}^{n} x_i = y$. Also, the denominator is the binomial probability of exactly y successes in n trials. Therefore, if $y = 0, 1, 2, \ldots, n$,

$$P(X_1 = x_1, \ldots, X_n = x_n | Y = y) = \begin{cases} \dfrac{p^y (1-p)^{n-y}}{\binom{n}{y} p^y (1-p)^{n-y}} = \dfrac{1}{\binom{n}{y}}, & \text{if } \sum_{i=1}^{n} x_i = y \\ 0, & \text{otherwise.} \end{cases}$$

It is important to note that the conditional distribution of $X_1,\ X_2,\ \ldots,\ X_n$ given Y *does not* depend upon p. That is, once Y is known, no other function of $X_1,\ X_2,\ \ldots,\ X_n$

will shed additional light on the possible value of p. In this sense Y contains all of the information about p. Therefore, the statistic Y is said to be *sufficient* for p. We generalize this idea in the following definition.

Definition 9.3

Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a probability distribution with unknown parameter θ. Then the statistic $U = g(Y_1, Y_2, \ldots, Y_n)$ is said to be *sufficient* for θ if the conditional distribution of $Y_1, Y_2, \ldots, Y_n$ given U does not depend on θ.

In many previous discussions we have considered the probability function $p(y)$ associated with a discrete random variable or the density function $f(y)$ for a continuous random variable to be functions of the argument y only. Our future discussions will be simplified if we adopt notation that will permit us to explicitly display the fact that the distribution associated with a random variable Y often depends on the value of a parameter θ. If Y is a discrete random variable that has a probability mass function that depends on the value of a parameter θ, instead of $p(y)$ we use the notation

$$p(y \mid \theta) = P(Y = y) \quad \text{where } \theta \text{ denotes the value of the parameter.}$$

Similarly, we will indicate the explicit dependence of the form of a continuous density function on the value of a parameter θ by writing the density function as $f(y \mid \theta)$ instead of the previously used $f(y)$.

Definition 9.3 tells us how to check whether a statistic is sufficient, but it really does not tell us how to find a sufficient statistic. Recall that in the discrete case the joint distribution of discrete random variables $Y_1, Y_2, \ldots, Y_n$ is given by a probability function $p(y_1, y_2, \ldots, y_n)$. If this joint probability function depends explicitly on the value of a parameter θ, we write it as $p(y_1, y_2, \ldots, y_n \mid \theta)$. This function gives the probability or *likelihood* of observing the event $(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n)$ when the value of the parameter is θ. In the continuous case when the joint distribution of $Y_1, Y_2, \ldots, Y_n$ depends on a parameter θ, we will write the joint density function as $f(y_1, y_2, \ldots, y_n \mid \theta)$. Henceforth, it will be convenient to have a single name for the function that defines the joint distribution of the variables $Y_1, Y_2, \ldots, Y_n$ observed in a sample.

Definition 9.4

Let $y_1, y_2, \ldots, y_n$ be sample observations taken on corresponding random variables $Y_1, Y_2, \ldots, Y_n$ whose distribution depends on a parameter θ. Then, if $Y_1, Y_2, \ldots, Y_n$ are discrete random variables, the *likelihood of the sample*, $L(y_1, y_2, \ldots, y_n \mid \theta)$, is defined to be the joint probability of $y_1, y_2, \ldots, y_n$. If $Y_1, Y_2, \ldots, Y_n$ are continuous random variables, the likelihood $L(y_1,\, y_2,\, \ldots,\, y_n \,|\, \theta)$ is defined to be the joint density evaluated at $y_1,\, y_2,\, \ldots,\, y_n$.

If the set of random variables $Y_1,\, Y_2,\, \ldots,\, Y_n$ denotes a random sample from a discrete distribution with probability function $p(y \,|\, \theta)$, then

$$\begin{aligned} L(y_1,\, y_2,\, \ldots, y_n \,|\, \theta) &= P(Y_1 = y_1,\, Y_2 = y_2,\, \ldots, Y_n = y_n \,|\, \theta) \\ &= p(y_1 \,|\, \theta) \times p(y_2 \,|\, \theta) \times \cdots \times p(y_n \,|\, \theta). \end{aligned}$$

Whereas if $Y_1,\, Y_2,\, \ldots,\, Y_n$ have a continuous distribution with density function $f(y \,|\, \theta)$, then

$$\begin{aligned} L(y_1, y_2, \ldots, y_n \,|\, \theta) &= f(y_1, y_2, \ldots, y_n \,|\, \theta) \\ &= f(y_1 \,|\, \theta) \times f(y_2 \,|\, \theta) \times \cdots \times f(y_n \,|\, \theta). \end{aligned}$$

To simplify notation, we will sometimes denote the likelihood by $L(\theta)$ instead of by $L(y_1,\, y_2,\, \ldots,\, y_n \,|\, \theta)$.

The following theorem relates the property of sufficiency to the likelihood $L(\theta)$.

Theorem 9.4 *Let U be a statistic based on the random sample $Y_1,\, Y_2,\, \ldots,\, Y_n$. Then U is a* sufficient statistic *for the estimation of a parameter θ if and only if the likelihood*

$L(\theta) = L(y_1, y_2, \ldots, y_n \mid \theta)$ can be factored into two nonnegative functions,

$$L(y_1, y_2, \ldots, y_n \mid \theta) = g(u, \theta) \times h(y_1,\, y_2,\, \ldots,\, y_n)$$

where $g(u, \theta)$ is a function only of u and θ and $h(y_1, y_2, \ldots, y_n)$ is not a function of θ.

Although the proof of Theorem 9.4 (also known as the *factorization criterion*) is beyond the scope of this book, we illustrate the usefulness of the theorem in the following example.

EXAMPLE 9.5 Let $Y_1,\, Y_2,\, \ldots,\, Y_n$ be a random sample in which Y_i possesses the probability density function

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

where $\theta > 0$, $i = 1, 2, \ldots, n$. Show that $\overline{Y}$ is a sufficient statistic for the estimation of θ.

Solution The likelihood $L(\theta)$ of the sample is the joint density

$$\begin{aligned} L(y_1,\, y_2,\, \ldots,\, y_n \mid \theta) &= f(y_1,\, y_2,\, \ldots,\, y_n \mid \theta) \\ &= f(y_1 \mid \theta) \times f(y_2 \mid \theta) \times \cdots \times f(y_n \mid \theta) \\ &= \frac{e^{-y_1/\theta}}{\theta} \times \frac{e^{-y_2/\theta}}{\theta} \times \cdots \times \frac{e^{-y_n/\theta}}{\theta} = \frac{e^{-\sum y_i/\theta}}{\theta^n} = \frac{e^{-n\overline{y}/\theta}}{\theta^n}. \end{aligned}$$

Notice that $L(\theta)$ is a function only of θ and $\overline{y}$ and that, if

$$g(\overline{y}, \theta) = \frac{e^{-n\overline{y}/\theta}}{\theta^n} \qquad \text{and} \qquad h(y_1, y_2, \ldots, y_n) = 1$$

then

$$L(y_1, y_2, \ldots, y_n \,|\, \theta) = g(\overline{y}, \theta) \times h(y_1, y_2, \ldots, y_n).$$

Hence Theorem 9.4 implies that $\overline{Y}$ is a sufficient statistic for the estimation of θ.

Theorem 9.4 can be used to show that there are many possible sufficient statistics for any one population parameter. First of all, according to Definition 9.3 or the factorization criterion (Theorem 9.4), the random sample itself is a sufficient statistic. Second, if $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a distribution with a density function with parameter θ, then the set of order statistics $Y_{(1)} \leq Y_{(2)} \leq \cdots \leq Y_{(n)}$, which is a function of $Y_1, Y_2, \ldots, Y_n$, is sufficient for θ. In Example 9.5, we decided that $\overline{Y}$ is a sufficient statistic for the estimation of θ. Theorem 9.4 could also have been used to show that $\sum_{i=1}^{n} Y_i$ is another sufficient statistic. Indeed, for the exponential distribution described in Example 9.5, any statistic that is a one–to–one function of $\overline{Y}$ is a sufficient statistic.

In our initial example of this section, involving the number of successes in n trials, $Y = \sum_{i=1}^{n} X_i$ reduces the data $X_1, X_2, \ldots, X_n$ to a single value that remains sufficient for p. Generally we would like to find a sufficient statistic that reduces the data in the sample as much as possible. Although many statistics are sufficient for the parameter

θ associated with a specific distribution, application of the factorization criterion typically leads to a statistic that provides the "best" summary of the information in the data. In Example 9.5, this statistic is $\overline{Y}$ (or some one–to–one function of it). In the next section we show how these sufficient statistics can be used to develop unbiased estimators with minimum variance.

EXERCISES

9.29 Let X_1, X_2, …, X_n denote n independent and identically distributed random variables such that

$$P(X_i = 1) = p \qquad \text{and} \qquad P(X_i = 0) = 1 - p$$

for each $i = 1, 2, \ldots, n$. Show that $\sum_{i=1}^{n} X_i$ is sufficient for p by using the factorization criterion given in Theorem 9.4.

9.30 Let Y_1, Y_2, …, Y_n denote a random sample from a normal distribution with mean μ and variance σ^2.

- **a.** If μ is unknown and σ^2 is known, show that $\overline{Y}$ is sufficient for μ.
- **b.** If μ is known and σ^2 is unknown, show that $\sum_{i=1}^{n}(Y_i - \mu)^2$ is sufficient for σ^2.
- **c.** If μ and σ^2 are both unknown, show that $\sum_{i=1}^{n} Y_i$ and $\sum_{i=1}^{n} Y_i^2$ are jointly sufficient for μ and σ^2. (Thus it follows that $\overline{Y}$ and $\sum_{i=1}^{n}(Y_i - \overline{Y})^2$ or $\overline{Y}$ and

S^2 are also jointly sufficient for μ and σ^2.)

9.31 Let Y_1, Y_2, $\ldots$, Y_n denote a random sample from a Poisson distribution with parameter λ. Show by conditioning that $\sum_{i=1}^{n} Y_i$ is sufficient for λ.

9.32 Let Y_1, Y_2, $\ldots$, Y_n denote a random sample from a Rayleigh distribution with parameter θ. (Refer to Exercise 6.30.) Show that $\sum_{i=1}^{n} Y_i^2$ is sufficient for θ.

9.33 Let Y_1, Y_2, $\ldots$, Y_n denote a random sample from a Weibull distribution with known m and unknown α. (Refer to Exercise 6.22.) Show that $\sum_{i=1}^{n} Y_i^m$ is sufficient for α.

9.34 If Y_1, Y_2, $\ldots$, Y_n denote a random sample from a geometric distribution with parameter p, show that $\overline{Y}$ is sufficient for p.

9.35 Let Y_1, Y_2, $\ldots$, Y_n denote independent and identically distributed random variables from a power family distribution with parameters α and θ. Then, by the result in Exercise 6.13, if $\alpha, \theta > 0$,

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

If θ is known, show that $\prod_{i=1}^{n} Y_i$ is sufficient for α.

9.36 Let Y_1, Y_2, $\ldots$, Y_n denote independent and identically distributed random variables from a Pareto distribution with parameters α and β. Then, by the result in Exercise 6.14, if $\alpha, \beta > 0$,

$$f(y \mid \alpha, \beta) = \begin{cases} \alpha\beta^{\alpha} y^{-(\alpha+1)}, & y \geq \beta \\ 0, & \text{elsewhere.} \end{cases}$$

If β is known, show that $\prod_{i=1}^{n} Y_i$ is sufficient for α.

9.37 Suppose that $Y_1, Y_2, \ldots, Y_n$ is a random sample from a probability density function in the (one-parameter) exponential family so that

$$f(y \mid \theta) = \begin{cases} a(\theta)b(y)e^{-[c(\theta)d(y)]}, & a \leq y \leq b \\ 0, & \text{elsewhere} \end{cases}$$

where a and b do not depend on θ. Show that $\sum_{i=1}^{n} d(Y_i)$ is sufficient for θ.

9.38 If $Y_1, Y_2, \ldots, Y_n$ denote a random sample from an exponential distribution with mean β, show that $f(y \mid \beta)$ is in the exponential family and that $\overline{Y}$ is sufficient for β.

9.39 Refer to Exercise 9.35. If θ is known, show that the power family of distributions is in the exponential family. What is a sufficient statistic for α? Does this contradict your answer to Exercise 9.35?

9.40 Refer to Exercise 9.36. If β is known, show that the Pareto distribution is in the exponential family. What is a sufficient statistic for α? Argue that there is no contradiction between your answer to this exercise and the answer you found in

Exercise 9.36.

***9.41** Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the uniform distribution over the interval $(0, \theta)$. Show that $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$ is sufficient for θ.

***9.42** Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the uniform distribution over the interval (θ_1, θ_2). Show that $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$ and $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$ are jointly sufficient for θ_1 and θ_2.

***9.43** Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the probability density function

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

Show that $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$ is sufficient for θ.

***9.44** Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with density function

$$f(y \mid \theta) = \begin{cases} \dfrac{3y^2}{\theta^3} & 0 \leq y \leq \theta \\ 0, & \text{elsewhere.} \end{cases}$$

Show that $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$ is sufficient for θ.

***9.45** Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with density function

$$f(y \mid \theta) = \begin{cases} \dfrac{2\theta^2}{y^3} & \theta < y < \infty \\ 0, & \text{elsewhere.} \end{cases}$$

Show that $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$ is sufficient for θ.

***9.46** Let Y_1, Y_2, $\ldots$, Y_n denote independent and identically distributed random variables from a power family distribution with parameters α and θ. Then, as in Exercise 9.35, if $\alpha, \theta > 0$,

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

Show that $\max(Y_1, Y_2, \ldots, Y_n)$ and $\prod_{i=1}^{n} Y_i$ are jointly sufficient for α and θ.

***9.47** Let Y_1, Y_2, $\ldots$, Y_n denote independent and identically distributed random variables from a Pareto distribution with parameters α and β. Then, as in Exercise 9.36, if $\alpha, \beta > 0$,

$$f(y \mid \alpha, \beta) = \begin{cases} \alpha \beta^{\alpha} y^{-(\alpha+1)}, & y \geq \beta \\ 0, & \text{elsewhere.} \end{cases}$$

Show that $\prod_{i=1}^{n} Y_i$ and $\min(Y_1, Y_2, \ldots, Y_n)$ are jointly sufficient for α and β.

9.5 The Rao–Blackwell Theorem and Minimum-Variance Unbiased Estimation

Sufficient statistics play an important role in finding good estimators for parameters. If $\hat{\theta}$ is an unbiased estimator for θ and if U is a statistic that is sufficient for θ then there is a function of U that is also an unbiased estimator for θ and has *no larger* variance than $\hat{\theta}$. If we seek unbiased estimators with small variances, we can restrict our search to estimators that are functions of sufficient statistics. The theoretical basis for the preceding remarks is provided in the following result, known as the *Rao–Blackwell theorem.*

Theorem 9.5 (The Rao–Blackwell Theorem.) *Let $\hat{\theta}$ be an unbiased estimator for θ such that $V(\hat{\theta}) < \infty$. If U is a sufficient statistic for θ, define $\hat{\theta}^* = E(\hat{\theta} \mid U)$. Then, for all θ,*

$$E(\hat{\theta}^*) = \theta \qquad \textit{and} \qquad V(\hat{\theta}^*) \leq V(\hat{\theta}).$$

Proof

Since U is sufficient for θ, the conditional distribution of any statistic (including $\hat{\theta}$), given U, does not depend on θ. Thus $\hat{\theta}^* = E(\hat{\theta} \mid U)$ is not a function of θ and is therefore a statistic.

Recall Theorems 5.14 and 5.15, where we considered how to find means and variances of random variables by using conditional means and variances. Since $\hat{\theta}$ is an unbiased estimator for θ, Theorem 5.14 implies that

$$E(\hat{\theta}^*) = E\left(E(\hat{\theta} \mid U)\right) = E(\hat{\theta}) = \theta.$$

Thus, $\hat{\theta}^*$ is an unbiased estimator for θ.

Theorem 5.15 implies that

$$\begin{aligned} V(\hat{\theta}) &= V\left(E(\hat{\theta} \,|\, U)\right) + E\left(V(\hat{\theta} \,|\, U)\right) \\ &= V\left(\hat{\theta}^*\right) + E\left(V(\hat{\theta} \,|\, U)\right). \end{aligned}$$

Since $V(\hat{\theta} \,|\, U = u) \geq 0$ *for all* u*, it follows that* $E\left(V(\hat{\theta} \,|\, U)\right) \geq 0$ *and therefore that* $V(\hat{\theta}) \geq V(\hat{\theta}^*)$*, as claimed.*

Theorem 9.5 implies that an unbiased estimator for θ with a small variance is, or can be made to be, a function of a sufficient statistic. If we have an unbiased estimator for θ, we might be able to improve it by using the result in Theorem 9.5. It might initially seem that the Rao–Blackwell theorem could be applied once to get a better unbiased estimator, and then reapplied to the resulting new estimator to get an even better unbiased estimator. If we apply the Rao–Blackwell theorem using the sufficient statistic U, then $\hat{\theta}^* = E(\hat{\theta} \,|\, U)$ will be a function of the statistic U, say, $\hat{\theta}^* = h(U)$. Suppose that we reapply the Rao–Blackwell theorem to $\hat{\theta}^*$ by using the same sufficient statistic U. Since, in general, $E(h(U) \,|\, U) = h(U)$, we see that by using the Rao–Blackwell theorem again our "new" estimator is just $h(U) = \hat{\theta}^*$. That is, if we use the same sufficient statistic in successive applications of the Rao–Blackwell theorem, we gain nothing after the first application. The only way that successive applications can lead to better unbiased estimators is if we use a different sufficient statistic when the theorem is reapplied. Thus, it is unnecessary to use the Rao–Blackwell theorem successively if we use the "right" sufficient statistic in our initial application.

Since many statistics are sufficient for a parameter θ associated with a distribution, which sufficient statistic should we use when we apply this theorem? For the distributions that we discuss in this text, the factorization criterion typically identifies a statistic U that best summarizes the information in the data about the parameter θ. Such statistics are called *minimal sufficient statistics.* Exercise 9.58 introduces a method for determining a minimal sufficient statistic that might be of interest to some readers. In a few of the subsequent exercises, you will see that this method usually yields the same sufficient statistics as those obtained from the factorization criterion. In the cases we consider, these statistics possess another property (completeness) that guarantees that, if we apply Theorem 9.5 using U, we not only get an estimator with a smaller variance, but actually obtain an unbiased estimator for θ with *minimum variance.* Such an estimator is called a *minimum variance unbiased estimator* (MVUE). See Casella and Berger (1990), Hogg and Craig (1995), or Mood, Graybill, and Boes (1974) for additional details.

Thus, if we start with an unbiased estimator for a parameter θ and the sufficient statistic obtained through the factorization criterion, application of the Rao–Blackwell theorem typically leads to an MVUE for the parameter. Direct computation of conditional expectations can be difficult. However, if U is the sufficient statistic that best summarizes the data and some function of U, say $h(U)$ can be found such that $E(h(U)) = \theta$, it follows that $h(U)$ is the MVUE for θ. We will illustrate this approach with several examples.

EXAMPLE 9.6 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a distribution where $P(Y_i = 1) = p$ and $P(Y_i = 0) = 1 - p$, with p unknown (such random variables are often called *Bernoulli* variables). Use the factorization criterion to find a sufficient statistic that best summarizes the data. Give an MVUE for p.

Solution Notice that the preceding probability function can be written as

$$P(Y_i = y_i) = p^{y_i}(1-p)^{1-y_i}, \quad y_i = 0, 1.$$

Thus, the likelihood $L(p)$ is

$$\begin{aligned} L(y_1, y_2, \ldots, y_n \mid p) &= p(y_1, y_2, \ldots, y_n \mid p) \\ &= p^{y_1}(1-p)^{1-y_1} \times p^{y_2}(1-p)^{1-y_2} \times \cdots \times p^{y_n}(1-p)^{1-y_n} \\ &= \underbrace{p^{\sum y_i}(1-p)^{n-\sum y_i}}_{g(\sum y_i,\, p)} \times \underbrace{1}_{h(y_1, y_2, \ldots, y_n)}. \end{aligned}$$

According to the factorization criterion, $U = \sum_{i=1}^{n} Y_i$ is sufficient for p. This statistic best summarizes the information about the parameter p. Notice that $E(U) = np$ or, equivalently, $E(U/n) = p$. Thus $U/n = \overline{Y}$ is an unbiased estimator for p. Since this estimator is a function of the sufficient statistic $\sum_{i=1}^{n} Y_i$, the estimator $\hat{p} = \overline{Y}$ is the

MVUE for p.

EXAMPLE 9.7 Suppose that $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the Weibull density function, given by

$$f(y \mid \theta) = \begin{cases} \left(\dfrac{2y}{\theta}\right) e^{-y^2/\theta}, & y > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

Find an MVUE for θ.

Solution We begin by using the factorization criterion to find the sufficient statistic that best summarizes the information about θ.

$$\begin{aligned} L(y_1, y_2, \ldots, y_n \mid \theta) &= f(y_1, y_2, \ldots, y_n \mid \theta) \\ &= \left(\frac{2}{\theta}\right)^n (y_1 \times y_2 \times \cdots \times y_n) \exp\left(-\frac{1}{\theta}\sum_{i=1}^n y_i^2\right) \\ &= \underbrace{\left(\frac{2}{\theta}\right)^n \exp\left(-\frac{1}{\theta}\sum_{i=1}^n y_i^2\right)}_{g(\sum y_i^2,\, \theta)} \times \underbrace{(y_1 \times y_2 \times \cdots \times y_n)}_{h(y_1, y_2, \ldots, y_n)}. \end{aligned}$$

Thus, $U = \sum_{i=1}^n Y_i^2$ is the minimal sufficient statistic for θ.

We now must find a function of this statistic that is unbiased for θ. Letting $W = Y_i^2$, we have

$$f_W(w) = f(\sqrt{w})\frac{d(\sqrt{w})}{dw} = \left(\frac{2}{\theta}\right)(\sqrt{w}e^{-w/\theta})\left(\frac{1}{2\sqrt{w}}\right) = \left(\frac{1}{\theta}\right)e^{-w/\theta}, \quad w > 0.$$

That is, Y_i^2 has an exponential distribution with parameter θ. Because

$$E(Y_i^2) = E(W) = \theta \qquad \text{and} \quad E\left(\sum_{i=1}^{n} Y_i^2\right) = n\theta$$

it follows that

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} Y_i^2$$

is an unbiased estimator of θ that is a function of the sufficient statistic $\sum_{i=1}^{n} Y_i^2$. Therefore, $\hat{\theta}$ is an MVUE of the Weibull parameter θ.

The following example illustrates the use of this technique for estimating two unknown parameters.

EXAMPLE 9.8 Suppose $Y_1, Y_2, \ldots, Y_n$ denotes a random sample from a normal distribution with unknown mean μ and variance σ^2. Find the MVUEs for μ and σ^2.

Solution Again, looking at the likelihood function, we have

$$L(y_1, y_2, \ldots, y_n \mid \mu, \sigma^2) = f(y_1, y_2, \ldots, y_n \mid \mu, \sigma^2)$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \mu)^2\right\}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n y_i^2 - 2\mu\sum_{i=1}^n y_i + n\mu^2\right)\right\}$$
$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left\{\frac{-n\mu^2}{2\sigma^2}\right\} \exp\left\{-\frac{1}{2\sigma^2}\left(\sum_{i=1}^n y_i^2 - 2\mu\sum_{i=1}^n y_i\right)\right\}.$$

Thus $\sum_{i=1}^{n} Y_i$ and $\sum_{i=1}^{n} Y_i^2$, jointly, are sufficient statistics for μ and σ^2.

We know from past work that $\overline{Y}$ is unbiased for μ and

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

is unbiased for σ^2. Because these estimators are functions of the statistics that best summarize the information about μ and σ^2, they are MVUEs for μ and σ^2.

The factorization criterion, together with the Rao–Blackwell theorem, can also be used to find minimum-variance unbiased estimators for functions of the parameters associated with a distribution. We illustrate the technique in the following example.

EXAMPLE 9.9 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the exponential density function given by

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

Find an MVUE of $V(Y_i)$.

Solution In Chapter 4 we determined that $E(Y_i) = \theta$ and that $V(Y_i) = \theta^2$. The factorization criterion implies that $\sum_{i=1}^{n} Y_i$ is the best sufficient statistic for θ. In fact, $\overline{Y}$ is the MVUE of θ. Therefore, it is tempting to use $\overline{Y}^2$ as an estimator of θ^2. But

$$E(\overline{Y}^2) = V(\overline{Y}) + [E(\overline{Y})]^2 = \frac{\theta^2}{n} + \theta^2 = \left(\frac{n+1}{n}\right)\theta^2.$$

It follows that $\overline{Y}^2$ is a biased estimate for θ^2. However,

$$\left(\frac{n}{n+1}\right)\overline{Y}^2$$

is an MVUE of θ^2 because it is an unbiased estimator for θ^2, which is a function of the sufficient statistic. No other unbiased estimator of θ^2 will have a smaller variance than this one.

A sufficient statistic for a parameter θ often can be used to construct an exact confidence interval for θ, if the probability distribution of the statistic can be found. The resulting intervals generally are the shortest that can be found with a specified confidence coefficient. We illustrate the technique with an example involving the Weibull distribution.

EXAMPLE 9.10 The following data, with measurements in hundreds of hours, represent the lengths of life of ten identical electronic components operating in a guidance control system for missiles:

.637	1.531	.733	2.256	2.364
1.601	.152	1.826	1.868	1.126.

The length of life of a component of this type is assumed to follow a Weibull distribution with density function given by

$$f(y \mid \theta) = \begin{cases} \left(\dfrac{2y}{\theta}\right) e^{-y^2/\theta}, & y > 0 \\ 0, & \text{elsewhere.} \end{cases}$$

Use the data to construct a 95% confidence interval for θ.

Solution We saw in Example 9.7 that the sufficient statistic that best summarizes the information about θ is $\sum_{i=1}^{n} Y_i^2$. We will use this statistic to form a pivotal quantity for constructing the desired confidence interval.

Recall from Example 9.7 that $W_i = Y_i^2$ has an exponential distribution with mean θ. Now consider the transformation $T_i = 2W_i/\theta$. Then

$$f_T(t) = f_W\left(\frac{\theta t}{2}\right)\frac{d(\theta t/2)}{dt} = \left(\frac{1}{\theta}\right)e^{-(\theta t/2)/\theta}\left(\frac{\theta}{2}\right) = \left(\frac{1}{2}\right)e^{-t/2}, \qquad t > 0.$$

Thus, for each $i = 1, 2, \ldots, n$, T_i has a χ^2 distribution with 2 degrees of freedom. Further, because the variables Y_i are independent, the variables T_i are independent, for $i = 1, 2, \ldots, n$. The sum of independent χ^2 random variables has a χ^2 distribution, with degrees of freedom equal to the sum of the degrees of freedom of the variables in the sum. Therefore, the quantity

$$\sum_{i=1}^{10} T_i = \frac{2}{\theta}\sum_{i=1}^{10} W_i = \frac{2}{\theta}\sum_{i=1}^{10} Y_i^2$$

has a χ^2 distribution with 20 degrees of freedom. Thus,

$$\frac{2}{\theta}\sum_{i=1}^{10} Y_i^2$$

is a pivotal quantity, and we can use the pivotal method (Section 8.5) to construct the desired confidence interval.

From Table 6, Appendix III, we can find two numbers a and b such that

$$P\left(a \le \frac{2}{\theta}\sum_{i=1}^{10} Y_i^2 \le b\right) = .95.$$

Manipulating the inequality to isolate θ in the middle, we have

$$\begin{aligned} .95 &= P\left(a \le \frac{2}{\theta}\sum_{i=1}^{10} Y_i^2 \le b\right) = P\left(\frac{1}{b} \le \frac{\theta}{2\sum_{i=1}^{10} Y_i^2} \le \frac{1}{a}\right) \\ &= P\left(\frac{2\sum_{i=1}^{10} Y_i^2}{b} \le \theta \le \frac{2\sum_{i=1}^{10} Y_i^2}{a}\right). \end{aligned}$$

From Table 6, Appendix III, the value that cuts off an area of .025 in the lower tail of the χ^2 distribution with 20 degrees of freedom is $a = 9.591$. The value that cuts off an area of .025 in the upper tail of the same distribution is $b = 34.170$. For the preceding data, $\sum_{i=1}^{10} Y_i^2 = 24.643$. Therefore, the 95% confidence interval for the Weibull parameter θ is

$$\left(\frac{2(24.643)}{34.170}, \frac{2(24.643)}{9.591}\right) \qquad \text{or} \qquad (1.442, 5.139).$$

This is a fairly wide interval for θ, but it is based on only ten observations.

In this section we have seen that the Rao–Blackwell theorem implies that unbiased estimators with small variances are functions of sufficient statistics. Generally speaking, the factorization criterion presented in Section 9.4 can be applied to find sufficient statistics that best summarize the information contained in sample data about parameters of interest. For the distributions that we consider in this text, a minimum-variance unbiased estimator (MVUE) for a target parameter θ can be found if we find some function of the best sufficient statistic and this function is an unbiased estimator for θ. This method often works well. However, sometimes a best sufficient statistic is a fairly complicated function of the observable random variables in the sample. In cases like these it may be difficult to find a function of the sufficient statistic that is an unbiased estimator for the target parameter. For this reason two additional methods of finding estimators—the method of moments and the method of maximum likelihood—are presented in the next two sections. A third important method for estimation, the method of least squares, is the topic of Chapter 11.

EXERCISES

9.48 Refer to Exercise 9.30(b). Find an MVUE of σ^2.

9.49 Refer to Exercise 9.12. Is the estimator of σ^2 given there an MVUE of σ^2?

9.50 Refer to Exercise 9.32. Use $\sum_{i=1}^{n} Y_i^2$ to find an MVUE of θ.

9.51 The number of breakdowns Y per day for a certain machine is a Poisson random variable with mean λ. The daily cost of repairing these breakdowns is given by $C = 3Y^2$. If $Y_1, Y_2, \ldots, Y_n$ denote the observed number of breakdowns for n independently selected days, find an MVUE for $E(C)$.

9.52 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the probability density function

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

a. Show that this density function is in the (one-parameter) exponential family and that $\sum_{i=1}^{n} -\ln(Y_i)$ is sufficient for θ. (See Exercise 9.37.)

b. If $W_i = -\ln(Y_i)$, show that W_i has an exponential distribution with mean $1/\theta$.

c. Use methods similar to those in Example 9.10 to show that $2\theta \sum_{i=1}^{n} W_i$ has a χ^2 distribution with $2n$ degrees of freedom.

d. Show that

$$E\left(\frac{1}{2\theta \sum_{i=1}^{n} W_i}\right) = \frac{1}{2(n-1)}.$$

[Hint: Recall Exercise 4.89.]

e. What is the MVUE for θ?

9.53 Refer to Exercise 9.41. Use $Y_{(n)}$ to find an MVUE of θ. (See Example 9.1.)

9.54 Refer to Exercise 9.43. Find a function of $Y_{(1)}$ that is a minimum variance unbiased estimator for θ.

9.55 Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with density function

$$f(y \mid \theta) = \begin{cases} \dfrac{3y^2}{\theta^3} & 0 \le y \le \theta \\ 0, & \text{elsewhere.} \end{cases}$$

In Exercise 9.44 you showed that $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$ is sufficient for θ.

a. Show that $Y_{(n)}$ has probability density function

$$f_{(n)}(y \mid \theta) = \begin{cases} \dfrac{3ny^{3n-1}}{\theta^{3n}} & 0 \le y \le \theta \\ 0, & \text{elsewhere.} \end{cases}$$

b. Find the UMVUE of θ.

9.56 Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a normal distribution with mean μ and variance 1.

a. Show that the MVUE of μ^2 is $\widehat{\mu^2} = \overline{Y}^2 - \frac{1}{n}$.

b. Derive the variance of $\widehat{\mu^2}$.

***9.57** In this exercise, we illustrate the direct use of the Rao–Blackwell theorem. Let $Y_1, Y_2, \ldots, Y_n$ be independent Bernoulli random variables with

$$p(y_i \mid p) = p^{y_i}(1-p)^{1-y_i}, \quad y_i = 0, 1.$$

That is, $P(Y_i = 1) = p$ and $P(Y_i = 0) = 1 - p$. Find the MVUE of $p(1-p)$, which is a term in the variance of Y_i or $W = \sum_{i=1}^{n} Y_i$, by the following steps.

a. Let

$$T = \begin{cases} 1, & \text{if } Y_1 = 1 \text{ and } Y_2 = 0 \\ 0, & \text{otherwise.} \end{cases}$$

Show that $E(T) = p(1-p)$.

b. Show that

$$P(T = 1 \mid W = w) = \frac{w(n-w)}{n(n-1)}.$$

c. Show that

$$E(T \mid W) = \frac{n}{n-1}\left[\frac{W}{n}\left(1 - \frac{W}{n}\right)\right] = \frac{n}{n-1}\overline{Y}(1-\overline{Y})$$

and hence that $n\overline{Y}(1-\overline{Y})/(n-1)$ is the MVUE of $p(1-p)$.

***9.58** The likelihood function $L(y_1, y_2, \ldots, y_n \mid \theta)$ takes on different values depending on the arguments $(y_1, y_2, \ldots, y_n)$. A method for deriving a *minimal* sufficient statistic developed by Lehmann and Scheffé uses the ratio of the likelihoods evaluated at two points, $(x_1, x_2, \ldots, x_n)$ and $(y_1, y_2, \ldots, y_n)$:

$$\frac{L(x_1, x_2, \ldots, x_n \mid \theta)}{L(y_1, y_2, \ldots, y_n \mid \theta)}.$$

Many times it is possible to find a function $g(x_1, x_2, \ldots, x_n)$ such that this ratio is free of the unknown parameter θ if and only if $g(x_1, x_2, \ldots, x_n) = g(y_1, y_2, \ldots, y_n)$. If such a function g can be found, then $g(Y_1, Y_2, \ldots, Y_n)$ is a minimal sufficient statistic for θ.

a. Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a Bernoulli distribution (see Example 9.6 and Exercise 9.57) with p unknown.

i. Show that

$$\frac{L(x_1, x_2, \ldots, x_n \mid p)}{L(y_1, y_2, \ldots, y_n \mid p)} = \left(\frac{p}{1-p}\right)^{\Sigma x_i - \Sigma y_i}.$$

ii. Argue that for this ratio to be independent of p, we must have

$$\sum_{i=1}^{n} x_i - \sum_{i=1}^{n} y_i = 0 \qquad \text{or} \qquad \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i.$$

iii. Using the method of Lehmann and Scheffé, what is a minimal sufficient statistic for p? How does this sufficient statistic compare to the sufficient statistic derived in Example 9.6 by using the factorization criterion?

b. Consider the Weibull density discussed in Example 9.7.

i. Show that

$$\frac{L(x_1, x_2, \ldots, x_n \mid \theta)}{L(y_1, y_2, \ldots, y_n \mid \theta)} = \left(\frac{x_1 x_2 \cdots x_n}{y_1 y_2 \cdots y_n}\right) \exp\left[-\frac{1}{\theta}\left(\sum_{i=1}^{n} x_i^2 - \sum_{i=1}^{n} y_i^2\right)\right].$$

ii. Argue that $\sum_{i=1}^{n} Y_i^2$ is a minimal sufficient statistic for θ.

***9.59** Refer to Exercise 9.58. Suppose that a sample of size n is taken from a normal population with mean μ and variance σ^2. Show that $\sum_{i=1}^{n} Y_i$ and $\sum_{i=1}^{n} Y_i^2$, jointly, form minimal sufficient statistics for μ and σ^2.

***9.60** Suppose that a statistic U has a probability density function that is positive over the interval $a \le U \le b$, and suppose that the density depends on a parameter θ that can range over the interval $\alpha_1 \le \theta \le \alpha_2$. Suppose also that $g(u)$ is continuous for u in the interval $[a, b]$. If $E[g(U)] = 0$ for all θ in the interval $[\alpha_1, \alpha_2]$ implies that $g(u)$ is identically zero, then the family of density functions $f_U(u \mid \theta)$ is said to be *complete.* (All of the statistics that we employed in in Section 9.5 have complete families of density functions.) Suppose that U is a sufficient statistic for θ, and $g_1(U)$ and $g_2(U)$ are both unbiased estimators of θ. Show that, if the family of density functions for U is complete, $g_1(U)$ must equal $g_2(U)$, and thus there is a *unique* function of U that is an unbiased estimator of θ.

Coupled with the Rao–Blackwell theorem, the property of completeness of $f_U(u \mid \theta)$, along with the sufficiency of U, assures us that there is a unique minimum-variance unbiased estimator of θ (UMVUE).

9.6 The Method of Moments

In this section we will discuss one of the oldest methods for deriving point estimators: the method of moments. A more sophisticated method, the method of maximum likelihood, is the topic of Section 9.7.

The method of moments is a very simple procedure for finding an estimator for one or more population parameters. Recall that the kth moment of a random variable, taken about the origin, is

$$\mu_k' = E(Y^k).$$

The corresponding kth sample moment is the average

$$m_k' = \frac{1}{n}\sum_{i=1}^{n} Y_i^k.$$

The method of moments is based on the intuitively appealing idea that sample moments should provide good estimates of the corresponding population moments. That is, m_k' should be a good estimator of μ_k', for $k = 1, 2, \ldots$. Then because the population moments $\mu_1', \mu_2', \ldots, \mu_k'$ are functions of the population parameters, we can equate corresponding population and sample moments and solve for the desired estimators.

Hence the method of moments can be stated as follows.

Method of Moments

Choose as estimates those values of the parameters that are solutions of the equations $\mu'_k = m'_k$, for $k = 1, 2, \ldots, t$, where t is the number of parameters to be estimated.

EXAMPLE 9.11 A random sample of n observations, Y_1, Y_2, $\ldots$, Y_n, is selected from a population in which Y_i, for $i = 1, 2, \ldots, n$, possesses a uniform probability density function over the interval $(0, \theta)$ where θ is unknown. Use the method of moments to estimate the parameter θ.

Solution The value of μ'_1 for a uniform random variable is

$$\mu'_1 = \mu = \frac{\theta}{2}.$$

The corresponding first sample moment is

$$m'_1 = \frac{1}{n}\sum_{i=1}^{n} Y_i = \overline{Y}.$$

Equating the corresponding population and sample moment, we obtain

$$\mu_1' = \frac{\theta}{2} = \overline{Y}.$$

The method of moments estimator for θ is the solution of the above equation. That is, $\hat{\theta} = 2\overline{Y}$.

For the distributions we consider in this text, the methods of Section 9.3 can be used to show that sample moments are consistent estimators of the corresponding population moments. Because the estimators obtained from the method of moments obviously are functions of the sample moments, estimators obtained using the method of moments are usually consistent estimators of their respective parameters.

EXAMPLE 9.12 Show that the estimator $\hat{\theta} = 2\overline{Y}$, derived in Example 9.11, is a consistent estimator for θ.

Solution In Example 9.1, we showed that $\hat{\theta} = 2\overline{Y}$ is an unbiased estimator for θ and that $V(\hat{\theta}) = \theta^2/3n$. Because $\lim_{n\to\infty} V(\hat{\theta}) = 0$, Theorem 9.1 implies that $\hat{\theta} = 2\overline{Y}$ is a consistent estimator for θ.

Although the estimator $\hat{\theta}$ derived in Example 9.11 is consistent, it is not necessarily the best estimator for θ. Indeed, the factorization criterion yields $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$ to be the best sufficient statistic for θ. Thus, according to the

Rao–Blackwell theorem, the method of moments estimator will have larger variance than an unbiased estimator based on $Y_{(n)}$. This, in fact, was shown to be the case in Example 9.1.

EXAMPLE 9.13 A random sample of n observations, $Y_1, Y_2, \ldots, Y_n$, is selected from a population where Y_i, for $i = 1, 2, \ldots, n$, possesses a gamma probability density function with parameters α and β (see Section 4.6 for the gamma probability density function). Find method of moments estimators for the unknown parameters α and β.

Solution Because we seek estimators for two parameters α and β, we must equate two pairs of population and sample moments.

The first two moments of the gamma distribution with parameters α and β are (see the inside of the back cover of the text, if necessary)

$$\mu_1' = \mu = \alpha\beta \qquad \text{and} \qquad \mu_2' = \sigma^2 + \mu^2 = \alpha\beta^2 + \alpha^2\beta^2.$$

Now equate these quantities to their corresponding sample moments, and solve for $\hat{\alpha}$ and $\hat{\beta}$. Thus,

$$\begin{aligned} \mu_1' &= \alpha\beta = m_1' = \overline{Y} \\ \mu_2' &= \alpha\beta^2 + \alpha^2\beta^2 = m_2' = \frac{1}{n}\sum_{i=1}^{n} Y_i^2. \end{aligned}$$

From the first equation we obtain $\hat{\beta} = \overline{Y}/\hat{\alpha}$. Substituting into the second equation and solving for $\hat{\alpha}$, we obtain

$$\hat{\alpha} = \frac{\overline{Y}^2}{(\sum Y_i^2/n) - \overline{Y}^2} = \frac{n\overline{Y}^2}{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}.$$

Substituting $\hat{\alpha}$ into the first equation, we obtain

$$\hat{\beta} = \frac{\overline{Y}}{\hat{\alpha}} = \frac{\sum_{i=1}^{n}(Y_i - \overline{Y})^2}{n\overline{Y}}.$$

The method of moments estimators $\hat{\alpha}$ and $\hat{\beta}$ in Example 9.13 are consistent. $\overline{Y}$ converges in probability to $E(Y_i) = \alpha\beta$, and $(1/n)\sum_{i=1}^{n} Y_i^2$ converges in probability to $E(Y_i^2) = \alpha\beta^2 + \alpha^2\beta^2$. Thus,

$$\hat{\alpha} = \frac{\overline{Y}^2}{\frac{1}{n}\sum_{i=1}^{n} Y_i^2 - \overline{Y}^2} \qquad \text{is a consistent estimator of} \qquad \frac{(\alpha\beta)^2}{\alpha\beta^2 + \alpha^2\beta^2 - (\alpha\beta)^2} = \alpha$$

and

$$\hat{\beta} = \frac{\overline{Y}}{\hat{\alpha}} \qquad \text{is a consistent estimator of} \qquad \frac{\alpha\beta}{\alpha} = \beta.$$

Using the factorization criterion, we can show $\sum_{i=1}^{n} Y_i$ and the product $\prod_{i=1}^{n} Y_i$ to be sufficient statistics for the gamma density function. Because the method of moments estimators $\hat{\alpha}$ and $\hat{\beta}$ are not functions of these sufficient statistics, we can find more efficient estimators for the parameters α and β. However, it is considerably more difficult to apply other methods to find estimators for these parameters.

To summarize, the method of moments finds estimators of unknown parameters by equating corresponding sample and population moments. The method is easy to employ and provides consistent estimators. However, the estimators derived by this method are often not functions of sufficient statistics.. As a result, method of moments estimators are sometimes not very efficient. In many cases the method of moments estimators are biased. The primary virtues of this method are its ease of use and that it sometimes yields estimators with reasonable properties.

EXERCISES

9.61 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the probability density function

$$f(y \mid \theta) = \begin{cases} (\theta+1)y^{\theta}, & 0 < y < 1;\ \theta > -1 \\ 0, & \text{elsewhere.} \end{cases}$$

Find an estimator for θ by the method of moments. Show that the estimator is consistent. Is the estimator a function of the sufficient statistic $-\sum_{i=1}^{n} \ln(Y_i)$ that we can obtain from the factorization criterion? What implications does this have?

9.62 Suppose that $Y_1, Y_2, \ldots, Y_n$ constitute a random sample from a Poisson distribution with mean λ. Find the method of moments estimator of λ.

9.63 If $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the normal distribution with known mean $\mu = 0$ and unknown variance σ^2, find the method of moments estimator of σ^2.

9.64 If $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the normal distribution with mean μ and variance σ^2, find the method of moments estimators of μ and σ^2.

9.65 An urn contains θ black balls and $N - \theta$ white balls. A sample of n balls is to be selected without replacement. Let Y denote the number of black balls in the sample. Show that $(N/n)Y$ is the method of moments estimator of θ.

9.66 Let $Y_1, Y_2, \ldots, Y_n$ constitute a random sample from the probability density function given by

$$f(y \mid \theta) = \begin{cases} \left(\dfrac{2}{\theta^2}\right)(\theta - y), & 0 \leq y \leq \theta \\ 0, & \text{elsewhere.} \end{cases}$$

a. Find an estimator for θ by using the method of moments.

b. Is this estimator a sufficient statistic for θ?

9.67 Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from the probability density function given by

$$f(y \mid \theta) = \begin{cases} \dfrac{\Gamma(2\theta)}{[\Gamma(\theta)]^2}(y^{\theta-1})(1-y)^{\theta-1}, & 0 \leq y \leq 1 \\ 0, & \text{elsewhere.} \end{cases}$$

Find the method of moments estimator for θ.

9.68 Let $X_1, X_2, X_3, \ldots$ be independent random variables such that $P(X_i = 1) = p$

and $P(X_i = 0) = 1 - p$ for each $i = 1, 2, 3, \ldots$. Let the random variable Y denote the number of trials necessary to obtain the first success, that is, the value of i for which $X_i = 1$ first occurs. Then Y has a geometric distribution with $P(Y = y) = (1-p)^{y-1}p$, for $y = 1, 2, 3, \ldots$. Find the method of moments estimator of p based on this single observation Y.

9.69 Let $Y_1, Y_2, \ldots, Y_n$ denote independent and identically distributed uniform random variables on the interval $(0, 3\theta)$. Derive the method of moments estimator for θ.

9.70 Let $Y_1, Y_2, \ldots, Y_n$ denote independent and identically distributed random variables from a power family distribution with parameters α and $\theta = 3$. Then, as in Exercise 9.35, if $\alpha > 0$,

$$f(y|\alpha) = \begin{cases} \alpha y^{\alpha-1}/3^{\alpha}, & 0 \le y \le 3 \\ 0, & \text{elsewhere.} \end{cases}$$

Show that $E(Y_1) = 3\alpha/(\alpha+1)$, and derive the method of moments estimator for α.

***9.71** Let $Y_1, Y_2, \ldots, Y_n$ denote independent and identically distributed random variables from a Pareto distribution with parameters α and β, where β is known.

Then, if $\alpha > 0$,

$$f(y|\alpha, \beta) = \begin{cases} \alpha\beta^\alpha y^{-(\alpha+1)}, & y \geq \beta \\ 0, & \text{elsewhere.} \end{cases}$$

Show that $E(Y_i) = \alpha\beta/(\alpha - 1)$ if $\alpha > 1$ and $E(Y_i)$ is undefined if $0 < \alpha < 1$. Thus, the method of moments estimator for α is undefined.

9.7 The Method of Maximum Likelihood

In Section 9.5 we presented a method for deriving a minimum-variance unbiased estimator for a target parameter: using the factorization criterion together with the Rao–Blackwell theorem. The method requires that we find some function of a minimal sufficient statistic that is an unbiased estimator for the target parameter. Although we have a method for finding a sufficient statistic, the determination of the function of the minimal sufficient statistic that gives us an unbiased estimator can be largely a matter of hit or miss. Section 9.6 contained a discussion of the method of moments. The method of moments is intuitive and easy to apply but does not usually lead to the best estimators. In this section we present a method, the method of maximum likelihood, that often leads to minimum-variance unbiased estimators.

We will use an example to illustrate the logic upon which the method of maximum likelihood is based. Suppose that we are confronted with a box that contains three balls. We know that each of the balls may be red or white, but we do not know the total number of either color. However, we are allowed to randomly sample two of the balls, without replacement. If our random sample yields two red balls, what would be a good estimate of the total number of red balls in the box? Obviously the number of red balls in the box must be two or three (if there were 0 or 1 red ball in the box, it would be impossible to obtain two red balls when sampling without replacement). If there are two red balls and one white ball in the box, the probability of randomly selecting two red balls is

$$\binom{2}{2}\binom{1}{0}\Big/\binom{3}{2} = \frac{1}{3}.$$

On the other hand, if there are three red balls in the box, the probability of randomly selecting two red balls is

$$\binom{3}{2}\Big/\binom{3}{2} = 1.$$

It should seem reasonable to choose three as the estimate of the number of red balls in the box, because this estimate *maximizes the probability* of obtaining the observed sample. Of course, it is possible for the box to contain only two red balls, but the observed outcome gives more credence to there being three red balls in the box.

This example illustrates a method for finding an estimator that can be applied to any situation. The technique, called the *method of maximum likelihood*, selects as estimates the values of the parameters that maximize the likelihood (the joint probability function or joint density function) of the observed sample. (See Definition 9.4.) Recall that we referred to this method of estimation in Chapter 3 where, in Examples 3.10

and 3.13 and Exercise 3.83, respectively, we found the maximum likelihood estimates of the parameter p based on single observations on binomial, geometric and negative binomial random variables.

Method of Maximum Likelihood

Suppose that the likelihood function depends on k parameters $\theta_1, \theta_2, \ldots, \theta_k$. Choose as estimates those values of the parameters that maximize the likelihood $L(y_1,\, y_2,\, \ldots,\, y_n \,|\, \theta_1, \theta_2, \ldots, \theta_k)$.

To emphasize the fact that the likelihood function is a function of the parameters $\theta_1, \theta_2, \ldots, \theta_k$, we sometimes write the likelihood function as $L(\theta_1, \theta_2, \ldots, \theta_k)$. It is common to refer to maximum-likelihood estimators as MLEs. We will illustrate the method with an example.

EXAMPLE 9.14 A binomial experiment consisting of n trials resulted in observations $y_1,\, y_2,\, \ldots,\, y_n$, where $y_i = 1$ if the ith trial was a success, and $y_i = 0$ otherwise. Find the maximum-likelihood estimator (MLE) of p, the probability of a success.

Solution The likelihood of the observed sample is the probability of observing $y_1, y_2, \ldots, y_n$. Hence

$$L(p) = L(y_1,\, y_2,\, \ldots,\, y_n \,|\, p) = p^y(1-p)^{n-y} \qquad \text{where} \qquad y = \sum_{i=1}^{n} y_i.$$

We now wish to find the value of p that maximizes $L(p)$. If $y = 0$, $L(p) = (1-p)^n$, and $L(p)$ is maximized when $p = 0$. Analogously, if $y = n$, $L(p) = p^n$, and $L(p)$ is maximized when $p = 1$. If $y = 1,\ 2, \ldots,\ n-1$, then $L(p) = p^y(1-p)^{n-y}$ is zero at $p = 0$ and $p = 1$ and is continuous for values of p between 0 and 1. Thus, for $y = 1,\ 2, \ldots,\ n-1$, we can find the value of p that maximizes $L(p)$ by setting the derivative $dL(p)/dp$ equal to zero and solving for p.

You will notice that $\ln[L(p)]$ is a monotonically increasing function of $L(p)$. Hence both $\ln[L(p)]$ and $L(p)$ are maximized for the same value of p. Because $L(p)$ is a product of functions of p, and finding the derivative of products is tedious, it is easier to find the value of p that maximizes $\ln[L(p)]$. We have

$$\ln[L(p)] = \ln\left[p^y(1-p)^{n-y}\right] = y\ln p + (n-y)\ln(1-p).$$

If $y = 1,\ 2, \ldots,\ n-1$, the derivative of $\ln[L(p)]$ with respect to p, is

$$\frac{d\ln[L(p)]}{dp} = y\left(\frac{1}{p}\right) + (n-y)\left(\frac{-1}{1-p}\right).$$

For $y = 1, 2, \ldots, n-1$, the value of p that maximizes (or minimizes) $\ln[L(p)]$ is the solution of the equation

$$\frac{y}{\hat{p}} - \frac{n-y}{1-\hat{p}} = 0.$$

Solving, we obtain the estimate $\hat{p} = y/n$. You can easily verify that this solution occurs when $\ln[L(p)]$ (and hence $L(p)$) achieves a maximum.

Since $L(p)$ is maximized at $p = 0$ when $y = 0$, at $p = 1$ when $y = n$ and at $p = y/n$ when $y = 1, 2, \ldots, n-1$, whatever the observed value of y, $L(p)$ is maximized when $p = y/n$.

The maximum-likelihood *estimator*, (MLE), $\hat{p} = Y/n$ is the fraction of successes in the total number of trials n. Hence the MLE of p is is actually the intuitive estimator for p that we used throughout Chapter 8.

EXAMPLE 9.15 Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a normal distribution with mean μ and variance σ^2. Find the maximum-likelihood estimators (MLEs) of μ and σ^2.

Solution Because $Y_1, Y_2, \ldots, Y_n$ are continuous random variables, $L(\mu, \sigma^2)$ is the joint density of the sample. Thus $L(\mu, \sigma^2) = f(y_1, y_2, \ldots, y_n \mid \mu, \sigma^2)$. In this case

$$\begin{aligned} L(\mu, \sigma^2) &= f(y_1, y_2, \ldots, y_n \mid \mu, \sigma^2) \\ &= f(y_1 \mid \mu, \sigma^2) \times f(y_2 \mid \mu, \sigma^2) \times \cdots \times f(y_n \mid \mu, \sigma^2) \end{aligned}$$

$$
\begin{aligned}
&= \left\{\frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(y_1-\mu)^2}{2\sigma^2}\right]\right\} \times \cdots \times \left\{\frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-(y_n-\mu)^2}{2\sigma^2}\right]\right\} \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2\right]
\end{aligned}
$$

[Recall that $\exp(w)$ is just another way of writing e^w.] Further,

$$
\ln\left[L(\mu,\,\sigma^2)\right] = -\frac{n}{2}\ln\sigma^2 - \frac{n}{2}\ln 2\pi - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i-\mu)^2.
$$

The MLEs of μ and σ^2 are the values that make $\ln\left[L(\mu,\,\sigma^2)\right]$ a maximum. Taking derivatives with respect to μ and σ^2, we obtain

$$
\frac{\partial\left\{\ln\left[L(\mu,\,\sigma^2)\right]\right\}}{\partial\mu} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i-\mu)
$$

and

$$
\frac{\partial\left\{\ln\left[L(\mu,\,\sigma^2)\right]\right\}}{\partial\sigma^2} = -\left(\frac{n}{2}\right)\left(\frac{1}{\sigma^2}\right) + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i-\mu)^2.
$$

Setting these derivatives equal to zero and solving simultaneously, we obtain from the first equation

$$
\frac{1}{\hat{\sigma}^2}\sum_{i=1}^{n}(y_i-\hat{\mu}) = 0 \qquad \text{or} \qquad \sum_{i=1}^{n} y_i - n\hat{\mu} = 0 \qquad \text{and} \qquad \hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i = \overline{y}.
$$

Substituting $\overline{y}$ for $\hat{\mu}$ in the second equation and solving for $\hat{\sigma}^2$, we have

$$-\left(\frac{n}{\hat{\sigma}^2}\right) + \frac{1}{\hat{\sigma}^4}\sum_{i=1}^{n}(y_i - \overline{y})^2 = 0 \qquad \text{or} \qquad \hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y})^2.$$

Thus $\overline{Y}$ and $\hat{\sigma}^2$ are the maximum-likelihood estimators of μ and σ^2, respectively. Notice that $\overline{Y}$ is unbiased for μ. Although $\hat{\sigma}^2$ is not unbiased for σ^2, it can easily be adjusted to the unbiased estimator S^2 (see Example 8.1).

EXAMPLE 9.16 Let $Y_1, Y_2, \ldots, Y_n$ be a random sample of observations from a uniform distribution with probability density function $f(y_i \mid \theta) = 1/\theta$, for $0 \leq y_i \leq \theta$ and $i = 1, 2, \ldots, n$. Find the MLE of θ.

Solution In this case the likelihood is given by

$$\begin{aligned} L(\theta) &= f(y_1, y_2, \ldots, y_n \mid \theta) = f(y_1 \mid \theta) \times f(y_2 \mid \theta) \times \cdots \times f(y_n \mid \theta) \\ &= \begin{cases} \dfrac{1}{\theta} \times \dfrac{1}{\theta} \times \cdots \times \dfrac{1}{\theta} = \dfrac{1}{\theta^n}, & \text{if all of the } y_i \text{ values are between 0 and } \theta \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

Obviously, $L(\theta)$ is not maximized when $L(\theta) = 0$. You will notice that $1/\theta^n$ is a monotonically decreasing function of θ. Hence nowhere in the interval $0 < \theta < \infty$ is $d[1/\theta^n]/d\theta$ equal to zero. However, $1/\theta^n$ increases as θ decreases, and $1/\theta^n$ is maximized by selecting θ to be as small as possible, subject to the constraint that all of the y_i values are between 0 and θ. The smallest value of θ that satisfies this constraint is the maximum observation in the set $y_1, y_2, \ldots, y_n$. That is, $\hat{\theta} = Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$ is the MLE for θ. This MLE for θ is not an unbiased estimator of θ, but it can be adjusted to be unbiased, as shown in Example 9.1.

We have seen that sufficient statistics that best summarize the data have desirable properties and often can be used to find an MVUE for parameters of interest. If U is *any* sufficient statistic for the estimation of a parameter θ, including the sufficient statistic obtained from the optimal use of the factorization criterion, the maximum-likelihood estimator is always some function of U. That is, the MLE depends on the sample observations only through the value of a sufficient statistic. To show this, we need only observe that, if U is a sufficient statistic for θ, the factorization criterion (Theorem 9.4) implies that the likelihood can be factored as

$$L(\theta) = L(y_1, y_2, \ldots, y_n \mid \theta) = g(u, \theta)h(y_1, y_2, \ldots, y_n)$$

where $g(u, \theta)$ is a function of only u and θ and $h(y_1, y_2, \ldots, y_n)$ *does not depend on* θ. Therefore, it follows that

$$\ln[L(\theta)] = \ln[g(u, \theta)] + \ln[h(y_1, y_2, \ldots, y_n)].$$

Notice that $\ln[h(y_1, y_2, \ldots, y_n)]$ does not depend on θ and, therefore, that maximizing

$\ln[L(\theta)]$ relative to θ is equivalent to maximizing $\ln[g(u, \theta)]$ relative to θ. Because $\ln[g(u, \theta)]$ depends on the data only through the value of the sufficient statistic U, it follows that $\hat{\theta}$ depends on the data only through the value of U. That is, if U is *any* sufficient statistic for estimating θ, the maximum-likelihood estimator (MLE) is always some function of U. Consequently, if a maximum-likelihood estimator for a parameter can be found and then adjusted by a constant to be unbiased, the resulting estimator often is an MVUE of the parameter in question. These realizations imply that the method of maximum likelihood can be a very useful tool in finding estimators with good properties.

Maximum-likelihood estimators (MLEs) have some additional properties that make this method of estimation particularly attractive. In Example 9.9, we considered estimation of θ^2, a function of the parameter θ. Functions of other parameters may also be of interest. For example, the variance of a binomial random variable is $np(1-p)$, a function of the parameter p. If Y has a Poisson distribution with mean λ, it follows that $P(Y = 0) = e^{-\lambda}$; we may wish to estimate this function of λ. Generally, if θ is the parameter associated with a distribution, we are sometimes interested in estimating some function of θ, say, $t(\theta)$, rather than θ itself. In Exercise 9.76, you will prove that, if $t(\theta)$ is a one-to-one function of θ and if $\hat{\theta}$ is the MLE for θ, then the MLE of $t(\theta)$ is given by

$$\widehat{t(\theta)} = t(\hat{\theta}).$$

This result, sometimes referred to as the *invariance property* of MLEs, also holds for any function of a parameter of interest (not just one-to-one functions). See Casella and Berger (1990) for details.

EXAMPLE 9.17 In Example 9.14, we found that the MLE of the binomial proportion p is given by $\hat{p} = Y/n$. What is the MLE for the variance of Y?

Solution The variance of a binomial random variable Y is given by $V(Y) = np(1-p)$. Since $V(Y)$ is a function of the binomial parameter p, namely, $V(Y) = t(p)$ with $t(p) = np(1-p)$, it follows that the MLE of $V(Y)$ is given by

$$\widehat{V(Y)} = \widehat{t(p)} = t(\hat{p}) = n\left(\frac{Y}{n}\right)\left(1 - \frac{Y}{n}\right).$$

This estimator is not unbiased. However, using the result in Exercise 9.57, we can easily adjust it to make it unbiased. Actually,

$$n\left(\frac{Y}{n}\right)\left(1 - \frac{Y}{n}\right)\left(\frac{n}{n-1}\right) = \left(\frac{n^2}{n-1}\right)\left(\frac{Y}{n}\right)\left(1 - \frac{Y}{n}\right)$$

is the UMVUE for $t(p) = np(1-p)$.

In the next section (optional), we summarize some of the convenient and useful large-sample properties of maximum-likelihood estimators.

EXERCISES

9.72 Suppose that $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the Poisson distribution with mean λ.

a. Find the maximum-likelihood estimator $\hat{\lambda}$ for λ.

b. Find the expected value and variance of $\hat{\lambda}$.

c. Show that the estimator of (a) is consistent for λ.

d. What is the MLE for $P(Y = 0) = e^{-\lambda}$?

9.73 Suppose that $Y_1, Y_2, \ldots, Y_n$ denote a random sample from an exponentially distributed population with mean θ. Find the MLE of the population variance θ^2. [Hint: Recall Example 9.9.]

9.74 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the density function given by

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

where r is a known positive constant.

a. Find a sufficient statistic for θ.

b. Find the maximum-likelihood estimator of θ.

c. Is the estimator in part (b) an MVUE for θ?

9.75 Suppose that $Y_1, Y_2, \ldots, Y_n$ constitute a random sample from a uniform distribution with probability density function

$$f(y \mid \theta) = \begin{cases} \dfrac{1}{2\theta + 1}, & 0 \leq y \leq 2\theta + 1 \\ 0, & \text{otherwise.} \end{cases}$$

a. Obtain the maximum-likelihood estimator of θ.

b. Obtain the MLE for the *variance* of the underlying distribution.

9.76 A certain type of electronic component has a lifetime Y (in hours) with probability density function given by

$$f(y \mid \theta) = \begin{cases} \left(\dfrac{1}{\theta^2}\right) y e^{-y/\theta}, & y > 0 \\ 0, & \text{otherwise.} \end{cases}$$

That is, Y has a gamma distribution with parameters $\alpha = 2$ and θ. Let $\hat{\theta}$ denote the maximum-likelihood estimator of θ. Suppose that three such components, tested independently, had lifetimes of 120, 130, and 128 hours.

a. Find the maximum-likelihood estimate of θ.

b. Find $E(\hat{\theta})$ and $V(\hat{\theta})$.

c. Suppose that θ actually equals 130. Give an approximate bound that you might expect for the error of estimation.

d. What is the MLE for the variance of Y?

9.77 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the density function given by

$$f(y \mid \alpha, \theta) = \begin{cases} \left(\dfrac{1}{\Gamma(\alpha)\theta^\alpha}\right) y^{\alpha-1} e^{-y/\theta}, & y > 0 \\ 0, & \text{elsewhere} \end{cases}$$

where $\alpha > 0$ is known

a. Find the maximum-likelihood estimator $\hat{\theta}$ of θ.

b. Find the expected value and variance of $\hat{\theta}$.

c. Show that $\hat{\theta}$ is consistent for θ.

d. What is the best (minimal) sufficient statistic for θ in this problem?

e. Suppose that $n = 5$ and $\alpha = 2$. Use the minimal sufficient statistic to construct a 90% confidence interval for θ. [Hint: Transform to a χ^2 distribution.]

9.78 Suppose that $X_1, X_2, \ldots, X_m$, representing yields per acre for corn variety A, constitute a random sample from a normal distribution with mean μ_1 and variance σ^2. Also, $Y_1 Y_2, \ldots, Y_n$, representing yields for corn variety B, constitute a random sample from a normal distribution with mean μ_2 and variance σ^2. If the X's and Y's are independent, find the maximum-likelihood estimator for the common variance σ^2. Assume that μ_1 and μ_2 are unknown.

9.79 A random sample of 100 voters selected from a large population revealed 30 favoring candidate A, 38 favoring candidate B, and 32 favoring candidate C. Find maximum-likelihood estimates for the proportions of voters in the population favoring candidates A, B, and C, respectively. Estimate the difference between the fractions favoring A and B, and place a two-standard-deviation bound on the error of estimation.

9.80 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from the probability density function

$$f(y \mid \theta) = \begin{cases} (\theta+1)y^{\theta}, & 0 < y < 1\,; \quad \theta > -1 \\ 0, & \text{elsewhere.} \end{cases}$$

Find the maximum-likelihood estimator for θ. Compare your answer to the method of moments estimator found in Exercise 9.61.

9.81 It is known that the probability p of tossing heads on an unbalanced coin is either $1/4$ or $3/4$. The coin is tossed twice and a value for Y, the number of heads, is observed. For each possible value of Y, which of the two values for p ($1/4$ or $3/4$) maximizes the probability that $Y = y$? Depending on the value of y actually observed, what is the maximum-likelihood estimator of p?

9.82 A random sample of 100 men produced a total of 25 who favored a controversial local issue. An independent random sample of 100 women produced a total of 30 who favored the issue. Assume that p_M is the true underlying proportion of men who favor the issue and that p_W is the true underlying proportion of women who favor of the issue. If it actually is true that $p_W = p_M = p$, find the maximum-likelihood estimator of the common proportion p.

***9.83** Find the maximum-likelihood estimator of θ based on a random sample of size n from a uniform distribution on the interval $(0, 2\theta)$.

***9.84** Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with density function

$$f(y \mid \theta) = \begin{cases} \dfrac{3y^2}{\theta^3} & 0 \le y \le \theta \\ 0, & \text{elsewhere.} \end{cases}$$

In Exercise 9.44 you showed that $Y_{(n)} = \max(Y_1, Y_2, \ldots, Y_n)$ is sufficient for θ.

a. Find the MLE for θ. Hint : see Example 9.16.

b. Find a function of the MLE in part (a) that is a pivotal quantity. Hint : see Exercise 9.55.

c. Use the pivotal quantity from part (b) to find a $(1-\alpha)100\%$ confidence interval for θ.

***9.85** Let $Y_1, Y_2, \ldots, Y_n$ be a random sample from a population with density function

$$f(y \mid \theta) = \begin{cases} \dfrac{2\theta^2}{y^3} & \theta < y < \infty \\ 0, & \text{elsewhere.} \end{cases}$$

In Exercise 9.45 you showed that $Y_{(1)} = \min(Y_1, Y_2, \ldots, Y_n)$ is sufficient for θ.

a. Find the MLE for θ. Hint : see Example 9.16.

b. Find a function of the MLE in part (a) that is a pivotal quantity.

c. Use the pivotal quantity from part (b) to find a $(1-\alpha)100\%$ confidence interval for θ.

***9.86** Suppose that $\hat{\theta}$ is the maximum-likelihood estimator for a parameter θ. Let $t(\theta)$ be a function of θ that possesses a unique inverse [that is, if $\beta = t(\theta)$, then $\theta = t^{-1}(\beta)$]. Show that $t(\hat{\theta})$ is the maximum-likelihood estimator of $t(\theta)$.

***9.87** A random sample of n items is selected from the large number of items produced by a certain production line in one day. Find the maximum-likelihood estimator of the ratio R, the proportion of defective items divided by the proportion of good items.

9.88 Consider a random sample of size n from a normal population with mean μ and variance σ^2, both unknown. Derive the maximum-likelihood estimator of σ.

9.8 Some Large-Sample Properties of MLEs (Optional)

Maximum-likelihood estimators also have interesting large–sample properties. Suppose that $t(\theta)$ is a differentiable function of θ. In Section 9.7, we argued by the invariance property that, if $\hat{\theta}$ is the MLE of θ, then the MLE of $t(\theta)$ is given by $t(\hat{\theta})$. Under some conditions of regularity that hold for the distributions we will consider, $t(\hat{\theta})$ is a *consistent* estimator for $t(\theta)$. In addition, for large sample sizes,

$$Z = \frac{t(\hat{\theta}) - t(\theta)}{\sqrt{\left[\frac{\partial t(\theta)}{\partial \theta}\right]^2 \bigg/ nE\left[-\frac{\partial^2 \ln f(Y \mid \theta)}{\partial \theta^2}\right]}}$$

has approximately a standard normal distribution. In this expression, the quantity $f(Y \mid \theta)$ in the denominator is the density function corresponding to the continuous distribution of interest, evaluated at the random value Y. In the discrete case, the analogous result holds with the probability function evaluated at the random value Y, $p(Y \mid \theta)$ substituted for the density $f(Y \mid \theta)$. If we desire a confidence interval for $t(\theta)$, we can use quantity Z as a pivotal quantity. If we proceed as in Section 8.6, we obtain the following approximate large sample $100(1 - \alpha)\%$ confidence interval for $t(\theta)$:

$$t(\hat{\theta}) \pm z_{\alpha/2}\sqrt{\left[\frac{\partial t(\theta)}{\partial \theta}\right]^2 \bigg/ nE\left[-\frac{\partial^2 \ln f(Y \mid \theta)}{\partial \theta^2}\right]} \approx$$
$$t(\hat{\theta}) \pm z_{\alpha/2}\sqrt{\left(\left[\frac{\partial t(\theta)}{\partial \theta}\right]^2 \bigg/ nE\left[-\frac{\partial^2 \ln f(Y \mid \theta)}{\partial \theta^2}\right]\right)\Bigg|_{\theta=\hat{\theta}}}.$$

We will illustrate this with the following example.

EXAMPLE 9.18 For random variable with a *Bernoulli* distribution, $p(y \mid p) = p^y(1-p)^{1-y}$, for $y = 0, 1$. If $Y_1, Y_2, \ldots, Y_n$ denote a random sample of size n from this distribution, derive a $100(1 - \alpha)\%$ confidence interval for $p(1 - p)$, the variance associated with this distribution.

Solution As in Example 9.14, the MLE of the parameter p is given by $\hat{p} = W/n$ where $W = \sum_{i=1}^{n} Y_i$. It follows that the MLE for $t(p) = p(1-p)$ is $\widehat{t(p)} = \hat{p}(1-\hat{p})$.

In this case,

$$\begin{aligned} t(p) &= p(1-p) = p - p^2 \qquad \text{and} \\ \frac{\partial t(p)}{\partial p} &= 1 - 2p. \end{aligned}$$

Also,

$$\begin{aligned} p(y \,|\, p) &= p^y (1-p)^{1-y} \\ \ln\left[p(y \,|\, p)\right] &= y(\ln p) + (1-y)\ln(1-p) \\ \frac{\partial \ln\left[p(y \,|\, p)\right]}{\partial p} &= \frac{y}{p} - \frac{1-y}{1-p} \\ \frac{\partial^2 \ln\left[p(y \,|\, p)\right]}{\partial p^2} &= -\frac{y}{p^2} - \frac{1-y}{(1-p)^2} \\ E\left[-\frac{\partial^2 \ln\left[p(Y \,|\, p)\right]}{\partial p^2}\right] &= E\left[\frac{Y}{p^2} + \frac{1-Y}{(1-p)^2}\right] \\ &= \frac{p}{p^2} + \frac{1-p}{(1-p)^2} = \frac{1}{p} + \frac{1}{1-p} = \frac{1}{p(1-p)}. \end{aligned}$$

Substituting into the earlier formula for the confidence interval for $t(\theta)$, we obtain

$$t(\hat{p}) \pm z_{\alpha/2} \sqrt{\left(\left[\frac{\partial t(p)}{\partial p}\right]^2 \Bigg/ nE\left[-\frac{\partial^2 \ln p(Y \,|\, p)}{\partial p^2}\right] \right)\Bigg|_{p=\hat{p}}}$$

$$= \quad \hat{p}(1-\hat{p}) \pm z_{\alpha/2}\sqrt{\left((1-2p)^2 \Big/ n \left[\frac{1}{p(1-p)}\right] \right)}\Bigg|_{p=\hat{p}}$$

$$= \quad \hat{p}(1-\hat{p}) \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})(1-2\hat{p})^2}{n}}$$

as the desired confidence interval for $p(1-p)$.

EXERCISES

***9.89** Consider the distribution discussed in Example 9.18. Use the method presented in Section 9.8 to derive a $100(1-\alpha)\%$ confidence interval for $t(p) = p$. Is the resulting interval familiar to you?

***9.90** Suppose that $Y_1, Y_2, \ldots, Y_n$ constitute a random sample of size n from an exponential distribution with mean θ. Find a $100(1-\alpha)\%$ confidence interval for $t(\theta) = \theta^2$.

***9.91** Let Y_1, Y_2, $\ldots$, Y_n denote a random sample of size n from a Poisson distribution with mean λ. Find a $100(1-\alpha)\%$ confidence interval for $t(\lambda) = e^{-\lambda} = P(Y = 0)$.

9.9 Summary

In this chapter we continued and extended the discussion of estimation begun in Chapter 8. Good estimators are consistent and efficient when compared to other estimators. The most efficient estimators, those with the smallest variances, are functions of the sufficient statistics that best summarize all of the information about the parameter of interest.

Two methods of finding estimators—the method of moments and the method of maximum likelihood—were presented. Moment estimators are consistent but generally not very efficient. Maximum-likelihood estimators, on the other hand, are consistent and, if adjusted to be unbiased, often lead to minimum-variance unbiased estimators. Since maximum-likelihood estimators have many good properties, this method is a popular method of estimation.

References and Further Readings

Casella, G., and Berger, R. L. *Statistical Inference.* Pacific Grove, CA: Brooks Cole, 1990.

Cramer, H. *Mathematical Methods of Statistics.* Princeton, NJ: Princeton University Press, 1963.

Hogg, R. V., and Craig, A. T. *Introduction to Mathematical Statistics*, 5th ed., Englewood Cliffs, NJ: Prentice Hall, 1995.

Lindgren, B. W. *Statistical Theory*, 4th ed. New York: Chapman and Hall, 1993.

Mood, A. M.; Graybill, F. A.; and Boes, D. *Introduction to the Theory of Statistics*, 3d ed. New York: McGraw-Hill, 1974.

Serfling, R. J. *Approximation Theorems of Mathematical Statistics.* New York:

John Wiley, 1980.

Wilks, S. S. *Mathematical Statistics.* New York: John Wiley, 1963.

SUPPLEMENTARY EXERCISES

9.92 Suppose that $Y_1, Y_2, \ldots, Y_n$ constitute a random sample from the density function

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

where θ is an unknown, positive constant.

a. Find an estimator $\hat{\theta}_1$ for θ by the method of moments.

b. Find an estimator $\hat{\theta}_2$ for θ by the method of maximum likelihood.

c. Adjust $\hat{\theta}_1$ and $\hat{\theta}_2$ so that they are unbiased. Find the efficiency of the adjusted $\hat{\theta}_1$ relative to the adjusted $\hat{\theta}_2$.

9.93 Refer to Exercise 9.30(b). Under the conditions outlined there, find the maximum-likelihood estimator of σ^2.

***9.94** Suppose that $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a Poisson distribution with mean λ. Find the MVUE of $P(Y_i = 0) = e^{-\lambda}$. [Hint: Make use of the Rao–Blackwell theorem.]

9.95 Suppose that a random sample of length-of-life measurements, $Y_1, Y_2, \ldots, Y_n$, is to be taken of components whose length of life has an exponential distribution with mean θ. It is frequently of interest to estimate

$$\overline{F}(t) = 1 - F(t) = e^{-t/\theta}$$

the *reliability* at time t of such a component. For any fixed value of t, find the maximum-likelihood estimator of $\overline{F}(t)$.

***9.96** The maximum-likelihood estimator obtained in Exercise 9.95 is a function of the minimal sufficient statistic for θ, but it is not unbiased. Use the Rao–Blackwell theorem to find the MVUE of $e^{-t/\theta}$ by the following steps.

a. Let

$$V = \begin{cases} 1, & Y_1 > t \\ 0, & \text{elsewhere.} \end{cases}$$

Show that V is an unbiased estimator of $e^{-t/\theta}$.

b. Because $U = \sum_{i=1}^{n} Y_i$ is the minimal sufficient statistic for θ, show that the conditional density function for Y_1, given $U = u$, is

$$f_{Y_1|U}(y_1 \mid u) = \begin{cases} \left(\dfrac{n-1}{u^{n-1}}\right)(u - y_1)^{n-2}, & 0 < y_1 < u \\ 0, & \text{elsewhere.} \end{cases}$$

c. Show that

$$E(V|U) = P(Y_1 > t|U) = \left(1 - \frac{t}{U}\right)^{n-1}.$$

This is the MVUE of $e^{-t/\theta}$ by the Rao–Blackwell theorem and by the fact that the density function for U is complete.

***9.97** Suppose that n integers are drawn at random and *with replacement* from the integers $1, 2, \ldots, N$. That is, each sampled integer has probability $1/N$ of taking on any of the values $1, 2, \ldots, N$, and the sampled values are independent.

a Find the method of moments estimator $\hat{N}_1$ of N.

b Find $E(\hat{N}_1)$ and $V(\hat{N}_1)$.

***9.98** Refer to Exercise 9.97.

a Find the maximum-likelihood estimator $\hat{N}_2$ of N.

b Show that $E(\hat{N}_2)$ is approximately $[n/(n+1)]N$. Adjust $\hat{N}_2$ to form an estimator $\hat{N}_3$ that is approximately unbiased for N.

c Find an approximate variance for $\hat{N}_3$ by using the fact that, for large N, the variance of the largest sampled integer is approximately

$$\frac{nN^2}{(n+1)^2(n+2)}.$$

d Show that for large N and $n > 1$, $V(\hat{N}_3) < V(\hat{N}_1)$.

***9.99** Suppose that enemy tanks have serial numbers $1, 2, \ldots, N$. A spy randomly observed five tanks (with replacement) with serial numbers 97, 64, 118, 210, and 57. Estimate N and place a bound on the error of estimation.

9.100 Let $Y_1, Y_2, \ldots, Y_n$ denote a random sample from a Poisson distribution with mean λ and define

$$W_n = \frac{\overline{Y} - \lambda}{\sqrt{\overline{Y}/n}}.$$

a. Show that the distribution of W_n converges to a standard normal distribution.

b. Use W_n and the result in part (a) to derive the formula for an approximate 95% confidence interval for λ.