Point Estimation

1 Introduction

Traditionally, problems of statistical inference are divided into **problems of estimation** and **tests of hypotheses**, though actually they are all decision problems and, hence, could be handled by the unified approach. The main difference between the two kinds of problems is that in problems of estimation we must determine the value of a parameter (or the values of several parameters) from a possible continuum of alternatives, whereas in tests of hypotheses we must decide whether to accept or reject a specific value or a set of specific values of a parameter (or those of several parameters).

DEFINITION 1. POINT ESTIMATION. *Using the value of a sample statistic to estimate the value of a population parameter is called* ***point estimation****. We refer to the value of the statistic as a* ***point estimate****.*

For example, if we use a value of $\overline{X}$ to estimate the mean of a population, an observed sample proportion to estimate the parameter θ of a binomial population, or a value of S^2 to estimate a population variance, we are in each case using a point estimate of the parameter in question. These estimates are called point estimates because in each case a single number, or a single point on the real axis, is used to estimate the parameter.

Correspondingly, we refer to the statistics themselves as **point estimators**. For instance, $\overline{X}$ may be used as a point estimator of μ, in which case $\bar{x}$ is a point estimate of this parameter. Similarly, S^2 may be used as a point estimator of σ^2, in which case s^2 is a point estimate of this parameter. Here we used the word "point" to distinguish between these estimators and estimates and the **interval estimators** and **interval estimates**.

Since estimators are random variables, one of the key problems of point estimation is to study their sampling distributions. For instance, when we estimate the variance of a population on the basis of a random sample, we can hardly expect that the value of S^2 we get will actually equal σ^2, but it would be reassuring, at least, to know whether we can expect it to be close. Also, if we must decide whether to use a sample mean or a sample median to estimate the mean of a population, it would be important to know, among other things, whether $\overline{X}$ or $\widetilde{X}$ is more likely to yield a value that is actually close.

From Chapter 10 of *John E. Freund's Mathematical Statistics with Applications*, Eighth Edition. Irwin Miller, Marylees Miller.

Various statistical properties of estimators can thus be used to decide which estimator is most appropriate in a given situation, which will expose us to the smallest risk, which will give us the most information at the lowest cost, and so forth. The particular properties of estimators that we shall discuss in Sections 2 through 6 are **unbiasedness**, **minimum variance**, **efficiency**, **consistency**, **sufficiency**, and **robustness**.

2 Unbiased Estimators

Perfect decision functions do not exist, and in connection with problems of estimation this means that there are no perfect estimators that always give the right answer. Thus, it would seem reasonable that an estimator should do so at least on the average; that is, its expected value should equal the parameter that it is supposed to estimate. If this is the case, the estimator is said to be **unbiased**; otherwise, it is said to be **biased**. Formally, this concept is expressed by means of the following definition.

DEFINITION 2. UNBIASED ESTIMATOR. *A statistic* $\hat{\Theta}$ *is an* ***unbiased estimator*** *of the parameter* θ *of a given distribution if and only if* $E(\hat{\Theta}) = \theta$ *for all possible values of* θ.

The following are some examples of unbiased and biased estimators.

EXAMPLE 1

Definition 2 requires that $E(\Theta) = \theta$ for all possible values of θ. To illustrate why this statement is necessary, show that unless $\theta = \frac{1}{2}$, the minimax estimator of the binomial parameter θ is biased.

Solution

Since $E(X) = n\theta$, it follows that

$$E\left(\frac{X + \frac{1}{2}\sqrt{n}}{n + \sqrt{n}}\right) = \frac{E\left(X + \frac{1}{2}\sqrt{n}\right)}{n + \sqrt{n}} = \frac{n\theta + \frac{1}{2}\sqrt{n}}{n + \sqrt{n}}$$

and it can easily be seen that this quantity does not equal θ unless $\theta = \frac{1}{2}$.

EXAMPLE 2

If X has the binomial distribution with the parameters n and θ, show that the sample proportion, $\frac{X}{n}$, is an unbiased estimator of θ.

Solution

Since $E(X) = n\theta$, it follows that

$$E\left(\frac{X}{n}\right) = \frac{1}{n} \cdot E(X) = \frac{1}{n} \cdot n\theta = \theta$$

and hence that $\frac{X}{n}$ is an unbiased estimator of θ.

EXAMPLE 3

If $X_1, X_2, \ldots, X_n$ constitute a random sample from the population given by

$$f(x) = \begin{cases} e^{-(x-\delta)} & \text{for } x > \delta \\ 0 & \text{elsewhere} \end{cases}$$

show that $\overline{X}$ is a biased estimator of δ.

Solution

Since the mean of the population is

$$\mu = \int_{\delta}^{\infty} x \cdot e^{-(x-\delta)} dx = 1 + \delta$$

it follows from the theorem "If $\overline{X}$ is the mean of a random sample of size n taken without replacement from a finite population of size N with the mean μ and the variance σ^2, then $E(\overline{X}) = \mu$ and $\text{var}(\overline{X}) = \frac{\sigma^2}{n} \cdot \frac{N-n}{N-1}$" that $E(\overline{X}) = 1 + \delta \neq \delta$ and hence that $\overline{X}$ is a biased estimator of δ.

When $\hat{\Theta}$, based on a sample of size n from a given population, is a biased estimator of θ, it may be of interest to know the extent of the **bias**, given by

$$b_n(\theta) = E(\hat{\Theta}) - \theta$$

Thus, for Example 1 the bias is

$$\frac{n\theta + \frac{1}{2}\sqrt{n}}{n + \sqrt{n}} - \theta = \frac{\frac{1}{2} - \theta}{\sqrt{n} + 1}$$

and it can be seen that it tends to be small when θ is close to $\frac{1}{2}$ and also when n is large.

DEFINITION 3. ASYMPTOTICALLY UNBIASED ESTIMATOR. *Letting* $b_n(\theta) = E(\hat{\Theta}) - \theta$ *express the* ***bias*** *of an estimator* $\hat{\Theta}$ *based on a random sample of size* n *from a given distribution, we say that* $\hat{\Theta}$ *is an* ***asymptotically unbiased estimator*** *of* θ *if and only if*

$$\lim_{n \to \infty} b_n(\theta) = 0$$

As far as Example 3 is concerned, the bias is $(1 + \delta) - \delta = 1$, but here there is something we can do about it. Since $E(\overline{X}) = 1 + \delta$, it follows that $E(\overline{X} - 1) = \delta$ and hence that $\overline{X} - 1$ is an unbiased estimator of δ. The following is another example where a minor modification of an estimator leads to an estimator that is unbiased.

EXAMPLE 4

If $X_1, X_2, \ldots, X_n$ constitute a random sample from a uniform population with $\alpha = 0$, show that the largest sample value (that is, the nth order statistic, Y_n) is a biased estimator of the parameter β. Also, modify this estimator of β to make it unbiased.

Solution

Substituting into the formula for $g_n(y_n) = \begin{cases} \frac{n}{\theta} \cdot e^{-y_n/\theta}[1 - e^{-y_n/\theta}]^{n-1} & \text{for } y_n > 0 \\ 0 & \text{elsewhere} \end{cases}$,
we find that the sampling distribution of Y_n is given by

$$\begin{aligned} g_n(y_n) &= n \cdot \frac{1}{\beta} \cdot \left(\int_0^{y_n} \frac{1}{\beta} dx \right)^{n-1} \\ &= \frac{n}{\beta^n} \cdot y_n^{n-1} \end{aligned}$$

for $0 < y_n < \beta$ and $g_n(y_n) = 0$ elsewhere, and hence that

$$\begin{aligned} E(Y_n) &= \frac{n}{\beta^n} \cdot \int_0^{\beta} y_n^n \, dy_n \\ &= \frac{n}{n+1} \cdot \beta \end{aligned}$$

Thus, $E(Y_n) \neq \beta$ and the nth order statistic is a biased estimator of the parameter β. However, since

$$\begin{aligned} E\left(\frac{n+1}{n} \cdot Y_n \right) &= \frac{n+1}{n} \cdot \frac{n}{n+1} \cdot \beta \\ &= \beta \end{aligned}$$

it follows that $\frac{n+1}{n}$ times the largest sample value is an unbiased estimator of the parameter β.

As unbiasedness is a desirable property of an estimator, we can explain why we divided by $n-1$ and not by n when we defined the sample variance: It makes S^2 an unbiased estimator of σ^2 for random samples from infinite populations.

Theorem 1. If S^2 is the variance of a random sample from an infinite population with the finite variance σ^2, then $E(S^2) = \sigma^2$.

Proof By definition of sample mean and sample variance,

$$\begin{aligned} E(S^2) &= E\left[\frac{1}{n-1} \cdot \sum_{i=1}^{n} (X_i - \overline{X})^2 \right] \\ &= \frac{1}{n-1} \cdot E\left[\sum_{i=1}^{n} \{(X_i - \mu) - (\overline{X} - \mu)\}^2 \right] \\ &= \frac{1}{n-1} \cdot \left[\sum_{i=1}^{n} E\{(X_i - \mu)^2\} - n \cdot E\{(\overline{X} - \mu)^2\} \right] \end{aligned}$$

Then, since $E\{(X_i - \mu)^2\} = \sigma^2$ and $E\{(\overline{X} - \mu)^2\} = \frac{\sigma^2}{n}$, it follows that

$$E(S^2) = \frac{1}{n-1} \cdot \left[\sum_{i=1}^{n} \sigma^2 - n \cdot \frac{\sigma^2}{n} \right] = \sigma^2$$

Although S^2 is an unbiased estimator of the variance of an infinite population, it is not an unbiased estimator of the variance of a finite population, and in neither case is S an unbiased estimator of σ. The bias of S as an estimator of σ is discussed, among others, in the book by E. S. Keeping listed among the references at the end of this chapter.

The discussion of the preceding paragraph illustrates one of the difficulties associated with the concept of unbiasedness. It may not be retained under functional transformations; that is, if $\hat{\Theta}$ is an unbiased estimator of θ, it does not necessarily follow that $\omega(\hat{\Theta})$ is an unbiased estimator of $\omega(\theta)$. Another difficulty associated with the concept of unbiasedness is that unbiased estimators are not necessarily unique. For instance, in Example 6 we shall see that $\frac{n+1}{n} \cdot Y_n$ is not the only unbiased estimator of the parameter β of Example 4, and in Exercise 8 we shall see that $\overline{X} - 1$ is not the only unbiased estimator of the parameter δ of Example 3.

3 Efficiency

If we have to choose one of several unbiased estimators of a given parameter, we usually take the one whose sampling distribution has the smallest variance. The estimator with the smaller variance is "more reliable."

DEFINITION 4. MINIMUM VARIANCE UNBIASED ESTIMATOR. *The estimator for the parameter* θ *of a given distribution that has the smallest variance of all unbiased estimators for* θ *is called the* ***minimum variance unbiased estimator****, or the* ***best unbiased estimator*** *for* θ.

If $\hat{\Theta}$ is an unbiased estimator of θ, it can be shown under very general conditions (referred to in the references at the end of the chapter) that the variance of $\hat{\Theta}$ must satisfy the inequality

$$\text{var}(\hat{\Theta}) \geqq \frac{1}{n \cdot E\left[\left(\frac{\partial \ln f(X)}{\partial \theta}\right)^2\right]}$$

where $f(x)$ is the value of the population density at x and n is the size of the random sample. This inequality, the **Cramér–Rao inequality**, leads to the following result.

THEOREM 2. If $\hat{\Theta}$ is an unbiased estimator of θ and

$$\text{var}(\hat{\Theta}) = \frac{1}{n \cdot E\left[\left(\frac{\partial \ln f(X)}{\partial \theta}\right)^2\right]}$$

then $\hat{\Theta}$ is a minimum variance unbiased estimator of θ.

Here, the quantity in the denominator is referred to as the **information** about θ that is supplied by the sample (see also Exercise 19). Thus, the smaller the variance is, the greater the information.

EXAMPLE 5

Show that $\overline{X}$ is a minimum variance unbiased estimator of the mean μ of a normal population.

Solution

Since

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \text{for } -\infty < x < \infty$$

it follows that

$$\ln f(x) = -\ln \sigma\sqrt{2\pi} - \frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2$$

so that

$$\frac{\partial \ln f(x)}{\partial \mu} = \frac{1}{\sigma}\left(\frac{x-\mu}{\sigma}\right)$$

and hence

$$E\left[\left(\frac{\partial \ln f(X)}{\partial \mu}\right)^2\right] = \frac{1}{\sigma^2} \cdot E\left[\left(\frac{X-\mu}{\sigma}\right)^2\right] = \frac{1}{\sigma^2} \cdot 1 = \frac{1}{\sigma^2}$$

Thus,

$$\frac{1}{n \cdot E\left[\left(\frac{\partial \ln f(X)}{\partial \mu}\right)^2\right]} = \frac{1}{n \cdot \frac{1}{\sigma^2}} = \frac{\sigma^2}{n}$$

and since $\overline{X}$ is unbiased and $\text{var}(\overline{X}) = \frac{\sigma^2}{n}$, it follows that $\overline{X}$ is a minimum variance unbiased estimator of μ.

It would be erroneous to conclude from this example that $\overline{X}$ is a minimum variance unbiased estimator of the mean of any population. Indeed, in Exercise 3 the reader will be asked to verify that this is not so for random samples of size $n = 3$ from the continuous uniform population with $\alpha = \theta - \frac{1}{2}$ and $\beta = \theta + \frac{1}{2}$.

As we have indicated, unbiased estimators of one and the same parameter are usually compared in terms of the size of their variances. If $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are two unbiased estimators of the parameter θ of a given population and the variance of $\hat{\Theta}_1$ is less than the variance of $\hat{\Theta}_2$, we say that $\hat{\Theta}_1$ is **relatively more efficient** than $\hat{\Theta}_2$. Also, we use the ratio

$$\frac{\text{var}(\hat{\Theta}_1)}{\text{var}(\hat{\Theta}_2)}$$

as a measure of the efficiency of $\hat{\Theta}_2$ relative to $\hat{\Theta}_1$.

EXAMPLE 6

In Example 4 we showed that if $X_1, X_2, \ldots, X_n$ constitute a random sample from a uniform population with $\alpha = 0$, then $\frac{n+1}{n} \cdot Y_n$ is an unbiased estimator of β.

(a) Show that $2\overline{X}$ is also an unbiased estimator of β.

(b) Compare the efficiency of these two estimators of β.

Solution

(a) Since the mean of the population is $\mu = \frac{\beta}{2}$ according to the theorem "The mean and the variance of the uniform distribution are given by $\mu = \frac{\alpha+\beta}{2}$ and $\sigma^2 = \frac{1}{12}(\beta-\alpha)^2$" it follows from the theorem "If $X_1, X_2, \ldots, X_n$ constitute a random sample from an infinite population with the mean μ and the variance σ^2, then $E(\overline{X}) = \mu$ and $\text{var}(\overline{X}) = \frac{\sigma^2}{n}$" that $E(\overline{X}) = \frac{\beta}{2}$ and hence that $E(2\overline{X}) = \beta$. Thus, $2\overline{X}$ is an unbiased estimator of β.

(b) First we must find the variances of the two estimators. Using the sampling distribution of Y_n and the expression for $E(Y_n)$ given in Example 4, we get

$$E(Y_n^2) = \frac{n}{\beta^n} \cdot \int_0^\beta y_n^{n+1}\, dy_n = \frac{n}{n+2} \cdot \beta^2$$

and

$$\text{var}(Y_n) = \frac{n}{n+2} \cdot \beta^2 - \left(\frac{n}{n+1} \cdot \beta\right)^2$$

If we leave the details to the reader in Exercise 27, it can be shown that

$$\text{var}\left(\frac{n+1}{n} \cdot Y_n\right) = \frac{\beta^2}{n(n+2)}$$

Since the variance of the population is $\sigma^2 = \frac{\beta^2}{12}$ according to the first stated theorem in the example, it follows from the above (second) theorem that $\text{var}(\overline{X}) = \frac{\beta^2}{12n}$ and hence that

$$\text{var}(2\overline{X}) = 4 \cdot \text{var}(\overline{X}) = \frac{\beta^2}{3n}$$

Therefore, the efficiency of $2\overline{X}$ relative to $\frac{n+1}{n} \cdot Y_n$ is given by

$$\frac{\text{var}\left(\frac{n+1}{n} \cdot Y_n\right)}{\text{var}(2\overline{X})} = \frac{\frac{\beta^2}{n(n+2)}}{\frac{\beta^2}{3n}} = \frac{3}{n+2}$$

and it can be seen that for $n > 1$ the estimator based on the nth order statistic is much more efficient than the other one. For $n = 10$, for example, the relative efficiency is only 25 percent, and for $n = 25$ it is only 11 percent.

EXAMPLE 7

When the mean of a normal population is estimated on the basis of a random sample of size $2n+1$, what is the efficiency of the median relative to the mean?

Solution

From the theorem on the previous page we know that $\overline{X}$ is unbiased and that

$$\text{var}(\overline{X}) = \frac{\sigma^2}{2n+1}$$

As far as $\widetilde{X}$ is concerned, it is unbiased by virtue of the symmetry of the normal distribution about its mean, and for large samples

$$\text{var}(\widetilde{X}) = \frac{\pi\sigma^2}{4n}$$

Thus, for large samples, the efficiency of the median relative to the mean is approximately

$$\frac{\text{var}(\overline{X})}{\text{var}(\widetilde{X})} = \frac{\dfrac{\sigma^2}{2n+1}}{\dfrac{\pi\sigma^2}{4n}} = \frac{4n}{\pi(2n+1)}$$

and the **asymptotic efficiency** of the median with respect to the mean is

$$\lim_{n\to\infty} \frac{4n}{\pi(2n+1)} = \frac{2}{\pi}$$

or about 64 percent.

The result of the preceding example may be interpreted as follows: For large samples, the mean requires only 64 percent as many observations as the median to estimate μ with the same reliability.

It is important to note that we have limited our discussion of relative efficiency to unbiased estimators. If we included biased estimators, we could always assure ourselves of an estimator with zero variance by letting its values equal the same constant regardless of the data that we may obtain. Therefore, if $\hat{\Theta}$ is not an unbiased estimator of a given parameter θ, we judge its merits and make efficiency comparisons on the basis of the **mean square error** $E[(\hat{\Theta}-\theta)^2]$ instead of the variance of $\hat{\Theta}$.

Exercises

1. If $X_1, X_2, \ldots, X_n$ constitute a random sample from a population with the mean μ, what condition must be imposed on the constants $a_1, a_2, \ldots, a_n$ so that

$$a_1X_1 + a_2X_2 + \cdots + a_nX_n$$

is an unbiased estimator of μ?

2. If $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are unbiased estimators of the same parameter θ, what condition must be imposed on the constants k_1 and k_2 so that

$$k_1\hat{\Theta}_1 + k_2\hat{\Theta}_2$$

is also an unbiased estimator of θ?

3. This question has been intentionally omitted for this edition.

4. This question has been intentionally omitted for this edition.

5. Given a random sample of size n from a population that has the known mean μ and the finite variance σ^2, show that

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

is an unbiased estimator of σ^2.

6. This question has been intentionally omitted for this edition.

7. Show that $\frac{X+1}{n+2}$ is a biased estimator of the binomial parameter θ. Is this estimator asymptotically unbiased?

8. With reference to Example 3, find an unbiased estimator of δ based on the smallest sample value (that is, on the first order statistic, Y_1).

9. With reference to Example 4, find an unbiased estimator of β based on the smallest sample value (that is, on the first order statistic, Y_1).

10. If $X_1, X_2, \ldots, X_n$ constitute a random sample from a normal population with $\mu = 0$, show that

$$\sum_{i=1}^{n} \frac{X_i^2}{n}$$

is an unbiased estimator of σ^2.

11. If X is a random variable having the binomial distribution with the parameters n and θ, show that $n \cdot \frac{X}{n} \cdot \left(1 - \frac{X}{n}\right)$ is a biased estimator of the variance of X.

12. If a random sample of size n is taken without replacement from the finite population that consists of the positive integers $1, 2, \ldots, k$, show that
(a) the sampling distribution of the nth order statistic, Y_n, is given by

$$f(y_n) = \frac{\binom{y_n - 1}{n-1}}{\binom{k}{n}}$$

for $y_n = n, \ldots, k$;
(b) $\frac{n+1}{n} \cdot Y_n - 1$ is an unbiased estimator of k.
See also Exercise 80.

13. Show that if $\hat{\Theta}$ is an unbiased estimator of θ and $\text{var}(\hat{\Theta}) \neq 0$, then $\hat{\Theta}^2$ is not an unbiased estimator of θ^2.

14. Show that the sample proportion $\frac{X}{n}$ is a minimum variance unbiased estimator of the binomial parameter θ. (*Hint*: Treat $\frac{X}{n}$ as the mean of a random sample of size n from a Bernoulli population with the parameter θ.)

15. Show that the mean of a random sample of size n is a minimum variance unbiased estimator of the parameter λ of a Poisson population.

16. If $\hat{\Theta}_1$ and $\hat{\Theta}_2$ are independent unbiased estimators of a given parameter θ and $\text{var}(\hat{\Theta}_1) = 3 \cdot \text{var}(\hat{\Theta}_2)$, find the constants a_1 and a_2 such that $a_1\hat{\Theta}_1 + a_2\hat{\Theta}_2$ is an unbiased estimator with minimum variance for such a linear combination.

17. Show that the mean of a random sample of size n from an exponential population is a minimum variance unbiased estimator of the parameter θ.

18. Show that for the unbiased estimator of Example 4, $\frac{n+1}{n} \cdot Y_n$, the Cramér–Rao inequality is not satisfied.

19. The information about θ in a random sample of size n is also given by

$$-n \cdot E\left[\frac{\partial^2 \ln f(X)}{\partial \theta^2}\right]$$

where $f(x)$ is the value of the population density at x, provided that the extremes of the region for which $f(x) \neq 0$ do not depend on θ. The derivation of this formula takes the following steps:
(a) Differentiating the expressions on both sides of

$$\int f(x)\, dx = 1$$

with respect to θ, show that

$$\int \frac{\partial \ln f(x)}{\partial \theta} \cdot f(x)\, dx = 0$$

by interchanging the order of integration and differentiation.
(b) Differentiating again with respect to θ, show that

$$E\left[\left(\frac{\partial \ln f(X)}{\partial \theta}\right)^2\right] = -E\left[\frac{\partial^2 \ln f(X)}{\partial \theta^2}\right]$$

20. Rework Example 5 using the alternative formula for the information given in Exercise 19.

21. If $\overline{X}_1$ is the mean of a random sample of size n from a normal population with the mean μ and the variance σ_1^2, $\overline{X}_2$ is the mean of a random sample of size n from a normal population with the mean μ and the variance σ_2^2, and the two samples are independent, show that
(a) $\omega \cdot \overline{X}_1 + (1-\omega) \cdot \overline{X}_2$, where $0 \leqq \omega \leqq 1$, is an unbiased estimator of μ;
(b) the variance of this estimator is a minimum when

$$\omega = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

22. With reference to Exercise 21, find the efficiency of the estimator of part (a) with $\omega = \frac{1}{2}$ relative to this estimator with

$$\omega = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

23. If $\overline{X}_1$ and $\overline{X}_2$ are the means of independent random samples of sizes n_1 and n_2 from a normal population with the mean μ and the variance σ^2, show that the variance of the unbiased estimator

$$\omega \cdot \overline{X}_1 + (1-\omega) \cdot \overline{X}_2$$

is a minimum when $\omega = \dfrac{n_1}{n_1+n_2}$.

24. With reference to Exercise 23, find the efficiency of the estimator with $\omega = \frac{1}{2}$ relative to the estimator with $\omega = \dfrac{n_1}{n_1+n_2}$.

25. If X_1, X_2, and X_3 constitute a random sample of size $n = 3$ from a normal population with the mean μ and the variance σ^2, find the efficiency of $\dfrac{X_1+2X_2+X_3}{4}$ relative to $\dfrac{X_1+X_2+X_3}{3}$ as estimates of μ.

26. If X_1 and X_2 constitute a random sample of size $n = 2$ from an exponential population, find the efficiency of $2Y_1$ relative to $\overline{X}$, where Y_1 is the first order statistic and $2Y_1$ and $\overline{X}$ are both unbiased estimators of the parameter θ.

27. Verify the result given for $\text{var}\left(\dfrac{n+1}{n} \cdot Y_n\right)$ in Example 6.

28. With reference to Example 3, we showed that $\overline{X} - 1$ is an unbiased estimator of δ, and in Exercise 8 the reader was asked to find another unbiased estimator of δ based on the smallest sample value. Find the efficiency of the first of these two estimators relative to the second.

29. With reference to Exercise 12, show that $2\overline{X} - 1$ is also an unbiased estimator of k, and find the efficiency of this estimator relative to the one of part (b) of Exercise 12 for
(a) $n = 2$; **(b)** $n = 3$.

30. Since the variances of the mean and the midrange are not affected if the same constant is added to each observation, we can determine these variances for random samples of size 3 from the uniform population

$$f(x) = \begin{cases} 1 & \text{for } \theta - \frac{1}{2} < x < \theta + \frac{1}{2} \\ 0 & \text{elsewhere} \end{cases}$$

by referring instead to the uniform population

$$f(x) = \begin{cases} 1 & \text{for } 0 < x < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Show that $E(X) = \frac{1}{2}, E(X^2) = \frac{1}{3}$, and $\text{var}(X) = \frac{1}{12}$ for this population so that for a random sample of size $n = 3$, $\text{var}(\overline{X}) = \frac{1}{36}$.

31. Show that if $\hat{\Theta}$ is a biased estimator of θ, then

$$E[(\hat{\Theta} - \theta)^2] = \text{var}(\hat{\Theta}) + [b(\theta)]^2$$

32. If $\hat{\Theta}_1 = \dfrac{X}{n}$, $\hat{\Theta}_2 = \dfrac{X+1}{n+2}$, and $\hat{\Theta}_3 = \dfrac{1}{3}$ are estimators of the parameter θ of a binomial population and $\theta = \frac{1}{2}$, for what values of n is
(a) the mean square error of $\hat{\Theta}_2$ less than the variance of $\hat{\Theta}_1$;
(b) the mean square error of $\hat{\Theta}_3$ less than the variance of $\hat{\Theta}_1$?

4 Consistency

In the preceding section we assumed that the variance of an estimator, or its mean square error, is a good indication of its chance fluctuations. The fact that these measures may not provide good criteria for this purpose is illustrated by the following example: Suppose that we want to estimate on the basis of one observation the parameter θ of the population given by

$$f(x) = \omega \cdot \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{x-\theta}{\sigma}\right)^2} + (1-\omega) \cdot \frac{1}{\pi} \cdot \frac{1}{1+(x-\theta)^2}$$

for $-\infty < x < \infty$ and $0 < \omega < 1$. Evidently, this population is a combination of a normal population with the mean θ and the variance σ^2 and a Cauchy population with $\alpha = \theta$ and $\beta = 1$. Now, if ω is very close to 1, say, $\omega = 1 - 10^{-100}$, and σ is very small, say, $\sigma = 10^{-100}$, the probability that a random variable having this distribution will take on a value that is very close to θ, and hence is a very good estimate of θ, is practically 1. Yet, since the variance of the Cauchy distribution does not exist, neither will the variance of this estimator.

The example of the preceding paragraph is a bit farfetched, but it suggests that we pay more attention to the probabilities with which estimators will take on values that are close to the parameters that they are supposed to estimate. Basing our argument on Chebyshev's theorem, when $n \to \infty$ the probability approaches 1 that the sample proportion $\frac{X}{n}$ will take on a value that differs from the binomial parameter θ by less than any arbitrary constant $c > 0$. Also using Chebyshev's theorem, we see that when $n \to \infty$ the probability approaches 1 that $\overline{X}$ will take on a value that differs from the mean of the population sampled by less than any arbitrary constant $c > 0$.

In both of these examples we are practically assured that, for large n, the estimators will take on values that are very close to the respective parameters. Formally, this concept of "closeness" is expressed by means of the following definition of **consistency**.

DEFINITION 5. CONSISTENT ESTIMATOR. *The statistic* $\hat{\Theta}$ *is a* ***consistent estimator*** *of the parameter* θ *of a given distribution if and only if for each* $c > 0$

$$\lim_{n \to \infty} P(|\hat{\Theta} - \theta| < c) = 1$$

Note that consistency is an **asymptotic property**, that is, a limiting property of an estimator. Informally, Definition 5 says that when n is sufficiently large, we can be practically certain that the error made with a consistent estimator will be less than any small preassigned positive constant. The kind of convergence expressed by the limit in Definition 5 is generally called **convergence in probability**.

Based on Chebyshev's theorem, $\frac{X}{n}$ is a consistent estimator of the binomial parameter θ and $\overline{X}$ is a consistent estimator of the mean of a population with a finite variance. In practice, we can often judge whether an estimator is consistent by using the following sufficient condition, which, in fact, is an immediate consequence of Chebyshev's theorem.

THEOREM 3. If $\hat{\Theta}$ is an unbiased estimator of the parameter θ and $\text{var}(\hat{\Theta}) \to 0$ as $n \to \infty$, then $\hat{\Theta}$ is a consistent estimator of θ.

EXAMPLE 8

Show that for a random sample from a normal population, the sample variance S^2 is a consistent estimator of σ^2.

Solution

Since S^2 is an unbiased estimator of σ^2 in accordance with Theorem 3, it remains to be shown that $\text{var}(S^2) \to 0$ as $n \to \infty$. Referring to the theorem "the random variable $\frac{(n-1)S^2}{\sigma^2}$ has a chi-square distribution with $n-1$ degrees of freedom", we find that for a random sample from a normal population

$$\text{var}(S^2) = \frac{2\sigma^4}{n-1}$$

It follows that $\text{var}(S^2) \to 0$ as $n \to \infty$, and we have thus shown that S^2 is a consistent estimator of the variance of a normal population.

It is of interest to note that Theorem 3 also holds if we substitute "asymptotically unbiased" for "unbiased." This is illustrated by the following example.

EXAMPLE 9

With reference to Example 3, show that the smallest sample value (that is, the first order statistic Y_1) is a consistent estimator of the parameter δ.

Solution

Substituting into the formula for $g_1(y_1)$, we find that the sampling distribution of Y_1 is given by

$$\begin{aligned} g_1(y_1) &= n \cdot e^{-(y_1-\delta)} \cdot \left[\int_{y_1}^{\infty} e^{-(x-\delta)}\, dx \right]^{n-1} \\ &= n \cdot e^{-n(y_1-\delta)} \end{aligned}$$

for $y_1 > \delta$ and $g_1(y_1) = 0$ elsewhere. Based on this result, it can easily be shown that $E(Y_1) = \delta + \frac{1}{n}$ and hence that Y_1 is an asymptotically unbiased estimator of δ. Furthermore,

$$\begin{aligned} P(|Y_1 - \delta| < c) &= P(\delta < Y_1 < \delta + c) \\ &= \int_{\delta}^{\delta+c} n \cdot e^{-n(y_1-\delta)}\, dy_1 \\ &= 1 - e^{-nc} \end{aligned}$$

Since $\lim_{n\to\infty} (1 - e^{-nc}) = 1$, it follows from Definition 5 that Y_1 is a consistent estimator of δ.

Theorem 3 provides a sufficient condition for the consistency of an estimator. It is not a necessary condition because consistent estimators need not be unbiased, or even asymptotically unbiased. This is illustrated by Exercise 41.

5 Sufficiency

An estimator $\hat{\Theta}$ is said to be **sufficient** if it utilizes all the information in a sample relevant to the estimation of θ, that is, if all the knowledge about θ that can be gained from the individual sample values and their order can just as well be gained from the value of $\hat{\Theta}$ alone.

Formally, we can describe this property of an estimator by referring to the conditional probability distribution or density of the sample values given $\hat{\Theta} = \hat{\theta}$, which is given by

$$f(x_1, x_2, \ldots, x_n | \hat{\theta}) = \frac{f(x_1, x_2, \ldots, x_n, \hat{\theta})}{g(\hat{\theta})} = \frac{f(x_1, x_2, \ldots, x_n)}{g(\hat{\theta})}$$

If it depends on θ, then particular values of $X_1, X_2, \ldots, X_n$ yielding $\hat{\Theta} = \hat{\theta}$ will be more probable for some values of θ than for others, and the knowledge of these sample values will help in the estimation of θ. On the other hand, if it does not depend on θ, then particular values of $X_1, X_2, \ldots, X_n$ yielding $\hat{\Theta} = \hat{\theta}$ will be just as likely for any value of θ, and the knowledge of these sample values will be of no help in the estimation of θ.

DEFINITION 6. SUFFICIENT ESTIMATOR. *The statistic $\hat{\Theta}$ is a* ***sufficient estimator*** *of the parameter θ of a given distribution if and only if for each value of $\hat{\Theta}$ the conditional probability distribution or density of the random sample $X_1, X_2, \ldots, X_n$, given $\hat{\Theta} = \theta$, is independent of θ.*

EXAMPLE 10

If $X_1, X_2, \ldots, X_n$ constitute a random sample of size n from a Bernoulli population, show that

$$\hat{\Theta} = \frac{X_1 + X_2 + \cdots + X_n}{n}$$

is a sufficient estimator of the parameter θ.

Solution

By the definition "**BERNOULLI DISTRIBUTION.** *A random variable X has a* ***Bernoulli distribution*** *and it is referred to as a Bernoulli random variable if and only if its probability distribution is given by $f(x;\theta) = \theta^x(1-\theta)^{1-x}$* for $x = 0, 1$",

$$f(x_i;\theta) = \theta^{x_i}(1-\theta)^{1-x_i} \qquad \text{for } x_i = 0, 1$$

so that

$$\begin{aligned} f(x_1, x_2, \ldots, x_n) &= \prod_{i=1}^{n} \theta^{x_i}(1-\theta)^{1-x_i} \\ &= \theta^{\sum_{i=1}^{n} x_i}(1-\theta)^{n-\sum_{i=1}^{n} x_i} \\ &= \theta^x(1-\theta)^{n-x} \\ &= \theta^{n\hat{\theta}}(1-\theta)^{n-n\hat{\theta}} \end{aligned}$$

for $x_i = 0$ or 1 and $i = 1, 2, \ldots, n$. Also, since

$$X = X_1 + X_2 + \cdots + X_n$$

is a binomial random variable with the parameters θ and n, its distribution is given by

$$b(x; n, \theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$$

and the transformation-of-variable technique yields

$$g(\hat{\theta}) = \binom{n}{n\hat{\theta}}\theta^{n\hat{\theta}}(1-\theta)^{n-n\hat{\theta}} \qquad \text{for } \hat{\theta} = 0, \frac{1}{n}, \ldots, 1$$

Now, substituting into the formula for $f(x_1, x_2, \ldots, x_n|\hat{\theta})$ on the previous page, we get

$$\begin{aligned}
\frac{f(x_1, x_2, \ldots, x_n, \hat{\theta})}{g(\hat{\theta})} &= \frac{f(x_1, x_2, \ldots, x_n)}{g(\hat{\theta})} \\
&= \frac{\theta^{n\hat{\theta}}(1-\theta)^{n-n\hat{\theta}}}{\binom{n}{n\hat{\theta}}\theta^{n\hat{\theta}}(1-\theta)^{n-n\hat{\theta}}} \\
&= \frac{1}{\binom{n}{n\hat{\theta}}} \\
&= \frac{1}{\binom{n}{x}} \\
&= \frac{1}{\binom{n}{x_1 + x_2 + \cdots + x_n}}
\end{aligned}$$

for $x_i = 0$ or 1 and $i = 1, 2, \ldots, n$. Evidently, this does not depend on θ and we have shown, therefore, that $\hat{\Theta} = \dfrac{X}{n}$ is a sufficient estimator of θ.

EXAMPLE 11

Show that $Y = \frac{1}{6}(X_1 + 2X_2 + 3X_3)$ is not a sufficient estimator of the Bernoulli parameter θ.

Solution

Since we must show that

$$f(x_1, x_2, x_3|y) = \frac{f(x_1, x_2, x_3, y)}{g(y)}$$

is not independent of θ for some values of X_1, X_2, and X_3, let us consider the case where $x_1 = 1, x_2 = 1$, and $x_3 = 0$. Thus, $y = \frac{1}{6}(1 + 2 \cdot 1 + 3 \cdot 0) = \frac{1}{2}$ and

$$\begin{aligned}
f\left(1, 1, 0 \middle| Y = \frac{1}{2}\right) &= \frac{P\left(X_1 = 1, X_2 = 1, X_3 = 0, Y = \frac{1}{2}\right)}{P\left(Y = \frac{1}{2}\right)} \\
&= \frac{f(1, 1, 0)}{f(1, 1, 0) + f(0, 0, 1)}
\end{aligned}$$

where

$$f(x_1, x_2, x_3) = \theta^{x_1 + x_2 + x_3}(1-\theta)^{3-(x_1 + x_2 + x_3)}$$

for $x_1 = 0$ or 1 and $i = 1,2,3$. Since $f(1,1,0) = \theta^2(1-\theta)$ and $f(0,0,1) = \theta(1-\theta)^2$, it follows that

$$f\left(1,1,0|Y=\frac{1}{2}\right) = \frac{\theta^2(1-\theta)}{\theta^2(1-\theta)+\theta(1-\theta)^2} = \theta$$

and it can be seen that this conditional probability depends on θ. We have thus shown that $Y = \frac{1}{6}(X_1 + 2X_2 + 3X_3)$ is not a sufficient estimator of the parameter θ of a Bernoulli population.

Because it can be very tedious to check whether a statistic is a sufficient estimator of a given parameter based directly on Definition 6, it is usually easier to base it instead on the following **factorization theorem**.

THEOREM 4. The statistic $\hat{\Theta}$ is a sufficient estimator of the parameter θ if and only if the joint probability distribution or density of the random sample can be factored so that

$$f(x_1, x_2, \ldots, x_n; \theta) = g(\hat{\theta}, \theta) \cdot h(x_1, x_2, \ldots, x_n)$$

where $g(\hat{\theta}, \theta)$ depends only on $\hat{\theta}$ and θ, and $h(x_1, x_2, \ldots, x_n)$ does not depend on θ.

A proof of this theorem may be found in more advanced texts; see, for instance, the book by Hogg and Tanis listed among the references at the end of this chapter. Here, let us illustrate the use of Theorem 4 by means of the following example.

EXAMPLE 12

Show that $\overline{X}$ is a sufficient estimator of the mean μ of a normal population with the known variance σ^2.

Solution

Making use of the fact that

$$f(x_1, x_2, \ldots, x_n; \mu) = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \cdot e^{-\frac{1}{2}\cdot\sum_{i=1}^{n}\left(\frac{x_i-\mu}{\sigma}\right)^2}$$

and that

$$\begin{aligned}\sum_{i=1}^{n}(x_i-\mu)^2 &= \sum_{i=1}^{n}[(x_i-\overline{x})-(\mu-\overline{x})]^2 \\ &= \sum_{i=1}^{n}(x_i-\overline{x})^2 + \sum_{i=1}^{n}(\overline{x}-\mu)^2 \\ &= \sum_{i=1}^{n}(x_i-\overline{x})^2 + n(\overline{x}-\mu)^2\end{aligned}$$

we get

$$f(x_1, x_2, \ldots, x_n; \mu) = \left\{ \frac{\sqrt{n}}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{1}{2}\left(\frac{\bar{x}-\mu}{\sigma/\sqrt{n}}\right)^2} \right\}$$

$$\times \left\{ \frac{1}{\sqrt{n}} \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^{n-1} \cdot e^{-\frac{1}{2}\cdot\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{\sigma}\right)^2} \right\}$$

where the first factor on the right-hand side depends only on the estimate $\bar{x}$ and the population mean μ, and the second factor does not involve μ. According to Theorem 4, it follows that $\overline{X}$ is a sufficient estimator of the mean μ of a normal population with the known variance σ^2.

Based on Definition 6 and Theorem 4, respectively, we have presented two ways of checking whether a statistic $\hat{\Theta}$ is a sufficient estimator of a given parameter θ. As we already said, the factorization theorem usually leads to easier solutions; but if we want to show that a statistic $\hat{\Theta}$ is not a sufficient estimator of a given parameter θ, it is nearly always easier to proceed with Definition 6. This was illustrated by Example 11.

Let us also mention the following important property of sufficient estimators. If $\hat{\Theta}$ is a sufficient estimator of θ, then any single-valued function $Y = u(\hat{\Theta})$, not involving θ, is also a sufficient estimator of θ, and therefore of $u(\theta)$, provided $y = u(\hat{\theta})$ can be solved to give the single-valued inverse $\hat{\theta} = w(y)$. This follows directly from Theorem 4, since we can write

$$f(x_1, x_2, \ldots, x_n; \theta) = g[w(y), \theta] \cdot h(x_1, x_2, \ldots, x_n)$$

where $g[w(y), \theta]$ depends only on y and θ. If we apply this result to Example 10, where we showed that $\hat{\Theta} = \frac{X}{n}$ is a sufficient estimator of the Bernoulli parameter θ, it follows that $X = X_1 + X_2 + \cdots + X_n$ is also a sufficient estimator of the mean $\mu = n\theta$ of a binomial population.

6 Robustness

In recent years, special attention has been paid to a statistical property called **robustness**. It is indicative of the extent to which estimation procedures (and, as we shall see later, other methods of inference) are adversely affected by violations of underlying assumptions. In other words, an estimator is said to be **robust** if its sampling distribution is not seriously affected by violations of assumptions. Such violations are often due to outliers caused by outright errors made, say, in reading instruments or recording the data or by mistakes in experimental procedures. They may also pertain to the nature of the populations sampled or their parameters. For instance, when estimating the average useful life of a certain electronic component, we may think that we are sampling an exponential population, whereas actually we are sampling a Weibull population, or when estimating the average income of a certain age group, we may use a method based on the assumption that we are sampling a normal population, whereas actually the population (income distribution) is highly skewed. Also, when estimating the difference between the average weights of two kinds of frogs, the difference between the mean I.Q.'s of two ethnic groups, and in general the difference $\mu_1 - \mu_2$ between the means of two populations, we may be assuming that the two populations have the same variance σ^2, whereas in reality $\sigma_1^2 \neq \sigma_2^2$.

As should be apparent, most questions of robustness are difficult to answer; indeed, much of the language used in the preceding paragraph is relatively imprecise. After all, what do we mean by "not seriously affected"? Furthermore, when we speak of violations of underlying assumptions, it should be clear that some violations are more serious than others. When it comes to questions of robustness, we are thus faced with all sorts of difficulties, mathematically and otherwise, and for the most part they can be resolved only by computer simulations.

Exercises

33. Use Definition 5 to show that Y_1, the first order statistic, is a consistent estimator of the parameter α of a uniform population with $\beta = \alpha + 1$.

34. With reference to Exercise 33, use Theorem 3 to show that $Y_1 - \frac{1}{n+1}$ is a consistent estimator of the parameter α.

35. With reference to the uniform population of Example 4, use the definition of consistency to show that Y_n, the nth order statistic, is a consistent estimator of the parameter β.

36. If $X_1, X_2, \ldots, X_n$ constitute a random sample of size n from an exponential population, show that $\overline{X}$ is a consistent estimator of the parameter θ.

37. With reference to Exercise 36, is X_n a consistent estimator of the parameter θ?

38. Show that the estimator of Exercise 21 is consistent.

39. Substituting "asymptotically unbiased" for "unbiased" in Theorem 3, show that $\frac{X+1}{n+2}$ is a consistent estimator of the binomial parameter θ.

40. Substituting "asymptotically unbiased" for "unbiased" in Theorem 3, use this theorem to rework Exercise 35.

41. To show that an estimator can be consistent without being unbiased or even asymptotically unbiased, consider the following estimation procedure: To estimate the mean of a population with the finite variance σ^2, we first take a random sample of size n. Then we randomly draw one of n slips of paper numbered from 1 through n, and if the number we draw is $2, 3, \ldots,$ or n, we use as our estimator the mean of the random sample; otherwise, we use the estimate n^2. Show that this estimation procedure is

(a) consistent;

(b) neither unbiased nor asymptotically unbiased.

42. If $X_1, X_2, \ldots, X_n$ constitute a random sample of size n from an exponential population, show that $\overline{X}$ is a sufficient estimator of the parameter θ.

43. If X_1 and X_2 are independent random variables having binomial distributions with the parameters θ and n_1 and θ and n_2, show that $\frac{X_1 + X_2}{n_1 + n_2}$ is a sufficient estimator of θ.

44. In reference to Exercise 43, is $\frac{X_1 + 2X_2}{n_1 + 2n_2}$ a sufficient estimator of θ?

45. After referring to Example 4, is the nth order statistic, Y_n, a sufficient estimator of the parameter β?

46. If X_1 and X_2 constitute a random sample of size $n = 2$ from a Poisson population, show that the mean of the sample is a sufficient estimator of the parameter λ.

47. If X_1, X_2, and X_3 constitute a random sample of size $n = 3$ from a Bernoulli population, show that $Y = X_1 + 2X_2 + X_3$ is not a sufficient estimator of θ. (*Hint*: Consider special values of X_1, X_2, and X_3.)

48. If $X_1, X_2, \ldots, X_n$ constitute a random sample of size n from a geometric population, show that $Y = X_1 + X_2 + \cdots + X_n$ is a sufficient estimator of the parameter θ.

49. Show that the estimator of Exercise 5 is a sufficient estimator of the variance of a normal population with the known mean μ.

7 The Method of Moments

As we have seen in this chapter, there can be many different estimators of one and the same parameter of a population. Therefore, it would seem desirable to have some general method, or methods, that yield estimators with as many desirable properties as possible. In this section and in Section 8 we shall present two such methods, the **method of moments**, which is historically one of the oldest methods, and the **method of maximum likelihood**. Furthermore, **Bayesian estimation** will be treated briefly in Section 9.

The method of moments consists of equating the first few moments of a population to the corresponding moments of a sample, thus getting as many equations as are needed to solve for the unknown parameters of the population.

DEFINITION 7. SAMPLE MOMENTS. *The* ***k*th *sample moment*** *of a set of observations* $x_1, x_2, \ldots, x_n$ *is the mean of their* k*th powers and it is denoted by* m'_k*; symbolically,*

$$m'_k = \frac{\sum_{i=1}^{n} x_i^k}{n}$$

Thus, if a population has r parameters, the method of moments consists of solving the system of equations

$$m'_k = \mu'_k \quad k = 1, 2, \ldots, r$$

for the r parameters.

EXAMPLE 13

Given a random sample of size n from a uniform population with $\beta = 1$, use the method of moments to obtain a formula for estimating the parameter α.

Solution

The equation that we shall have to solve is $m'_1 = \mu'_1$, where $m'_1 = \bar{x}$ and $\mu'_1 = \frac{\alpha+\beta}{2} = \frac{\alpha+1}{2}$. Thus,

$$\bar{x} = \frac{\alpha+1}{2}$$

and we can write the estimate of α as

$$\hat{\alpha} = 2\bar{x} - 1$$

EXAMPLE 14

Given a random sample of size n from a gamma population, use the method of moments to obtain formulas for estimating the parameters α and β.

Solution

The system of equations that we shall have to solve is

$$m'_1 = \mu'_1 \quad \text{and} \quad m'_2 = \mu'_2$$

where $\mu'_1 = \alpha\beta$ and $\mu'_2 = \alpha(\alpha+1)\beta^2$. Thus,

$$m'_1 = \alpha\beta \quad \text{and} \quad m'_2 = \alpha(\alpha+1)\beta^2$$