CHAPTER 7

Sufficient Statistics

7.1 Measures of Quality of Estimators

In Chapter 6 we presented some procedures for finding point estimates, interval estimates, and tests of statistical hypotheses. In this and the next two chapters, we provide reasons why certain statistics are used in these various statistical inferences. We begin by considering desirable properties of a point estimate.

Now it would seem that if $y = u(x_1, x_2, \ldots, x_n)$ is to qualify as a good point estimate of θ, there should be a great probability that the statistic $Y = u(X_1, X_2, \ldots, X_n)$ will be close to θ; that is, θ should be a sort of rallying point for the numbers $y = u(x_1, x_2, \ldots, x_n)$. This can be achieved in one way by selecting $Y = u(X_1, X_2, \ldots, X_n)$ in such a way that not only is Y an unbiased estimator of θ, but also the variance of Y is as small as it can be made. We do this because the variance of Y is a measure of the intensity of the concentration of the probability for Y in the neighborhood of the point $\theta = E(Y)$. Accordingly, we define an unbiased minimum variance estimator of the parameter θ in the following manner.

Definition 1. For a given positive integer n, $Y = u(X_1, X_2, \ldots, X_n)$ will be called an *unbiased minimum variance* estimator of the par-

ameter θ if Y is unbiased, that is, $E(Y) = \theta$, and if the variance of Y is less than or equal to the variance of every other unbiased estimator of θ.

For illustration, let $X_1, X_2, \ldots, X_9$ denote a random sample from a distribution that is $N(\theta, 1)$, $-\infty < \theta < \infty$. Since the statistic $\bar{X} = (X_1 + X_2 + \cdots + X_9)/9$ is $N(\theta, \frac{1}{9})$, $\bar{X}$ is an unbiased estimator of θ. The statistic X_1 is $N(\theta, 1)$, so X_1 is also an unbiased estimator of θ. Although the variance $\frac{1}{9}$ of $\bar{X}$ is less than the variance 1 of X_1, we cannot say, with $n = 9$, that $\bar{X}$ is the unbiased minimum variance estimator of θ; that definition requires that the comparison be made with every unbiased estimator of θ. To be sure, it is quite impossible to tabulate all other unbiased estimators of this parameter θ, so other methods must be developed for making the comparisons of the variances. A beginning on this problem will be made in this chapter.

Let us now discuss the problem of point estimation of a parameter from a slightly different standpoint. Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from a distribution that has the p.d.f. $f(x; \theta)$, $\theta \in \Omega$. The distribution may be either of the continuous or the discrete type. Let $Y = u(X_1, X_2, \ldots, X_n)$ be a statistic on which we wish to base a point estimate of the parameter θ. Let $\delta(y)$ be that function of the observed value of the statistic Y which is the point estimate of θ. Thus the function δ *decides* the value of our point estimate of θ and δ is called a *decision function* or a *decision rule*. One value of the decision function, say $\delta(y)$, is called a *decision*. Thus a numerically determined point estimate of a parameter θ is a decision. Now a decision may be correct or it may be wrong. It would be useful to have a measure of the seriousness of the difference, if any, between the true value of θ and the point estimate $\delta(y)$. Accordingly, with each pair, $[\theta, \delta(y)]$, $\theta \in \Omega$, we will associate a nonnegative number $\mathscr{L}[\theta, \delta(y)]$ that reflects this seriousness. We call the function $\mathscr{L}$ the *loss function*. The expected (mean) value of the loss function is called the *risk function*. If $g(y; \theta)$, $\theta \in \Omega$, is the p.d.f. of Y, the risk function $R(\theta, \delta)$ is given by

$$R(\theta, \delta) = E\{\mathscr{L}[\theta, \delta(Y)]\} = \int_{-\infty}^{\infty} \mathscr{L}[\theta, \delta(y)]g(y; \theta)\, dy$$

if Y is a random variable of the continuous type. It would be desirable to select a decision function that minimizes the risk $R(\theta, \delta)$ for all values of θ, $\theta \in \Omega$. But this is usually impossible because the decision function δ that minimizes $R(\theta, \delta)$ for one value of θ may not minimize

$R(\theta, \delta)$ for another value of θ. Accordingly, we need either to restrict our decision function to a certain class or to consider methods of ordering the risk functions. The following example, while very simple, dramatizes these difficulties.

Example 1. Let $X_1, X_2, \ldots, X_{25}$ be a random sample from a distribution that is $N(\theta, 1)$, $-\infty < \theta < \infty$. Let $Y = \bar{X}$, the mean of the random sample, and let $\mathscr{L}[\theta, \delta(y)] = [\theta - \delta(y)]^2$. We shall compare the two decision functions given by $\delta_1(y) = y$ and $\delta_2(y) = 0$ for $-\infty < y < \infty$. The corresponding risk functions are

$$R(\theta, \delta_1) = E[(\theta - Y)^2] = \tfrac{1}{25}$$

and

$$R(\theta, \delta_2) = E[(\theta - 0)^2] = \theta^2.$$

Obviously, if, in fact, $\theta = 0$, then $\delta_2(y) = 0$ is an excellent decision and we have $R(0, \delta_2) = 0$. However, if θ differs from zero by very much, it is equally clear that $\delta_2(y) = 0$ is a poor decision. For example, if, in fact, $\theta = 2$, $R(2, \delta_2) = 4 > R(2, \delta_1) = \frac{1}{25}$. In general, we see that $R(\theta, \delta_2) < R(\theta, \delta_1)$, provided that $-\frac{1}{5} < \theta < \frac{1}{5}$ and that otherwise $R(\theta, \delta_2) \geq R(\theta, \delta_1)$. That is, one of these decision functions is better than the other for some values of θ and the other decision functions are better for other values of θ. If, however, we had restricted our consideration to decision functions δ such that $E[\delta(Y)] = \theta$ for all values of θ, $\theta \in \Omega$, then the decision $\delta_2(y) = 0$ is not allowed. Under this restriction and with the given $\mathscr{L}[\theta, \delta(y)]$, the risk function is the variance of the unbiased estimator $\delta(Y)$, and we are confronted with the problem of finding the unbiased minimum variance estimator. Later in this chapter we show that the solution is $\delta(y) = y = \bar{x}$.

Suppose, however, that we do not want to restrict ourselves to decision functions δ, such that $E[\delta(Y)] = \theta$ for all values of θ, $\theta \in \Omega$. Instead, let us say that the decision function that minimizes the maximum of the risk function is the best decision function. Because, in this example, $R(\theta, \delta_2) = \theta^2$ is unbounded, $\delta_2(y) = 0$ is not, in accordance with this criterion, a good decision function. On the other hand, with $-\infty < \theta < \infty$, we have

$$\max_{\theta} R(\theta, \delta_1) = \max_{\theta} \left(\tfrac{1}{25}\right) = \tfrac{1}{25}.$$

Accordingly, $\delta_1(y) = y = \bar{x}$ seems to be a very good decision in accordance with this criterion because $\frac{1}{25}$ is small. As a matter of fact, it can be proved that δ_1 is the best decision function, as measured by the *minimax criterion*, when the loss function is $\mathscr{L}[\theta, \delta(y)] = [\theta - \delta(y)]^2$.

In this example we illustrated the following:

1. Without some restriction on the decision function, it is difficult to

find a decision function that has a risk function which is uniformly less than the risk function of another decision function.

2. A principle of selecting a best decision function, called the *minimax principle*. This principle may be stated as follows: If the decision function given by $\delta_0(y)$ is such that, for all $\theta \in \Omega$,

$$\max_{\theta} R[\theta, \delta_0(y)] \leq \max_{\theta} R[\theta, \delta(y)]$$

for every other decision function $\delta(y)$, then $\delta_0(y)$ is called a *minimax decision function*.

With the restriction $E[\delta(Y)] = \theta$ and the loss function $\mathscr{L}[\theta, \delta(y)] = [\theta - \delta(y)]^2$, the decision function that minimizes the risk function yields an unbiased estimator with minimum variance. If, however, the restriction $E[\delta(Y)] = \theta$ is replaced by some other condition, the decision function $\delta(Y)$, if it exists, which minimizes $E\{[\theta - \delta(Y)]^2\}$ uniformly in θ is sometimes called the *minimum mean-square-error estimator*. Exercises 7.6, 7.7, and 7.8 provide examples of this type of estimator.

There are two additional observations about decision rules and loss functions that should be made at this point. First, since Y is a statistic, the decision rule $\delta(Y)$ is also a statistic, and we could have started directly with a decision rule based on the observations in a random sample, say $\delta_1(X_1, X_2, \ldots, X_n)$. The risk function is then given by

$$\begin{aligned} R(\theta, \delta_1) &= E\{\mathscr{L}[\theta, \delta_1(X_1, X_2, \ldots, X_n)]\} \\ &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \mathscr{L}[\theta, \delta_1(x_1, x_2, \ldots, x_n)] \\ &\qquad \times f(x_1; \theta) \cdots f(x_n; \theta)\, dx_1 \cdots dx_n \end{aligned}$$

if the random sample arises from a continuous-type distribution. We did not do this because, as you will see in this chapter, it is rather easy to find a good statistic, say Y, upon which to base all of the statistical inferences associated with a particular model. Thus we thought it more appropriate to start with a statistic that would be familiar, like the m.l.e. $Y = \bar{X}$ in Example 1. The second decision rule of that example could be written $\delta_2(X_1, X_2, \ldots, X_n) = 0$, a constant no matter what values of $X_1, X_2, \ldots, X_n$ are observed.

The second observation is that we have only used one loss function, namely the *square-error loss function* $\mathscr{L}(\theta, \delta) = (\theta - \delta)^2$.

The *absolute-error loss function* $\mathscr{L}(\theta, \delta) = |\theta - \delta|$ is another popular one. The loss function defined by

$$\mathscr{L}(\theta, \delta) = 0, \qquad |\theta - \delta| \leq a,$$
$$= b, \qquad |\theta - \delta| > a,$$

where a and b are positive constants, is sometimes referred to as the *goal post loss function*. The reason for this terminology is that football fans recognize it is like kicking a field goal: There is no loss (actually a three-point gain) if within a units of the middle but b units of loss (zero points awarded) if outside that restriction. In addition, loss functions can be asymmetric as well as symmetric as the three previous ones have been. That is, for example, it might be more costly to underestimate the value of θ than to overestimate it. (Many of us think about this type of loss function when estimating the time it takes us to reach an airport to catch a plane.) Some of these loss functions are considered when studying Bayesian estimates in Chapter 8.

Let us close this section with an interesting illustration that raises a question leading to the likelihood principle which many statisticians believe is a quality characteristic that estimators should enjoy. Suppose that two statisticians, A and B, observe 10 independent trials of a random experiment ending in success or failure. Let the probability of success on each trial be θ, where $0 < \theta < 1$. Let us say that each statistician observes one success in these 10 trials. Suppose, however, that A had decided to take $n = 10$ such observations in advance and found only one success while B had decided to take as many observations as needed to get the first success, which happened on the 10th trial. The model of A is that Y is $b(n = 10, \theta)$ and $y = 1$ is observed. On the other hand, B is considering the random variable Z that has a geometric p.d.f. $g(z) = (1 - \theta)^{z-1}\theta$, $z = 1, 2, 3, \ldots$, and $z = 10$ is observed. In either case, the relative frequency of success is

$$\frac{y}{n} = \frac{1}{z} = \frac{1}{10},$$

which could be used as an estimate of θ.

Let us observe, however, that one of the corresponding estimators, Y/n and $1/Z$, is biased. We have

$$E\left(\frac{Y}{10}\right) = \frac{1}{10}E(Y) = \frac{1}{10}(10\theta) = \theta$$

while

$$E\left(\frac{1}{Z}\right) = \sum_{z=1}^{\infty} \frac{1}{z}(1-\theta)^{z-1}\theta$$

$$= \theta + \tfrac{1}{2}(1-\theta)\theta + \tfrac{1}{3}(1-\theta)^2\theta + \cdots > \theta.$$

That is, $1/Z$ is a biased estimator while $Y/10$ is unbiased. Thus A is using an unbiased estimator while B is not. Should we adjust B's estimator so that it too is unbiased?

It is interesting to note that if we maximize the two respective likelihood functions, namely

$$L_1(\theta) = \binom{10}{y}\theta^y(1-\theta)^{10-y}$$

and

$$L_2(\theta) = (1-\theta)^{z-1}\theta,$$

with $n = 10$, $y = 1$, and $z = 10$, we get exactly the same answer, $\hat{\theta} = \frac{1}{10}$. This must be the case, because in each situation we are maximizing $(1-\theta)^9\theta$. Many statisticians believe that this is the way it should be and accordingly adopt the *likelihood principle:*

Suppose two different sets of data from possibly two different random experiments lead to respective likelihood ratios, $L_1(\theta)$ and $L_2(\theta)$, that are proportional to each other. These two data sets provide the same information about the parameter θ and a statistician should obtain the same estimate of θ from either.

In our special illustration, we note that $L_1(\theta) \propto L_2(\theta)$, and the likelihood principle states that statisticians A and B should make the same inference. Thus believers in the likelihood principle would not adjust the second estimator to make it unbiased.

EXERCISES

7.1. Show that the mean $\bar{X}$ of a random sample of size n from a distribution having p.d.f. $f(x;\theta) = (1/\theta)e^{-(x/\theta)}$, $0 < x < \infty$, $0 < \theta < \infty$, zero elsewhere, is an unbiased estimator of θ and has variance θ^2/n.

7.2. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a normal distribution with mean zero and variance θ, $0 < \theta < \infty$. Show that $\sum_1^n X_i^2/n$ is an unbiased estimator of θ and has variance $2\theta^2/n$.

7.3. Let $Y_1 < Y_2 < Y_3$ be the order statistics of a random sample of size 3 from the uniform distribution having p.d.f. $f(x; \theta) = 1/\theta, 0 < x < \theta, 0 < \theta < \infty$, zero elsewhere. Show that $4Y_1$, $2Y_2$, and $\frac{4}{3}Y_3$ are all unbiased estimators of θ. Find the variance of each of these unbiased estimators.

7.4. Let Y_1 and Y_2 be two independent unbiased estimators of θ. Say the variance of Y_1 is twice the variance of Y_2. Find the constants k_1 and k_2 so that $k_1Y_1 + k_2Y_2$ is an unbiased estimator with smallest possible variance for such a linear combination.

7.5. In Example 1 of this section, take $\mathscr{L}[\theta, \delta(y)] = |\theta - \delta(y)|$. Show that $R(\theta, \delta_1) = \frac{1}{5}\sqrt{2/\pi}$ and $R(\theta, \delta_2) = |\theta|$. Of these two decision functions δ_1 and δ_2, which yields the smaller maximum risk?

7.6. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a Poisson distribution with parameter θ, $0 < \theta < \infty$. Let $Y = \sum_1^n X_i$ and let $\mathscr{L}[\theta, \delta(y)] = [\theta - \delta(y)]^2$. If we restrict our considerations to decision functions of the form $\delta(y) = b + y/n$, where b does not depend upon y, show that $R(\theta, \delta) = b^2 + \theta/n$. What decision function of this form yields a uniformly smaller risk than every other decision function of this form? With this solution, say δ, and $0 < \theta < \infty$, determine $\max_\theta R(\theta, \delta)$ if it exists.

7.7. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that is $N(\mu, \theta), 0 < \theta < \infty$, where μ is unknown. Let $Y = \sum_1^n (X_i - \bar{X})^2/n = S^2$ and let $\mathscr{L}[\theta, \delta(y)] = [\theta - \delta(y)]^2$. If we consider decision functions of the form $\delta(y) = by$, where b does not depend upon y, show that $R(\theta, \delta) = (\theta^2/n^2)[(n^2 - 1)b^2 - 2n(n - 1)b + n^2]$. Show that $b = n/(n + 1)$ yields a minimum risk for decision functions of this form. Note that $nY/(n + 1)$ is not an unbiased estimator of θ. With $\delta(y) = ny/(n + 1)$ and $0 < \theta < \infty$, determine $\max_\theta R(\theta, \delta)$ if it exists.

7.8. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that is $b(1, \theta), 0 \leq \theta \leq 1$. Let $Y = \sum_1^n X_i$ and let $\mathscr{L}[\theta, \delta(y)] = [\theta - \delta(y)]^2$. Consider decision functions of the form $\delta(y) = by$, where b does not depend upon y. Prove that $R(\theta, \delta) = b^2n\theta(1 - \theta) + (bn - 1)^2\theta^2$. Show that

$$\max_\theta R(\theta, \delta) = \frac{b^4n^2}{4[b^2n - (bn - 1)^2]},$$

provided that the value b is such that $b^2n \geq 2(bn - 1)^2$. Prove that $b = 1/n$ does not minimize $\max_\theta R(\theta, \delta)$.

7.9. Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with mean $\theta > 0$.

(a) Statistician A observes the sample to be the values $x_1, x_2, \ldots, x_n$ with sum $y_1 = \Sigma x_i$. Find the m.l.e. of θ.

(b) Statistician B loses the sample values $x_1, x_2, \ldots, x_n$ but remembers the sum y_1 and the fact that the sample arose from a Poisson distribution. Thus B decides to create some fake observations which he calls $z_1, z_2, \ldots, z_n$ (as he knows they will probably not equal the original x-values) as follows. He notes that the conditional probability of independent Poisson random variables $Z_1, Z_2, \ldots, Z_n$ being equal to $z_1, z_2, \ldots, z_n$, given $\Sigma z_i = y_1$, is

$$\frac{\frac{\theta^{z_1}e^{-\theta}}{z_1!}\frac{\theta^{z_2}e^{-\theta}}{z_2!}\cdots\frac{\theta^{z_n}e^{-\theta}}{z_n!}}{\frac{(n\theta)^{y_1}e^{-n\theta}}{y_1!}} = \frac{y_1!}{z_1!\,z_2!\cdots z_n!}\left(\frac{1}{n}\right)^{z_1}\left(\frac{1}{n}\right)^{z_2}\cdots\left(\frac{1}{n}\right)^{z_n}$$

since $Y_1 = \Sigma Z_i$ has a Poisson distribution with mean $n\theta$. The latter distribution is multinomial with y_1 independent trials, each terminating in one of n mutually exclusive and exhaustive ways, each of which has the same probability $1/n$. Accordingly, B runs such a multinomial experiment y_1 independent trials and obtains $z_1, z_2, \ldots, z_n$. Find the likelihood function using these z-values. Is it proportional to that of statistician A?

Hint: Here the likelihood function is the product of this conditional p.d.f. and the p.d.f. of $Y_1 = \Sigma Z_i$.

7.2 A Sufficient Statistic for a Parameter

Suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a distribution that has p.d.f. $f(x; \theta)$, $\theta \in \Omega$. In Chapter 6 and Section 7.1 we constructed statistics to make statistical inferences as illustrated by point and interval estimation and tests of statistical hypotheses. We note that a statistic, say $Y = u(X_1, X_2, \ldots, X_n)$, is a form of data reduction. For illustration, instead of listing all of the individual observations $X_1, X_2, \ldots, X_n$, we might prefer to give only the sample mean $\bar{X}$ or the sample variance S^2. Thus statisticians look for ways of reducing a set of data so that these data can be more easily understood without losing the meaning associated with the entire set of observations.

It is interesting to note that a statistic $Y = u(X_1, X_2, \ldots, X_n)$ really partitions the sample space of $X_1, X_2, \ldots, X_n$. For illustration, suppose we say that the sample was observed and $\bar{x} = 8.32$. There are many points in the sample space which have that same mean of 8.32,

and we can consider them as belonging to the set $\{(x_1, x_2, \ldots, x_n): \bar{x} = 8.32\}$. As a matter of fact, all points on the hyperplane

$$x_1 + x_2 + \cdots + x_n = (8.32)n$$

yield the mean of $\bar{x} = 8.32$, so this hyperplane is that set. However, there are many values that $\bar{X}$ can take and thus there are many such sets. So, in this sense, the sample mean $\bar{X}$—or any statistic $Y = u(X_1, X_2, \ldots, X_n)$—partitions the sample space into a collection of sets.

Often in the study of statistics the parameter θ of the model is unknown; thus we desire to make some statistical inference about it. In this section we consider a statistic denoted by $Y_1 = u_1(X_1, X_2, \ldots, X_n)$, which we call a *sufficient statistic* and which we find is good for making those inferences. This sufficient statistic partitions the sample space in such a way that, given

$$(X_1, X_2, \ldots, X_n) \in \{(x_1, x_2, \ldots, x_n): u_1(x_1, x_2, \ldots, x_n) = y_1\},$$

the conditional probability of $X_1, X_2, \ldots, X_n$ does not depend upon θ. Intuitively, this means that once the set determined by $Y_1 = y_1$ is fixed, the distribution of another statistic, say $Y_2 = u_2(X_1, X_2, \ldots, X_n)$, does not depend upon the parameter θ because the conditional distribution of $X_1, X_2, \ldots, X_n$ does not depend upon θ. Hence it is impossible to use Y_2, given $Y_1 = y_1$, to make a statistical inference about θ. So, in a sense, Y_1 *exhausts* all the information about θ that is contained in the sample. This is why we call $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ a sufficient statistic.

To understand clearly the definition of a sufficient statistic for a parameter θ, we start with an illustration.

Example 1. Let $X_1, X_2, \ldots, X_n$ denote a random sample from the distribution that has p.d.f.

$$f(x; \theta) = \theta^x(1 - \theta)^{1-x}, \qquad x = 0, 1; \quad 0 < \theta < 1;$$
$$= 0 \qquad \text{elsewhere.}$$

The statistic $Y_1 = X_1 + X_2 + \cdots + X_n$ has the p.d.f.

$$g_1(y_1; \theta) = \binom{n}{y_1}\theta^{y_1}(1 - \theta)^{n-y_1}, \qquad y_1 = 0, 1, \ldots, n,$$
$$= 0 \qquad \text{elsewhere.}$$

What is the conditional probability

$$\Pr(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n | Y_1 = y_1) = P(A|B),$$

say, where $y_1 = 0, 1, 2, \ldots, n$? Unless the sum of the integers $x_1, x_2, \ldots, x_n$ (each of which equals zero or 1) is equal to y_1, the conditional probability obviously equals zero because $A \cap B = \varnothing$. But in the case $y_1 = \Sigma x_i$, we have that $A \subset B$ so that $A \cap B = A$ and $P(A|B) = P(A)/P(B)$; thus the conditional probability equals

$$\frac{\theta^{x_1}(1-\theta)^{1-x_1}\theta^{x_2}(1-\theta)^{1-x_2}\cdots\theta^{x_n}(1-\theta)^{1-x_n}}{\binom{n}{y_1}\theta^{y_1}(1-\theta)^{n-y_1}} = \frac{\theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}}{\binom{n}{\sum x_i}\theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}}$$

$$= \frac{1}{\binom{n}{\sum x_i}}.$$

Since $y_1 = x_1 + x_2 + \cdots + x_n$ equals the number of 1's in the n independent trials, this conditional probability is the probability of selecting a particular arrangement of y_1 1's and $(n - y_1)$ zeros. Note that this conditional probability does *not* depend upon the value of the parameter θ.

In general, let $g_1(y_1; \theta)$ be the p.d.f. of the statistic $Y_1 = u_1(X_1, X_2, \ldots, X_n)$, where $X_1, X_2, \ldots, X_n$ is a random sample arising from a distribution of the discrete type having p.d.f. $f(x; \theta)$, $\theta \in \Omega$. The conditional probability of $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$, given $Y_1 = y_1$, equals

$$\frac{f(x_1; \theta)f(x_2; \theta)\cdots f(x_n; \theta)}{g_1[u_1(x_1, x_2, \ldots, x_n); \theta]},$$

provided that $x_1, x_2, \ldots, x_n$ are such that the fixed $y_1 = u_1(x_1, x_2, \ldots, x_n)$, and equals zero otherwise. We say that $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ is a *sufficient statistic* for θ if and only if this ratio does not depend upon θ. While, with distributions of the continuous type, we cannot use the same argument, we do, in this case, accept the fact that if this ratio does not depend upon θ, then the conditional distribution of $X_1, X_2, \ldots, X_n$, given $Y_1 = y_1$, does not depend upon θ. Thus, in both cases, we use the same definition of a sufficient statistic for θ.

Definition 2. Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from a distribution that has p.d.f. $f(x; \theta)$, $\theta \in \Omega$. Let

$Y_1 = u_1(X_1, X_2, \ldots, X_n)$ be a statistic whose p.d.f. is $g_1(y_1; \theta)$. Then Y_1 is a *sufficient statistic* for θ if and only if

$$\frac{f(x_1; \theta)f(x_2; \theta) \cdots f(x_n; \theta)}{g_1[u_1(x_1, x_2, \ldots, x_n); \theta]} = H(x_1, x_2, \ldots, x_n),$$

where $H(x_1, x_2, \ldots, x_n)$ does not depend upon $\theta \in \Omega$.

Remark. In most cases in this book, $X_1, X_2, \ldots, X_n$ do represent the observations of a random sample; that is, they are i.i.d. It is not necessary, however, in more general situations, that these random variables be independent; as a matter of fact, they do not need to be identically distributed. Thus, more generally, the definition of sufficiency of a statistic $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ would be extended to read that

$$\frac{f(x_1, x_2, \ldots, x_n; \theta)}{g_1[u_1(x_1, x_2, \ldots, x_n); \theta]} = H(x_1, x_2, \ldots, x_n)$$

does not depend upon $\theta \in \Omega$, where $f(x_1, x_2, \ldots, x_n; \theta)$ is the joint p.d.f. of $X_1, X_2, \ldots, X_n$. There are even a few situations in which we need an extension like this one in this book.

We now give two examples that are illustrative of the definition.

Example 2. Let $X_1, X_2, \ldots, X_n$ be a random sample from a gamma distribution with $\alpha = 2$ and $\beta = \theta > 0$. Since the m.g.f. associated with this distribution is $M(t) = (1 - \theta t)^{-2}$, $t < 1/\theta$, the m.g.f. of $Y_1 = \sum_{i=1}^{n} X_i$ is

$$\begin{aligned} E[e^{t(X_1 + X_2 + \cdots + X_n)}] &= E(e^{tX_1})E(e^{tX_2}) \cdots E(e^{tX_n}) \\ &= [(1 - \theta t)^{-2}]^n = (1 - \theta t)^{-2n}. \end{aligned}$$

Thus Y_1 has a gamma distribution with $\alpha = 2n$ and $\beta = \theta$, so that its p.d.f. is

$$\begin{aligned} g_1(y_1; \theta) &= \frac{1}{\Gamma(2n)\theta^{2n}} y_1^{2n-1} e^{-y_1/\theta}, \qquad 0 < y_1 < \infty, \\ &= 0 \qquad \text{elsewhere.} \end{aligned}$$

Thus we have that the ratio in Definition 2 equals

$$\frac{\left[\frac{x_1^{2-1}e^{-x_1/\theta}}{\Gamma(2)\theta^2}\right]\left[\frac{x_2^{2-1}e^{-x_2/\theta}}{\Gamma(2)\theta^2}\right] \cdots \left[\frac{x_n^{2-1}e^{-x_n/\theta}}{\Gamma(2)\theta^2}\right]}{\frac{(x_1 + x_2 + \cdots + x_n)^{2n-1}e^{-(x_1 + x_2 + \cdots + x_n)/\theta}}{\Gamma(2n)\theta^{2n}}} = \frac{\Gamma(2n)}{[\Gamma(2)]^n} \frac{x_1 x_2 \cdots x_n}{(x_1 + x_2 + \cdots x_n)^{2n-1}},$$

where $0 < x_i < \infty$, $i = 1, 2, \ldots, n$. Since this ratio does not depend upon θ, the sum Y_1 is a sufficient statistic for θ.

Example 3. Let $Y_1 < Y_2 < \cdots < Y_n$ denote the order statistics of a random sample of size n from the distribution with p.d.f.

$$f(x; \theta) = e^{-(x-\theta)} I_{(\theta, \infty)}(x).$$

Here we use the indicator function of set A defined by

$$I_A(x) = 1, \qquad x \in A,$$
$$= 0, \qquad x \notin A.$$

This means, of course, that $f(x; \theta) = e^{-(x-\theta)}$, $\theta < x < \infty$, and zero elsewhere. The p.d.f. of $Y_1 = \min(X_i)$ is

$$g_1(y_1; \theta) = ne^{-n(y_1-\theta)} I_{(\theta,\infty)}(y_1).$$

Thus we have that

$$\frac{\prod_{i=1}^{n} e^{-(x_i-\theta)} I_{(\theta,\infty)}(x_i)}{ne^{-n(\min x_i-\theta)} I_{(\theta,\infty)}(\min x_i)} = \frac{e^{-x_1-x_2-\cdots-x_n}}{ne^{-n \min x_i}}$$

since $\prod_{i=1}^{n} I_{(\theta,\infty)}(x_i) = I_{(\theta,\infty)}(\min x_i)$, because when $\theta < \min x_i$, then $\theta < x_i$, $i = 1, 2, \ldots, n$, and at least one x-value is less than or equal to θ when $\min x_i \leq \theta$. Since this ratio does not depend upon θ, the first order statistic Y_1 is a sufficient statistic for θ.

If we are to show, by means of the definition, that a certain statistic Y_1 is or is not a sufficient statistic for a parameter θ, we must first of all know the p.d.f. of Y_1, say $g_1(y_1; \theta)$. In some instances it may be quite tedious to find this p.d.f. Fortunately, this problem can be avoided if we will but prove the following *factorization theorem* of Neyman.

Theorem 1. *Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that has* p.d.f. *$f(x; \theta)$, $\theta \in \Omega$. The statistic $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ is a sufficient statistic for θ if and only if we can find two nonnegative functions, k_1 and k_2, such that*

$$f(x_1; \theta) f(x_2; \theta) \cdots f(x_n; \theta)$$
$$= k_1[u_1(x_1, x_2, \ldots, x_n); \theta] k_2(x_1, x_2, \ldots, x_n),$$

where $k_2(x_1, x_2, \ldots, x_n)$ does not depend upon θ.

Proof. We shall prove the theorem when the random variables

are of the continuous type. Assume that the factorization is as stated in the theorem. In our proof we shall make the one-to-one transformation $y_1 = u_1(x_1, \ldots, x_n)$, $y_2 = u_2(x_1, \ldots, x_n), \ldots, y_n = u_n(x_1, \ldots, x_n)$ having the inverse functions $x_1 = w_1(y_1, \ldots, y_n)$, $x_2 = w_2(y_1, \ldots, y_n), \ldots, x_n = w_n(y_1, \ldots, y_n)$ and Jacobian J. The joint p.d.f. of the statistics $Y_1, Y_2, \ldots, Y_n$ is then given by

$$g(y_1, y_2, \ldots, y_n; \theta) = k_1(y_1; \theta)k_2(w_1, w_2, \ldots, w_n)|J|,$$

where $w_i = w_i(y_1, y_2, \ldots, y_n)$, $i = 1, 2, \ldots, n$. The p.d.f. of Y_1, say $g_1(y_1; \theta)$, is given by

$$g_1(y_1; \theta) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(y_1, y_2, \ldots, y_n; \theta)\, dy_2 \cdots dy_n$$

$$= k_1(y_1; \theta) \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |J| k_2(w_1, w_2, \ldots, w_n)\, dy_2 \cdots dy_n.$$

Now the function k_2 does not depend upon θ. Nor is θ involved in either the Jacobian J or the limits of integration. Hence the $(n-1)$-fold integral in the right-hand member of the preceding equation is a function of y_1 alone, say $m(y_1)$. Thus

$$g_1(y_1; \theta) = k_1(y_1; \theta)m(y_1).$$

If $m(y_1) = 0$, then $g_1(y_1; \theta) = 0$. If $m(y_1) > 0$, we can write

$$k_1[u_1(x_1, \ldots, x_n); \theta] = \frac{g_1[u_1(x_1, \ldots, x_n); \theta]}{m[u_1(x_1, \ldots, x_n)]},$$

and the assumed factorization becomes

$$f(x_1; \theta) \cdots f(x_n; \theta) = g_1[u_1(x_1, \ldots, x_n); \theta] \frac{k_2(x_1, \ldots, x_n)}{m[u_1(x_1, \ldots, x_n)]}.$$

Since neither the function k_2 nor the function m depends upon θ, then in accordance with the definition, Y_1 is a sufficient statistic for the parameter θ.

Conversely, if Y_1 is a sufficient statistic for θ, the factorization can be realized by taking the function k_1 to be the p.d.f. of Y_1, namely the function g_1. This completes the proof of the theorem.

Example 4. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distri-

bution that is $N(\theta, \sigma^2)$, $-\infty < \theta < \infty$, where the variance $\sigma^2 > 0$ is known. If $\bar{x} = \sum_1^n x_i/n$, then

$$\sum_{i=1}^{n} (x_i - \theta)^2 = \sum_{i=1}^{n} [(x_i - \bar{x}) + (\bar{x} - \theta)]^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2 + n(\bar{x} - \theta)^2$$

because

$$2\sum_{i=1}^{n} (x_i - \bar{x})(\bar{x} - \theta) = 2(\bar{x} - \theta)\sum_{i=1}^{n} (x_i - \bar{x}) = 0.$$

Thus the joint p.d.f. of $X_1, X_2, \ldots, X_n$ may be written

$$\left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n \exp\left[-\sum_{i=1}^{n} (x_i - \theta)^2/2\sigma^2\right]$$

$$= \{\exp[-n(\bar{x} - \theta)^2/2\sigma^2]\}\left\{\frac{\exp\left[-\sum_{i=1}^{n} (x_i - \bar{x})^2/2\sigma^2\right]}{(\sigma\sqrt{2\pi})^n}\right\}.$$

Since the first factor of the right-hand member of this equation depends upon $x_1, x_2, \ldots, x_n$ only through $\bar{x}$, and since the second factor does not depend upon θ, the factorization theorem implies that the mean $\bar{X}$ of the sample is, for any particular value of σ^2, a sufficient statistic for θ, the mean of the normal distribution.

We could have used the definition in the preceding example because we know that $\bar{X}$ is $N(\theta, \sigma^2/n)$. Let us now consider an example in which the use of the definition is inappropriate.

Example 5. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution with p.d.f.

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

where $0 < \theta$. We shall use the factorization theorem to prove that the product $u_1(X_1, X_2, \ldots, X_n) = X_1 X_2 \cdots X_n$ is a sufficient statistic for θ. The joint p.d.f. of $X_1, X_2, \ldots, X_n$ is

$$\theta^n (x_1 x_2 \cdots x_n)^{\theta - 1} = [\theta^n (x_1 x_2 \cdots x_n)^{\theta}]\left(\frac{1}{x_1 x_2 \cdots x_n}\right),$$

where $0 < x_i < 1$, $i = 1, 2, \ldots, n$. In the factorization theorem let

$$k_1[u_1(x_1, x_2, \ldots, x_n); \theta] = \theta^n (x_1 x_2 \cdots x_n)^{\theta}$$

and

$$k_2(x_1, x_2, \ldots, x_n) = \frac{1}{x_1 x_2 \cdots x_n}.$$

Since $k_2(x_1, x_2, \ldots, x_n)$ does not depend upon θ, the product $X_1 X_2 \cdots X_n$ is a sufficient statistic for θ.

There is a tendency for some readers to apply incorrectly the factorization theorem in those instances in which the domain of positive probability density depends upon the parameter θ. This is due to the fact that they do not give proper consideration to the domain of the function $k_2(x_1, x_2, \ldots, x_n)$. This will be illustrated in the next example.

Example 6. In Example 3 with $f(x; \theta) = e^{-(x-\theta)} I_{(\theta,\infty)}(x)$, it was found that the first order statistic Y_1 is a sufficient statistic for θ. To illustrate our point about not considering the domain of the function, take $n = 3$ and note that

$$e^{-(x_1-\theta)} e^{-(x_2-\theta)} e^{-(x_3-\theta)} = [e^{-3 \max x_i + 3\theta}][e^{-x_1 - x_2 - x_3 + 3 \max x_i}]$$

or a similar expression. Certainly, in the latter formula, there is no θ in the second factor and it might be assumed that $Y_3 = \max X_i$ is a sufficient statistic for θ. Of course, this is incorrect because we should have written the joint p.d.f. of X_1, X_2, X_3 as

$$[e^{-(x_1-\theta)} I_{(\theta,\infty)}(x_1)][e^{-(x_2-\theta)} I_{(\theta,\infty)}(x_2)][e^{-(x_3-\theta)} I_{(\theta,\infty)}(x_3)]$$
$$= [e^{3\theta} I_{(\theta,\infty)}(\min x_i)][e^{-x_1 - x_2 - x_3}]$$

because $I_{(\theta,\infty)}(\min x_i) = I_{(\theta,\infty)}(x_1) I_{(\theta,\infty)}(x_2) I_{(\theta,\infty)}(x_3)$. A similar statement cannot be made with $\max x_i$. Thus $Y_1 = \min X_i$ is the sufficient statistic for θ, not $Y_3 = \max X_i$.

EXERCISES

7.10. Let $X_1, X_2, \ldots, X_n$ be a random sample from the normal distribution $N(0, \theta)$, $0 < \theta < \infty$. Show that $\sum_1^n X_i^2$ is a sufficient statistic for θ.

7.11. Prove that the sum of the observations of a random sample of size n from a Poisson distribution having parameter θ, $0 < \theta < \infty$, is a sufficient statistic for θ.

7.12. Show that the nth order statistic of a random sample of size n from the uniform distribution having p.d.f. $f(x; \theta) = 1/\theta$, $0 < x < \theta$, $0 < \theta < \infty$, zero elsewhere, is a sufficient statistic for θ. Generalize this result by

considering the p.d.f. $f(x; \theta) = Q(\theta)M(x)$, $0 < x < \theta$, $0 < \theta < \infty$, zero elsewhere. Here, of course,

$$\int_0^\theta M(x)\,dx = \frac{1}{Q(\theta)}.$$

7.13. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a geometric distribution that has p.d.f. $f(x; \theta) = (1 - \theta)^x\theta$, $x = 0, 1, 2, \ldots$, $0 < \theta < 1$, zero elsewhere. Show that $\sum_1^n X_i$ is a sufficient statistic for θ.

7.14. Show that the sum of the observations of a random sample of size n from a gamma distribution that has p.d.f. $f(x; \theta) = (1/\theta)e^{-x/\theta}$, $0 < x < \infty$, $0 < \theta < \infty$, zero elsewhere, is a sufficient statistic for θ.

7.15. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a beta distribution with parameters $\alpha = \theta > 0$ and $\beta = 2$. Show that the product $X_1 X_2 \cdots X_n$ is a sufficient statistic for θ.

7.16. Show that the product of the sample observations is a sufficient statistic for $\theta > 0$ if the random sample is taken from a gamma distribution with parameters $\alpha = \theta$ and $\beta = 6$.

7.17. What is the sufficient statistic for θ if the sample arises from a beta distribution in which $\alpha = \beta = \theta > 0$?

7.3 Properties of a Sufficient Statistic

Suppose that a random sample $X_1, X_2, \ldots, X_n$ is taken from a distribution with p.d.f. $f(x; \theta)$ that depends upon one parameter $\theta \in \Omega$. Say that a sufficient statistic $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ for θ exists and has p.d.f. $g_1(y_1; \theta)$. Now consider two statisticians, A and B. The first statistician, A, has all of the observed data $x_1, x_2, \ldots, x_n$; but the second, B, has only the value y_1 of the sufficient statistic. Clearly, A has as much information as does B. However, it turns out that B is as well off as A in making statistical inferences about θ in the following sense. Since the conditional probability of $X_1, X_2, \ldots, X_n$, given $Y_1 = y_1$, does not depend upon θ, statistician B can create some pseudo observations, say $Z_1, Z_2, \ldots, Z_n$, that provide a likelihood function that is proportional to that based on $X_1, X_2, \ldots, X_n$ with the factor $g_1(y_1; \theta)$ being common to each likelihood. The other factors of the two likelihood functions do not depend upon θ. Hence, in either case, inferences, like the m.l.e. of θ, would be based upon the sufficient statistic Y_1.

To make this clear, we provide two illustrations. The first is based

upon Example 1 of Section 7.2. There the ratio of the likelihood function and the p.d.f. of Y_1 is

$$\frac{L(\theta)}{g_1(y_1;\theta)}=\frac{1}{\binom{n}{y_1}},$$

where $y_1=\sum_{i=1}^{n} x_i$. Recall that each x_i is equal to zero or 1, and thus y_1 is the sum of y_1 *ones* and $(n-y_1)$ *zeros*. Say we know only the value y_1 and not $x_1, x_2, \ldots, x_n$; so we create pseudovalues $z_1, z_2, \ldots, z_n$ by arranging at random y_1 *ones* and $(n-y_1)$ *zeros* so that the probability of each arrangement is $p=1\Big/\binom{n}{y_1}$. Thus the probability that these z-values equal the original x-values is p, and hence it is highly unlikely, namely with probability p, that those two sets of values would be equal. Yet the two likelihood functions are proportional, namely

$$\left[\binom{n}{y_1}\theta^{y_1}(1-\theta)^{n-y_1}\right]\frac{1}{\binom{n}{\sum x_i}} \propto \left[\binom{n}{y_1}\theta^{y_1}(1-\theta)^{n-y_1}\right]\frac{1}{\binom{n}{\sum z_i}}$$

because $y_1=\sum_{i=1}^{n} x_i=\sum_{i=1}^{n} z_i$. Clearly, the m.l.e. of θ, using either expression, is y_1/n.

The next illustration refers back to Exercise 7.9. There the sample arose from a Poisson distribution with parameter $\theta>0$. It turns out that $Y_1=\sum_{i=1}^{n} X_i$ is a sufficient statistic for θ (see Exercise 7.11). In Exercise 7.9 we found that

$$\frac{L(\theta)}{g_1(y_1;\theta)}=\frac{y_1!}{x_1!\,x_2!\cdots x_n!}\left(\frac{1}{n}\right)^{x_1}\left(\frac{1}{n}\right)^{x_2}\cdots\left(\frac{1}{n}\right)^{x_n},$$

when $L(\theta)$ is the likelihood function based upon $x_1, x_2, \ldots, x_n$. Since this is a multinomial distribution that does not depend upon θ, we can generate some values of $Z_1, Z_2, \ldots, Z_n$, say $z_1, z_2, \ldots, z_n$, that have this multinomial distribution. It is interesting to note that while in the previous examples the z-values provided an arrangement of the x-values, here the z-values do not need to equal those x-values. That is, the values $z_1, z_2, \ldots, z_n$ do not necessarily provide an arrangement of $x_1, x_2, \ldots, x_n$. It is, however, true that $\Sigma z_i=\Sigma x_i=y_1$. Of course,

from the way the z-values were obtained, the two likelihood functions enjoy the property of being proportional, namely

$$g_1(y_1;\theta)\frac{y_1!}{x_1!\,x_2!\cdots x_n!}\left(\frac{1}{n}\right)^{x_1}\left(\frac{1}{n}\right)^{x_2}\cdots\left(\frac{1}{n}\right)^{x_n}$$

$$\propto g_1(y_1;\theta)\frac{y_1!}{z_1!\,z_2!\cdots z_n!}\left(\frac{1}{n}\right)^{z_1}\left(\frac{1}{n}\right)^{z_2}\cdots\left(\frac{1}{n}\right)^{z_n}.$$

Thus, for illustration, using either of these likelihood functions, the m.l.e. of θ is y_1/n because this is the value of θ that maximizes $g_1(y_1;\theta)$.

Since we have considered how the statistician knowing only the value of the sufficient statistic can create a sample that satisfies the likelihood principle; and thus, in this sense, she is as well off as the statistician that has knowledge of all the data. So let us now state a fairly obvious theorem that relates the m.l.e. of θ to a sufficient statistic.

Theorem 2. *Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that has* p.d.f. *$f(x;\theta)$, $\theta\in\Omega$. If a sufficient statistic $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ for θ exists and if a maximum likelihood estimator $\hat{\theta}$ of θ also exists uniquely, then $\hat{\theta}$ is a function of $Y_1 = u_1(X_1, X_2, \ldots, X_n)$.*

Proof. Let $g_1(y_1;\theta)$ be the p.d.f. of Y_1. Then by the definition of sufficiency, the likelihood function

$$L(\theta; x_1, x_2, \ldots, x_n) = f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta)$$

$$= g_1[u_1(x_1, \ldots, x_n);\theta]H(x_1, \ldots, x_n),$$

where $H(x_1, \ldots, x_n)$ does not depend upon θ. Thus L and g_1, as functions of θ, are maximized simultaneously. Since there is one and only one value of θ that maximizes L and hence $g_1[u_1(x_1, \ldots, x_n);\theta]$, that value of θ must be a function of $u_1(x_1, x_2, \ldots, x_n)$. Thus the m.l.e. $\hat{\theta}$ is a function of the sufficient statistic $Y_1 = u_1(X_1, X_2, \ldots, X_n)$.

Let us consider another important property possessed by a sufficient statistic $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ for θ. The conditional p.d.f. of a second statistic, say $Y_2 = u_2(X_1, X_2, \ldots, X_n)$, given $Y_1 = y_1$, does not depend upon θ. On intuitive grounds, we might surmise that the conditional p.d.f. of Y_2, given some linear function $aY_1 + b$, $a \neq 0$, of Y_1, does not depend upon θ. That is, it seems as though the

random variable $aY_1 + b$ is also a sufficient statistic for θ. This conjecture is correct. In fact, every function $Z = u(Y_1)$, or $Z = u[u_1(X_1, X_2, \ldots, X_n)] = v(X_1, X_2, \ldots, X_n)$, not involving θ, with a single-valued inverse $Y_1 = w(Z)$, is also a sufficient statistic for θ. To prove this, we write, in accordance with the factorization theorem,

$$f(x_1; \theta) \cdots f(x_n; \theta) = k_1[u_1(x_1, x_2, \ldots, x_n); \theta]k_2(x_1, x_2, \ldots, x_n).$$

However, we find that $y_1 = w(z)$ or, equivalently, $u_1(x_1, x_2, \ldots, x_n) = w[v(x_1, x_2, \ldots, x_n)]$, which is not a function of θ. Hence

$$f(x_1; \theta) \cdots f(x_n; \theta) = k_1\{w[v(x_1, \ldots, x_n)]; \theta\}k_2(x_1, x_2, \ldots, x_n).$$

Since the first factor of the right-hand member of this equation is a function of $z = v(x_1, \ldots, x_n)$ and θ, while the second factor does not depend upon θ, the factorization theorem implies that $Z = u(Y_1)$ is also a sufficient statistic for θ.

Possibly, the preceding observation is obvious if we think about the sufficient statistic Y_1 partitioning the sample space in such a way that the conditional probability of $X_1, X_2, \ldots, X_n$, given $Y_1 = y_1$, does not depend upon θ. We say this because every function $Z = u(Y_1)$ with a single-valued inverse $Y_1 = w(Z)$ would partition the sample space in exactly the same way, that is, the set of points

$$\{(x_1, x_2, \ldots, x_n) : u_1(x_1, x_2, \ldots, x_n) = y_1\},$$

for each y_1, is exactly the same as

$$\{(x_1, x_2, \ldots, x_n) : v(x_1, x_2, \ldots, x_n) = u(y_1)\}$$

because $w[v(x_1, x_2, \ldots, x_n)] = u_1(x_1, x_2, \ldots, x_n) = y_1$.

Remark. Throughout the discussion of sufficient statistics, as a matter of fact throughout much of the mathematics of statistical inference, we hope the reader recognizes the importance of the assumption of having a certain model. Clearly, when we say that a statistician having the value of a certain statistic (here sufficient) is as well off in making statistical inferences as the statistician who has all of the data, we depend upon the fact that a certain model is true. For illustration, knowing that we have i.i.d. variables, each with p.d.f. $f(x; \theta)$, is extremely important; because if that $f(x; \theta)$ is incorrect or if the independence assumption does not hold, our resulting inferences could be very bad. The statistician with all the data could—and should—check to see if the model is reasonably good. Such procedures checking the model are often called *model diagnostics*, the discussion of which we leave to a more applied course in statistics.

We now consider a result of Rao and Blackwell from which we see that we need consider only functions of the sufficient statistic in finding the unbiased point estimates of parameters. In showing this, we can refer back to a result of Section 2.2: If X_1 and X_2 are random variables and certain expectations exist, then

$$E[X_2] = E[E(X_2|X_1)]$$

and

$$\text{var}\,(X_2) \geq \text{var}\,[E(X_2|X_1)].$$

For the adaptation in context of sufficient statistics, we let the sufficient statistic Y_1 be X_1 and Y_2, an unbiased statistic of θ, be X_2. Thus, with $E(Y_2|y_1) = \varphi(y_1)$, we have

$$\theta = E(Y_2) = E[\varphi(Y_1)]$$

and

$$\text{var}\,(Y_2) \geq \text{var}\,[\varphi(Y_1)].$$

That is, through this conditioning, the function $\varphi(Y_1)$ of the sufficient statistic Y_1 is an unbiased estimator of θ having smaller variance than that of the unbiased estimator Y_2. We summarize this discussion more formally in the following theorem, which can be attributed to Rao and Blackwell.

Theorem 3. *Let $X_1, X_2, \ldots, X_n$, n a fixed positive integer, denote a random sample from a distribution (continuous or discrete) that has* p.d.f. *$f(x; \theta)$, $\theta \in \Omega$. Let $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ be a sufficient statistic for θ, and let $Y_2 = u_2(X_1, X_2, \ldots, X_n)$, not a function of Y_1 alone, be an unbiased estimator of θ. Then $E(Y_2|y_1) = \varphi(y_1)$ defines a statistic $\varphi(Y_1)$. This statistic $\varphi(Y_1)$ is a function of the sufficient statistic for θ; it is an unbiased estimator of θ; and its variance is less than that of Y_2.*

This theorem tells us that in our search for an unbiased minimum variance estimator of a parameter, we may, if a sufficient statistic for the parameter exists, restrict that search to functions of the sufficient statistic. For if we begin with an unbiased estimator Y_2 that is not a function of the sufficient statistic Y_1 alone, then we can always improve on this by computing $E(Y_2|y_1) = \varphi(y_1)$ so that $\varphi(Y_1)$ is an unbiased estimator with smaller variance than that of Y_2.

After Theorem 3 many students believe that it is necessary to find

first some unbiased estimator Y_2 in their search for $\varphi(Y_1)$, an unbiased estimator of θ based upon the sufficient statistic Y_1. This is not the case at all, and Theorem 3 simply convinces us that we can restrict our search for a best estimator to functions of Y_1. It frequently happens that $E(Y_1) = a\theta + b$, where $a \neq 0$ and b are constants, and thus $(Y_1 - b)/a$ is a function of Y_1 that is an unbiased estimator of θ. That is, we can usually find an unbiased estimator based on Y_1 without first finding an estimator Y_2. In the next two sections we discover that, in most instances, if there is one function $\varphi(Y_1)$ that is unbiased, $\varphi(Y_1)$ is the only unbiased estimator based on the sufficient statistic Y_1.

Remark. Since the unbiased estimator $\varphi(Y_1)$, where $\varphi(y_1) = E(Y_2|y_1)$, has variance smaller than that of the unbiased estimator Y_2 of θ, students sometimes reason as follows. Let the function $\Upsilon(y_3) = E[\varphi(Y_1)|Y_3 = y_3]$, where Y_3 is another statistic, which is not sufficient for θ. By the Rao–Blackwell theorem, we have that $E[\Upsilon(Y_3)] = \theta$ and $\Upsilon(Y_3)$ has a smaller variance than does $\varphi(Y_1)$. Accordingly, $\Upsilon(Y_3)$ must be better than $\varphi(Y_1)$ as an unbiased estimator of θ. But this is *not* true because Y_3 is not sufficient; thus θ is present in the conditional distribution of Y_1, given $Y_3 = y_3$, and the conditional mean $\Upsilon(y_3)$. So although indeed $E[\Upsilon(Y_3)] = \theta$, $\Upsilon(Y_3)$ is not even a statistic because it involves the unknown parameter θ and hence cannot be used as an estimator.

Example 1. Let X_1, X_2, X_3 be a random sample from an exponential distribution with mean $\theta > 0$, so that the joint p.d.f. is

$$\left(\frac{1}{\theta}\right)^3 e^{-(x_1 + x_2 + x_3)/\theta}, \qquad 0 < x_i < \infty,$$

$i = 1, 2, 3$, zero elsewhere. From the factorization theorem, we see that $Y_1 = X_1 + X_2 + X_3$ is a sufficient statistic for θ. Of course,

$$E(Y_1) = E(X_1 + X_2 + X_3) = 3\theta,$$

and thus $Y_1/3 = \bar{X}$ is a function of the sufficient statistic that is an unbiased estimator of θ.

In addition, let $Y_2 = X_2 + X_3$ and $Y_3 = X_3$. The one-to-one transformation defined by

$$x_1 = y_1 - y_2, \qquad x_2 = y_2 - y_3, \qquad x_3 = y_3$$

has Jacobian equal to 1 and the joint p.d.f. of Y_1, Y_2, Y_3 is

$$g(y_1, y_2, y_3; \theta) = \left(\frac{1}{\theta}\right)^3 e^{-y_1/\theta}, \qquad 0 < y_3 < y_2 < y_1 < \infty,$$

zero elsewhere. The marginal p.d.f. of Y_1 and Y_3 is found by integrating out y_2 to obtain

$$g_{13}(y_1, y_3; \theta) = \left(\frac{1}{\theta}\right)^3 (y_1 - y_3)e^{-y_1/\theta}, \qquad 0 < y_3 < y_1 < \infty,$$

zero elsewhere. The p.d.f. of Y_3 alone is

$$g_3(y_3; \theta) = \frac{1}{\theta} e^{-y_3/\theta}, \qquad 0 < y_3 < \infty,$$

zero elsewhere, since $Y_3 = X_3$ is an observation of a random sample from this exponential distribution.

Accordingly, the conditional p.d.f. of Y_1, given $Y_3 = y_3$, is

$$g_{1|3}(y_1|y_3) = \frac{g_{13}(y_1, y_3; \theta)}{g_3(y_3; \theta)}$$

$$= \left(\frac{1}{\theta}\right)^2 (y_1 - y_3)e^{-(y_1 - y_3)/\theta}, \qquad 0 < y_3 < y_1 < \infty,$$

zero elsewhere. Thus

$$E\left(\frac{Y_1}{3}|y_3\right) = E\left(\frac{Y_1 - Y_3}{3}|y_3\right) + E\left(\frac{Y_3}{3}|y_3\right)$$

$$= \left(\frac{1}{3}\right)\int_{y_3}^{\infty} \left(\frac{1}{\theta}\right)^2 (y_1 - y_3)^2 e^{-(y_1 - y_3)/\theta}\, dy_1 + \frac{y_3}{3}$$

$$= \left(\frac{1}{3}\right)\frac{\Gamma(3)\theta^3}{\theta^2} + \frac{y_3}{3} = \frac{2\theta}{3} + \frac{y_3}{3} = \Upsilon(y_3).$$

Of course, $E[\Upsilon(Y_3)] = \theta$ and $\text{var}\,[\Upsilon(Y_3)] \leq \text{var}\,(Y_1/3)$, but $\Upsilon(Y_3)$ is not a statistic as it involves θ and cannot be used as an estimator of θ. This illustrates the preceding remark.

EXERCISES

7.18. In each of the Exercises 7.10, 7.11, 7.13, and 7.14, show that the m.l.e. of θ is a function of the sufficient statistic for θ.

7.19. Let $Y_1 < Y_2 < Y_3 < Y_4 < Y_5$ be the order statistics of a random sample of size 5 from the uniform distribution having p.d.f. $f(x; \theta) = 1/\theta$, $0 < x < \theta$, $0 < \theta < \infty$, zero elsewhere. Show that $2Y_3$ is an unbiased estimator of θ. Determine the joint p.d.f. of Y_3 and the sufficient statistic Y_5 for θ. Find the conditional expectation $E(2Y_3|y_5) = \varphi(y_5)$. Compare the variances of $2Y_3$ and $\varphi(Y_5)$.

Hint: All of the integrals needed in this exercise can be evaluated by making a change of variable such as $z = y/\theta$ and using the results associated with the beta p.d.f.; see Section 4.4.

7.20. If X_1, X_2 is a random sample of size 2 from a distribution having p.d.f. $f(x; \theta) = (1/\theta)e^{-x/\theta}$, $0 < x < \infty$, $0 < \theta < \infty$, zero elsewhere, find the joint p.d.f. of the sufficient statistic $Y_1 = X_1 + X_2$ for θ and $Y_2 = X_2$. Show that Y_2 is an unbiased estimator of θ with variance θ^2. Find $E(Y_2|y_1) = \varphi(y_1)$ and the variance of $\varphi(Y_1)$.

7.21. Let the random variables X and Y have the joint p.d.f. $f(x, y) = (2/\theta^2)e^{-(x+y)/\theta}$, $0 < x < y < \infty$, zero elsewhere.

(a) Show that the mean and the variance of Y are, respectively, $3\theta/2$ and $5\theta^2/4$.

(b) Show that $E(Y|x) = x + \theta$. In accordance with the theory, the expected value of $X + \theta$ is that of Y, namely, $3\theta/2$, and the variance of $X + \theta$ is less than that of Y. Show that the variance of $X + \theta$ is in fact $\theta^2/4$.

7.22. In each of Exercises 7.10, 7.11, and 7.12, compute the expected value of the given sufficient statistic and, in each case, determine an unbiased estimator of θ that is a function of that sufficient statistic alone.

7.4 Completeness and Uniqueness

Let $X_1, X_2, \ldots, X_n$ be a random sample from the Poisson distribution that has p.d.f.

$$f(x; \theta) = \frac{\theta^x e^{-\theta}}{x!}, \qquad x = 0, 1, 2, \ldots; \quad 0 < \theta;$$

$$= 0 \qquad \text{elsewhere.}$$

From Exercise 7.11 of Section 7.2 we know that $Y_1 = \sum_{i=1}^{n} X_i$ is a sufficient statistic for θ and its p.d.f. is

$$g_1(y_1; \theta) = \frac{(n\theta)^{y_1} e^{-n\theta}}{y_1!}, \qquad y_1 = 0, 1, 2, \ldots,$$

$$= 0 \qquad \text{elsewhere.}$$

Let us consider the family $\{g_1(y_1; \theta) : 0 < \theta\}$ of probability density functions. Suppose that the function $u(Y_1)$ of Y_1 is such that $E[u(Y_1)] = 0$ for every $\theta > 0$. We shall show that this requires $u(y_1)$ to be zero at every point $y_1 = 0, 1, 2, \ldots$. That is,

$$E[u(Y_1)] = 0, \qquad 0 < \theta,$$

implies that

$$0 = u(0) = u(1) = u(2) = u(3) = \cdots.$$

We have for all $\theta > 0$ that

$$0 = E[u(Y_1)] = \sum_{y_1=0}^{\infty} u(y_1)\frac{(n\theta)^{y_1}e^{-n\theta}}{y_1!}$$

$$= e^{-n\theta}\left[u(0) + u(1)\frac{n\theta}{1!} + u(2)\frac{(n\theta)^2}{2!} + \cdots\right].$$

Since $e^{-n\theta}$ does not equal zero, we have that

$$0 = u(0) + [nu(1)]\theta + \left[\frac{n^2u(2)}{2}\right]\theta^2 + \cdots.$$

However, if such an infinite series converges to zero for all $\theta > 0$, then each of the coefficients must equal zero. That is,

$$u(0) = 0, \qquad nu(1) = 0, \qquad \frac{n^2u(2)}{2} = 0, \ldots$$

and thus $0 = u(0) = u(1) = u(2) = \cdots$, as we wanted to show. Of course, the condition $E[u(Y_1)] = 0$ for all $\theta > 0$ does not place any restriction on $u(y_1)$ when y_1 is not a nonnegative integer. So we see that, in this illustration, $E[u(Y_1)] = 0$ for all $\theta > 0$ requires that $u(y_1)$ equals zero except on a set of points that has probability zero for each p.d.f. $g_1(y_1; \theta)$, $0 < \theta$. From the following definition we observe that the family $\{g_1(y_1; \theta) : 0 < \theta\}$ is complete.

Definition 3. Let the random variable Z of either the continuous type or the discrete type have a p.d.f. that is one member of the family $\{h(z; \theta) : \theta \in \Omega\}$. If the condition $E[u(Z)] = 0$, for every $\theta \in \Omega$, requires that $u(z)$ be zero except on a set of points that has probability zero for each p.d.f. $h(z; \theta)$, $\theta \in \Omega$, then the family $\{h(z; \theta) : \theta \in \Omega\}$ is called a *complete family* of probability density functions.

Remark. In Section 1.9 it was noted that the existence of $E[u(X)]$ implies that the integral (or sum) converges absolutely. This absolute convergence was tacitly assumed in our definition of completeness and it is needed to prove that certain families of probability density functions are complete.

In order to show that certain families of probability density functions of the continuous type are complete, we must appeal to the same type of theorem in analysis that we used when we claimed that the moment-generating function uniquely determines a distribution. This is illustrated in the next example.

Example 1. Let Z have a p.d.f. that is a member of the family $\{h(z;\theta): 0 < \theta < \infty\}$, where

$$h(z;\theta) = \frac{1}{\theta} e^{-z/\theta}, \qquad 0 < z < \infty,$$
$$= 0 \qquad \text{elsewhere.}$$

Let us say that $E[u(Z)] = 0$ for every $\theta > 0$. That is,

$$\frac{1}{\theta}\int_0^\infty u(z)e^{-z/\theta}\,dz = 0, \qquad \text{for } \theta > 0.$$

Readers acquainted with the theory of transforms will recognize the integral in the left-hand member as being essentially the Laplace transform of $u(z)$. In that theory we learn that the only function $u(z)$ transforming to a function of θ which is identically equal to zero is $u(z) = 0$, except (in our terminology) on a set of points that has probability zero for each $h(z;\theta)$, $0 < \theta$. That is, the family $\{h(z;\theta): 0 < \theta < \infty\}$ is complete.

Let the parameter θ in the p.d.f. $f(x;\theta)$, $\theta \in \Omega$, have a sufficient statistic $Y_1 = u_1(X_1, X_2, \ldots, X_n)$, where $X_1, X_2, \ldots, X_n$ is a random sample from this distribution. Let the p.d.f. of Y_1 be $g_1(y_1;\theta)$, $\theta \in \Omega$. It has been seen that, if there is any unbiased estimator Y_2 (not a function of Y_1 alone) of θ, then there is at least one function of Y_1 that is an unbiased estimator of θ, and our search for a best estimator of θ may be restricted to functions of Y_1. Suppose it has been verified that a certain function $\varphi(Y_1)$, not a function of θ, is such that $E[\varphi(Y_1)] = \theta$ for all values of θ, $\theta \in \Omega$. Let $\psi(Y_1)$ be another function of the sufficient statistic Y_1 alone, so that we also have $E[\psi(Y_1)] = \theta$ for all values of θ, $\theta \in \Omega$. Hence

$$E[\varphi(Y_1) - \psi(Y_1)] = 0, \qquad \theta \in \Omega.$$

If the family $\{g_1(y_1;\theta): \theta \in \Omega\}$ is complete, the function $\varphi(y_1) - \psi(y_1) = 0$, except on a set of points that has probability zero. That is, for every other unbiased estimator $\psi(Y_1)$ of θ, we have

$$\varphi(y_1) = \psi(y_1)$$

except possibly at certain special points. Thus, in this sense [namely $\varphi(y_1) = \psi(y_1)$, except on a set of points with probability zero], $\varphi(Y_1)$ is the unique function of Y_1, which is an unbiased estimator of θ. In accordance with the Rao–Blackwell theorem, $\varphi(Y_1)$ has a smaller variance than every other unbiased estimator of θ. That is, the

statistic $\varphi(Y_1)$ is the unbiased minimum variance estimator of θ. This fact is stated in the following theorem of Lehmann and Scheffé.

Theorem 4. *Let $X_1, X_2, \ldots, X_n$, n a fixed positive integer, denote a random sample from a distribution that has* p.d.f. *$f(x; \theta)$, $\theta \in \Omega$, let $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ be a sufficient statistic for θ, and let the family $\{g_1(y_1; \theta) : \theta \in \Omega\}$ of probability density functions be complete. If there is a function of Y_1 that is an unbiased estimator of θ, then this function of Y_1 is the unique unbiased minimum variance estimator of θ. Here "unique" is used in the sense described in the preceding paragraph.*

The statement that Y_1 is a sufficient statistic for a parameter θ, $\theta \in \Omega$, and that the family $\{g_1(y_1; \theta) : \theta \in \Omega\}$ of probability density functions is complete is lengthy and somewhat awkward. We shall adopt the less descriptive, but more convenient, terminology that Y_1 is a *complete sufficient statistic* for θ. In the next section we study a fairly large class of probability density functions for which a complete sufficient statistic Y_1 for θ can be determined by inspection.

EXERCISES

7.23. If $az^2 + bz + c = 0$ for more than two values of z, then $a = b = c = 0$. Use this result to show that the family $\{b(2, \theta) : 0 < \theta < 1\}$ is complete.

7.24. Show that each of the following families is not complete by finding at least one nonzero function $u(x)$ such that $E[u(X)] = 0$, for all $\theta > 0$.

(a) $f(x; \theta) = \dfrac{1}{2\theta}, \qquad -\theta < x < \theta, \quad \text{where } 0 < \theta < \infty,$

$= 0 \qquad \text{elsewhere.}$

(b) $N(0, \theta)$, where $0 < \theta < \infty$.

7.25. Let $X_1, X_2, \ldots, X_n$ represent a random sample from the discrete distribution having the probability density function

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

$$= 0 \qquad \text{elsewhere.}$$

Show that $Y_1 = \sum_1^n X_i$ is a complete sufficient statistic for θ. Find the unique function of Y_1 that is the unbiased minimum variance estimator of θ.

Hint: Display $E[u(Y_1)] = 0$, show that the constant term $u(0)$ is equal to zero, divide both members of the equation by $\theta \neq 0$, and repeat the argument.

7.26. Consider the family of probability density functions $\{h(z; \theta) : \theta \in \Omega\}$, where $h(z; \theta) = 1/\theta$, $0 < z < \theta$, zero elsewhere.

(a) Show that the family is complete provided that $\Omega = \{\theta : 0 < \theta < \infty\}$. *Hint:* For convenience, assume that $u(z)$ is continuous and note that the derivative of $E[u(Z)]$ with respect to θ is equal to zero also.

(b) Show that this family is not complete if $\Omega = \{\theta : 1 < \theta < \infty\}$. *Hint:* Concentrate on the interval $0 < z < 1$ and find a nonzero function $u(z)$ on that interval such that $E[u(Z)] = 0$ for all $\theta > 1$.

7.27. Show that the first order statistic Y_1 of a random sample of size n from the distribution having p.d.f. $f(x; \theta) = e^{-(x-\theta)}$, $\theta < x < \infty$, $-\infty < \theta < \infty$, zero elsewhere, is a complete sufficient statistic for θ. Find the unique function of this statistic which is the unbiased minimum variance estimator of θ.

7.28. Let a random sample of size n be taken from a distribution of the discrete type with p.d.f. $f(x; \theta) = 1/\theta$, $x = 1, 2, \ldots, \theta$, zero elsewhere, where θ is an unknown positive integer.

(a) Show that the largest observation, say Y, of the sample is a complete sufficient statistic for θ.

(b) Prove that

$$[Y^{n+1} - (Y-1)^{n+1}]/[Y^n - (Y-1)^n]$$

is the unique unbiased minimum variance estimator of θ.

7.5 The Exponential Class of Probability Density Functions

Consider a family $\{f(x; \theta) : \theta \in \Omega\}$ of probability density functions, where Ω is the interval set $\Omega = \{\theta : \gamma < \theta < \delta\}$, where γ and δ are known constants, and where

$$\begin{aligned} f(x; \theta) &= \exp[p(\theta)K(x) + S(x) + q(\theta)], \qquad a < x < b, \\ &= 0 \qquad \text{elsewhere.} \end{aligned} \tag{1}$$

A p.d.f. of the form (1) is said to be a member of the *exponential class* of probability density functions of the continuous type. If, in addition,

1. neither a nor b depends upon θ, $\gamma < \theta < \delta$,
2. $p(\theta)$ is a nontrivial continuous function of θ, $\gamma < \theta < \delta$,
3. each of $K'(x) \not\equiv 0$ and $S(x)$ is a continuous function of x, $a < x < b$,

we say that we have a *regular case* of the exponential class. A p.d.f.

$$f(x;\theta) = \exp[p(\theta)K(x) + S(x) + q(\theta)], \qquad x = a_1, a_2, a_3, \ldots,$$
$$= 0 \qquad \text{elsewhere,}$$

is said to represent a regular case of the exponential class of probability density functions of the discrete type if

1. The set $\{x : x = a_1, a_2, \ldots\}$ does not depend upon θ.
2. $p(\theta)$ is a nontrivial continuous function of θ, $\gamma < \theta < \delta$.
3. $K(x)$ is a nontrivial function of x on the set $\{x : x = a_1, a_2, \ldots\}$.

For example, each member of the family $\{f(x;\theta) : 0 < \theta < \infty\}$, where $f(x;\theta)$ is $N(0,\theta)$, represents a regular case of the exponential class of the continuous type because

$$f(x;\theta) = \frac{1}{\sqrt{2\pi\theta}} e^{-x^2/2\theta}$$
$$= \exp\left(-\frac{1}{2\theta}x^2 - \ln\sqrt{2\pi\theta}\right), \qquad -\infty < x < \infty.$$

Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that has a p.d.f. that represents a regular case of the exponential class of the continuous type. The joint p.d.f. of $X_1, X_2, \ldots, X_n$ is

$$\exp\left[p(\theta)\sum_1^n K(x_i) + \sum_1^n S(x_i) + nq(\theta)\right]$$

for $a < x_i < b$, $i = 1, 2, \ldots, n$, $\gamma < \theta < \delta$, and is zero elsewhere. At points of positive probability density, this joint p.d.f. may be written as the product of the two nonnegative functions

$$\exp\left[p(\theta)\sum_1^n K(x_i) + nq(\theta)\right]\exp\left[\sum_1^n S(x_i)\right].$$

In accordance with the factorization theorem (Theorem 1, Section 7.2) $Y_1 = \sum_1^n K(X_i)$ is a sufficient statistic for the parameter θ. To prove that $Y_1 = \sum_1^n K(X_i)$ is a sufficient statistic for θ in the discrete case, we take the joint p.d.f. of $X_1, X_2, \ldots, X_n$ to be positive on a discrete set of points, say, when $x_i \in \{x : x = a_1, a_2, \ldots\}$, $i = 1, 2, \ldots, n$. We then use the factorization theorem. It is left as an exercise to show that in either the continuous or the discrete case the p.d.f. of Y_1 is of the form

$$g_1(y_1;\theta) = R(y_1)\exp[p(\theta)y_1 + nq(\theta)]$$

at points of positive probability density. The points of positive probability density and the function $R(y_1)$ do not depend upon θ.

At this time we use a theorem in analysis to assert that the family $\{g_1(y_1; \theta) : \gamma < \theta < \delta\}$ of probability density functions is complete. This is the theorem we used when we asserted that a moment-generating function (when it exists) uniquely determines a distribution. In the present context it can be stated as follows.

Theorem 5. *Let $f(x; \theta)$, $\gamma < \theta < \delta$, be a* p.d.f. *which represents a regular case of the exponential class. Then if $X_1, X_2, \ldots, X_n$ (where n is a fixed positive integer) is a random sample from a distribution with* p.d.f. *$f(x; \theta)$, the statistic $Y_1 = \sum_1^n K(X_i)$ is a sufficient statistic for θ and the family $\{g_1(y_1; \theta) : \gamma < \theta < \delta\}$ of probability density functions of Y_1 is complete. That is, Y_1 is a complete sufficient statistic for θ.*

This theorem has useful implications. In a regular case of form (1), we can see by inspection that the sufficient statistic is $Y_1 = \sum_1^n K(X_i)$. If we can see how to form a function of Y_1, say $\varphi(Y_1)$, so that $E[\varphi(Y_1)] = \theta$, then the statistic $\varphi(Y_1)$ is unique and is the unbiased minimum variance estimator of θ.

Example 1. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a normal distribution that has p.d.f.

$$f(x; \theta) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\theta)^2}{2\sigma^2}\right], \qquad -\infty < x < \infty, \quad -\infty < \theta < \infty,$$

or

$$f(x; \theta) = \exp\left(\frac{\theta}{\sigma^2}x - \frac{x^2}{2\sigma^2} - \ln\sqrt{2\pi\sigma^2} - \frac{\theta^2}{2\sigma^2}\right).$$

Here σ^2 is any fixed positive number. This is a regular case of the exponential class with

$$p(\theta) = \frac{\theta}{\sigma^2}, \qquad K(x) = x,$$

$$S(x) = -\frac{x^2}{2\sigma^2} - \ln\sqrt{2\pi\sigma^2}, \qquad q(\theta) = -\frac{\theta^2}{2\sigma^2}.$$

Accordingly, $Y_1 = X_1 + X_2 + \cdots + X_n = n\bar{X}$ is a complete sufficient statistic for the mean θ of a normal distribution for every fixed value of the variance σ^2. Since $E(Y_1) = n\theta$, then $\varphi(Y_1) = Y_1/n = \bar{X}$ is the only function of Y_1 that is an unbiased estimator of θ; and being a function of the sufficient statistic

Y_1, it has a minimum variance. That is, $\bar{X}$ is the unique unbiased minimum variance estimator of θ. Incidentally, since Y_1 is a one-to-one function of $\bar{X}$, $\bar{X}$ itself is also a complete sufficient statistic for θ.

Example 2. Consider a Poisson distribution with parameter $\theta, 0 < \theta < \infty$. The p.d.f. of this distribution is

$$f(x; \theta) = \frac{\theta^x e^{-\theta}}{x!} = \exp[(\ln \theta)x - \ln(x!) - \theta], \qquad x = 0, 1, 2, \ldots,$$
$$= 0 \qquad \text{elsewhere.}$$

In accordance with Theorem 5, $Y_1 = \sum_1^n X_i$ is a complete sufficient statistic for θ. Since $E(Y_1) = n\theta$, the statistic $\varphi(Y_1) = Y_1/n = \bar{X}$, which is also a complete sufficient statistic for θ, is the unique unbiased minimum variance estimator of θ.

EXERCISES

7.29. Write the p.d.f.

$$f(x; \theta) = \frac{1}{6\theta^4} x^3 e^{-x/\theta}, \qquad 0 < x < \infty, \quad 0 < \theta < \infty,$$

zero elsewhere, in the exponential form. If $X_1, X_2, \ldots, X_n$ is a random sample from this distribution, find a complete sufficient statistic Y_1 for θ and the unique function $\varphi(Y_1)$ of this statistic that is the unbiased minimum variance estimator of θ. Is $\varphi(Y_1)$ itself a complete sufficient statistic?

7.30. Let $X_1, X_2, \ldots, X_n$ denote a random sample of size $n > 1$ from a distribution with p.d.f. $f(x; \theta) = \theta e^{-\theta x}$, $0 < x < \infty$, zero elsewhere, and $\theta > 0$. Then $Y = \sum_1^n X_i$ is a sufficient statistic for θ. Prove that $(n-1)/Y$ is the unbiased minimum variance estimator of θ.

7.31. Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from a distribution with p.d.f. $f(x; \theta) = \theta x^{\theta - 1}$, $0 < x < 1$, zero elsewhere, and $\theta > 0$.

(a) Show that the *geometric mean* $(X_1 X_2 \cdots X_n)^{1/n}$ of the sample is a complete sufficient statistic for θ.

(b) Find the maximum likelihood estimator of θ, and observe that it is a function of this geometric mean.

7.32. Let $\bar{X}$ denote the mean of the random sample $X_1, X_2, \ldots, X_n$ from a gamma-type distribution with parameters $\alpha > 0$ and $\beta = \theta > 0$. Compute $E[X_1 | \bar{x}]$.

Hint: Can you find directly a function $\psi(\bar{X})$ of $\bar{X}$ such that $E[\psi(\bar{X})] = \theta$? Is $E(X_1|\bar{x}) = \psi(\bar{x})$? Why?

7.33. Let X be a random variable with a p.d.f. of a regular case of the exponential class. Show that $E[K(X)] = -q'(\theta)/p'(\theta)$, provided these derivatives exist, by differentiating both members of the equality

$$\int_a^b \exp[p(\theta)K(x) + S(x) + q(\theta)]\,dx = 1$$

with respect to θ. By a second differentiation, find the variance of $K(X)$.

7.34. Given that $f(x; \theta) = \exp[\theta K(x) + S(x) + q(\theta)]$, $a < x < b$, $\gamma < \theta < \delta$, represents a regular case of the exponential class, show that the moment-generating function $M(t)$ of $Y = K(X)$ is $M(t) = \exp[q(\theta) - q(\theta + t)]$, $\gamma < \theta + t < \delta$.

7.35. Given, in the preceding exercise, that $E(Y) = E[K(X)] = \theta$. Prove that Y is $N(\theta, 1)$.

Hint: Consider $M'(0) = \theta$ and solve the resulting differential equation.

7.36. If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution that has a p.d.f. which is a regular case of the exponential class, show that the p.d.f. of $Y_1 = \sum_1^n K(X_i)$ is of the form $g_1(y_1; \theta) = R(y_1)\exp[p(\theta)y_1 + nq(\theta)]$.

Hint: Let $Y_2 = X_2, \ldots, Y_n = X_n$ be $n - 1$ auxiliary random variables. Find the joint p.d.f. of $Y_1, Y_2, \ldots, Y_n$ and then the marginal p.d.f. of Y_1.

7.37. Let Y denote the median and let $\bar{X}$ denote the mean of a random sample of the size $n = 2k + 1$ from a distribution that is $N(\mu, \sigma^2)$. Compute $E(Y|\bar{X} = \bar{x})$.

Hint: See Exercise 7.32.

7.38. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with p.d.f. $f(x; \theta) = \theta^2 x e^{-\theta x}$, $0 < x < \infty$, where $\theta > 0$.

(a) Argue that $Y = \sum_1^n X_i$ is a complete sufficient statistic for θ.

(b) Compute $E(1/Y)$ and find the function of Y which is the unique unbiased minimum variance estimator of θ.

7.39. Let $X_1, X_2, \ldots, X_n$, $n > 2$, be a random sample from the binomial distribution $b(1, \theta)$.

(a) Show that $Y_1 = X_1 + X_2 + \cdots + X_n$ is a complete sufficient statistic for θ.

(b) Find the function $\varphi(Y_1)$ which is the unbiased minimum variance estimator of θ.

(c) Let $Y_2 = (X_1 + X_2)/2$ and compute $E(Y_2)$.

(d) Determine $E(Y_2|Y_1 = y_1)$.

7.6 Functions of a Parameter

Up to this point we have sought an unbiased and minimum variance estimator of a parameter θ. Not always, however, are we interested in θ but rather in a function of θ. This will be illustrated in the following examples.

Example 1. Let $X_1, X_2, \ldots, X_n$ denote the observations of a random sample of size $n > 1$ from a distribution that is $b(1, \theta)$, $0 < \theta < 1$. We know that if $Y = \sum_1^n X_i$, then Y/n is the unique unbiased minimum variance estimator of θ. Now the variance of Y/n is $\theta(1 - \theta)/n$. Suppose that an unbiased and minimum variance estimator of this variance is sought. Because Y is a sufficient statistic for θ, it is known that we can restrict our search to functions of Y. Consider the statistic $(Y/n)(1 - Y/n)/n$. This statistic is suggested by the fact that Y/n is an estimator of θ. The expectation of this statistic is given by

$$\frac{1}{n} E\left[\frac{Y}{n}\left(1 - \frac{Y}{n}\right)\right] = \frac{1}{n^2} E(Y) - \frac{1}{n^3} E(Y^2).$$

Now $E(Y) = n\theta$ and $E(Y^2) = n\theta(1 - \theta) + n^2\theta^2$. Hence

$$\frac{1}{n} E\left[\frac{Y}{n}\left(1 - \frac{Y}{n}\right)\right] = \frac{n-1}{n}\frac{\theta(1-\theta)}{n}.$$

If we multiply both members of this equation by $n/(n - 1)$, we find that the statistic $(Y/n)(1 - Y/n)/(n - 1)$ is the unique unbiased minimum variance estimator of the variance of Y/n.

A somewhat different, but also very important problem in point estimation is considered in the next example. In the example the distribution of a random variable X is described by a p.d.f. $f(x; \theta)$ that depends upon $\theta \in \Omega$. The problem is to estimate the fractional part of the probability for this distribution which is at, or to the left of, a fixed point c. Thus we seek an unbiased minimum variance estimator of $F(c; \theta)$, where $F(x; \theta)$ is the distribution function of X.

Example 2. Let $X_1, X_2, \ldots, X_n$ be a random sample of size $n > 1$ from a distribution that is $N(\theta, 1)$. Suppose that we wish to find an unbiased minimum variance estimator of the function of θ defined by

$$\Pr(X \leq c) = \int_{-\infty}^{c} \frac{1}{\sqrt{2\pi}} e^{-(x-\theta)^2/2}\, dx = \Phi(c - \theta),$$

where c is a fixed constant. There are many unbiased estimators of $\Phi(c - \theta)$. We first exhibit one of these, say $u(X_1)$, a function of X_1 alone. We shall then

compute the conditional expectation, $E[u(X_1)|\bar{X} = \bar{x}] = \varphi(\bar{x})$, of this unbiased statistic, given the sufficient statistic $\bar{X}$, the mean of the sample. In accordance with the theorems of Rao–Blackwell and Lehmann–Scheffé, $\varphi(\bar{X})$ is the unique unbiased minimum variance estimator of $\Phi(c - \theta)$.

Consider the function $u(x_1)$, where

$$u(x_1) = 1, \qquad x_1 \leq c,$$
$$= 0, \qquad x_1 > c.$$

The expected value of the random variable $u(X_1)$ is given by

$$E[u(X_1)] = \int_{-\infty}^{\infty} u(x_1) \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x_1 - \theta)^2}{2}\right] dx_1$$
$$= \int_{-\infty}^{c} (1) \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x_1 - \theta)^2}{2}\right] dx_1,$$

because $u(x_1) = 0$, $x_1 > c$. But the latter integral has the value $\Phi(c - \theta)$. That is, $u(X_1)$ is an unbiased estimator of $\Phi(c - \theta)$.

We shall next discuss the joint distribution of X_1 and $\bar{X}$ and the conditional distribution of X_1, given $\bar{X} = \bar{x}$. This conditional distribution will enable us to compute $E[u(X_1)|\bar{X} = \bar{x}] = \varphi(\bar{x})$. In accordance with Exercise 4.92, Section 4.7, the joint distribution of X_1 and $\bar{X}$ is bivariate normal with means θ and θ, variances $\sigma_1^2 = 1$ and $\sigma_2^2 = 1/n$, and correlation coefficient $\rho = 1/\sqrt{n}$. Thus the conditional p.d.f. of X_1, given $\bar{X} = \bar{x}$, is normal with linear conditional mean

$$\theta + \frac{\rho\sigma_1}{\sigma_2}(\bar{x} - \theta) = \bar{x}$$

and with variance

$$\sigma_1^2(1 - \rho^2) = \frac{n - 1}{n}.$$

The conditional expectation of $u(X_1)$, given $\bar{X} = \bar{x}$, is then

$$\varphi(\bar{x}) = \int_{-\infty}^{\infty} u(x_1) \sqrt{\frac{n}{n-1}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{n(x_1 - \bar{x})^2}{2(n-1)}\right] dx_1$$
$$= \int_{-\infty}^{c} \sqrt{\frac{n}{n-1}} \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{n(x_1 - \bar{x})^2}{2(n-1)}\right] dx_1.$$

The change of variable $z = \sqrt{n}(x_1 - \bar{x})/\sqrt{n-1}$ enables us to write, with $c' = \sqrt{n}(c - \bar{x})/\sqrt{n-1}$, this conditional expectation is

$$\varphi(\bar{x}) = \int_{-\infty}^{c'} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz = \Phi(c') = \Phi\left[\frac{\sqrt{n}(c - \bar{x})}{\sqrt{n-1}}\right].$$

Thus the unique, unbiased, and minimum variance estimator of $\Phi(c - \theta)$ is, for every fixed constant c, given by $\varphi(\bar{X}) = \Phi[\sqrt{n}(c - \bar{X})/\sqrt{n-1}]$.

Remark. We should like to draw the attention of the reader to a rather important fact. This has to do with the adoption of a *principle*, such as the principle of unbiasedness and minimum variance. A principle is not a theorem; and seldom does a principle yield satisfactory results in all cases. So far, this principle has provided quite satisfactory results. To see that this is not always the case, let X have a Poisson distribution with parameter θ, $0 < \theta < \infty$. We may look upon X as a random sample of size 1 from this distribution. Thus X is a complete sufficient statistic for θ. We seek the estimator of $e^{-2\theta}$ that is unbiased and has minimum variance. Consider $Y = (-1)^X$. We have

$$E(Y) = E[(-1)^X] = \sum_{x=0}^{\infty} \frac{(-\theta)^x e^{-\theta}}{x!} = e^{-2\theta}.$$

Accordingly, $(-1)^X$ is the unbiased minimum variance estimator of $e^{-2\theta}$. Here this estimator leaves much to be desired. We are endeavoring to elicit some information about the number $e^{-2\theta}$, where $0 < e^{-2\theta} < 1$. Yet our point estimate is either -1 or $+1$, each of which is a very poor estimate of a number between zero and 1. We do not wish to leave the reader with the impression that an unbiased minimum variance estimator is *bad*. That is not the case at all. We merely wish to point out that if one tries hard enough, he can find instances where such a statistic is *not good*. Incidentally, the maximum likelihood estimator of $e^{-2\theta}$ is, in the case where the sample size equals 1, e^{-2X}, which is probably a much better estimator in practice than is the unbiased estimator $(-1)^X$.

EXERCISES

7.40. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that is $N(\theta, 1)$, $-\infty < \theta < \infty$. Find the unbiased minimum variance estimator of θ^2.

Hint: First determine $E(\bar{X}^2)$.

7.41. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that is $N(0, \theta)$. Then $Y = \sum X_i^2$ is a complete sufficient statistic for θ. Find the unbiased minimum variance estimator of θ^2.

7.42. In the notation of Example 2 of this section, is there an unbiased minimum variance estimator of $\Pr(-c \leq X \leq c)$? Here $c > 0$.

7.43. Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with parameter $\theta > 0$. Find the unbiased minimum variance estimator of $\Pr(X \leq 1) = (1 + \theta)e^{-\theta}$.

Hint: Let $u(x_1) = 1$, $x_1 \leq 1$, zero elsewhere, and find $E[u(X_1)|Y = y]$, where $Y = \sum_1^n X_i$. Make use of Example 2, Section 4.2.

7.44. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a Poisson distribution with parameter $\theta > 0$. From the Remark of this section, we know that $E[(-1)^{X_1}] = e^{-2\theta}$.

(a) Show that $E[(-1)^{X_1}|Y_1 = y_1] = (1 - 2/n)^{y_1}$, where $Y_1 = X_1 + X_2 + \cdots + X_n$.

Hint: First show that the conditional p.d.f. of $X_1, X_2, \ldots, X_{n-1}$, given $Y_1 = y_1$, is multinomial, and hence that of X_1 given $Y_1 = y_1$ is $b(y_1, 1/n)$.

(b) Show that the m.l.e. of $e^{-2\theta}$ is $e^{-2\bar{X}}$.

(c) Since $y_1 = n\bar{x}$, show that $(1 - 2/n)^{y_1}$ is approximately equal to $e^{-2\bar{x}}$ when n is large.

7.45. Let a random sample of size n be taken from a distribution that has the p.d.f. $f(x; \theta) = (1/\theta) \exp(-x/\theta) I_{(0,\infty)}(x)$. Find the m.l.e. and the unbiased minimum variance estimator of $\Pr(X \leq 2)$.

7.7 The Case of Several Parameters

In many of the interesting problems we encounter, the p.d.f. may not depend upon a single parameter θ, but perhaps upon two (or more) parameters, say θ_1 and θ_2, where $(\theta_1, \theta_2) \in \Omega$, a two-dimensional parameter space. We now define joint sufficient statistics for the parameters. For the moment we shall restrict ourselves to the case of two parameters.

Definition 4. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that has p.d.f. $f(x; \theta_1, \theta_2)$, where $(\theta_1, \theta_2) \in \Omega$. Let $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ and $Y_2 = u_2(X_1, X_2, \ldots, X_n)$ be two statistics whose joint p.d.f. is $g_{12}(y_1, y_2; \theta_1, \theta_2)$. The statistics Y_1 and Y_2 are called *joint sufficient statistics* for θ_1 and θ_2 if and only if

$$\frac{f(x_1; \theta_1, \theta_2) f(x_2; \theta_1, \theta_2) \cdots f(x_n; \theta_1, \theta_2)}{g_{12}[u_1(x_1, \ldots, x_n), u_2(x_1, \ldots, x_n); \theta_1, \theta_2]} = H(x_1, x_2, \ldots, x_n),$$

where $H(x_1, x_2, \ldots, x_n)$ does not depend upon θ_1 or θ_2.

As may be anticipated, the factorization theorem can be extended. In our notation it can be stated in the following manner. The statistics $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ and $Y_2 = u_2(X_1, X_2, \ldots, X_n)$ are joint suffi-

cient statistics for the parameters θ_1 and θ_2 if and only if we can find two nonnegative functions k_1 and k_2 such that

$$f(x_1; \theta_1, \theta_2)f(x_2; \theta_1, \theta_2)\cdots f(x_n; \theta_1, \theta_2)$$
$$= k_1[u_1(x_1, x_2, \ldots, x_n), u_2(x_1, x_2, \ldots, x_n); \theta_1, \theta_2]k_2(x_1, x_2, \ldots, x_n),$$

where the function $k_2(x_1, x_2, \ldots, x_n)$ does not depend upon both or either of θ_1 and θ_2.

Example 1. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution having p.d.f.

$$f(x; \theta_1, \theta_2) = \frac{1}{2\theta_2}, \qquad \theta_1 - \theta_2 < x < \theta_1 + \theta_2,$$
$$= 0 \qquad \text{elsewhere,}$$

where $-\infty < \theta_1 < \infty$, $0 < \theta_2 < \infty$. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics. The joint p.d.f. of Y_1 and Y_n is given by

$$g_{1n}(y_1, y_n; \theta_1, \theta_2) = \frac{n(n-1)}{(2\theta_2)^n}(y_n - y_1)^{n-2}, \qquad \theta_1 - \theta_2 < y_1 < y_n < \theta_1 + \theta_2,$$

and equals zero elsewhere. Accordingly, the joint p.d.f. of $X_1, X_2, \ldots, X_n$ can be written, for points of positive probability density,

$$\left(\frac{1}{2\theta_2}\right)^n = \frac{n(n-1)[\max(x_i) - \min(x_i)]^{n-2}}{(2\theta_2)^n} \times \left(\frac{1}{n(n-1)[\max(x_i) - \min(x_i)]^{n-2}}\right).$$

Since $\min(x_i) \leq x_j \leq \max(x_i)$, $j = 1, 2, \ldots, n$, the last factor does not depend upon the parameters. Either the definition or the factorization theorem assures us that Y_1 and Y_n are joint sufficient statistics for θ_1 and θ_2.

The extension of the notion of joint sufficient statistics for more than two parameters is a natural one. Suppose that a certain p.d.f. depends upon m parameters. Let a random sample of size n be taken from the distribution that has this p.d.f. and define m statistics. These m statistics are called joint sufficient statistics for the m parameters if and only if the ratio of the joint p.d.f. of the observations of the random sample and the joint p.d.f. of these m statistics does not depend upon the m parameters, whatever the fixed values of the m statistics. Again the factorization theorem is readily extended.

There is an extension of the Rao–Blackwell theorem that can be

adapted to joint sufficient statistics for several parameters, but that extension will not be included in this book. However, the concept of a complete family of probability density functions is generalized as follows: Let

$$\{f(v_1, v_2, \ldots, v_k; \theta_1, \theta_2, \ldots, \theta_m) : (\theta_1, \theta_2, \ldots, \theta_m) \in \Omega\}$$

denote a family of probability density functions of k random variables $V_1, V_2, \ldots, V_k$ that depends upon m parameters $(\theta_1, \theta_2, \ldots, \theta_m) \in \Omega$. Let $u(v_1, v_2, \ldots, v_k)$ be a function of $v_1, v_2, \ldots, v_k$ (but not a function of any or all of the parameters). If

$$E[u(V_1, V_2, \ldots, V_k)] = 0$$

for all $(\theta_1, \theta_2, \ldots, \theta_m) \in \Omega$ implies that $u(v_1, v_2, \ldots, v_k) = 0$ at all points $(v_1, v_2, \ldots, v_k)$, except on a set of points that has probability zero for all members of the family of probability density functions, we shall say that the family of probability density functions is a complete family.

The remainder of our treatment of the case of several parameters will be restricted to probability density functions that represent what we shall call regular cases of the exponential class. Let $X_1, X_2, \ldots, X_n$, $n > m$, denote a random sample from a distribution that depends on m parameters and has a p.d.f. of the form

$$f(x; \theta_1, \theta_2, \ldots, \theta_m) = \exp\left[\sum_{j=1}^{m} p_j(\theta_1, \theta_2, \ldots, \theta_m)K_j(x) + S(x) + q(\theta_1, \theta_2, \ldots, \theta_m)\right] \tag{1}$$

for $a < x < b$, and equals zero elsewhere.

A p.d.f. of the form (1) is said to be a member of the *exponential class* of probability density functions of the continuous type. If, in addition,

1. neither a nor b depends upon any or all of the parameters $\theta_1, \theta_2, \ldots, \theta_m$,
2. the $p_j(\theta_1, \theta_2, \ldots, \theta_m)$, $j = 1, 2, \ldots, m$, are nontrivial, functionally independent, continuous functions of θ_j, $\gamma_j < \theta_j < \delta_j$, $j = 1, 2, \ldots, m$,
3. the $K_j(x)$, $j = 1, 2, \ldots, m$, are continuous for $a < x < b$ and no one is a linear homogeneous function of the others,

4. $S(x)$ is a continuous function of x, $a < x < b$, we say that we have a *regular case* of the exponential class.

The joint p.d.f. of $X_1, X_2, \ldots, X_n$ is given, at points of positive probability density, by

$$\exp\left[\sum_{j=1}^{m} p_j(\theta_1, \ldots, \theta_m) \sum_{i=1}^{n} K_j(x_i) + \sum_{i=1}^{n} S(x_i) + nq(\theta_1, \ldots, \theta_m)\right]$$

$$= \exp\left[\sum_{j=1}^{m} p_j(\theta_1, \ldots, \theta_m) \sum_{i=1}^{n} K_j(x_i) + nq(\theta_1, \ldots, \theta_m)\right]$$

$$\times \exp\left[\sum_{i=1}^{n} S(x_i)\right].$$

In accordance with the factorization theorem, the statistics

$$Y_1 = \sum_{i=1}^{n} K_1(X_i), \qquad Y_2 = \sum_{i=1}^{n} K_2(X_i), \ldots, Y_m = \sum_{i=1}^{n} K_m(X_i)$$

are joint sufficient statistics for the m parameters $\theta_1, \theta_2, \ldots, \theta_m$. It is left as an exercise to prove that the joint p.d.f. of $Y_1, \ldots, Y_m$ is of the form

$$R(y_1, \ldots, y_m) \exp\left[\sum_{j=1}^{m} p_j(\theta_1, \ldots, \theta_m) y_j + nq(\theta_1, \ldots, \theta_m)\right] \quad (2)$$

at points of positive probability density. These points of positive probability density and the function $R(y_1, \ldots, y_m)$ do not depend upon any or all of the parameters $\theta_1, \theta_2, \ldots, \theta_m$. Moreover, in accordance with a theorem in analysis, it can be asserted that, in a regular case of the exponential class, the family of probability density functions of these joint sufficient statistics $Y_1, Y_2, \ldots, Y_m$ is complete when $n > m$. In accordance with a convention previously adopted, we shall refer to $Y_1, Y_2, \ldots, Y_m$ as *joint complete sufficient statistics for the parameters* $\theta_1, \theta_2, \ldots, \theta_m$.

Example 2. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that is $N(\theta_1, \theta_2)$, $-\infty < \theta_1 < \infty, 0 < \theta_2 < \infty$. Thus the p.d.f. $f(x; \theta_1, \theta_2)$ of the distribution may be written as

$$f(x; \theta_1, \theta_2) = \exp\left(\frac{-1}{2\theta_2} x^2 + \frac{\theta_1}{\theta_2} x - \frac{\theta_1^2}{2\theta_2} - \ln \sqrt{2\pi\theta_2}\right).$$

Therefore, we can take $K_1(x) = x^2$ and $K_2(x) = x$. Consequently, the statistics

$$Y_1 = \sum_1^n X_i^2 \qquad \text{and} \qquad Y_2 = \sum_1^n X_i$$

are joint complete sufficient statistics for θ_1 and θ_2. Since the relations

$$Z_1 = \frac{Y_2}{n} = \bar{X}, \qquad Z_2 = \frac{Y_1 - Y_2^2/n}{n-1} = \frac{\sum (X_i - \bar{X})^2}{n-1}$$

define a one-to-one transformation, Z_1 and Z_2 are also joint complete sufficient statistics for θ_1 and θ_2. Moreover,

$$E(Z_1) = \theta_1 \quad \text{and} \quad E(Z_2) = \theta_2.$$

From completeness, we have that Z_1 and Z_2 are the only functions of Y_1 and Y_2 that are unbiased estimators of θ_1 and θ_2, respectively.

A p.d.f.

$$f(x; \theta_1, \theta_2, \ldots, \theta_m) = \exp\left[\sum_{j=1}^m p_j(\theta_1, \theta_2, \ldots, \theta_m)K_j(x) + S(x) + q(\theta_1, \theta_2, \ldots, \theta_m)\right], \qquad x = a_1, a_2, a_3, \ldots,$$

zero elsewhere, is said to represent a regular case of the exponential class of probability density functions of the discrete type if

1. the set $\{x : x = a_1, a_2, \ldots\}$ does not depend upon any or all of the parameters $\theta_1, \theta_2, \ldots, \theta_m$,
2. the $p_j(\theta_1, \theta_2, \ldots, \theta_m)$, $j = 1, 2, \ldots, m$, are nontrivial, functionally independent, and continuous functions of θ_j, $\gamma_j < \theta_j < \delta_j$, $j = 1, 2, \ldots, m$,
3. the $K_j(x)$, $j = 1, 2, \ldots, m$, are nontrivial functions of x on the set $\{x : x = a_1, a_2, \ldots\}$ and no one is a linear function of the others.

Let $X_1, X_2, \ldots, X_n$ denote a random sample from a discrete-type distribution that represents a regular case of the exponential class. Then the statements made above in connection with the random variable of the continuous type are also valid here.

Not always do we sample from a distribution of one random variable X. We could, for instance, sample from a distribution of two random variables V and W with joint p.d.f. $f(v, w; \theta_1, \theta_2, \ldots, \theta_m)$. Recall that by a random sample $(V_1, W_1), (V_2, W_2), \ldots, (V_n, W_n)$ from a distribution of this sort, we mean that the joint p.d.f. of these $2n$ random variables is given by

$$f(v_1, w_1; \theta_1, \ldots, \theta_m)f(v_2, w_2; \theta_1, \ldots, \theta_m) \cdots f(v_n, w_n; \theta_1, \ldots, \theta_m).$$

In particular, suppose that the random sample is taken from a distribution that has the p.d.f. of V and W of the exponential class

$$f(v, w; \theta_1, \ldots, \theta_m) = \exp\left[\sum_{j=1}^{m} p_j(\theta_1, \ldots, \theta_m)K_j(v, w) + S(v, w) + q(\theta_1, \ldots, \theta_m)\right] \quad (3)$$

for $a < v < b$, $c < w < d$, and equals zero elsewhere, where a, b, c, d do not depend on the parameters and conditions similar to 1 to 4, p. 343, are imposed. Then the m statistics

$$Y_1 = \sum_{i=1}^{n} K_1(V_i, W_i), \ldots, \qquad Y_m = \sum_{i=1}^{n} K_m(V_i, W_i)$$

are joint complete sufficient statistics for the m parameters $\theta_1, \theta_2, \ldots, \theta_m$.

EXERCISES

7.46. Let $Y_1 < Y_2 < Y_3$ be the order statistics of a random sample of size 3 from the distribution with p.d.f.

$$f(x; \theta_1, \theta_2) = \frac{1}{\theta_2}\exp\left(-\frac{x - \theta_1}{\theta_2}\right),$$

$$\theta_1 < x < \infty, \quad -\infty < \theta_1 < \infty, \quad 0 < \theta_2 < \infty,$$

zero elsewhere. Find the joint p.d.f. of $Z_1 = Y_1$, $Z_2 = Y_2$, and $Z_3 = Y_1 + Y_2 + Y_3$. The corresponding transformation maps the space $\{(y_1, y_2, y_3): \theta_1 < y_1 < y_2 < y_3 < \infty\}$ onto the space

$$\{(z_1, z_2, z_3): \theta_1 < z_1 < z_2 < (z_3 - z_1)/2 < \infty\}$$

Show that Z_1 and Z_3 are joint sufficient statistics for θ_1 and θ_2.

7.47. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution that has a p.d.f. of form (1) of this section. Show that $Y_1 = \sum_{i=1}^{n} K_1(X_i)$, $\ldots, Y_m = \sum_{i=1}^{n} K_m(X_i)$ have a joint p.d.f. of form (2) of this section.

7.48. Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ denote a random sample of size n from a bivariate normal distribution with means μ_1 and μ_2, positive variances σ_1^2 and σ_2^2, and correlation coefficient ρ. Show that $\sum_1^n X_i$, $\sum_1^n Y_i$, $\sum_1^n X_i^2$, $\sum_1^n Y_i^2$, and $\sum_1^n X_i Y_i$ are joint complete sufficient statistics for the five

parameters. Are $\bar{X} = \sum_1^n X_i/n$, $\bar{Y} = \sum_1^n Y_i/n$, $S_1^2 = \sum_1^n (X_i - \bar{X})^2/n$, $S_2^2 = \sum_1^n (Y_i - \bar{Y})^2/n$, and $\sum_1^n (X_i - \bar{X})(Y_i - \bar{Y})/nS_1S_2$ also joint complete sufficient statistics for these parameters?

7.49. Let the p.d.f. $f(x; \theta_1, \theta_2)$ be of the form

$$\exp [p_1(\theta_1, \theta_2)K_1(x) + p_2(\theta_1, \theta_2)K_2(x) + S(x) + q(\theta_1, \theta_2)], \quad a < x < b,$$

zero elsewhere. Let $K_1'(x) = cK_2'(x)$. Show that $f(x; \theta_1, \theta_2)$ can be written in the form

$$\exp [p(\theta_1, \theta_2)K(x) + S(x) + q_1(\theta_1, \theta_2)], \quad a < x < b,$$

zero elsewhere. This is the reason why it is required that no one $K_j'(x)$ be a linear homogeneous function of the others, that is, so that the number of sufficient statistics equals the number of parameters.

7.50. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample $X_1, X_2, \ldots, X_n$ of size n from a distribution of the continuous type with p.d.f. $f(x)$. Show that the ratio of the joint p.d.f. of $X_1, X_2, \ldots, X_n$ and that of $Y_1 < Y_2 < \cdots < Y_n$ is equal to $1/n!$, which does not depend upon the underlying p.d.f. This suggests that $Y_1 < Y_2 < \cdots < Y_n$ are joint sufficient statistics for the unknown "parameter" f.

7.51. Let $X_1, X_2, \ldots, X_n$ be a random sample from the uniform distribution with p.d.f. $f(x; \theta_1, \theta_2) = 1/(2\theta_2)$, $\theta_1 - \theta_2 < x < \theta_1 + \theta_2$, where $-\infty < \theta_1 < \infty$ and $\theta_2 > 0$, and the p.d.f. is equal to zero elsewhere.

(a) Show that $Y_1 = \min(X_i)$ and $Y_n = \max(X_i)$, the joint sufficient statistics for θ_1 and θ_2, are complete.

(b) Find the unbiased minimum variance estimators of θ_1 and θ_2.

7.52. Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\theta_1, \theta_2)$.

(a) If the constant b is defined by the equation $\Pr(X \leq b) = 0.90$, find the m.l.e. and the unbiased minimum variance estimator of b.

(b) If c is a given constant, find the m.l.e. and the unbiased minimum variance estimator of $\Pr(X \leq c)$.

7.8 Minimal Sufficient and Ancillary Statistics

In the study of statistics, it is clear that we want to reduce the data contained in the entire sample as much as possible without losing relevant information about the important characteristics of the underlying distribution. That is, a large collection of numbers in the sample is not as meaningful as a few good summary statistics of those data. Sufficient statistics, if they exist, are valuable because we know

that the statistician with those summary measures is as well off as the statistician with the entire sample. Sometimes, however, there are several sets of joint sufficient statistics, and thus we would like to find the simplest one of these sets. For illustration, in a sense, the observations $X_1, X_2, \ldots, X_n, n > 2$, of a random sample from $N(\theta_1, \theta_2)$ could be thought of as joint sufficient statistics for θ_1 and θ_2. We know, however, that we can use $\bar{X}$ and S^2 as joint sufficient statistics for those parameters, which is a great simplification over using $X_1, X_2, \ldots, X_n$, particularly if n is large.

In most instances in this chapter, we have been able to find a single sufficient statistic for one parameter or two joint sufficient statistics for two parameters. Possibly the most complicated case considered so far is given in Exercise 7.48, in which we find five joint sufficient statistics for five parameters. Exercise 7.50 suggests the possibility of using the order statistics of a random sample for some completely unknown distribution of the continuous type.

What we would like to do is to change from one set of joint sufficient statistics to another, always reducing the number of statistics involved until we cannot go any further without losing the sufficiency of the resulting statistics. Those statistics that are there at the end of this process are called minimal sufficient statistics for the parameters. That is, *minimal sufficient statistics* are those that are sufficient for the parameters and are functions of every other set of sufficient statistics for those same parameters. Often, if there are k parameters, we can find k joint sufficient statistics that are minimal. In particular, if there is one parameter, we can often find a single sufficient statistic which is minimal. Most of the earlier examples that we have considered illustrate this point, but this is not always the case as shown by the following example.

Example 1. Let $X_1, X_2, \ldots, X_n$ be a random sample from the uniform distribution over the interval $(\theta - 1, \theta + 1)$ having p.d.f.

$$f(x; \theta) = (\tfrac{1}{2})I_{(\theta-1,\theta+1)}(x), \qquad \text{where } -\infty < \theta < \infty.$$

The joint p.d.f. of $X_1, X_2, \ldots, X_n$ equals the product of $(\frac{1}{2})^n$ and certain indicator functions, namely

$$(\tfrac{1}{2})^n \prod_{i=1}^{n} I_{(\theta-1,\theta+1)}(x_i) = (\tfrac{1}{2})^n \{I_{(\theta-1,\theta+1)}[\min(x_i)]\}\{I_{(\theta-1,\theta+1)}[\max(x_i)]\},$$

because $\theta - 1 < \min(x_i) \leq x_j \leq \max(x_i) < \theta + 1, j = 1, 2, \ldots, n$. Thus the order statistics $Y_1 = \min(X_i)$ and $Y_n = \max(X_i)$ are the sufficient statistics for θ. These two statistics actually are minimal for this one parameter, as

we cannot reduce the number of them to less than two and still have sufficiency.

There is an observation that helps us observe that almost all the sufficient statistics that we have studied thus far are minimal. We have noted that the m.l.e. $\hat{\theta}$ of θ is a function of one or more sufficient statistics, when the latter exist. Suppose that this m.l.e. $\hat{\theta}$ is also sufficient. Since this sufficient statistic $\hat{\theta}$ is a function of the other sufficient statistics, it must be minimal. For example, we have

1. The m.l.e. $\hat{\theta} = \bar{X}$ of θ in $N(\theta, \sigma^2)$, σ^2 known, is a minimal sufficient statistic for θ.
2. The m.l.e. $\hat{\theta} = \bar{X}$ of θ in a Poisson distribution with mean θ is a minimal sufficient statistic for θ.
3. The m.l.e. $\hat{\theta} = Y_n = \max(X_i)$ of θ in the uniform distribution over $(0, \theta)$ is a minimal sufficient statistic for θ.
4. The maximum likelihood estimators $\hat{\theta}_1 = \bar{X}$ and $\hat{\theta}_2 = S^2$ of θ_1 and θ_2 in $N(\theta_1, \theta_2)$ are joint minimal sufficient statistics for θ_1 and θ_2.

From these examples we see that the minimal sufficient statistics do not need to be unique, for any one-to-one transformation of them also provides minimal sufficient statistics. For illustration, in 4, the ΣX_i and ΣX_i^2 are also minimal sufficient statistics for θ_1 and θ_2.

Example 2. Consider the model given in Example 1. There we noted that $Y_1 = \min(X_i)$ and $Y_n = \max(X_i)$ are joint sufficient statistics. Also, we have

$$\theta - 1 < Y_1 < Y_n < \theta + 1$$

or, equivalently,

$$Y_n - 1 < \theta < Y_1 + 1.$$

Hence, to maximize the likelihood function so that it equals $(\frac{1}{2})^n$, θ can be any value between $Y_n - 1$ and $Y_1 + 1$. For example, many statisticians take the m.l.e. to be the mean of these two end points, namely

$$\hat{\theta} = \frac{Y_n - 1 + Y_1 + 1}{2} = \frac{Y_1 + Y_n}{2},$$

which is the midrange. We recognize, however, that this m.l.e. is not unique. Some might argue that since $\hat{\theta}$ is an m.l.e. of θ and since it is a function of the joint sufficient statistics, Y_1 and Y_n, for θ, it will be a minimal sufficient statistic. This is not the case at all, for $\hat{\theta}$ is not even sufficient. Note that the m.l.e. must itself be a sufficient statistic for the parameter before it can be considered the minimal sufficient statistic.

There is also a relationship between a minimal sufficient statistic and completeness that is explained more fully in the 1950 article by Lehmann and Scheffé. Let us say simply and without explanation that for the cases in this book, complete sufficient statistics are minimal sufficient statistics. The converse is not true, however, by noting that in Example 1 we have

$$E\left[\frac{Y_n - Y_1}{2} - \frac{n-1}{n+1}\right] = 0, \qquad \text{for all } \theta.$$

That is, there is a nonzero function of those minimal sufficient statistics, Y_1 and Y_n, whose expectation is zero for all θ.

There are other statistics that almost seem opposites of sufficient statistics. That is, while sufficient statistics contain all the information about the parameters, these other statistics, called *ancillary statistics*, have distributions free of the parameters and seemingly contain no information about those parameters. As an illustration, we know that the variance S^2 of a random sample from $N(\theta, 1)$ has a distribution that does not depend upon θ and hence is an ancillary statistic. Another example is the ratio $Z = X_1/(X_1 + X_2)$, where X_1, X_2 is a random sample from a gamma distribution with known parameter $\alpha > 0$ and unknown parameter $\beta = \theta$, because Z has a beta distribution that is free of θ. There are a great number of examples of ancillary statistics, and we provide some rules that make them rather easy to find with certain models.

First consider the situation in which there is a location parameter. That is, let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution that has a p.d.f. of the form $f(x - \theta)$, for every real θ; that is, θ is a location parameter. Let $Z = u(X_1, X_2, \ldots, X_n)$ be a statistic such that

$$u(x_1 + d, x_2 + d, \ldots, x_n + d) = u(x_1, x_2, \ldots, x_n),$$

for all real d. The one-to-one transformation defined by $W_i = X_i - \theta$, $i = 1, 2, \ldots, n$, requires that the joint p.d.f. of $W_1, W_2, \ldots, W_n$ be

$$f(w_1)f(w_2)\cdots f(w_n),$$

which does not depend upon θ. In addition, we have, because of the special functional nature of $u(x_1, x_2, \ldots, x_n)$, that

$$Z = u(W_1 + \theta, W_2 + \theta, \ldots, W_n + \theta) = u(W_1, W_2, \ldots, W_n)$$

is a function of $W_1, W_2, \ldots, W_n$ alone (not of θ). Hence Z must have

a distribution that does not depend upon θ because, for illustration, the m.g.f. of Z, namely

$$E(e^{tZ}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{tu(x_1, \ldots, x_n)} f(x_1 - \theta) \cdots f(x_n - \theta)\, dx_1 \cdots dx_n$$

$$= \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} e^{tu(w_1, \ldots, w_n)} f(w_1) \cdots f(w_n)\, dw_1 \cdots dw_n$$

is free of θ. We call $Z = u(X_1, X_2, \ldots, X_n)$ a *location-invariant statistic*. We immediately see that we can construct many examples of location-invariant statistics: the sample variance $= S^2$, the sample range $= Y_n - Y_1$, the mean deviation from the sample median $= (1/n)\,\Sigma|X_i - \text{median}\,(X_i)|$, $X_1 + X_2 - X_3 - X_4$, $X_1 + X_3 - 2X_2$, $(1/n)\,\Sigma\,[X_i - \min\,(X_i)]$, and so on.

We now consider a *scale-invariant statistic*. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution that has a p.d.f. of the form $(1/\theta)f(x/\theta)$, for all $\theta > 0$; that is, θ is a scale parameter. Say that $Z = u(X_1, X_2, \ldots, X_n)$ is a statistic such that

$$u(cx_1, cx_2, \ldots, cx_n) = u(x_1, x_2, \ldots, x_n)$$

for all $c > 0$. The one-to-one transformation defined by $W_i = X_i/\theta$, $i = 1, 2, \ldots, n$, requires the following: (1) that the joint p.d.f. of $W_1, W_2, \ldots, W_n$ be equal to

$$f(w_1)f(w_2) \cdots f(w_n),$$

and (2) that the statistic Z be equal to

$$Z = u(\theta W_1, \theta W_2, \ldots, \theta W_n) = u(W_1, W_2, \ldots, W_n).$$

Since neither the joint p.d.f. of $W_1, W_2, \ldots, W_n$ nor Z contain θ, the distribution of Z must not depend upon θ. There are also many examples of scale-invariant statistics like this: $Z: X_1/(X_1 + X_2)$, $X_1^2/\sum_1^n X_i^2$, $\min\,(X_i)/\max\,(X_i)$, and so on.

Finally, the location and the scale parameters can be combined in a p.d.f. of the form $(1/\theta_2)f[(x - \theta_1)/\theta_2]$, $-\infty < \theta_1 < \infty$, $0 < \theta_2 < \infty$. Through a one-to-one transformation defined by $W_i = (X_i - \theta_1)/\theta_2$, $i = 1, 2, \ldots, n$, it is easy to show that a statistic $Z = u(X_1, X_2, \ldots, X_n)$ such that

$$u(cx_1 + d, \ldots, cx_n + d) = u(x_1, \ldots, x_n)$$

for $-\infty < d < \infty$, $0 < c < \infty$, has a distribution that does not depend upon θ_1 and θ_2. Statistics like this $Z = u(X_1, X_2, \ldots, X_n)$ are *location-and-scale-invariant statistics*. Again there are many examples: $[\max(X_i) - \min(X_i)]/S$, $\sum_{i=1}^{n-1}(X_{i+1} - X_i)^2/S^2$, $(X_i - \bar{X})/S$, $|X_i - X_j|/S$, $i \neq j$, and so on.

Thus these location-invariant, scale-invariant, and location-and-scale-invariant statistics provide good illustrations, with the appropriate model for the p.d.f., of ancillary statistics. Since an ancillary statistic and a complete (minimal) sufficient statistic are such opposites, we might believe that there is, in some sense, no relationship between the two. This is true and in the next section we show that they are independent statistics.

EXERCISES

7.53. Let $X_1, X_2, \ldots, X_n$ be a random sample from each of the following distributions involving the parameter θ. In each case find the m.l.e. of θ and show that it is a sufficient statistic for θ and hence a minimal sufficient statistic.

(a) $b(1, \theta)$, where $0 \leq \theta \leq 1$.
(b) Poisson with mean $\theta > 0$.
(c) Gamma with $\alpha = 3$ and $\beta = \theta > 0$.
(d) $N(\theta, 1)$, where $-\infty < \theta < \infty$.
(e) $N(0, \theta)$, where $0 < \theta < \infty$.

7.54. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample of size n from the uniform distribution over the closed interval $[-\theta, \theta]$ having p.d.f. $f(x; \theta) = (1/2\theta)I_{[-\theta,\theta]}(x)$.

(a) Show that Y_1 and Y_n are joint sufficient statistics for θ.
(b) Argue that the m.l.e. of θ equals $\hat{\theta} = \max(-Y_1, Y_n)$.
(c) Demonstrate that the m.l.e. $\hat{\theta}$ is a sufficient statistic for θ and thus is a minimal sufficient statistic for θ.

7.55. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample of size n from a distribution with p.d.f.

$$f(x; \theta_1, \theta_2) = \left(\frac{1}{\theta_2}\right) e^{-(x-\theta_1)/\theta_2} I_{(\theta_1,\infty)}(x),$$

where $-\infty < \theta_1 < \infty$ and $0 < \theta_2 < \infty$. Find joint minimal sufficient statistics for θ_1 and θ_2.

7.56. With random samples from each of the distributions given in Exercises 7.53(d), 7.54, and 7.55, define at least two ancillary statistics that are different from the examples given in the text. These examples illustrate, respectively, location-invariant, scale-invariant, and location-and-scale-invariant statistics.

7.9 Sufficiency, Completeness, and Independence

We have noted that if we have a sufficient statistic Y_1 for a parameter θ, $\theta \in \Omega$, then $h(z|y_1)$, the conditional p.d.f. of another statistic Z, given $Y_1 = y_1$, does not depend upon θ. If, moreover, Y_1 and Z are independent, the p.d.f. $g_2(z)$ of Z is such that $g_2(z) = h(z|y_1)$, and hence $g_2(z)$ must not depend upon θ either. So the independence of a statistic Z and the sufficient statistic Y_1 for a parameter θ means that the distribution of Z does not depend upon $\theta \in \Omega$. That is, Z is an ancillary statistic.

It is interesting to investigate a converse of that property. Suppose that the distribution of an ancillary statistic Z does not depend upon θ; then, are Z and the sufficient statistic Y_1 for θ independent? To begin our search for the answer, we know that the joint p.d.f. of Y_1 and Z is $g_1(y_1; \theta)h(z|y_1)$, where $g_1(y_1; \theta)$ and $h(z|y_1)$ represent the marginal p.d.f. of Y_1 and the conditional p.d.f. of Z given $Y_1 = y_1$, respectively. Thus the marginal p.d.f. of Z is

$$\int_{-\infty}^{\infty} g_1(y_1; \theta)h(z|y_1)\,dy_1 = g_2(z),$$

which, by hypothesis, does not depend upon θ. Because

$$\int_{-\infty}^{\infty} g_2(z)g_1(y_1; \theta)\,dy_1 = g_2(z),$$

it follows, by taking the difference of the last two integrals, that

$$\int_{-\infty}^{\infty} [g_2(z) - h(z|y_1)]g_1(y_1; \theta)\,dy_1 = 0 \tag{1}$$

for all $\theta \in \Omega$. Since Y_1 is a sufficient statistic for θ, $h(z|y_1)$ does not depend upon θ. By assumption, $g_2(z)$ and hence $g_2(z) - h(z|y_1)$ do not depend upon θ. Now if the family $\{g_1(y_1; \theta) : \theta \in \Omega\}$ is complete, Equation (1) would require that

$$g_2(z) - h(z|y_1) = 0 \qquad \text{or} \qquad g_2(z) = h(z|y_1).$$

That is, the joint p.d.f. of Y_1 and Z must be equal to

$$g_1(y_1; \theta)h(z|y_1) = g_1(y_1; \theta)g_2(z).$$

Accordingly, Y_1 and Z are independent, and we have proved the following theorem, which was considered in special cases by Neyman and Hogg and proved in general by Basu.

Theorem 6. *Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution having a* p.d.f. *$f(x; \theta)$, $\theta \in \Omega$, where Ω is an interval set. Let $Y_1 = u_1(X_1, X_2, \ldots, X_n)$ be a sufficient statistic for θ, and let the family $\{g_1(y_1; \theta) : \theta \in \Omega\}$ of probability density functions of Y_1 be complete. Let $Z = u(X_1, X_2, \ldots, X_n)$ be any other statistic (not a function of Y_1 alone). If the distribution of Z does not depend upon θ, then Z is independent of the sufficient statistic Y_1.*

In the discussion above, it is interesting to observe that if Y_1 is a sufficient statistic for θ, then the independence of Y_1 and Z implies that the distribution of Z does not depend upon θ whether $\{g_1(y_1; \theta) : \theta \in \Omega\}$ is or is not complete. However, in the converse, to prove the independence from the fact that $g_2(z)$ does not depend upon θ, we definitely need the completeness. Accordingly, if we are dealing with situations in which we know that the family $\{g(y_1; \theta) : \theta \in \Omega\}$ is complete (such as a regular case of the exponential class), we can say that the statistic Z is independent of the sufficient statistic Y_1 if, and only if, the distribution of Z does not depend upon θ (i.e., Z is an ancillary statistic).

It should be remarked that the theorem (including the special formulation of it for regular cases of the exponential class) extends immediately to probability density functions that involve m parameters for which there exist m joint sufficient statistics. For example, let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution having the p.d.f. $f(x; \theta_1, \theta_2)$ that represents a regular case of the exponential class such that there are two joint complete sufficient statistics for θ_1 and θ_2. Then any other statistic $Z = u(X_1, X_2, \ldots, X_n)$ is independent of the joint complete sufficient statistics if and only if the distribution of Z does not depend upon θ_1 or θ_2.

We give an example of the theorem that provides an alternative proof of the independence of $\bar{X}$ and S^2, the mean and the variance of a random sample of size n from a distribution that is $N(\mu, \sigma^2)$. This proof is presented as if we did not know that nS^2/σ^2 is $\chi^2(n-1)$ because that fact and the independence were established in the same argument (see Section 4.8).

Example 1. Let $X_1, X_2, \ldots, X_n$ denote a random sample of size n from a distribution that is $N(\mu, \sigma^2)$. We know that the mean $\bar{X}$ of the sample is, for every known σ^2, a complete sufficient statistic for the parameter μ, $-\infty < \mu < \infty$. Consider the statistic

$$S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2,$$

which is location-invariant. Thus S^2 must have a distribution that does not depend upon μ; and hence, by the theorem, S^2 and $\bar{X}$, the complete sufficient statistic for μ, are independent.

Example 2. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the distribution having p.d.f.

$$f(x; \theta) = e^{-(x-\theta)}, \qquad \theta < x < \infty, \quad -\infty < \theta < \infty.$$
$$= 0 \qquad \text{elsewhere.}$$

Here the p.d.f. is of the form $f(x - \theta)$, where $f(x) = e^{-x}$, $0 < x < \infty$, zero elsewhere. Moreover, we know (Exercise 7.27) that the first order statistic $Y_1 = \min(X_i)$ is a complete sufficient statistic for θ. Hence Y_1 must be independent of each location-invariant statistic $u(X_1, X_2, \ldots, X_n)$, enjoying the property that

$$u(x_1 + d, x_2 + d, \ldots, x_n + d) = u(x_1, x_2, \ldots, x_n)$$

for all real d. Illustrations of such statistics are S^2, the sample range, and

$$\frac{1}{n}\sum_{i=1}^{n}[X_i - \min(X_i)].$$

Example 3. Let X_1, X_2 denote a random sample of size $n = 2$ from a distribution with p.d.f.

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

The p.d.f. is of the form $(1/\theta)f(x/\theta)$, where $f(x) = e^{-x}$, $0 < x < \infty$, zero elsewhere. We know (Section 7.5) that $Y_1 = X_1 + X_2$ is a complete sufficient statistic for θ. Hence Y_1 is independent of every scale-invariant statistic $u(X_1, X_2)$ with the property $u(cx_1, cx_2) = u(x_1, x_2)$. Illustrations of these are X_1/X_2 and $X_1/(X_1 + X_2)$, statistics that have F and beta distributions, respectively.

Example 4. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that is $N(\theta_1, \theta_2)$, $-\infty < \theta_1 < \infty$, $0 < \theta_2 < \infty$. In Example 2,

Section 7.7, it was proved that the mean $\bar{X}$ and the variance S^2 of the sample are joint complete sufficient statistics for θ_1 and θ_2. Consider the statistic

$$Z = \frac{\sum_{1}^{n-1} (X_{i+1} - X_i)^2}{\sum_{1}^{n} (X_i - \bar{X})^2} = u(X_1, X_2, \ldots, X_n),$$

which satisfies the property that $u(cx_1 + d, \ldots, cx_n + d) = u(x_1, \ldots, x_n)$. That is, the ancillary statistic Z is independent of both $\bar{X}$ and S^2.

Let $N(\theta_1, \theta_3)$ and $N(\theta_2, \theta_4)$ denote two normal distributions. Recall that in Example 2, Section 6.5, a statistic, which was denoted by T, was used to test the hypothesis that $\theta_1 = \theta_2$, provided that the unknown variances θ_3 and θ_4 were equal. The hypothesis that $\theta_1 = \theta_2$ is rejected if the computed $|T| \geq c$, where the constant c is selected so that $\alpha_2 = \Pr(|T| \geq c; \theta_1 = \theta_2, \theta_3 = \theta_4)$ is the assigned significance level of the test. We shall show that, if $\theta_3 = \theta_4$, F of Exercise 6.52 and T are independent. Among other things, this means that if these two tests based on F and T, respectively, are performed sequentially, with significance levels α_1 and α_2, the probability of accepting both these hypotheses, when they are true, is $(1 - \alpha_1)(1 - \alpha_2)$. Thus the significance level of this joint test is $\alpha = 1 - (1 - \alpha_1)(1 - \alpha_2)$.

The independence of F and T, when $\theta_3 = \theta_4$, can be established by an appeal to sufficiency and completeness. The three statistics $\bar{X}$, $\bar{Y}$, and $\sum_{1}^{n} (X_i - \bar{X})^2 + \sum_{1}^{m} (Y_i - \bar{Y})^2$ are joint complete sufficient statistics for the three parameters θ_1, θ_2, and $\theta_3 = \theta_4$. Obviously, the distribution of F does not depend upon θ_1, θ_2, or $\theta_3 = \theta_4$, and hence F is independent of the three joint complete sufficient statistics. However, T is a function of these three joint complete sufficient statistics alone, and, accordingly, T is independent of F. It is important to note that these two statistics are independent whether $\theta_1 = \theta_2$ or $\theta_1 \neq \theta_2$. This permits us to calculate probabilities other than the significance level of the test. For example, if $\theta_3 = \theta_4$ and $\theta_1 \neq \theta_2$, then

$$\Pr(c_1 < F < c_2, |T| \geq c) = \Pr(c_1 < F < c_2) \Pr(|T| \geq c).$$

The second factor in the right-hand member is evaluated by using the probabilities for what is called a noncentral t-distribution. Of course, if $\theta_3 = \theta_4$ and the difference $\theta_1 - \theta_2$ is large, we would want the preceding probability to be close to 1 because the event $\{c_1 < F < c_2, |T| \geq c\}$ leads to a correct decision, namely accept $\theta_3 = \theta_4$ and reject $\theta_1 = \theta_2$.

In this section we have given several examples in which the complete sufficient statistics are independent of ancillary statistics. Thus, in those cases, the ancillary statistics provide no information about the parameters. However, if the sufficient statistics are not complete, the ancillary statistics could provide some information as the following example demonstrates.

Example 5. We refer back to Examples 1 and 2 of Section 7.8. There the first and nth order statistics, Y_1 and Y_n, were minimal sufficient statistics for θ, where the sample arose from an underlying distribution having p.d.f. $(\frac{1}{2})I_{(\theta-1,\theta+1)}(x)$. Often $T_1 = (Y_1 + Y_n)/2$ is used as an estimator of θ as it is a function of those sufficient statistics which is unbiased. Let us find a relationship between T_1 and the ancillary statistic $T_2 = Y_n - Y_1$.

The joint p.d.f. of Y_1 and Y_n is

$$g(y_1, y_n; \theta) = n(n-1)(y_n - y_1)^{n-2}/2^n, \qquad \theta - 1 < y_1 < y_n < \theta + 1,$$

zero elsewhere. Accordingly, the joint p.d.f. of T_1 and T_2 is, since the absolute value of the Jacobian equals 1,

$$h(t_1, t_2; \theta) = n(n-1)t_2^{n-2}/2^n, \qquad \theta - 1 + \frac{t_2}{2} < t_1 < \theta + 1 - \frac{t_2}{2}, \quad 0 < t_2 < 2,$$

zero elsewhere. Thus the p.d.f. of T_2 is

$$h_2(t_2; \theta) = n(n-1)t_2^{n-2}(2 - t_2)/2^n, \qquad 0 < t_2 < 2,$$

zero elsewhere, which of course is free of θ as T_2 is an ancillary statistic. Thus the conditional p.d.f. of T_1, given $T_2 = t_2$, is

$$h_{1|2}(t_1|t_2; \theta) = \frac{1}{2 - t_2}, \qquad \theta - 1 + \frac{t_2}{2} < t_1 < \theta + 1 - \frac{t_2}{2}, \quad 0 < t_2 < 2,$$

zero elsewhere. Note that this is uniform on the interval $(\theta - 1 + t_2/2, \theta + 1 - t_2/2)$; so the conditional mean and variance of T_1 are, respectively,

$$E(T_1|t_2) = \theta \qquad \text{and} \qquad \text{var}\,(T_1|t_2) = \frac{(2 - t_2)^2}{12}.$$

That is, given $T_2 = t_2$, we know something about the conditional variance of T_1. In particular, if that observed value of T_2 is large (close to 2), that variance is small and we can place more reliance on the estimator T_1. On the other hand, a small value of t_2 means that we have less confidence in T_1 as an estimator of θ. It is extremely interesting to note that this conditional variance does not depend upon the sample size n but only on the given value of $T_2 = t_2$. Of course, as the sample size increases, T_2 tends to become larger and, in those cases, T_1 has smaller conditional variance.

While Example 5 is a special one demonstrating mathematically that an ancillary statistic can provide some help in point estimation, this does actually happen in practice too. For illustration, we know that if the sample size is large enough, then

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n-1}}$$

has an approximate standard normal distribution. Of course, if the sample arises from a normal distribution, $\bar{X}$ and S are independent and T has a t-distribution with $n-1$ degrees of freedom. Even if the sample arises from a symmetric distribution, $\bar{X}$ and S are uncorrelated and T has an approximate t-distribution and certainly an approximate standard normal distribution with sample sizes around 30 or 40. On the other hand, if the sample arises from a highly skewed distribution (say to the right), then $\bar{X}$ and S are highly correlated and the probability $\Pr(-1.96 < T < 1.96)$ is not necessarily close to 0.95 unless the sample size is extremely large (certainly much greater than 30). Intuitively, one can understand why this correlation exists if the underlying distribution is highly skewed to the right. While S has a distribution free of μ (and hence is an ancillary), a large value of S implies a large value of $\bar{X}$, since the underlying p.d.f. is like the one depicted in Figure 7.1. Of course, a small value of $\bar{X}$ (say less than the mode) requires a relatively small value of S. This means that unless n is extremely large, it is risky to say that

$$\bar{x} - \frac{1.96s}{\sqrt{n-1}}, \qquad \bar{x} + \frac{1.96s}{\sqrt{n-1}}$$

provides an approximate 95 percent confidence interval with data from a very skewed distribution. As a matter of fact, the authors have seen situations in which this confidence coefficient is closer to 70 percent, rather than 95 percent, with sample sizes of 30 to 40.

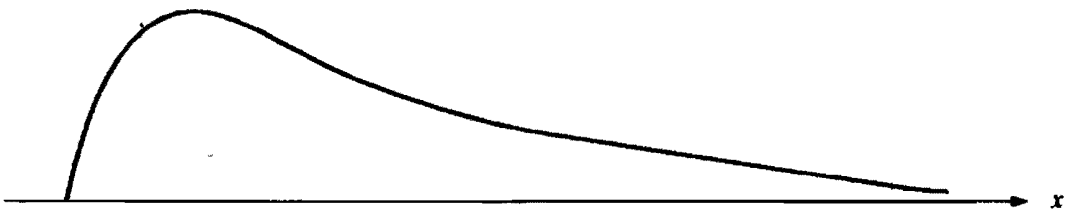

FIGURE 7.1

EXERCISES

7.57. Let $Y_1 < Y_2 < Y_3 < Y_4$ denote the order statistics of a random sample of size $n = 4$ from a distribution having p.d.f. $f(x; \theta) = 1/\theta$, $0 < x < \theta$, zero elsewhere, where $0 < \theta < \infty$. Argue that the complete sufficient statistic Y_4 for θ is independent of each of the statistics Y_1/Y_4 and $(Y_1 + Y_2)/(Y_3 + Y_4)$.

Hint: Show that the p.d.f. is of the form $(1/\theta)f(x/\theta)$, where $f(x) = 1$, $0 < x < 1$, zero elsewhere.

7.58. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample from the normal distribution $N(\theta, \sigma^2)$, $-\infty < \theta < \infty$. Show that the distribution of $Z = Y_n - \bar{Y}$ does not depend upon θ. Thus $\bar{Y} = \sum_1^n Y_i/n$, a complete sufficient statistic for θ, is independent of Z.

7.59. Let $X_1, X_2, \ldots, X_n$ be a random sample from the normal distribution $N(\theta, \sigma^2)$, $-\infty < \theta < \infty$. Prove that a necessary and sufficient condition that the statistics $Z = \sum_1^n a_i X_i$ and $Y = \sum_1^n X_i$, a complete sufficient statistic for θ, be independent is that $\sum_1^n a_i = 0$.

7.60. Let X and Y be random variables such that $E(X^k)$ and $E(Y^k) \neq 0$ exist for $k = 1, 2, 3, \ldots$. If the ratio X/Y and its denominator Y are independent, prove that $E[(X/Y)^k] = E(X^k)/E(Y^k)$, $k = 1, 2, 3, \ldots$.

Hint: Write $E(X^k) = E[Y^k(X/Y)^k]$.

7.61. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample of size n from a distribution that has p.d.f. $f(x; \theta) = (1/\theta)e^{-x/\theta}$, $0 < x < \infty$, $0 < \theta < \infty$, zero elsewhere. Show that the ratio $R = nY_1 \Big/ \sum_1^n Y_i$ and its denominator (a complete sufficient statistic for θ) are independent. Use the result of the preceding exercise to determine $E(R^k)$, $k = 1, 2, 3, \ldots$.

7.62. Let $X_1, X_2, \ldots, X_5$ be a random sample of size 5 from the distribution that has p.d.f. $f(x) = e^{-x}$, $0 < x < \infty$, zero elsewhere. Show that $(X_1 + X_2)/(X_1 + X_2 + \cdots + X_5)$ and its denominator are independent.

Hint: The p.d.f. $f(x)$ is a member of $\{f(x; \theta) : 0 < \theta < \infty\}$, where $f(x; \theta) = (1/\theta)e^{-x/\theta}$, $0 < x < \infty$, zero elsewhere.

7.63. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample from the normal distribution $N(\theta_1, \theta_2)$, $-\infty < \theta_1 < \infty$, $0 < \theta_2 < \infty$. Show that the joint complete sufficient statistics $\bar{X} = \bar{Y}$ and S^2 for θ_1 and θ_2 are independent of each of $(Y_n - \bar{Y})/S$ and $(Y_n - Y_1)/S$.

7.64. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample from a distribution with the p.d.f.

$$f(x; \theta_1, \theta_2) = \frac{1}{\theta_2} \exp\left(-\frac{x - \theta_1}{\theta_2}\right),$$

$\theta_1 < x < \infty$, zero elsewhere, where $-\infty < \theta_1 < \infty, 0 < \theta_2 < \infty$. Show that the joint complete sufficient statistics Y_1 and $\bar{X} = \bar{Y}$ for θ_1 and θ_2 are independent of $(Y_2 - Y_1) \Big/ \sum_1^n (Y_i - Y_1)$.

7.65. Let $X_1, X_2, \ldots, X_5$ be a random sample of size $n = 5$ from the normal distribution $N(0, \theta)$.

(a) Argue that the ratio $R = (X_1^2 + X_2^2)/(X_1^2 + \cdots + X_5^2)$ and its denominator $(X_1^2 + \cdots + X_5^2)$ are independent.

(b) Does $5R/2$ have an F-distribution with 2 and 5 degrees of freedom? Explain your answer.

(c) Compute $E(R)$ using Exercise 7.60.

7.66. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample of size n from a distribution having p.d.f.

$$f(x; \theta) = (1/\theta) \exp\left(\frac{-x}{\theta}\right), \qquad 0 < x < \infty,$$

and equal zero elsewhere, where $0 < \theta < \infty$. Show that $W = \sum_1^n Y_i$ and $Z = nY_1 \Big/ \sum_1^n Y_i$ are independent. Find $E(Z^k), k = 1, 2, 3, \ldots$ using the result of Exercise 7.60. What is the distribution of Z?

7.67. Referring to Example 5 of this section, determine c so that

$$\Pr(-c < T_1 - \theta < c | T_2 = t_2) = 0.95.$$

Use this result to find a 95 percent confidence interval for θ, given $T_2 = t_2$; and note how its length is smaller when the range t_2 is larger.

ADDITIONAL EXERCISES

7.68. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with p.d.f. $f(x; \theta) = \theta e^{-\theta x}$, $0 < x < \infty$, zero elsewhere where $0 < \theta$.

(a) What is the complete sufficient statistic, say Y, for θ?

(b) What function of Y is an unbiased estimator of θ?

7.69. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample of size n from a distribution with p.d.f. $f(x; \theta) = 1/\theta$, $0 < x < \theta$, zero

elsewhere. The statistic Y_n is a complete sufficient statistic for θ and it has p.d.f.

$$g(y_n; \theta) = \frac{ny_n^{n-1}}{\theta^n}, \qquad 0 < y_n < \theta,$$

and zero elsewhere.

(a) Find the distribution function $H_n(z; \theta)$ of $Z = n(\theta - Y_n)$.

(b) Find the $\lim_{n\to\infty} H_n(z; \theta)$ and thus the limiting distribution of Z.

7.70. Let $X_1, \ldots, X_n$; $Y_1, \ldots, Y_n$; $Z_1, \ldots, Z_n$ be respective independent random samples from three normal distributions $N(\mu_1 = \alpha + \beta, \sigma^2)$ $N(\mu_2 = \beta + \gamma, \sigma^2)$, $N(\mu_3 = \alpha + \gamma, \sigma^2)$. Find a point estimator for β that is based on $\bar{X}$, $\bar{Y}$, $\bar{Z}$. Is this estimator unique? Why? If σ^2 is unknown, explain how to find a confidence interval for β.

7.71. Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with mean θ. Find the conditional expectation $E(X_1 + 2X_2 + 3X_3 | \sum_1^n X_i)$.

7.72. Let $X_1, X_2, \ldots, X_n$ be a random sample of size n from the normal distribution $N(\theta, 1)$. Find the unbiased minimum variance estimator of θ^2.

7.73. Let $X_1, X_2, \ldots, X_n$ be a random sample from a Poisson distribution with mean θ. Find the unbiased minimum variance estimator of θ^2.

7.74. We consider a random sample $X_1, X_2, \ldots, X_n$ from a distribution with p.d.f. $f(x; \theta) = (1/\theta) \exp(-x/\theta)$, $0 < x < \infty$, zero elsewhere, where $0 < \theta$. Possibly, in a life testing situation, however, we only observe the first r order statistics, $Y_1 < Y_2 < \cdots < Y_r$.

(a) Record the joint p.d.f. of these order statistics and denote it by $L(\theta)$.

(b) Under these conditions, find the m.l.e., $\hat{\theta}$, by maximizing $L(\theta)$.

(c) Find the m.g.f. and p.d.f. of $\hat{\theta}$.

(d) With a slight extension of the definition of sufficiency, is $\hat{\theta}$ a sufficient statistic?

(e) Find the unbiased minimum variance estimator for θ.

(f) Show that $Y_1/\hat{\theta}$ and $\hat{\theta}$ are independent.

7.75. Let us repeat Bernoulli trials with parameter θ until k successes occur. If Y is the number of trials needed:

(a) Show that the p.d.f. of Y is $g(y; \theta) = \binom{y-1}{k-1}\theta^k(1-\theta)^{y-k}$, $y = k, k+1, \ldots$, zero elsewhere, where $0 \leq \theta \leq 1$.

(b) Prove that this family of probability density functions is complete.

(c) Demonstrate that $E[(k-1)/(Y-1)] = \theta$.

(d) Is it possible to find another statistic, which is a function of Y alone, that is unbiased? Why?

7.76. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with p.d.f. $f(x; \theta) = \theta^x(1-\theta)$, $x = 0, 1, 2, \ldots$, zero elsewhere, where $0 \leq \theta \leq 1$.

(a) Find the m.l.e., $\hat{\theta}$, of θ.

(b) Show that $\sum_{i=1}^{n} X_i$ is a complete sufficient statistic for θ.

(c) Determine the unbiased minimum variance estimator of θ.

7.77. If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution with p.d.f. $f(x; \theta) = \frac{1}{2}\theta^3 x^2 e^{-\theta x}$, $0 < x < \infty$, zero elsewhere, where $0 < \theta < \infty$:

(a) Find the m.l.e., $\hat{\theta}$, of θ. Is $\hat{\theta}$ unbiased?

Hint: First find the p.d.f. of $Y = \sum_{i=1}^{n} X_i$ and then compute $E(\hat{\theta})$.

(b) Argue that Y is a complete sufficient statistic for θ.

(c) Find the unbiased minimum variance estimator of θ.

(d) Show that X_1/Y and Y are independent.

(e) What is the distribution of X_1/Y?

CHAPTER 8

More About Estimation

8.1 Bayesian Estimation

In Chapter 6 we introduced point and interval estimation for various parameters. In Chapter 7 we observed how such inferences should be based upon sufficient statistics for the parameters if they exist. In this chapter we introduce other concepts related to estimation and begin this by considering *Bayesian estimates*, which are also based upon sufficient statistics if the latter exist.

In introducing the interesting and sometimes controversial Bayesian method of estimation, the student should constantly keep in mind that making statistical inferences from the data does not strictly follow a mathematical approach. Clearly, up to now, we have had to construct models before we have been able to make such inferences. These models are subjective, and the resulting inference depends greatly on the model selected. For illustration, two statisticians could very well select different models for exactly the same situation and make different inferences with exactly the same data. Most statisticians would use some type of model diagnostics to see if