CHAPTER 6

Introduction to Statistical Inference

6.1 Point Estimation

The first five chapters of this book deal with certain concepts and problems of probability theory. Throughout we have carefully distinguished between a sample space $\mathscr{C}$ of outcomes and the space $\mathscr{A}$ of one or more random variables defined on $\mathscr{C}$. With this chapter we begin a study of some problems in statistics and here we are more interested in the number (or numbers) by which an outcome is represented than we are in the outcome itself. Accordingly, we shall adopt a frequently used convention. We shall refer to a random variable X as the outcome of a random experiment and we shall refer to the space of X as the sample space. Were it not so awkward, we would call X the numerical outcome. Once the experiment has been performed and it is found that $X = x$, we shall call x the experimental value of X for that performance of the experiment.

This convenient terminology can be used to advantage in more general situations. To illustrate this, let a random experiment be repeated n independent times and under identical conditions. Then the random variables $X_1, X_2, \ldots, X_n$ (each of which assigns a numerical value to an outcome) constitute (Section 4.1) the observations of a random sample. If we are more concerned with the numerical representations of the outcomes than with the outcomes themselves, it seems natural to refer to $X_1, X_2, \ldots, X_n$ as the outcomes. And what more appropriate name can we give to the space of a random sample than the sample space? Once the experiment has been performed the indicated number of times and it is found that $X_1 = x_1$, $X_2 = x_2, \ldots, X_n = x_n$, we shall refer to $x_1, x_2, \ldots, x_n$ as the experimental values of $X_1, X_2, \ldots, X_n$ or as the sample data.

We shall use the terminology of the two preceding paragraphs, and in this section we shall give some examples of *statistical inference*. These examples will be built around the notion of a *point estimate* of an unknown parameter in a p.d.f.

Let a random variable X have a p.d.f. that is of known functional form but in which the p.d.f. depends upon an unknown parameter θ that may have any value in a set Ω. This will be denoted by writing the p.d.f. in the form $f(x; \theta)$, $\theta \in \Omega$. The set Ω will be called the *parameter space*. Thus we are confronted, not with one distribution of probability, but with a *family* of distributions. To each value of θ, $\theta \in \Omega$, there corresponds one member of the family. A family of probability density functions will be denoted by the symbol $\{f(x; \theta) : \theta \in \Omega\}$. Any member of this family of probability density functions will be denoted by the symbol $f(x; \theta)$, $\theta \in \Omega$. We shall continue to use the special symbols that have been adopted for the normal, the chi-square, and the binomial distributions. We may, for instance, have the family $\{N(\theta, 1) : \theta \in \Omega\}$, where Ω is the set $-\infty < \theta < \infty$. One member of this family of distributions is the distribution that is $N(0, 1)$. Any arbitrary member is $N(\theta, 1)$, $-\infty < \theta < \infty$.

Let us consider a family of probability density functions $\{f(x; \theta) : \theta \in \Omega\}$. It may be that the experimenter needs to select precisely *one* member of the family as being the p.d.f. of his random variable. That is, he needs a *point* estimate of θ. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that has a p.d.f. which is one member (but which member we do not know) of the family $\{f(x; \theta) : \theta \in \Omega\}$ of probability density functions. That is, our sample

arises from a distribution that has the p.d.f. $f(x;\theta):\theta\in\Omega$. Our problem is that of defining a statistic $Y_1 = u_1(X_1, X_2, \ldots, X_n)$, so that if $x_1, x_2, \ldots, x_n$ are the observed experimental values of $X_1, X_2, \ldots, X_n$, then the number $y_1 = u_1(x_1, x_2, \ldots, x_n)$ will be a good point estimate of θ.

The following illustration should help motivate one principle that is often used in finding point estimates.

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

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

where $0 \le \theta \le 1$. The probability that $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$ is the joint p.d.f.

$$\theta^{x_1}(1-\theta)^{1-x_1}\theta^{x_2}(1-\theta)^{1-x_2}\cdots\theta^{x_n}(1-\theta)^{1-x_n} = \theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i},$$

where x_i equals zero or 1, $i = 1, 2, \ldots, n$. This probability, which is the joint p.d.f. of $X_1, X_2, \ldots, X_n$, may be regarded as a function of θ and, when so regarded, is denoted by $L(\theta)$ and called the *likelihood function*. That is,

$$L(\theta) = \theta^{\Sigma x_i}(1-\theta)^{n-\Sigma x_i}, \qquad 0 \le \theta \le 1.$$

We might ask what value of θ would maximize the probability $L(\theta)$ of obtaining this particular observed sample $x_1, x_2, \ldots, x_n$. Certainly, this maximizing value of θ would seemingly be a good estimate of θ because it would provide the largest probability of this particular sample. Since the likelihood function $L(\theta)$ and its logarithm, $\ln L(\theta)$, are maximized for the same value θ, either $L(\theta)$ or $\ln L(\theta)$ can be used. Here

$$\ln L(\theta) = \left(\sum_1^n x_i\right)\ln\theta + \left(n - \sum_1^n x_i\right)\ln(1-\theta);$$

so we have

$$\frac{d\ln L(\theta)}{d\theta} = \frac{\sum x_i}{\theta} - \frac{n - \sum x_i}{1-\theta} = 0,$$

provided that θ is not equal to zero or 1. This is equivalent to the equation

$$(1-\theta)\sum_1^n x_i = \theta\left(n - \sum_1^n x_i\right),$$

whose solution for θ is $\sum_1^n x_i/n$. That $\sum_1^n x_i/n$ actually maximizes $L(\theta)$ and $\ln L(\theta)$ can be easily checked, even in the cases in which all of $x_1, x_2, \ldots, x_n$

equal zero together or 1 together. That is, $\sum_1^n x_i/n$ is the value of θ that maximizes $L(\theta)$. The corresponding statistic,

$$\hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} X_i = \bar{X},$$

is called the *maximum likelihood estimator* of θ. The observed value of $\hat{\theta}$, namely $\sum_1^n x_i/n$, is called the *maximum likelihood estimate* of θ. For a simple example, suppose that $n = 3$, and $x_1 = 1, x_2 = 0, x_3 = 1$, then $L(\theta) = \theta^2(1 - \theta)$ and the observed $\hat{\theta} = \frac{2}{3}$ is the maximum likelihood estimate of θ.

The principle of the *method of maximum likelihood* can now be formulated easily. Consider a random sample $X_1, X_2, \ldots, X_n$ from a distribution having p.d.f. $f(x; \theta)$, $\theta \in \Omega$. The joint p.d.f. of $X_1, X_2, \ldots, X_n$ is $f(x_1; \theta)f(x_2; \theta) \cdots f(x_n; \theta)$. This joint p.d.f. may be regarded as a function of θ. When so regarded, it is called the likelihood function L of the random sample, and we write

$$L(\theta; x_1, x_2, \ldots, x_n) = f(x_1; \theta)f(x_2; \theta) \cdots f(x_n; \theta), \qquad \theta \in \Omega.$$

Suppose that we can find a nontrivial function of $x_1, x_2, \ldots, x_n$, say $u(x_1, x_2, \ldots, x_n)$, such that, when θ is replaced by $u(x_1, x_2, \ldots, x_n)$, the likelihood function L is maximized. That is, $L[u(x_1, x_2, \ldots, x_n); x_1, x_2, \ldots, x_n]$ is at least as great as $L(\theta; x_1, x_2, \ldots, x_n)$ for every $\theta \in \Omega$. Then the statistic $u(X_1, X_2, \ldots, X_n)$ will be called a *maximum likelihood estimator* (hereafter abbreviated m.l.e.) of θ and will be denoted by the symbol $\hat{\theta} = u(X_1, X_2, \ldots, X_n)$. We remark that in many instances there will be a unique m.l.e. $\hat{\theta}$ of a parameter θ, and often it may be obtained by the process of differentiation.

Example 2. Let $X_1, X_2, \ldots, X_n$ be a random sample from the normal distribution $N(\theta, 1)$, $-\infty < \theta < \infty$. Here

$$L(\theta; x_1, x_2, \ldots, x_n) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left[-\sum_1^n \frac{(x_i - \theta)^2}{2}\right].$$

This function L can be maximized by setting the first derivative of L, with respect to θ, equal to zero and solving the resulting equation for θ. We note, however, that each of the functions L and $\ln L$ is maximized at the same value of θ. So it may be easier to solve

$$\frac{d \ln L(\theta; x_1, x_2, \ldots, x_n)}{d\theta} = 0.$$

For this example,

$$\frac{d \ln L(\theta; x_1, x_2, \ldots, x_n)}{d\theta} = \sum_1^n (x_i - \theta).$$

If this derivative is equated to zero, the solution for the parameter θ is $u(x_1, x_2, \ldots, x_n) = \sum_1^n x_i/n$. That $\sum_1^n x_i/n$ actually maximizes L is easily shown.

Thus the statistic

$$\hat{\theta} = u(X_1, X_2, \ldots, X_n) = \frac{1}{n}\sum_1^n X_i = \bar{X}$$

is the unique m.l.e. of the mean θ.

It is interesting to note that in both Examples 1 and 2, it is true that $E(\hat{\theta}) = \theta$. That is, in each of these cases, the expected value of the estimator is equal to the corresponding parameter, which leads to the following definition.

Definition 1. Any statistic whose mathematical expectation is equal to a parameter θ is called an *unbiased* estimator of the parameter θ. Otherwise, the statistic is said to be *biased.*

Example 3. Let

$$f(x; \theta) = \frac{1}{\theta}, \quad 0 < x \leq \theta, \quad 0 < \theta < \infty,$$

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

and let $X_1, X_2, \ldots, X_n$ denote a random sample from this distribution. Note that we have taken $0 < x \leq \theta$ instead of $0 < x < \theta$ so as to avoid a discussion of supremum versus maximum. Here

$$L(\theta; x_1, x_2, \ldots, x_n) = \frac{1}{\theta^n}, \qquad 0 < x_i \leq \theta,$$

which is an ever-decreasing function of θ. The maximum of such functions cannot be found by differentiation but by selecting θ as small as possible. Now $\theta \geq$ each x_i; in particular, then, $\theta \geq \max(x_i)$. Thus L can be made no larger than

$$\frac{1}{[\max(x_i)]^n}$$

and the unique m.l.e. $\hat{\theta}$ of θ in this example is the nth order statistic $\max(X_i)$. It can be shown that $E[\max(X_i)] = n\theta/(n+1)$. Thus, in this instance, the m.l.e. of the parameter θ is biased. That is, the property of unbiasedness is not in general a property of a m.l.e.

While the m.l.e. $\hat{\theta}$ of θ in Example 3 is a biased estimator, results in Chapter 5 show that the nth order statistic $\hat{\theta} = \max(X_i) = Y_n$ converges in probability to θ. Thus, in accordance with the following definition, we say that $\hat{\theta} = Y_n$ is a consistent estimator of θ.

Definition 2. Any statistic that converges in probability to a parameter θ is called a *consistent* estimator of that parameter θ.

Consistency is a desirable property of an estimator; and, in all cases of practical interest, maximum likelihood estimators are consistent.

The preceding definitions and properties are easily generalized. Let $X, Y, \ldots, Z$ denote random variables that may or may not be independent and that may or may not be identically distributed. Let the joint p.d.f. $g(x, y, \ldots, z; \theta_1, \theta_2, \ldots, \theta_m)$, $(\theta_1, \theta_2, \ldots, \theta_m) \in \Omega$, depend on m parameters. This joint p.d.f., when regarded as a function of $(\theta_1, \theta_2, \ldots, \theta_m) \in \Omega$, is called the likelihood function of the random variables. Then those functions $u_1(x, y, \ldots, z)$, $u_2(x, y, \ldots, z)$, $\ldots$, $u_m(x, y, \ldots, z)$ that maximize this likelihood function with respect to $\theta_1, \theta_2, \ldots, \theta_m$, respectively, define the maximum likelihood estimators

$$\hat{\theta}_1 = u_1(X, Y, \ldots, Z), \qquad \hat{\theta}_2 = u_2(X, Y, \ldots, Z), \ldots,$$
$$\hat{\theta}_m = u_m(X, Y, \ldots, Z)$$

of the m parameters.

Example 4. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that is $N(\theta_1, \theta_2)$, $-\infty < \theta_1 < \infty, 0 < \theta_2 < \infty$. We shall find $\hat{\theta}_1$ and $\hat{\theta}_2$, the maximum likelihood estimators of θ_1 and θ_2. The logarithm of the likelihood function may be written in the form

$$\ln L(\theta_1, \theta_2; x_1, \ldots, x_n) = -\frac{\sum_1^n (x_i - \theta_1)^2}{2\theta_2} - \frac{n \ln (2\pi\theta_2)}{2}.$$

We observe that we may maximize by differentiation. We have

$$\frac{\partial \ln L}{\partial \theta_1} = \frac{\sum_1^n (x_i - \theta_1)}{\theta_2}, \qquad \frac{\partial \ln L}{\partial \theta_2} = \frac{\sum_1^n (x_i - \theta_1)^2}{2\theta_2^2} - \frac{n}{2\theta_2}.$$

If we equate these partial derivatives to zero and solve simultaneously the two equations thus obtained, the solutions for θ_1 and θ_2 are found to be $\sum_1^n x_i/n = \bar{x}$ and $\sum_1^n (x_i - \bar{x})^2/n = s^2$, respectively. It can be verified that these

solutions maximize L. Thus the maximum likelihood estimators of $\theta_1 = \mu$ and $\theta_2 = \sigma^2$ are, respectively, the mean and the variance of the sample, namely $\hat{\theta}_1 = \bar{X}$ and $\hat{\theta}_2 = S^2$. Whereas $\hat{\theta}_1$ is an unbiased estimator of θ_1, the estimator $\hat{\theta}_2 = S^2$ is biased because

$$E(\hat{\theta}_2) = \frac{\sigma^2}{n} E\left(\frac{n\hat{\theta}_2}{\sigma^2}\right) = \frac{\sigma^2}{n} E\left(\frac{nS^2}{\sigma^2}\right) = \frac{(n-1)\sigma^2}{n} = \frac{(n-1)\theta_2}{n}.$$

However, in Chapter 5 it has been shown that $\hat{\theta}_1 = \bar{X}$ and $\hat{\theta}_2 = S^2$ converge in probability to θ_1 and θ_2, respectively, and thus they are consistent estimators of θ_1 and θ_2.

Suppose that we wish to estimate a function of θ, say $h(\theta)$. For convenience, let us say that $\eta = h(\theta)$ defines a one-to-one transformation. Then the value of η, say $\hat{\eta}$, that maximizes the likelihood function $L(\theta)$, or equivalently $L[\theta = h^{-1}(\eta)]$, is selected so that $\hat{\theta} = h^{-1}(\hat{\eta})$, where $\hat{\theta}$ is the m.l.e. of θ. Thus $\hat{\eta}$ is taken so that $\hat{\eta} = h(\hat{\theta})$; that is,

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

This result is called the *invariance property of a maximum likelihood estimator*. For illustration, if $\eta = \theta^3$, where θ is the mean of $N(\theta, 1)$, then $\hat{\eta} = \bar{X}^3$. While there is a little complication if $h(\theta)$ is not one-to-one, we still use the fact that $\hat{\eta} = h(\hat{\theta})$. Thus if $\bar{X}$ is the mean of the sample from $b(1, \theta)$, so that $\hat{\theta} = \bar{X}$ and if $\eta = \theta(1 - \theta)$, then $\hat{\eta} = \bar{X}(1 - \bar{X})$. These ideas can be extended to more than one parameter. For illustration, in Example 4, if $\eta = \theta_1 + 2\sqrt{\theta_2}$, then $\hat{\eta} = \bar{X} + 2S$.

Sometimes it is impossible to find maximum likelihood estimators in a convenient closed form and numerical methods must be used to maximize the likelihood function. For illustration, suppose that $X_1, X_2, \ldots, X_n$ is a random sample from a gamma distribution with parameters $\alpha = \theta_1$ and $\beta = \theta_2$, where $\theta_1 > 0$, $\theta_2 > 0$. It is difficult to maximize

$$L(\theta_1, \theta_2; x_1, \ldots, x_n) = \left[\frac{1}{\Gamma(\theta_1)\theta_2^{\theta_1}}\right]^n (x_1 x_2 \cdots x_n)^{\theta_1 - 1} \exp\left(-\sum_1^n x_i/\theta_2\right)$$

with respect to θ_1 and θ_2, owing to the presence of the gamma function $\Gamma(\theta_1)$. Thus numerical methods must be used to maximize L once $x_1, x_2, \ldots, x_n$ are observed.

There are other ways, however, to obtain easily point estimates of

θ_1 and θ_2. For illustration, in the gamma distribution situation, let us simply equate the first two moments of the distribution to the corresponding moments of the sample. This seems like a reasonable way in which to find estimators, since the empirical distribution $F_n(x)$ converges in probability to $F(x)$, and hence corresponding moments should be about equal. Here in this illustration we have

$$\theta_1\theta_2 = \bar{X}, \quad \theta_1\theta_2^2 = S^2,$$

the solutions of which are

$$\tilde{\theta}_1 = \frac{\bar{X}^2}{S^2} \qquad \text{and} \qquad \tilde{\theta}_2 = \frac{S^2}{\bar{X}}.$$

We say that these latter two statistics, $\tilde{\theta}_1$ and $\tilde{\theta}_2$, are respective estimators of θ_1 and θ_2 found by the *method of moments*.

To generalize the discussion of the preceding paragraph, let $X_1, X_2, \ldots, X_n$ be a random sample of size n from a distribution with p.d.f. $f(x; \theta_1, \theta_2, \ldots, \theta_r)$, $(\theta_1, \ldots, \theta_r) \in \Omega$. The expectation $E(X^k)$ is frequently called the kth moment of the distribution, $k = 1, 2, 3, \ldots$. The sum $M_k = \sum_1^n X_i^k/n$ is the kth moment of the sample, $k = 1, 2, 3, \ldots$. The method of moments can be described as follows. Equate $E(X^k)$ to M_k, beginning with $k = 1$ and continuing until there are enough equations to provide unique solutions for $\theta_1, \theta_2, \ldots, \theta_r$, say $h_i(M_1, M_2, \ldots)$, $i = 1, 2, \ldots, r$, respectively. It should be noted that this could be done in an equivalent manner by equating $\mu = E(X)$ to $\bar{X}$ and $E[(X - \mu)^k]$ to $\sum_1^n (X_i - \bar{X})^k/n$, $k = 2, 3$, and so on until unique solutions for $\theta_1, \theta_2, \ldots, \theta_r$ are obtained. This alternative procedure was used in the preceding illustration. In most practical cases, the estimator $\tilde{\theta}_i = h_i(M_1, M_2, \ldots)$ of θ_i, found by the method of moments, is a consistent estimator of θ_i, $i = 1, 2, \ldots, r$.

EXERCISES

6.1. Let $X_1, X_2, \ldots, X_n$ represent a random sample from each of the distributions having the following probability density functions:

(a) $f(x; \theta) = \theta^x e^{-\theta}/x!$, $x = 0, 1, 2, \ldots, 0 \leq \theta < \infty$, zero elsewhere, where $f(0; 0) = 1$.

(b) $f(x; \theta) = \theta x^{\theta - 1}$, $0 < x < 1$, $0 < \theta < \infty$, zero elsewhere.

(c) $f(x; \theta) = (1/\theta)e^{-x/\theta}$, $0 < x < \infty$, $0 < \theta < \infty$, zero elsewhere.

(d) $f(x;\theta) = \frac{1}{2}e^{-|x-\theta|}$, $-\infty < x < \infty$, $-\infty < \theta < \infty$.
(e) $f(x;\theta) = e^{-(x-\theta)}$, $\theta \le x < \infty$, $-\infty < \theta < \infty$, zero elsewhere.
In each case find the m.l.e. $\hat{\theta}$ of θ.

6.2. Let $X_1, X_2, \ldots, X_n$ be i.i.d., each with the distribution having p.d.f. $f(x;\theta_1,\theta_2) = (1/\theta_2)e^{-(x-\theta_1)/\theta_2}$, $\theta_1 \le x < \infty$, $-\infty < \theta_1 < \infty$, $0 < \theta_2 < \infty$, zero elsewhere. Find the maximum likelihood estimators of θ_1 and θ_2.

6.3. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample from a distribution with p.d.f. $f(x;\theta) = 1$, $\theta - \frac{1}{2} \le x \le \theta + \frac{1}{2}$, $-\infty < \theta < \infty$, zero elsewhere. Show that every statistic $u(X_1, X_2, \ldots, X_n)$ such that

$$Y_n - \tfrac{1}{2} \le u(X_1, X_2, \ldots, X_n) \le Y_1 + \tfrac{1}{2}$$

is a m.l.e. of θ. In particular, $(4Y_1 + 2Y_n + 1)/6$, $(Y_1 + Y_n)/2$, and $(2Y_1 + 4Y_n - 1)/6$ are three such statistics. Thus uniqueness is not in general a property of a m.l.e.

6.4. Let X_1, X_2, and X_3 have the multinomial distribution in which $n = 25$, $k = 4$, and the unknown probabilities are θ_1, θ_2, and θ_3, respectively. Here we can, for convenience, let $X_4 = 25 - X_1 - X_2 - X_3$ and $\theta_4 = 1 - \theta_1 - \theta_2 - \theta_3$. If the observed values of the random variables are $x_1 = 4$, $x_2 = 11$, and $x_3 = 7$, find the maximum likelihood estimates of θ_1, θ_2, and θ_3.

6.5. The *Pareto distribution* is frequently used as a model in study of incomes and has the distribution function

$$F(x;\theta_1,\theta_2) = 1 - (\theta_1/x)^{\theta_2}, \qquad \theta_1 \le x, \text{ zero elsewhere,}$$

$$\text{where } \theta_1 > 0 \quad \text{and} \quad \theta_2 > 0.$$

If $X_1, X_2, \ldots, X_n$ is a random sample from this distribution, find the maximum likelihood estimators of θ_1 and θ_2.

6.6. Let Y_n be a statistic such that $\lim_{n\to\infty} E(Y_n) = \theta$ and $\lim_{n\to\infty} \sigma^2_{Y_n} = 0$. Prove that Y_n is a consistent estimator of θ.

Hint: $\Pr(|Y_n - \theta| \ge \epsilon) \le E[(Y_n - \theta)^2]/\epsilon^2$ and $E[(Y_n - \theta)^2] = [E(Y_n - \theta)]^2 + \sigma^2_{Y_n}$. Why?

6.7. For each of the distributions in Exercise 6.1, find an estimator of θ by the method of moments and show that it is consistent.

6.8. If a random sample of size n is taken from a distribution having p.d.f. $f(x;\theta) = 2x/\theta^2$, $0 < x \le \theta$, zero elsewhere, find:
(a) The m.l.e. $\hat{\theta}$ for θ.
(b) The constant c so that $E(c\hat{\theta}) = \theta$.
(c) The m.l.e. for the median of the distribution.

6.9. Let $X_1, X_2, \ldots, X_n$ be i.i.d., each with a distribution with p.d.f. $f(x; \theta) = (1/\theta)e^{-x/\theta}$, $0 < x < \infty$, zero elsewhere. Find the m.l.e. of $\Pr(X \leq 2)$.

6.10. Let X have a binomial distribution with parameters n and p. The variance of X/n is $p(1-p)/n$; this is sometimes estimated by the m.l.e. $\frac{X}{n}\left(1 - \frac{X}{n}\right)\Big/n$. Is this an unbiased estimator of $p(1-p)/n$? If not, can you construct one by multiplying this one by a constant?

6.11. Let the table

x	0	1	2	3	4	5
Frequency	6	10	14	13	6	1

represent a summary of a sample of size 50 from a binomial distribution having $n = 5$. Find the m.l.e. of $\Pr(X \geq 3)$.

6.12. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample of size n from the uniform distribution of the continuous type over the closed interval $[\theta - \rho, \theta + \rho]$. Find the maximum likelihood estimators for θ and ρ. Are these two unbiased estimators?

6.13. Let X_1, X_2, X_3, X_4, X_5 be a random sample from a Cauchy distribution with median θ, that is, with p.d.f.

$$f(x; \theta) = \frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}, \qquad -\infty < x < \infty,$$

where $-\infty < \theta < \infty$. If $x_1 = -1.94$, $x_2 = 0.59$, $x_3 = -5.98$, $x_4 = -0.08$, $x_5 = -0.77$, find by numerical methods the m.l.e. of θ.

6.2 Confidence Intervals for Means

Suppose that we are willing to accept as a fact that the (numerical) outcome X of a random experiment is a random variable that has a normal distribution with known variance σ^2 but unknown mean μ. That is, μ is some constant, but its value is unknown. To elicit some information about μ, we decide to repeat the random experiment under identical conditions n independent times, n being a fixed positive integer. Let the random variables $X_1, X_2, \ldots, X_n$ denote, respectively, the outcomes to be obtained on these n repetitions of the experiment. If our assumptions are fulfilled, we then have under consideration a random sample $X_1, X_2, \ldots, X_n$ from a distribution that is $N(\mu, \sigma^2)$, σ^2 known. Consider the maximum likelihood estima-

tor of μ, namely $\hat{\mu} = \bar{X}$. Of course, $\bar{X}$ is $N(\mu, \sigma^2/n)$ and $(\bar{X} - \mu)/(\sigma/\sqrt{n})$ is $N(0, 1)$. Thus

$$\Pr\left(-2 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 2\right) = 0.954.$$

However, the events

$$-2 < \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} < 2,$$

$$\frac{-2\sigma}{\sqrt{n}} < \bar{X} - \mu < \frac{2\sigma}{\sqrt{n}},$$

and

$$\bar{X} - \frac{2\sigma}{\sqrt{n}} < \mu < \bar{X} + \frac{2\sigma}{\sqrt{n}}$$

are equivalent. Thus these events have the same probability. That is,

$$\Pr\left(\bar{X} - \frac{2\sigma}{\sqrt{n}} < \mu < \bar{X} + \frac{2\sigma}{\sqrt{n}}\right) = 0.954.$$

Since σ is a known number, each of the random variables $\bar{X} - 2\sigma/\sqrt{n}$ and $\bar{X} + 2\sigma/\sqrt{n}$ is a statistic. The interval $(\bar{X} - 2\sigma/\sqrt{n}, \bar{X} + 2\sigma/\sqrt{n})$ is a random interval. In this case, both end points of the interval are statistics. The immediately preceding probability statement can be read: Prior to the repeated independent performances of the random experiment, the probability is 0.954 that the random interval $(\bar{X} - 2\sigma/\sqrt{n}, \bar{X} + 2\sigma/\sqrt{n})$ includes the unknown fixed point (parameter) μ.

Up to this point, only probability has been involved; the determination of the p.d.f. of $\bar{X}$ and the determination of the random interval were problems of probability. Now the problem becomes statistical. Suppose the experiment yields $X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n$. Then the sample value of $\bar{X}$ is $\bar{x} = (x_1 + x_2 + \cdots + x_n)/n$, a known number. Moreover, since σ is known, the interval $(\bar{x} - 2\sigma/\sqrt{n}, \bar{x} + 2\sigma/\sqrt{n})$ has known endpoints. Obviously, we cannot say that 0.954 is the probability that the particular interval $(\bar{x} - 2\sigma/\sqrt{n}, \bar{x} + 2\sigma/\sqrt{n})$ includes the parameter μ, for μ, although unknown, is some constant, and this particular interval either does or does not include μ. However, the fact that we had such a high probability, prior to the performance of the experiment, that the random interval $(\bar{X} - 2\sigma/\sqrt{n}, \bar{X} + 2\sigma/\sqrt{n})$ includes the fixed point (parameter) μ, leads us to have some

reliance on the particular interval $(\bar{x} - 2\sigma/\sqrt{n}, \bar{x} + 2\sigma/\sqrt{n})$. This reliance is reflected by calling the known interval $(\bar{x} - 2\sigma/\sqrt{n}, \bar{x} + 2\sigma/\sqrt{n})$ a 95.4 percent *confidence interval* for μ. The number 0.954 is called the *confidence coefficient*. The confidence coefficient is equal to the probability that the random interval includes the parameter. One may, of course, obtain an 80, a 90, or a 99 percent confidence interval for μ by using 1.282, 1.645, or 2.576, respectively, instead of the constant 2.

A statistical inference of this sort is an example of *interval estimation* of a parameter. Note that the interval estimate of μ is found by taking a good (here maximum likelihood) estimate $\bar{x}$ of μ and adding and subtracting twice the standard deviation of $\bar{X}$, namely $2\sigma/\sqrt{n}$, which is small if n is large. If σ were not known, the end points of the random interval would not be statistics. Although the probability statement about the random interval remains valid, the sample data would not yield an interval with known end points.

Example 1. If in the preceding discussion $n = 40$, $\sigma^2 = 10$, and $\bar{x} = 7.164$, then $(7.164 - 1.282\sqrt{\frac{10}{40}}, 7.164 + 1.282\sqrt{\frac{10}{40}})$, or $(6.523, 7.805)$, is an 80 percent confidence interval for μ. Thus we have an interval estimate of μ.

In the next example we show how the central limit theorem may be used to help us find an approximate confidence interval for μ when our sample arises from a distribution that is not normal.

Example 2. Let $\bar{X}$ denote the mean of a random sample of size 25 from a distribution having variance $\sigma^2 = 100$, and mean μ. Since $\sigma/\sqrt{n} = 2$, then approximately

$$\Pr\left(-1.96 < \frac{\bar{X} - \mu}{2} < 1.96\right) = 0.95,$$

or

$$\Pr(\bar{X} - 3.92 < \mu < \bar{X} + 3.92) = 0.95.$$

Let the observed mean of the sample be $\bar{x} = 67.53$. Accordingly, the interval from $\bar{x} - 3.92 = 63.61$ to $\bar{x} + 3.92 = 71.45$ is an approximate 95 percent confidence interval for the mean μ.

Let us now turn to the problem of finding a confidence interval for the mean μ of a normal distribution when we are not so fortunate as to know the variance σ^2. From Section 4.8, we know that

$$T = \frac{\sqrt{n}(\bar{X} - \mu)/\sigma}{\sqrt{nS^2/[\sigma^2(n-1)]}} = \frac{\bar{X} - \mu}{S/\sqrt{n-1}}$$

has a t-distribution with $n - 1$ degrees of freedom, whatever the value

of $\sigma^2 > 0$. For a given positive integer n and a probability of 0.95, say, we can find a number b from Table IV in Appendix B, such that

$$\Pr\left(-b < \frac{\bar{X} - \mu}{S/\sqrt{n-1}} < b\right) = 0.95,$$

which can be written in the form

$$\Pr\left(\bar{X} - \frac{bS}{\sqrt{n-1}} < \mu < \bar{X} + \frac{bS}{\sqrt{n-1}}\right) = 0.95.$$

Then the interval $[\bar{X} - (bS/\sqrt{n-1}), \bar{X} + (bS/\sqrt{n-1})]$ is a random interval having probability 0.95 of including the unknown fixed point (parameter) μ. If the experimental values of $X_1, X_2, \ldots, X_n$ are $x_1, x_2, \ldots, x_n$ with $s^2 = \sum_1^n (x_i - \bar{x})^2/n$, where $\bar{x} = \sum_1^n x_i/n$, then the interval $[\bar{x} - (bs/\sqrt{n-1}), \bar{x} + (bs/\sqrt{n-1})]$ is a 95 percent confidence interval for μ for every $\sigma^2 > 0$. Again this interval estimate of μ is found by adding and subtracting a quantity, here $bs/\sqrt{n-1}$, to the point estimate $\bar{x}$.

Example 3. If in the preceding discussion $n = 10$, $\bar{x} = 3.22$, and $s = 1.17$, then the interval $[3.22 - (2.262)(1.17)/\sqrt{9},\ 3.22 + (2.262)(1.17)/\sqrt{9}]$ or $(2.34, 4.10)$ is a 95 percent confidence interval for μ.

Remark. If one wishes to find a confidence interval for μ and if the variance σ^2 of the nonnormal distribution is unknown (unlike Example 2 of this section), he may with large samples proceed as follows. If certain weak conditions are satisfied, then S^2, the variance of a random sample of size $n \geq 2$, converges in probability to σ^2. Then in

$$\frac{\sqrt{n}(\bar{X} - \mu)/\sigma}{\sqrt{nS^2/(n-1)\sigma^2}} = \frac{\sqrt{n-1}(\bar{X} - \mu)}{S}$$

the numerator of the left-hand member has a limiting distribution that is $N(0, 1)$ and the denominator of that member converges in probability to 1. Thus $\sqrt{n-1}(\bar{X} - \mu)/S$ has a limiting distribution that is $N(0, 1)$. This fact enables us to find approximate confidence intervals for μ when our conditions are satisfied. This procedure works particularly well when the underlying nonnormal distribution is symmetric, because then $\bar{X}$ and S^2 are uncorrelated (the proof of which is beyond the level of the text). As the underlying distribution becomes more skewed, however, the sample size must be larger to achieve good approximations to the desired probabilities. A similar procedure can be followed in the next section when seeking confidence intervals for the difference of the means of two nonnormal distributions.

We shall now consider the problem of determining a confidence interval for the unknown parameter p of a binomial distribution when the parameter n is known. Let Y be $b(n, p)$, where $0 < p < 1$ and n is known. Then p is the mean of Y/n. We shall use a result of Example 1, Section 5.5, to find an approximate 95.4 percent confidence interval for the mean p. There we found that

$$\Pr\left[-2 < \frac{Y - np}{\sqrt{n(Y/n)(1 - Y/n)}} < 2\right] = 0.954,$$

approximately. Since

$$\frac{Y - np}{\sqrt{n(Y/n)(1 - Y/n)}} = \frac{(Y/n) - p}{\sqrt{(Y/n)(1 - Y/n)/n}},$$

the probability statement above can easily be written in the form

$$\Pr\left[\frac{Y}{n} - 2\sqrt{\frac{(Y/n)(1 - Y/n)}{n}} < p < \frac{Y}{n} + 2\sqrt{\frac{(Y/n)(1 - Y/n)}{n}}\right] = 0.954,$$

approximately. Thus, for large n, if the experimental value of Y is y, the interval

$$\left[\frac{y}{n} - 2\sqrt{\frac{(y/n)(1 - y/n)}{n}}, \quad \frac{y}{n} + 2\sqrt{\frac{(y/n)(1 - y/n)}{n}}\right]$$

provides an approximate 95.4 percent confidence interval for p.

A more complicated approximate 95.4 percent confidence interval can be obtained from the fact that $Z = (Y - np)/\sqrt{np(1 - p)}$ has a limiting distribution that is $N(0, 1)$, and the fact that the event $-2 < Z < 2$ is equivalent to the event

$$\frac{Y + 2 - 2\sqrt{[Y(n - Y)/n] + 1}}{n + 4} < p < \frac{Y + 2 + 2\sqrt{[Y(n - Y)/n] + 1}}{n + 4}. \tag{1}$$

The first of these facts was established in Chapter 5, and the proof of inequalities (1) is left as an exercise. Thus an experimental value y of Y may be used in inequalities (1) to determine an approximate 95.4 percent confidence interval for p.

If one wishes a 95 percent confidence interval for p that does not depend upon limiting distribution theory, he or she may use the following approach. (This approach is quite general and can be used in other instances; see Exercise 6.21.) Determine two *increasing*

functions of p, say $c_1(p)$ and $c_2(p)$, such that for each value of p we have, at least approximately,

$$\Pr[c_1(p) < Y < c_2(p)] = 0.95.$$

The reason that this may be approximate is due to the fact that Y has a distribution of the discrete type and thus it is, in general, impossible to achieve the probability 0.95 exactly. With $c_1(p)$ and $c_2(p)$ increasing functions, they have single-valued inverses, say $d_1(y)$ and $d_2(y)$, respectively. Thus the events $c_1(p) < Y < c_2(p)$ and $d_2(Y) < p < d_1(Y)$ are equivalent and we have, at least approximately,

$$\Pr[d_2(Y) < p < d_1(Y)] = 0.95.$$

In the case of the binomial distribution, the functions $c_1(p)$, $c_2(p)$, $d_2(y)$, and $d_1(y)$ cannot be found explicitly, but a number of books provide tables of $d_2(y)$ and $d_1(y)$ for various values of n.

Example 4. If, in the preceding discussion, we take $n = 100$ and $y = 20$, the first approximate 95.4 percent confidence interval is given by $(0.2 - 2\sqrt{(0.2)(0.8)/100},\ 0.2 + 2\sqrt{(0.2)(0.8)/100})$ or $(0.12, 0.28)$. The approximate 95.4 percent confidence interval provided by inequalities (1) is

$$\left(\frac{22 - 2\sqrt{(1600/100) + 1}}{104}, \frac{22 + 2\sqrt{(1600/100) + 1}}{104}\right)$$

or $(0.13, 0.29)$. By referring to the appropriate tables found elsewhere, we find that an approximate 95 percent confidence interval has the limits $d_2(20) = 0.13$ and $d_1(20) = 0.29$. Thus, in this example, we see that all three methods yield results that are in substantial agreement.

Remark. The fact that the variance of Y/n is a function of p caused us some difficulty in finding a confidence interval for p. Another way of handling the problem is to try to find a function $u(Y/n)$ of Y/n, whose variance is essentially free of p. In Section 5.4, we proved that

$$u\left(\frac{Y}{n}\right) = \arcsin\sqrt{\frac{Y}{n}}$$

has an approximate normal distribution with mean $\arcsin\sqrt{p}$ and variance $1/4n$. Hence we could find an approximate 95.4 percent confidence interval by using

$$\Pr\left(-2 < \frac{\arcsin\sqrt{Y/n} - \arcsin\sqrt{p}}{\sqrt{1/4n}} < 2\right) = 0.954$$

and solving the inequalities for p.

Example 5. Suppose that we sample from a distribution with unknown

mean μ and variance $\sigma^2 = 225$. We want to find the sample size n so that $\bar{x} \pm 1$ (which means $\bar{x} - 1$ to $\bar{x} + 1$) serves as a 95 percent confidence interval for μ. Using the fact that the sample mean of the observations, $\bar{X}$, is approximately $N(\mu, \sigma^2/n)$, we see that the interval given by $\bar{x} \pm 1.96(15/\sqrt{n})$ will serve as an approximate 95 percent confidence interval for μ. That is, we want

$$1.96\left(\frac{15}{\sqrt{n}}\right) = 1$$

or, equivalently,

$$\sqrt{n} = 29.4, \qquad \text{and thus} \qquad n \approx 864.36$$

or $n = 865$ because n must be an integer. Suppose, however, we could not afford to take 865 observations. In that case, the accuracy or confidence level could possibly be relaxed some. For illustration, rather than requiring $\bar{x} \pm 1$ to be a 95 percent confidence interval for μ, possibly $\bar{x} \pm 2$ would be a satisfactory 80 percent one. If this modification is acceptable, we now have

$$1.282\left(\frac{15}{\sqrt{n}}\right) = 2$$

or, equivalently,

$$\sqrt{n} = 9.615 \qquad \text{and} \qquad n \approx 92.4.$$

Since n must be an integer, we would probably use 93 in practice. Most likely, the persons involved in this project would find this is a more reasonable sample size.

EXERCISES

6.14. Let the observed value of the mean $\bar{X}$ of a random sample of size 20 from a distribution that is $N(\mu, 80)$ be 81.2. Find a 95 percent confidence interval for μ.

6.15. Let $\bar{X}$ be the mean of a random sample of size n from a distribution that is $N(\mu, 9)$. Find n such that $\Pr(\bar{X} - 1 < \mu < \bar{X} + 1) = 0.90$, approximately.

6.16. Let a random sample of size 17 from the normal distribution $N(\mu, \sigma^2)$ yield $\bar{x} = 4.7$ and $s^2 = 5.76$. Determine a 90 percent confidence interval for μ.

6.17. Let $\bar{X}$ denote the mean of a random sample of size n from a distribution that has mean μ and variance $\sigma^2 = 10$. Find n so that the probability is approximately 0.954 that the random interval $(\bar{X} - \frac{1}{2}, \bar{X} + \frac{1}{2})$ includes μ.

6.18. Let $X_1, X_2, \ldots, X_9$ be a random sample of size 9 from a distribution that is $N(\mu, \sigma^2)$.

(a) If σ is known, find the length of a 95 percent confidence interval for μ if this interval is based on the random variable $\sqrt{9}(\bar{X} - \mu)/\sigma$.

(b) If σ is unknown, find the expected value of the length of a 95 percent confidence interval for μ if this interval is based on the random variable $\sqrt{8}(\bar{X} - \mu)/S$.

Hint: Write $E(S) = (\sigma/\sqrt{n})E[(nS^2/\sigma^2)^{1/2}]$.

(c) Compare these two answers.

6.19. Let $X_1, X_2, \ldots, X_n, X_{n+1}$ be a random sample of size $n + 1, n > 1$, from a distribution that is $N(\mu, \sigma^2)$. Let $\bar{X} = \sum_1^n X_i/n$ and $S^2 = \sum_1^n (X_i - \bar{X})^2/n$. Find the constant c so that the statistic $c(\bar{X} - X_{n+1})/S$ has a t-distribution. If $n = 8$, determine k such that $\Pr(\bar{X} - kS < X_9 < \bar{X} + kS) = 0.80$. The observed interval $(\bar{x} - ks, \bar{x} + ks)$ is often called an 80 percent *prediction interval* for X_9.

6.20. Let Y be $b(300, p)$. If the observed value of Y is $y = 75$, find an approximate 90 percent confidence interval for p.

6.21. Let $\bar{X}$ be the mean of a random sample of size n from a distribution that is $N(\mu, \sigma^2)$, where the positive variance σ^2 is known. Use the fact that $\Phi(2) - \Phi(-2) = 0.954$ to find, for each μ, $c_1(\mu)$ and $c_2(\mu)$ such that $\Pr[c_1(\mu) < \bar{X} < c_2(\mu)] = 0.954$. Note that $c_1(\mu)$ and $c_2(\mu)$ are increasing functions of μ. Solve for the respective functions $d_1(\bar{x})$ and $d_2(\bar{x})$; thus we also have that $\Pr[d_2(\bar{X}) < \mu < d_1(\bar{X})] = 0.954$. Compare this with the answer obtained previously in the text.

6.22. In the notation of the discussion of the confidence interval for p, show that the event $-2 < Z < 2$ is equivalent to inequalities (1).

Hint: First observe that $-2 < Z < 2$ is equivalent to $Z^2 < 4$, which can be written as an inequality involving a quadratic expression in p.

6.23. Let $\bar{X}$ denote the mean of a random sample of size 25 from a gamma-type distribution with $\alpha = 4$ and $\beta > 0$. Use the central limit theorem to find an approximate 0.954 confidence interval for μ, the mean of the gamma distribution.

Hint: Base the confidence interval on the random variable $(\bar{X} - 4\beta)/(4\beta^2/25)^{1/2} = 5\bar{X}/2\beta - 10$.

6.24. Let $\bar{x}$ be the observed mean of a random sample of size n from a distribution having mean μ and known variance σ^2. Find n so that $\bar{x} - \sigma/4$ to $\bar{x} + \sigma/4$ is an approximate 95 percent confidence interval for μ.

6.25. Assume a binomial model for a certain random variable. If we desire a 90 percent confidence interval for p that is at most 0.02 in length, find n.

Hint: Note that $\sqrt{(y/n)(1-y/n)} \leq \sqrt{(\frac{1}{2})(1-\frac{1}{2})}$.

6.26. It is known that a random variable X has a Poisson distribution with parameter μ. A sample of 200 observations from this population has a mean equal to 3.4. Construct an approximate 90 percent confidence interval for μ.

6.27. Let $Y_1 < Y_2 < \cdots < Y_n$ denote the order statistics of a random sample of size n from a distribution that has p.d.f. $f(x) = 3x^2/\theta^3$, $0 < x < \theta$, zero elsewhere.
(a) Show that $\Pr(c < Y_n/\theta < 1) = 1 - c^{3n}$, where $0 < c < 1$.
(b) If n is 4 and if the observed value of Y_4 is 2.3, what is a 95 percent confidence interval for θ?

6.28. Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$, where both parameters μ and σ^2 are unknown. A *confidence interval* for σ^2 can be found as follows. We know that nS^2/σ^2 is $\chi^2(n-1)$. Thus we can find constants a and b so that $\Pr(nS^2/\sigma^2 < b) = 0.975$ and $\Pr(a < nS^2/\sigma^2 < b) = 0.95$.
(a) Show that this second probability statement can be written as $\Pr(nS^2/b < \sigma^2 < nS^2/a) = 0.95$.
(b) If $n = 9$ and $s^2 = 7.63$, find a 95 percent confidence interval for σ^2.
(c) If μ is known, how would you modify the preceding procedure for finding a confidence interval for σ^2?

6.29. Let $X_1, X_2, \ldots, X_n$ be a random sample from a gamma distribution with known parameter $\alpha = 3$ and unknown $\beta > 0$. Discuss the construction of a confidence interval for β.
Hint: What is the distribution of $2\sum_{i=1}^{n} X_i/\beta$? Follow the procedure outlined in Exercise 6.28.

6.3 Confidence Intervals for Differences of Means

The random variable T may also be used to obtain a confidence interval for the difference $\mu_1 - \mu_2$ between the means of two normal distributions, say $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, when the distributions have the same, but unknown, variance σ^2.

Remark. Let X have a normal distribution with unknown parameters μ_1 and σ^2. A modification can be made in conducting the experiment so that the variance of the distribution will remain the same but the mean of the distribution will be changed; say, increased. After the modification has been effected, let the random variable be denoted by Y, and let Y have a normal distribution with unknown parameters μ_2 and σ^2. Naturally, it is hoped that

μ_2 is greater than μ_1, that is, that $\mu_1 - \mu_2 < 0$. Accordingly, one seeks a confidence interval for $\mu_1 - \mu_2$ in order to make a statistical inference.

A confidence interval for $\mu_1 - \mu_2$ may be obtained as follows: Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ denote, respectively, independent random samples from the two distributions, $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, respectively. Denote the means of the samples by $\bar{X}$ and $\bar{Y}$ and the variances of the samples by S_1^2 and S_2^2, respectively. It should be noted that these four statistics are independent. The independence of $\bar{X}$ and S_1^2 (and, inferentially that of $\bar{Y}$ and S_2^2) was established in Section 4.8; the assumption that the two samples are independent accounts for the independence of the others. Thus $\bar{X}$ and $\bar{Y}$ are normally and independently distributed with means μ_1 and μ_2 and variances σ^2/n and σ^2/m, respectively. In accordance with Section 4.7, their difference $\bar{X} - \bar{Y}$ is normally distributed with mean $\mu_1 - \mu_2$ and variance $\sigma^2/n + \sigma^2/m$. Then the random variable

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\sigma^2/n + \sigma^2/m}}$$

is normally distributed with zero mean and unit variance. This random variable may serve as the numerator of a T random variable. Further, nS_1^2/σ^2 and mS_2^2/σ^2 have independent chi-square distributions with $n - 1$ and $m - 1$ degrees of freedom, respectively, so that their sum $(nS_1^2 + mS_2^2)/\sigma^2$ has a chi-square distribution with $n + m - 2$ degrees of freedom, provided that $m + n - 2 > 0$. Because of the independence of $\bar{X}$, $\bar{Y}$, S_1^2, and S_2^2, it is seen that

$$\sqrt{\frac{nS_1^2 + mS_2^2}{\sigma^2(n + m - 2)}}$$

may serve as the denominator of a T random variable. That is, the random variable

$$T = \frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\dfrac{nS_1^2 + mS_2^2}{n + m - 2}\left(\dfrac{1}{n} + \dfrac{1}{m}\right)}}$$

has a t-distribution with $n + m - 2$ degrees of freedom. As in the previous section, we can (once n and m are specified positive integers with $n + m - 2 > 0$) find a positive number b from Table IV of Appendix B such that

$$\Pr(-b < T < b) = 0.95.$$

If we set

$$R = \sqrt{\frac{nS_1^2 + mS_2^2}{n+m-2}\left(\frac{1}{n}+\frac{1}{m}\right)},$$

this probability may be written in the form

$$\Pr\left[(\bar{X}-\bar{Y}) - bR < \mu_1 - \mu_2 < (\bar{X}-\bar{Y}) + bR\right] = 0.95.$$

It follows that the random interval

$$\left[(\bar{X}-\bar{Y}) - b\sqrt{\frac{nS_1^2 + mS_2^2}{n+m-2}\left(\frac{1}{n}+\frac{1}{m}\right)},\right.$$

$$\left.(\bar{X}-\bar{Y}) + b\sqrt{\frac{nS_1^2 + mS_2^2}{n+m-2}\left(\frac{1}{n}+\frac{1}{m}\right)}\right]$$

has probability 0.95 of including the unknown fixed point $(\mu_1 - \mu_2)$. As usual, the experimental values of $\bar{X}$, $\bar{Y}$, S_1^2, and S_2^2, namely $\bar{x}$, $\bar{y}$, s_1^2, and s_2^2, will provide a 95 percent confidence interval for $\mu_1 - \mu_2$ when the variances of the two normal distributions are unknown but equal. A consideration of the difficulty encountered when the unknown variances of the two normal distributions are not equal is assigned to one of the exercises.

Example 1. It may be verified that if in the preceding discussion $n = 10$, $m = 7$, $\bar{x} = 4.2$, $\bar{y} = 3.4$, $s_1^2 = 49$, $s_2^2 = 32$, then the interval $(-5.16, 6.76)$ is a 90 percent confidence interval for $\mu_1 - \mu_2$.

Let Y_1 and Y_2 be two independent random variables with binomial distributions $b(n_1, p_1)$ and $b(n_2, p_2)$, respectively. Let us now turn to the problem of finding a confidence interval for the difference $p_1 - p_2$ of the means of Y_1/n_1 and Y_2/n_2 when n_1 and n_2 are known. Since the mean and the variance of $Y_1/n_1 - Y_2/n_2$ are, respectively, $p_1 - p_2$ and $p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2$, then the random variable given by the ratio

$$\frac{(Y_1/n_1 - Y_2/n_2) - (p_1 - p_2)}{\sqrt{p_1(1-p_1)/n_1 + p_2(1-p_2)/n_2}}$$

has mean zero and variance 1 for all positive integers n_1 and n_2. Moreover, since both Y_1 and Y_2 have approximate normal distributions for large n_1 and n_2, one suspects that the ratio has an approximate normal distribution. This is actually the case, but it will not be

proved here. Moreover, if $n_1/n_2 = c$, where c is a fixed positive constant, the result of Exercise 6.36 shows that the random variable

$$\frac{(Y_1/n_1)(1 - Y_1/n_1)/n_1 + (Y_2/n_2)(1 - Y_2/n_2)/n_2}{p_1(1 - p_1)/n_1 + p_2(1 - p_2)/n_2} \tag{1}$$

converges in probability to 1 as $n_2 \to \infty$ (and thus $n_1 \to \infty$, since $n_1/n_2 = c$, $c > 0$). In accordance with Theorem 6, Section 5.5, the random variable

$$W = \frac{(Y_1/n_1 - Y_2/n_2) - (p_1 - p_2)}{U},$$

where

$$U = \sqrt{(Y_1/n_1)(1 - Y_1/n_1)/n_1 + (Y_2/n_2)(1 - Y_2/n_2)/n_2},$$

has a limiting distribution that is $N(0, 1)$. The event $-2 < W < 2$, the probability of which is approximately equal to 0.954, is equivalent to the event

$$\frac{Y_1}{n_1} - \frac{Y_2}{n_2} - 2U < p_1 - p_2 < \frac{Y_1}{n_1} - \frac{Y_2}{n_2} + 2U.$$

Accordingly, the experimental values y_1 and y_2 of Y_1 and Y_2, respectively, will provide an approximate 95.4 percent confidence interval for $p_1 - p_2$.

Example 2. If, in the preceding discussion, we take $n_1 = 100$, $n_2 = 400$, $y_1 = 30$, $y_2 = 80$, then the experimental values of $Y_1/n_1 - Y_2/n_2$ and U are 0.1 and $\sqrt{(0.3)(0.7)/100 + (0.2)(0.8)/400} = 0.05$, respectively. Thus the interval (0, 0.2) is an approximate 95.4 percent confidence interval for $p_1 - p_2$.

EXERCISES

6.30. Let two independent random samples, each of size 10, from two normal distributions $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$ yield $\bar{x} = 4.8$, $s_1^2 = 8.64$, $\bar{y} = 5.6$, $s_2^2 = 7.88$. Find a 95 percent confidence interval for $\mu_1 - \mu_2$.

6.31. Let two independent random variables Y_1 and Y_2, with binomial distributions that have parameters $n_1 = n_2 = 100$, p_1, and p_2, respectively, be observed to be equal to $y_1 = 50$ and $y_2 = 40$. Determine an approximate 90 percent confidence interval for $p_1 - p_2$.

6.32. Discuss the problem of finding a confidence interval for the difference $\mu_1 - \mu_2$ between the two means of two normal distributions if the variances σ_1^2 and σ_2^2 are known but not necessarily equal.

6.33. Discuss Exercise 6.32 when it is assumed that the variances are unknown and unequal. This is a very difficult problem, and the discussion should point out exactly where the difficulty lies. If, however, the variances are unknown but their ratio σ_1^2/σ_2^2 is a known constant k, then a statistic that is a T random variable can again be used. Why?

6.34. As an illustration of Exercise 6.33, one can let $X_1, X_2, \ldots, X_9$ and $Y_1, Y_2, \ldots, Y_{12}$ represent two independent random samples from the respective normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$. It is given that $\sigma_1^2 = 3\sigma_2^2$, but σ_2^2 is unknown. Define a random variable which has a t-distribution that can be used to find a 95 percent interval for $\mu_1 - \mu_2$.

6.35. Let $\bar{X}$ and $\bar{Y}$ be the means of two independent random samples, each of size n, from the respective distributions $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, where the common variance is known. Find n such that

$$\Pr(\bar{X} - \bar{Y} - \sigma/5 < \mu_1 - \mu_2 < \bar{X} - \bar{Y} + \sigma/5) = 0.90$$

6.36. Under the conditions given, show that the random variable defined by ratio (1) of the text converges in probability to 1.

6.37. Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be two independent random samples from the respective normal distributions $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, where the four parameters are unknown. To construct a *confidence interval for the ratio*, σ_1^2/σ_2^2, of the variances, form the quotient of the two independent chi-square variables, each divided by its degrees of freedom, namely

$$F = \frac{\dfrac{mS_2^2}{\sigma_2^2}\Big/(m-1)}{\dfrac{nS_1^2}{\sigma_1^2}\Big/(n-1)},$$

where S_1^2 and S_2^2 are the respective sample variances.

(a) What kind of distribution does F have?

(b) From the appropriate table, a and b can be found so that $\Pr(F < b) = 0.975$ and $\Pr(a < F < b) = 0.95$.

(c) Rewrite the second probability statement as

$$\Pr\left[a\frac{nS_1^2/(n-1)}{mS_2^2/(m-1)} < \frac{\sigma_1^2}{\sigma_2^2} < b\frac{nS_1^2/(n-1)}{mS_2^2/(m-1)}\right] = 0.95.$$

The observed values, s_1^2 and s_2^2, can be inserted in these inequalities to provide a 95 percent confidence interval for σ_1^2/σ_2^2.

6.4 Tests of Statistical Hypotheses

The two principal areas of statistical inference are the areas of estimation of parameters and of tests of statistical hypotheses. The

problem of estimation of parameters, both point and interval estimation, has been treated. In Sections 6.4 and 6.5 some aspects of statistical hypotheses and tests of statistical hypotheses will be considered. The subject will be introduced by way of example.

Example 1. Let it be known that the outcome X of a random experiment is $N(\theta, 100)$. For instance, X may denote a score on a test, which score we assume to be normally distributed with mean θ and variance 100. Let us say the past experience with this random experiment indicates that $\theta = 75$. Suppose, owing possibly to some research in the area pertaining to this experiment, some changes are made in the method of performing this random experiment. It is then suspected that no longer does $\theta = 75$ but that now $\theta > 75$. There is as yet no formal experimental evidence that $\theta > 75$; hence the statement $\theta > 75$ is a conjecture or a *statistical hypothesis*. In admitting that the statistical hypothesis $\theta > 75$ may be false, we allow, in effect, the possibility that $\theta \leq 75$. Thus there are actually two statistical hypotheses. First, that the unknown parameter $\theta \leq 75$; that is, there has been no increase in θ. Second, that the unknown parameter $\theta > 75$. Accordingly, the parameter space is $\Omega = \{\theta : -\infty < \theta < \infty\}$. We denote the first of these hypotheses by the symbols $H_0 : \theta \leq 75$ and the second by the symbols $H_1 : \theta > 75$. Since the values $\theta > 75$ are alternatives to those where $\theta \leq 75$, the hypothesis $H_1 : \theta > 75$ is called the *alternative hypothesis*. Needless to say, H_0 could be called the alternative to H_1; however, the conjecture, here $\theta > 75$, that is made by the research worker is usually taken to be the alternative hypothesis. In any case the problem is to decide which of these hypotheses is to be accepted. To reach a decision, the random experiment is to be repeated a number of independent times, say n, and the results observed. That is, we consider a random sample $X_1, X_2, \ldots, X_n$ from a distribution that is $N(\theta, 100)$, and we devise a rule that will tell us what decision to make once the experimental values, say $x_1, x_2, \ldots, x_n$, have been determined. Such a rule is called a *test* of the hypothesis $H_0 : \theta \leq 75$ against the alternative hypothesis $H_1 : \theta > 75$. There is no bound on the number of rules or tests that can be constructed. We shall consider three such tests. Our tests will be constructed around the following notion. We shall partition the sample space $\mathscr{A}$ into a subset C and its complement C^*. If the experimental values of $X_1, X_2, \ldots, X_n$, say $x_1, x_2, \ldots, x_n$, are such that the point $(x_1, x_2, \ldots, x_n) \in C$, we shall reject the hypothesis H_0 (accept the hypothesis H_1). If we have $(x_1, x_2, \ldots, x_n) \in C^*$, we shall accept the hypothesis H_0 (reject the hypothesis H_1).

Test 1. Let $n = 25$. The sample space $\mathscr{A}$ is the set

$$\{(x_1, x_2, \ldots, x_{25}) : -\infty < x_i < \infty, i = 1, 2, \ldots, 25\}.$$

Let the subset C of the sample space be

$$C = \{(x_1, x_2, \ldots, x_{25}) : x_1 + x_2 + \cdots + x_{25} > (25)(75)\}.$$

We shall reject the hypothesis H_0 if and only if our 25 experimental values are such that $(x_1, x_2, \ldots, x_{25}) \in C$. If $(x_1, x_2, \ldots, x_{25})$ is not an element of C, we shall accept the hypothesis H_0. This subset C of the sample space that leads to the rejection of the hypothesis $H_0 : \theta \leq 75$ is called the *critical region* of Test 1. Now $\sum_1^{25} x_i > (25)(75)$ if and only if $\bar{x} > 75$, where $\bar{x} = \sum_1^{25} x_i/25$. Thus we can much more conveniently say that we shall reject the hypothesis $H_0 : \theta \leq 75$ and accept the hypothesis $H_1 : \theta > 75$ if and only if the experimentally determined value of the sample mean $\bar{x}$ is greater than 75. If $\bar{x} \leq 75$, we accept the hypothesis $H_0 : \theta \leq 75$. Our test then amounts to this: We shall reject the hypothesis $H_0 : \theta \leq 75$ if the mean of the sample exceeds the maximum value of the mean of the distribution when the hypothesis H_0 is true.

It would help us to evaluate a test of a statistical hypothesis if we knew the probability of rejecting that hypothesis (and hence of accepting the alternative hypothesis). In our Test 1, this means that we want to compute the probability

$$\Pr[(X_1, \ldots, X_{25}) \in C] = \Pr(\bar{X} > 75).$$

Obviously, this probability is a function of the parameter θ and we shall denote it by $K_1(\theta)$. The function $K_1(\theta) = \Pr(\bar{X} > 75)$ is called the *power function* of Test 1, and the value of the power function at a parameter point is called the *power* of Test 1 at that point. Because $\bar{X}$ is $N(\theta, 4)$, we have

$$K_1(\theta) = \Pr\left(\frac{\bar{X} - \theta}{2} > \frac{75 - \theta}{2}\right) = 1 - \Phi\left(\frac{75 - \theta}{2}\right).$$

So, for illustration, we have, by Table III of Appendix B, that the power at $\theta = 75$ is $K_1(75) = 0.500$. Other powers are $K_1(73) = 0.159$, $K_1(77) = 0.841$, and $K_1(79) = 0.977$. The graph of $K_1(\theta)$ of Test 1 is depicted in Figure 6.1. Among other things, this means that, if $\theta = 75$, the probability of rejecting the hypothesis $H_0 : \theta \leq 75$ is $\frac{1}{2}$. That is, if $\theta = 75$ so that H_0 is true, the

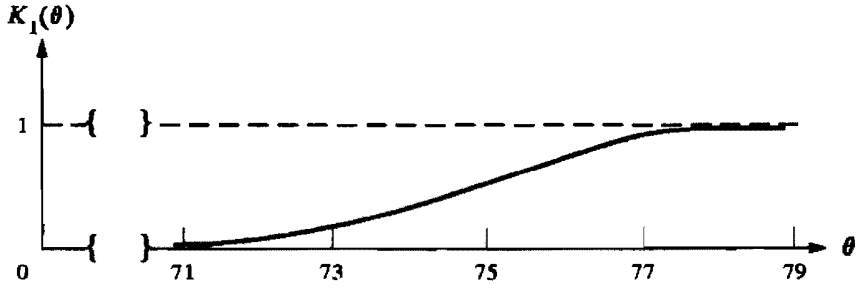

FIGURE 6.1

probability of rejecting this true hypothesis H_0 is $\frac{1}{2}$. Many statisticians and research workers find it very undesirable to have such a high probability as $\frac{1}{2}$ assigned to this kind of mistake: namely the rejection of H_0 when H_0 is a true hypothesis. Thus Test 1 does not appear to be a very satisfactory test. Let us try to devise another test that does not have this objectionable feature. We shall do this by making it more difficult to reject the hypothesis H_0, with the hope that this will give a smaller probability of rejecting H_0 when that hypothesis is true.

Test 2. Let $n = 25$. We shall reject the hypothesis $H_0 : \theta \leq 75$ and accept the hypothesis $H_1 : \theta > 75$ if and only if $\bar{x} > 78$. Here the critical region is $C = \{(x_1, \ldots, x_{25}) : x_1 + \cdots + x_{25} > (25)(78)\}$. The power function of Test 2 is, because $\bar{X}$ is $N(\theta, 4)$,

$$K_2(\theta) = \Pr(\bar{X} > 78) = 1 - \Phi\left(\frac{78 - \theta}{2}\right).$$

Some values of the power function of Test 2 are $K_2(73) = 0.006$, $K_2(75) = 0.067$, $K_2(77) = 0.309$, and $K_2(79) = 0.691$. That is, if $\theta = 75$, the probability of rejecting $H_0 : \theta \leq 75$ is 0.067; this is much more desirable than the corresponding probability $\frac{1}{2}$ that resulted from Test 1. However, if H_0 is false and, in fact, $\theta = 77$, the probability of rejecting $H_0 : \theta \leq 75$ (and hence of accepting $H_1 : \theta > 75$) is only 0.309. In certain instances, this low probability 0.309 of a correct decision (the acceptance of H_1 when H_1 is true) is objectionable. That is, Test 2 is not wholly satisfactory. Perhaps we can overcome the undesirable features of Tests 1 and 2 if we proceed as in Test 3.

Test 3. Let us first select a power function $K_3(\theta)$ that has the features of a small value at $\theta = 75$ and a large value at $\theta = 77$. For instance, take $K_3(75) = 0.159$ and $K_3(77) = 0.841$. To determine a test with such a power function, let us reject $H_0 : \theta \leq 75$ if and only if the experimental value $\bar{x}$ of the mean of a random sample of size n is greater than some constant c. Thus the critical region is $C = \{(x_1, x_2, \ldots, x_n) : x_1 + x_2 + \cdots + x_n > nc\}$. It should be noted that the sample size n and the constant c have not been determined as yet. However, since $\bar{X}$ is $N(\theta, 100/n)$, the power function is

$$K_3(\theta) = \Pr(\bar{X} > c) = 1 - \Phi\left(\frac{c - \theta}{10/\sqrt{n}}\right).$$

The conditions $K_3(75) = 0.159$ and $K_3(77) = 0.841$ require that

$$1 - \Phi\left(\frac{c - 75}{10/\sqrt{n}}\right) = 0.159, \qquad 1 - \Phi\left(\frac{c - 77}{10/\sqrt{n}}\right) = 0.841.$$

Equivalently, from Table III of Appendix B, we have

$$\frac{c - 75}{10/\sqrt{n}} = 1, \qquad \frac{c - 77}{10/\sqrt{n}} = -1.$$

The solution to these two equations in n and c is $n = 100$, $c = 76$. With these values of n and c, other powers of Test 3 are $K_3(73) = 0.001$ and $K_3(79) = 0.999$. It is important to observe that although Test 3 has a more desirable power function than those of Tests 1 and 2, a certain "price" has been paid—a sample size of $n = 100$ is required in Test 3, whereas we had $n = 25$ in the earlier tests.

Remark. Throughout the text we frequently say that we accept the hypothesis H_0 if we do not reject H_0 in favor of H_1. If this decision is made, it certainly does not mean that H_0 is true or that we even believe that it is true. All it means is, based upon the data at hand, that we are not convinced that the hypothesis H_0 is wrong. Accordingly, the statement "We accept H_0" would possibly be better read as "We do not reject H_0." However, because it is in fairly common use, we use the statement "We accept H_0," but read it with this remark in mind.

We have now illustrated the following concepts:

1. A statistical hypothesis.
2. A test of a hypothesis against an alternative hypothesis and the associated concept of the critical region of the test.
3. The power of a test.

These concepts will now be formally defined.

Definition 3. A *statistical hypothesis* is an assertion about the distribution of one or more random variables. If the statistical hypothesis completely specifies the distribution, it is called a *simple statistical hypothesis*; if it does not, it is called a *composite statistical hypothesis*.

If we refer to Example 1, we see that both $H_0 : \theta \leq 75$ and $H_1 : \theta > 75$ are composite statistical hypotheses, since neither of them completely specifies the distribution. If there, instead of $H_0 : \theta \leq 75$, we had $H_0 : \theta = 75$, then H_0 would have been a simple statistical hypothesis.

Definition 4. A *test* of a statistical hypothesis is a rule which, when the experimental sample values have been obtained, leads to a decision to accept or to reject the hypothesis under consideration.

Definition 5. Let C be that subset of the sample space which, in accordance with a prescribed test, leads to the rejection of the hypothesis under consideration. Then C is called the *critical region* of the test.

Definition 6. The *power function* of a test of a statistical hypothesis H_0 against an alternative hypothesis H_1 is that function, defined for all distributions under consideration, which yields the probability that the sample point falls in the critical region C of the test, that is, a function that yields the probability of rejecting the hypothesis under consideration. The value of the power function at a parameter point is called the *power* of the test at that point.

Definition 7. Let H_0 denote a hypothesis that is to be tested against an alternative hypothesis H_1 in accordance with a prescribed test. The *significance level* of the test (or the *size* of the critical region C) is the maximum value (actually supremum) of the power function of the test when H_0 is true.

If we refer again to Example 1, we see that the significance levels of Tests 1, 2, and 3 of that example are 0.500, 0.067, and 0.159, respectively. An additional example may help clarify these definitions.

Example 2. It is known that the random variable X has a p.d.f. of the form

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

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

It is desired to test the simple hypothesis $H_0 : \theta = 2$ against the alternative simple hypothesis $H_1 : \theta = 4$. Thus $\Omega = \{\theta : \theta = 2, 4\}$. A random sample X_1, X_2 of size $n = 2$ will be used. The test to be used is defined by taking the critical region to be $C = \{(x_1, x_2) : 9.5 \leq x_1 + x_2 < \infty\}$. The power function of the test and the significance level of the test will be determined.

There are but two probability density functions under consideration, namely, $f(x; 2)$ specified by H_0 and $f(x; 4)$ specified by H_1. Thus the power function is defined at but two points $\theta = 2$ and $\theta = 4$. The power function of the test is given by $\Pr[(X_1, X_2) \in C]$. If H_0 is true, that is, $\theta = 2$, the joint p.d.f. of X_1 and X_2 is

$$f(x_1; 2)f(x_2; 2) = \tfrac{1}{4}e^{-(x_1 + x_2)/2}, \qquad 0 < x_1 < \infty, \quad 0 < x_2 < \infty,$$

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

and

$$\Pr[(X_1, X_2) \in C] = 1 - \Pr[(X_1, X_2) \in C^*]$$

$$= 1 - \int_0^{9.5} \int_0^{9.5 - x_2} \tfrac{1}{4} e^{-(x_1 + x_2)/2}\, dx_1\, dx_2$$

$$= 0.05, \quad \text{approximately.}$$

If H_1 is true, that is, $\theta = 4$, the joint p.d.f. of X_1 and X_2 is

$$f(x_1; 4)f(x_2; 4) = \tfrac{1}{16}e^{-(x_1 + x_2)/4}, \qquad 0 < x_1 < \infty, \quad 0 < x_2 < \infty,$$
$$= 0 \qquad \text{elsewhere,}$$

and

$$\Pr[(X_1, X_2) \in C] = 1 - \int_0^{9.5} \int_0^{9.5 - x_2} \tfrac{1}{16}e^{-(x_1 + x_2)/4}\, dx_1\, dx_2$$
$$= 0.31, \quad \text{approximately.}$$

Thus the power of the test is given by 0.05 for $\theta = 2$ and by 0.31 for $\theta = 4$. That is, the probability of rejecting H_0 when H_0 is true is 0.05, and the probability of rejecting H_0 when H_0 is false is 0.31. Since the significance level of this test (or the size of the critical region) is the power of the test when H_0 is true, the significance level of this test is 0.05.

The fact that the power of this test, when $\theta = 4$, is only 0.31 immediately suggests that a search be made for another test which, with the same power when $\theta = 2$, would have a power greater than 0.31 when $\theta = 4$. However later, it will be clear that such a search would be fruitless. That is, there is no test with a significance level of 0.05 and based on a random sample of size $n = 2$ that has greater power at $\theta = 4$. The only manner in which the situation may be improved is to have recourse to a random sample of size n greater than 2.

Our computations of the powers of this test at the two points $\theta = 2$ and $\theta = 4$ were purposely done the hard way to focus attention on fundamental concepts. A procedure that is computationally simpler is the following. When the hypothesis H_0 is true, the random variable X is $\chi^2(2)$. Thus the random variable $X_1 + X_2 = Y$, say, is $\chi^2(4)$. Accordingly, the power of the test when H_0 is true is given by

$$\Pr(Y \geq 9.5) = 1 - \Pr(Y < 9.5) = 1 - 0.95 = 0.05,$$

from Table II of Appendix B. When the hypothesis H_1 is true, the random variable $X/2$ is $\chi^2(2)$; so the random variable $(X_1 + X_2)/2 = Z$, say, is $\chi^2(4)$. Accordingly, the power of the test when H_1 is true is given by

$$\Pr(X_1 + X_2 \geq 9.5) = \Pr(Z \geq 4.75)$$
$$= \int_{4.75}^{\infty} \tfrac{1}{4}ze^{-z/2}\, dz,$$

which is equal to 0.31, approximately.

Remark. The rejection of the hypothesis H_0 when that hypothesis is true is, of course, an incorrect decision or an error. This incorrect decision is often called a type I error; accordingly, the significance level of the test is the probability of committing an error of type I. The acceptance of H_0 when H_0

is false (H_1 is true) is called an error of type II. Thus the probability of a type II error is 1 minus the power of the test when H_1 is true. Frequently, it is disconcerting to the student to discover that there are so many names for the same thing. However, since all of them are used in the statistical literature, we feel obligated to point out that "significance level," "size of the critical region," "power of the test when H_0 is true," and "the probability of committing an error of type I" are all equivalent.

EXERCISES

6.38. Let X have a p.d.f. of the form $f(x; \theta) = \theta x^{\theta-1}, 0 < x < 1$, zero elsewhere, where $\theta \in \{\theta : \theta = 1, 2\}$. To test the simple hypothesis $H_0 : \theta = 1$ against the alternative simple hypothesis $H_1 : \theta = 2$, use a random sample X_1, X_2 of size $n = 2$ and define the critical region to be $C = \{(x_1, x_2) : \frac{3}{4} \leq x_1 x_2\}$. Find the power function of the test.

6.39. Let X have a binomial distribution with parameters $n = 10$ and $p \in \{p : p = \frac{1}{4}, \frac{1}{2}\}$. The simple hypothesis $H_0 : p = \frac{1}{2}$ is rejected, and the alternative simple hypothesis $H_1 : p = \frac{1}{4}$ is accepted, if the observed value of X_1, a random sample of size 1, is less than or equal to 3. Find the power function of the test.

6.40. Let X_1, X_2 be a random sample of size $n = 2$ from the distribution having p.d.f. $f(x; \theta) = (1/\theta)e^{-x/\theta}, 0 < x < \infty$, zero elsewhere. We reject $H_0 : \theta = 2$ and accept $H_1 : \theta = 1$ if the observed values of X_1, X_2, say x_1, x_2, are such that

$$\frac{f(x_1; 2)f(x_2; 2)}{f(x_1; 1)f(x_2; 1)} \leq \frac{1}{2}.$$

Here $\Omega = \{\theta : \theta = 1, 2\}$. Find the significance level of the test and the power of the test when H_0 is false.

6.41. Sketch, as in Figure 6.1, the graphs of the power functions of Tests 1, 2, and 3 of Example 1 of this section.

6.42. Let us assume that the life of a tire in miles, say X, is normally distributed with mean θ and standard deviation 5000. Past experience indicates that $\theta = 30{,}000$. The manufacturer claims that the tires made by a new process have mean $\theta > 30{,}000$, and it is very possible that $\theta = 35{,}000$. Let us check his claim by testing $H_0 : \theta = 30{,}000$ against $H_1 : \theta > 30{,}000$. We shall observe n independent values of X, say $x_1, \ldots, x_n$, and we shall reject H_0 (thus accept H_1) if and only if $\bar{x} \geq c$. Determine n and c so that the power function $K(\theta)$ of the test has the values $K(30{,}000) = 0.01$ and $K(35{,}000) = 0.98$.

6.43. Let X have a Poisson distribution with mean θ. Consider the simple hypothesis $H_0 : \theta = \frac{1}{2}$ and the alternative composite hypothesis $H_1 : \theta < \frac{1}{2}$.

Thus $\Omega = \{\theta : 0 < \theta \leq \frac{1}{2}\}$. Let $X_1, \ldots, X_{12}$ denote a random sample of size 12 from this distribution. We reject H_0 if and only if the observed value of $Y = X_1 + \cdots + X_{12} \leq 2$. If $K(\theta)$ is the power function of the test, find the powers $K(\frac{1}{2})$, $K(\frac{1}{3})$, $K(\frac{1}{4})$, $K(\frac{1}{6})$, and $K(\frac{1}{12})$. Sketch the graph of $K(\theta)$. What is the significance level of the test?

6.44. Let Y have a binomial distribution with parameters n and p. We reject $H_0 : p = \frac{1}{2}$ and accept $H_1 : p > \frac{1}{2}$ if $Y \geq c$. Find n and c to give a power function $K(p)$ which is such that $K(\frac{1}{2}) = 0.10$ and $K(\frac{2}{3}) = 0.95$, approximately.

6.45. Let $Y_1 < Y_2 < Y_3 < Y_4$ be the order statistics of a random sample of size $n = 4$ from a distribution with p.d.f. $f(x; \theta) = 1/\theta$, $0 < x < \theta$, zero elsewhere, where $0 < \theta$. The hypothesis $H_0 : \theta = 1$ is rejected and $H_1 : \theta > 1$ accepted if the observed $Y_4 \geq c$.

(a) Find the constant c so that the significance level is $\alpha = 0.05$.

(b) Determine the power function of the test.

6.5 Additional Comments About Statistical Tests

All of the alternative hypotheses considered in Section 6.4 were *one-sided hypotheses*. For illustration, in Exercise 6.42 we tested $H_0 : \theta = 30{,}000$ against the one-sided alternative $H_1 : \theta > 30{,}000$, where θ is the mean of a normal distribution having standard deviation $\sigma = 5000$. The test associated with this situation, namely reject H_0 if and only if the sample mean $\bar{X} \geq c$, is a *one-sided test*. For convenience, we often call $H_0 : \theta = 30{,}000$ the *null hypothesis* because, as in this exercise, it suggests that the new process has not changed the mean of the distribution. That is, the new process has been used without consequence if in fact the mean still equals 30,000; hence the terminology null hypothesis is appropriate. So in Exercise 6.42 we are testing a simple null hypothesis against a composite one-sided alternative with a one-sided test.

This does suggest that there could be two-sided alternative hypotheses. For illustration, in Exercise 6.42, suppose there is the possibility that the new process might decrease the mean. That is, say that we simply do not know whether with the new process $\theta > 30{,}000$ or $\theta < 30{,}000$; or there has been no change and the null hypothesis $H_0 : \theta = 30{,}000$ is still true. Then we would want to test $H_0 : \theta = 30{,}000$ against the two-sided alternative $H_1 : \theta \neq 30{,}000$. To help see how to construct a *two-sided test* for H_0 against H_1, consider the following argument.

In dealing with a test of $H_0: \theta = 30{,}000$ against the one-sided alternative $\theta > 30{,}000$, we used $\bar{X} \geq c$ or, equivalently,

$$Z = \frac{\bar{X} - 30{,}000}{\sigma/\sqrt{n}} \geq \frac{c - 30{,}000}{\sigma/\sqrt{n}} = c_1,$$

where since $\bar{X}$ is $N(\theta = 30{,}000, \sigma^2/n)$ under H_0, Z is $N(0, 1)$; and we could select $c_1 = 1.645$ to have a test of significance level $\alpha = 0.05$. That is, if $\bar{X}$ is $1.645\sigma/\sqrt{n}$ greater than the mean $\theta = 30{,}000$, we would reject H_0 and accept H_1 and the significance level would be equal to $\alpha = 0.05$. To test $H_0: \theta = 30{,}000$ against $H_1: \theta \neq 30{,}000$, let us again use $\bar{X}$ through Z and reject H_0 if $\bar{X}$ or Z is too large or too small. Namely, if we reject H_0 and accept H_1 when

$$|Z| = \left|\frac{\bar{X} - 30{,}000}{\sigma/\sqrt{n}}\right| \geq 1.96,$$

the significance level $\alpha = 0.05$ because this is the probability of $|Z| \geq 1.96$ when H_0 is true.

It is interesting to note that the latter test is the equivalent of saying that we reject H_0 and accept H_1 if 30,000 is not in the (two-sided) confidence interval for the mean θ. Or equivalently, if

$$\bar{X} - 1.96\frac{\sigma}{\sqrt{n}} < 30{,}000 < \bar{X} + 1.96\frac{\sigma}{\sqrt{n}},$$

then we accept $H_0: \theta = 30{,}000$ because those two inequalities are equivalent to

$$\left|\frac{\bar{X} - 30{,}000}{\sigma/\sqrt{n}}\right| < 1.96,$$

which leads to the acceptance of $H_0: \theta = 30{,}000$.

Once we recognize this relationship between confidence intervals and tests of hypotheses, we can use all those statistics that we used to construct confidence intervals to test hypotheses, not only against two-sided alternatives but one-sided ones as well. Without listing all of these in a table, we give enough of them so that the principle can be understood.

Example 1. Let $\bar{X}$ and S^2 be the mean and the variance of a random sample of size n coming from $N(\mu, \sigma^2)$. To test, at significance level $\alpha = 0.05$, $H_0: \mu = \mu_0$ against the two-sided alternative $H_1: \mu \neq \mu_0$, reject if

$$|T| = \left|\frac{\bar{X} - \mu_0}{S/\sqrt{n-1}}\right| \geq b,$$

where b is the 97.5th percentile of the t-distribution with $n-1$ degrees of freedom.

Example 2. Let independent random samples be taken from $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, respectively. Say these have the respective sample characteristics n, $\bar{X}$, S_1^2 and m, $\bar{Y}$, S_2^2. At $\alpha = 0.05$, reject $H_0 : \mu_1 = \mu_2$ and accept the one-sided alternative $H_1 : \mu_1 > \mu_2$ if

$$T = \frac{\bar{X} - \bar{Y} - 0}{\sqrt{\frac{nS_1^2 + mS_2^2}{n+m-2}\left(\frac{1}{n} + \frac{1}{m}\right)}} \geq c.$$

Note that $\bar{X} - \bar{Y}$ has a normal distribution with mean zero under H_0. So c is taken as the 95th percentile of a t-distribution with $n + m - 2$ degrees of freedom to provide $\alpha = 0.05$.

Example 3. Say Y is $b(n, p)$. To test $H_0 : p = p_0$ against $H_1 : p < p_0$, we use either

$$Z_1 = \frac{(Y/n) - p_0}{\sqrt{p_0(1-p_0)/n}} \leq c \qquad \text{or} \qquad Z_2 = \frac{(Y/n) - p_0}{\sqrt{(Y/n)(1 - Y/n)/n}} \leq c.$$

If n is large, both Z_1 and Z_2 have approximate standard normal distributions provided that $H_0 : p = p_0$ is true. Hence c is taken to be -1.645 to give an approximate significance level of $\alpha = 0.05$. Some statisticians use Z_1 and others Z_2. We do not have strong preference one way or the other because the two methods provide about the same numerical result. As one might suspect, using Z_1 provides better probabilities for power calculations if the true p is close to p_0 while Z_2 is better if H_0 is clearly false. However, with a two-sided alternative hypothesis, Z_2 does provide a better relationship with the confidence interval for p. That is, $|Z_2| < 2$ is equivalent to p_0 being in the interval from

$$\frac{Y}{n} - 2\sqrt{\frac{(Y/n)(1 - Y/n)}{n}} \quad \text{to} \quad \frac{Y}{n} + 2\sqrt{\frac{(Y/n)(1 - Y/n)}{n}},$$

which is the interval that provides a 95.4 percent confidence interval for p as considered in Section 6.2.

In closing this section, we introduce the concepts of *randomized tests* and *p-values* through an example and remarks that follow the example.

Example 4. Let $X_1, X_2, \ldots, X_{10}$ be a random sample of size $n = 10$ from a Poisson distribution with mean θ. A critical region for testing $H_0 : \theta = 0.1$ against $H_1 : \theta > 0.1$ is given by $Y = \sum_{i=1}^{10} X_i \geq 3$. The statistic Y has a Poisson

distribution with mean 10θ. Thus, with $\theta = 0.1$ so that the mean of Y is 1, the significance level of the test is

$$\Pr(Y \geq 3) = 1 - \Pr(Y \leq 2) = 1 - 0.920 = 0.080.$$

If the critical region defined by $\sum_1^{10} x_i \geq 4$ is used, the significance level is

$$\alpha = \Pr(Y \geq 4) = 1 - \Pr(Y \leq 3) = 1 - 0.981 = 0.019.$$

If a significance level of about $\alpha = 0.05$, say, is desired, most statisticians would use one of these tests; that is, they would adjust the significance level to that of one of these convenient tests. However, a significance level of $\alpha = 0.05$ can be achieved exactly by rejecting H_0 if $\sum_1^{10} x_i \geq 4$ or if $\sum_1^{10} x_i = 3$ and an auxiliary independent random experiment resulted in "success," where the probability of success is selected to be equal to

$$\frac{0.050 - 0.019}{0.080 - 0.019} = \frac{31}{61}.$$

This is due to the fact that, when $\theta = 0.1$ so that the mean of Y is 1,

$$\begin{aligned}\Pr(Y \geq 4) + \Pr(Y = 3 \text{ and success}) &= 0.019 + \Pr(Y = 3)\Pr(\text{success}) \\ &= 0.019 + (0.061)\tfrac{31}{61} = 0.05.\end{aligned}$$

The process of performing the auxiliary experiment to decide whether to reject or not when $Y = 3$ is sometimes referred to as a *randomized test.*

Remarks. Not many statisticians like randomized tests in practice, because the use of them means that two statisticians could make the same assumptions, observe the same data, apply the same test, and yet make different decisions. Hence they usually adjust their significance level so as not to randomize. As a matter of fact, many statisticians report what are commonly called *p-values* (for *probability values*). For illustration, if in Example 4 the observed Y is $y = 4$, the p-value is 0.019; and if it is $y = 3$, the p-value is 0.080. That is, the p-value is the observed "tail" probability of a statistic being at least as extreme as the particular observed value when H_0 is true. Hence, more generally, if $Y = u(X_1, X_2, \ldots, X_n)$ is the statistic to be used in a test of H_0 and if the critical region is of the form

$$u(x_1, x_2, \ldots, x_n) \leq c,$$

an observed value $u(x_1, x_2, \ldots, x_n) = d$ would mean that the

$$p\text{-value} = \Pr(Y \leq d; H_0).$$

That is, if $G(y)$ is the distribution function of $Y = u(X_1, X_2, \ldots, X_n)$, provided that H_0 is true, the p-value is equal to $G(d)$ in this case. However,

$G(Y)$, in the continuous case, is uniformly distributed on the unit interval, so an observed value $G(d) \leq 0.05$ would be equivalent to selecting c, so that

$$\Pr[u(X_1, X_2, \ldots, X_n) \leq c; H_0] = 0.05$$

and observing that $d \leq c$. Most computer programs automatically print out the p-value of a test.

Example 5. Let $X_1, X_2, \ldots, X_{25}$ be a random sample from $N(\mu, \sigma^2 = 4)$. To test $H_0: \mu = 77$ against the one-sided alternative hypothesis $H_1: \mu < 77$, say we observe the 25 values and determine that $\bar{x} = 76.1$. The variance of $\bar{X}$ is $\sigma^2/n = 4/25 = 0.16$; so we know that $Z = (\bar{X} - 77)/0.4$ is $N(0, 1)$ provided that $\mu = 77$. Since the observed value of this test statistic is $z = (76.1 - 77)/0.4 = -2.25$, the p-value of the test is $\Phi(-2.25) = 1 - 0.988 = 0.012$. Accordingly, if we were using a significance level of $\alpha = 0.05$, we would reject H_0 and accept $H_1: \mu < 77$ because $0.012 < 0.05$.

EXERCISES

6.46. Assume that the weight of cereal in a "10-ounce box" is $N(\mu, \sigma^2)$. To test $H_0: \mu = 10.1$ against $H_1: \mu > 10.1$, we take a random sample of size $n = 16$ and observe that $\bar{x} = 10.4$ and $s = 0.4$.

(a) Do we accept or reject H_0 at the 5 percent significance level?

(b) What is the approximate p-value of this test?

6.47. Each of 51 golfers hit three golf balls of brand X and three golf balls of brand Y in a random order. Let X_i and Y_i equal the averages of the distances traveled by the brand X and brand Y golf balls hit by the ith golfer, $i = 1, 2, \ldots, 51$. Let $W_i = X_i - Y_i$, $i = 1, 2, \ldots, 51$. Test $H_0: \mu_W = 0$ against $H_1: \mu_W > 0$, where μ_W is the mean of the differences. If $\bar{w} = 2.07$ and $s_w^2 = 84.63$, would H_0 be accepted or rejected at an $\alpha = 0.05$ significance level? What is the p-value of this test?

6.48. Among the data collected for the World Health Organization air quality monitoring project is a measure of suspended particles in $\mu g/m^3$. Let X and Y equal the concentration of suspended particles in $\mu g/m^3$ in the city center (commercial district) for Melbourne and Houston, respectively. Using $n = 13$ observations of X and $m = 16$ observations of Y, we shall test $H_0: \mu_X = \mu_Y$ against $H_1: \mu_X < \mu_Y$.

(a) Define the test statistic and critical region, assuming that the variances are equal. Let $\alpha = 0.05$.

(b) If $\bar{x} = 72.9, s_x = 25.6, \bar{y} = 81.7$, and $s_y = 28.3$, calculate the value of the test statistic and state your conclusion.

6.49. Let p equal the proportion of drivers who use a seat belt in a state that

does not have a mandatory seat belt law. It was claimed that $p = 0.14$. An advertising campaign was conducted to increase this proportion. Two months after the campaign, $y = 104$ out of a random sample of $n = 590$ drivers were wearing their seat belts. Was the campaign successful?
(a) Define the null and alternative hypotheses.
(b) Define a critical region with an $\alpha = 0.01$ significance level.
(c) Determine the approximate p-value and state your conclusion.

6.50. A machine shop that manufactures toggle levers has both a day and a night shift. A toggle lever is defective if a standard nut cannot be screwed onto the threads. Let p_1 and p_2 be the proportion of defective levers among those manufactured by the day and night shifts, respectively. We shall test the null hypothesis, $H_0 : p_1 = p_2$, against a two-sided alternative hypothesis based on two random samples, each of 1000 levers taken from the production of the respective shifts.
(a) Define the test statistic which has an approximate $N(0, 1)$ distribution. Sketch a standard normal p.d.f. illustrating the critical region having $\alpha = 0.05$.
(b) If $y_1 = 37$ and $y_2 = 53$ defectives were observed for the day and night shifts, respectively, calculate the value of the test statistic and the approximate p-value (note that this is a two-sided test). Locate the calculated test statistic on your figure in part (a) and state your conclusion.

6.51. In Exercise 6.28 we found a confidence interval for the variance σ^2 using the variance S^2 of a random sample of size n arising from $N(\mu, \sigma^2)$, where the mean μ is unknown. In testing $H_0 : \sigma^2 = \sigma_0^2$ against $H_1 : \sigma^2 > \sigma_0^2$, use the critical region defined by $nS^2/\sigma_0^2 \geq c$. That is, reject H_0 and accept H_1 if $S^2 \geq c\sigma_0^2/n$. If $n = 13$ and the significance level $\alpha = 0.025$, determine c.

6.52. In Exercise 6.37, in finding a confidence interval for the ratio of the variances of two normal distributions, we used a statistic $[nS_1^2/(n-1)]/[mS_2^2/(m-1)]$, which has an F-distribution when those two variances are equal. If we denote that statistic by F, we can test $H_0 : \sigma_1^2 = \sigma_2^2$ against $H_1 : \sigma_1^2 > \sigma_2^2$ using the critical region $F \geq c$. If $n = 13$, $m = 11$, and $\alpha = 0.05$, find c.

6.6 Chi-Square Tests

In this section we introduce tests of statistical hypotheses called *chi-square tests*. A test of this sort was originally proposed by Karl Pearson in 1900, and it provided one of the earlier methods of statistical inference.

Let the random variable X_i be $N(\mu_i, \sigma_i^2)$, $i = 1, 2, \ldots, n$, and let $X_1, X_2, \ldots, X_n$ be mutually independent. Thus the joint p.d.f. of these variables is

$$\frac{1}{\sigma_1\sigma_2\cdots\sigma_n(2\pi)^{n/2}} \exp\left[-\frac{1}{2}\sum_1^n \left(\frac{x_i-\mu_i}{\sigma_i}\right)^2\right], \qquad -\infty < x_i < \infty.$$

The random variable that is defined by the exponent (apart from the coefficient $-\frac{1}{2}$) is $\sum_1^n (X_i - \mu_i)^2/\sigma_i^2$, and this random variable is $\chi^2(n)$. In Section 4.10 we generalized this joint normal distribution of probability to n random variables that are *dependent* and we call the distribution a *multivariate normal distribution*. In Section 10.8, it will be shown that a certain exponent in the joint p.d.f. (apart from a coefficient of $-\frac{1}{2}$) defines a random variable that is $\chi^2(n)$. This fact is the mathematical basis of the chi-square tests.

Let us now discuss some random variables that have approximate chi-square distributions. Let X_1 be $b(n, p_1)$. Since the random variable $Y = (X_1 - np_1)/\sqrt{np_1(1-p_1)}$ has, as $n \to \infty$, a limiting distribution that is $N(0, 1)$, we would strongly suspect that the limiting distribution of $Z = Y^2$ is $\chi^2(1)$. This is, in fact, the case, as will now be shown. If $G_n(y)$ represents the distribution function of Y, we know that

$$\lim_{n\to\infty} G_n(y) = \Phi(y), \qquad -\infty < y < \infty,$$

where $\Phi(y)$ is the distribution function of a distribution that is $N(0, 1)$. Let $H_n(z)$ represent, for each positive integer n, the distribution function of $Z = Y^2$. Thus, if $z \geq 0$,

$$\begin{aligned} H_n(z) = \Pr(Z \leq z) &= \Pr(-\sqrt{z} \leq Y \leq \sqrt{z}) \\ &= G_n(\sqrt{z}) - G_n[(-\sqrt{z})-]. \end{aligned}$$

Accordingly, since $\Phi(y)$ is everywhere continuous,

$$\begin{aligned} \lim_{n\to\infty} H_n(z) &= \Phi(\sqrt{z}) - \Phi(-\sqrt{z}) \\ &= 2\int_0^{\sqrt{z}} \frac{1}{\sqrt{2\pi}} e^{-w^2/2}\, dw. \end{aligned}$$

If we change the variable of integration in this last integral by writing $w^2 = v$, then

$$\lim_{n\to\infty} H_n(z) = \int_0^{z^2} \frac{1}{\Gamma(\frac{1}{2})2^{1/2}} v^{1/2-1} e^{-v/2}\, dv,$$

provided that $z \geq 0$. If $z < 0$, then $\lim_{n\to\infty} H_n(z) = 0$. Thus $\lim_{n\to\infty} H_n(z)$ is equal to the distribution function of a random variable that is $\chi^2(1)$. This is the desired result.

Let us now return to the random variable X_1 which is $b(n, p_1)$. Let $X_2 = n - X_1$ and let $p_2 = 1 - p_1$. If we denote Y^2 by Q_1 instead of Z, we see that Q_1 may be written as

$$Q_1 = \frac{(X_1 - np_1)^2}{np_1(1-p_1)} = \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_1 - np_1)^2}{n(1-p_1)}$$

$$= \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_2 - np_2)^2}{np_2}$$

because $(X_1 - np_1)^2 = (n - X_2 - n + np_2)^2 = (X_2 - np_2)^2$. Since Q_1 has a limiting chi-square distribution with 1 degree of freedom, we say, when n is a positive integer, that Q_1 has an approximate chi-square distribution with 1 degree of freedom. This result can be generalized as follows.

Let $X_1, X_2, \ldots, X_{k-1}$ have a multinomial distribution with the parameters $n, p_1, \ldots, p_{k-1}$, as in Section 3.1. As a convenience, let $X_k = n - (X_1 + \cdots + X_{k-1})$ and let $p_k = 1 - (p_1 + \cdots + p_{k-1})$. Define Q_{k-1} by

$$Q_{k-1} = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i}.$$

It is proved in a more advanced course that, as $n \to \infty$, Q_{k-1} has a limiting distribution that is $\chi^2(k-1)$. If we accept this fact, we can say that Q_{k-1} has an approximate chi-square distribution with $k-1$ degrees of freedom when n is a positive integer. Some writers caution the user of this approximation to be certain that n is large enough that each np_i, $i = 1, 2, \ldots, k$, is at least equal to 5. In any case it is important to realize that Q_{k-1} does not have a chi-square distribution, only an approximate chi-square distribution.

The random variable Q_{k-1} may serve as the basis of the tests of certain statistical hypotheses which we now discuss. Let the sample

space $\mathscr{A}$ of a random experiment be the union of a finite number k of mutually disjoint sets $A_1, A_2, \ldots, A_k$. Furthermore, let $P(A_i) = p_i$, $i = 1, 2, \ldots, k$, where $p_k = 1 - p_1 - \cdots - p_{k-1}$, so that p_i is the probability that the outcome of the random experiment is an element of the set A_i. The random experiment is to be repeated n independent times and X_i will represent the number of times the outcome is an element of the set A_i. That is, $X_1, X_2, \ldots, X_k = n - X_1 - \cdots - X_{k-1}$ are the frequencies with which the outcome is, respectively, an element of $A_1, A_2, \ldots, A_k$. Then the joint p.d.f. of $X_1, X_2, \ldots, X_{k-1}$ is the multinomial p.d.f. with the parameters $n, p_1, \ldots, p_{k-1}$. Consider the simple hypothesis (concerning this multinomial p.d.f.) $H_0 : p_1 = p_{10}$, $p_2 = p_{20}, \ldots, p_{k-1} = p_{k-1,0}$ $(p_k = p_{k0} = 1 - p_{10} - \cdots - p_{k-1,0})$, where $p_{10}, \ldots, p_{k-1,0}$ are specified numbers. It is desired to test H_0 against all alternatives.

If the hypothesis H_0 is true, the random variable

$$Q_{k-1} = \sum_{1}^{k} \frac{(X_i - np_{i0})^2}{np_{i0}}$$

has an approximate chi-square distribution with $k - 1$ degrees of freedom. Since, when H_0 is true, np_{i0} is the expected value of X_i, one would feel intuitively that experimental values of Q_{k-1} should not be too large if H_0 is true. With this in mind, we may use Table II of Appendix B, with $k - 1$ degrees of freedom, and find c so that $\Pr(Q_{k-1} \geq c) = \alpha$, where α is the desired significance level of the test. If, then, the hypothesis H_0 is rejected when the observed value of Q_{k-1} is at least as great as c, the test of H_0 will have a significance level that is approximately equal to α.

Some illustrative examples follow.

Example 1. One of the first six positive integers is to be chosen by a random experiment (perhaps by the cast of a die). Let $A_i = \{x : x = i\}$, $i = 1, 2, \ldots, 6$. The hypothesis $H_0 : P(A_i) = p_{i0} = \frac{1}{6}$, $i = 1, 2, \ldots, 6$, will be tested, at the approximate 5 percent significance level, against all alternatives. To make the test, the random experiment will be repeated, under the same conditions, 60 independent times. In this example $k = 6$ and $np_{i0} = 60(\frac{1}{6}) = 10$, $i = 1, 2, \ldots, 6$. Let X_i denote the frequency with which the random experiment terminates with the outcome in A_i, $i = 1, 2, \ldots, 6$, and let $Q_5 = \sum_{1}^{6} (X_i - 10)^2/10$. If H_0 is true, Table II, with $k - 1 = 6 - 1 = 5$ degrees of freedom, shows that we have $\Pr(Q_5 \geq 11.1) = 0.05$. Now suppose that

the experimental frequencies of $A_1, A_2, \ldots, A_6$ are, respectively, 13, 19, 11, 8, 5, and 4. The observed value of Q_5 is

$$\frac{(13-10)^2}{10}+\frac{(19-10)^2}{10}+\frac{(11-10)^2}{10}+\frac{(8-10)^2}{10}$$
$$+\frac{(5-10)^2}{10}+\frac{(4-10)^2}{10}=15.6$$

Since $15.6 > 11.1$, the hypothesis $P(A_i)=\frac{1}{6}$, $i=1,2,\ldots,6$, is rejected at the (approximate) 5 percent significance level.

Example 2. A point is to be selected from the unit interval $\{x: 0<x<1\}$ by a random process. Let $A_1=\{x: 0<x\leq\frac{1}{4}\}$, $A_2=\{x:\frac{1}{4}<x\leq\frac{1}{2}\}$, $A_3=\{x:\frac{1}{2}<x\leq\frac{3}{4}\}$, and $A_4=\{x:\frac{3}{4}<x<1\}$. Let the probabilities p_i, $i=1,2,3,4$, assigned to these sets under the hypothesis be determined by the p.d.f. $2x$, $0<x<1$, zero elsewhere. Then these probabilities are, respectively,

$$p_{10}=\int_0^{1/4}2x\,dx=\tfrac{1}{16},\qquad p_{20}=\tfrac{3}{16},\qquad p_{30}=\tfrac{5}{16},\qquad p_{40}=\tfrac{7}{16}.$$

Thus the hypothesis to be tested is that p_1, p_2, p_3, and $p_4=1-p_1-p_2-p_3$ have the preceding values in a multinomial distribution with $k=4$. This hypothesis is to be tested at an approximate 0.025 significance level by repeating the random experiment $n=80$ independent times under the same conditions. Here the np_{i0}, $i=1,2,3,4$, are, respectively, 5, 15, 25, and 35. Suppose the observed frequencies of A_1, A_2, A_3, and A_4 are 6, 18, 20, and 36, respectively. Then the observed value of $Q_3=\sum_1^4(X_i-np_{i0})^2/(np_{i0})$ is

$$\frac{(6-5)^2}{5}+\frac{(18-15)^2}{15}+\frac{(20-25)^2}{25}+\frac{(36-35)^2}{35}=\frac{64}{35}=1.83,$$

approximately. From Table II, with $4-1=3$ degrees of freedom, the value corresponding to a 0.025 significance level is $c=9.35$. Since the observed value of Q_3 is less than 9.35, the hypothesis is accepted at the (approximate) 0.025 level of significance.

Thus far we have used the chi-square test when the hypothesis H_0 is a simple hypothesis. More often we encounter hypotheses H_0 in which the multinomial probabilities $p_1, p_2, \ldots, p_k$ are not completely specified by the hypothesis H_0. That is, under H_0, these probabilities are functions of unknown parameters. For illustration, suppose that a certain random variable Y can take on any real value. Let us partition the space $\{y: -\infty<y<\infty\}$ into k mutually disjoint sets $A_1, A_2, \ldots, A_k$ so that the events $A_1, A_2, \ldots, A_k$ are mutually exclu-

sive and exhaustive. Let H_0 be the hypothesis that Y is $N(\mu, \sigma^2)$ with μ and σ^2 unspecified. Then each

$$p_i = \int_{A_i} \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left[-(y-\mu)^2/2\sigma^2\right] dy, \qquad i = 1, 2, \ldots, k,$$

is a function of the unknown parameters μ and σ^2. Suppose that we take a random sample $Y_1, \ldots, Y_n$ of size n from this distribution. If we let X_i denote the frequency of A_i, $i = 1, 2, \ldots, k$, so that $X_1 + \cdots + X_k = n$, the random variable

$$Q_{k-1} = \sum_{i=1}^{k} \frac{(X_i - np_i)^2}{np_i}$$

cannot be computed once $X_1, \ldots, X_k$ have been observed, since each p_i, and hence Q_{k-1}, is a function of the unknown parameters μ and σ^2.

There is a way out of our trouble, however. We have noted that Q_{k-1} is a function of μ and σ^2. Accordingly, choose the values of μ and σ^2 that minimize Q_{k-1}. Obviously, these values depend upon the observed $X_1 = x_1, \ldots, X_k = x_k$ and are called *minimum chi-square estimates* of μ and σ^2. These point estimates of μ and σ^2 enable us to compute numerically the estimates of each p_i. Accordingly, if these values are used, Q_{k-1} can be computed once $Y_1, Y_2, \ldots, Y_n$, and hence $X_1, X_2, \ldots, X_k$, are observed. However, a very important aspect of the fact, which we accept without proof, is that now Q_{k-1} is approximately $\chi^2(k-3)$. That is, the number of degrees of freedom of the limiting chi-square distribution of Q_{k-1} is reduced by one for each parameter estimated by the experimental data. This statement applies not only to the problem at hand but also to more general situations. Two examples will now be given. The first of these examples will deal with the test of the hypothesis that two multinominal distributions are the same.

Remark. In many instances, such as that involving the mean μ and the variance σ^2 of a normal distribution, minimum chi-square estimates are difficult to compute. Hence other estimates, such as the maximum likelihood estimates $\hat{\mu} = \bar{Y}$ and $\widehat{\sigma^2} = S^2$, are used to evaluate p_i and Q_{k-1}. In general, Q_{k-1} is not minimized by maximum likelihood estimates, and thus its computed value is somewhat greater than it would be if minimum chi-square estimates were used. Hence, when comparing it to a critical value listed in the chi-square table with $k-3$ degrees of freedom, there is a greater chance of rejecting than there would be if the actual minimum of Q_{k-1} is used.

Accordingly, the approximate significance level of such a test will be somewhat higher than that value found in the table. This modification should be kept in mind and, if at all possible, each p_i should be estimated using the frequencies $X_1, \ldots, X_k$ rather than using directly the observations $Y_1, Y_2, \ldots, Y_n$ of the random sample.

Example 3. Let us consider two multinomial distributions with parameters $n_j, p_{1j}, p_{2j}, \ldots, p_{kj}$, $j = 1, 2$, respectively. Let X_{ij}, $i = 1, 2, \ldots, k$, $j = 1, 2$, represent the corresponding frequencies. If n_1 and n_2 are large and the observations from one distribution are independent of those from the other, the random variable

$$\sum_{j=1}^{2} \sum_{i=1}^{k} \frac{(X_{ij} - n_j p_{ij})^2}{n_j p_{ij}}$$

is the sum of two independent random variables, each of which we treat as though it were $\chi^2(k-1)$; that is, the random variable is approximately $\chi^2(2k-2)$. Consider the hypothesis

$$H_0 : p_{11} = p_{12}, p_{21} = p_{22}, \ldots, p_{k1} = p_{k2},$$

where each $p_{i1} = p_{i2}$, $i = 1, 2, \ldots, k$, is unspecified. Thus we need point estimates of these parameters. The maximum likelihood estimator of $p_{i1} = p_{i2}$, based upon the frequencies X_{ij}, is $(X_{i1} + X_{i2})/(n_1 + n_2)$, $i = 1, 2, \ldots, k$. Note that we need only $k - 1$ point estimates, because we have a point estimate of $p_{k1} = p_{k2}$ once we have point estimates of the first $k - 1$ probabilities. In accordance with the fact that has been stated, the random variable

$$\sum_{j=1}^{2} \sum_{i=1}^{k} \frac{\{X_{ij} - n_j[(X_{i1} + X_{i2})/(n_1 + n_2)]\}^2}{n_j[(X_{i1} + X_{i2})/(n_1 + n_2)]}$$

has an approximate χ^2 distribution with $2k - 2 - (k - 1) = k - 1$ degrees of freedom. Thus we are able to test the hypothesis that two multinomial distributions are the same; this hypothesis is rejected when the computed value of this random variable is at least as great as an appropriate number from Table II, with $k - 1$ degrees of freedom.

The second example deals with the subject of *contingency tables.*

Example 4. Let the result of a random experiment be classified by two attributes (such as the color of the hair and the color of the eyes). That is, one attribute of the outcome is one and only one of certain mutually exclusive and exhaustive events, say $A_1, A_2, \ldots, A_a$; and the other attribute of the outcome is also one and only one of certain mutually exclusive and exhaustive events, say $B_1, B_2, \ldots, B_b$. Let $p_{ij} = P(A_i \cap B_j)$, $i = 1, 2, \ldots, a$; $j = 1, 2, \ldots, b$. The random experiment is to be repeated n independent times

and X_{ij} will denote the frequency of the event $A_i \cap B_j$. Since there are $k = ab$ such events as $A_i \cap B_j$, the random variable

$$Q_{ab-1} = \sum_{j=1}^{b} \sum_{i=1}^{a} \frac{(X_{ij} - np_{ij})^2}{np_{ij}}$$

has an approximate chi-square distribution with $ab - 1$ degrees of freedom, provided that n is large. Suppose that we wish to test the independence of the A attribute and the B attribute; that is, we wish to test the hypothesis $H_0 : P(A_i \cap B_j) = P(A_i)P(B_j)$, $i = 1, 2, \ldots, a$; $j = 1, 2, \ldots, b$. Let us denote $P(A_i)$ by $p_{i.}$ and $P(B_j)$ by $p_{.j}$; thus

$$p_{i.} = \sum_{j=1}^{b} p_{ij}, \qquad p_{.j} = \sum_{i=1}^{a} p_{ij},$$

and

$$1 = \sum_{j=1}^{b} \sum_{i=1}^{a} p_{ij} = \sum_{j=1}^{b} p_{.j} = \sum_{i=1}^{a} p_{i.}.$$

Then the hypothesis can be formulated as $H_0 : p_{ij} = p_{i.}p_{.j}$, $i = 1, 2, \ldots, a$; $j = 1, 2, \ldots, b$. To test H_0, we can use Q_{ab-1} with p_{ij} replaced by $p_{i.}p_{.j}$. But if $p_{i.}$, $i = 1, 2, \ldots, a$, and $p_{.j}$, $j = 1, 2, \ldots, b$, are unknown, as they frequently are in applications, we cannot compute Q_{ab-1} once the frequencies are observed. In such a case we estimate these unknown parameters by

$$\hat{p}_{i.} = \frac{X_{i.}}{n}, \quad \text{where} \quad X_{i.} = \sum_{j=1}^{b} X_{ij}, \qquad i = 1, 2, \ldots, a,$$

and

$$\hat{p}_{.j} = \frac{X_{.j}}{n}, \quad \text{where} \quad X_{.j} = \sum_{i=1}^{a} X_{ij}, \qquad j = 1, 2, \ldots, b.$$

Since $\sum_i p_{i.} = \sum_j p_{.j} = 1$, we have estimated only $a - 1 + b - 1 = a + b - 2$ parameters. So if these estimates are used in Q_{ab-1}, with $p_{ij} = p_{i.}p_{.j}$, then, according to the rule that has been stated in this section, the random variable

$$\sum_{j=1}^{b} \sum_{i=1}^{a} \frac{[X_{ij} - n(X_{i.}/n)(X_{.j}/n)]^2}{n(X_{i.}/n)(X_{.j}/n)}$$

has an approximate chi-square distribution with $ab - 1 - (a + b - 2) = (a - 1)(b - 1)$ degrees of freedom provided that H_0 is true. The hypothesis H_0 is then rejected if the computed value of this statistic exceeds the constant c, where c is selected from Table II so that the test has the desired significance level α.

In each of the four examples of this section, we have indicated that the statistic used to test the hypothesis H_0 has an approximate chi-square distribution, provided that n is sufficiently large and H_0 is

true. To compute the power of any of these tests for values of the parameters not described by H_0, we need the distribution of the statistic when H_0 is not true. In each of these cases, the statistic has an approximate distribution called a *noncentral chi-square distribution*. The noncentral chi-square distribution will be discussed in Section 10.3.

EXERCISES

6.53. A number is to be selected from the interval $\{x : 0 < x < 2\}$ by a random process. Let $A_i = \{x : (i-1)/2 < x \leq i/2\}$, $i = 1, 2, 3$, and let $A_4 = \{x : \frac{3}{2} < x < 2\}$. A certain hypothesis assigns probabilities p_{i0} to these sets in accordance with $p_{i0} = \int_{A_i} (\frac{1}{2})(2 - x)\, dx$, $i = 1, 2, 3, 4$. This hypothesis (concerning the multinomial p.d.f. with $k = 4$) is to be tested, at the 5 percent level of significance, by a chi-square test. If the observed frequencies of the sets A_i, $i = 1, 2, 3, 4$, are, respectively, 30, 30, 10, 10, would H_0 be accepted at the (approximate) 5 percent level of significance?

6.54. Let the following sets be defined: $A_1 = \{x : -\infty < x \leq 0\}$, $A_i = \{x : i - 2 < x \leq i - 1\}$, $i = 2, \ldots, 7$, and $A_8 = \{x : 6 < x < \infty\}$. A certain hypothesis assigns probabilities p_{i0} to these sets A_i in accordance with

$$p_{i0} = \int_{A_i} \frac{1}{2\sqrt{2\pi}} \exp\left[-\frac{(x-3)^2}{2(4)}\right] dx, \qquad i = 1, 2, \ldots, 7, 8.$$

This hypothesis (concerning the multinomial p.d.f. with $k = 8$) is to be tested, at the 5 percent level of significance, by a chi-square test. If the observed frequencies of the sets A_i, $i = 1, 2, \ldots, 8$, are, respectively, 60, 96, 140, 210, 172, 160, 88, and 74, would H_0 be accepted at the (approximate) 5 percent level of significance?

6.55. A die was cast $n = 120$ independent times and the following data resulted:

Spots up	1	2	3	4	5	6
Frequency	b	20	20	20	20	40-b

If we use a chi-square test, for what values of b would the hypothesis that the die is unbiased be rejected at the 0.025 significance level?

6.56. Consider the problem from genetics of crossing two types of peas. The Mendelian theory states that the probabilities of the classifications (a) round and yellow, (b) wrinkled and yellow, (c) round and green, and (d) wrinkled and green are $\frac{9}{16}$, $\frac{3}{16}$, $\frac{3}{16}$, and $\frac{1}{16}$, respectively. If, from 160 independent observations, the observed frequencies of these respective

classifications are 86, 35, 26, and 13, are these data consistent with the Mendelian theory? That is, test, with $\alpha = 0.01$, the hypothesis that the respective probabilities are $\frac{9}{16}$, $\frac{3}{16}$, $\frac{3}{16}$, and $\frac{1}{16}$.

6.57. Two different teaching procedures were used on two different groups of students. Each group contained 100 students of about the same ability. At the end of the term, an evaluating team assigned a letter grade to each student. The results were tabulated as follows.

	Grade					
Group	A	B	C	D	F	Total
I	15	25	32	17	11	100
II	9	18	29	28	16	100

If we consider these data to be independent observations from two respective multinomial distributions with $k = 5$, test, at the 5 percent significance level, the hypothesis that the two distributions are the same (and hence the two teaching procedures are equally effective).

6.58. Let the result of a random experiment be classified as one of the mutually exclusive and exhaustive ways A_1, A_2, A_3 and also as one of the mutually exclusive and exhaustive ways B_1, B_2, B_3, B_4. Two hundred independent trials of the experiment result in the following data:

	B_1	B_2	B_3	B_4
A_1	10	21	15	6
A_2	11	27	21	13
A_3	6	19	27	24

Test, at the 0.05 significance level, the hypothesis of independence of the A attribute and the B attribute, namely $H_0 : P(A_i \cap B_j) = P(A_i)P(B_j)$, $i = 1, 2, 3$ and $j = 1, 2, 3, 4$, against the alternative of dependence.

6.59. A certain genetic model suggests that the probabilities of a particular trinomial distribution are, respectively, $p_1 = p^2$, $p_2 = 2p(1 - p)$, and $p_3 = (1 - p)^2$, where $0 < p < 1$. If X_1, X_2, X_3 represent the respective frequencies in n independent trials, explain how we could check on the adequacy of the genetic model.

6.60. Let the result of a random experiment be classified as one of the mutually exclusive and exhaustive ways A_1, A_2, A_3 and also as one of the

mutually and exhaustive ways B_1, B_2, B_3, B_4. Say that 180 independent trials of the experiment result in the following frequencies:

	B_1	B_2	B_3	B_4
A_1	$15 - 3k$	$15 - k$	$15 + k$	$15 + 3k$
A_2	15	15	15	15
A_3	$15 + 3k$	$15 + k$	$15 - k$	$15 - 3k$

where k is one of the integers 0, 1, 2, 3, 4, 5. What is the smallest value of k that will lead to the rejection of the independence of the A attribute and the B attribute at the $\alpha = 0.05$ significance level?

6.61. It is proposed to fit the Poisson distribution to the following data

x	0	1	2	3	$3 < x$
Frequency	20	40	16	18	6

(a) Compute the corresponding chi-square goodness-of-fit statistic. *Hint:* In computing the mean, treat $3 < x$ as $x = 4$.
(b) How many degrees of freedom are associated with this chi-square?
(c) Do these data result in the rejection of the Poisson model at the $\alpha = 0.05$ significance level?

ADDITIONAL EXERCISES

6.62. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample of size n from the distribution having p.d.f. $f(x) = 2x/\theta^2$, $0 < x < \theta$, zero elsewhere.
(a) If $0 < c < 1$, show that $\Pr(c < Y_n/\theta < 1) = 1 - c^{2n}$.
(b) If $n = 5$ and if the observed value of Y_n is 1.8, find a 99 percent confidence interval for θ.

6.63. If 0.35, 0.92, 0.56, and 0.71 are the four observed values of a random sample from a distribution having p.d.f. $f(x; \theta) = \theta x^{\theta - 1}$, $0 < x < 1$, zero elsewhere, find an estimate for θ.

6.64. Let the table

x	0	1	2	3	4	5
Frequency	6	10	14	13	6	1

represent a summary of a random sample of size 50 from a Poisson distribution. Find the maximum likelihood estimate of Pr $(X = 2)$.

6.65. Let X be $N(\mu, 100)$. To test $H_0 : \mu = 80$ against $H_1 : \mu > 80$, let the critical region be defined by $C = \{(x_1, x_2, \ldots, x_{25}) : \bar{x} \geq 83\}$, where $\bar{x}$ is the sample mean of a random sample of size $n = 25$ from this distribution.
(a) How is the power function $K(\mu)$ defined for this test?
(b) What is the significance level of this test?
(c) What are the values of $K(80)$, $K(83)$, and $K(86)$?
(d) Sketch the graph of the power function.
(e) What is the p-value corresponding to $\bar{x} = 83.41$?

6.66. Let X equal the yield of alfalfa in tons per acre per year. Assume that X is $N(1.5, 0.09)$. It is hoped that new fertilizer will increase the average yield. We shall test the null hypothesis $H_0 : \mu = 1.5$ against the alternative hypothesis $H_1 : \mu > 1.5$. Assume that the variance continues to equal $\sigma^2 = 0.09$ with the new fertilizer. Using $\bar{X}$, the mean of a random sample of size n, as the test statistic, reject H_0 if $\bar{x} \geq c$. Find n and c so that the power function $K(\mu) = \Pr(\bar{X} \geq c : \mu)$ is such that $\alpha = K(1.5) = 0.05$ and $K(1.7) = 0.95$.

6.67. A random sample of 100 observations from a Poisson distribution has a mean equal to 6.25. Construct an approximate 95 percent confidence interval for the mean of the distribution.

6.68. Say that a random sample of size 25 is taken from a binomial distribution with parameters $n = 5$ and p. These data are then lost, but we recall that the relative frequency of the value 5 was $\frac{6}{25}$. Under these conditions, how would you estimate p? Is this suggested estimate unbiased?

6.69. When 100 tacks were thrown on a table, 60 of them landed point up. Obtain a 95 percent confidence interval for the probability that a tack of this type will land point up. Assume independence.

6.70. Let $X_1, X_2, \ldots, X_8$ be a random sample of size $n = 8$ from a Poisson distribution with mean μ. Reject the simple null hypothesis $H_0 : \mu = 0.5$ and accept $H_1 : \mu > 0.5$ if the observed sum $\sum_{i=1}^{8} x_i \geq 8$.
(a) Compute the significance level α of the test.
(b) Find the power function $K(\mu)$ of the test as a sum of Poisson probabilities.
(c) Using the Appendix, determine $K(0.75)$, $K(1)$, and $K(1.25)$.

6.71. Let p denote the probability that, for a particular tennis player, the first serve is good. Since $p = 0.40$, this player decided to take lessons in order to increase p. When the lessons are completed, the hypothesis

$H_0 : p = 0.40$ will be tested against $H_1 : p > 0.40$ based on $n = 25$ trials. Let y equal the number of first serves that are good, and let the critical region be defined by $C = \{y : y \geq 13\}$.

(a) Determine $\alpha = \Pr(Y \geq 13; p = 0.40)$.

(b) Find $\beta = \Pr(Y < 13)$ when $p = 0.60$; that is, $\beta = \Pr(Y \leq 12; p = 0.60)$.

6.72. The mean birth weight in the United States is $\mu = 3315$ grams with a standard deviation of $\sigma = 575$. Let X equal the birth weight in grams in Jerusalem. Assume that the distribution of X is $N(\mu, \sigma^2)$. We shall test the null hypothesis $H_0 : \mu = 3315$ against the alternative hypothesis $H_1 : \mu < 3315$ using a random sample of size $n = 30$.

(a) Define a critical region that has a significance level of $\alpha = 0.05$.

(b) If the random sample of $n = 30$ yielded $\bar{x} = 3189$ and $s = 488$, what is your conclusion?

(c) What is the approximate p-value of your test?

6.73. Let $Y_1 < Y_2 < \cdots < Y_5$ be the order statistics of a random sample of size 5 from the distribution having p.d.f. $f(x) = \exp[-(x - \theta)/\beta]/\beta$, $\theta < x < \infty$, zero elsewhere. Discuss the construction of a 90 percent confidence interval for β if θ is known.

6.74. Three independent random samples, each of size 6, are drawn from three normal distributions having common unknown variance. We find the three sample variances to be 10, 14, and 8, respectively.

(a) Compute an unbiased estimate of the common variance.

(b) Determine a 90 percent confidence interval for the common variance.

6.75. Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$.

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

(b) If c is given constant, find the m.l.e. of $\Pr(X \leq c)$.

6.76. Let $\bar{X}_1$, $\bar{X}_2$, and $\bar{X}_3$ and S_1^2, S_2^2, and S_3^2 denote the means and the variances of three independent random samples, each of size 10, from a normal distribution with mean μ and variance σ^2. Find the constant c so that

$$\Pr\left(\frac{\bar{X}_1 + \bar{X}_2 - 2\bar{X}_3}{\sqrt{10S_1^2 + 10S_2^2 + 10S_3^2}} \leq c\right) = 0.95.$$

6.77. Let Y be $b(192, p)$. We reject $H_0 : p = 0.75$ and accept $H_1 : p > 0.75$ if and only if $Y \geq 152$. Use the normal approximation to determine:

(a) $\alpha = \Pr(Y \geq 152; p = 0.75)$.

(b) $\beta = \Pr(Y < 152)$ when $p = 0.80$.

6.78. Let Y be $b(100, p)$. To test $H_0 : p = 0.08$ against $H_1 : p < 0.08$, we reject H_0 and accept H_1 if and only if $Y \leq 6$.

(a) Determine the significance level α of the test.
(b) Find the probability of the type II error if in fact $p = 0.04$.

6.79. Let $X_1, X_2, \ldots, X_n$ be a random sample from a Bernoulli distribution with parameter p. If p is restricted so that we know that $\frac{1}{2} \leq p \leq 1$, find the m.l.e. of this parameter.

6.80. Consider two Bernoulli distributions with unknown parameters p_1 and p_2, respectively. If Y and Z equal the numbers of successes in two independent random samples, each of sample size n, from the respective distributions, determine the maximum likelihood estimators of p_1 and p_2 if we know that $0 \leq p_1 \leq p_2 \leq 1$.

6.81. Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be n i.i.d. pairs of random variables, each with the bivariate normal distribution having five parameters μ_1, μ_2, σ_1^2, σ_2^2, and ρ.
(a) Show that $Z_i = X_i - Y_i$ is $N(\mu, \sigma^2)$, where $\mu = \mu_1 - \mu_2$ and $\sigma^2 = \sigma_1^2 - 2\rho\sigma_1\sigma_2 + \sigma_2^2$, $i = 1, 2, \ldots, n$.
(b) Since all five parameters are unknown, μ and σ^2 are unknown. To test $H_0 : \mu = 0$ ($H_0 : \mu_1 = \mu_2$) against $H_1 : \mu > 0$ ($H_1 : \mu_1 > \mu_2$), construct a t-test based upon the mean and the variance of the n differences $Z_1, Z_2, \ldots, Z_n$. This is often called a *paired t-test*.

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-