CHAPTER 9

Theory of Statistical Tests

9.1 Certain Best Tests

In Chapter 6 we introduced many concepts associated with tests of statistical hypotheses. In this chapter we consider some methods of constructing good statistical tests, beginning with testing a simple hypothesis H_0 against a simple alternative hypothesis H_1. Thus, in all instances, the parameter space is a set that consists of exactly two points. Under this restriction, we shall do three things:

1. Define a best test for testing H_0 against H_1.
2. Prove a theorem that provides a method of determining a best test.
3. Give two examples.

Before we define a best test, one important observation should be made. Certainly, a test specifies a critical region; but it can also be said that a choice of a critical region defines a test. For instance, if one is given the critical region $C = \{(x_1, x_2, x_3) : x_1^2 + x_2^2 + x_3^2 \geq 1\}$, the test is determined: Three random variables X_1, X_2, X_3 are to be considered; if the observed values are x_1, x_2, x_3, accept H_0 if

$x_1^2 + x_2^2 + x_3^2 < 1$; otherwise, reject H_0. That is, the terms "test" and "critical region" can, in this sense, be used interchangeably. Thus, if we define a best critical region, we have defined a best test.

Let $f(x; \theta)$ denote the p.d.f. of a random variable X. Let $X_1, X_2, \ldots, X_n$ denote a random sample from this distribution, and consider the two simple hypotheses $H_0 : \theta = \theta'$ and $H_1 : \theta = \theta''$. Thus $\Omega = \{\theta : \theta = \theta', \theta''\}$. We now define a best critical region (and hence a best test) for testing the simple hypothesis H_0 against the alternative simple hypothesis H_1. In this definition the symbols $\Pr[(X_1, X_2, \ldots, X_n) \in C; H_0]$ and $\Pr[(X_1, X_2, \ldots, X_n) \in C; H_1]$ mean $\Pr[(X_1, X_2, \ldots, X_n) \in C]$ when, respectively, H_0 and H_1 are true.

Definition 1. Let C denote a subset of the sample space. Then C is called a *best critical region* of size α for testing the simple hypothesis $H_0 : \theta = \theta'$ against the alternative simple hypothesis $H_1 : \theta = \theta''$ if, for every subset A of the sample space for which $\Pr[(X_1, \ldots, X_n) \in A; H_0] = \alpha$:

(a) $\Pr[(X_1, X_2, \ldots, X_n) \in C; H_0] = \alpha$.
(b) $\Pr[(X_1, X_2, \ldots, X_n) \in C; H_1] \geq \Pr[(X_1, X_2, \ldots, X_n) \in A; H_1]$.

This definition states, in effect, the following: First assume H_0 to be true. In general, there will be a multiplicity of subsets A of the sample space such that $\Pr[(X_1, X_2, \ldots, X_n) \in A] = \alpha$. Suppose that there is one of these subsets, say C, such that when H_1 is true, the power of the test associated with C is at least as great as the power of the test associated with each other A. Then C is defined as a best critical region of size α for testing H_0 against H_1.

In the following example we shall examine this definition in some detail and in a very simple case.

Example 1. Consider the one random variable X that has a binomial distribution with $n = 5$ and $p = \theta$. Let $f(x; \theta)$ denote the p.d.f. of X and let $H_0 : \theta = \frac{1}{2}$ and $H_1 : \theta = \frac{3}{4}$. The following tabulation gives, at points of positive probability density, the values of $f(x; \frac{1}{2})$, $f(x; \frac{3}{4})$, and the ratio $f(x; \frac{1}{2})/f(x; \frac{3}{4})$.

x	0	1	2	3	4	5
$f(x; \frac{1}{2})$	$\frac{1}{32}$	$\frac{5}{32}$	$\frac{10}{32}$	$\frac{10}{32}$	$\frac{5}{32}$	$\frac{1}{32}$
$f(x; \frac{3}{4})$	$\frac{1}{1024}$	$\frac{15}{1024}$	$\frac{90}{1024}$	$\frac{270}{1024}$	$\frac{405}{1024}$	$\frac{243}{1024}$
$f(x; \frac{1}{2})/f(x; \frac{3}{4})$	32	$\frac{32}{3}$	$\frac{32}{9}$	$\frac{32}{27}$	$\frac{32}{81}$	$\frac{32}{243}$

We shall use one random value of X to test the simple hypothesis $H_0 : \theta = \frac{1}{2}$ against the alternative simple hypothesis $H_1 : \theta = \frac{3}{4}$, and we shall first assign the significance level of the test to be $\alpha = \frac{1}{32}$. We seek a best critical region of size $\alpha = \frac{1}{32}$. If $A_1 = \{x : x = 0\}$ and $A_2 = \{x : x = 5\}$, then $\Pr(X \in A_1; H_0) = \Pr(X \in A_2; H_0) = \frac{1}{32}$ and there is no other subset A_3 of the space $\{x : x = 0, 1, 2, 3, 4, 5\}$ such that $\Pr(X \in A_3; H_0) = \frac{1}{32}$. Then either A_1 or A_2 is the best critical region C of size $\alpha = \frac{1}{32}$ for testing H_0 against H_1. We note that $\Pr(X \in A_1; H_0) = \frac{1}{32}$ and that $\Pr(X \in A_1; H_1) = \frac{1}{1024}$. Thus, if the set A_1 is used as a critical region of size $\alpha = \frac{1}{32}$, we have the intolerable situation that the probability of rejecting H_0 when H_1 is true (H_0 is false) is much less than the probability of rejecting H_0 when H_0 is true.

On the other hand, if the set A_2 is used as a critical region, then $\Pr(X \in A_2; H_0) = \frac{1}{32}$ and $\Pr(X \in A_2; H_1) = \frac{243}{1024}$. That is, the probability of rejecting H_0 when H_1 is true is much greater than the probability of rejecting H_0 when H_0 is true. Certainly, this is a more desirable state of affairs, and actually A_2 is the best critical region of size $\alpha = \frac{1}{32}$. The latter statement follows from the fact that, when H_0 is true, there are but two subsets, A_1 and A_2, of the sample space, each of whose probability measure is $\frac{1}{32}$ and the fact that

$$\frac{243}{1024} = \Pr(X \in A_2; H_1) > \Pr(X \in A_1; H_1) = \frac{1}{1024}.$$

It should be noted, in this problem, that the best critical region $C = A_2$ of size $\alpha = \frac{1}{32}$ is found by including in C the point (or points) at which $f(x; \frac{1}{2})$ is *small* in comparison with $f(x; \frac{3}{4})$. This is seen to be true once it is observed that the ratio $f(x; \frac{1}{2})/f(x; \frac{3}{4})$ is a minimum at $x = 5$. Accordingly, the ratio $f(x; \frac{1}{2})/f(x; \frac{3}{4})$, which is given in the last line of the above tabulation, provides us with a precise tool by which to find a best critical region C for certain given values of α. To illustrate this, take $\alpha = \frac{6}{32}$. When H_0 is true, each of the subsets $\{x : x = 0, 1\}$, $\{x : x = 0, 4\}$, $\{x : x = 1, 5\}$, $\{x : x = 4, 5\}$ has probability measure $\frac{6}{32}$. By direct computation it is found that the best critical region of this size is $\{x : x = 4, 5\}$. This reflects the fact that the ratio $f(x; \frac{1}{2})/f(x; \frac{3}{4})$ has its two smallest values for $x = 4$ and $x = 5$. The power of this test, which has $\alpha = \frac{6}{32}$, is

$$\Pr(X = 4, 5; H_1) = \frac{405}{1024} + \frac{243}{1024} = \frac{648}{1024}.$$

The preceding example should make the following theorem, due to Neyman and Pearson, easier to understand. It is an important theorem because it provides a systematic method of determining a best critical region.

Neyman–Pearson Theorem. *Let $X_1, X_2, \ldots, X_n$, where n is a fixed positive integer, denote a random sample from a distribution that has* p.d.f. $f(x; \theta)$. *Then the joint* p.d.f. *of $X_1, X_2, \ldots, X_n$ is*

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

Let θ' and θ'' be distinct fixed values of θ so that $\Omega = \{\theta : \theta = \theta', \theta''\}$, and let k be a positive number. Let C be a subset of the sample space such that:

(a) $\dfrac{L(\theta'; x_1, x_2, \ldots, x_n)}{L(\theta''; x_1, x_2, \ldots, x_n)} \le k,$ *for each point* $(x_1, x_2, \ldots, x_n) \in C$.

(b) $\dfrac{L(\theta'; x_1, x_2, \ldots, x_n)}{L(\theta''; x_1, x_2, \ldots, x_n)} \ge k,$ *for each point* $(x_1, x_2, \ldots, x_n) \in C^*$.

(c) $\alpha = \Pr[(X_1, X_2, \ldots, X_n) \in C; H_0]$.

Then C is a best critical region of size α for testing the simple hypothesis $H_0 : \theta = \theta'$ against the alternative simple hypothesis $H_1 : \theta = \theta''$.

Proof. We shall give the proof when the random variables are of the continuous type. If C is the only critical region of size α, the theorem is proved. If there is another critical region of size α, denote it by A. For convenience, we shall let $\int \cdots_R \int L(\theta; x_1, \ldots, x_n)\, dx_1 \cdots dx_n$ be denoted by $\int_R L(\theta)$. In this notation we wish to show that

$$\int_C L(\theta'') - \int_A L(\theta'') \ge 0.$$

Since C is the union of the disjoint sets $C \cap A$ and $C \cap A^*$ and A is the union of the disjoint sets $A \cap C$ and $A \cap C^*$, we have

$$\begin{aligned} \int_C L(\theta'') - \int_A L(\theta'') &= \int_{C \cap A} L(\theta'') + \int_{C \cap A^*} L(\theta'') - \int_{A \cap C} L(\theta'') - \int_{A \cap C^*} L(\theta'') \\ &= \int_{C \cap A^*} L(\theta'') - \int_{A \cap C^*} L(\theta''). \qquad (1) \end{aligned}$$

However, by the hypothesis of the theorem, $L(\theta'') \ge (1/k)L(\theta')$ at each point of C, and hence at each point of $C \cap A^*$; thus

$$\int_{C \cap A^*} L(\theta'') \ge \frac{1}{k} \int_{C \cap A^*} L(\theta').$$

But $L(\theta'') \le (1/k)L(\theta')$ at each point of C^*, and hence at each point of $A \cap C^*$; accordingly,

$$\int_{A\cap C^*} L(\theta'') \le \frac{1}{k} \int_{A\cap C^*} L(\theta').$$

These inequalities imply that

$$\int_{C\cap A^*} L(\theta'') - \int_{A\cap C^*} L(\theta'') \ge \frac{1}{k} \int_{C\cap A^*} L(\theta') - \frac{1}{k} \int_{A\cap C^*} L(\theta');$$

and, from Equation (1), we obtain

$$\int_{C} L(\theta'') - \int_{A} L(\theta'') \ge \frac{1}{k} \left[\int_{C\cap A^*} L(\theta') - \int_{A\cap C^*} L(\theta') \right]. \qquad (2)$$

However,

$$\begin{aligned} &\int_{C\cap A^*} L(\theta') - \int_{A\cap C^*} L(\theta') \\ &= \int_{C\cap A^*} L(\theta') + \int_{C\cap A} L(\theta') - \int_{A\cap C} L(\theta') - \int_{A\cap C^*} L(\theta') \\ &= \int_{C} L(\theta') - \int_{A} L(\theta') \\ &= \alpha - \alpha = 0. \end{aligned}$$

If this result is substituted in inequality (2), we obtain the desired result,

$$\int_{C} L(\theta'') - \int_{A} L(\theta'') \ge 0.$$

If the random variables are of the discrete type, the proof is the same, with integration replaced by summation.

Remark. As stated in the theorem, conditions (a), (b), and (c) are sufficient ones for region C to be a best critical region of size α. However, they are also necessary. We discuss this briefly. Suppose there is a region A of size α that does not satisfy (a) and (b) and that is as powerful at $\theta = \theta''$ as C, which satisfies (a), (b), and (c). Then expression (1) would be zero, since the power at θ'' using A is equal to that using C. It can be proved that to have expression (1) equal zero A must be of the same form as C. As a matter of fact, in the continuous case, A and C would essentially be the same region; that is, they could differ only by a set having probability zero. However, in the discrete case, if $\Pr[L(\theta') = kL(\theta''); H_0]$ is positive, A and C could be different sets, but each would necessarily enjoy conditions (a), (b), and (c) to be a best critical region of size α.

One aspect of the theorem to be emphasized is that if we take C to be the set of all points $(x_1, x_2, \ldots, x_n)$ which satisfy

$$\frac{L(\theta'; x_1, x_2, \ldots, x_n)}{L(\theta''; x_1, x_2, \ldots, x_n)} \le k, \qquad k > 0,$$

then, in accordance with the theorem, C will be a best critical region. This inequality can frequently be expressed in one of the forms (where c_1 and c_2 are constants)

$$u_1(x_1, x_2, \ldots, x_n; \theta', \theta'') \le c_1,$$

or

$$u_2(x_1, x_2, \ldots, x_n; \theta', \theta'') \ge c_2.$$

Suppose that it is the first form, $u_1 \le c_1$. Since θ' and θ'' are given constants, $u_1(X_1, X_2, \ldots, X_n; \theta', \theta'')$ is a statistic; and if the p.d.f. of this statistic can be found when H_0 is true, then the significance level of the test of H_0 against H_1 can be determined from this distribution. That is,

$$\alpha = \Pr [u_1(X_1, X_2, \ldots, X_n; \theta', \theta'') \le c_1; H_0].$$

Moreover, the test may be based on this statistic; for, if the observed values of $X_1, X_2, \ldots, X_n$ are $x_1, x_2, \ldots, x_n$, we reject H_0 (accept H_1) if $u_1(x_1, x_2, \ldots, x_n) \le c_1$.

A positive number k determines a best critical region C whose size is $\alpha = \Pr [(X_1, X_2, \ldots, X_n) \in C; H_0]$ for that particular k. It may be that this value of α is unsuitable for the purpose at hand; that is, it is too large or too small. However, if there is a statistic $u_1(X_1, X_2, \ldots, X_n)$, as in the preceding paragraph, whose p.d.f. can be determined when H_0 is true, we need not experiment with various values of k to obtain a desirable significance level. For if the distribution of the statistic is known, or can be found, we may determine c_1 such that $\Pr [u_1(X_1, X_2, \ldots, X_n) \le c_1; H_0]$ is a desirable significance level.

An illustrative example follows.

Example 2. Let $X_1, X_2, \ldots, X_n$ denote a random sample from the distribution that has the p.d.f.

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

It is desired to test the simple hypothesis $H_0 : \theta = \theta' = 0$ against the alternative simple hypothesis $H_1 : \theta = \theta'' = 1$. Now

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

$$= \exp\left(-\sum_1^n x_i + \frac{n}{2}\right).$$

If $k > 0$, the set of all points $(x_1, x_2, \dots, x_n)$ such that

$$\exp\left(-\sum_1^n x_i + \frac{n}{2}\right) \le k$$

is a best critical region. This inequality holds if and only if

$$-\sum_1^n x_i + \frac{n}{2} \le \ln k$$

or, equivalently,

$$\sum_1^n x_i \ge \frac{n}{2} - \ln k = c.$$

In this case, a best critical region is the set $C = \left\{(x_1, x_2, \dots, x_n): \sum_1^n x_i \ge c\right\}$, where c is a constant that can be determined so that the size of the critical region is a desired number α. The event $\sum_1^n X_i \ge c$ is equivalent to the event $\bar{X} \ge c/n = c_1$, say, so the test may be based upon the statistic $\bar{X}$. If H_0 is true, that is, $\theta = \theta' = 0$, then $\bar{X}$ has a distribution that is $N(0, 1/n)$. For a given positive integer n, the size of the sample, and a given significance level α, the number c_1 can be found from Table III in Appendix B, so that $\Pr(\bar{X} \ge c_1; H_0) = \alpha$. Hence, if the experimental values of $X_1, X_2, \dots, X_n$ were, respectively, $x_1, x_2, \dots, x_n$, we would compute $\bar{x} = \sum_1^n x_i/n$. If $\bar{x} \ge c_1$, the simple hypothesis $H_0 : \theta = \theta' = 0$ would be rejected at the significance level α; if $\bar{x} < c_1$, the hypothesis H_0 would be accepted. The probability of rejecting H_0, when H_0 is true, is α; the probability of rejecting H_0, when H_0 is false, is the value of the power of the test at $\theta = \theta'' = 1$. That is,

$$\Pr(\bar{X} \ge c_1; H_1) = \int_{c_1}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{1/n}} \exp\left[-\frac{(\bar{x}-1)^2}{2(1/n)}\right] d\bar{x}.$$

For example, if $n = 25$ and if α is selected to be 0.05, then from Table III we find that $c_1 = 1.645/\sqrt{25} = 0.329$. Thus the power of this best test of H_0 against H_1 is 0.05, when H_0 is true, and is

$$\int_{0.329}^{\infty} \frac{1}{\sqrt{2\pi}\sqrt{\frac{1}{25}}} \exp\left[-\frac{(\bar{x}-1)^2}{2(\frac{1}{25})}\right] d\bar{x} = \int_{-3.355}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-w^2/2}\, dw = 0.999+,$$

when H_1 is true.

There is another aspect of this theorem that warrants special mention. It has to do with the number of parameters that appear in the p.d.f. Our notation suggests that there is but one parameter. However, a careful review of the proof will reveal that nowhere was this needed or assumed. The p.d.f. may depend upon any finite number of parameters. What is essential is that the hypothesis H_0 and the alternative hypothesis H_1 be simple, namely that they completely specify the distributions. With this in mind, we see that the simple hypotheses H_0 and H_1 do not need to be hypotheses about the parameters of a distribution, nor, as a matter of fact, do the random variables $X_1, X_2, \ldots, X_n$ need to be independent. That is, if H_0 is the simple hypothesis that the joint p.d.f. is $g(x_1, x_2, \ldots, x_n)$, and if H_1 is the alternative simple hypothesis that the joint p.d.f. is $h(x_1, x_2, \ldots, x_n)$, then C is a best critical region of size α for testing H_0 against H_1 if, for $k > 0$:

1′. $\dfrac{g(x_1, x_2, \ldots, x_n)}{h(x_1, x_2, \ldots, x_n)} \le k$ for $(x_1, x_2, \ldots, x_n) \in C$.

2′. $\dfrac{g(x_1, x_2, \ldots, x_n)}{h(x_1, x_2, \ldots, x_n)} \ge k$ for $(x_1, x_2, \ldots, x_n) \in C^*$.

3′. $\alpha = \Pr[(X_1, X_2, \ldots, X_n) \in C; H_0]$.

An illustrative example follows.

Example 3. Let $X_1, \ldots, X_n$ denote a random sample from a distribution which has a p.d.f. $f(x)$ that is positive on and only on the nonnegative integers. It is desired to test the simple hypothesis

$$H_0: f(x) = \frac{e^{-1}}{x!}, \qquad x = 0, 1, 2, \ldots,$$
$$= 0 \qquad \text{elsewhere},$$

against the alternative simple hypothesis

$$H_1: f(x) = (\tfrac{1}{2})^{x+1}, \qquad x = 0, 1, 2, \ldots,$$
$$= 0 \qquad \text{elsewhere}.$$

Here

$$\frac{g(x_1, \ldots, x_n)}{h(x_1, \ldots, x_n)} = \frac{e^{-n}/(x_1!\, x_2! \cdots x_n!)}{(\frac{1}{2})^n(\frac{1}{2})^{x_1+x_2+\cdots+x_n}}$$

$$= \frac{(2e^{-1})^n 2^{\Sigma x_i}}{\prod_1^n (x_i!)}.$$

If $k > 0$, the set of points $(x_1, x_2, \ldots, x_n)$ such that

$$\left(\sum_1^n x_i\right)\ln 2 - \ln\left[\prod_1^n (x_i!)\right] \le \ln k - n\ln(2e^{-1}) = c$$

is a best critical region C. Consider the case of $k = 1$ and $n = 1$. The preceding inequality may be written $2^{x_1}/x_1! \le e/2$. This inequality is satisfied by all points in the set $C = \{x_1 : x_1 = 0, 3, 4, 5, \ldots\}$. Thus the power of the test when H_0 is true is

$$\Pr(X_1 \in C; H_0) = 1 - \Pr(X_1 = 1, 2; H_0) = 0.448,$$

approximately, in accordance with Table I of Appendix B. The power of the test when H_1 is true is given by

$$\Pr(X_1 \in C; H_1) = 1 - \Pr(X_1 = 1, 2; H_1)$$

$$= 1 - (\tfrac{1}{4} + \tfrac{1}{8}) = 0.625.$$

Remark. In the notation of this section, say C is a critical region such that

$$\alpha = \int_C L(\theta') \quad \text{and} \quad \beta = \int_{C^*} L(\theta''),$$

so that here α and β equal the respective probabilities of the type I and type II errors associated with C. Let d_1 and d_2 be two given positive constants. Consider a certain linear function of α and β, namely

$$d_1 \int_C L(\theta') + d_2 \int_{C^*} L(\theta'') = d_1 \int_C L(\theta') + d_2\left[1 - \int_C L(\theta'')\right]$$

$$= d_2 + \int_C [d_1 L(\theta') - d_2 L(\theta'')].$$

If we wished to minimize this expression, we would select C to be the set of all $(x_1, x_2, \ldots, x_n)$ such that

$$d_1 L(\theta') - d_2 L(\theta'') < 0$$

or, equivalently,

$$\frac{L(\theta')}{L(\theta'')} < \frac{d_2}{d_1}, \qquad \text{for all } (x_1, x_2, \ldots, x_n) \in C,$$

which according to the Neyman–Pearson theorem provides a best critical region with $k = d_2/d_1$. That is, this critical region C is one that minimizes $d_1\alpha + d_2\beta$. There could be others, for example, including points on which $L(\theta')/L(\theta'') = d_2/d_1$, but these would still be best critical regions according to the Neyman–Pearson theorem.

EXERCISES

9.1. In Example 2 of this section, let the simple hypotheses read $H_0 : \theta = \theta' = 0$ and $H_1 : \theta = \theta'' = -1$. Show that the best test of H_0 against H_1 may be carried out by use of the statistic $\bar{X}$, and that if $n = 25$ and $\alpha = 0.05$, the power of the test is 0.999+ when H_1 is true.

9.2. Let the random variable X have the p.d.f. $f(x; \theta) = (1/\theta)e^{-x/\theta}$, $0 < x < \infty$, zero elsewhere. Consider the simple hypothesis $H_0 : \theta = \theta' = 2$ and the alternative hypothesis $H_1 : \theta = \theta'' = 4$. Let X_1, X_2 denote a random sample of size 2 from this distribution. Show that the best test of H_0 against H_1 may be carried out by use of the statistic $X_1 + X_2$ and that the assertion in Example 2 of Section 6.4 is correct.

9.3. Repeat Exercise 9.2 when $H_1 : \theta = \theta'' = 6$. Generalize this for every $\theta'' > 2$.

9.4. Let $X_1, X_2, \ldots, X_{10}$ be a random sample of size 10 from a normal distribution $N(0, \sigma^2)$. Find a best critical region of size $\alpha = 0.05$ for testing $H_0 : \sigma^2 = 1$ against $H_1 : \sigma^2 = 2$. Is this a best critical region of size 0.05 for testing $H_0 : \sigma^2 = 1$ against $H_1 : \sigma^2 = 4$? Against $H_1 : \sigma^2 = \sigma_1^2 > 1$?

9.5. If $X_1, X_2, \ldots, X_n$ is a random sample from a distribution having p.d.f. of the form $f(x; \theta) = \theta x^{\theta - 1}$, $0 < x < 1$, zero elsewhere, show that a best critical region for testing $H_0 : \theta = 1$ against $H_1 : \theta = 2$ is

$$C = \left\{(x_1, x_2, \ldots, x_n) : c \le \prod_{i=1}^{n} x_i\right\}.$$

9.6. Let $X_1, X_2, \ldots, X_{10}$ be a random sample from a distribution that is $N(\theta_1, \theta_2)$. Find a best test of the simple hypothesis $H_0 : \theta_1 = \theta_1' = 0$, $\theta_2 = \theta_2' = 1$ against the alternative simple hypothesis $H_1 : \theta_1 = \theta_1'' = 1$, $\theta_2 = \theta_2'' = 4$.

9.7. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a normal distribution $N(\theta, 100)$. Show that $C = \left\{(x_1, x_2, \ldots, x_n) : c \le \bar{x} = \sum_{1}^{n} x_i/n\right\}$ is a best critical region for testing $H_0 : \theta = 75$ against $H_1 : \theta = 78$. Find n and c so that

$$\Pr[(X_1, X_2, \ldots, X_n) \in C; H_0] = \Pr(\bar{X} \ge c; H_0) = 0.05$$

and

$$\Pr[(X_1, X_2, \ldots, X_n) \in C; H_1] = \Pr(\bar{X} \ge c; H_1) = 0.90, \text{ approximately.}$$

9.8. If $X_1, X_2, \ldots, X_n$ is a random sample from a beta distribution with parameters $\alpha = \beta = \theta > 0$, find a best critical region for testing $H_0: \theta = 1$ against $H_1: \theta = 2$.

9.9. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution having the p.d.f. $f(x; p) = p^x(1-p)^{1-x}$, $x = 0, 1$, zero elsewhere. Show that $C = \left\{(x_1, \ldots, x_n) : \sum_1^n x_i \le c\right\}$ is a best critical region for testing $H_0: p = \frac{1}{2}$ against $H_1: p = \frac{1}{3}$. Use the central limit theorem to find n and c so that approximately $\Pr\left(\sum_1^n X_i \le c; H_0\right) = 0.10$ and $\Pr\left(\sum_1^n X_i \le c; H_1\right) = 0.80$.

9.10. Let $X_1, X_2, \ldots, X_{10}$ denote a random sample of size 10 from a Poisson distribution with mean θ. Show that the critical region C defined by $\sum_1^{10} x_i \ge 3$ is a best critical region for testing $H_0: \theta = 0.1$ against $H_1: \theta = 0.5$. Determine, for this test, the significance level α and the power at $\theta = 0.5$.

9.2 Uniformly Most Powerful Tests

This section will take up the problem of a test of a simple hypothesis H_0 against an alternative composite hypothesis H_1. We begin with an example.

Example 1. Consider the p.d.f.

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

of Example 2, Section 6.4, and later of Exercise 9.3. It is desired to test the simple hypothesis $H_0: \theta = 2$ against the alternative composite hypothesis $H_1: \theta > 2$. Thus $\Omega = \{\theta : \theta \ge 2\}$. A random sample, X_1, X_2, of size $n = 2$ will be used, and the critical region is $C = \{(x_1, x_2) : 9.5 \le x_1 + x_2 < \infty\}$. It was shown in the example cited that the significance level of the test is approximately 0.05 and that the power of the test when $\theta = 4$ is approximately 0.31. The power function $K(\theta)$ of the test for all $\theta \ge 2$ will now be obtained. We have

$$K(\theta) = 1 - \int_0^{9.5} \int_0^{9.5 - x_2} \frac{1}{\theta^2} \exp\left(-\frac{x_1 + x_2}{\theta}\right) dx_1\, dx_2$$
$$= \left(\frac{\theta + 9.5}{\theta}\right) e^{-9.5/\theta}, \qquad 2 \le \theta.$$

For example, $K(2) = 0.05$, $K(4) = 0.31$, and $K(9.5) = 2/e$. It is known (Exercise 9.3) that $C = \{(x_1, x_2) : 9.5 \le x_1 + x_2 < \infty\}$ is a best critical region of size 0.05 for testing the simple hypothesis $H_0 : \theta = 2$ against each simple hypothesis in the composite hypothesis $H_1 : \theta > 2$.

The preceding example affords an illustration of a test of a simple hypothesis H_0 that is a best test of H_0 against every simple hypothesis in the alternative composite hypothesis H_1. We now define a critical region, when it exists, which is a best critical region for testing a simple hypothesis H_0 against an alternative composite hypothesis H_1. It seems desirable that this critical region should be a best critical region for testing H_0 against each simple hypothesis in H_1. That is, the power function of the test that corresponds to this critical region should be at least as great as the power function of any other test with the same significance level for every simple hypothesis in H_1.

Definition 2. The critical region C is a *uniformly most powerful critical region* of size α for testing the simple hypothesis H_0 against an alternative composite hypothesis H_1 if the set C is a best critical region of size α for testing H_0 against each simple hypothesis in H_1. A test defined by this critical region C is called a *uniformly most powerful test*, with significance level α, for testing the simple hypothesis H_0 against the alternative composite hypothesis H_1.

As will be seen presently, uniformly most powerful tests do not always exist. However, when they do exist, the Neyman–Pearson theorem provides a technique for finding them. Some illustrative examples are given here.

Example 2. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that is $N(0, \theta)$, where the variance θ is an unknown positive number. It will be shown that there exists a uniformly most powerful test with significance level α for testing the simple hypothesis $H_0 : \theta = \theta'$, where θ' is a fixed positive number, against the alternative composite hypothesis $H_1 : \theta > \theta'$. Thus $\Omega = \{\theta : \theta \geq \theta'\}$. The joint p.d.f. of $X_1, X_2, \ldots, X_n$ is

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

Let θ'' represent a number greater than θ', and let k denote a positive number. Let C be the set of points where

$$\frac{L(\theta'; x_1, x_2, \ldots, x_n)}{L(\theta''; x_1, x_2, \ldots, x_n)} \leq k,$$

that is, the set of points where

$$\left(\frac{\theta''}{\theta'}\right)^{n/2} \exp\left[-\left(\frac{\theta''-\theta'}{2\theta'\theta''}\right)\sum_1^n x_i^2\right] \le k$$

or, equivalently,

$$\sum_1^n x_i^2 \ge \frac{2\theta'\theta''}{\theta''-\theta'}\left[\frac{n}{2}\ln\left(\frac{\theta''}{\theta'}\right) - \ln k\right] = c.$$

The set $C = \left\{(x_1, x_2, \dots, x_n) : \sum_1^n x_i^2 \ge c\right\}$ is then a best critical region for testing the simple hypothesis $H_0 : \theta = \theta'$ against the simple hypothesis $\theta = \theta''$. It remains to determine c, so that this critical region has the desired size α. If H_0 is true, the random variable $\sum_1^n X_i^2/\theta'$ has a chi-square distribution with n degrees of freedom. Since $\alpha = \Pr\left(\sum_1^n X_i^2/\theta' \ge c/\theta'; H_0\right)$, c/θ' may be read from Table II in Appendix B and c determined. Then $C = \left\{(x_1, x_2, \dots, x_n) : \sum_1^n x_i^2 \ge c\right\}$ is a best critical region of size α for testing $H_0 : \theta = \theta'$ against the hypothesis $\theta = \theta''$. Moreover, for each number θ'' greater than θ', the foregoing argument holds. That is, if θ''' is another number greater than θ', then $C = \left\{(x_1, \dots, x_n) : \sum_1^n x_i^2 \ge c\right\}$ is a best critical region of size α for testing $H_0 : \theta = \theta'$ against the hypothesis $\theta = \theta'''$. Accordingly, $C = \left\{(x_1, \dots, x_n) : \sum_1^n x_i^2 \ge c\right\}$ is a uniformly most powerful critical region of size α for testing $H_0 : \theta = \theta'$ against $H_1 : \theta > \theta'$. If $x_1, x_2, \dots, x_n$ denote the experimental values of $X_1, X_2, \dots, X_n$, then $H_0 : \theta = \theta'$ is rejected at the significance level α, and $H_1 : \theta > \theta'$ is accepted, if $\sum_1^n x_i^2 \ge c$; otherwise, $H_0 : \theta = \theta'$ is accepted.

If, in the preceding discussion, we take $n = 15$, $\alpha = 0.05$, and $\theta' = 3$, then here the two hypotheses will be $H_0 : \theta = 3$ and $H_1 : \theta > 3$. From Table II, $c/3 = 25$ and hence $c = 75$.

Example 3. Let $X_1, X_2, \dots, X_n$ denote a random sample from a distribution that is $N(\theta, 1)$, where the mean θ is unknown. It will be shown that there is no uniformly most powerful test of the simple hypothesis $H_0 : \theta = \theta'$, where θ' is a fixed number, against the alternative composite hypothesis $H_1 : \theta \ne \theta'$. Thus $\Omega = \{\theta : -\infty < \theta < \infty\}$. Let θ'' be a number not equal to θ'. Let k be a positive number and consider

$$\frac{(1/2\pi)^{n/2}\exp\left[-\sum_1^n (x_i-\theta')^2/2\right]}{(1/2\pi)^{n/2}\exp\left[-\sum_1^n (x_i-\theta'')^2/2\right]} \le k.$$

The preceding inequality may be written as

$$\exp\left\{-(\theta''-\theta')\sum_1^n x_i + \frac{n}{2}[(\theta'')^2-(\theta')^2]\right\} \le k$$

or

$$(\theta''-\theta')\sum_1^n x_i \ge \frac{n}{2}[(\theta'')^2-(\theta')^2] - \ln k.$$

This last inequality is equivalent to

$$\sum_1^n x_i \ge \frac{n}{2}(\theta''+\theta') - \frac{\ln k}{\theta''-\theta'},$$

provided that $\theta'' > \theta'$, and it is equivalent to

$$\sum_1^n x_i \le \frac{n}{2}(\theta''+\theta') - \frac{\ln k}{\theta''-\theta'}$$

if $\theta'' < \theta'$. The first of these two expressions defines a best critical region for testing $H_0 : \theta = \theta'$ against the hypothesis $\theta = \theta''$ provided that $\theta'' > \theta'$, while the second expression defines a best critical region for testing $H_0 : \theta = \theta'$ against the hypothesis $\theta = \theta''$ provided that $\theta'' < \theta'$. That is, a best critical region for testing the simple hypothesis against an alternative simple hypothesis, say $\theta = \theta' + 1$, will not serve as a best critical region for testing $H_0 : \theta = \theta'$ against the alternative simple hypothesis $\theta = \theta' - 1$, say. By definition, then, there is no uniformly most powerful test in the case under consideration.

It should be noted that had the alternative composite hypothesis been either $H_1 : \theta > \theta'$ or $H_1 : \theta < \theta'$, a uniformly most powerful test would exist in each instance.

Example 4. In Exercise 9.10 the reader was asked to show that if a random sample of size $n = 10$ is taken from a Poisson distribution with mean θ, the critical region defined by $\sum_1^{10} x_i \ge 3$ is a best critical region for testing $H_0 : \theta = 0.1$ against $H_1 : \theta = 0.5$. This critical region is also a uniformly most powerful one for testing $H_0 : \theta = 0.1$ against $H_1 : \theta > 0.1$ because, with $\theta'' > 0.1$,

$$\frac{(0.1)^{\Sigma x_i}e^{-10(0.1)}/(x_1!\,x_2!\cdots x_n!)}{(\theta'')^{\Sigma x_i}e^{-10(\theta'')}/(x_1!\,x_2!\cdots x_n!)} \le k$$

is equivalent to

$$\left(\frac{0.1}{\theta''}\right)^{\Sigma x_i} e^{-10(0.1-\theta'')} \le k.$$

The preceding inequality may be written as

$$\left(\sum_1^n x_i\right)(\ln 0.1 - \ln \theta'') \le \ln k + 10(0.1 - \theta'')$$

or, since $\theta'' > 0.1$, equivalently as

$$\sum_1^n x_i \ge \frac{\ln k + 1 - 10\theta''}{\ln 0.1 - \ln \theta''}.$$

Of course, $\sum_1^{10} x_i \ge 3$ is of the latter form.

Let us make an observation, although obvious when pointed out, that is important. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a distribution that has p.d.f. $f(x; \theta)$, $\theta \in \Omega$. Suppose that $Y = u(X_1, X_2, \ldots, X_n)$ is a sufficient statistic for θ. In accordance with the factorization theorem, the joint p.d.f. of $X_1, X_2, \ldots, X_n$ may be written

$$L(\theta; x_1, x_2, \ldots, x_n) = k_1[u(x_1, x_2, \ldots, x_n); \theta]k_2(x_1, x_2, \ldots, x_n),$$

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

$$\frac{L(\theta'; x_1, x_2, \ldots, x_n)}{L(\theta''; x_1, x_2, \ldots, x_n)} = \frac{k_1[u(x_1, x_2, \ldots, x_n); \theta']}{k_1[u(x_1, x_2, \ldots, x_n); \theta'']}$$

depends upon $x_1, x_2, \ldots, x_n$ only through $u(x_1, x_2, \ldots, x_n)$. Accordingly, if there is a sufficient statistic $Y = u(X_1, X_2, \ldots, X_n)$ for θ and if a best test or a uniformly most powerful test is desired, there is no need to consider tests which are based upon any statistic other than the sufficient statistic. This result supports the importance of sufficiency.

Often, when $\theta'' < \theta'$ the ratio

$$\frac{L(\theta'; x_1, x_2, \ldots, x_n)}{L(\theta''; x_1, x_2, \ldots, x_n)},$$

which depends upon $x_1, x_2, \ldots, x_n$ only through $y = u(x_1, x_2, \ldots, x_n)$, is an increasing function of $y = u(x_1, x_2, \ldots, x_n)$. In such a case we say that we have a *monotone likelihood ratio* in the statistic $Y = u(X_1, X_2, \ldots, X_n)$.

Example 5. Let $X_1, X_2, \ldots, X_n$ be a random sample from a Bernoulli distribution with parameter $p = \theta$, where $0 < \theta < 1$. Let $\theta'' < \theta'$. Then the ratio

$$\frac{L(\theta'; x_1, x_2, \ldots, x_n)}{L(\theta''; x_1, x_2, \ldots, x_n)} = \frac{(\theta')^{\Sigma x_i}(1-\theta')^{n-\Sigma x_i}}{(\theta'')^{\Sigma x_i}(1-\theta'')^{n-\Sigma x_i}} = \left[\frac{\theta'(1-\theta'')}{\theta''(1-\theta')}\right]^{\Sigma x_i}\left(\frac{1-\theta'}{1-\theta''}\right)^n.$$

Since $\theta'/\theta'' > 1$ and $(1-\theta'')/(1-\theta') > 1$, so that $\theta'(1-\theta'')/\theta''(1-\theta') > 1$, the ratio is an increasing function of $y = \Sigma x_i$. Thus we have a monotone likelihood ratio in the statistic $Y = \Sigma X_i$.

We can generalize Example 5 by noting the following. Suppose that the random sample $X_1, X_2, \ldots, X_n$ arises from a p.d.f. representing a regular case of the exponential class, namely

$$\begin{aligned} f(x; \theta) &= \exp\left[p(\theta)K(x) + S(x) + q(\theta)\right], \qquad x \in \mathscr{A}, \\ &= 0 \qquad \text{elsewhere,} \end{aligned}$$

where the space $\mathscr{A}$ of X is free of θ. Further assume that $p(\theta)$ is an increasing function of θ. Then

$$\begin{aligned} \frac{L(\theta')}{L(\theta'')} &= \frac{\exp\left[p(\theta')\sum_{i=1}^{n} K(x_i) + \sum_{i=1}^{n} S(x_i) + nq(\theta')\right]}{\exp\left[p(\theta'')\sum_{i=1}^{n} K(x_i) + \sum_{i=1}^{n} S(x_i) + nq(\theta'')\right]} \\ &= \exp\left\{[p(\theta') - p(\theta'')]\sum_{i=1}^{n} K(x_i) + n[q(\theta') - q(\theta'')]\right\}. \end{aligned}$$

If $\theta'' < \theta'$, $p(\theta)$ being an increasing function requires this ratio to be an increasing function of $y = \sum_{i=1}^{n} K(x_i)$. Thus we have a monotone likelihood ratio in the statistic $Y = \sum_{i=1}^{n} K(X_i)$. Moreover, if we test $H_0: \theta = \theta'$ against $H_1: \theta < \theta'$, then, with $\theta'' < \theta'$, we see that

$$\frac{L(\theta')}{L(\theta'')} \le k$$

is equivalent to $\Sigma K(x_i) \le c$ for every $\theta'' < \theta'$. That is, this provides a uniformly most powerful critical region.

If, in the preceding situation with monotone likelihood ratio, we test $H_0: \theta = \theta'$ against $H_1: \theta > \theta'$, then $\Sigma K(x_i) \ge c$ would be a uniformly most powerful critical region. From the likelihood ratios displayed in Examples 2, 3, 4, and 5 we see immediately that the respective critical regions

$$\sum_{i=1}^{n} x_i^2 \geq c, \qquad \sum_{i=1}^{n} x_i \geq c, \qquad \sum_{i=1}^{n} x_i \geq c, \qquad \sum_{i=1}^{n} x_i \geq c$$

are uniformly most powerful for testing $H_0 : \theta = \theta'$ against $H_1 : \theta > \theta'$.

There is a final remark that should be made about uniformly most powerful tests. Of course, in Definition 2, the word *uniformly* is associated with θ; that is, C is a best critical region of size α for testing $H_0 : \theta = \theta_0$ against all θ values given by the composite alternative H_1. However, suppose that the form of such a region is

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

Then this form provides uniformly most powerful critical regions for all attainable α values by, of course, appropriately changing the value of c. That is, there is a certain uniformity property, also associated with α, that is not always noted in statistics texts.

EXERCISES

9.11. Let X have the p.d.f. $f(x; \theta) = \theta^x(1 - \theta)^{1-x}$, $x = 0, 1$, zero elsewhere. We test the simple hypothesis $H_0 : \theta = \frac{1}{4}$ against the alternative composite hypothesis $H_1 : \theta < \frac{1}{4}$ by taking a random sample of size 10 and rejecting $H_0 : \theta = \frac{1}{4}$ if and only if the observed values $x_1, x_2, \ldots, x_{10}$ of the sample observations are such that $\sum_{1}^{10} x_i \leq 1$. Find the power function $K(\theta)$, $0 < \theta \leq \frac{1}{4}$, of this test.

9.12. Let X have a p.d.f. of the form $f(x; \theta) = 1/\theta, 0 < x < \theta$, zero elsewhere. Let $Y_1 < Y_2 < Y_3 < Y_4$ denote the order statistics of a random sample of size 4 from this distribution. Let the observed value of Y_4 be y_4. We reject $H_0 : \theta = 1$ and accept $H_1 : \theta \neq 1$ if either $y_4 \leq \frac{1}{2}$ or $y_4 \geq 1$. Find the power function $K(\theta)$, $0 < \theta$, of the test.

9.13. Consider a normal distribution of the form $N(\theta, 4)$. The simple hypothesis $H_0 : \theta = 0$ is rejected, and the alternative composite hypothesis $H_1 : \theta > 0$ is accepted if and only if the observed mean $\bar{x}$ of a random sample of size 25 is greater than or equal to $\frac{3}{5}$. Find the power function $K(\theta)$, $0 \leq \theta$, of this test.

9.14. Consider the two normal distributions $N(\mu_1, 400)$ and $N(\mu_2, 225)$. Let $\theta = \mu_1 - \mu_2$. Let $\bar{x}$ and $\bar{y}$ denote the observed means of two independent random samples, each of size n, from these two distributions. We reject $H_0 : \theta = 0$ and accept $H_1 : \theta > 0$ if and only if $\bar{x} - \bar{y} \geq c$. If $K(\theta)$ is the power function of this test, find n and c so that $K(0) = 0.05$ and $K(10) = 0.90$, approximately.

9.15. If, in Example 2 of this section, $H_0 : \theta = \theta'$, where θ' is a fixed positive number, and $H_1 : \theta < \theta'$, show that the set $\left\{(x_1, x_2, \ldots, x_n) : \sum_1^n x_i^2 \le c\right\}$ is a uniformly most powerful critical region for testing H_0 against H_1.

9.16. If, in Example 2 of this section, $H_0 : \theta = \theta'$, where θ' is a fixed positive number, and $H_1 : \theta \neq \theta'$, show that there is no uniformly most powerful test for testing H_0 against H_1.

9.17. Let $X_1, X_2, \ldots, X_{25}$ denote a random sample of size 25 from a normal distribution $N(\theta, 100)$. Find a uniformly most powerful critical region of size $\alpha = 0.10$ for testing $H_0 : \theta = 75$ against $H_1 : \theta > 75$.

9.18. Let $X_1, X_2, \ldots, X_n$ denote a random sample from a normal distribution $N(\theta, 16)$. Find the sample size n and a uniformly most powerful test of $H_0 : \theta = 25$ against $H_1 : \theta < 25$ with power function $K(\theta)$ so that approximately $K(25) = 0.10$ and $K(23) = 0.90$.

9.19. Consider a distribution having a p.d.f. of the form $f(x; \theta) = \theta^x(1-\theta)^{1-x}$, $x = 0, 1$, zero elsewhere. Let $H_0 : \theta = \frac{1}{20}$ and $H_1 : \theta > \frac{1}{20}$. Use the central limit theorem to determine the sample size n of a random sample so that a uniformly most powerful test of H_0 against H_1 has a power function $K(\theta)$, with approximately $K(\frac{1}{20}) = 0.05$ and $K(\frac{1}{10}) = 0.90$.

9.20. Illustrative Example 1 of this section dealt with a random sample of size $n = 2$ from a gamma distribution with $\alpha = 1$, $\beta = \theta$. Thus the m.g.f. of the distribution is $(1 - \theta t)^{-1}$, $t < 1/\theta$, $\theta \ge 2$. Let $Z = X_1 + X_2$. Show that Z has a gamma distribution with $\alpha = 2$, $\beta = \theta$. Express the power function $K(\theta)$ of Example 1 in terms of a single integral. Generalize this for a random sample of size n.

9.21. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with p.d.f. $f(x; \theta) = \theta x^{\theta-1}$, $0 < x < \infty$, zero elsewhere, where $\theta > 0$. Find a sufficient statistic for θ and show that a uniformly most powerful test of $H_0 : \theta = 6$ against $H_1 : \theta < 6$ is based on this statistic.

9.22. Let X have the p.d.f. $f(x; \theta) = \theta^x(1-\theta)^{1-x}$, $x = 0, 1$, zero elsewhere. We test $H_0 : \theta = \frac{1}{2}$ against $H_1 : \theta < \frac{1}{2}$ by taking a random sample $X_1, X_2, \ldots, X_5$ of size $n = 5$ and rejecting H_0 if $Y = \sum_1^5 X_i$ is observed to be less than or equal to a constant c.

(a) Show that this is a uniformly most powerful test.
(b) Find the significance level when $c = 1$.
(c) Find the significance level when $c = 0$.
(d) By using a *randomized test*, modify the tests given in parts (b) and (c) to find a test with significance level $\alpha = \frac{2}{32}$.

9.3 Likelihood Ratio Tests

The notion of using the magnitude of the ratio of two probability density functions as the basis of a best test or of a uniformly most powerful test can be modified, and made intuitively appealing, to provide a method of constructing a test of a composite hypothesis against an alternative composite hypothesis or of constructing a test of a simple hypothesis against an alternative composite hypothesis when a uniformly most powerful test does not exist. This method leads to tests called *likelihood ratio tests*. A likelihood ratio test, as just remarked, is not necessarily a uniformly most powerful test, but it has been proved in the literature that such a test often has desirable properties.

A certain terminology and notation will be introduced by means of an example.

Example 1. Let the random variable X be $N(\theta_1, \theta_2)$ and let the parameter space be $\Omega = \{(\theta_1, \theta_2) : -\infty < \theta_1 < \infty, 0 < \theta_2 < \infty\}$. Let the composite hypothesis be $H_0 : \theta_1 = 0, \theta_2 > 0$, and let the alternative composite hypothesis be $H_1 : \theta_1 \neq 0, \theta_2 > 0$. The set $\omega = \{(\theta_1, \theta_2) : \theta_1 = 0, 0 < \theta_2 < \infty\}$ is a subset of Ω and will be called the *subspace* specified by the hypothesis H_0. Then, for instance, the hypothesis H_0 may be described as $H_0 : (\theta_1, \theta_2) \in \omega$. It is proposed that we test H_0 against all alternatives in H_1.

Let $X_1, X_2, \ldots, X_n$ denote a random sample of size $n > 1$ from the distribution of this example. The joint p.d.f. of $X_1, X_2, \ldots, X_n$ is, at each point in Ω,

$$L(\theta_1, \theta_2; x_1, \ldots, x_n) = \left(\frac{1}{2\pi\theta_2}\right)^{n/2} \exp\left[-\frac{\sum_1^n (x_i - \theta_1)^2}{2\theta_2}\right] = L(\Omega).$$

At each point $(\theta_1, \theta_2) \in \omega$, the joint p.d.f. of $X_1, X_2, \ldots, X_n$ is

$$L(0, \theta_2; x_1, \ldots, x_n) = \left(\frac{1}{2\pi\theta_2}\right)^{n/2} \exp\left[-\frac{\sum_1^n x_i^2}{2\theta_2}\right] = L(\omega).$$

The joint p.d.f., now denoted by $L(\omega)$, is not completely specified, since θ_2 may be any positive number; nor is the joint p.d.f., now denoted by $L(\Omega)$, completely specified, since θ_1 may be any real number and θ_2 any positive number. Thus the ratio of $L(\omega)$ to $L(\Omega)$ could not provide a basis for a test of H_0 against H_1. Suppose, however, that we modify this ratio in the following manner. We shall find the maximum of $L(\omega)$ in ω, that is, the maximum of $L(\omega)$ with respect to θ_2. And we shall find the maximum of

$L(\Omega)$ in Ω, that is, the maximum of $L(\Omega)$ with respect to θ_1 and θ_2. The ratio of these maxima will be taken as the criterion for a test of H_0 against H_1. Let the maximum of $L(\omega)$ in ω be denoted by $L(\hat{\omega})$ and let the maximum of $L(\Omega)$ in Ω be denoted by $L(\hat{\Omega})$. Then the criterion for the test of H_0 against H_1 is the likelihood ratio

$$\lambda(x_1, x_2, \ldots, x_n) = \lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})}.$$

Since $L(\omega)$ and $L(\Omega)$ are probability density functions, $\lambda \geq 0$; and since ω is a subset of Ω, $\lambda \leq 1$.

In our example the maximum, $L(\hat{\omega})$, of $L(\omega)$ is obtained by first setting

$$\frac{d \ln L(\omega)}{d\theta_2} = -\frac{n}{2\theta_2} + \frac{\sum_1^n x_i^2}{2\theta_2^2}$$

equal to zero and solving for θ_2. The solution of θ_2 is $\sum_1^n x_i^2/n$, and this number maximizes $L(\omega)$. Thus the maximum is

$$L(\hat{\omega}) = \left(\frac{1}{2\pi \sum_1^n x_i^2/n}\right)^{n/2} \exp\left[-\frac{\sum_1^n x_i^2}{2\sum_1^n x_i^2/n}\right]$$

$$= \left(\frac{ne^{-1}}{2\pi \sum_1^n x_i^2}\right)^{n/2}.$$

On the other hand, by using Example 4, Section 6.1, the maximum, $L(\hat{\Omega})$, of $L(\Omega)$ is obtained by replacing θ_1 and θ_2 by $\sum_1^n x_i/n = \bar{x}$ and $\sum_1^n (x_i - \bar{x})^2/n$, respectively. That is

$$L(\hat{\Omega}) = \left[\frac{1}{2\pi \sum_1^n (x_i - \bar{x})^2/n}\right]^{n/2} \exp\left[-\frac{\sum_1^n (x_i - \bar{x})^2}{2\sum_1^n (x_i - \bar{x})^2/n}\right]$$

$$= \left[\frac{ne^{-1}}{2\pi \sum_1^n (x_i - \bar{x})^2}\right]^{n/2}.$$

Thus here

$$\lambda = \left[\frac{\sum_1^n (x_i - \bar{x})^2}{\sum_1^n x_i^2}\right]^{n/2}.$$

Because $\sum_1^n x_i^2 = \sum_1^n (x_i - \bar{x})^2 + n\bar{x}^2$, λ may be written

$$\lambda = \frac{1}{\left\{1 + \left[n\bar{x}^2 \Big/ \sum_1^n (x_i - \bar{x})^2\right]\right\}^{n/2}}.$$

Now the hypothesis H_0 is $\theta_1 = 0$, $\theta_2 > 0$. If the observed number $\bar{x}$ were zero, the experiment tends to confirm H_0. But if $\bar{x} = 0$ and $\sum_1^n x_i^2 > 0$, then $\lambda = 1$. On the other hand, if $\bar{x}$ and $n\bar{x}^2 \Big/ \sum_1^n (x_i - \bar{x})^2$ deviate considerably from zero, the experiment tends to negate H_0. Now the greater the deviation of $n\bar{x}^2 \Big/ \sum_1^n (x_i - \bar{x})^2$ from zero, the smaller λ becomes. That is, if λ is used as a test criterion, then an intuitively appealing critical region for testing H_0 is a set defined by $0 \le \lambda \le \lambda_0$, where λ_0 is a positive proper fraction. Thus we reject H_0 if $\lambda \le \lambda_0$. A test that has the critical region $\lambda \le \lambda_0$ is a *likelihood ratio test*. In this example $\lambda \le \lambda_0$ when and only when

$$\frac{\sqrt{n}|\bar{x}|}{\sqrt{\sum_1^n (x_i - \bar{x})^2/(n-1)}} \ge \sqrt{(n-1)(\lambda_0^{-2/n} - 1)} = c.$$

If $H_0 : \theta_1 = 0$ is true, the results in Section 4.8 show that the statistic

$$t(X_1, X_2, \ldots, X_n) = \frac{\sqrt{n}(\bar{X} - 0)}{\sqrt{\sum_1^n (X_i - \bar{X})^2/(n-1)}}$$

has a t-distribution with $n - 1$ degrees of freedom. Accordingly, in this example the likelihood ratio test of H_0 against H_1 may be based on a T-statistic. For a given positive integer n, Table IV in Appendix B may be used (with $n - 1$ degrees of freedom) to determine the number c such that $\alpha = \Pr[|t(X_1, X_2, \ldots, X_n)| \ge c; H_0]$ is the desired significance level of the test. If the experimental values of $X_1, X_2, \ldots, X_n$ are, respectively, $x_1, x_2, \ldots, x_n$, then we reject H_0 if and only if $|t(x_1, x_2, \ldots, x_n)| \ge c$. If, for instance, $n = 6$ and $\alpha = 0.05$, then from Table IV, $c = 2.571$.

The preceding example should make the following generalization easier to read: Let $X_1, X_2, \ldots, X_n$ denote n independent random variables having, respectively, the probability density functions $f_i(x_i; \theta_1, \theta_2, \ldots, \theta_m)$, $i = 1, 2, \ldots, n$. The set that consists of all parameter points $(\theta_1, \theta_2, \ldots, \theta_m)$ is denoted by Ω, which we have called the parameter space. Let ω be a subset of the parameter space Ω. We wish to test the (simple or composite) hypothesis

$H_0: (\theta_1, \theta_2, \ldots, \theta_m) \in \omega$ against all alternative hypotheses. Define the likelihood functions

$$L(\omega) = \prod_{i=1}^{n} f_i(x_i; \theta_1, \theta_2, \ldots, \theta_m), \qquad (\theta_1, \theta_2, \ldots, \theta_m) \in \omega,$$

and

$$L(\Omega) = \prod_{i=1}^{n} f_i(x_i; \theta_1, \theta_2, \ldots, \theta_m), \qquad (\theta_1, \theta_2, \ldots, \theta_m) \in \Omega.$$

Let $L(\hat{\omega})$ and $L(\hat{\Omega})$ be the maxima, which we assume to exist, of these two likelihood functions. The ratio of $L(\hat{\omega})$ to $L(\hat{\Omega})$ is called the *likelihood ratio* and is denoted by

$$\lambda(x_1, x_2, \ldots, x_n) = \lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})}.$$

Let λ_0 be a positive proper function. The *likelihood ratio test principle* states that the hypothesis $H_0: (\theta_1, \theta_2, \ldots, \theta_m) \in \omega$ is rejected if and only if

$$\lambda(x_1, x_2, \ldots, x_n) = \lambda \le \lambda_0.$$

The function λ defines a random variable $\lambda(X_1, X_2, \ldots, X_n)$, and the significance level of the test is given by

$$\alpha = \Pr\,[\lambda(X_1, X_2, \ldots, X_n) \le \lambda_0; H_0].$$

The likelihood ratio test principle is an intuitive one. However, the principle does lead to the same test, when testing a simple hypothesis H_0 against an alternative simple hypothesis H_1, as that given by the Neyman–Pearson theorem (Exercise 9.25). Thus it might be expected that a test based on this principle has some desirable properties.

An example of the preceding generalization will be given.

Example 2. Let the independent random variables X and Y have distributions that are $N(\theta_1, \theta_3)$ and $N(\theta_2, \theta_3)$, where the means θ_1 and θ_2 and common variance θ_3 are unknown. Then $\Omega = \{(\theta_1, \theta_2, \theta_3): -\infty < \theta_1 < \infty, -\infty < \theta_2 < \infty, 0 < \theta_3 < \infty\}$. Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ denote independent random samples from these distributions. The hypothesis $H_0: \theta_1 = \theta_2$, unspecified, and θ_3 unspecified, is to be tested against all alternatives. Then $\omega = \{(\theta_1, \theta_2, \theta_3): -\infty < \theta_1 = \theta_2 < \infty, 0 < \theta_3 < \infty\}$. Here

$X_1, X_2, \ldots, X_n, Y_1, Y_2, \ldots, Y_m$ are $n + m > 2$ mutually independent random variables having the likelihood functions

$$L(\omega) = \left(\frac{1}{2\pi\theta_3}\right)^{(n+m)/2} \exp\left[-\frac{\sum_1^n (x_i - \theta_1)^2 + \sum_1^m (y_i - \theta_1)^2}{2\theta_3}\right]$$

and

$$L(\Omega) = \left(\frac{1}{2\pi\theta_3}\right)^{(n+m)/2} \exp\left[-\frac{\sum_1^n (x_i - \theta_1)^2 + \sum_1^m (y_i - \theta_2)^2}{2\theta_3}\right].$$

If

$$\frac{\partial \ln L(\omega)}{\partial \theta_1} \quad \text{and} \quad \frac{\partial \ln L(\omega)}{\partial \theta_3}$$

are equated to zero, then (Exercise 9.26)

$$\sum_1^n (x_i - \theta_1) + \sum_1^m (y_i - \theta_1) = 0, \tag{1}$$

$$-(n+m) + \frac{1}{\theta_3}\left[\sum_1^n (x_i - \theta_1)^2 + \sum_1^m (y_i - \theta_1)^2\right] = 0.$$

The solutions for θ_1 and θ_3 are, respectively,

$$u = \frac{\sum_1^n x_i + \sum_1^m y_i}{n+m}$$

and

$$w = \frac{\sum_1^n (x_i - u)^2 + \sum_1^m (y_i - u)^2}{n+m},$$

and u and w maximize $L(\omega)$. The maximum is

$$L(\hat{\omega}) = \left(\frac{e^{-1}}{2\pi w}\right)^{(n+m)/2}.$$

In like manner, if

$$\frac{\partial \ln L(\Omega)}{\partial \theta_1}, \quad \frac{\partial \ln L(\Omega)}{\partial \theta_2}, \quad \frac{\partial \ln L(\Omega)}{\partial \theta_3}$$

are equated to zero, then (Exercise 9.27)

$$\sum_1^n (x_i - \theta_1) = 0,$$

$$\sum_1^m (y_i - \theta_2) = 0, \tag{2}$$

$$-(n+m) + \frac{1}{\theta_3}\left[\sum_1^n (x_i - \theta_1)^2 + \sum_1^m (y_i - \theta_2)^2\right] = 0.$$

The solutions for θ_1, θ_2, and θ_3 are, respectively,

$$u_1 = \frac{\sum_1^n x_i}{n},$$

$$u_2 = \frac{\sum_1^m y_i}{m},$$

$$w' = \frac{\sum_1^n (x_i - u_1)^2 + \sum_1^m (y_i - u_2)^2}{n+m},$$

and u_1, u_2, and w' maximize $L(\Omega)$. The maximum is

$$L(\hat{\Omega}) = \left(\frac{e^{-1}}{2\pi w'}\right)^{(n+m)/2},$$

so that

$$\lambda(x_1, \ldots, x_n, y_1, \ldots, y_m) = \lambda = \frac{L(\hat{\omega})}{L(\hat{\Omega})} = \left(\frac{w'}{w}\right)^{(n+m)/2}.$$

The random variable defined by $\lambda^{2/(n+m)}$ is

$$\frac{\sum_1^n (X_i - \bar{X})^2 + \sum_1^m (Y_i - \bar{Y})^2}{\sum_1^n \{X_i - [(n\bar{X} + m\bar{Y})/(n+m)]\}^2 + \sum_1^m \{Y_i - [(n\bar{X} + m\bar{Y})/(n+m)]\}^2}.$$

Now

$$\begin{aligned}\sum_1^n \left(X_i - \frac{n\bar{X} + m\bar{Y}}{n+m}\right)^2 &= \sum_1^n \left[(X_i - \bar{X}) + \left(\bar{X} - \frac{n\bar{X} + m\bar{Y}}{n+m}\right)\right]^2 \\ &= \sum_1^n (X_i - \bar{X})^2 + n\left(\bar{X} - \frac{n\bar{X} + m\bar{Y}}{n+m}\right)^2\end{aligned}$$

and

$$\sum_1^m \left(Y_i - \frac{n\bar{X}+m\bar{Y}}{n+m}\right)^2 = \sum_1^m \left[(Y_i - \bar{Y}) + \left(\bar{Y} - \frac{n\bar{X}+m\bar{Y}}{n+m}\right)\right]^2$$

$$= \sum_1^m (Y_i - \bar{Y})^2 + m\left(\bar{Y} - \frac{n\bar{X}+m\bar{Y}}{n+m}\right)^2.$$

But

$$n\left(\bar{X} - \frac{n\bar{X}+m\bar{Y}}{n+m}\right)^2 = \frac{m^2n}{(n+m)^2}(\bar{X}-\bar{Y})^2$$

and

$$m\left(\bar{Y} - \frac{n\bar{X}+m\bar{Y}}{n+m}\right)^2 = \frac{n^2m}{(n+m)^2}(\bar{X}-\bar{Y})^2.$$

Hence the random variable defined by $\lambda^{2/(n+m)}$ may be written

$$\frac{\sum_1^n (X_i - \bar{X})^2 + \sum_1^m (Y_i - \bar{Y})^2}{\sum_1^n (X_i - \bar{X})^2 + \sum_1^m (Y_i - \bar{Y})^2 + [nm/(n+m)](\bar{X}-\bar{Y})^2}$$

$$= \frac{1}{1 + \dfrac{[nm/(n+m)](\bar{X}-\bar{Y})^2}{\sum_1^n (X_i - \bar{X})^2 + \sum_1^m (Y_i - \bar{Y})^2}}.$$

If the hypothesis $H_0: \theta_1 = \theta_2$ is true, the random variable

$$T = \frac{\sqrt{\dfrac{nm}{n+m}}(\bar{X}-\bar{Y})}{\sqrt{\dfrac{\sum_1^n (X_i - \bar{X})^2 + \sum_1^m (Y_i - \bar{Y})^2}{n+m-2}}}$$

has, in accordance with Section 6.3, a t-distribution with $n + m - 2$ degrees of freedom. Thus the random variable defined by $\lambda^{2/(n+m)}$ is

$$\frac{n+m-2}{(n+m-2)+T^2}.$$

The test of H_0 against all alternatives may then be based on a t-distribution with $n + m - 2$ degrees of freedom.

The likelihood ratio principle calls for the rejection of H_0 if and only if $\lambda \le \lambda_0 < 1$. Thus the significance level of the test is

$$\alpha = \Pr[\lambda(X_1, \ldots, X_n, Y_1, \ldots, Y_m) \le \lambda_0; H_0].$$

However, $\lambda(X_1, \ldots, X_n, Y_1, \ldots, Y_m) \le \lambda_0$ is equivalent to $|T| \ge c$, and so

$$\alpha = \Pr(|T| \ge c; H_0).$$

For given values of n and m, the number c is determined from Table IV in Appendix B (with $n + m - 2$ degrees of freedom) in such a manner as to yield a desired α. Then H_0 is rejected at a significance level α if and only if $|t| \ge c$, where t is the experimental value of T. If, for instance, $n = 10$, $m = 6$, and $\alpha = 0.05$, then $c = 2.145$.

In each of the two examples of this section it was found that the likelihood ratio test could be based on a statistic which, when the hypothesis H_0 is true, has a t-distribution. To help us compute the powers of these tests at parameter points other than those described by the hypothesis H_0, we turn to the following definition.

Definition 3. Let the random variable W be $N(\delta, 1)$; let the random variable V be $\chi^2(r)$, and W and V be independent. The quotient

$$T = \frac{W}{\sqrt{V/r}}$$

is said to have a *noncentral t-distribution* with r degrees of freedom and noncentrality parameter δ. If $\delta = 0$, we say that T has a central t-distribution.

In the light of this definition, let us reexamine the statistics of the examples of this section. In Example 1 we had

$$t(X_1, \ldots, X_n) = \frac{\sqrt{n}\,\bar{X}}{\sqrt{\sum_1^n (X_i - \bar{X})^2/(n-1)}}$$

$$= \frac{\sqrt{n}\,\bar{X}/\sigma}{\sqrt{\sum_1^n (X_i - \bar{X})^2/[\sigma^2(n-1)]}}.$$

Here $W_1 = \sqrt{n}\,\bar{X}/\sigma$ is $N(\sqrt{n}\,\theta_1/\sigma, 1)$, $V_1 = \sum_1^n (X_i - \bar{X})^2/\sigma^2$ is $\chi^2(n-1)$, and W_1 and V_1 are independent. Thus, if $\theta_1 \ne 0$, we see, in accordance with the definition, that $t(X_1, \ldots, X_n)$ has a noncentral t-distribution with $n-1$ degrees of freedom and noncentrality parameter $\delta_1 = \sqrt{n}\,\theta_1/\sigma$. In Example 2 we had

$$T = \frac{W_2}{\sqrt{V_2/(n+m-2)}},$$

where

$$W_2 = \sqrt{\frac{nm}{n+m}}(\bar{X} - \bar{Y}) \Big/ \sigma$$

and

$$V_2 = \frac{\sum_1^n (X_i - \bar{X})^2 + \sum_1^m (Y_i - \bar{Y})^2}{\sigma^2}.$$

Here W_2 is $N[\sqrt{nm/(n+m)}(\theta_1 - \theta_2)/\sigma, 1]$, V_2 is $\chi^2(n+m-2)$, and W_2 and V_2 are independent. Accordingly, if $\theta_1 \neq \theta_2$, T has a noncentral t-distribution with $n+m-2$ degrees of freedom and noncentrality parameter $\delta_2 = \sqrt{nm/(n+m)}(\theta_1 - \theta_2)/\sigma$. It is interesting to note that $\delta_1 = \sqrt{n}\,\theta_1/\sigma$ measures the deviation of θ_1 from $\theta_1 = 0$ in units of the standard deviation $\sigma/\sqrt{n}$ of $\bar{X}$. The noncentrality parameter $\delta_2 = \sqrt{nm/(n+m)}(\theta_1 - \theta_2)/\sigma$ is equal to the deviation of $\theta_1 - \theta_2$ from $\theta_1 - \theta_2 = 0$ in units of the standard deviation $\sigma\sqrt{(n+m)/nm}$ of $\bar{X} - \bar{Y}$.

There are various tables of the noncentral t-distribution, but they are much too cumbersome to be included in this book. However, with the aid of such tables, we can determine the power functions of these tests as functions of the noncentrality parameters.

In Example 2, in testing the equality of the means of two normal distributions, it was assumed that the unknown variances of the distributions were equal. Let us now consider the problem of testing the equality of these two unknown variances.

Example 3. We are given the independent random samples $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ from the distributions, which are $N(\theta_1, \theta_3)$ and $N(\theta_2, \theta_4)$, respectively. We have

$$\Omega = \{(\theta_1, \theta_2, \theta_3, \theta_4) : -\infty < \theta_1, \theta_2 < \infty, 0 < \theta_3, \theta_4 < \infty\}.$$

The hypothesis $H_0 : \theta_3 = \theta_4$, unspecified, with θ_1 and θ_2 also unspecified, is to be tested against all alternatives. Then

$$\omega = \{(\theta_1, \theta_2, \theta_3, \theta_4) : -\infty < \theta_1, \theta_2 < \infty, 0 < \theta_3 = \theta_4 < \infty\}.$$

It is easy to show (see Exercise 9.30) that the statistic defined by $\lambda = L(\hat{\omega})/L(\hat{\Omega})$ is a function of the statistic

$$F = \frac{\sum_1^n (X_i - \bar{X})^2/(n-1)}{\sum_1^m (Y_i - \bar{Y})^2/(m-1)}.$$

If $\theta_3 = \theta_4$, this statistic F has an F-distribution with $n - 1$ and $m - 1$ degrees of freedom. The hypothesis that $(\theta_1, \theta_2, \theta_3, \theta_4) \in \omega$ is rejected if the computed $F \le c_1$ or if the computed $F \ge c_2$. The constants c_1 and c_2 are usually selected so that, if $\theta_3 = \theta_4$,

$$\Pr(F \le c_1) = \Pr(F \ge c_2) = \frac{\alpha_1}{2},$$

where α_1 is the desired significance level of this test.

Often, under H_0, it is difficult to determine the distribution of $\lambda = \lambda(X_1, X_2, \ldots, X_n)$ or the distribution of an equivalent statistic upon which to base the likelihood ratio test. Hence it is impossible to find λ_0 such that $\Pr[\lambda \le \lambda_0; H_0]$ equals an appropriate value of α. The fact that the maximum likelihood estimators in a regular case have a joint normal distribution does, however, provide a solution. Using this fact, in a more advanced course, we can show that $-2 \ln \lambda$ has, given H_0 is true, an approximate chi-square distribution with r degrees of freedom, where r = the dimension of Ω − the dimension of ω. For illustration, in Example 1, the dimension of $\Omega = 2$ and the dimension of $\omega = 1$ and $r = 2 - 1 = 1$.

Also, in that example, note that

$$-2 \ln \lambda = n \ln \left\{1 + \frac{n\bar{x}^2}{\sum (x_i - \bar{x})^2}\right\} = n \ln \left\{1 + \frac{\bar{x}^2}{s^2}\right\}.$$

Hence, with n large so that $\bar{x}^2/s^2$ is close to zero under $H_0 : \theta_1 = 0$, let us approximate the right-hand member by two terms of a Taylor's series expanded about zero:

$$-2 \ln \lambda \approx 0 + \frac{n\bar{x}^2}{s^2}.$$

Since n is large, we can replace n by $n - 1$ to get the approximation

$$-2 \ln \lambda \approx \left(\frac{\bar{x}}{s/\sqrt{n-1}}\right)^2 = t^2.$$

But $T = \bar{X}/(S/\sqrt{n-1})$ under $H_0 : \theta_1 = 0$ has a t-distribution with $n - 1$ degrees of freedom. Moreover, with large $n - 1$, the distribution of T is approximately $N(0, 1)$ and the square of a standardized normal variable is $\chi^2(1)$, which is in agreement with the stated result. Exercise 9.31 provides another illustration of the fact that $-2 \ln \lambda$ has an approximate chi-square distribution.

EXERCISES

9.23. In Example 1 let $n = 10$, and let the experimental values of the random variables yield $\bar{x} = 0.6$ and $\sum_1^{10} (x_i - \bar{x})^2 = 3.6$. If the test derived in that example is used, do we accept or reject $H_0: \theta_1 = 0$ at the 5 percent significance level?

9.24. In Example 2 let $n = m = 8$, $\bar{x} = 75.2$, $\bar{y} = 78.6$, $\sum_1^8 (x_i - \bar{x})^2 = 71.2$, $\sum_1^8 (y_i - \bar{y})^2 = 54.8$. If we use the test derived in that example, do we accept or reject $H_0: \theta_1 = \theta_2$ at the 5 percent significance level?

9.25. Show that the likelihood ratio principle leads to the same test, when testing a simple hypothesis H_0 against an alternative simple hypothesis H_1, as that given by the Neyman–Pearson theorem. Note that there are only two points in Ω.

9.26. Verify Equations (1) of Example 2 of this section.

9.27. Verify Equations (2) of Example 2 of this section.

9.28. Let $X_1, X_2, \ldots, X_n$ be a random sample from the normal distribution $N(\theta, 1)$. Show that the likelihood ratio principle for testing $H_0: \theta = \theta'$, where θ' is specified, against $H_1: \theta \neq \theta'$ leads to the inequality $|\bar{x} - \theta'| \geq c$. Is this a uniformly most powerful test of H_0 against H_1?

9.29. Let $X_1, X_2, \ldots, X_n$ be a random sample from the normal distribution $N(\theta_1, \theta_2)$. Show that the likelihood ratio principle for testing $H_0: \theta_2 = \theta_2'$ specified, and θ_1 unspecified, against $H_1: \theta_2 \neq \theta_2'$, θ_1 unspecified, leads to a test that rejects when $\sum_1^n (x_i - \bar{x})^2 \leq c_1$ or $\sum_1^n (x_i - \bar{x})^2 \geq c_2$, where $c_1 < c_2$ are selected appropriately.

9.30. Let $X_1, \ldots, X_n$ and $Y_1, \ldots, Y_m$ be independent random samples from the distributions $N(\theta_1, \theta_3)$ and $N(\theta_2, \theta_4)$, respectively.

(a) Show that the likelihood ratio for testing $H_0: \theta_1 = \theta_2, \theta_3 = \theta_4$ against all alternatives is given by

$$\frac{\left[\sum_1^n (x_i - \bar{x})^2/n\right]^{n/2} \left[\sum_1^m (y_i - \bar{y})^2/m\right]^{m/2}}{\left\{\left[\sum_1^n (x_i - u)^2 + \sum_1^m (y_i - u)^2\right]\Big/(m + n)\right\}^{(n+m)/2}},$$

where $u = (n\bar{x} + m\bar{y})/(n + m)$.

(b) Show that the likelihood ratio test for testing $H_0 : \theta_3 = \theta_4$, θ_1 and θ_2 unspecified, against $H_1 : \theta_3 \neq \theta_4$, θ_1 and θ_2 unspecified, can be based on the random variable

$$F = \frac{\sum_1^n (X_i - \bar{X})^2/(n-1)}{\sum_1^m (Y_i - \bar{Y})^2/(m-1)}.$$

(c) If $\theta_3 = \theta_4$, argue that the F-statistic in part (b) is independent of the T-statistic of Example 2 of this section.

9.31. Let n independent trials of an experiment be such that $x_1, x_2, \ldots, x_k$ are the respective numbers of times that the experiment ends in the mutually exclusive and exhaustive events $A_1, A_2, \ldots, A_k$. If $p_i = P(A_i)$ is constant throughout the n trials, then the probability of that particular sequence of trials is $L = p_1^{x_1} p_2^{x_2} \cdots p_k^{x_k}$.

(a) Recalling that $p_1 + p_2 + \cdots + p_k = 1$, show that the likelihood ratio for testing $H_0 : p_i = p_{i0} > 0$, $i = 1, 2, \ldots, k$, against all alternatives is given by

$$\lambda = \prod_{i=1}^{k} \left(\frac{(p_{i0})^{x_i}}{(x_i/n)^{x_i}} \right).$$

(b) Show that

$$-2 \ln \lambda = \sum_{i=1}^{k} \frac{x_i (x_i - np_{0i})^2}{(np_i')^2},$$

where p_i' is between p_{0i} and x_i/n.

Hint: Expand $\ln p_{i0}$ in a Taylor's series with the remainder in the term involving $(p_{i0} - x_i/n)^2$.

(c) For large n, argue that $x_i/(np_i')^2$ is approximated by $1/(np_{i0})$ and hence

$$-2 \ln \lambda \approx \sum_{i=1}^{k} \frac{(x_i - np_{0i})^2}{np_{0i}}, \qquad \text{when } H_0 \text{ is true.}$$

In Section 6.6 we said the right-hand member of this last equation defines a statistic that has an approximate chi-square distribution with $k - 1$ degrees of freedom. Note that

$$\text{dimension of } \Omega - \text{dimension of } \omega = (k-1) - 0 = k - 1.$$

9.32. Let $Y_1 < Y_2 < \cdots < Y_5$ be the order statistics of a random sample of size $n = 5$ from a distribution with p.d.f. $f(x; \theta) = \frac{1}{2} e^{-|x - \theta|}$, $-\infty < x < \infty$, for all real θ. Find the likelihood ratio test λ for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta \neq \theta_0$.

9.33. Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_m$ be independent random samples from the two normal distributions $N(0, \theta_1)$ and $N(0, \theta_2)$.

(a) Find the likelihood ratio λ for testing the composite hypothesis $H_0 : \theta_1 = \theta_2$ against the composite alternative $H_1 : \theta_1 \neq \theta_2$.

(b) This λ is a function of what F-statistic that would actually be used in this test?

9.34. A random sample $X_1, X_2, \ldots, X_n$ arises from a distribution given by

$$H_0 : f(x; \theta) = \frac{1}{\theta}, \qquad 0 < x < \theta, \quad \text{zero elsewhere,}$$

or

$$H_1 : f(x; \theta) = \frac{1}{\theta} e^{-x/\theta}, \qquad 0 < x < \infty, \quad \text{zero elsewhere.}$$

Determine the likelihood ratio (λ) test associated with the test of H_0 against H_1.

9.35. Let X and Y be two independent random variables with respective probability density functions

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

zero elsewhere, $i = 1, 2$. To test $H_0 : \theta_1 = \theta_2$ against $H_1 : \theta_1 \neq \theta_2$, two independent random samples of sizes n_1 and n_2, respectively, were taken from these distributions. Find the likelihood ratio λ and show that λ can be written as a function of a statistic having an F-distribution, under H_0.

9.36. Consider the two uniform distributions with respective probability density functions

$$f(x; \theta_i) = \frac{1}{2\theta_i}, \qquad -\theta_i < x < \theta_i,$$

zero elsewhere, $i = 1, 2$. The null hypothesis is $H_0 : \theta_1 = \theta_2$ while the alternative is $H_1 : \theta_1 \neq \theta_2$. Let $X_1 < X_2 < \cdots < X_{n_1}$ and $Y_1 < Y_2 < \cdots < Y_{n_2}$ be the order statistics of two independent random samples from the two distributions, respectively. Using the likelihood ratio λ, find the statistic used to test H_0 against H_1. Find the distribution of $-2 \ln \lambda$ when H_0 is true. Note that in this nonregular case the number of degrees of freedom is two times the difference of the dimensions of Ω and ω.

9.4 The Sequential Probability Ratio Test

In Section 9.1 we proved a theorem that provided us with a method for determining a best critical region for testing a simple hypothesis against an alternative simple hypothesis. The theorem was as follows. Let $X_1, X_2, \ldots, X_n$ be a random sample with fixed sample size n from a distribution that has p.d.f. $f(x; \theta)$, where $\theta \in \{\theta : \theta = \theta', \theta''\}$ and θ' and θ'' are known numbers. Let the joint p.d.f. of $X_1, X_2, \ldots, X_n$ be denoted by

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

a notation that reveals both the parameter θ and the sample size n. If we reject $H_0 : \theta = \theta'$ and accept $H_1 : \theta = \theta''$ when and only when

$$\frac{L(\theta', n)}{L(\theta'', n)} \le k,$$

where $k > 0$, then this is a best test of H_0 against H_1.

Let us now suppose that the sample size n is *not* fixed in advance. In fact, let the sample size be a random variable N with sample space $\{n : n = 1, 2, 3, \ldots\}$. An interesting procedure for testing the simple hypothesis $H_0 : \theta = \theta'$ against the simple hypothesis $H_1 : \theta = \theta''$ is the following. Let k_0 and k_1 be two positive constants with $k_0 < k_1$. Observe the independent outcomes $X_1, X_2, X_3, \ldots$ in sequence, say $x_1, x_2, x_3, \ldots$, and compute

$$\frac{L(\theta', 1)}{L(\theta'', 1)}, \frac{L(\theta', 2)}{L(\theta'', 2)}, \frac{L(\theta', 3)}{L(\theta'', 3)}, \ldots.$$

The hypothesis $H_0 : \theta = \theta'$ is rejected (and $H_1 : \theta = \theta''$ is accepted) if and only if there exists a positive integer n so that $(x_1, x_2, \ldots, x_n)$ belongs to the set

$$C_n = \left\{(x_1, \ldots, x_n) : k_0 < \frac{L(\theta', j)}{L(\theta'', j)} < k_1, j = 1, \ldots, n-1, \right.$$
$$\left. \text{and} \quad \frac{L(\theta', n)}{L(\theta'', n)} \le k_0 \right\}.$$

On the other hand, the hypothesis $H_0 : \theta = \theta'$ is accepted (and $H_1 : \theta = \theta''$ is rejected) if and only if there exists a positive integer n so that $(x_1, x_2, \ldots, x_n)$ belongs to the set

$$B_n = \left\{(x_1, \ldots, x_n) : k_0 < \frac{L(\theta', j)}{L(\theta'', j)} < k_1, j = 1, 2, \ldots, n-1, \right.$$
$$\left. \text{and} \quad \frac{L(\theta', n)}{L(\theta'', n)} \ge k_1 \right\}.$$

That is, we continue to observe sample observations as long as

$$k_0 < \frac{L(\theta', n)}{L(\theta'', n)} < k_1. \tag{1}$$

We stop these observations in one of two ways:

1. With rejection of $H_0 : \theta = \theta'$ as soon as

$$\frac{L(\theta', n)}{L(\theta'', n)} \le k_0,$$

or

2. with acceptance of $H_0 : \theta = \theta'$ as soon as

$$\frac{L(\theta', n)}{L(\theta'', n)} \ge k_1.$$

A test of this kind is called Wald's *sequential probability ratio test*. Now, frequently inequality (1) can be conveniently expressed in an equivalent form

$$c_0(n) < u(x_1, x_2, \ldots, x_n) < c_1(n),$$

where $u(X_1, X_2, \ldots, X_n)$ is a statistic and $c_0(n)$ and $c_1(n)$ depend on the constants $k_0, k_1, \theta', \theta''$, and on n. Then the observations are stopped and a decision is reached as soon as

$$u(x_1, x_2, \ldots, x_n) \le c_0(n) \qquad \text{or} \qquad u(x_1, x_2, \ldots, x_n) \ge c_1(n).$$

We now give an illustrative example.

Example 1. Let X have a p.d.f.

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

In the preceding discussion of a sequential probability ratio test, let $H_0 : \theta = \frac{1}{3}$ and $H_1 : \theta = \frac{2}{3}$; then, with $\sum x_i = \sum_1^n x_i$,

$$\frac{L(\frac{1}{3}, n)}{L(\frac{2}{3}, n)} = \frac{(\frac{1}{3})^{\sum x_i}(\frac{2}{3})^{n-\sum x_i}}{(\frac{2}{3})^{\sum x_i}(\frac{1}{3})^{n-\sum x_i}} = 2^{n-2\sum x_i}.$$

If we take logarithms to the base 2, the inequality

$$k_0 < \frac{L(\frac{1}{3}, n)}{L(\frac{2}{3}, n)} < k_1,$$

with $0 < k_0 < k_1$, becomes

$$\log_2 k_0 < n - 2\sum_1^n x_i < \log_2 k_1,$$

or, equivalently,

$$c_0(n) = \frac{n}{2} - \frac{1}{2}\log_2 k_1 < \sum_1^n x_i < \frac{n}{2} - \frac{1}{2}\log_2 k_0 = c_1(n).$$

Note that $L(\frac{1}{3}, n)/L(\frac{2}{3}, n) \le k_0$ if and only if $c_1(n) \le \sum_1^n x_i$; and $L(\frac{1}{3}, n)/L(\frac{2}{3}, n) \ge k_1$ if and only if $c_0(n) \ge \sum_1^n x_i$. Thus we continue to observe outcomes as long as $c_0(n) < \sum_1^n x_i < c_1(n)$. The observation of outcomes is discontinued with the first value n of N for which either $c_1(n) \le \sum_1^n x_i$ or $c_0(n) \ge \sum_1^n x_i$. The inequality $c_1(n) \le \sum_1^n x_i$ leads to the rejection of $H_0 : \theta = \frac{1}{3}$ (the acceptance of H_1), and the inequality $c_0(n) \ge \sum_1^n x_i$ leads to the acceptance of $H_0 : \theta = \frac{1}{3}$ (the rejection of H_1).

Remarks. At this point, the reader undoubtedly sees that there are many questions that should be raised in connection with the sequential probability ratio test. Some of these questions are possibly among the following:

1. What is the probability of the procedure continuing indefinitely?
2. What is the value of the power function of this test at each of the points $\theta = \theta'$ and $\theta = \theta''$?
3. If θ'' is one of several values of θ specified by an alternative composite hypothesis, say $H_1 : \theta > \theta'$, what is the power function at each point $\theta \ge \theta'$?
4. Since the sample size N is a random variable, what are some of the properties of the distribution of N? In particular, what is the expected value $E(N)$ of N?
5. How does this test compare with tests that have a fixed sample size n?

A course in sequential analysis would investigate these and many other problems. However, in this book our objective is largely that of acquainting the reader with this kind of test procedure. Accordingly, we assert that the answer to question 1 is zero. Moreover, it can be proved that if $\theta = \theta'$ or if $\theta = \theta''$, $E(N)$ is smaller, for this sequential procedure, than the sample size of a fixed-sample-size test which has the same values of the power function at those points. We now consider question 2 in some detail.

In this section we shall denote the power of the test when H_0 is true by the symbol α and the power of the test when H_1 is true by the symbol $1 - \beta$. Thus α is the probability of committing a type I error (the rejection of H_0 when H_0 is true), and β is the probability of committing a type II error (the acceptance of H_0 when H_0 is false). With the sets C_n and B_n as previously defined, and with random variables of the continuous type, we then have

$$\alpha = \sum_{n=1}^{\infty} \int_{C_n} L(\theta', n), \qquad 1 - \beta = \sum_{n=1}^{\infty} \int_{C_n} L(\theta'', n).$$

Since the probability is 1 that the procedure will terminate, we also have

$$1 - \alpha = \sum_{n=1}^{\infty} \int_{B_n} L(\theta', n), \qquad \beta = \sum_{n=1}^{\infty} \int_{B_n} L(\theta'', n).$$

If $(x_1, x_2, \ldots, x_n) \in C_n$, we have $L(\theta', n) \le k_0 L(\theta'', n)$; hence it is clear that

$$\alpha = \sum_{n=1}^{\infty} \int_{C_n} L(\theta', n) \le \sum_{n=1}^{\infty} \int_{C_n} k_0 L(\theta'', n) = k_0(1 - \beta).$$

Because $L(\theta', n) \ge k_1 L(\theta'', n)$ at each point of the set B_n, we have

$$1 - \alpha = \sum_{n=1}^{\infty} \int_{B_n} L(\theta', n) \ge \sum_{n=1}^{\infty} \int_{B_n} k_1 L(\theta'', n) = k_1 \beta.$$

Accordingly, it follows that

$$\frac{\alpha}{1 - \beta} \le k_0, \qquad k_1 \le \frac{1 - \alpha}{\beta}, \tag{2}$$

provided that β is not equal to zero or 1.

Now let α_a and β_a be preassigned proper fractions; some typical values in the applications are 0.01, 0.05, and 0.10. If we take

$$k_0 = \frac{\alpha_a}{1 - \beta_a}, \qquad k_1 = \frac{1 - \alpha_a}{\beta_a},$$

then inequalities (2) become

$$\frac{\alpha}{1 - \beta} \le \frac{\alpha_a}{1 - \beta_a}, \qquad \frac{1 - \alpha_a}{\beta_a} \le \frac{1 - \alpha}{\beta}; \tag{3}$$

or, equivalently,

$$\alpha(1 - \beta_a) \le (1 - \beta)\alpha_a, \qquad \beta(1 - \alpha_a) \le (1 - \alpha)\beta_a.$$

If we add corresponding members of the immediately preceding inequalities, we find that

$$\alpha + \beta - \alpha\beta_a - \beta\alpha_a \le \alpha_a + \beta_a - \beta\alpha_a - \alpha\beta_a$$

and hence

$$\alpha + \beta \le \alpha_a + \beta_a.$$

That is, the sum $\alpha + \beta$ of the probabilities of the two kinds of errors is bounded above by the sum $\alpha_a + \beta_a$ of the preassigned numbers. Moreover, since α and β are positive proper fractions, inequalities (3) imply that

$$\alpha \le \frac{\alpha_a}{1-\beta_a}, \qquad \beta \le \frac{\beta_a}{1-\alpha_a};$$

consequently, we have an upper bound on each of α and β. Various investigations of the sequential probability ratio test seem to indicate that in most practical cases, the values of α and β are quite close to α_a and β_a. This prompts us to approximate the power function at the points $\theta = \theta'$ and $\theta = \theta''$ by α_a and $1 - \beta_a$, respectively.

Example 2. Let X be $N(\theta, 100)$. To find the sequential probability ratio test for testing $H_0: \theta = 75$ against $H_1: \theta = 78$ such that each of α and β is approximately equal to 0.10, take

$$k_0 = \frac{0.10}{1-0.10} = \frac{1}{9}, \qquad k_1 = \frac{1-0.10}{0.10} = 9.$$

Since

$$\frac{L(75, n)}{L(78, n)} = \frac{\exp\left[-\sum (x_i - 75)^2/2(100)\right]}{\exp\left[-\sum (x_i - 78)^2/2(100)\right]} = \exp\left(-\frac{6\sum x_i - 459n}{200}\right),$$

the inequality

$$k_0 = \frac{1}{9} < \frac{L(75, n)}{L(78, n)} < 9 = k_1$$

can be rewritten, by taking logarithms, as

$$-\ln 9 < \frac{6\sum x_i - 459\mathrm{n}}{200} < \ln 9.$$

This inequality is equivalent to the inequality

$$c_0(n) = \tfrac{153}{2}n - \tfrac{100}{3}\ln 9 < \sum_1^n x_i < \tfrac{153}{2}n + \tfrac{100}{3}\ln 9 = c_1(n).$$

Moreover, $L(75, n)/L(78, n) \le k_0$ and $L(75, n)/L(78, n) \ge k_1$ are equivalent to the inequalities $\sum_1^n x_i \ge c_1(n)$ and $\sum_1^n x_i \le c_0(n)$, respectively. Thus the observation of outcomes is discontinued with the first value n of N for which either $\sum_1^n x_i \ge c_1(n)$ or $\sum_1^n x_i \le c_0(n)$. The inequality $\sum_1^n x_i \ge c_1(n)$ leads to the rejection of $H_0: \theta = 75$, and the inequality $\sum_1^n x_i \le c_0(n)$ leads to the acceptance of $H_0: \theta = 75$. The power of the test is approximately 0.10 when H_0 is true, and approximately 0.90 when H_1 is true.

Remark. It is interesting to note that a sequential probability ratio test can be thought of as a *random-walk procedure*. For illustrations, the final inequalities of Examples 1 and 2 can be rewritten as

$$-\log_2 k_1 < \sum_1^n 2(x_i - 0.5) < -\log_2 k_0$$

and

$$-\frac{100}{3}\ln 9 < \sum_1^n (x_i - 76.5) < \frac{100}{3}\ln 9,$$

respectively. In each instance, we can think of starting at the point zero and taking random steps until one of the boundaries is reached. In the first situation the random steps are $2(X_1 - 0.5), 2(X_2 - 0.5), 2(X_3 - 0.5), \ldots$ and hence are of the same length, 1, but with random directions. In the second instance, both the length and the direction of the steps are random variables, $X_1 - 76.5, X_2 - 76.5, X_3 - 76.5, \ldots$.

In recent years, there has been much attention to improving quality of products using statistical methods. One such simple method was developed by Walter Shewhart in which a sample of size n of the items being produced is taken and they are measured, resulting in n values. The mean $\bar{x}$ of these n measurements has an approximate normal distribution with mean μ and variance σ^2/n. In practice, μ and σ^2 must be estimated, but in this discussion, we assume that they are known. From theory we know that the probability is 0.997 that $\bar{x}$ is between

$$\text{LCL} = \mu - \frac{3\sigma}{\sqrt{n}} \qquad \text{and} \qquad \text{UCL} = \mu + \frac{3\sigma}{\sqrt{n}}.$$

These two values are called the lower (LCL) and upper (UCL) control limits, respectively. Samples like this are taken periodically, resulting in a sequence of means, say $\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots$. These are usually plotted; and if they are between the LCL and UCL, we say that the process is *in control.* If one falls outside the limits, this would suggest that the mean μ has shifted, and the process would be investigated.

It was recognized by some that there could be a shift in the mean, say from μ to $\mu + (\sigma/\sqrt{n})$; and it would still be difficult to detect that shift with a single sample mean as now the probability of a single $\bar{x}$ exceeding UCL is only about 0.023. This means that we would need about $1/0.023 \approx 43$ samples, each of size n, on the average before detecting such a shift. This seems too long; so statisticians recognized that they should be cumulating experience as the sequence $\bar{x}_1, \bar{x}_2, \bar{x}_3, \ldots$ is observed in order to help them detect the shift sooner. It is the practice to compute the standardized variable $Z = (\bar{X} - \mu)/(\sigma/\sqrt{n})$; thus we state the problem in these terms and provide the solution given by a sequential probability ratio test.

Here Z is $N(\theta, 1)$, and we wish to test $H_0 : \theta = 0$ against $H_1 : \theta = 1$ using the sequence of i.i.d. random variables $Z_1, Z_2, \ldots, Z_m, \ldots$. We use m rather than n, as the latter is the size of the samples taken periodically. We have

$$\frac{L(0, m)}{L(1, m)} = \frac{\exp[-\sum z_i^2/2]}{\exp[-\sum (z_i - 1)^2/2]} = \exp\left[-\sum_{i=1}^{m} (z_i - 0.5)\right].$$

Thus

$$k_0 < \exp\left[-\sum_{i=1}^{m} (z_i - 0.5)\right] < k_1$$

can be rewritten as

$$h = -\ln k_0 > \sum_{i=1}^{m} (z_i - 0.5) > -\ln k_1 = -h.$$

It is true that $-\ln k_0 = \ln k_1$ when $\alpha_a = \beta_a$. Often, $h = -\ln k_0$ is taken to be about 4 or 5, suggesting that $\alpha_a = \beta_a$ is small, like 0.01. As $\Sigma (z_i - 0.5)$ is cumulating the sum of $z_i - 0.5$, $i = 1, 2, 3, \ldots$, these procedures are often called CUSUMS. If the CUSUM $= \Sigma (z_i - 0.5)$ exceeds h, we would investigate the process, as it seems that the mean has shifted upward. If this shift is to $\theta = 1$, the theory associated with these procedures shows that we need only 8 or 9 samples on the average, rather than 43, to detect this shift. For more information about these methods, the reader is referred to one of the many books on quality improvement through statistical methods. What we would like to emphasize here is that, through sequential methods (not only the sequential probability ratio test), we should take advantage of all past experience that we can gather in making inferences.

EXERCISES

9.37. Let X be $N(0, \theta)$ and, in the notation of this section, let $\theta' = 4$, $\theta'' = 9$, $\alpha_a = 0.05$, and $\beta_a = 0.10$. Show that the sequential probability ratio test can be based upon the statistic $\sum_1^n X_i^2$. Determine $c_0(n)$ and $c_1(n)$.

9.38. Let X have a Poisson distribution with mean θ. Find the sequential probability ratio test for testing $H_0: \theta = 0.02$ against $H_1: \theta = 0.07$. Show that this test can be based upon the statistic $\sum_1^n X_i$. If $\alpha_a = 0.20$ and $\beta_a = 0.10$, find $c_0(n)$ and $c_1(n)$.

9.39. Let the independent random variables Y and Z be $N(\mu_1, 1)$ and $N(\mu_2, 1)$, respectively. Let $\theta = \mu_1 - \mu_2$. Let us observe independent observations from each distribution, say $Y_1, Y_2, \ldots$ and $Z_1, Z_2, \ldots$. To test sequentially the hypothesis $H_0: \theta = 0$ against $H_1: \theta = \frac{1}{2}$, use the sequence $X_i = Y_i - Z_i$, $i = 1, 2, \ldots$. If $\alpha_a = \beta_a = 0.05$, show that the test can be based upon $\bar{X} = \bar{Y} - \bar{Z}$. Find $c_0(n)$ and $c_1(n)$.

9.40. Say that a manufacturing process makes about 3 percent defective items, which is considered satisfactory for this particular product. The managers would like to decrease this to about 1 percent and clearly want to guard against a substantial increase, say to 5 percent. To monitor the process, periodically $n = 100$ items are taken and the number X of defectives counted. Assume that X is $b(n = 100, p = \theta)$. Based on a sequence $X_1, X_2, \ldots, X_m, \ldots$, determine a sequential probability ratio test that tests $H_0: \theta = 0.01$ against $H_1: \theta = 0.05$. (Note that $\theta = 0.03$, the present level, is in between these two values.) Write this test in the form

$$h_0 > \sum_{i=1}^{m} (x_i - nd) > h_1$$

and determine d, h_0, and h_1 if $\alpha_a = \beta_a = 0.02$.

9.5 Minimax, Bayesian, and Classification Procedures

In Chapters 7 and 8 we considered several procedures which may be used in problems of point estimation. Among these were decision function procedures (in particular, minimax decisions) and Bayesian procedures. In this section, we apply these same principles to the problem of testing a simple hypothesis H_0 against an alternative simple hypothesis H_1. It is important to observe that each of these procedures yields, in accordance with the Neyman–Pearson theorem, a best test of H_0 against H_1.

We first investigate the decision function approach to the problem of testing a simple hypothesis against a simple alternative hypothesis. Let the joint p.d.f. of n random variables $X_1, X_2, \ldots, X_n$ depend upon the parameter θ. Here n is a fixed positive integer. This p.d.f. is denoted by $L(\theta; x_1, x_2, \ldots, x_n)$ or, for brevity, by $L(\theta)$. Let θ' and θ'' be distinct and fixed values of θ. We wish to test the simple hypothesis $H_0 : \theta = \theta'$ against the simple hypothesis $H_1 : \theta = \theta''$. Thus the parameter space is $\Omega = \{\theta : \theta = \theta', \theta''\}$. In accordance with the decision function procedure, we need a function δ of the observed values of $X_1, \ldots, X_n$ (or, of the observed value of a statistic Y) that decides which of the two values of θ, θ' or θ'', to accept. That is, the function δ selects either $H_0 : \theta = \theta'$ or $H_1 : \theta = \theta''$. We denote these decisions by $\delta = \theta'$ and $\delta = \theta''$, respectively. Let $\mathscr{L}(\theta, \delta)$ represent the loss function associated with this decision problem. Because the pairs $(\theta = \theta', \delta = \theta')$ and $(\theta = \theta'', \delta = \theta'')$ represent correct decisions, we shall always take $\mathscr{L}(\theta', \theta') = \mathscr{L}(\theta'', \theta'') = 0$. On the other hand, if either $\delta = \theta''$ when $\theta = \theta'$ or $\delta = \theta'$ when $\theta = \theta''$, then a positive value should be assigned to the loss function; that is, $\mathscr{L}(\theta', \theta'') > 0$ and $\mathscr{L}(\theta'', \theta') > 0$.

It has previously been emphasized that a test of $H_0 : \theta = \theta'$ against $H_1 : \theta = \theta''$ can be described in terms of a critical region in the sample space. We can do the same kind of thing with the decision function. That is, we can choose a subset C of the sample space and if $(x_1, x_2, \ldots, x_n) \in C$, we can make the decision $\delta = \theta''$; whereas, if $(x_1, x_2, \ldots, x_n) \in C^*$, the complement of C, we make the decision $\delta = \theta'$. Thus a given critical region C determines the decision function. In this sense, we may denote the risk function by $R(\theta, C)$ instead of $R(\theta, \delta)$. That is, in a notation used in Section 9.1,

$$R(\theta, C) = R(\theta, \delta) = \int_{C \cup C^*} \mathscr{L}(\theta, \delta) L(\theta).$$

Since $\delta = \theta''$ if $(x_1, \ldots, x_n) \in C$ and $\delta = \theta'$ if $(x_1, \ldots, x_n) \in C^*$, we have

$$R(\theta, C) = \int_C \mathscr{L}(\theta, \theta'') L(\theta) + \int_{C^*} \mathscr{L}(\theta, \theta') L(\theta). \tag{1}$$

If, in Equation (1), we take $\theta = \theta'$, then $\mathscr{L}(\theta', \theta') = 0$ and hence

$$R(\theta', C) = \int_C \mathscr{L}(\theta', \theta'') L(\theta') = \mathscr{L}(\theta', \theta'') \int_C L(\theta').$$

On the other hand, if in Equation (1) we let $\theta = \theta''$, then $\mathscr{L}(\theta'', \theta'') = 0$ and, accordingly,

$$R(\theta'', C) = \int_{C^*} \mathscr{L}(\theta'', \theta')L(\theta'') = \mathscr{L}(\theta'', \theta') \int_{C^*} L(\theta'').$$

It is enlightening to note that, if $K(\theta)$ is the power function of the test associated with the critical region C, then

$$R(\theta', C) = \mathscr{L}(\theta', \theta'')K(\theta') = \mathscr{L}(\theta', \theta'')\alpha,$$

where $\alpha = K(\theta')$ is the significance level; and

$$R(\theta'', C) = \mathscr{L}(\theta'', \theta')[1 - K(\theta'')] = \mathscr{L}(\theta'', \theta')\beta,$$

where $\beta = 1 - K(\theta'')$ is the probability of the type II error.

Let us now see if we can find a minimax solution to our problem. That is, we want to find a critical region C so that

$$\max [R(\theta', C), R(\theta'', C)]$$

is minimized. We shall show that the solution is the region

$$C = \left\{(x_1, \ldots, x_n) : \frac{L(\theta'; x_1, \ldots, x_n)}{L(\theta''; x_1, \ldots, x_n)} \le k\right\},$$

provided the positive constant k is selected so that $R(\theta', C) = R(\theta'', C)$. That is, if k is chosen so that

$$\mathscr{L}(\theta', \theta'') \int_C L(\theta') = \mathscr{L}(\theta'', \theta') \int_{C^*} L(\theta''),$$

then the critical region C provides a minimax solution. In the case of random variables of the continuous type, k can always be selected so that $R(\theta', C) = R(\theta'', C)$. However, with random variables of the discrete type, we may need to consider an auxiliary random experiment when $L(\theta')/L(\theta'') = k$ in order to achieve the exact equality $R(\theta', C) = R(\theta'', C)$.

To see that this region C is the minimax solution, consider every other region A for which $R(\theta', C) \ge R(\theta', A)$. Obviously, a region A for which $R(\theta', C) < R(\theta', A)$ is not a candidate for a minimax solution, for then $R(\theta', C) = R(\theta'', C) < \max [R(\theta', A), R(\theta'', A)]$. Since $R(\theta', C) \ge R(\theta', A)$ means that

$$\mathscr{L}(\theta', \theta'') \int_C L(\theta') \ge \mathscr{L}(\theta', \theta'') \int_A L(\theta'),$$

we have

$$\alpha = \int_C L(\theta') \geq \int_A L(\theta').$$

That is, the significance level of the test associated with the critical region A is less than or equal to α. But C, in accordance with the Neyman–Pearson theorem, is a best critical region of size α. Thus

$$\int_C L(\theta'') \geq \int_A L(\theta'')$$

and

$$\int_{C^*} L(\theta'') \leq \int_{A^*} L(\theta'').$$

Accordingly,

$$\mathscr{L}(\theta'', \theta') \int_{C^*} L(\theta'') \leq \mathscr{L}(\theta'', \theta') \int_{A^*} L(\theta''),$$

or, equivalently,

$$R(\theta'', C) \leq R(\theta'', A).$$

That is,

$$R(\theta', C) = R(\theta'', C) \leq R(\theta'', A).$$

This means that

$$\max [R(\theta', C), R(\theta'', C)] \leq R(\theta'', A).$$

Then certainly,

$$\max [R(\theta', C), R(\theta'', C)] \leq \max [R(\theta', A), R(\theta'', A)],$$

and the critical region C provides a minimax solution, as we wanted to show.

Example 1. Let $X_1, X_2, \ldots, X_{100}$ denote a random sample of size 100 from a distribution that is $N(\theta, 100)$. We again consider the problem of testing $H_0: \theta = 75$ against $H_1: \theta = 78$. We seek a minimax solution with $\mathscr{L}(75, 78) = 3$ and $\mathscr{L}(78, 75) = 1$. Since $L(75)/L(78) \leq k$ is equivalent to $\bar{x} \geq c$, we want to determine c, and thus k, so that

$$3 \Pr (\bar{X} \geq c; \theta = 75) = \Pr (\bar{X} < c; \theta = 78).$$

Because $\bar{X}$ is $N(\theta, 1)$, the preceding equation can be rewritten as

$$3[1 - \Phi(c - 75)] = \Phi(c - 78).$$

If we use Table III of the appendix, we see, by trial and error, that the solution is $c = 76.8$, approximately. The significance level of the test is $1 - \Phi(1.8) = 0.036$, approximately, and the power of the test when H_1 is true is $1 - \Phi(-1.2) = 0.885$, approximately.

Next, let us consider the Bayesian approach to the problem of testing the simple hypothesis $H_0 : \theta = \theta'$ against the simple hypothesis $H_1 : \theta = \theta''$. We continue to use the notation already presented in this section. In addition, we recall that we need the p.d.f. $h(\theta)$ of the random variable Θ. Since the parameter space consists of but two points, θ' and θ'', Θ is a random variable of the discrete type; and we have $h(\theta') + h(\theta'') = 1$. Since $L(\theta; x_1, x_2, \ldots, x_n) = L(\theta)$ is the conditional p.d.f. of $X_1, X_2, \ldots, X_n$, given $\Theta = \theta$, the joint p.d.f. of $X_1, X_2, \ldots, X_n$ and Θ is

$$h(\theta)L(\theta; x_1, x_2, \ldots, x_n) = h(\theta)L(\theta).$$

Because

$$\sum_{\Omega} h(\theta)L(\theta) = h(\theta')L(\theta') + h(\theta'')L(\theta'')$$

is the marginal p.d.f. of $X_1, X_2, \ldots, X_n$, the conditional p.d.f. of Θ, given $X_1 = x_1, \ldots, X_n = x_n$, is

$$k(\theta|x_1, \ldots, x_n) = \frac{h(\theta)L(\theta)}{h(\theta')L(\theta') + h(\theta'')L(\theta'')}.$$

Now a Bayes' solution to a decision problem is defined in Section 8.1 as a $\delta(y)$ such that $E\{\mathscr{L}[\theta, \delta(y)]|Y = y\}$ is a minimum. In this problem if $\delta = \theta'$, the conditional expectation of $\mathscr{L}(\theta, \delta)$, given $X_1 = x_1, \ldots, X_n = x_n$, is

$$\sum_{\Omega} \mathscr{L}(\theta, \theta')k(\theta|x_1, \ldots, x_n) = \frac{\mathscr{L}(\theta'', \theta')h(\theta'')L(\theta'')}{h(\theta')L(\theta') + h(\theta'')L(\theta'')},$$

because $\mathscr{L}(\theta', \theta') = 0$; and if $\delta = \theta''$, this expectation is

$$\sum_{\Omega} \mathscr{L}(\theta, \theta'')k(\theta|x_1, \ldots, x_n) = \frac{\mathscr{L}(\theta', \theta'')h(\theta')L(\theta')}{h(\theta')L(\theta') + h(\theta'')L(\theta'')},$$

because $\mathscr{L}(\theta'', \theta'') = 0$. Accordingly, the Bayes' solution requires that the decision $\delta = \theta''$ be made if

$$\frac{\mathscr{L}(\theta', \theta'')h(\theta')L(\theta')}{h(\theta')L(\theta') + h(\theta'')L(\theta'')} < \frac{\mathscr{L}(\theta'', \theta')h(\theta'')L(\theta'')}{h(\theta')L(\theta') + h(\theta'')L(\theta'')},$$

or, equivalently, if

$$\frac{L(\theta')}{L(\theta'')} < \frac{\mathscr{L}(\theta'', \theta')h(\theta'')}{\mathscr{L}(\theta', \theta'')h(\theta')}. \tag{2}$$

If the sign of inequality in expression (2) is reversed, we make the decision $\delta = \theta'$; and if the two members of expression (2) are equal, we can use some auxiliary random experiment to make the decision. It is important to note that expression (2) describes, in accordance with the Neyman–Pearson theorem, a best test.

Example 2. In addition to the information given in Example 1, suppose that we know the prior probabilities for $\theta = \theta' = 75$ and for $\theta = \theta'' = 78$ to be given, respectively, by $h(75) = \frac{1}{7}$ and $h(78) = \frac{6}{7}$. Then the Bayes' solution is, in this case,

$$\frac{L(75)}{L(78)} < \frac{(1)(\frac{6}{7})}{(3)(\frac{1}{7})} = 2,$$

which is equivalent to $\bar{x} > 76.3$, approximately. The power of the test when H_0 is true is $1 - \Phi(1.3) = 0.097$, approximately, and the power of the test when H_1 is true is $1 - \Phi(-1.7) = \Phi(1.7) = 0.955$, approximately.

In summary, we make the following comments. In testing the simple hypothesis $H_0 : \theta = \theta'$ against the simple hypothesis $H_1 : \theta = \theta''$, it is emphasized that each principle leads to critical regions of the form

$$\left\{(x_1, x_2, \ldots, x_n) : \frac{L(\theta'; x_1, \ldots, x_n)}{L(\theta''; x_1, \ldots, x_n)} \le k\right\},$$

where k is a positive constant. In the classical approach, we determine k by requiring that the power function of the test have a certain value at the point $\theta = \theta'$ or at the point $\theta = \theta''$ (usually, the value α at the point $\theta = \theta'$). The minimax decision requires k to be selected so that

$$\mathscr{L}(\theta', \theta'') \int_C L(\theta') = \mathscr{L}(\theta'', \theta') \int_{C^*} L(\theta'').$$

Finally, the Bayes' procedure requires that

$$k = \frac{\mathscr{L}(\theta'', \theta')h(\theta'')}{\mathscr{L}(\theta', \theta'')h(\theta')}.$$

Each of these tests is a best test for testing a simple hypothesis $H_0 : \theta = \theta'$ against a simple alternative hypothesis $H_1 : \theta = \theta''$.

The summary above has an interesting application to the problem of *classification*, which can be described as follows. An investigator makes a number of measurements on an item and wants to place it into one of several categories (or classify it). For convenience in our discussion, we assume that only two measurements, say X and Y, are made on the item to be classified. Moreover, let X and Y have a joint p.d.f. $f(x, y; \theta)$, where the parameter θ represents one or more parameters. In our simplification, suppose that there are only two possible joint distributions (categories) for X and Y, which are indexed by the parameter values θ' and θ'', respectively. In this case, the problem then reduces to one of observing $X = x$ and $Y = y$ and then testing the hypothesis $\theta = \theta'$ against the hypothesis $\theta = \theta''$, with the classification of X and Y being in accord with which hypothesis is accepted. From the Neyman–Pearson theorem, we know that a best decision of this sort is of the form: If

$$\frac{f(x, y; \theta')}{f(x, y; \theta'')} \le k,$$

choose the distribution indexed by θ''; that is, we classify (x, y) as coming from the distribution indexed by θ''. Otherwise, choose the distribution indexed by θ'; that is, we classify (x, y) as coming from the distribution indexed by θ'. Here k can be selected by considering the power function, a minimax decision, or a Bayes' procedure. We favor the latter if the losses and prior probabilities are known.

Example 3. Let (x, y) be an observation of the random pair (X, Y), which has a bivariate normal distribution with parameters μ_1, μ_2, σ_1^2, σ_2^2, and ρ. In Section 3.5 that joint p.d.f. is given by

$$f(x, y; \mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-q(x,y;\mu_1,\mu_2)/2},$$

$$-\infty < x < \infty, \quad -\infty < y < \infty,$$

where $\sigma_1 > 0$, $\sigma_2 > 0$, $-1 < \rho < 1$, and

$$q(x, y; \mu_1, \mu_2) = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right].$$

Assume that σ_1^2, σ_2^2, and ρ are known but that we do not know whether the respective means of (X, Y) are (μ_1', μ_2') or (μ_1'', μ_2''). The inequality

$$\frac{f(x, y; \mu_1', \mu_2', \sigma_1^2, \sigma_2^2, \rho)}{f(x, y; \mu_1'', \mu_2'', \sigma_1^2, \sigma_2^2, \rho)} \le k$$

is equivalent to

$$\tfrac{1}{2}[q(x, y; \mu_1'', \mu_2'') - q(x, y; \mu_1', \mu_2')] \le \ln k.$$

Moreover, it is clear that the difference in the left-hand member of this inequality does not contain terms involving x^2, xy, and y^2. In particular, this inequality is the same as

$$\frac{1}{1-\rho^2}\left\{\left[\frac{\mu_1' - \mu_1''}{\sigma_1^2} - \frac{\rho(\mu_2' - \mu_2'')}{\sigma_1\sigma_2}\right]x + \left[\frac{\mu_2' - \mu_2''}{\sigma_2^2} - \frac{\rho(\mu_1' - \mu_1'')}{\sigma_1\sigma_2}\right]y\right\}$$

$$\le \ln k + \tfrac{1}{2}[q(0, 0; \mu_1', \mu_2') - q(0, 0; \mu_1'', \mu_2'')], \quad (3)$$

or, for brevity,

$$ax + by \le c.$$

That is, if this linear function of x and y in the left-hand member of inequality (3) is less than or equal to a certain constant, we would classify that (x, y) as coming from the bivariate normal distribution with means μ_1'' and μ_2''. Otherwise, we would classify (x, y) as arising from the bivariate normal distribution with means μ_1' and μ_2'. Of course, if the prior probabilities and losses are given, k and thus c can be found easily; this will be illustrated in Exercise 9.43.

Once the rule for classification is established, the statistician might be interested in the two probabilities of misclassifications using that rule. The first of these two is associated with the classification of (x, y) as arising from the distribution indexed by θ'' if, in fact, it comes from that index by θ'. The second misclassification is similar, but with the interchange of θ' and θ''. In the preceding example, the probabilities of these respective misclassifications are

$$\Pr(aX + bY \le c; \mu_1', \mu_2') \qquad \text{and} \qquad \Pr(aX + bY > c; \mu_1'', \mu_2'').$$

Fortunately, the distribution of $Z = aX + bY$ is easy to determine, so each of these probabilities is easy to calculate. The m.g.f. of Z is

$$E(e^{tZ}) = E[e^{t(aX+bY)}] = E(e^{atX+btY}).$$

Hence in the joint m.g.f. of X and Y found in Section 3.5, simply replace t_1 by at and t_2 by bt to obtain

$$E(e^{tZ}) = \exp\left[\mu_1 at + \mu_2 bt + \frac{\sigma_1^2(at)^2 + 2\rho\sigma_1\sigma_2(at)(bt) + \sigma_2^2(bt)^2}{2}\right]$$

$$= \exp\left[(a\mu_1 + b\mu_2)t + \frac{(a^2\sigma_1^2 + 2ab\rho\sigma_1\sigma_2 + b^2\sigma_2^2)t^2}{2}\right].$$

However, this is the m.g.f. of the normal distribution

$$N(a\mu_1 + b\mu_2, a^2\sigma_1^2 + 2ab\rho\sigma_1\sigma_2 + b^2\sigma_2^2).$$

With this information, it is easy to compute the probabilities of misclassifications, and this will also be demonstrated in Exercise 9.43.

One final remark must be made with respect to the use of the important classification rule established in Example 3. In most instances the parameter values μ_1', μ_2' and μ_1'', μ_2'' as well as σ_1^2, σ_2^2, and ρ are unknown. In such cases the statistician has usually observed a random sample (frequently called a *training sample*) from each of the two distributions. Let us say the samples have sizes n' and n'', respectively, with sample characteristics

$$\bar{x}', \bar{y}', (s_x')^2, (s_y')^2, r' \qquad \text{and} \qquad \bar{x}'', \bar{y}'', (s_x'')^2, (s_y'')^2, r''.$$

Accordingly, if in inequality (3) the parameters μ_1', μ_2', μ_1'', μ_2'', σ_1^2, σ_2^2, and $\rho\sigma_1\sigma_2$ are replaced by the unbiased estimates

$$\bar{x}', \bar{y}', \bar{x}'', \bar{y}'', \frac{n'(s_x')^2 + n''(s_x'')^2}{n' + n'' - 2}, \frac{n'(s_y')^2 + n''(s_y'')^2}{n' + n'' - 2},$$
$$\frac{n'r's_x's_y' + n''r''s_x''s_y''}{n' + n'' - 2},$$

the resulting expression in the left-hand member is frequently called Fisher's *linear discriminant function*. Since those parameters have been estimated, the distribution theory associated with $aX + bY$ is not appropriate for Fisher's function. However, if n' and n'' are large, the distribution of $aX + bY$ does provide an approximation.

Although we have considered only bivariate distributions in this section, the results can easily be extended to multivariate normal distributions after a study of Sections 4.10, 10.8, and 10.9.

EXERCISES

9.41. Let $X_1, X_2, \ldots, X_{20}$ be a random sample of size 20 from a distribution which is $N(\theta, 5)$. Let $L(\theta)$ represent the joint p.d.f. of $X_1, X_2, \ldots, X_{20}$. The problem is to test $H_0: \theta = 1$ against $H_1: \theta = 0$. Thus $\Omega = \{\theta : \theta = 0, 1\}$.

(a) Show that $L(1)/L(0) \le k$ is equivalent to $\bar{x} \le c$.

(b) Find c so that the significance level is $\alpha = 0.05$. Compute the power of this test if H_1 is true.

(c) If the loss function is such that $\mathscr{L}(1, 1) = \mathscr{L}(0, 0) = 0$ and $\mathscr{L}(1, 0) = \mathscr{L}(0, 1) > 0$, find the minimax test. Evaluate the power function of this test at the points $\theta = 1$ and $\theta = 0$.

(d) If, in addition, the prior probabilities of $\theta = 1$ and $\theta = 0$ are, respectively, $h(1) = \frac{3}{4}$ and $h(0) = \frac{1}{4}$, find the Bayes' test. Evaluate the power function of this test at the points $\theta = 1$ and $\theta = 0$.

9.42. Let $X_1, X_2, \ldots, X_{10}$ be a random sample of size 10 from a Poisson distribution with parameter θ. Let $L(\theta)$ be the joint p.d.f. of $X_1, X_2, \ldots, X_{10}$. The problem is to test $H_0: \theta = \frac{1}{2}$ against $H_1: \theta = 1$.

(a) Show that $L(\frac{1}{2})/L(1) \le k$ is equivalent to $y = \sum_1^{10} x_i \ge c$.

(b) In order to make $\alpha = 0.05$, show that H_0 is rejected if $y > 9$ and, if $y = 9$, reject H_0 with probability $\frac{1}{2}$ (using some auxiliary random experiment).

(c) If the loss function is such that $\mathscr{L}(\frac{1}{2}, \frac{1}{2}) = \mathscr{L}(1, 1) = 0$ and $\mathscr{L}(\frac{1}{2}, 1) = 1$ and $\mathscr{L}(1, \frac{1}{2}) = 2$ show that the minimax procedure is to reject H_0 if $y > 6$ and, if $y = 6$, reject H_0 with probability 0.08 (using some auxiliary random experiment).

(d) If, in addition, we are given that the prior probabilities of $\theta = \frac{1}{2}$ and $\theta = 1$ are $h(\frac{1}{2}) = \frac{1}{3}$ and $h(1) = \frac{2}{3}$, respectively, show that the Bayes' solution is to reject H_0 if $y > 5.2$, that is, reject H_0 if $y \ge 6$.

9.43. In Example 3 let $\mu_1' = \mu_2' = 0$, $\mu_1'' = \mu_2'' = 1$, $\sigma_1^2 = 1$, $\sigma_2^2 = 1$, and $\rho = \frac{1}{2}$.

(a) Evaluate inequality (3) when the prior probabilities are $h(\mu_1', \mu_2') = \frac{1}{3}$ and $h(\mu_1'', \mu_2'') = \frac{2}{3}$ and the losses are $\mathscr{L}[\theta = (\mu_1', \mu_2'), \delta = (\mu_1'', \mu_2'')] = 4$ and $\mathscr{L}[\theta = (\mu_1'', \mu_2''), \delta = (\mu_1', \mu_2')] = 1$.

(b) Find the distribution of the linear function $aX + bY$ that results from part (a).

(c) Compute $\Pr(aX + bY \le c; \mu_1' = \mu_2' = 0)$ and $\Pr(aX + bY > c; \mu_1'' = \mu_2'' = 1)$.

9.44. Let X and Y have the joint p.d.f.

$$f(x, y; \theta_1, \theta_2) = \frac{1}{\theta_1\theta_2} \exp\left(-\frac{x}{\theta_1} - \frac{y}{\theta_2}\right), \qquad 0 < x < \infty, \quad 0 < y < \infty,$$

zero elsewhere, where $0 < \theta_1$, $0 < \theta_2$. An observation (x, y) arises from the joint distribution with parameters equal to either $(\theta_1' = 1, \theta_2' = 5)$ or $(\theta_1'' = 3, \theta_2'' = 2)$. Determine the form of the classification rule.

9.45. Let X and Y have a joint bivariate normal distribution. An observation (x, y) arises from the joint distribution with parameters equal to either

$$\mu_1' = \mu_2' = 0, \quad (\sigma_1^2)' = (\sigma_2^2)' = 1, \quad \rho' = \tfrac{1}{2}$$

or

$$\mu_1'' = \mu_2'' = 1, \quad (\sigma_1^2)'' = 4, \quad (\sigma_2^2)'' = 9, \quad \rho'' = \tfrac{1}{2}.$$

Show that the classification rule involves a second degree polynomial in x and y.

9.46. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with one of the two probability density functions $(1/\rho)f_i[(x-\theta)/\rho]$, $-\infty < \theta < \infty$, $\rho > 0$, $i = 1, 2$. We wish to decide from which of these distributions the sample arose. We assign the respective prior probabilities p_1 and p_2 to f_1 and f_2, where $p_1 + p_2 = 1$. If the prior p.d.f. assigned to the nuisance parameters θ and ρ is $g(\theta, \rho)$, the posterior probability of f_i is proportional to $p_i I(f_i|x_1, \ldots, x_n)$, where

$$I(f_i|x_1, \ldots, x_n) = \int_0^\infty \int_{-\infty}^\infty \left(\frac{1}{\rho}\right)^n f_i\left(\frac{x_1-\theta}{\rho}\right)\cdots f_i\left(\frac{x_n-\theta}{\rho}\right) g(\theta, \rho)\, d\theta\, d\rho,$$
$$i = 1, 2.$$

If the losses associated with the two wrong decisions are equal, we would select the p.d.f. with the largest posterior probability.

(a) If $g(\theta; \rho)$ is a vague noninformative prior proportional to $1/\rho$, show that

$$I(f_i|x_1, \ldots, x_n) = \int_0^\infty \int_{-\infty}^\infty \left(\frac{1}{\rho}\right)^{n+1} f_i\left(\frac{x_1-\theta}{\rho}\right)\cdots f_i\left(\frac{x_n-\theta}{\rho}\right) d\theta\, d\rho$$
$$= \int_0^\infty \int_{-\infty}^\infty \lambda^{n-2} f_i(\lambda x_1 - u)\cdots f_i(\lambda x_n - u)\, du\, d\lambda$$

by changing variables through $\theta = \mu/\lambda$, $\rho = 1/\lambda$. Hájek and Šidák show that using this last expression, the Bayesian procedure of selecting f_2 over f_1 if

$$p_2 I(f_2|x_1, \ldots, x_n) \ge p_1 I(f_1|x_1, \ldots, x_n)$$

provides a most powerful location and scale invariant test of one model against another.

(b) Evaluate $I(f_i|x_1, \ldots, x_n)$, $i = 1, 2$, given in (a) for $f_1(x) = \frac{1}{2}$, $-1 < x < 1$, zero elsewhere, and $f_2(x)$ is the p.d.f. of $N(0, 1)$. Show that the most powerful location and scale invariant test for selecting the normal distribution over the uniform is of the form $(Y_n - Y_1)/S \le k$, where $Y_1 < Y_2 < \cdots < Y_n$ are the order statistics and S is the sample standard deviation.

ADDITIONAL EXERCISES

9.47. Consider a random sample $X_1, X_2, \ldots, X_n$ from a distribution with p.d.f. $f(x; \theta) = \theta(1-x)^{\theta-1}$, $0 < x < 1$, zero elsewhere, where $\theta > 0$.

(a) Find the form of the uniformly most powerful test of $H_0: \theta = 1$ against $H_1: \theta > 1$.

(b) What is the likelihood ratio λ for testing $H_0: \theta = 1$ against $H_1: \theta \neq 1$?

9.48. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with p.d.f.

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

(a) Find a complete sufficient statistic for θ.

(b) If $\alpha = \beta = \frac{1}{10}$, find the sequential probability ratio test of $H_0 : \theta = 2$ against $H_1 : \theta = 3$.

9.49. Let X have a Poisson p.d.f. with parameter θ. We shall use a random sample of size n to test $H_0 : \theta = 1$ against $H_1 : \theta \neq 1$.

(a) Find the likelihood ratio λ for making this test.

(b) Show that λ can be expressed in terms of $\bar{X}$, the mean of the sample, so that the test can be based upon $\bar{X}$.

9.50. Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ be independent random samples from two normal distributions $N(\mu_1, \sigma^2)$ and $N(\mu_2, \sigma^2)$, respectively, where σ^2 is the common but unknown variance.

(a) Find the likelihood ratio λ for testing $H_0 : \mu_1 = \mu_2 = 0$ against all alternatives.

(b) Rewrite λ so that it is a function of a statistic Z which has a well-known distribution.

(c) Give the distribution of Z under both null and alternative hypotheses.

9.51. Let $X_1, \ldots, X_n$ denote a random sample from a gamma-type distribution with alpha equal to 2 and beta equal to θ. Let $H_0 : \theta = 1$ and $H_1 : \theta > 1$.

(a) Show that there exists a uniformly most powerful test for H_0 against H_1, determine the statistic Y upon which the test may be based, and indicate the nature of the best critical region.

(b) Find the p.d.f. of the statistic Y in part (a). If we want a significance level of 0.05, write an equation which can be used to determine the critical region. Let $K(\theta)$, $\theta \geq 1$, be the power function of the test. Express the power function as an integral.

9.52. Let $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ be a random sample from a bivariate normal distribution with $\mu_1, \mu_2, \sigma_1^2 = \sigma_2^2 = \sigma^2, \rho = \frac{1}{2}$, where μ_1, μ_2, and $\sigma^2 > 0$ are unknown real numbers. Find the likelihood ratio λ for testing $H_0 : \mu_1 = \mu_2 = 0$, σ^2 unknown against all alternatives. The likelihood ratio λ is a function of what statistic that has a well-known distribution?

9.53. Let $W' = (W_1, W_2)$ be an observation from one of two bivariate normal distributions, I and II, each with $\mu_1 = \mu_2 = 0$ but with the respective variance–covariance matrices

$$\mathbf{V}_1 = \begin{pmatrix} 1 & 0 \\ 0 & 4 \end{pmatrix} \quad \text{and} \quad \mathbf{V}_2 = \begin{pmatrix} 3 & 0 \\ 0 & 12 \end{pmatrix}.$$

How would you classify W into I or II?

9.54. Let X be Poisson θ. Find the sequential probability ratio test for testing $H_0 : \theta = 0.05$ against $H_1 : \theta = 0.03$. Write this in the form

$c_0(n) < \sum_{i=1}^{n} X_i < c_1(n)$, determining $c_0(n)$ and $c_1(n)$ when $\alpha_a = 0.10$ and $\beta_a = 0.05$.

9.55. Let X and Y have the joint p.d.f.

$$f(x, y; \theta_1, \theta_2) = \frac{1}{\theta_1\theta_2} \exp\left(-\frac{x}{\theta_1} - \frac{y}{\theta_2}\right), \qquad 0 < x < \infty, \quad 0 < y < \infty,$$

zero elsewhere, where $0 < \theta_1$, $0 < \theta_2$. An observation (x, y) arises from the joint distribution with $\theta_1' = 10$, $\theta_2' = 5$ or $\theta_1'' = 3$, $\theta_2'' = 2$. Determine the form of the classification rule.

CHAPTER 10

Inferences About Normal Models

10.1 The Distributions of Certain Quadratic Forms

A homogeneous polynomial of degree 2 in n variables is called a *quadratic* form in those variables. If both the variables and the coefficients are real, the form is called a *real quadratic* form. Only real quadratic forms will be considered in this book. To illustrate, the form $X_1^2 + X_1X_2 + X_2^2$ is a quadratic form in the two variables X_1 and X_2; the form $X_1^2 + X_2^2 + X_3^2 - 2X_1X_2$ is a quadratic form in the three variables X_1, X_2, and X_3; but the form $(X_1 - 1)^2 + (X_2 - 2)^2 = X_1^2 + X_2^2 - 2X_1 - 4X_2 + 5$ is not a quadratic form in X_1 and X_2, although it is a quadratic form in the variables $X_1 - 1$ and $X_2 - 2$.

Let $\bar{X}$ and S^2 denote, respectively, the mean and the variance of a random sample $X_1, X_2, \ldots, X_n$ from an arbitrary distribution. Thus

$$nS^2 = \sum_1^n (X_i - \bar{X})^2 = \sum_1^n \left(X_i - \frac{X_1 + X_2 + \cdots + X_n}{n}\right)^2$$