Chapter 3
Uniformly Most Powerful Tests

3.1 Stating the Problem

We now begin the study of the statistical problem that forms the principal subject of this book, the problem of hypothesis testing. As the term suggests, one wishes to decide whether or not some hypothesis that has been formulated is correct. The choice here lies between only two decisions: accepting or rejecting the hypothesis. A decision procedure for such a problem is called a *test* of the hypothesis in question.

The decision is to be based on the value of a certain random variable X, the distribution P_θ of which is known to belong to a class $\mathcal{P} = \{P_\theta, \theta \in \Omega\}$. We shall assume that if θ were known, one would also know whether or not the hypothesis is true. The distributions of $\mathcal{P}$ can then be classified into those for which the hypothesis is true and those for which it is false. The resulting two mutually exclusive classes are denoted by H and K, and the corresponding subsets of Ω by Ω_H and Ω_K respectively, so that $H \cup K = \mathcal{P}$ and $\Omega_H \cup \Omega_K = \Omega$. Mathematically, the hypothesis is equivalent to the statement that P_θ is an element of H. It is therefore convenient to identify the hypothesis with this statement and to use the letter H also to denote the hypothesis. Analogously we call the distributions in K the alternatives to H, so that K is the *class of alternatives*.

Let the decisions of accepting or rejecting H be denoted by d_0 and d_1 respectively. A nonrandomized test procedure assigns to each possible value x of X one of these two decisions and thereby divides the sample space into two complementary regions A and R. If X falls into A, the hypothesis is accepted; otherwise it is rejected. The set A is called the region of acceptance, and the set R the region of rejection or *critical region*.

When performing a test one may arrive at the correct decision, or one may commit one of two errors: rejecting the hypothesis when it is true (error of the first kind or Type 1 error) or accepting it when it is false (error of the second kind or Type 2 error). The consequences of these are often quite different. For example, if one tests for the presence of some disease, incorrectly deciding on the necessity of treatment

may cause the patient discomfort and financial loss. On the other hand, failure to diagnose the presence of the ailment may lead to the patient's death.

It is desirable to carry out the test in a manner which keeps the probabilities of the two types of error to a minimum. Unfortunately, when the number of observations is given, both probabilities cannot be controlled simultaneously. It is customary therefore to assign a bound to the probability of incorrectly rejecting H when it is true and to attempt to minimize the other probability subject to this condition. Thus one selects a number α between 0 and 1, called the *level of significance*, and imposes the condition that

$$P_\theta\{\delta(X) = d_1\} = P_\theta\{X \in R\} \le \alpha \quad \text{for all} \quad \theta \in \Omega_H. \tag{3.1}$$

Subject to this condition, it is desired to minimize $P_\theta\{\delta(X) = d_0\}$ for θ in Ω_K or, equivalently, to maximize

$$P_\theta\{\delta(X) = d_1\} = P_\theta\{X \in R\} \quad \text{for all} \quad \theta \in \Omega_K. \tag{3.2}$$

Although (3.1) usually implies that

$$\sup_{\Omega_H} P_\theta\{X \in R\} = \alpha, \tag{3.3}$$

it is convenient to introduce a term for the left-hand side of (3.3): it is called the *size* of the test or critical region R. The condition (3.1) therefore restricts consideration to test whose size does not exceed the given level of significance. The probability of rejection (3.2) evaluated for a given θ in Ω_K is called the *power* of the test against the alternative θ. Considered as a function of θ for all $\theta \in \Omega$, the probability (3.2) is called the *power function* of the test and is denoted by $\beta(\theta)$.

Although we may formally decide between accepting H when $X \in A$ and rejecting H when $X \in R$, it must be emphasized that "accepting" H does not prove that H is true. Failure to reject H may result from insufficient data or poor power, so that accepting H should be interpreted as the data provide insufficient evidence against the null hypothesis.

The choice of a level of significance α is usually somewhat arbitrary, since in most situations there is no precise limit to the probability of an error of the first kind that can be tolerated.[1] Standard values, such as 0.01 or 0.05, were originally chosen to effect a reduction in the tables needed for carrying out various tests. By habit, and because of the convenience of standardization in providing a common frame of reference, these values gradually became entrenched as the conventional levels to use. This is unfortunate, since the choice of significance level should also take into consideration the power that the test will achieve against the alternatives of interest. There is little point in carrying out an experiment which has only a small chance

---

[1] The standard way to remove the arbitrary choice of α is to report the *p*-value of the test, defined as the smallest level of significance leading to rejection of the null hypothesis. This approach will be discussed toward the end of Section 3.3.

of detecting the effect being sought when it exists. Surveys by Cohen (1962) and Freiman et al. (1978) suggest that this is in fact the case for many studies. Ideally, the sample size should then be increased to permit adequate values for both significance level and power. If that is not feasible one may wish to use higher values of α than the customary ones. The opposite possibility, that one would like to decrease α, arises when the latter is so close to 1 that α can be lowered appreciably without a significant loss of power (cf. Problem 3.11). Rules for choosing α in relation to the attainable power are discussed by Lehmann (1958), Arrow (1960), and Sanathanan (1974), and from a Bayesian point of view by Savage (1962, pp. 64–66). See also Rosenthal and Rubin (1985).

Another consideration that may enter into the specification of a significance level is the attitude toward the hypothesis before the experiment is performed. If one firmly believes the hypothesis to be true, extremely convincing evidence will be required before one is willing to give up this belief, and the significance level will accordingly be set very low. (A low significance level results in the hypothesis being rejected only for a set of values of the observations whose total probability under hypothesis is small, so that such values would be most unlikely to occur if H were true.)

Let us next consider the structure of a randomized test. For any values x, such a test chooses between the two decisions, rejection or acceptance, with certain probabilities that depend on x and will be denoted by $\phi(x)$ and $1 - \phi(x)$ respectively. If the value of X is x, a random experiment is performed with two possible outcomes R and $\bar{R}$, the probabilities of which are $\phi(x)$ and $1 - \phi(x)$. If in this experiment R occurs, the hypothesis is rejected, otherwise it is accepted. A randomized test is therefore completely characterized by a function ϕ, the *critical function*, with $0 \le \phi(x) \le 1$ for all x. If ϕ takes on only the values 1 and 0, one is back in the case of a nonrandomized test. The set of points x for which $\phi(x) = 1$ is then just the region of rejection, so that in a nonrandomized test ϕ is simply the indicator function of the critical region.

If the distribution of X is P_θ, and the critical function ϕ is used, the probability of rejection is

$$E_\theta \phi(X) = \int \phi(x) \, dP_\theta(x),$$

the conditional probability $\phi(x)$ of rejection given x, integrated with respect to the probability distribution of X. The problem is to select ϕ so as to maximize the power

$$\beta_\phi(\theta) = E_\theta \phi(X) \qquad \text{for all} \quad \theta \in \Omega_K \tag{3.4}$$

subject to the condition

$$E_\theta \phi(X) \le \alpha \qquad \text{for all} \quad \theta \in \Omega_H. \tag{3.5}$$

A level α test that maximizes (3.4) is called a *most powerful* (MP) level α test. The same difficulty now arises that presented itself in the general discussion of Chapter 1. Typically, the test that maximizes the power against a particular alternative in K

depends on this alternative, so that some additional principal has to be introduced to define what is meant by an optimum test. There is one important exception: if K contains only one distribution, that is, if one is concerned with a single alternative, the problem is completely specified by (3.4) and (3.5). It then reduces to the mathematical problem of maximizing an integral subject to certain side conditions. The theory of this problem, and its statistical applications, constitutes the principle subject of the present chapter. In special cases it may of course turn out that the same test maximizes the power of all alternatives in K even when there is more than one. Examples of such *uniformly most powerful* (UMP) tests will be given in Sections 3.4 and 3.7.

In the above formulation the problem can be considered a special case of the general decision problem with two types of losses. Corresponding to the two kinds of error, one can introduce the two-component loss functions,

$$
\begin{aligned}
&L_1(\theta, d_1) = 1 \text{ or } 0 &&\text{as} \quad \theta \in \Omega_H \text{ or } \theta \in \Omega_K, \\
&L_1(\theta, d_0) = 0 &&\text{for all } \theta
\end{aligned}
$$

and

$$
\begin{aligned}
&L_2(\theta, d_0) = 0 \text{ or } 1 &&\text{as} \quad \theta \in \Omega_H \text{ or } \theta \in \Omega_K, \\
&L_2(\theta, d_1) = 0 &&\text{for all } \theta .
\end{aligned}
$$

With this definition the minimization of $EL_2(\theta, \delta(X))$ subject to the restriction $EL_1(\theta, \delta(X)) \leq \alpha$ is exactly equivalent to the problem of hypothesis testing as given above.

The formal loss functions L_1 and L_2 clearly do not represent in general the true losses. The loss resulting from an incorrect acceptance of the hypothesis, for example, will not be the same for all alternatives. The more the alternative differs from the hypothesis, the more serious are the consequences of such an error. As was discussed earlier, we have purposely foregone the more detailed approach implied by this criticism. Rather than working with a loss function which in practice one does not know, it seems preferable to base the theory on the simpler and intuitively appealing notion of error. It will be seen later that at least some of the results can be justified also in the more elaborate formulation.

3.2   The Neyman–Pearson Fundamental Lemma

A class of distributions is called *simple* if it contains a single distribution, and otherwise it is said to be *composite*. The problem of hypothesis testing is completely specified by (3.4) and (3.5) if K is simple. Its solution is easiest and can be given explicitly when the same is true of H. Let the distributions under a simple hypothesis H and alternative K be P_0 and P_1, and suppose for a moment that these distributions are discrete with $P_i\{X = x\} = P_i(x)$ for $i = 0, 1$. If at first one restricts attention to nonrandomized tests, the optimum test is defined as the critical region S satisfying

$$\sum_{x \in S} P_0(x) \le \alpha \qquad (3.6)$$

and

$$\sum_{x \in S} P_1(x) = \text{maximum} .$$

It is easy to see which points should be included in S. To each point are attached two values, its probability under P_0 and under P_1. The selected points are to have a total value not exceeding α on the one scale, and as large as possible on the other. This is a situation that occurs in many contexts. A buyer with a limited budget who wants to get "the most for his money" will rate the items according to their *value per dollar*. In order to travel a given distance in the shortest possible time, one must choose the quickest mode of transportation, that is, the one that yields the largest number of *miles per hour*. Analogously in the present problem the most valuable points x are those with the highest value of

$$r(x) = \frac{P_1(x)}{P_0(x)}.$$

The points are therefore rated according to the value of this ratio and selected for S in this order, as many as one can afford under restriction (3.6). Formally this means that S is the set of all points x for which $r(x) > c$, where c is determined by the condition

$$P_0\{X \in S\} = \sum_{x : r(x) > c} P_0(x) = \alpha .$$

Here a difficulty is seen to arise. It may happen that when a certain point is included, the value α has not yet been reached but that it would be exceeded if the point were also included. The exact value α can then either not be achieved at all, or it can be attained only by breaking the preference order established by $r(x)$. The resulting optimization problem has no explicit solution. (Algorithms for obtaining the maximizing set S are given by the theory of linear programming.) The difficulty can be avoided, however, by a modification which does not require violation of the r-order and which does lead to a simple explicit solution, namely by permitting randomization.[2] This makes it possible to split the next point, including only a portion of it, and thereby to obtain the exact value α without breaking the order of preference that has been established for inclusion of the various sample points. These considerations are formalized in the following theorem, the *fundamental lemma of Neyman and Pearson*.

---

[2] In practice, typically neither the breaking of the r-order nor randomization is considered acceptable. The common solution, instead, is to adopt a value of α that can be attained exactly and therefore does not present this problem.

Theorem 3.2.1 *Let P_0 and P_1 be probability distributions possessing densities p_0 and p_1 respectively with respect to a measure μ.*[3]

(i) Existence. *For testing $H : p_0$ against the alternative $K : p_1$ there exists a test ϕ and a constant k such that*

$$E_0\phi(X) = \alpha \tag{3.7}$$

and

$$\phi(x) = \begin{cases} 1 \ when & p_1(x) > kp_0(x), \\ 0 \ when & p_1(x) < kp_0(x). \end{cases} \tag{3.8}$$

(ii) Sufficient condition for a most powerful test. *If a test satisfies (3.7) and (3.8) for some k, then it is most powerful for testing p_0 against p_1 at level α.*

(iii) Necessary condition for a most powerful test. *If ϕ is most powerful at level α for testing p_0 against p_1, then for some k it satisfies (3.8) a.e. μ. It also satisfies (3.7) unless there exists a test of size $< \alpha$ and with power 1.*

PROOF. For $\alpha = 0$ and $\alpha = 1$ the theorem is easily seen to be true provided the value $k = +\infty$ is admitted in (3.8) and $0 \cdot \infty$ is interpreted as 0. Throughout the proof we shall therefore assume $0 < \alpha < 1$.

(i): Let $\alpha(c) = P_0\{p_1(X) > cp_0(X)\}$. Since the probability is computed under P_0, the inequality needs to be considered only for the set where $p_0(x) > 0$, so that $\alpha(c)$ is the probability that the random variable $p_1(X)/p_0(X)$ exceeds c. Thus $1 - \alpha(c)$ is a cumulative distribution function, and $\alpha(c)$ is nonincreasing and continuous on the right, $\alpha(c^-) - \alpha(c) = P_0\{p_1(X)/p_0(X) = c\}$, $\alpha(-\infty) = 1$, and $\alpha(\infty) = 0$. Given any $0 < \alpha < 1$, let c_0 be such that $\alpha(c_0) \le \alpha \le \alpha(c_0^-)$, and consider the test ϕ defined by

$$\phi(x) = \begin{cases} 1 & \text{when } p_1(x) > c_0 p_0(x), \\ \frac{\alpha - \alpha(c_0)}{\alpha(c_0^-) - \alpha(c_0)} & \text{when } p_1(x) = c_0 p_0(x), \\ 0 & \text{when } p_1(x) < c_0 p_0(x). \end{cases}$$

Here the middle expression is meaningful unless $\alpha(c_0) = \alpha(c_0^-)$; since then $P_0\{p_1(X) = c_0 p_0(X)\} = 0$, ϕ is defined a.e. The size of ϕ is

$$E_0\phi(X) = P_0\left\{\frac{p_1(X)}{p_0(X)} > c_0\right\} + \frac{\alpha - \alpha(c_0)}{\alpha(c_0^-) - \alpha(c_0)} P_0\left\{\frac{p_1(X)}{p_0(X)} = c_0\right\} = \alpha,$$

so that c_0 can be taken as the k of the theorem.

(ii): Suppose that ϕ is a test satisfying (3.7) and (3.8) and that ϕ^* is any other test with $E_0\phi^*(X) \le \alpha$. Denote by S^+ and S^- the sets in the sample space where $\phi(x) - \phi^*(x) > 0$ and < 0, respectively. If x is in S^+, $\phi(x)$ must be > 0 and $p_1(x) \ge kp_0(x)$. In the same way $p_1(x) \le kp_0(x)$ for all x in S^-, and hence

$$\int (\phi - \phi^*)(p_1 - kp_0)\, d\mu = \int_{S^+ \cup S^-} (\phi - \phi^*)(p_1 - kp_0)\, d\mu \ge 0.$$

---

[3] There is no loss of generality in this assumption, since one can take $\mu = P_0 + P_1$.

The difference in power between ϕ and $\phi*$ therefore satisfies

$$\int (\phi - \phi^*) p_1 \, d\mu \geq k \int (\phi - \phi^*) p_0 \, d\mu \geq 0,$$

as was to be proved.

(iii): Let ϕ^* be most powerful at level α for testing p_0 against p_1, and let ϕ satisfy (3.7) and (3.8). Let S be the intersection of the set $S^+ \cup S^-$, on which ϕ and ϕ^* differ, with the set $\{x : p_1(x) \neq kp_0(x)\}$, and suppose that $\mu(S) > 0$. Since $(\phi - \phi^*)(p_1 - kp_0)$ is positive on S, it follows from Problem 2.4 that

$$\int_{S^+ \cup S^-} (\phi - \phi^*)(p_1 - kp_0) \, d\mu = \int_S (\phi - \phi^*)(p_1 - kp_0) \, d\mu > 0$$

and hence ϕ is more powerful against p_1 than ϕ^*. This is a contradiction, and therefore $\mu(S) = 0$, as was to be proved.

If ϕ^* were of size $< \alpha$ and power < 1, it would be possible to include in the rejection region additional points or portions of points and thereby to increase the power until either the power is 1 or the size is α. Thus either $E_0\phi^*(X) = \alpha$ or $E_1\phi^*(X) = 1$. ∎

The proof of part (iii) shows that the most powerful test is uniquely determined by (3.7) and (3.8) except on the set on which $p_1(x) = kp_0(x)$. On this set, ϕ can be defined arbitrarily provided the resulting test has size α. Actually, we have shown that it is always possible to define ϕ to be constant over this boundary set. In the trivial case when there exists a test of power 1, the constant k of (3.8) is 0, and one will accept H for all points for which $p_1(x) = kp_0(x)$ even though the test may then have size $< \alpha$.

It follows from these remarks that the most powerful test is determined uniquely (up to sets of measure zero) by (3.7) and (3.8) whenever the set on which $p_1(x) = kp_0(x)$ has μ-measure zero. This unique test is then clearly nonrandomized. More generally, it is seen that randomization is not required except possibly on the boundary set, where it may be necessary to randomize in order to get the size equal to α. When there exists a test of power 1, (3.7) and (3.8) will determine a most powerful test, but it may not be unique in that there may exist a test also most powerful and satisfying (3.7) and (3.8) for some $\alpha' < \alpha$.

Corollary 3.2.1 *Let β denote the power of the most powerful level-α test $(0 < \alpha < 1)$ for testing P_0 against P_1. Then $\alpha < \beta$ unless $P_0 = P_1$.*

PROOF. Since the level-α test given by $\phi(x) \equiv \alpha$ has power α, it is seen that $\alpha \leq \beta$. If $\alpha = \beta < 1$, the test $\phi(x) \equiv \alpha$ is most powerful and by Theorem 3.2.1(iii) must satisfy (3.8). Then $p_0(x) = p_1(x)$ a.e. μ and hence $P_0 = P_1$. ∎

An alternative method for proving some of the results of this section is based on the following geometric representation of the problem of testing a simple hypothesis

against a simple alternative. Let N be the set of all points (α, β) for which there exists a test ϕ such that

$$\alpha = E_0\phi(X), \qquad \beta = E_1\phi(X).$$

This set is convex, contains the points (0,0) and (1,1), and is symmetric with respect to the point $(\frac{1}{2}, \frac{1}{2})$ in the sense that with any point (α, β) it also contains the point $(1 - \alpha, 1 - \beta)$. In addition, the set N is closed. [This follows from the weak compactness theorem for critical functions, Theorem A.5.1 of the Appendix; the argument is the same as that in the proof of Theorem 3.6.1(i).]

For each value $0 < \alpha_0 < 1$, the level-α_0 tests are represented by the points whose abscissa is $\leq \alpha_o$. The most powerful of these tests (whose existence follows from the fact that N is closed) corresponds to the point on the upper boundary of N with abscissa α_0. This is the only point corresponding to a most powerful level-α_0 test unless there exists a point $(\alpha, 1)$ in N with $\alpha < \alpha_0$ (Figure 3.1b).

As an example of this geometric approach, consider the following alternative proof of Corollary 3.2.1. Suppose that for some $0 < \alpha_0 < 1$ the power of the most powerful level-α_0 test is α_0. Then it follows from the convexity of N that $(\alpha, \beta) \in N$ implies $\beta \leq \alpha$, and hence from the symmetry of N that N consists exactly of the line segment connecting the points (0,0) and (1,1). This means that $\int \phi p_o \, d\mu = \int \phi p_1 \, d\mu$ for all ϕ and hence that $p_0 = p_1$ (a.e.μ), as was to be proved. A proof of Theorem 3.2.1 along these lines is given in a more general setting in the proof of Theorem 3.6.1.

Example 3.2.1 (Normal Location Model) Suppose X is an observation from $N(\xi, \sigma^2)$, with σ^2 known. The null hypothesis specifies $\xi = 0$ and the alternative specifies $\xi = \xi_1$ for some $\xi_1 > 0$. Then, the likelihood ratio is given by

$$\frac{p_1(x)}{p_0(x)} = \frac{\exp[-\frac{1}{2\sigma^2}(x - \xi_1)^2]}{\exp[-\frac{1}{2\sigma^2}x^2]} = \exp[\frac{\xi_1 x}{\sigma^2} - \frac{\xi_1^2}{2\sigma^2}]. \tag{3.9}$$
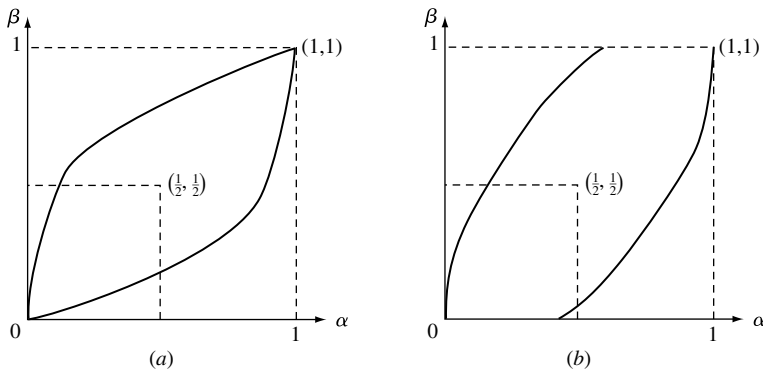
Figure 3.1 Possible Values of (α, β) for Varying ϕ

Since the exponential function is strictly increasing and $\xi_1 > 0$, the set of x where $p_1(x)/p_0(x) > k$ is equivalent to the set of x where $x > k'$. In order to determine k', the level constraint

$$P_0\{X > k'\} = \alpha$$

must be satisfied, and so $k' = \sigma z_{1-\alpha}$, where $z_{1-\alpha}$ is the $1 - \alpha$ quantile of the standard normal distribution. Therefore, the most powerful (MP) level α test rejects if $X > \sigma z_{1-\alpha}$. Several points are worth mentioning. First, the MP level α test is unique up to sets of Lebesgue measure 0, by the Neyman–Pearson Lemma, Theorem 3.2.1 (iii). Second, since this test is MP for any alternative $\xi_1 > 0$, the test is UMP level α against the composite alternatives $\xi_1 > 0$. Third, by a similar argument, the test that rejects if $X < \sigma z_\alpha$ is UMP level α against the composite alternatives $\xi_1 < 0$. Finally, it follows that no UMP level α test exists against all two-sided alternatives $\xi_1 \neq 0$. ∎

3.3   *p*-values

Testing at a fixed level α, as described in Sections 3.1 and 3.2, is one of two standard (non-Bayesian) approaches to the evaluation of hypotheses. To explain the other, suppose that, under P_0, the distribution of $p_1(X)/p_0(X)$ is continuous. Then, the most powerful level α test is nonrandomized and rejects if $p_1(X)/p_0(X) > k$, where $k = k(\alpha)$ is determined by (3.7). For varying α, the resulting tests provide an example of the typical situation in which the rejection regions R_α are nested in the sense that

$$R_\alpha \subseteq R_{\alpha'} \quad \text{if } \alpha < \alpha' . \tag{3.10}$$

When this is the case,[4] it is good practice to determine not only whether the hypothesis is accepted or rejected at the given significance level, but also to determine the smallest significance level, or more formally

$$\hat{p} = \hat{p}(X) = \inf\{\alpha : \ X \in R_\alpha\} , \tag{3.11}$$

at which the hypothesis would be rejected for the given observation. This number, the so-called *p-value* gives an idea of how strongly the data contradict the hypothesis.[5] It also enables others to reach a verdict based on the significance level of their choice.

---

[4] See Problems 3.17 and 3.63 for examples where optimal nonrandomized tests need not be nested.

[5] One could generalize the definition of *p*-value to include randomized level α tests ϕ_α assuming that they are nested in the sense that $\phi_\alpha(x) \leq \phi_{\alpha'}(x)$ for all x and $\alpha < \alpha'$. Simply define $\hat{p} = \inf\{\alpha : \ \phi_\alpha(X) = 1\}$; in words, $\hat{p}$ is the smallest level of significance where the hypothesis is rejected with probability one.

Example 3.3.1 (Continuation of Example) 3.2.1 Let Φ denote the standard normal c.d.f. Then, the rejection region can be written as

$$R_\alpha = \{X : \ X > \sigma z_{1-\alpha}\} = \{X : \ \Phi(\frac{X}{\sigma}) > 1 - \alpha\} = \{X : \ 1 - \Phi(\frac{X}{\sigma}) < \alpha\} \ .$$

For a given observed value of X, the inf over all α where the last inequality holds is

$$\hat{p} = 1 - \Phi(\frac{X}{\sigma}) \ .$$

Alternatively, the p-value is $P_0\{X \geq x\}$, where x is the observed value of X. Note that, under $\xi = 0$, the distribution of $\hat{p}$ is given by

$$P_0\{\hat{p} \leq u\} = P_0\{1 - \Phi(\frac{X}{\sigma}) \leq u\} = P_0\{\Phi(\frac{X}{\sigma}) \geq 1 - u\} = u \ ,$$

because $\Phi(X/\sigma)$ is uniformly distributed on (0,1) (see Problem 3.22); therefore, $\hat{p}$ is uniformly distributed on (0,1). ∎

A general property of p-values is given in the following lemma, which applies to both simple and composite null hypotheses.

Lemma 3.3.1 *Suppose X has distribution P_θ for some $\theta \in \Omega$, and the null hypothesis H specifies $\theta \in \Omega_H$. Assume the rejection regions satisfy (3.10).*
(i) If

$$\sup_{\theta \in \Omega_H} P_\theta\{X \in R_\alpha\} \leq \alpha \quad \text{for all } 0 < \alpha < 1, \tag{3.12}$$

then the distribution of $\hat{p}$ under $\theta \in \Omega_H$ satisfies

$$P_\theta\{\hat{p} \leq u\} \leq u \quad \text{for all } 0 \leq u \leq 1 \ . \tag{3.13}$$

(ii) If, for $\theta \in \Omega_H$,

$$P_\theta\{X \in R_\alpha\} = \alpha \quad \text{for all } 0 < \alpha < 1 \ , \tag{3.14}$$

then

$$P_\theta\{\hat{p} \leq u\} = u \quad \text{for all } 0 \leq u \leq 1 \ ;$$

i.e., $\hat{p}$ is uniformly distributed over (0, 1).

PROOF. (i) If $\theta \in \Omega_H$, then the event $\{\hat{p} \leq u\}$ implies $\{X \in R_v\}$ for all $u < v$. The result follows by letting $v \to u$.

(ii) Since the event $\{X \in R_u\}$ implies $\{\hat{p} \leq u\}$, it follows that

$$P_\theta\{\hat{p} \leq u\} \geq P_\theta\{X \in R_u\} \ .$$

Therefore, if (3.14) holds, then $P_\theta\{\hat{p} \leq u\} \geq u$, and the result follows from (i). ∎

Example 3.3.2 Suppose X takes values $1, 2, \ldots, 10$. Under H, the distribution is uniform, i.e., $p_0(j) = \frac{1}{10}$ for $j = 1, \ldots, 10$. Under K, suppose $p_1(j) = j/55$. The MP level $\alpha = i/10$ test rejects if $X \geq 11 - i$. However, unless α is a multiple of $1/10$, the MP level α test is randomized. If we want to restrict attention to nonrandomized procedures, consider the conservative approach by defining

$$R_\alpha = \{X \geq 11 - i\} \quad \text{if} \quad \frac{i}{10} \leq \alpha < \frac{i+1}{10} \; .$$

If the observed value of X is x, then the *p*-value is given by $(11 - x)/10$. Then, the distribution of $\hat{p}$ under H is given by

$$P\{\hat{p} \leq u\} = P\{\frac{11 - X}{10} \leq u\} = P\{X \geq 11 - 10u\} \leq u \; , \qquad (3.15)$$

and the last inequality is an equality if and only if u is of the form $i/10$ for some integer $i = 0, 1, \ldots, 10$, i.e., the levels for which the MP test is nonrandomized (Problem 3.21). ∎

In general, we say a *p*-value is valid if it satisfies (3.13) for all $\theta \in \Omega_H$, even if $\hat{p}$ is not specified through a family of rejection regions. If (3.13) holds, then the test that rejects H if $\hat{p} \leq \alpha$ is level α. A direct approach is given next.

Example 3.3.3 (Constructing P-values for a Simple H) Suppose for a given family indexed by $\theta \in \Omega$, the problem is to test a simple null hypothesis $\theta = \theta_0$. Here θ is quite general; it can be real-valued, vector-valued, or even be function-valued. (As an example, if θ corresponds to a c.d.f. on the real line, θ_0 could specify the uniform distribution.) Let $T = T(X)$ be any real-valued test statistic, and let its c.d.f. be denoted by $F_\theta(\cdot)$. Then, $\hat{p} = F_{\theta_0}(T)$ serves as a valid *p*-value for testing the null hypothesis $\theta = \theta_0$. To see why, note that if Y is any real-valued random variable with c.d.f. $G(\cdot)$, then (Problem 3.23)

$$P\{G(Y) \leq u\} \leq u \quad \text{for all } 0 \leq u \leq 1 \; .$$

Hence, for any $0 \leq u \leq 1$,

$$P_{\theta_0}\{\hat{p} \leq u\} = P_{\theta_0}\{F_{\theta_0}(T) \leq u\} \leq u \; ,$$

verifying (3.13). A test based on $F_{\theta_0}(T)$ is appropriate if small values of T indicate departures from H. Similarly, if large values of T indicate departures from H, then $1 - F_{\theta_0}^-(T)$ is a valid *p*-value, where $F_{\theta_0}^-(t) = P_{\theta_0}\{T < t\}$. ∎

P-values, with the additional information they provide, are typically more appropriate than fixed levels in scientific problems, whereas a fixed predetermined α is unavoidable when acceptance or rejection of H implies an imminent concrete decision. A review of some of the issues arising in this context, with references to the

literature, are given in Kruskal (1978), Wasserstein and Lazar (2016), and Kuffner and Walker (2019).

3.4 Distributions with Monotone Likelihood Ratio

The case that both the hypothesis and the class of alternatives are simple is mainly of theoretical interest, since problems arising in applications typically involve a parametric family of distributions depending on one or more parameters. In the simplest situation of this kind the distributions depend on a single real-valued parameter θ, and the hypothesis is one sided, say $H : \theta \leq \theta_0$. In general, the most powerful test of H against an alternative $\theta_1 > \theta_0$ depends on θ_1 and is then not UMP. However, a UMP test does exist if an additional assumption is satisfied. The real-valued parameter family of densities $p_\theta(x)$ is said to have *monotone likelihood ratio*[6] if there exists a real-valued function $T(x)$ such that for any $\theta < \theta'$ the distributions P_θ and $P_{\theta'}$ are distinct, and the ratio $p_{\theta'}(x)/p_\theta(x)$ is a nondecreasing function of $T(x)$.

Theorem 3.4.1 *Let θ be a real parameter, and let the random variable X have probability density $p_\theta(x)$ with monotone likelihood ratio in $T(x)$.*

  (i) *For testing $H : \theta \leq \theta_0$ against $K : \theta > \theta_0$, there exists a UMP test, which is given by*

$$\phi(x) = \begin{cases} 1 \ when \ T(x) > C, \\ \gamma \ when \ T(x) = C, \\ 0 \ when \ T(x) < C, \end{cases} \tag{3.16}$$

where C and γ are determined by

$$E_{\theta_0}\phi(X) = \alpha. \tag{3.17}$$

  (ii) *The power function*

$$\beta(\theta) = E_\theta\phi(X)$$

of this test is strictly increasing for all points θ for which $0 < \beta(\theta) < 1$.

  (iii) *For all θ', the test determined by (3.16) and (3.17) is UMP for testing $H' : \theta \leq \theta'$ against $K' : \theta > \theta'$ at level $\alpha' = \beta(\theta')$.*

  (iv) *For any $\theta < \theta_0$ the test minimizes $\beta(\theta)$ (the probability of an error of the first kind) among all tests satisfying (3.17).*

PROOF. (i) and (ii): Consider first the hypothesis $H_0 : \theta = \theta_0$ and some simple alternative $\theta_1 > \theta_0$. The most desirable points for rejection are those for which $r(x) = p_{\theta_1}(x)/p_{\theta_0}(x) = g[T(x)]$ is sufficiently large. If $T(x) < T(x')$, then $r(x) \leq r(x')$

---

[6] This definition is in terms of specific versions of the densities p_θ. If instead the definition is to be given in terms of the distribution P_θ, various null-set considerations enter which are discussed in Pfanzagl (1967).

and x' is at least as desirable as x. Thus the test which rejects for large values of $T(x)$ is most powerful. As in the proof of Theorem 3.2.1(i), it is seen that there exist C and γ such that (3.16) and (3.17) hold. By Theorem 3.2.1(ii), the resulting test is also most powerful for testing $P_{\theta'}$ against $P_{\theta''}$ at level $\alpha' = \beta(\theta')$ provided $\theta' < \theta''$. Part (ii) of the present theorem now follows from Corollary 3.2.1. Since $\beta(\theta)$ is therefore nondecreasing the test satisfies

$$E_\theta \phi(X) \leq \alpha \quad \text{for} \quad \theta \leq \theta_0. \tag{3.18}$$

The class of tests satisfying (3.18) is contained in the class satisfying $E_{\theta_0} \phi(X) \leq \alpha$. Since the given test maximizes $\beta(\theta_1)$ within this wider class, it also maximizes $\beta(\theta_1)$ subject to (3.18); since it is independent of the particular alternative $\theta_1 > \theta_0$ chosen, it is UMP against K.

(iii) is proved by an analogous argument.

(iv) follows from the fact that the test which minimizes the power for testing a simple hypothesis against a simple alternative is obtained by applying the fundamental lemma (Theorem 3.2.1) with all inequalities reversed.

By interchanging inequalities throughout, one obtains in an obvious manner the solution of the dual problem, $H : \theta \geq \theta_0$, $K : \theta < \theta_0$. ∎

The proof of (i) and (ii) exhibits the basic property of families with monotone likelihood ratio: every pair of parameter values $\theta_0 < \theta_1$ establishes essentially the same preference order of the sample points (in the sense of the preceding section). A few examples of such families, and hence of UMP one-sided tests, will be given below. However, the main applications of Theorem 3.4.1 will come later, when such families appear as the set of conditional distributions given a sufficient statistic (Chapters 4 and 5) and as distributions of a maximal invariant (Chapters 6 and 7).

Example 3.4.1 (Hypergeometric) From a lot containing N items of a manufactured product, a sample of size n is selected at random, and each item in the sample is inspected. If the total number of defective items in the lot is D, the number X of defectives found in the sample has the *hypergeometric* distribution

$$P\{X = x\} = P_D(x) = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}}, \quad \max(0, n + D - N) \leq x \leq \min(n, D).$$

Interpreting $P_D(x)$ as a density with respect to the measure μ that assigns to any set on the real line as measure the number of integers $0, 1, 2, \ldots$ that it contains, and nothing that for values of x within its range

$$\frac{P_{D+1}(x)}{P_D(x)} = \begin{cases} \frac{D+1}{N-D}\frac{N-D-n+x}{D+1-x} & \text{if } n + D + 1 - N \leq x \leq D, \\ 0 \text{ or } \infty & \text{if } x = n + D - N \text{ or } D + 1, \end{cases}$$

it is seen that the distributions satisfy the assumption of monotone likelihood ratios with $T(x) = x$. Therefore there exists a UMP test for testing the hypothesis

$H : D \leq D_0$ against $K : D > D_0$, which rejects H when X is too large, and an analogous test for testing $H' : D \geq D_0$. ∎

An important class of families of distributions that satisfy the assumptions of Theorem 3.4.1 are the *one-parameter exponential families*.

Corollary 3.4.1 *Let θ be a real parameter, and let X have probability density (with respect to some measure μ)*

$$p_\theta(x) = C(\theta)e^{Q(\theta)T(x)}h(x), \tag{3.19}$$

where Q is strictly monotone. Then there exists a UMP test ϕ for testing $H : \theta \leq \theta_0$ against $K : \theta > \theta_0$. If Q is increasing,

$$\phi(x) = 1, \gamma, 0 \quad as \quad T(x) >, =, < C,$$

where C and γ are determined by $E_{\theta_0}\phi(X) = \alpha$. If Q is decreasing, the inequalities are reversed.

A converse of Corollary 3.4.1 is given by Pfanzagl (1968), who shows under weak regularity conditions that the existence of UMP tests against one-sided alternatives for all sample sizes and one value of α implies an exponential family.

As in Example 3.4.1, we shall denote the right-hand side of (3.19) by $P_\theta(x)$ instead of $p_\theta(x)$ when it is a probability, that is, when X is discrete and μ is counting measure.

Example 3.4.2 (Binomial) The binomial distributions $b(p, n)$ with

$$P_p(x) = \binom{n}{x} p^x (1 - p)^{n-x}$$

satisfy (3.19) with $T(x) = x$, $\theta = p$, $Q(p) = \log[p/(1 - p)]$. The problem of testing $H : p \geq p_0$ arises, for instance, in the situation of Example 3.4.1 if one supposes that the production process is in statistical control, so that the various items constitute independent trials with constant probability p of being defective. The number of defectives X in a sample of size n is then sufficient statistic for the distribution of the variables X_i $(i = 1, \ldots, n)$, where X_i is 1 or 0 as the ith item drawn is defective or not, and X is distributed as $b(p, n)$. There exists therefore a UMP test of H, which rejects H when X is too small.

An alternative sampling plan which is sometimes used in binomial situations is *inverse binomial sampling*. Here the experiment is continued until a specified number m of successes—for example, cures effected by some new medical treatment—has been obtained. If Y_i denotes the number of trials after the $(i - 1)$st success, up to but not including the ith success, the probability that $Y_i = y$ is pq^y for $y = 0, 1, \ldots$, so that the joint distribution of $Y_1, \ldots, Y_m$ is

$$P_p(y_1, \ldots, y_m) = p^m q^{\sum y_i}, \qquad y_k = 0, 1, \ldots, \quad k = 1, \ldots, m.$$

This is an exponential family with $T(y) = \sum y_i$ and $Q(p) = \log(1 - p)$. Since $Q(p)$ is a decreasing function of p, the UMP test of $H : p \leq p_0$ rejects H when T is too small. This is what one would expect, since the realization of m successes in only a few more than m trials indicates a high value of p. The test statistic T, which is the number of trials required in excess of m to get m successes, has the negative binomial distribution (Problem 1.1(i))

$$P(t) = \binom{m + t - 1}{m - 1} p^m q^t, \qquad t = 0, 1, \ldots .. \blacksquare$$

Example 3.4.3 (Poisson) If $X_1, \ldots, X_n$ are independent Poisson variables with $E(X_i) = \lambda$, their joint distribution is

$$P_\lambda(x_1, \ldots, x_n) = \frac{\lambda^{x_1 + \cdots + x_n}}{x_1! \cdots x_n!} e^{-n\lambda}.$$

This constitutes an exponential family with $T(x) = \sum x_i$, and $Q(\lambda) = \log \lambda$. One-sided hypotheses concerning λ might arise if λ is a bacterial density and the X's are a number of bacterial counts, or if the X's denote the number of α-particles produced in equal time intervals by a radioactive substance, etc. The UMP test of the hypothesis $\lambda \leq \lambda_0$ rejects when $\sum X_i$ is too large. Here the test statistic $\sum X_i$ has itself a Poisson distribution with parameter $n\lambda$.

Instead of observing the radioactive material for given time periods or counting the number of bacteria in given areas of a slide, one can adopt an inverse sampling method. The experiment is then continued, or the area over which the bacteria are counted is enlarged, until a count of m has been obtained. The observations consist of the times $T_1, \ldots, T_m$ that it takes for the first occurrence, from the first to the second, and so on. If one is dealing with a Poisson process and the number of occurrences in a time or space interval τ has the distribution

$$P(x) = \frac{(\lambda\tau)^x}{x!} e^{-\lambda\tau}, \qquad x = 0, 1, \ldots,$$

then the observed times are independently distributed, each with the exponential density $\lambda e^{-\lambda t}$ for $t \geq 0$ (Problem 1.1(ii)). The joint densities

$$p_\lambda(t_1, \ldots, t_m) = \lambda^m \exp\left(-\lambda \sum_{i=1}^{m} t_i\right), \qquad t_1, \ldots, t_m \geq 0,$$

form an exponential family with $T(t_1, \ldots, t_m) = \sum t_i$ and $Q(\lambda) = -\lambda$. The UMP test of $H : \lambda \leq \lambda_0$ rejects when $T = \sum T_i$ is too small. Since $2\lambda T_i$ has density $\frac{1}{2} e^{-u/2}$ for $u \geq 0$, which is the density of a χ^2-distribution with 2 degrees of freedom, $2\lambda T$ has a χ^2-distribution with $2m$ degrees of freedom. The boundary of the rejection region can therefore be determined from a table of χ^2. $\blacksquare$

The formulation of the problem of hypothesis testing given at the beginning of the chapter takes account of the losses resulting from wrong decisions only in terms of the two types of error. To obtain a more detailed description of the problem of testing $H : \theta \leq \theta_0$ against the alternatives $\theta > \theta_0$, one can consider it as a decision problem with the decisions d_0 and d_1 of accepting and rejecting H and a loss function $L(\theta, d_i) = L_i(\theta)$. Typically, $L_0(\theta)$ will be 0 for $\theta \leq \theta_0$ and strictly increasing for $\theta \geq \theta_0$, and $L_1(\theta)$ will be strictly decreasing for $\theta \leq \theta_0$ and equal to 0 for $\theta \geq \theta_0$. The difference then satisfies

$$L_1(\theta) - L_0(\theta) \gtrless 0 \quad \text{as} \quad \theta \lessgtr \theta_0. \tag{3.20}$$

The following theorem is a special case of complete class results of Karlin and Rubin (1956) and Brown et al. (1976).

Theorem 3.4.2 (i) *Under the assumptions of Theorem 3.4.1, the family of tests given by (3.16) and (3.17) with $0 \leq \alpha \leq 1$ is essentially complete provided the loss function satisfies (3.20).*

(ii) *This family is also minimal essentially complete if the set of points x for which $p_\theta(x) > 0$ is independent of θ.*

PROOF. (i): The risk function of any test ϕ is

$$R(\theta, \phi) = \int p_\theta(x)\{\phi(x)L_1(\theta) + [1 - \phi(x)]L_0(\theta)\} \, d\mu(x)$$

$$= \int p_\theta(x)\{L_0(\theta) + [L_1(\theta) - L_0(\theta)]\phi(x)\} \, d\mu(x),$$

and hence the difference of two risk functions is

$$R(\theta, \phi') - R(\theta, \phi) = [L_1(\theta) - L_0(\theta)] \int (\phi' - \phi)p_\theta \, d\mu.$$

This is ≤ 0 for all θ if

$$\beta_{\phi'}(\theta) - \beta_\phi(\theta) = \int (\phi' - \phi)p_\theta \, d\mu \gtrless 0 \quad \text{for} \quad \theta \gtrless \theta_0.$$

Given any test ϕ, let $E_{\theta_0}\phi(X) = \alpha$. It follows from Theorem 3.4.1(i) that there exists a UMP level-α test ϕ' for testing $\theta = \theta_0$ against $\theta > \theta_0$, which satisfies (3.16) and (3.17). By Theorem 3.4.1(iv), ϕ' also minimizes the power for $\theta < \theta_0$. Thus the two risk functions satisfy $R(\theta, \phi') \leq R(\theta, \phi)$ for all θ, as was to be proved.

(ii): Let ϕ_α and $\phi_{\alpha'}$ be of sizes $\alpha < \alpha'$ and UMP for testing θ_0 against $\theta > \theta_0$. Then $\beta_{\phi_\alpha}(\theta) < \beta_{\phi_{\alpha'}}(\theta)$ for all $\theta > \theta_0$ unless $\beta_{\phi_\alpha}(\theta) = 1$. By considering the problem of testing $\theta = \theta_0$ against $\theta < \theta_0$, it is seen analogously that this inequality also holds for all $\theta < \theta_0$ unless $\beta_{\phi_{\alpha'}}(\theta) = 0$. Since the exceptional possibilities are excluded by

the assumptions, it follows that $R(\theta, \phi') \underset{>}{\overset{<}{}} R(\theta, \phi)$ as $\theta \underset{<}{\overset{>}{}} \theta_0$. Hence each of the two risk functions is better than the other for some values of θ. ∎

The class of tests previously derived as UMP at the various significance levels α is now seen to constitute an essentially complete class for a much more general decision problem, in which the loss function is only required to satisfy certain broad qualitative conditions. From this point of view, the formulation involving the specification of a level of significance can be considered a simple way of selecting a particular procedure from an essentially complete family.

The property of monotone likelihood ratio defines a very strong ordering of a family of distributions. For later use, we consider also the following somewhat weaker definition. A family of cumulative distribution functions F_θ on the real line is said to be *stochastically increasing* (and the same term is applied to random variables possessing these distributions) if the distributions are distinct and if $\theta < \theta'$ implies $F_\theta(x) \geq F_{\theta'}(x)$ for all x. If then X and X' have distributions F_θ and F'_θ, respectively, it follows that $P\{X > x\} \leq P\{X' > x\}$ for all x so that X' tends to have larger values than X. In this case the variable X' is said to be *stochastically larger* than X. This relationship is made more intuitive by the following characterization of the stochastic ordering of two distributions.

Lemma 3.4.1 *Let F_0 and F_1 be two cumulative distribution functions on the real line. Then $F_1(x) \leq F_0(x)$ for all x if and only if there exist two nondecreasing functions f_0 and f_1, and a random variable V such that (a) $f_0(v) \leq f_1(v)$ for all v, and (b) the distributions of $f_0(V)$ and $f_1(V)$ are F_0 and F_1, respectively.*

PROOF. Suppose first that the required f_0, f_1 and V exist. Then

$$F_1(x) = P\{f_1(V) \leq x\} \leq P\{f_0(V) \leq x\} = F_0(x)$$

for all x. Conversely, suppose that $F_1(x) \leq F_0(x)$ for all x, and let $f_i(y) = \inf\{x : F_i(x - 0) \leq y \leq F_1(x)\}, i = 0, 1$. These functions are nondecreasing and for $f_i = f, F_i = F$ satisfy

$$f[F(x)] \leq x \text{ and } F[f(y)] \geq y \quad \text{for all } x \text{ and } y.$$

It follows that $y \leq F(x_0)$ implies $f(y) \leq f[F(x_0)] \leq x_0$ and that conversely $f(y) \leq x_0$ implies $F[f(y)] \leq F(x_0)]$ and hence $y \leq F(x_0)$, so that the two inequalities $f(y) \leq x_0$ and $y \leq F(x_0)$ are equivalent. Let V be uniformly distributed on $(0,1)$. Then $P\{f_i(V) \leq x\} = P\{V \leq F_i(x)\} = F_i(x)$. Since $F_i(x) \leq F_0(x)$ for all x implies $f_0(y) \leq f_1(y)$ for all y, this completes the proof. ∎

One of the simplest examples of a stochastically ordered family is a location parameter family, that is, a family satisfying

$$F_\theta(x) = F(x - \theta).$$

To see that this is stochastically increasing, let X be a random variable with distribution $F(x)$. Then $\theta < \theta'$ implies

$$F(x - \theta) = P\{X \le x - \theta\} \ge P\{X \le x - \theta'\} = F(x - \theta'),$$

as was to be shown.

Another example is furnished by families with monotone likelihood ratio. This is seen from the following lemma, which establishes some basic properties of these families.

Lemma 3.4.2 *Let $p_\theta(x)$ be a family of densities on the real line with monotone likelihood ratio in x.*

(i) *If ψ is a nondecreasing function of x, then $E_\theta \psi(X)$ is a nondecreasing function of θ; if $X_1, \ldots, X_n$ are independently distributed with density p_θ and ψ' is a function of $x_1, \ldots, x_n$ which is nondecreasing in each of its arguments, then $E_\theta \psi'(X_1, \ldots, X_n)$ is a nondecreasing function of θ.*

(ii) *For any $\theta < \theta'$, the cumulative distribution functions of X under θ and θ' satisfy*

$$F_{\theta'}(x) \le F_\theta(x) \qquad \text{for all } x.$$

(iii) *Let ψ be a function with a single change of sign. More specifically, suppose there exists a value x_0 such that $\psi(x) \le 0$ for $x < x_0$ and $\psi(x) \ge 0$ for $x \ge x_0$. Then there exists θ_0 such that $E_\theta \psi(X) \le 0$ for $\theta < \theta_0$ and $E_\theta \psi(X) \ge 0$ for $\theta > \theta_0$, unless $E_\theta \psi(X)$ is either positive for all θ or negative for all θ.*

(iv) *Suppose that $p_\theta(x)$ is positive for all θ and all x, that $p_{\theta'}(x)/p_\theta(x)$ is strictly increasing in x for $\theta < \theta'$, and that $\psi(x)$ is as in (iii) and is $\ne 0$ with positive probability. If $E_{\theta_0} \psi(X) = 0$, then $E_\theta \psi(X) < 0$ for $\theta < \theta_0$ and > 0 for $\theta > \theta_0$.*

PROOF. (i): Let $\theta < \theta'$, and let A and B be the sets for which $p_{\theta'}(x) < p_\theta(x)$ and $p_{\theta'}(x) > p_\theta(x)$ respectively. If $a = \sup_A \psi(x)$ and $b = \inf_B \psi(x)$, then $b - a \ge 0$ and

$$\int \psi(p_{\theta'} - p_\theta)\, d\mu \ge a \int_A (p_{\theta'} - p_\theta)\, d\mu + b \int_B (p_{\theta'} - p_\theta)\, d\mu$$

$$= (b - a) \int_B (p_{\theta'} - p_\theta)\, d\mu \ge 0,$$

which proves the first assertion. The result for general n follows by induction.

(ii): This follows from (i) by letting $\psi(x) = 1$ for $x > x_0$ and $\psi(x) = 0$ otherwise.

(iii): We shall show first that for any $\theta' < \theta''$, $E_{\theta'} \psi(X) > 0$ implies $E_{\theta''} \psi(X) \ge 0$. If $p_{\theta''}(x_0)/p_{\theta'}(x_0) = \infty$, then $p_{\theta'}(x) = 0$ for $x \ge x_0$ and hence $E_{\theta'} \psi(X) \le 0$. Suppose therefore that $p_{\theta''}(x_0)/p_{\theta'}(x_0) = c < \infty$. Then $\psi(x) \ge 0$ on the set $S = \{x : p_{\theta'}(x) = 0 \text{ and } p_{\theta''}(x) > 0\}$, and

$$E_{\theta''}\psi(X) \geq \int_{\tilde{S}} \psi \frac{p_{\theta''}}{p_{\theta'}} p_{\theta'} \, d\mu$$

$$\geq \int_{-\infty}^{x_0-} c\psi p_{\theta'} \, d\mu + \int_{x_0}^{\infty} c\psi p_{\theta'} \, d\mu = cE_{\theta'}\psi(X) \geq 0.$$

The result now follows by letting $\theta_0 = \inf\{\theta : E_\theta\psi(X) > 0\}$.

(iv): The proof is analogous to that of (iii). ∎

Part (ii) of the lemma shows that any family of distributions with monotone likelihood ratio in x is stochastically increasing. That the converse does not hold is shown for example by the Cauchy densities

$$\frac{1}{\pi} \frac{1}{1 + (x - \theta)^2}.$$

The family is stochastically increasing, since θ is a location parameter; however, the likelihood ratio is not monotone. Conditions under which a location parameter family possesses monotone likelihood ratio are given in Example 8.2.1.

Lemma 3.4.2 is a special case of a theorem of Karlin (1957, 1968) relating the number of sign changes of $E_\theta\psi(X)$ to those of $\psi(x)$ when the densities $p_\theta(x)$ are *totally positive* (defined in Problem 3.55). The application of totally positive—or equivalently, variation diminishing—distributions to statistics is discussed by Brown et al. (1981); see also Problem 3.58.

3.5 Confidence Bounds

The theory of UMP one-sided tests can be applied to the problem of obtaining a lower or upper bound for a real-valued parameter θ. The problem of setting a lower bound arises, for example, when θ is the breaking strength of a new alloy; that of setting an upper bound, when θ is the toxicity of drug or the probability of an undesirable event. The discussion of lower and upper bounds is completely parallel, and it is therefore enough to consider the case of a lower bound, say $\underline{\theta}$.

Since $\underline{\theta} = \underline{\theta}(X)$ will be a function of the observations, it cannot be required to fall below θ with certainty, but only with specified high probability. One selects a number $1 - \alpha$, the *confidence level*, and restricts attention to bounds $\underline{\theta}$ satisfying

$$P_\theta\{\underline{\theta}(X) \leq \theta\} \geq 1 - \alpha \quad \text{for all } \theta. \tag{3.21}$$

The function $\underline{\theta}$ is called a lower *confidence bound* for θ at confidence level $1 - \alpha$; the infimum of the left-hand side of (3.21), which in practice will be equal to $1 - \alpha$, is called the *confidence coefficient* of $\underline{\theta}$.

Subject to (3.21), $\underline{\theta}$ should underestimate θ by as little as possible. One can ask, for example, that the probability of $\underline{\theta}$ falling below any $\theta' < \theta$ should be a minimum.

A function $\underline{\theta}$ for which

$$P_\theta\{\underline{\theta}(X) \le \theta'\} = \text{ minimum} \tag{3.22}$$

for all $\theta' < \theta$ subject to (3.21) is a uniformly *most accurate* lower confidence bound for θ at confidence level $1 - \alpha$.

Let $L(\theta, \underline{\theta})$ be a measure of the loss resulting from underestimating θ, so that for each fixed θ the function $L(\theta, \underline{\theta})$ is defined and nonnegative for $\underline{\theta} < \theta$, and is nonincreasing in this second argument. One would then wish to minimize

$$E_\theta L(\theta, \underline{\theta}) \tag{3.23}$$

subject to (3.21). It can be shown that a uniformly most accurate lower confidence bound $\underline{\theta}$ minimizes (3.23) subject to (3.21) for every such loss function L. (See Problem 3.49.)

The derivation of uniformly most accurate confidence bounds is facilitated by introducing the following more general concept, which will be considered in more detail in Chapter 5. A family of subsets $S(x)$ of the parameter space Ω is said to constitute a family of *confidence sets* at confidence level $1 - \alpha$ if

$$P_\theta\{\theta \in S(X)\} \ge 1 - \alpha \quad \text{ for all } \quad \theta \in \Omega, \tag{3.24}$$

that is, if the random sets $S(X)$ covers the true parameter point with probability $\ge 1 - \alpha$. A lower confidence bound corresponds to the special case that $S(x)$ is a one-sided interval

$$S(x) = \{\theta : \underline{\theta}(x) \le \theta < \infty\}.$$

Theorem 3.5.1 (i) *For each $\theta_0 \in \Omega$ let $A(\theta_0)$ be the acceptance region of a level-α test for testing $H(\theta_0) : \theta = \theta_0$, and for each sample point x let $S(x)$ denote the set of parameter values*

$$S(x) = \{\theta : x \in A(\theta), \theta \in \Omega\}.$$

Then $S(x)$ is a family of confidence sets for θ at confidence level $1 - \alpha$.

(ii) *If for all θ_0, $A(\theta_0)$ is UMP for testing $H(\theta_0)$ at level α against the alternatives $K(\theta_0)$, then for each $\theta_0 \notin \Omega$, $S(X)$ minimizes the probability*

$$P_\theta\{\theta_0 \in S(X)\} \quad \text{ for all } \quad \theta \in K(\theta_0)$$

among all level $1 - \alpha$ families of confidence sets for θ.

PROOF. (i) By definition of $S(x)$,

$$\theta \in S(x) \quad \text{ if and only if } \quad x \in A(\theta), \tag{3.25}$$

and hence

$$P_\theta\{\theta \in S(X)\} = P_\theta\{X \in A(\theta)\} \geq 1 - \alpha.$$

(ii) If $S^*(x)$ is any other family of confidence sets at level $1 - \alpha$, and if $A^*(\theta) = \{x : \theta \in S^*(x)\}$, then

$$P_\theta\{X \in A^*(\theta)\} = P_\theta\{\theta \in S^*(X)\} \geq 1 - \alpha$$

so that $A^*(\theta_0)$ is the acceptance region of a level-α test of $H(\theta_0)$. It follows from the assumed property of $A(\theta_0)$ that for any $\theta \in K(\theta_0)$

$$P_\theta\{X \in A^*(\theta_0)\} \geq P_\theta\{X \in A(\theta_0)\}$$

and hence that

$$P_\theta\{\theta_0 \in S^*(X)\} \geq P_\theta\{\theta_0 \in S(X)\},$$

as was to be proved. ∎

The equivalence (3.25) shows the structure of the confidence sets $S(x)$ as the totality of parameter values θ for which the hypothesis $H(\theta)$ is accepted when x is observed. A confidence set can therefore be viewed as a combined statement regarding the tests of the various hypotheses $H(\theta)$, which exhibits the values for which the hypothesis is accepted [$\theta \in S(x)$] and those for which it is rejected [$\theta \in \bar{S}(x)$]. Such a method of constructing confidence sets for parameters is known as "test inversion".

Note that a lower confidence bound $\underline{\theta}$ satisfying (3.21) corresponds to the interval $[\underline{\theta}, \infty)$. However, one can typically also conclude the open interval $(\underline{\theta}, \infty)$ contains the true θ with probability at least $1 - \alpha$. This occurs under the conditions of Corollary 3.5.1 below. So, the resulting confidence interval may or may not include the endpoint, and those obtained through "test inversion" will include the endpoint or not depending on the exact construction of the tests. The first definition where the confidence set includes the lower endpoint was a convenient way to initiate the discussion. In practice, confidence intervals are usually presented with the endpoints included. Certainly, if the open interval satisfies the coverage constraint, so does its closure.

Corollary 3.5.1 *Let the family of densities $p_\theta(x)$, $\theta \in \Omega$, have monotone likelihood ratio in $T(x)$, and suppose that the cumulative distribution function $F_\theta(t)$ of $T = T(X)$ is a continuous function in each of the variables t and θ when the other is fixed.*

(i) There exists a uniformly most accurate confidence bound $\underline{\theta}$ for θ at each confidence level $1 - \alpha$.

(ii) If x denotes the observed values of X and $t = T(x)$, and if the equation

$$F_\theta(t) = 1 - \alpha \tag{3.26}$$

has a solution $\theta = \hat{\theta}$ in Ω then this solution is unique and $\underline{\theta}(x) = \hat{\theta}$.

PROOF. (i): There exists for each θ_0 a constant $C(\theta_0)$ such that

$$P_{\theta_0}\{T > C(\theta_0)\} = \alpha,$$

and by Theorem 3.4.1, $T > C(\theta_0)$ is a UMP level-α rejection region for testing $\theta = \theta_0$ against $\theta > \theta_0$. By Corollary 3.2.1, the power of this test against any alternative $\theta_1 > \theta_0$ exceeds α, and hence $C(\theta_0) < C(\theta_1)$ so that the function C is strictly increasing; it is also continuous. Let $A(\theta_0)$ denote the acceptance region $T \leq C(\theta_0)$, and let $S(x)$ be defined by (3.25). If follows from the monotonicity of the function C that $S(x)$ consists of those values $\theta \in \Omega$ which satisfy $\underline{\theta} \leq \theta$, where

$$\underline{\theta} = \inf\{\theta : T(x) \leq C(\theta)\}.$$

By Theorem 3.5.1, the sets $\{\theta : \underline{\theta}(x) \leq \theta\}$, restricted to possible values of the parameter, constitute a family of confidence sets at level $1 - \alpha$, which minimize $P_\theta\{\underline{\theta} \leq \theta'\}$ for all $\theta \in K(\theta')$, that is, for all $\theta > \theta'$. This shows $\underline{\theta}$ to be a uniformly most accurate confidence bound for θ.

(ii): It follows from Corollary 3.2.1 that $F_\theta(t)$ is a strictly decreasing function of θ at any point t for which $0 < F_\theta(t) < 1$, and hence that (3.26) can have at most one solution. Suppose now that t is the observed value of T and that the equation $F_\theta(t) = 1 - \alpha$ has the solution $\hat{\theta} \in \Omega$. Then $F_{\hat{\theta}}(t) = 1 - \alpha$, and by definition of the function C, $C(\hat{\theta}) = t$. The inequality $t \leq C(\theta)$ is then equivalent to $C(\hat{\theta}) \leq C(\theta)$ and hence to $\hat{\theta} \leq \theta$. It follows that $\underline{\theta} = \hat{\theta}$, as was to be proved. ∎

Under the same assumptions, the corresponding upper confidence bound with confidence coefficient $1 - \alpha$ is the solution $\bar{\theta}$ of the equation $P_\theta\{T \geq t\} = 1 - \alpha$ or equivalently of $F_\theta(t) = \alpha$.

Example 3.5.1 (Exponential waiting times) To determine an upper bound for the degree of radioactivity λ of a radioactive substance, the substance is observed until a count of m has been obtained on a Geiger counter. Under the assumptions of Example 3.4.3, the joint probability density of the times $T_i(i = 1, \ldots, m)$ elapsing between the $(i - 1)$st count and the ith one is

$$p(t_1, \ldots, t_m) = \lambda^m e^{-\lambda \sum t_i}, \qquad t_1, \ldots, t_m \geq 0.$$

If $T = \sum T_i$ denotes the total time of observation, then $2\lambda T$ has a χ^2-distribution with $2m$ degrees of freedom, and, as was shown in Example 3.4.3, the acceptance region of the most powerful test of $H(\lambda_0) : \lambda = \lambda_0$ against $\lambda < \lambda_0$ is $2\lambda_0 T \leq C$, where C is determined by the equation

$$\int_0^C \chi^2_{2m} = 1 - \alpha .$$

The set $S(t_1, \ldots, t_m)$ defined by (3.25) is then the set of values λ such that $\lambda \leq C/2T$, and it follows from Theorem 3.5.1 that $\bar{\lambda} = C/2T$ is a uniformly most accurate upper confidence bound for λ. This result can also be obtained through Corollary 3.5.1. ∎

If the variables X or T are discrete, Corollary 3.5.1 cannot be applied directly since the distribution functions $F_\theta(t)$ are not continuous, and for most values of θ_0 the optimum test of $H : \theta = \theta_0$ is randomized. However, any randomized test based on X has the following representation as a nonrandomized test depending on X and an independent variable U distributed uniformly over (0, 1). Given a critical function ϕ, consider the rejection region

$$R = \{(x, u) : u \leq \phi(x)\}.$$

Then

$$P\{(X, U) \in R\} = P\{U \leq \phi(X)\} = E_\phi(X),$$

whatever the distribution of X, so that R has the same power function as ϕ and the two tests are equivalent. The pair of variables (X, U) has a particularly simple representation when X is integer-valued. In this case the statistic

$$T = X + U$$

is equivalent to the pair (X, U), since with probability 1

$$X = [T], \qquad U = T - [T],$$

where $[T]$ denotes the largest integer $\leq T$. The distribution of T is continuous, and confidence bounds can be based on this statistic.

Example 3.5.2 (Binomial) An upper bound is required for a binomial probability p—for example, the probability that a batch of polio vaccine manufactured according to a certain procedure contains any live virus. Let $X_1, \ldots, X_n$ denote the outcome of n trials, X_i being 1 or 0 with probabilities p and q respectively, and let $X = \sum X_i$. Then $T = X + U$ has probability density

$$\binom{n}{[t]} p^{[t]} q^{n-[t]}, \qquad 0 \leq t < n + 1.$$

This satisfies the conditions of Corollary 3.5.1, and the upper confidence bound $\bar{p}$ is therefore the solution, if it exists, of the equation

$$P_p\{T < t\} = \alpha, \tag{3.27}$$

where t is the observed value of T. A solution does exist for all values $\alpha \leq t \leq n + \alpha$. For $n + \alpha < t$, the hypothesis $H(p_0) : p = p_0$ is accepted against the alternative $p < p_0$ for all values of p_0 and hence $\bar{p} = 1$. For $t < \alpha$, $H(p_0)$ is rejected for all

values of p_0 and the confidence set $S(t)$ is therefore empty. Consider instead the sets $S^*(t)$ which are equal to $S(t)$ for $t \geq \alpha$ and which for $t < \alpha$ consist of the single point $p = 0$. They are also confidence sets at level $1 - \alpha$, since for all p,

$$P_p\{p \in S^*(T)\} \geq P_p\{p \in S(T)\} = 1 - \alpha.$$

On the other hand, $P_p\{p' \in S^*(T)\} = P_p\{p' \in S(T)\}$ for all $p' > 0$ and hence

$$P_p\{p' \in S^*(T)\} = P_p\{p' \in S(T)\} \qquad \text{for all} \quad p' > p.$$

Thus the family of sets $S^*(t)$ minimizes the probability of covering p' for all $p' > p$ at confidence level $1 - \alpha$. The associated confidence bound $\bar{p}^*(t) = \bar{p}(t)$ for $t \geq \alpha$ and $\bar{p}^*(t) = 0$ for $t < \alpha$ is therefore a uniformly most accurate upper confidence bound for p at level $1 - \alpha$.

In practice, so as to avoid randomization and obtain a bound not dependent on the extraneous variable U, one usually replaces T by $X + 1 = [T] + 1$. Since $\bar{p}^*(t)$ is a nondecreasing function of t, the resulting upper confidence bound $\bar{p}^*([t] + 1)$ is then somewhat larger than necessary; as a compensation it also gives a correspondingly higher probability of not falling below the true p.

Equivalently, rather than finding the solution to (3.27), first note that for $x = [t]$,

$$P_p\{T < t\} \leq P_p\{T < x + 1\} = P_p\{X \leq x\} .$$

Therefore, a conservative solution is to find the value of p, say $\hat{p}_U$, satisfying

$$P_p\{X \leq x\} = \alpha ,$$

where x is the observed value of X. When $x = n$, there is no solution, but then $\hat{p}_U = 1$ serves as the upper confidence bound. Otherwise, the value $\hat{p}_U$ of p for which

$$\sum_{j=0}^{x} \binom{n}{j} p^j (1-p)^{n-j} = \alpha$$

serves as an upper $1 - \alpha$ confidence bound for p. For $x = 0$, the solution is $\hat{p}_U = 1 - \alpha^{1/n}$. In fact, for $0 < x < n$, $\hat{p}_U$ can be expressed as the $1 - \alpha$ quantile of the Beta distribution with parameters $x + 1$ and $n - x$ (Problem 3.48). Similarly, a lower $1 - \alpha$ confidence bound is the value $\hat{p}_L$ satisfying

$$\sum_{j=x}^{n} \binom{n}{j} p^j (1-p)^{n-j} = \alpha .$$

When $0 < x < n$, $\hat{p}_L$ can be expressed as the α quantile of the Beta distribution with parameters x and $n - x + 1$. If $x = 0$, then we take $\hat{p}_U = 0$, and if $x = n$,

the solution is $\alpha^{1/n}$. The interval $[\hat{p}_L, \hat{p}_U]$ then serves as a level $1 - 2\alpha$ confidence interval for p.

The conservative solution $\hat{p}_U$ dates back to Clopper and Pearson (1934). References to tables for the confidence bounds and a careful discussion of various approximations can be found in Hall (1982) and Blyth (1986). Two-sided intervals will be discussed in Example 5.5.4. Large-sample approaches will be discussed in Example 11.3.4. ∎

More generally, when X or T is discrete, a reasonable approach to constructing a $1 - \alpha$ upper confidence bound for some parameter θ that avoids randomization is the following (whether or not T is optimal in any sense). Let $F_\theta(t) = P_\theta\{T \leq t\}$. Assume, for fixed t, $F_\theta(t)$ is continuous and strictly monotone decreasing in θ. Suppose $\bar{\theta}$ satisfies

$$F_{\bar{\theta}}(t) = \alpha , \tag{3.28}$$

where t is the observed value of T.[7] Then, $\bar{\theta}$ serves as a level $1 - \alpha$ upper confidence bound for θ. In Figure 3.2, $F_\theta(t)$ is plotted as a function of θ with $T = t$ fixed. The solution to (3.28) is shown as $\bar{\theta}$. Alternatively, Figure 3.3 shows how the confidence bounds may be obtained from the inverse function $F_\theta^{-1}(\alpha)$. (In Figure 3.3, the functions displayed are linear in θ, as would be the case in a location model, though this is not generally the case.)

The confidence bounds can also be derived simply from "test inversion". Indeed, the test of $H(\theta_0)$ against $\theta < \theta_0$ that rejects $H(\theta_0)$ for small values of T has (possibly conservative) p-value given by $F_{\theta_0}(t)$; see Example 3.3.3. Therefore, any θ_0 for which $F_{\theta_0}(t) > \alpha$ should be included in the confidence region, while any θ_0 for which $F_{\theta_0}(t) \leq \alpha$ should not be included in the confidence region. Such an approach is consistent with the solution to (3.28) when such a solution exists, but it also applies if no solution exists (as can happen in the binomial example). Similarly, a level $1 - \alpha$ lower confidence band for θ, $\underline{\theta}$ may be obtained as the solution to the equation $F_\theta^-(t) = 1 - \alpha$; if no solution exists, the region consists of all θ_0 for which $F_{\theta_0}^-(t) < 1 - \alpha$.

Let $\underline{\theta}$ and $\bar{\theta}$ be lower and upper bounds for θ with confidence coefficients $1 - \alpha_1$ and $1 - \alpha_2$ respectively and suppose that $\underline{\theta}(x) < \bar{\theta}(x)$ for all x. This will be the case under the assumptions of Corollary 3.5.1 if $\alpha_1 + \alpha_2 < 1$. The intervals $(\underline{\theta}, \bar{\theta})$ are then *confidence intervals* for θ with confidence coefficient $1 - \alpha_1 - \alpha_2$; that is, they contain the true parameter value with probability $1 - \alpha_1 - \alpha_2$, since
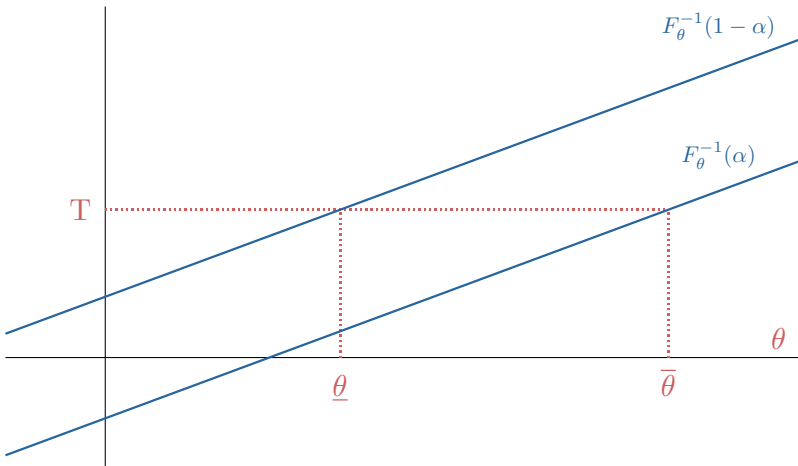
$$P_\theta\{\underline{\theta} \leq \theta \leq \bar{\theta}\} = 1 - \alpha_1 - \alpha_2 \qquad \text{for all } \theta.$$

---

[7] An explicit solution for the value of θ satisfying (3.28) may be unavailable, and one may resort to approximate numerical approaches, such as by discretization of θ or the "automatic percentile" method of DiCiccio and Romano (1989).
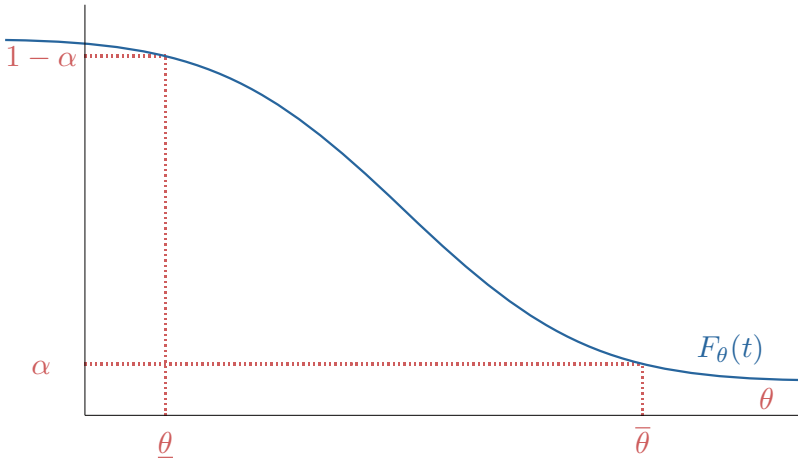
Figure 3.2 Confidence Bounds from F_θ

Figure 3.3 Confidence Bounds from F_θ^{-1}

If $\underline{\theta}$ and $\bar{\theta}$ are uniformly most accurate, they minimize $E_\theta L_1(\theta, \underline{\theta})$ and $E_\theta L_2(\theta, \bar{\theta})$ at their respective levels for any function L_1 that is nonincreasing in $\underline{\theta}$ for $\underline{\theta} < \theta$ and 0 for $\underline{\theta} \geq \theta$ and any L_2 that is nondecreasing in $\bar{\theta}$ for $\bar{\theta} > \theta$ and 0 for $\bar{\theta} \leq \theta$. Letting

$$L(\theta; \underline{\theta}, \bar{\theta}) = L_1(\theta, \underline{\theta}) + L_2(\theta, \bar{\theta}),$$

the intervals $(\underline{\theta}, \bar{\theta})$ therefore minimize $E_\theta L(\theta; \underline{\theta}, \bar{\theta})$ subject to

$$P_\theta\{\underline{\theta} > \theta\} \leq \alpha_1, \qquad P_\theta\{\bar{\theta} < \theta\} \leq \alpha_2.$$

An example of such a loss function is

$$L(\theta; \underline{\theta}, \bar{\theta}) = \begin{cases} \bar{\theta} - \underline{\theta} \text{ if } \underline{\theta} \le \theta \le \bar{\theta}, \\ \bar{\theta} - \theta \text{ if } \theta < \underline{\theta}, \\ \theta - \underline{\theta} \text{ if } \bar{\theta} < \theta, \end{cases}$$

which provides a natural measure of the accuracy of the intervals. Other possible measures are the actual length $\bar{\theta} - \underline{\theta}$ of the intervals, or, for example, $a(\theta - \underline{\theta})^2 + b(\bar{\theta} - \theta)^2$, which gives an indication of the distance of the two endpoints from the true value.[8]

An important limiting case corresponds to the levels $\alpha_1 = \alpha_2 = \frac{1}{2}$. Under the assumptions of Corollary 3.5.1 and if the region of positive density is independent of θ so that tests of power 1 are impossible when $\alpha < 1$, the upper and lower confidence bounds $\bar{\theta}$ and $\underline{\theta}$ coincide in this case. The common bound satisfies

$$P_\theta\{\underline{\theta} \le \theta\} = P_\theta\{\underline{\theta} \ge \theta\} = \frac{1}{2},$$

and the estimate $\underline{\theta}$ of θ is therefore as likely to underestimate as to overestimate the true value. An estimate with this property is said to be *median unbiased*. (For the relation of this to other concepts of unbiasedness, see Problem 1.3.) It follows from the above result for arbitrary α_1 and α_2 that among all median unbiased estimates, $\underline{\theta}$ minimizes $EL(\theta, \underline{\theta})$ for any *monotone* loss function, that is, any loss function which for fixed θ has a minimum of 0 at $\underline{\theta} = \theta$ and is nondecreasing as $\underline{\theta}$ moves away from θ in either direction. By taking in particular $L(\theta, \underline{\theta}) = 0$ when $|\theta - \underline{\theta}| \le \triangle$ and $= 1$ otherwise, it is seen that among all median unbiased estimates, $\underline{\theta}$ minimizes the probability of differing from θ by more than any given amount; more generally it maximizes the probability

$$P_\theta\{-\triangle_1 \le \theta - \underline{\theta} < \triangle_2\}$$

for any $\triangle_1, \triangle_2 \ge 0$.

A more detailed assessment of the position of θ than that provided by confidence bounds or intervals corresponding to a fixed level $\gamma = 1 - \alpha$ is obtained by stating confidence bounds for a number of levels, for example upper confidence bounds corresponding to values such as $\gamma = 0.05, 0.1, 0.25, 0.5, 0.75, 0.9, 0.95$. These constitute a set of *standard confidence bounds*,[9] from which different specific intervals or bounds can be obtained in the obvious manner.

---

[8] Proposed by Wolfowitz (1950).

[9] Suggested by Tukey (1949b).

3.6   A Generalization of the Fundamental Lemma

The following is a useful extension of Theorem 3.2.1 to the case of more than one side condition.

Theorem 3.6.1 *Let $f_1, \ldots, f_{m+1}$ be real-valued functions defined on a Euclidean space $\mathcal{X}$ and integrable μ, and suppose that for given constants $c_1, \ldots, c_m$ there exists a critical function ϕ satisfying*

$$\int \phi f_i \, d\mu = c_i, \qquad i = 1, \ldots, m. \tag{3.29}$$

Let $\mathcal{C}$ be the class of critical functions ϕ for which (3.29) holds.
(i) Among all members of $\mathcal{C}$ there exists one that maximizes

$$\int \phi f_{m+1} \, d\mu.$$

(ii) A sufficient condition for a member of $\mathcal{C}$ to maximize

$$\int \phi f_{m+1} \, d\mu$$

is the existence of constants $k_1, \ldots, k_m$ such that

$$\phi(x) = 1 \qquad \text{when} \quad f_{m+1}(x) > \sum_{i=1}^{m} k_i f_i(x),$$

$$\tag{3.30}$$

$$\phi(x) = 0 \qquad \text{when} \quad f_{m+1}(x) < \sum_{i=1}^{m} k_i f_i(x).$$

(iii) If a member of $\mathcal{C}$ satisfies (3.30) with $k_1, \ldots, k_m \geq 0$, then it maximizes

$$\int \phi f_{m+1} \, d\mu$$

among all critical functions satisfying

$$\int \phi f_i \, d\mu \leq c_i, \qquad i = 1, \ldots, m. \tag{3.31}$$

(iv) The set M of points in m-dimensional space whose coordinates are

$$\left(\int \phi f_1 \, d\mu, \ldots, \int \phi f_m \, d\mu \right)$$

for some critical function ϕ is convex and closed. If $(c_1, \ldots, c_m)$ is an inner point[10] of M, then there exist constants $k_1, \ldots, k_m$ and a test ϕ satisfying (3.29) and (3.30), and a necessary condition for a member of $\mathcal{C}$ to maximize

$$\int \phi f_{m+1} \, d\mu$$

is that (3.30) holds a.e. μ.

Here the term "inner point of M" in statement (iv) can be interpreted as meaning a point interior to M relative to m-space or relative to the smallest linear space (of dimension $\leq m$) containing M. The theorem is correct with both interpretations but is stronger with respect to the latter, for which it will be proved.

We also note that exactly analogous results hold for the minimization of $\int \phi f_{m+1} \, d\mu$.

PROOF. (i): Let $\{\phi_n\}$ be a sequence of functions in $\mathcal{C}$ such that $\int \phi_n f_{m+1} \, d\mu$ tends to $\sup_\phi \int \phi f_{m+1} \, d\mu$. By the weak compactness theorem for critical functions (Theorem 3.4.2 of the Appendix), there exists a subsequence $\{\phi_{n_i}\}$ and a critical function ϕ such that

$$\int \phi_{n_i} f_k \, d\mu \to \int \phi f_k \, d\mu \quad \text{for} \quad k = 1, \cdots, m+1.$$

It follows that ϕ is in $\mathcal{C}$ and maximizes the integral with respect to $f_{m+1} \, d\mu$ within $\mathcal{C}$.

(ii) and (iii) are proved exactly as was part (ii) of Theorem 3.2.1.

(iv): That M is closed follows again from the weak compactness theorem, and its convexity is a consequence of the fact that if ϕ_1 and ϕ_2 are critical functions, so is $\alpha\phi_1 + (1 - \alpha)\phi_2$ for any $0 \leq \alpha \leq 1$. If N (see Figure 3.4) is the totality of points in $(m + 1)$-dimensional space with coordinates

$$\left(\int \phi f_1 \, d\mu, \ldots, \int \phi f_{m+1} \, d\mu \right),$$

where ϕ ranges over the class of all critical functions, then N is convex and closed by the same argument. Denote the coordinates of a general point in M and N by $(u_1, \ldots, u_m)$ and $(u_1, \ldots, u_{m+1})$ respectively. The points of N, the first m coordinates of which are $c_1, \ldots, c_m$, form a closed interval $[c^*, c^{**}]$.

---

[10] A discussion of the problem when this assumption is not satisfied is given by Dantzig and Wald (1951).
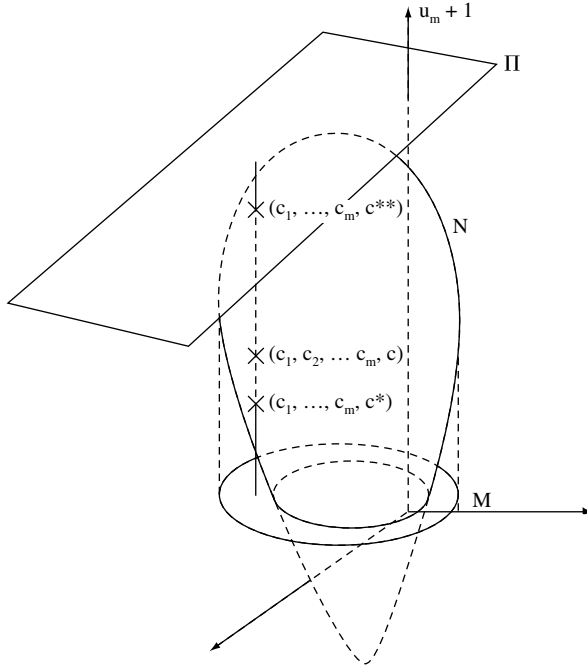
Figure 3.4 The sets M and N

Assume first that $c^* < c^{**}$. Since $(c_1, \ldots, c_m, c^{**})$ is a boundary point of N, there exists a hyperplane $\prod$ through it such that every point on N lies below or on $\prod$. Let the equation of $\prod$ be

$$\sum_{i=1}^{m+1} k_i u_i = \sum_{i=1}^{m} k_i c_i + k_{m+1} c^{**}.$$

Since $(c_1, \ldots, c_m)$ is an inner point of M, the coefficient $k_{m+1} \neq 0$. To see this, let $c^* < c < c^{**}$, so that $(c_1, \ldots c_m, c)$ is an inner point of N. Then there exists a sphere with this point as center lying entirely in N and hence below $\prod$. It follows that the point $(c_1, \ldots c_m, c)$ does not lie on $\prod$ and hence that $k_{m+1} \neq 0$. We may therefore take $k_{m+1} = -1$ and see that for any point of N

$$u_{m+1} - \sum_{i=1}^{m} k_i u_i \leq c^{**}_{m+1} - \sum_{i=1}^{m} k_i c_i.$$

That is, all critical functions ϕ satisfy

$$\int \phi \left(f_{m+1} - \sum_{i=1}^{m} k_i f_i \right) d\mu \leq \int \phi^{**} \left(f_{m+1} - \sum_{i=1}^{m} k_i f_i \right) d\mu,$$

where ϕ^{**} is the test giving rise to the point $(c_1, \ldots, c_m, c^{**})$. Thus ϕ^{**} is the critical function that maximizes the left-hand side of this inequality. Since the integral in question is maximized by putting ϕ equal to 1 when the integrand is positive and equal to 0 when it is negative, ϕ^{**} satisfies (3.30) a.e. μ.

If $c^* = c^{**}$, let $(c'_1, \ldots, c'_m)$ be any point of M other than $(c_1, \ldots, c_m)$. We shall now show that there exists exactly one real number c' such that $(c'_1, \ldots, c'_m, c')$ is in N. Suppose to the contrary that $(c'_1, \ldots, c'_m, \underline{c}')$ and $(c'_1, \ldots, c'_m, \bar{c}')$ are both in N, and consider any point $(c''_1, \ldots, c''_m, c'')$ of N such that $(c_1, \ldots, c_m)$ is an interior point of the line segment joining $(c'_1, \ldots, c'_m)$ and $(c''_1, \ldots, c''_m)$. Such a point exists since $(c_1, \ldots, c_m)$ is an inner point of M. Then the convex set spanned by the three points $(c'_1, \ldots, c'_m, \underline{c}')$, $(c'_1, \ldots, c'_m, \bar{c}')$, and $(c''_1, \ldots, c''_m, c'')$ is contained in N and contains points $(c_1, \ldots, c_m, \underline{c})$ and $(c_1, \ldots, c_m, \bar{c})$ with $\underline{c} < \bar{c}$, which is a contradiction. Since N is convex, contains the origin, and has at most one point on any vertical line $u_1 = c'_1, \ldots, \quad u_m = c'_m$, it is contained in a hyperplane, which passes through the origin and is not parallel to the u_{m+1}-axis. It follows that

$$\int \phi f_{m+1}\, d\mu = \sum_{i=1}^{m} k_i \int \phi f_i \, d\mu$$

for all ϕ. This arises of course only in the trivial case that

$$f_{m+1} = \sum_{i=1}^{m} k_i f_i \qquad \text{a.e. } \mu,$$

and (3.30) is satisfied vacuously. ∎

Corollary 3.6.1 *Let $p_1, \ldots, p_m, p_{m+1}$ be probability densities with respect to a measure μ, and let $0 < \alpha < 1$. Then there exists a test ϕ such that $E_i\phi(X) = \alpha$ $(i = 1, \ldots, m)$ and $E_{m+1}\phi(X) > \alpha$, unless $p_{m+1} = \sum_{i=1}^{m} k_i p_i$, a.e. μ.*

PROOF. The proof will be by induction over m. For $m = 1$ the result reduces to Corollary 3.2.1. Assume now that it has been proved for any set of m distributions, and consider the case of $m + 1$ densities $p_1, \ldots, p_{m+1}$. If $p_1, \ldots, p_m$ are linearly dependent, the number of p_i can be reduced and the result follows from the induction hypothesis. Assume therefore that $p_1, \ldots, p_m$ are linearly independent. Then for each $j = 1, \ldots, m$ there exist by the induction hypothesis tests ϕ_j and ϕ'_j such that $E_i\phi_j(X) = E_i\phi'_j(X) = \alpha$ for all $i = 1, \ldots, j-1, j+1, \ldots, m$ and $E_j\phi_j(X) < \alpha < E_j\phi'_j(X)$. It follows that the point of m-space for which all m coordinates are equal to α is an inner point of M, so that Theorem 3.6.1(iv) is applicable. The test $\phi(x) \equiv \alpha$ is such that $E_i\phi(X) = \alpha$ for $i = 1, \ldots, m$. If among all tests satisfying the side conditions this one is most powerful, it has to satisfy (3.30). Since $0 < \alpha < 1$, this implies

$$p_{m+1} = \sum_{i=1}^{m} k_i \, p_i \qquad \text{a.e.} \mu,$$

as was to be proved. ∎

The most useful parts of Theorems 3.2.1 and 3.6.1 are the parts (ii), which give sufficient conditions for a critical function to maximize an integral subject to certain side conditions. These results can be derived very easily as follows by the method of undetermined multipliers.

Lemma 3.6.1 *Let $F_1, \ldots, F_{m+1}$ be real-valued functions defined over a space U, and consider the problem of maximizing $F_{m+1}(u)$ subject to $F_i(u) = c_i$ $(i = 1, \ldots, m)$. A sufficient condition for a point u^0 satisfying the side conditions to be a solution of the given problem is that among all points of U it maximizes*

$$F_{m+1}(u) - \sum_{i=1}^{m} k_i \, F_i(u)$$

for some $k_1, \ldots, k_m$.

When applying the lemma one usually carries out the maximization for arbitrary k's, and then determines the constants so as to satisfy the side conditions.

PROOF. If u is any point satisfying the side conditions, then

$$F_{m+1}(u) - \sum_{i=1}^{m} k_i \, F_i(u) \le F_{m+1}(u^0) - \sum_{i=1}^{m} k_i \, F_i(u^0),$$

and hence $F_{m+1}(u) \le F_{m+1}(u^0)$.

As an application consider the problem treated in Theorem 3.6.1. Let U be the space of critical functions ϕ, and let $F_i(\phi) = \int \phi f_i \, d\mu$. Then a sufficient condition for ϕ to maximize $F_{m+1}(\phi)$, subject to $F_i(\phi) = c_i$, is that it maximizes $F_{m+1}(\phi) - \sum k_i F_i(\phi) = \int (f_{m+1} - \sum k_i f_i) \phi \, d\mu$. This is achieved by setting $\phi(x) = 1$ or 0 as $f_{m+1}(x) >$ or $< \sum k_i f_i(x)$.

3.7 Two-Sided Hypotheses

UMP tests exist not only for one sided but also for certain two-sided hypotheses of the form

$$H : \theta \le \theta_1 \text{ or } \theta \ge \theta_2 \qquad (\theta_1 < \theta_2). \tag{3.32}$$

This problem arises when trying to demonstrate equivalence (or sometimes called bioequivalence) of treatments; for example, a new drug may be declared equivalent

to the current standard drug if the difference in therapeutic effect is small, meaning θ lies in a small interval about 0. Such testing problems also occur when one wishes to determine whether given specifications have been met concerning the proportion of an ingredient in a drug or some other compound, or whether a measuring instrument, for example a scale, is properly balanced. One then sets up the hypothesis that θ does not lie within the required limits, so that an error of the first kind consists in declaring θ to be satisfactory when in fact it is not. In practice, the decision to accept H will typically be accompanied by a statement of whether θ is believed to be $\leq \theta_1$ or $\geq \theta_2$. The implications of H are, however, frequently sufficiently important so that acceptance will in any case be followed by a more detailed investigation. If a manufacturer tests each precision instrument before releasing it and the test indicates an instrument to be out of balance, further work will be done to get it properly adjusted. If in a scientific investigation the inequalities $\theta \leq \theta_1$ and $\theta \geq \theta_2$ contradict some assumptions that have been formulated, a more complex theory may be needed and further experimentation will be required. In such situations there may be only two basic choices, to act as if $\theta_1 < \theta < \theta_2$ or to carry out some further investigation, and the formulation of the problem as that of testing the hypothesis H may be appropriate. In the present section, the existence of a UMP test of H will be proved for one-parameter exponential families.

Theorem 3.7.1 (i) *For testing the hypothesis $H : \theta \leq \theta_1$ or $\theta \geq \theta_2$ $(\theta_1 < \theta_2)$ against the alternatives $K : \theta_1 < \theta < \theta_2$ in the one-parameter exponential family (3.19) there exists a UMP test given by*

$$\phi(x) = \begin{cases} 1 & \text{when } C_1 < T(x) < C_2 \;\; (C_1 < C_2), \\ \gamma_i & \text{when } T(x) = C_i, \quad i = 1, 2, \\ 0 & \text{when } T(x) < C_1 \text{ or } > C_2, \end{cases} \tag{3.33}$$

where the C's and γ's are determined by

$$E_{\theta_1}\phi(X) = E_{\theta_2}\phi(X) = \alpha. \tag{3.34}$$

(ii) *This test minimizes $E_\theta\phi(X)$ subject to (3.34) for all $\theta < \theta_1$ and $> \theta_2$.*

(iii) *For $0 < \alpha < 1$ the power function of this test has a maximum at a point θ_0 between θ_1 and θ_2 and decreases strictly as θ tends away from θ_0 in either direction, unless there exist two values t_1, t_2 such that $P_\theta\{T(X) = t_1\} + P_\theta\{T(X) = t_2\} = 1$ for all θ.*

PROOF. (i): One can restrict attention to the sufficient statistic $T = T(X)$, the distribution of which by Lemma 2.7.2 is

$$dP_\theta(t) = C(\theta)e^{Q(\theta)t}d\nu(t),$$

where $Q(\theta)$ is assumed to be strictly increasing. Let $\theta_1 < \theta' < \theta_2$, and consider first the problem of maximizing $E_{\theta'}\psi(T)$ subject to (3.34) with $\phi(x) = \psi[T(x)]$. If M denotes the set of all points $E_{\theta_1}\psi(T), E_{\theta_2}\psi(T))$ as ψ ranges over the totality of critical

functions, then the point (α, α) is an inner point of M. This follows from the fact that by Corollary 3.2.1 the set M contains points (α, u_1) and (α, u_2) with $u_1 < \alpha < u_2$ and that it contains all points (u, u) with $0 < u < 1$. Hence by part (iv) of Theorem 3.6.1 there exist constants k_1, k_2 and test $\psi_0(t)$ and that $\phi_0(x) = \psi_0[T(x)]$ satisfies (3.34) and that $\psi_0(t) = 1$ when

$$k_1 C(\theta_1)e^{Q(\theta_1)t} + k_2 C(\theta_2)e^{Q(\theta_2)t} < C(\theta')e^{Q(\theta')t}$$

and therefore when

$$a_1 e^{b_1 t} + a_2 e^{b_2 t} < 1 \qquad (b_1 < 0 < b_2),$$

and $\psi_0(t) = 0$ when the left-hand side is > 1. Here the a's cannot both be ≤ 0, since then the test would always reject. If one of the a's is ≤ 0 and the other one is > 0, then the left-hand side is strictly monotone, and the test is of the one-sided type considered in Corollary 3.4.1, which has a strictly monotone power function and hence cannot satisfy (3.34). Therefore, since both a's are positive, the test satisfies (3.33). It follows from Lemma 3.7.1 below that the C's and γ's are uniquely determined by (3.33) and (3.34), and hence from Theorem 3.6.1(iii) that the test is UMP subject to the weaker restriction $E_{\theta_i} \psi(T) \leq \alpha \ (i = 1, 2)$. To complete the proof that this test is UMP for testing H, it is necessary to show that it satisfies $E_\theta \psi(T) \leq \alpha$ for $\theta \leq \theta_1$ and $\theta \geq \theta_2$. This follows from (ii) by comparison with the test $\psi(t) \equiv \alpha$.

(ii): Let $\theta' < \theta_1$, and apply Theorem 3.6.1(iv) to minimize $E_{\theta'} \phi(X)$ subject to (3.34). Dividing through by $e^{Q(\theta_1)t}$, the desired test is seen to have a rejection region of the form

$$a_1 e^{b_1 t} + a_2 e^{b_2 t} < 1 \qquad (b_1 < 0 < b_2).$$

Thus it coincides with the test $\psi_0(t)$ obtained in (i). By Theorem 3.6.1(iv) the first and third conditions of (3.33) are also necessary, and the optimum test is therefore unique provided $P\{T = C_i\} = 0$.

(iii): Without loss of generality let $Q(\theta) = \theta$. It follows from (i) and the continuity of $\beta(\theta) = E_\theta \phi(X)$ that either $\beta(\theta)$ satisfies (iii) or there exist three points $\theta' < \theta'' < \theta'''$ such that $\beta(\theta'') \leq \beta(\theta') = \beta(\theta''') = c$, say. Then $0 < c < 1$, since $\beta(\theta') = 0$ (or 1) implies $\phi(t) = 0$ (or 1) a.e. ν and this is excluded by (3.34). As is seen by the proof of (i), the test minimizes $E_{\theta''} \phi(X)$ subject to $E_{\theta'} \phi(X) = E_{\theta'''} \phi(X) = c$ for all $\theta' < \theta'' < \theta'''$. However, unless T takes on at most two values with probability 1 or all θ, $p_{\theta'}, p_{\theta''}, p_{\theta'''}$ are linearly independent, which by Corollary 3.6.1 implies $\beta(\theta'') > c$. ∎

In order to determine the C's and γ's, one will in practice start with some trial values C_1^*, γ_1^*, find C_2^*, γ_2^* such that $\beta^*(\theta_1) = \alpha$, and compute $\beta^*(\theta_2)$, which will usually be either too large or too small. For the selection of the next trial values it is then helpful to note that if $\beta^*(\theta_2) < \alpha$, the correct acceptance region is to the right of the one chosen, that is, it satisfies either $C_1 > C_1^*$ or $C_1 = C_1^*$ and $\gamma_1 < \gamma_1^*$, and that the converse holds if $\beta^*(\theta_2) > \alpha$. This is a consequence of the following lemma.

Lemma 3.7.1 *Let $p_\theta(x)$ satisfy the assumptions of Lemma 3.4.2(iv).*

(i) *If ϕ and ϕ^* are two tests satisfying (3.33) and $E_{\theta_1}\phi(T) = E_{\theta_1}\phi^*(T)$, and if ϕ^* is to the right of ϕ, then $\beta(\theta) < $ or $> \beta^*(\theta)$ as $\theta > \theta_1$ or $< \theta_1$.*

(ii) *If ϕ and ϕ^* satisfy (3.33) and (3.34), then $\phi = \phi^*$ with probability one.*

PROOF. (i): The result follows from Lemma 3.4.2(iv) with $\psi = \phi^* - \phi$. (ii): Since $E_{\theta_1}\phi(T) = E_{\theta_1}\phi^*(T)$, ϕ^* lies either to the left or the right of ϕ, and application of (i) completes the proof. ∎

Although a UMP test exists for testing that $\theta \leq \theta_1$ or $\geq \theta_2$ in an exponential family, the same is not true for the dual hypothesis $H : \theta_1 \leq \theta \leq \theta_2$ or for testing $\theta = \theta_0$ (Problem 3.59). There do, however, exist UMP unbiased tests of these hypotheses, as will be shown in Chapter 4.

3.8   Least Favorable Distributions

It is a consequence of Theorem 3.2.1 that there always exists a most powerful test for testing a simple hypothesis against a simple alternative. More generally, consider the case of a Euclidean sample space; probability densities $f_\theta, \theta \in \omega$, and g with respect to a measure μ; and the problem of testing $H : f_\theta, \theta \in \omega$, against the simple alternative $K : g$. The existence of a most powerful level α test then follows from the weak compactness theorem for critical functions (Theorem A.5.1 of the Appendix) as in Theorem 3.6.1(i).

Theorem 3.2.1 also provides an explicit construction for the most powerful test in the case of a simple hypothesis. We shall now extend this theorem to composite hypotheses in the direction of Theorem 3.6.1 by the method of undetermined multipliers. However, in the process of extension the result becomes much less explicit. Essentially it leaves open the determination of the multipliers, which now take the form of an arbitrary distribution. In specific problems this usually still involves considerable difficulty. From another point of view the method of attack, as throughout the theory of hypothesis testing, is to reduce the composite hypothesis to a simple one. This is achieved by considering weighted averages of the distributions of H. The composite hypothesis H is replaced by the simple hypothesis H_Λ that the probability density of X is given by

$$h_\Lambda(x) = \int_\omega f_\theta(x)\, d\Lambda(\theta),$$

where Λ is a probability distribution over ω. The problem of finding a suitable Λ is frequently made easier by the following consideration. Since H provides no information concerning θ and since H_Λ is to be equivalent to H for the purpose of testing against g, knowledge of the distribution Λ should provide as little help for this task as possible. To make this precise suppose that θ is known to have a distribution Λ. Then the maximum power β_Λ that can be attained against g is that of

the most powerful test ϕ_Λ for testing H_Λ against g. The distribution Λ is said to be *least favorable* (at level-α) if for all Λ' the inequality $\beta_\Lambda \leq \beta_{\Lambda'}$ holds.

Theorem 3.8.1 *Let a σ-field be defined over ω such that the densities $f_\theta(x)$ are jointly measurable in θ and x. Suppose that over this σ-field there exists a probability distribution Λ such that the most powerful level-α test ϕ_Λ for testing H_Λ against g is of size $\leq \alpha$ also with respect to the original hypothesis H.*

(i) *The test ϕ_Λ is most powerful for testing H against g.*

(ii) *If ϕ_Λ is the unique most powerful level α for testing H_Λ against g, it is also the unique most powerful test of H against g.*

(iii) *The distribution Λ is least favorable.*

PROOF. We note first that h_Λ is again a density with respect to μ, since by Fubini's theorem (Theorem 2.2.4)

$$\int h_\Lambda(x)\, d\mu(x) = \int_\omega d\Lambda(\theta) \int f_\theta(x)\, d\mu(x) = \int_\omega d\Lambda(\theta) = 1.$$

Suppose that ϕ_Λ is a level-α test for testing H, and let ϕ^* be any other level-α test. Then since $E_\theta \phi^*(X) \leq \alpha$ for all $\theta \in \omega$, we have

$$\int \phi^*(x) h_\Lambda(x)\, d\mu(x) = \int_\omega E_\theta \phi^*(X) d\Lambda(\theta) \leq \alpha.$$

Therefore $\phi*$ is a level-α test also for testing H_Λ and its power cannot exceed that of ϕ_Λ. This proves (i) and (ii). If Λ' is any distribution, it follows further that ϕ_Λ is a level-α test also for testing $H_{\Lambda'}$, and hence that its power against g cannot exceed that of the most powerful test, which by definition is $\beta_{\Lambda'}$. ∎

The conditions of this theorem can be given a somewhat different form by noting that ϕ_Λ can satisfy $\int_\omega E_\theta \phi_\Lambda(X)\, d\Lambda(\theta) = \alpha$ and $E_\theta \phi_\Lambda(X) \leq \alpha$ for all $\theta \in \omega$ only if the set of θ'''s with $E_\theta \phi_\Lambda(X) = \alpha$ has Λ-measure one.

Corollary 3.8.1 *Suppose that Λ is a probability distribution over ω and that ω' is a subset of ω with $\Lambda(\omega') = 1$. Let ϕ_Λ be a test such that*

$$\phi_\Lambda(x) = \begin{cases} 1 \text{ if } g(x) > k \int f_\theta(x)\, d\Lambda(\theta), \\ 0 \text{ if } g(x) < k \int f_\theta(x)\, d\Lambda(\theta). \end{cases} \tag{3.35}$$

Then ϕ_Λ is a most powerful level-α for testing H against g provided

$$E_{\theta'} \phi_\Lambda(X) = \sup_{\theta \in \omega} E_\theta \phi_\Lambda(X) = \alpha \quad \text{for} \quad \theta' \in \omega'. \tag{3.36}$$

Theorems 3.4.1 and 3.7.1 constitute two simple applications of Theorem 3.8.1. The set ω' over which the least favorable distribution Λ is concentrated consists of the single point θ_0 in the first of these examples and of the two points θ_1 and θ_2 in the second. This is what one might expect, since in both cases these are the distributions

of H that appear to be "closest" to K. Another example in which the least favorable distribution is concentrated at a single point is the following.

Example 3.8.1   (Sign test) The quality of items produced by a manufacturing process is measured by a characteristic X such as the tensile strength of a piece of material, or the length of life or brightness of a light bulb. For an item to be satisfactory X must exceed a given constant u, and one wishes to test the hypothesis $H : p \geq p_0$, where

$$p = P\{X \leq u\}$$

is the probability of an item being defective. Let $X_1, \ldots, X_n$ be the measurements of n sample items, so that the X's are independently distributed with common distribution about which no knowledge is assumed. Any distribution on the real line can be characterized by the probability p together with the conditional probability distributions P_- and P_+ of X given $X \leq u$ and $X > u$ respectively. If the distributions P_- and P_+ have probability densities p_- and $p+$, for example with respect to $\mu = P_- + P_+$, then the joint density of $X_1, \ldots, X_n$ at a sample point $x_1, \ldots, x_n$ satisfying

$$x_{i_1}, \ldots, x_{i_m} \leq u < x_{j_1}, \ldots, x_{j_{n-m}}$$

is

$$p^m (1 - p)^{n-m} p_-(x_{i_1}) \cdots p_-(x_{i_m}) p_+(x_{j_1}) \cdots p_+(x_{j_{n-m}}).$$

Consider now a fixed alternative to H, say (p_1, P_-, P_+), with $p_1 < p_0$. One would then expect the least favorable distribution Λ over H to assign probability 1 to the distribution (p_0, P_-, P_+) since this appears to be closest to the selected alternative. With this choice of Λ, the test (3.35) becomes

$$\phi_\Lambda(x) = 1 \text{ or } 0 \quad \text{as} \quad \left(\frac{p_1}{p_0}\right)^m \left(\frac{q_1}{q_0}\right)^{n-m} > \text{ or } < C,$$

and hence as $m <$ or $> C$. The test therefore rejects when the number M of defectives is sufficiently small, or more precisely, when $M < C$ and with probability γ when $M = C$, where

$$P\{M < C\} + \gamma P\{M = C\} = \alpha \quad \text{for} \quad p = p_0. \tag{3.37}$$

The distribution of M is the binomial distribution $b(p, n)$, and does not depend on P_+ and P_-. As a consequence, the power function of the test depends only on p and is a decreasing function of p, so that under H it takes on its maximum for $p = p_0$. This proves Λ to be least favorable and ϕ_Λ to be most powerful. Since the test is independent of the particular alternative chosen, it is UMP.

Expressed in terms of the variables $Z_i = X_i - u$, the test statistic M is the number of variables ≤ 0, and the test is the so-called *sign test* (cf. Section 4.9). It is an example of a *nonparametric* test, since it is derived without assuming a given functional form for the distribution of the X's such as the normal, uniform, or Poisson, in which only certain parameters are unknown.

The above argument applies, with only the obvious modifications, to the case that an item is satisfactory if X lies within certain limits: $u < X < v$. This occurs, for example, if X is the length of a metal part or the proportion of an ingredient in a chemical compound, for which certain tolerances have been specified. More generally the argument applies also to the situation in which X is vector-valued. Suppose that an item is satisfactory only when X lies in a certain set S, for example, if all the dimensions of a metal part or the proportions of several ingredients lie within specified limits. The probability of a defective is then

$$p = P\{X \in S^c\},$$

and P_- and P_+ denote the conditional distributions of X given $X \in S$ and $X \in S^c$ respectively. As before, there exists a UMP test of $H : p \geq p_0$, and it rejects H when the number M of defectives is sufficiently small, with the boundary of the test being determined by (3.37). ∎

A distribution Λ satisfying the conditions of Theorem 3.8.1 exists in most of the usual statistical problems, and in particular under the following assumptions. Let the sample space be Euclidean, let ω be a closed Borel set in s-dimensional Euclidean space, and suppose that $f_\theta(x)$ is a continuous function of θ for almost all x. Then given any g there exists a distribution Λ satisfying the conditions of Theorem 3.8.1 provided

$$\lim_{n \to \infty} \int_S f_{\theta_n}(x) \, d\mu(x) = 0$$

for every bounded set S in the sample space and for every sequence of vectors θ_n whose distance from the origin tends to infinity.

From this it follows as did Corollaries 1 and 4 from Theorems 3.2.1 and 3.6.1, that if the above conditions hold and if $0 < \alpha < 1$, there exists a test of power $\beta > \alpha$ for testing $H : f_\theta, \theta \in \omega$, against g unless $g = \int f_\theta \, d\Lambda(\theta)$ for some Λ. An example of the latter possibility is obtained by letting f_θ and g be the normal densities $N(\theta, \sigma_0^2)$ and $N(0, \sigma_1^2)$ respectively with $\sigma_0^2 < \sigma_1^2$. (See the following section.)

The above and related results concerning the existence and structure of least favorable distributions are given in Lehmann (1952b) (with the requirement that ω be closed mistakenly omitted), in Reinhardt (1961), and in Krafft and Witting (1967), where the relation to linear programming is explored.

3.9   Applications to Normal Distributions

3.9.1   Univariate Normal Models

Because of their wide applicability, the problems of testing the mean ξ and variance σ^2 of a normal distribution are of particular importance. Here and in similar problems later, the parameter not being tested is assumed to be unknown, but will not be shown explicitly in a statement of the hypothesis. We shall write, for example, $\sigma \leq \sigma_0$ instead of the more complete statement $\sigma \leq \sigma_0, -\infty < \xi < \infty$. The standard (likelihood ratio) tests of the two hypotheses $\sigma \leq \sigma_0$ and $\xi \leq \xi_0$ are given by the rejection regions

$$\sum(x_i - \bar{x})^2 \geq C \tag{3.38}$$

and

$$\frac{\sqrt{n}(\bar{x} - \xi_0)}{\sqrt{\frac{1}{n-1}\sum(x_i - \bar{x})^2}} \geq C. \tag{3.39}$$

The corresponding tests for the hypotheses $\sigma \geq \sigma_0$ and $\xi \geq \xi_o$ are obtained from the rejection regions (3.38) and (3.39) by reversing the inequalities. As will be shown in later chapters, these four tests are UMP both within the class of unbiased and within the class of invariant test (but see Section 13.2 for problems arising when the assumption of normality does not hold exactly). However, at the usual significance levels only the first of them is actually UMP.

Example 3.9.1   (One-sided tests of variance.) Let $X_1, \ldots, X_n$ be a sample from $N(\xi, \sigma^2)$, and consider first the hypotheses $H_1 : \sigma \geq \sigma_0$ and $H_2 : \sigma \leq \sigma_0$, and a simple alternative $K : \xi = \xi_1, \sigma = \sigma_1$. It seems reasonable to suppose that the least favorable distribution Λ in the (ξ, σ)-plane is concentrated on the line $\sigma = \sigma_0$. Since $Y = \sum X_i/n = \bar{X}$ and $U = \sum(X_i - \bar{X})^2$ are sufficient statistics for the parameters (ξ, σ), attention can be restricted to these variables. Their joint density under H_Λ is

$$C_o u^{(n-3)/2} \exp\left(-\frac{u}{2\sigma_0^2}\right) \int \exp\left[-\frac{n}{2\sigma_o^2}(y - \xi)^2\right] d\Lambda(\xi),$$

while under K it is

$$C_1 u^{(n-3)/2} \exp\left(-\frac{u}{2\sigma_1^2}\right) \exp\left[-\frac{n}{2\sigma_1^2}(y - \xi_1)^2\right].$$

The choice of Λ is seen to affect only the distribution of Y. A least favorable Λ should therefore have the property that the density of Y under H_Λ,

$$\int \frac{\sqrt{n}}{\sqrt{2\pi\sigma_0^2}} \exp\left[-\frac{n}{2\sigma_0^2}(y-\xi)^2\right] d\Lambda(\xi),$$

comes as close as possible to the alternative density,

$$\frac{\sqrt{n}}{\sqrt{2\pi\sigma_1^2}} \exp\left[-\frac{n}{2\sigma_1^2}(y-\xi_1)^2\right].$$

At this point one must distinguish between H_1 and H_2. In the first case $\sigma_1 < \sigma_0$. By suitable choice of Λ the mean of Y can be made equal to ξ_1, but the variance will if anything be increased over its initial value σ_0^2. This suggests that the least favorable distribution assigns probability 1 to the point $\xi = \xi_1$, since in this way the distribution of Y is normal both under H and K with the same mean in both cases and the smallest possible difference between the variances. The situation is somewhat different for H_2, for which $\sigma_0 < \sigma_1$. If the least favorable distribution Λ has a density, say Λ', the density of Y under H_Λ becomes

$$\int_{-\infty}^{\infty} \frac{\sqrt{n}}{\sqrt{2\pi\sigma_0}} \exp\left[-\frac{n}{2\sigma_0^2}(y-\xi)^2\right] \Lambda'(\xi)\, d\xi.$$

This is the probability density of the sum of two independent random variables, one distributed as $N(0, \sigma_0^2/n)$ and the other with density $\Lambda'(\xi)$. If Λ is taken to be $N(\xi_1, (\sigma_1^2 - \sigma_0^2)/n)$, the distribution of Y under H_Λ becomes $N(\xi_1, \sigma_1^2/n)$, the same as under K.

We now apply Corollary 3.8.1 with the distributions Λ suggested above. For H_1 it is more convenient to work with the original variables than with Y and U. Substitution in (3.35) gives $\phi(x) = 1$ when

$$\frac{(2\pi\sigma_1^2)^{-n/2} \exp\left[-\frac{1}{2\sigma_1^2}\sum(x_i - \xi_1)^2\right]}{(2\pi\sigma_0^2)^{-n/2} \exp\left[-\frac{1}{2\sigma_0^2}\sum(x_i - \xi_1)^2\right]} > C,$$

that is, when

$$\sum(x_i - \xi_1)^2 \le C. \tag{3.40}$$

To justify the choice of Λ, one must show that

$$P\left\{\sum(X_i - \xi_1)^2 \le C|\xi, \sigma\right\}$$

takes on its maximum over the half plane $\sigma \ge \sigma_0$ at the point $\xi = \xi_1$, $\sigma = \sigma_0$. For any fixed σ, the above is the probability of the sample point falling in a sphere radius, computed under the assumption that the X's are independently distributed as $N(\xi, \sigma^2)$. This probability is maximized when the center of the sphere coincides

with that of the distribution, that is, when $\xi = \xi_1$. (This follows for example from Problem 7.15.) The probability then becomes

$$P\left\{\sum\left(\frac{x_i - \xi_1}{\sigma}\right)^2 \leq \frac{C}{\sigma^2}\,\Big|\,\xi_1, \sigma\right\} = P\left\{\sum V_i^2 \leq \frac{C}{\sigma^2}\right\},$$

where $V_1, \ldots, V_n$ are independently distributed as $N(0, 1)$. This is a decreasing function of σ and therefore takes on its maximum when $\sigma = \sigma_0$.

In the case of H_2, application of Corollary 3.8.1 to the sufficient statistics (Y, U) gives $\phi(y, u) = 1$ when

$$\frac{C_1 u^{(n-3)/2} \exp\left(-\frac{u}{2\sigma_1^2}\right) \exp\left[-\frac{n}{2\sigma_1^2}(y - \xi_1)^2\right]}{C_0 u^{(n-3)/2} \exp\left(-\frac{u}{2\sigma_0^2}\right) \int \exp\left[-\frac{n}{2\sigma_0^2}(y - \xi)^2\right] \Lambda'(\xi)\,d\xi}$$

$$= C' \exp\left[-\frac{u}{2}\left(\frac{1}{\sigma_1^2} - \frac{1}{\sigma_0^2}\right)\right] \geq C,$$

that is, when

$$u = \sum(x_i - \bar{x})^2 \geq C. \tag{3.41}$$

Since the distribution of $\sum(X_i - \bar{X})^2/\sigma^2$ does not depend on ξ or σ, the probability $P\{\sum(X_i - \bar{X})^2 \geq C \mid \xi, \sigma\}$ is independent of ξ and increases with σ so that the conditions of Corollary 3.8.1 are satisfied. The test (3.41), being independent of ξ_1 and σ_1, is UMP for testing $\sigma \leq \sigma_0$ against $\sigma > \sigma_0$. It is also seen to coincide with the likelihood ratio test (3.38). On the other hand, the most powerful test (3.40) for testing $\sigma \geq \sigma_0$ against $\sigma < \sigma_0$ does depend on the value ξ_1 of ξ under the alternative.

It has been tacitly assumed so far that $n > 1$. If $n = 1$, the argument applies without change with respect to H_1, leading to (3.40) with $n = 1$. However, in the discussion of H_2 the statistic U now drops out, and Y coincides with the single observation X. Using the same Λ as before, one sees that X has the same distribution under H_Λ as under K, and the test ϕ_Λ therefore becomes $\phi_\Lambda(x) \equiv \alpha$. This satisfies the conditions of Corollary 3.8.1 and is therefore the most powerful test for the given problem. It follows that a single observation is of no value for testing the hypothesis H_2, as seems intuitively obvious, but that it could be used to test H_1 if the class of alternatives were sufficiently restricted. ∎

The corresponding derivation for the hypothesis $\xi \leq \xi_0$ is less straightforward. It turns out[11] that Student's test given by (3.39) is most powerful if the level of significance α is $\geq \frac{1}{2}$, regardless of the alternative $\xi_1 > \xi_0, \sigma_1$. This test is therefore UMP for $\alpha \geq \frac{1}{2}$. On the other hand, when $\alpha < \frac{1}{2}$ the most powerful test of H rejects when $\sum(x_i - a)^2 \leq b$, where the constants a and b depend on the alternative (ξ_1, σ_1) and on α. Thus for the significance levels that are of interest, a UMP test of H does

---

[11] See Lehmann and Stein (1948).

not exist. No new problem arises for the hypothesis $\xi \geq \xi_0$, since this reduces to the case just considered through the transformation $Y_i = \xi_0 - (X_i - \xi_0)$.

3.9.2 Multivariate Normal Models

Let X denote a $k \times 1$ random vector whose ith component, X_i, is a real-valued random variable. The mean of X, denoted $E(X)$, is a vector with ith component $E(X_i)$ (assuming it exists). The covariance matrix of X, denoted Σ, is the $k \times k$ matrix with (i, j) entry $Cov(X_i, X_j)$. Σ is well-defined iff $E(|X|^2) < \infty$, where $|\cdot|$ denotes the Euclidean norm. Note that if A is an $m \times k$ matrix, then the $m \times 1$ vector $Y = AX$ has mean (vector) $AE(X)$ and covariance matrix $A \Sigma A^\top$, where $A^\top$ is the transpose of A (Problem 3.68).

The multivariate generalization of a real-valued normally distributed random variable is a random vector $X = (X_1, \ldots, X_k)^\top$ with the *multivariate normal probability density*

$$\frac{\sqrt{|A|}}{(2\pi)^{\frac{1}{2}k}} \exp\left[-\frac{1}{2} \sum \sum a_{ij}(x_i - \xi_i)(x_j - \xi_j)\right], \qquad (3.42)$$

where the matrix $A = (a_{ij})$ is positive definite, and $|A|$ denotes its determinant. The means and covariance matrix of the X's are given by

$$E(X_i) = \xi_i, \qquad E(X_i - \xi_i)(X_j - \xi_j) = \sigma_{ij}, \quad (\sigma_{ij}) = A^{-1}. \qquad (3.43)$$

The column vector $\xi = (\xi_1, \ldots, \xi_k)^\top$ is the mean vector and $\Sigma = A^{-1}$ is the covariance matrix of X.

Such a definition only applies when A is nonsingular, in which case we say that X has a nonsingular multivariate normal distribution. More generally, we say that Y has a multivariate normal distribution if $Y = BX + \mu$ for some $m \times k$ matrix of constants B and $m \times 1$ constant vector μ, where X has some nonsingular multivariate normal distribution. Then, Y is multivariate normal if and only if $\sum_{i=1}^{m} c_i Y_i$ is univariate normal for all c, where $N(\xi, \sigma^2)$ with $\sigma = 0$ is interpreted to be the distribution that is point mass at ξ. Basic properties of the multivariate normal distribution are given in Anderson (2003).

Example 3.9.2 (One-sided tests of a combination of means.) Assume X is multivariate normal with unknown mean $\xi = (\xi_1, \ldots, \xi_k)^\top$ and known covariance matrix Σ. Assume $a = (a_1, \ldots, a_k)^\top$ is a fixed vector with $a^\top \Sigma a > 0$. The problem is to test

$$H : \sum_{i=1}^{k} a_i \xi_i \leq \delta \quad \text{vs.} \quad K : \sum_{i=1}^{k} a_k \xi_i > \delta \, .$$

We will show that a UMP level α test exists, which rejects when $\sum_i a_i X_i > \sigma z_{1-\alpha}$, where $\sigma^2 = a^\top \Sigma a$. To see why,[12] we will consider four cases of increasing generality.

Case 1. If $k = 1$ and the problem is to test the mean of X_1, the result follows by Problem 3.1.

Case 2. Consider now general k, so that $(X_1, \ldots, X_k)$ has mean $(\xi_1, \ldots, \xi_k)$ and covariance matrix Σ. However, consider the special case $(a_1, \ldots, a_k) = (1, 0, \ldots, 0)$. Also, assume X_1 and $(X_2, \ldots, X_k)$ are independent. Then, for any fixed alternative $(\xi_1', \ldots, \xi_k')$ with $\xi_1' > \delta$, the least favorable distribution concentrates on the single point $(\delta, \xi_2', \ldots, \xi_k')$ (Problem 3.70).

Case 3. As in case 2, consider $a_1 = 1$ and $a_i = 0$ if $i > 1$, but now allow Σ to be an arbitrary covariance matrix. We can reduce the problem to case 2 by an appropriate linear transformation. Simply let $Y_1 = X_1$ and, for $i > 1$, let

$$Y_i = X_i - \frac{Cov(X_1, X_i)}{Var(X_1)} X_1 .$$

Then, it is easily checked that $Cov(Y_1, Y_i) = 0$ if $i > 1$. Moreover, Y is just a 1:1 transformation of X. But, the problem of testing $E(Y_1) = E(X_1)$ based on $Y = (Y_1, \ldots, Y_k)$ is in the form already studied in case 2, and the UMP test rejects for large values of $Y_1 = X_1$.

Case 4. Now, consider arbitrary $(a_1, \ldots, a_k)$ satisfying $a^\top \Sigma a > 0$. Let $Z = OX$, where O is any orthogonal matrix with first row $(a_1, \ldots, a_k)$. Then, $E(Z_1) = \sum_{i=1}^k a_i \xi_i$, and the problem of testing $E(Z_1) \leq \delta$ versus $E(Z_1) > \delta$ reduces to case 3. Hence, the UMP test rejects for large values of $Z_1 = \sum_{i=1}^k a_i X_i$. ∎

Example 3.9.3 (Equivalence tests of a combination of means.) As in Example 3.9.2, assume X is multivariate normal $N(\xi, \Sigma)$ with unknown mean vector ξ and known covariance matrix Σ. Fix $\delta > 0$ and any vector $a = (a_1, \ldots, a_k)^\top$ satisfying $a^\top \Sigma a > 0$. Consider testing

$$H : |\sum_{i=1}^k a_i \xi_i| \geq \delta \quad vs \quad K : |\sum_{i=1}^k a_i \xi_i| < \delta .$$

Then, a UMP level α test also exists and it rejects H if

$$|\sum_{i=1}^k a_i X_i| < C ,$$

where $C = C(\alpha, \delta, \sigma)$ satisfies

---

[12] Proposition 15.2 of van der Vaart (1998) provides an alternative proof in the case Σ is invertible.

$$\Phi\left(\frac{C-\delta}{\sigma}\right) - \Phi\left(\frac{-C-\delta}{\sigma}\right) = \alpha \tag{3.44}$$

and $\sigma^2 = a^\top \Sigma a$. Hence, the power of this test against an alternative $(\xi_1, \dots, \xi_k)$ with $|\sum_i a_i \xi_i| = \delta' < \delta$ is

$$\Phi\left(\frac{C-\delta'}{\sigma}\right) - \Phi\left(\frac{-C-\delta'}{\sigma}\right) .$$

To see why, we again consider four cases of increasing generality. *Case 1.* Suppose $k = 1$, so that $X_1 = X$ is $N(\xi, \sigma^2)$ and we are testing $|\xi| \geq \delta$ versus $|\xi| < \delta$. (This case follows by Theorem 3.7.1, but we argue independently so that the argument applies to the other cases as well.) Fix an alternative $\xi = m$ with $|m| < \delta$. Reduce the composite null hypothesis to a simple one via a least favorable distribution that places mass p on $N(\delta, \sigma^2)$ and mass $1 - p$ on $N(-\delta, \sigma^2)$. The value of p will be chosen shortly so that such a distribution is least favorable (and will be seen to depend on m, α, σ and δ). By the Neyman–Pearson Lemma, the MP test of

$$pN(\delta, \sigma^2) + (1 - p)N(-\delta, \sigma^2) \quad vs \quad N(m, \sigma^2)$$

rejects for small values of

$$\frac{p \exp\left[-\frac{1}{2\sigma^2}(X - \delta)^2\right] + (1 - p)\exp\left[-\frac{1}{2\sigma^2}(X + \delta)^2\right]}{\exp\left[-\frac{1}{2\sigma^2}(X - m)^2\right]} , \tag{3.45}$$

or equivalently for small values of $f(X)$, where

$$f(x) = p \exp[(\delta - m)X/\sigma^2] + (1 - p)\exp[-(\delta + m)X/\sigma^2] .$$

We can now choose p so that $f(C) = f(-C)$, so that p must satisfy

$$\frac{p}{1 - p} = \frac{\exp[(\delta + m)C/\sigma^2] - \exp[-(\delta + m)C/\sigma^2]}{\exp[(\delta - m)C/\sigma^2] - \exp[-(\delta - m)C/\sigma^2]}. \tag{3.46}$$

Since $\delta - m > 0$ and $\delta + m > 0$, both the numerator and denominator of the right side of (3.46) are positive, so the right side is a positive number. On the other hand, $p/(1 - p)$ is a nondecreasing function of p with range $[0, \infty)$ as p varies from 0 to 1. Thus, p is well-defined. Also, observe $f''(x) \geq 0$ for all x. It follows that (for this special choice of C)

$$\{X : f(X) \leq f(C)\} = \{X : |X| \leq C\}$$

is the rejection region of the MP test. Such a test is easily seen to be level α for the original composite null hypothesis because its power function is symmetric and decreases away from zero. Thus, the result follows by Theorem 3.8.1.

Case 2. Consider now general k, so that $(X_1, \ldots, X_k)$ has mean $(\xi_1, \ldots, \xi_k)$ and covariance matrix Σ. However, consider the special case $(a_1, \ldots, a_k) = (1, 0, \ldots, 0)$, so we are testing $|\xi_1| \geq \delta$ versus $|\xi_1| < \delta$. Also, assume X_1 and $(X_2, \ldots, X_k)$ are independent, so that the first row and first column of Σ are zero except the first entry, which is σ^2 (assumed positive). Using the same reasoning as case 1, fix an alternative $m = (m_1, \ldots, m_k)$ with $|m_1| < \delta$ and consider testing

$$pN\left((\delta, m_2, \ldots, m_k), \Sigma\right) + (1 - p)N\left((-\delta, m_2, \ldots, m_k), \Sigma\right)$$

versus $N\left((m_1, \ldots, m_k), \Sigma\right)$. The likelihood ratio is in fact the same as (3.45) because each term is now multiplied by the density of $(X_2, \ldots, X_k)$ (by independence), and these densities cancel. The UMP test from case 1, which rejects when $|X_1| \leq C$, is UMP in this situation as well.

Case 3. As in case 2, consider $a_1 = 1$ and $a_i = 0$ if $i > 1$, but now allow Σ to be an arbitrary covariance matrix. By transforming X to Y as in Case 3 of Example 3.9.2, the result follows (Problem 3.71).

Case 4. Now, consider arbitrary $(a_1, \ldots, a_k)$ satisfying $a^\top \Sigma a > 0$. As in case 4 of Example 3.9.2, transform X to Z and the result follows (Problem 3.71).

3.10 Problems

Section 3.2

Problem 3.1 Let $X_1, \ldots, X_n$ be a sample from the normal distribution $N(\xi, \sigma^2)$.

(i) If $\sigma = \sigma_0$ (known), there exists a UMP test for testing $H : \xi \leq \xi_0$ against $\xi > \xi_0$, which rejects when $\sum(X_i - \xi_0)$ is too large.
(ii) If $\xi = \xi_0$ (known), there exists a UMP test for testing $H : \sigma \leq \sigma_0$ against $K : \sigma > \sigma_0$, which rejects when $\sum(X_i - \xi_0)^2$ is too large.

Problem 3.2 *UMP test for $U(0, \theta)$.* Let $X = (X_1, \ldots, X_n)$ be a sample from the uniform distribution on $(0, \theta)$.

(i) For testing $H : \theta \leq \theta_0$ against $K : \theta > \theta_0$ any test is UMP at level α for which $E_{\theta_0}\phi(X) = \alpha$, $E_\theta\phi(X) \leq \alpha$ for $\theta \leq \theta_0$, and $\phi(x) = 1$ when $\max(x_1, \ldots, x_n) > \theta_0$.
(ii) For testing $H : \theta = \theta_0$ against $K : \theta \neq \theta_0$ a unique UMP test exists, and is given by $\phi(x) = 1$ when $\max(x_1, \ldots, x_n) > \theta_0$ or $\max(x_1, \ldots, x_n) \leq \theta_0 \sqrt[n]{\alpha}$, and $\phi(x) = 0$ otherwise.

[(i): For each $\theta > \theta_0$ determine the ordering established by $r(x) = p_\theta(x)/p_{\theta_0}(x)$ and use the fact that many points are equivalent under this ordering.

(ii): Determine the UMP tests for testing $\theta = \theta_0$ against $\theta < \theta_0$ and combine this result with that of part (i).]

Problem 3.3 Suppose N i.i.d. random variables are generated from the same known strictly increasing absolutely continuous cdf $F(\cdot)$. We are told only X, the maximum of these random variables. Is there a UMP size α test of

$$H_0 : N \le 5 \text{ versus } H_1 : N > 5?$$

If so, find it.

Problem 3.4 *UMP test for exponential densities.* Let $X_1, \ldots, X_n$ be a sample from the exponential distribution $E(a, b)$ of Problem 1.18, and let $X_{(1)} = \min(X_1, \ldots, X_n)$.

(i) Determine the UMP test for testing $H : a = a_0$ against $K : a \ne a_0$ when b is assumed known.

(ii) The power of any MP level-α test of $H : a = a_0$ against $K : a = a_1 < a_0$ is given by
$$\beta^*(a_1) = 1 - (1 - \alpha)e^{-n(a_0 - a_1)/b}.$$

(iii) For the problem of part (i), when b is unknown, the power of any level α test which rejects when
$$\frac{X_{(1)} - a_0}{\sum[X_i - X_{(1)}]} \le C_1 \text{ or } \ge C_2$$

against any alternative (a_1, b) with $a_1 < a_0$ is equal to $\beta^*(a_1)$ of part (ii) (independent of the particular choice of C_1 and C_2).

(iv) The test of part (iii) is a UMP level-α test of $H : a = a_0$ against $K : a \ne a_0$ (b unknown).

(v) Determine the UMP test for testing $H : a = a_0, b = b_0$ against the alternatives $a < a_0, b < b_0$.

(vi) Explain the (very unusual) existence in this case of a UMP test in the presence of a nuisance parameter [part(iv)] and for a hypothesis specifying two parameters [part(v)].

[(i) The variables $Y_i = e^{-X_i/b}$ are a sample from the uniform distribution on $(0, e^{-a/b})$.]

Note. For more general versions of parts (ii)–(iv), see Takeuchi (1969) and Kabe and Laurent (1981).

Problem 3.5 In the proof of Theorem 3.2.1(i), consider the set of c satisfying $\alpha(c) \le \alpha \le \alpha(c - 0)$. If there is only one such c, c is unique; otherwise, there is an interval of such values $[c_1, c_2]$. Argue that, in this case, if $\alpha(c)$ is continuous at c_2, then $P_i(C) = 0$ for $i = 0, 1$, where

$$C = \left\{ x : p_0(x) > 0 \text{ and } c_1 < \frac{p_1(x)}{p_0(x)} \leq c_2 \right\}.$$

If $\alpha(c)$ is not continuous at c_2, then the result is false.

Problem 3.6 Let P_0, P_1, P_2 be the probability distributions assigning to the integers $1, \ldots, 6$ the following probabilities:

	1	2	3	4	5	6
P_0	0.03	0.02	0.02	0.01	0	0.92
P_1	0.06	0.05	0.08	0.02	0.01	0.78
P_2	0.09	0.05	0.12	0	0.02	0.72

Determine whether there exists a level-α test of $H : P = P_0$ which is UMP against the alternatives P_1 and P_2 when (i) $\alpha = 0.01$; (ii) $\alpha = 0.05$; (iii) $\alpha = 0.07$.

Problem 3.7 Let the distribution of X be given by

x	0	1	2	3
$P_\theta(X = x)$	θ	2θ	$0.9 - 2\theta$	$0.1 - \theta$

where $0 < \theta < 0.1$. For testing $H : \theta = 0.05$ against $\theta > 0.05$ at level $\alpha = 0.05$, determine which of the following tests (if any) is UMP:

  (i) $\phi(0) = 1, \phi(1) = \phi(2) = \phi(3) = 0$;
 (ii) $\phi(1) = 0.5, \phi(0) = \phi(2) = \phi(3) = 0$;
(iii) $\phi(3) = 1, \phi(0) = \phi(1) = \phi(2) = 0$.

Problem 3.8 A random variable X has the *Pareto distribution* $P(c, \tau)$ if its density is $c\tau^c/x^{c+1}, 0 < \tau < x, 0 < C$.

  (i) Show that this defines a probability density.
 (ii) If X has distribution $P(c, \tau)$, then $Y = \log X$ has exponential distribution $E(\xi, b)$ with $\xi = \log \tau, b = 1/c$.
(iii) If $X_1, \ldots, X_n$ is a sample from $P(c, \tau)$, use (ii) and Problem 3.4 to obtain UMP tests of (a) $H : \tau = \tau_0$ against $\tau \neq \tau_0$ when b is known; (b) $H : c = c_0, \tau = \tau$ against $c > c_0, \tau < \tau_0$.

Problem 3.9 Let X be distributed according to $P_\theta, \theta \in \Omega$, and let T be sufficient for θ. If $\varphi(X)$ is any test of a hypothesis concerning θ, then $\psi(T)$ given by $\psi(t) = E[\varphi(X) \mid t]$ is a test depending on T only, and its power function is identical with that of $\varphi(X)$.

Problem 3.10 In the notation of Section 3.2, consider the problem of testing $H_0 : P = P_0$ against $H_1 : P = P_1$, and suppose that known probabilities $\pi_0 = \pi$ and $\pi_1 = 1 - \pi$ can be assigned to H_0 and H_1 prior to the experiment.

(i)   The overall probability of an error resulting from the use of a test φ is

$$\pi E_0 \varphi(X) + (1 - \pi) E_1 [1 - \varphi(X)].$$

(ii) The *Bayes test* minimizing this probability is given by (3.8) with $k = \pi_0 / \pi_1$.

(iii) The conditional probability of H_i given $X = x$, the *posterior probability* of H_i is

$$\frac{\pi_i p_i(x)}{\pi_0 p_0(x) + \pi_1 p_1(x)},$$

and the Bayes test therefore decides in favor of the hypothesis with the larger posterior probability.

Problem 3.11   (i) For testing $H_0 : \theta = 0$ against $H_1 : \theta = \theta_1$ when X is $N(\theta, 1)$, given any $0 < \alpha < 1$ and any $0 < \pi < 1$ (in the notation of the preceding problem), there exists θ_1 and x such that (a) H_0 is rejected when $X = x$ but (b) $P(H_0 \mid x)$ is arbitrarily close to 1.

(ii) The paradox of part (i) is due to the fact that α is held constant while the power against θ_1 is permitted to get arbitrarily close to 1. The paradox disappears if α is determined so that the probabilities of type I and type II error are equal [but see Berger and Sellke (1987)].

[For a discussion of such paradoxes, see Lindley (1957), Bartlett (1957), Schafer (1982, 1988) and Robert (1993).]

Problem 3.12   Let $X_1, \ldots, X_n$ be independently distributed, each uniformly over the integers $1, 2, \ldots, \theta$. Determine whether there exists a UMP test for testing $H : \theta = \theta_0$, at level $1/\theta_0^n$ against the alternatives (i) $\theta > \theta_0$; (ii) $\theta < \theta_0$; (iii) $\theta \neq \theta_0$.

Problem 3.13   The following example shows that the power of a test can sometimes be increased by selecting a random rather than a fixed sample size even when the randomization does not depend on the observations. Let $X_1, \ldots, X_n$ be independently distributed as $N(\theta, 1)$, and consider the problem of testing $H : \theta = 0$ against $K : \theta = \theta_1 > 0$.

(i)   The power of the most powerful test as a function of the sample size n is not necessarily concave.

(ii) In particular for $\alpha = 0.005$, $\theta_1 = \frac{1}{2}$, better power is obtained by taking 2 or 16 observations with probability $\frac{1}{2}$ each than by taking a fixed sample of 9 observations.

(iii) The power can be increased further if the test is permitted to have different significance levels α_1 and α_2 for the two sample sizes and it is required only that the expected significance level be equal to $\alpha = 0.005$. Examples are: (a) with probability $\frac{1}{2}$ take $n_1 = 2$ observations and perform the test of significance at level $\alpha_1 = 0.001$, or take $n_2 = 16$ observations and perform the test at level $\alpha_2 = 0.009$; (b) with probability $\frac{1}{2}$ take $n_1 = 0$ or $n_2 = 18$ observations and let the respective significance levels be $\alpha_1 = 0$, $\alpha_2 = 0.01$.

Note. This and related examples were discussed by Kruskal in a seminar held at Columbia University in 1954. A more detailed investigation of the phenomenon has been undertaken by Cohen (1958).

Problem 3.14 If the sample space $\mathcal{X}$ is Euclidean and P_0, P_1 have densities with respect to Lebesgue measure, there exists a nonrandomized most powerful test for testing P_0 against P_1 at every significance level α.[13] [This is a consequence of Theorem 3.2.1 and the following lemma.[14] Let $f \geq 0$ and $\int_A f(x)\,dx = a$. Given any $0 \leq b \leq a$, there exists a subset B of A such that $\int_B f(x)\,dx = b$.]

Problem 3.15 *Fully informative statistics*. A statistic T is *fully informative* if for every decision problem the decision procedures based only on T form an essentially complete class. If $\mathcal{P}$ is dominated and T is fully informative, then T is sufficient. [Consider any pair of distributions P_0, $P_1 \in \mathcal{P}$ with densities p_0, p_1, and let $g_i = p_i/(p_0 + p_1)$. Suppose that T is fully informative, and let $\mathcal{A}_t$ be the subfield induced by T. Then $\mathcal{A}_t$ contains the subfield induced by (g_0, g_1) since it contains every rejection which is unique most powerful for testing P_0 against P_1 (or P_1 against P_0) at some level α. Therefore, T is sufficient for every pair of distributions (P_0, P_1), and hence by Problem 2.11 it is sufficient for $\mathcal{P}$.]

Problem 3.16 Based on X with distribution indexed by $\theta \in \Omega$, the problem is to test $\theta \in \omega$ versus $\theta \in \omega'$. Suppose there exists a test ϕ such that $E_\theta[\phi(X)] \leq \beta$ for all θ in ω, where $\beta < \alpha$. Show there exists a level α test $\phi^*(X)$ such that

$$E_\theta[\phi(X)] \leq E_\theta[\phi^*(X)] \,,$$

for all θ in ω' and this inequality is strict if $E_\theta[\phi(X)] < 1$.

Problem 3.17 *A counterexample*. Typically, as α varies the most powerful level α tests for testing a hypothesis H against a simple alternative are nested in the sense that the associated rejection regions, say R_α, satisfy $R_\alpha \subseteq R_{\alpha'}$, for any $\alpha < \alpha'$. Even if the most powerful tests are nonrandomized, this may be false. Suppose X takes values 1, 2, and 3 with probabilities 0.85, 0.1, and 0.05 under H and probabilities 0.7, 0.2, and 0.1 under K respectively.
(i) At any level < 0.15, the MP test is not unique.
(ii) At $\alpha = 0.05$ and $\alpha' = 0.1$, there exist unique nonrandomized MP tests and they are not nested.
(iii) At these levels there exist MP tests ϕ and ϕ' that are nested in the sense that $\phi(x) \leq \phi'(x)$ for all x. [This example appears as Example 10.16 in Romano and Siegel (1986).]

Problem 3.18 Under the setup of Theorem 3.2.1, show that there always exist MP tests that are nested in the sense of Problem 3.17(iii).

---

[13] For more general results concerning the possibility of dispensing with randomized procedures, see Dvoretzky et al. (1951).

[14] For a proof of this lemma see Halmos (1974, p. 174). The lemma is a special case of a theorem of Lyapounov (1940); see Blackwell (1951a).

Problem 3.19 Suppose $X_1, \ldots, X_n$ are i.i.d. $N(\xi, \sigma^2)$ with σ known. For testing $\xi = 0$ versus $\xi \neq 0$, the average power of a test $\phi = \phi(X_1, \ldots, X_n)$ is given by

$$\int_{-\infty}^{\infty} E_\xi(\phi) d\Lambda(\xi) ,$$

where Λ is a probability distribution on the real line. Suppose that Λ is symmetric about 0; that is, $\Lambda\{E\} = \Lambda\{-E\}$ for all Borel sets E. Show that, among α level tests, the one maximizing average power rejects for large values of $|\sum_i X_i|$. Show that this test need not maximize average power if Λ is not symmetric.

Problem 3.20 Let f_θ, $\theta \in \Omega$, denote a family of densities with respect to a measure μ. (We assume Ω is endowed with a σ-field so that the densities $f_\theta(x)$ are jointly measurable in θ and x.) Consider the problem of testing a simple null hypothesis $\theta = \theta_0$ against the composite alternatives $\Omega_K = \{\theta : \theta \neq \theta_0\}$. Let Λ be a probability distribution on Ω_K.
(i) As explicitly as possible, find a test ϕ that maximizes $\int_{\Omega_K} E_\theta(\phi) d\Lambda(\theta)$, subject to it being level α.
(ii) Let $h(x) = \int f_\theta(x) d\Lambda(\theta)$. Consider the nonrandomized ϕ test that rejects if and only if $h(x)/f_{\theta_0}(x) > k$, and suppose $\mu\{x : h(x) = k f_\theta(x)\} = 0$. Then, ϕ is admissible at level $\alpha = E_{\theta_0}(\phi)$ in the sense that it is impossible that there exists another level α test ϕ' such that $E_\theta(\phi') \geq E_\theta(\phi)$ for all θ.
(iii) Show that the test of Problem 3.19 is admissible.

Section 3.3

Problem 3.21 In Example 3.21, show that p-value is indeed given by $\hat{p} = \hat{p}(X) = (11 - X)/10$. Also, graph the c.d.f. of $\hat{p}$ under H and show that the last inequality in (3.15) is an equality if and only if u is of the form $0, \ldots, 10$.

Problem 3.22 Suppose X has a continuous distribution function F. Show that $F(X)$ is uniformly distributed on $(0, 1)$. [The transformation from X to $F(X)$ is known as the *probability integral transformation*.]

Problem 3.23 (i) Show that if Y is any random variable with c.d.f. $G(\cdot)$, then

$$P\{G(Y) \leq u\} \leq u \quad \text{for all } 0 \leq u \leq 1 .$$

If $G^-(t) = P\{Y < t\}$, then show

$$P\{1 - G^-(Y) \leq u\} \leq u \quad \text{for all } 0 \leq u \leq 1 .$$

(ii) In Example 3.3.3, show that $F_{\theta_0}(T)$ and $1 - F_{\theta_0}^-(T)$ are both valid p-values, in the sense that (3.13) holds.

Problem 3.24 Under the setup of Lemma 3.3.1, suppose the rejection regions are defined by

$$R_\alpha = \{X : T(X) \geq k(\alpha)\} \tag{3.47}$$

for some real-valued statistic $T(X)$ and $k(\alpha)$ satisfying

$$\sup_{\theta \in \Omega_H} P_\theta\{T(X) \geq k(\alpha)\} = \alpha .$$

Then, show

$$\hat{p} = \sup_{\theta \in \Omega_H} P\{T(X) \geq t\} ,$$

where t is the observed value of $T(X)$.

Problem 3.25 Under the setup of Lemma 3.3.1, show that there exists a real-valued statistic $T(X)$ so that the rejection region is necessarily of the form (3.47). [*Hint*: Let $T(X) = -\hat{p}$.]

Problem 3.26 (i) If $\hat{p}$ is uniform on $(0, 1)$, show that $-2\log(\hat{p})$ has the Chi-squared distribution with 2 degrees of freedom.
(ii) Suppose $\hat{p}_1, \ldots, \hat{p}_s$ are i.i.d. uniform on $(0, 1)$. Let $F = -2\log(\hat{p}_1 \cdots \hat{p}_s)$. Argue that F has the Chi-squared distribution with $2s$ degrees of freedom. What can you say about F if the $\hat{p}_i$ are independent and satisfy $P\{\hat{p}_i \leq u\} \leq u$ for all $0 \leq u \leq 1$? [Fisher (1934a) proposed F as a means of combining p-values from independent experiments.]

Section 3.4

Problem 3.27 Let X be the number of successes in n independent trials with probability p of success, and let $\phi(x)$ be the UMP test (3.16) for testing $p \leq p_0$ against $p > p_0$ at the level of significance α.

 (i) For $n = 6$, $p_0 = 0.25$ and the levels $\alpha = 0.05, 0.1, 0.2$ determine C and γ, and the power of the test against $p_1 = 0.3, 0.4, 0.5, 0.6, 0.7$.
 (ii) If $p_0 = 0.2$ and $\alpha = 0.05$, and it is desired to have power $\beta \geq 0.9$ against $p_1 = 0.4$, determine the necessary sample size (a) by using tables of the binomial distribution, (b) by using the normal approximation.[15]
(iii) Use the normal approximation to determine the sample size required when $\alpha = 0.05$, $\beta = 0.9$, $p_0 = 0.01$, $p_1 = 0.02$.

Problem 3.28 (i) A necessary and sufficient condition for densities $p_\theta(x)$ to have monotone likelihood ratio in x, if the mixed second derivative $\partial^2 \log p_\theta(x)/\partial\theta\,\partial x$ exists, is that this derivative is ≥ 0 for all θ and x.

---

[15] Tables and approximations are discussed, for example, in Chapter 3 of Johnson and Kotz (1969).

(ii) An equivalent condition is that

$$p_\theta(x) \frac{\partial^2 p_\theta(x)}{\partial \theta \, \partial x} \geq \frac{\partial p_\theta(x)}{\partial \theta} \, \frac{\partial p_\theta(x)}{\partial x} \qquad \text{for all } \theta \text{ and } x.$$

Problem 3.29 Let the probability density p_θ of X have monotone likelihood ratio in $T(x)$, and consider the problem of testing $H : \theta \leq \theta_0$ against $\theta > \theta_0$. If the distribution of T is continuous, the p-value $\hat{p}$ of the UMP test is given by $\hat{p} = P_{\theta_0}\{T \geq t\}$, where t is the observed value of T. This holds also without the assumption of continuity if for randomized tests $\hat{p}$ is defined as the smallest significance level at which the hypothesis is rejected with probability 1. Show that, for any $\theta \leq \theta_0$, $P_\theta\{\hat{p} \leq u\} \leq u$ for any $0 \leq u \leq 1$.

Problem 3.30 Let $X_1, \ldots, X_n$ be independently distributed with density $(2\theta)^{-1} e^{-x/2\theta}$, $x \geq 0$, and let $Y_1 \leq \cdots \leq Y_n$ be the ordered X's. Assume that Y_1 becomes available first, then Y_2, and so on, and that observation is continued until Y_r has been observed. On the basis of $Y_1, \ldots, Y_r$ it is desired to test $H : \theta \geq \theta_0 = 1000$ at level $\alpha = 0.05$ against $\theta < \theta_0$.

 (i) Determine the rejection region when $r = 4$, and find the power of the test against $\theta_1 = 500$.
(ii) Find the value of r required to get power $\beta \geq 0.95$ against the alternative.

   [In Problem 2.15 the distribution of $[\sum_{i=1}^{r} Y_i + (n-r)Y_r]/\theta$ was found to be χ^2 with $2r$ degrees of freedom.]

Problem 3.31 When a Poisson process with rate λ is observed for a time interval of length τ, the number X of events occurring has the Poisson distribution $P(\lambda\tau)$. Under an alternative scheme, the process is observed until r events have occurred, and the time T of observation is then a random variable such that $2\lambda T$ has a χ^2-distribution with $2r$ degrees of freedom. For testing $H : \lambda \leq \lambda_0$ at level α one can, under either design, obtain a specified power β against an alternative λ_1 by choosing τ and r sufficiently large.

 (i) The ratio of the time of observation required for this purpose under the first design to the expected time required under the second is $\lambda\tau/r$.
(ii) Determine for which values of λ each of the two designs is preferable when $\lambda_0 = 1, \lambda_1 = 2, \alpha = 0.05, \beta = 9$.

Problem 3.32 Let $X = (X_1, \ldots, X_n)$ be a sample from the uniform distribution $U(\theta, \theta + 1)$.

 (i) For testing $H : \theta \leq \theta_0$ against $K : \theta > \theta_0$ at level α, there exists a UMP test which rejects when $\min(X_1, \ldots, X_n) > \theta_0 + C(\alpha)$ or $\max(X_1, \ldots, X_n) > \theta_0 + 1$ for suitable $C(\alpha)$.
(ii) The family $U(\theta, \theta + 1)$ does not have monotone likelihood ratio. [Additional results for this family are given in Birnbaum (1954b) and Pratt (1958).]

[(ii) By Theorem 3.4.1, monotone likelihood ratio implies that the family of UMP test of $H : \theta \leq \theta_0$ against $K : \theta > \theta_0$ generated as α varies from 0 to 1 is independent of θ_0.]

Problem 3.33 Let X be a single observation from the Cauchy density given at the end of Section 3.4.

 (i) Show that no UMP test exists for testing $\theta = 0$ against $\theta > 0$.
(ii) Determine the totality of different shapes the MP level-α rejection region for testing $\theta = \theta_0$ against $\theta = \theta_1$ can take on for varying α and $\theta_1 - \theta_0$.

Problem 3.34 Let X_i be independently distributed as $N(i\Delta, 1), i = 1, \ldots, n$. Show that there exists a UMP test of $H : \Delta \leq 0$ against $K : \Delta > 0$, and determine it as explicitly as possible.

Problem 3.35 Suppose a time series $X_0, X_1, X_2, \ldots$ evolves in the following way. The process starts at 0, so $X_0 = 0$. For any $i \geq 1$, conditional on $X_0, \ldots, X_{i-1}, X_i = \rho X_{i-1} + \epsilon_i$, where the ϵ_i are i.i.d. standard normal. You observe $X_0, X_1, X_2, \ldots, X_n$. For testing the null hypothesis $\rho = 0$ versus a fixed alternative $\rho = \rho' > 0$, determine a MP level α test. Determine whether or not there exists a uniformly most powerful test against all $\rho > 0$.

Note. The following problems (and some in later chapters) refer to the gamma, Pareto, Weibull, and inverse Gaussian distributions. For more information about these distributions, see Chapters 17, 19, 20, and 25 respectively of Johnson and Kotz (1970).

Problem 3.36 Let $X_1, \ldots, X_n$ be a sample from the *gamma distribution* $\Gamma(g, b)$ with density

$$\frac{1}{\Gamma(g)b^g}x^{g-1}e^{-x/b}, \qquad 0 < x, \quad 0 < b, g.$$

Show that there exists a UMP test for testing

 (i) $H : b \leq b_0$ against $b > b_0$ when g is known;
(ii) $H : g \leq g_0$ against $g > g_0$ when b is known.

In each case give the form of the rejection region.

Problem 3.37 A random variable X has the *Weibull distribution* $W(b, c)$ if its density is

$$\frac{c}{b}\left(\frac{x}{b}\right)^{c-1}e^{-(x/b)^c}, \qquad x > 0, b, c > 0.$$

Show that this defines a probability density. If $X_1, \ldots, X_n$ is a sample from $W(b, c)$, with the shape parameter c known, show that there exists a UMP test of $H : b \leq b_0$ against $b > b_0$ and give its form.

Problem 3.38 Consider a single observation X from $W(1, c)$.

  (i) The family of distributions does not have a monotone likelihood ratio in x.
 (ii) The most powerful test of $H : c = 1$ against $c = 2$ rejects when $X < k_1$ and when $X > k_2$. Show how to determine k_1 and k_2.
(iii) Generalize (ii) to arbitrary alternatives $c_1 > 1$, and show that a UMP test of $H : c = 1$ against $c > 1$ does not exist.
 (iv) For any $c_1 > 1$, the power function of the MP test of $H : c = 1$ against $c = c_1$ is an increasing function of c.

Problem 3.39 Let $X_1, \ldots, X_n$ be a sample from the *inverse Gaussian* distribution $I(\mu, \tau)$ with density

$$\sqrt{\frac{\tau}{2\pi x^3}} \exp\left(-\frac{\tau}{2x\mu^2}(x - \mu)^2\right), \qquad x > 0, \quad \tau, \mu > 0.$$

Show that there exists a UMP test for testing

  (i) $H : \mu \le \mu_0$ against $\mu > \mu_0$ when τ is known;
 (ii) $H : \tau \le \tau_0$ against $\tau > \tau_0$ when μ is known. In each case give the form of the rejection region.
(iii) The distribution of $V = r(X_i - \mu)^2/X_i\mu^2$ is χ_1^2 and hence that of $\tau \sum[(X_i - \mu)^2/X_i\mu^2]$ is χ_n^2.

[Let $Y = \min(X_i, \mu^2/X_i)$, $Z = \tau(Y - \mu)^2/\mu^2 Y$. Then $Z = V$ and Z is χ_1^2 [Shuster (1968)].] *Note*. The UMP test for (ii) is discussed in Chhikara and Folks (1976).

Problem 3.40 Let $X_1, \cdots, X_n$ be a sample from a location family with common density $f(x - \theta)$, where the location parameter $\theta \in \mathbf{R}$ and $f(\cdot)$ is known. Consider testing the null hypothesis that $\theta = \theta_0$ versus an alternative $\theta = \theta_1$ for some $\theta_1 > \theta_0$. Suppose there exists a most powerful level α test of the form: reject the null hypothesis iff $T = T(X_1, \cdots, X_n) > C$, where C is a constant and $T(X_1, \ldots, X_n)$ is location equivariant, i.e., $T(X_1 + c, \ldots, X_n + c) = T(X_1, \ldots, X_n) + c$ for all constants c. Is the test also most powerful level α for testing the null hypothesis $\theta \le \theta_0$ against the alternative $\theta = \theta_1$. Prove or give a counterexample.

Problem 3.41 *Extension of Lemma* 3.4.2. Let P_0 and P_1 be two distributions with densities p_0, p_1 such that $p_1(x)/p_0(x)$ is a nondecreasing function of a real-valued statistic $T(x)$.

  (i) If $T = T(X)$ has probability density p_i' when the original distribution of X is P_i, then $p_1'(t)/p_0'(t)$ is nondecreasing in t.
 (ii) $E_0\psi(T) \le E_1\psi(T)$ for any nondecreasing function ψ.
(iii) If $p_1(x)/p_0(x)$ is a strictly increasing function of $t = T(x)$, so is $p_1'(t)/p_0'(t)$, and $E_0\psi(T) < E_1\psi(T)$ unless $\psi[T(x)]$ is constant a.e. $(P_0 + P_1)$ or $E_0\psi(T) = E_1\psi(T) = \pm\infty$.

(iv) For any distinct distributions with densities p_0, p_1,

$$-\infty \leq E_0 \log \left[\frac{p_1(X)}{p_0(X)} \right] < E_1 \log \left[\frac{p_1(X)}{p_0(X)} \right] \leq \infty.$$

[(i): Without loss of generality suppose that $p_1(x)/p_0(x) = T(x)$. Then for any integrable ϕ,

$$\int \phi(t) p_1'(t) \, dv(t) = \int \phi[T(x)] T(x) p_0(x) \, d\mu(x) = \int \phi(t) t p_0'(t) \, dv(t),$$

and hence $p_1'(t)/p_0'(t) = t$ a.e.
(iv): The possibility $E_0 \log[p_1(X)/p_0(X)] = \infty$ is excluded, since by the convexity of the function log,

$$E_0 \log \left[\frac{p_1(X)}{p_0(X)} \right] < \log E_0 \left[\frac{p_1(X)}{p_0(X)} \right] = 0.$$

Similarly for E_1. The strict inequality now follows from (iii) with $T(x) = p_1(x)/p_0(x)$.]

Problem 3.42 F_0, F_1 are two cumulative distribution functions on the real line, then $F_i(x) \leq F_0(x)$ for all x if and only if $E_0 \psi(X) \leq E_1 \psi(X)$ for any nondecreasing function ψ.

Problem 3.43 Let F and G be two continuous, strictly increasing c.d.f.s, and let $k(u) = G[F^{-1}(u)]$, $0 < u < 1$.
(i) Show F and G are stochastically ordered, say $F(x) \leq G(x)$ for all x, if and only if $k(u) \leq u$ for all $0 < u < 1$.
(ii) If F and G have densities f and g, then show they are monotone likelihood ratio ordered, say g/f nondecreasing, if and only if k is convex.
(iii) Use (i) and (ii) to give an alternative proof of the fact that MLR implies stochastic ordering.

Problem 3.44 Let $f(x)/[1 - F(x)]$ be the "mortality" of a subject at time x given that it has survived to this time. A c.d.f. F is said to be smaller than G in the hazard ordering if

$$\frac{g(x)}{1 - G(x)} \leq \frac{f(x)}{1 - F(x)} \quad \text{for all } x \, . \tag{3.48}$$

(i) Show that (3.48) is equivalent to

$$\frac{1 - F(x)}{1 - G(x)} \quad \text{is nonincreasing.} \tag{3.49}$$

(ii) Show that (3.48) holds if and only if k is starshaped. [A function k defined on an interval $I \subseteq [0, \infty)$ is starshaped on I if $k(\lambda x) \leq \lambda k(x)$ whenever $x \in I$, $\lambda x \in I$, $0 \leq \lambda \leq 1$. Problems 3.43 and 3.44 are based on Lehmann and Rojo (1992).]

Section 3.5

Problem 3.45 Typically, lower confidence bounds $\underline{\theta}(X)$ satisfying (3.21) also satisfy

$$P_\theta\{\underline{\theta}(X) < \theta\} \geq 1 - \alpha \qquad \text{for all } \theta$$

so that θ is strictly greater than $\underline{\theta}(X)$ with probability $\geq 1 - \alpha$. A similar issue of course also applies to upper confidence bounds. Investigate conditions where one can claim the endpoints are open endpoints. What happens in Example 3.5.2 for both the uniformly most accurate upper confidence bound, as well as the Clopper-Pearson solution?

Problem 3.46 In Example 3.5.2, what is an explicit formula for the uniformly most accurate upper bound at level $1 - \alpha$ when $X = 0$ and $U = u$? Compare it to the Clopper-Pearson bound in the same situation.

Problem 3.47 (i) For $n = 5, 10$ and $1 - \alpha = 0.95$, graph the upper confidence limits $\bar{p}$ and $\bar{p}^*$ of Example 3.5.2 as functions of $t = x + u$.
(ii) For the same values of n and $\alpha_1 = \alpha_2 = 0.05$, graph the lower and upper confidence limits $\underline{p}$ and $\bar{p}$.

Problem 3.48 (i) Suppose $U_1, \ldots, U_n$ are i.i.d. $U(0, 1)$ and let $U_{(k)}$ denote the kth largest value (or kth order statistic). Find the density of $U_{(k)}$ and show that

$$P\{U_{(k)} \leq p\} = \int_0^p \frac{n!}{(k-1)!(n-k)!} u^{k-1}(1-u)^{n-k} du \;,$$

which in turn is equal to

$$\sum_{j=k}^n \binom{n}{j} p^j (1-p)^{n-j} \;.$$

(ii) Use (i) to show that, in Example 3.5.2 with $0 < x < n$, the Clopper-Pearson solution $\hat{p}_U$ for an upper $1 - \alpha$ confidence bound for p can be expressed as the $1 - \alpha$ quantile of the Beta distribution with parameters $x + 1$ and $n - x$.

Problem 3.49 *Confidence bounds with minimum risk.* Let $L(\theta, \underline{\theta})$ be nonnegative and nonincreasing in its second argument for $\underline{\theta} < \theta$, and equal to 0 for $\underline{\theta} \geq \theta$. If $\underline{\theta}$ and $\underline{\theta}^*$ are two lower confidence bounds for θ such that

$$P_0\{\underline{\theta} \leq \theta'\} \leq P_\theta\{\underline{\theta}^* \leq \theta'\} \qquad \text{for all} \quad \theta' \leq \theta,$$

then

$$E_\theta L(\theta, \underline{\theta}) \le E_\theta L(\theta, \underline{\theta}^*).$$

[Define two cumulative distribution functions F and F^* by $F(u) = P_\theta\{\theta \le u\}/P_\theta\{\underline{\theta}^* \le \theta\}$, $F^*(u) = P_\theta\{\underline{\theta}^* \le u\}/P_\theta\{\underline{\theta}^* \le \theta\}$ for $u < \theta$, $F(u) = F^*(u) = 1$ for $u \ge \theta$. Then $F(u) \le F^*(u)$ for all u, and it follows from Problem 3.42 that

$$E_\theta[L(\theta, \underline{\theta})] = P_\theta\{\theta^* \le \theta\} \int L(\theta, u) dF(u)$$

$$\le P_\theta\{\underline{\theta}^* \le \theta\} \int L(\theta, u) dF^*(u) = E_\theta[L(\theta, \underline{\theta}^*)].]$$

Section 3.6

Problem 3.50 If $\beta(\theta)$ denotes the power function of the UMP test of Corollary 3.4.1, and if the function Q of (3.19) is differentiable, then $\beta'(\theta) > 0$ for all θ for which $Q'(\theta) > 0$.

[To show that $\beta'(\theta_0) > 0$, consider the problem of maximizing, subject to $E_{\theta_0}\phi(X) = \alpha$, the derivative $\beta'(\theta_0)$ or equivalently the quantity $E_{\theta_0}[T(X) \phi(X)]$.]

Problem 3.51 *Optimum selection procedures.* On each member of a population n measurements $(X_1, \ldots, X_n) = X$ are taken, for example the scores of n aptitude tests which are administered to judge the qualifications of candidates for a certain training program. A future measurement Y such as the score in a final test at the end of the program is of interest but unavailable. The joint distribution of X and Y is assumed known.

 (i) One wishes to select a given proportion α of the candidates in such a way as to maximize the expectation of Y for the selected group. This is achieved by selecting the candidates for which $E(Y|x) \ge C$, where C is determined by the condition that the probability of a member being selected is α. When $E(Y|x) = C$, it may be necessary to randomize in order to get the exact value α.
(ii) If instead the problem is to maximize the probability with which in the selected population Y is greater than or equal to some preassigned score y_0, one selects the candidates for which the conditional probability $P\{Y \ge y_0|x\}$ is sufficiently large.

[(i): Let $\phi(x)$ denote the probability with which a candidate with measurements x is to be selected. Then the problem is that of maximizing

$$\int \left[\int y p^{Y|x}(y) \, \phi(x) dy \right] p^x(x) dx$$

subject to

$$\int \phi(x)p^x(x)dx = \alpha.]$$

Problem 3.52 The following example shows that Corollary 3.6.1 does not extend to a countably infinite family of distributions. Let p_n be the uniform probability density on $[0, 1 + 1/n]$, and p_0 the uniform density on $(0, 1)$.

(i) Then p_0 is linearly independent of $(p_1, p_2, \ldots)$, that is, there do not exist constants $c_1, c_2, \ldots$ such that $p_0 = \sum c_n p_n$.
(ii) There does not exist a test ϕ such that $\int \phi p_n = \alpha$ for $n = 1, 2, \ldots$ but $\int \phi p_0 > \alpha$.

Problem 3.53 Let $F_1, \ldots, F_{m+1}$ be real-valued functions defined over a space U. A sufficient condition for u_0 to maximize F_{m+1} subject to $F_i(u) \leq c_i (i = 1, \ldots, m)$ is that it satisfies these side conditions, that it maximizes $F_{m+1}(u) - \sum k_i F_i(u)$ for some constants $k_i \geq 0$, and that $F_i(u_o) = c_i$ for those values i for which $k_i > 0$.

Section 3.7

Problem 3.54 For a random variable X with binomial distribution $b(p, n)$, determine the constants C_i, γ $(i = 1, 2)$ in the UMP test (3.33) for testing $H : p \leq 0.2$ or ≤ 0.7 when $\alpha = 0.1$ and $n = 15$. Find the power of the test against the alternative $p = 0.4$.

Problem 3.55 *Totally positive families.* A family of distributions with probability densities $p_\theta(x)$, θ and x real-valued and varying over Ω and $\mathcal{X}$, respectively, is said to be totally positive of order $r (\mathrm{TP}_r)$ if for all $x_1 < \cdots < x_n$ and $\theta_1 < \cdots < \theta_n$

$$\triangle_n = \begin{vmatrix} p_{\theta_1}(x_1) & \cdots & p_{\theta_1}(x_n) \\ p_{\theta_n}(x_1) & \cdots & p_{\theta_n}(x_n) \end{vmatrix} \geq 0 \quad \text{for all} \quad n = 1, 2, \ldots, r. \tag{3.50}$$

It is said to be strictly totally positive of order r (STP_r) if strict inequality holds in (3.50). The family is said to be (strictly) totally positive of infinity if (3.50) holds for all $n = 1, 2, \ldots$. These definitions apply not only to probability densities but to any real-valued functions $p_\theta(x)$ of two real variables.

(i) For $r = 1$, (3.50) states that $p_\theta(x) \geq 0$; for $r = 2$, that $p_\theta(x)$ has monotone likelihood ratio in x.
(ii) If $a(\theta) > 0$, $b(x) > 0$, and $p_\theta(x)$ is STP_r then so is $a(\theta)b(x)p_\theta(x)$.
(iii) If a and b are real-valued functions mapping Ω and $\mathcal{X}$ onto Ω' and $\mathcal{X}'$ and are strictly monotone in the same direction, and if $p_\theta(x)$ is STP_r, then $p_{\theta'}(x')$ with $\theta' = a^{-1}(\theta)$ and $x' = b^{-1}(x)$ is $(STP)_r$ over $(\Omega', \mathcal{X}')$.

Problem 3.56 *Exponential families.* The exponential family (3.19) with $T(x) = x$ and $Q(\theta) = \theta$ is STP_∞, with Ω the natural parameter space and $\mathcal{X} = (-\infty, \infty)$.

[That the determinant $|e^{\theta_i x_j}|$, $i, j = 1, \ldots, n$, is positive can be proved by induction. Divide the ith column by $e^{\theta_1 x_i}$, $i = 1, \ldots, n$; subtract in the resulting determinant the $(n-1)$st column from the nth, the $(n-2)$nd from the $(n-1)$st, $\ldots$, the 1st from the 2nd; and expand the determinant obtained in this way by the first row. Then $\triangle_n$ is seen to have the same sign as

$$\triangle'_n = |e^{\eta_i x_j} - e^{\eta_i x_j - 1}|, \qquad i, j = 2, \ldots, n,$$

where $\eta_i = \theta_i - \theta_1$. If this determinant is expanded by the first column, one obtains a sum of the form

$$a_2(e^{\eta_2 x_2} - e^{\eta_2 x_1}) + \cdots + a_n(e^{\eta_n x_2} - e^{\eta_n x_1}) = h(x_2) - h(x_1)$$
$$= (x_2 - x_1)h'(y_2),$$

where $x_1 < y_2 < x_2$. Rewriting $h'(y_2)$ as a determinant of which all columns but the first coincide with those of $\triangle'_n$ and proceeding in the same manner with the columns, one reduces the determinant to $|e^{\eta_i y_j}|$, $i, j = 2, \ldots, n$, which is positive by the induction hypothesis.]

Problem 3.57 STP_3. Let θ and x be real-valued, and suppose that the probability densities $p_\theta(x)$ are such that $p_{\theta'}(x)/p_\theta(x)$ is strictly increasing in x for $\theta < \theta'$. Then the following two conditions are equivalent: (a) For $\theta_1 < \theta_2 < \theta_3$ and $k_1, k_2, k_3 > 0$, let

$$g(x) = k_1 p_{\theta_1}(x) - k_2 p_{\theta_2}(x) + k_3 p_{\theta_3}(x).$$

If $g(x_1) - g(x_3) = 0$, then the function g is positive outside the interval (x_1, x_3) and negative inside. (b) The determinant $\triangle_3$ given by (3.50) is positive for all $\theta_1 < \theta_2 < \theta_3$, $x_1 < x_2 < x_3$. [It follows from (a) that the equation $g(x) = 0$ has at most two solutions.]

[That (b) implies (a) can be seen for $x_1, < x_2 < x_3$ by considering the determinant

$$\begin{vmatrix} g(x_1) & g(x_2) & g(x_3) \\ p_{\theta_2}(x_1) & p_{\theta_2}(x_2) & p_{\theta_2}(x_3) \\ p_{\theta_3}(x_1) & p_{\theta_3}(x_2) & p_{\theta_3}(x_3) \end{vmatrix}$$

Suppose conversely that (a) holds. Monotonicity of the likelihood ratios implies that the rank of $\triangle_3$ is at least two, so that there exist constants k_1, k_2, k_3 such that $g(x_1) = g(x_3) = 0$. That the k's are positive follows again from the monotonicity of the likelihood ratios.]

Problem 3.58 *Extension of Theorem* 3.7.1. The conclusions of Theorem 3.7.1 remain valid if the density of a sufficient statistic T (which without loss of generality will be taken to be X), say $p_\theta(x)$, is STP_3 and is continuous in x for each θ.

[The two properties of exponential families that are used in the proof of Theorem 3.7.1 are continuity in x and (a) of the preceding problem.]

Problem 3.59 For testing the hypothesis $H' : \theta_1 \leq \theta \leq \theta_2 (\theta_1 \leq \theta_2)$ against the alternatives $\theta < \theta_1$ or $\theta > \theta_2$, or the hypothesis $\theta = \theta_0$ against the alternatives $\theta \neq \theta_0$, in an exponential family or more generally in a family of distributions satisfying the assumptions of Problem 3.58, a UMP test does not exist.

[This follows from a consideration of the UMP tests for the one-sided hypotheses $H_1 : \theta \geq \theta_1$ and $H_2 : \theta \leq \theta_2$.]

Problem 3.60 Let f, g be two probability densities with respect to μ. For testing the hypothesis $H : \theta \leq \theta_0$ or $\theta \geq \theta_1 (0 < \theta_0 < \theta_1 < 1)$ against the alternatives $\theta_0 < \theta < \theta_1$, in the family $\mathcal{P} = \{\theta f(x) + (1 - \theta)g(x), 0 \leq \theta \leq 1\}$, the test $\varphi(x) \equiv \alpha$ is UMP at level α.

Section 3.8

Problem 3.61 Let the variables $X_i (i = 1, \ldots, s)$ be independently distributed with Poisson distribution $P(\lambda_i)$. For testing the hypothesis $H : \sum \lambda_j \leq a$ (for example, that the combined radioactivity of a number of pieces of radioactive material does not exceed a), there exists a UMP test, which rejects when $\sum X_j > C$.

[If the joint distribution of the X's is factored into the marginal distribution of $\sum X_j$ (Poisson with mean $\sum \lambda_j$) times the conditional distribution of the variables $Y_i = X_j / \sum X_j$ given $\sum X_j$ (multinomial with probabilities $p_i = \lambda_i / \sum \lambda_j$), the argument is analogous to that given in Example 3.8.1.]

Problem 3.62 *Confidence bounds for a median.* Let $X_1, \ldots, X_n$ be a sample from a continuous cumulative distribution functions F. Let ξ be the unique median of F if it exists, or more generally let $\xi = \inf\{\xi' : F(\xi') \geq \frac{1}{2}\}$.

(i) If the ordered X's are $X_{(1)} < \cdots < X_{(n)}$, a uniformly most accurate lower confidence bound for ξ is $\underline{\xi} = X_{(k)}$ with probability ρ, $\underline{\xi} = X_{(k+1)}$ with probability $1 - \rho$, where k and ρ are determined by

$$\rho \sum_{j=k}^{n} \binom{n}{j} \frac{1}{2^n} + (1 - \rho) \sum_{j=k+1}^{n} \binom{n}{j} \frac{1}{2^n} = 1 - \alpha.$$

(ii) This bound has confidence coefficient $1 - \alpha$ for any median of F.
(iii) Determine most accurate lower confidence bounds for the $100p$-percentile ξ of F defined by $\xi = \inf\{\xi' : F(\xi') = p\}$.

[For fixed ξ_0, the problem of testing $H : \xi = \xi_0$ to against $K : \xi > \xi_0$ is equivalent to testing $H' : p = \frac{1}{2}$ against $K' : p < \frac{1}{2}$.]

Problem 3.63 A *counterexample*. Typically, as α varies, the most powerful level α tests for testing a hypothesis H against a simple alternative are nested in the sense that the associated rejection regions, say R_α, satisfy $R_\alpha \subseteq R_{\alpha'}$, for any $\alpha < \alpha'$. The following example shows that this need not be satisfied for composite H. Let X take on the values 1, 2, 3, 4 with probabilities under distributions P_0, P_1, Q:

	1	2	3	4
P_0	$\frac{2}{13}$	$\frac{4}{13}$	$\frac{3}{13}$	$\frac{4}{13}$
P_1	$\frac{4}{13}$	$\frac{2}{13}$	$\frac{1}{13}$	$\frac{6}{13}$
Q	$\frac{4}{13}$	$\frac{3}{13}$	$\frac{2}{13}$	$\frac{4}{13}$

Then the most powerful test for testing the hypothesis that the distribution of X is P_0 or P_1 against the alternative that it is Q rejects at level $\alpha = \frac{5}{13}$ when $X = 1$ or 3, and at level $\alpha = \frac{6}{13}$ when $X = 1$ or 2.

Problem 3.64 Let X and Y be the number of successes in two sets of n binomial trials with probabilities p_1 and p_2 of success.

(i) The most powerful test of the hypothesis $H : p_2 \leq p_1$ against an alternative (p_1', p_2') with $p_1' < p_2'$ and $p_1' + p_2' = 1$ at level $\alpha < \frac{1}{2}$ rejects when $Y - X > C$ and with probability γ when $Y - X = C$.

(ii) This test is not UMP against the alternatives $p_1 < p_2$.

[(i): Take the distribution Λ assigning probability 1 to the point $p_1 = p_2 = \frac{1}{2}$ as an a priori distribution over H. The most powerful test against (p_1', p_2') is then the one proposed above. To see that Λ is least favorable, consider the probability of rejection $\beta(p_1, p_2)$ for $p_1 = p_2 = p$. By symmetry this is given by

$$2\beta(p, p) = P\{|Y - X| > C\} + \gamma P\{|Y - X| = C\}.$$

Let X_i be 1 or 0 as the ith trial in the first series is a success or failure, and let Y_1, be defined analogously with respect to the second series. Then $Y - X = \sum_{i-1}^n (Y_i - X_i)$, and the fact that $2\beta(p, p)$ attains its maximum for $p = \frac{1}{2}$ can be proved by induction over n.

(ii): Since $\beta(p, p) < \alpha$ for $p \neq 1$, the power $\beta(p_1, p_2)$ is $< \alpha$ for alternatives $p_1 < p_2$ sufficiently close to the line $p_1 = p_2$. That the test is not UMP now follows from a comparison with $\phi(x, y) \equiv \alpha$.]

Problem 3.65 *Sufficient statistics with nuisance parameters*.

(i) A statistic T is said to be *partially sufficient* for θ in the presence of a nuisance parameter η if the parameter space is the direct product of the set of possible θ- and η-values, and if the following two conditions hold: (a) the conditional distribution given $T = t$ depends only on η; (b) the marginal distribution of T depends only on θ. If these conditions are satisfied, there exists a UMP test for testing the composite hypothesis $H : \theta = \theta_0$ against the composite class of alternatives $\theta = \theta_1$, which depends only on T.

(ii) Part (i) provides an alternative proof that the test of Example 3.8.1 is UMP.

[Let $\psi_0(t)$ be the most powerful level α test for testing θ_0 against θ_1 that depends only on t, let $\phi(x)$ be any level-α test, and let $\psi(t) = E_{\eta_1}[\phi(X) \mid t]$. Since $E_{\theta_i}\psi(T) = E_{\theta_i,\eta_1}\phi(X)$, it follows that ψ is a level-α test of H and its power, and therefore the power of ϕ, does not exceed the power of ψ_0.]

Note. For further discussion of this and related concepts of partial sufficiency see Fraser (1956), Dawid (1975), Sprott (1975), Basu (1978), and Barndorff-Nielsen (1978).

Section 3.9

Problem 3.66 Let $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ be independent samples from $N(\xi, 1)$ and $N(\eta, 1)$, and consider the hypothesis $H : \eta \leq \xi$ against $K : \eta > \xi$. There exists a UMP test, and it rejects the hypothesis when $\bar{Y} - \bar{X}$ is too large.

[If $\xi_1 < \eta_1$ is a particular alternative, the distribution assigning probability 1 to the point $\eta = \xi = (m\xi_1 + n\eta_1)/(m + n)$ is least favorable.]

Problem 3.67 Let $X_1, \ldots, X_m$; $Y_1, \ldots, Y_n$ be independently, normally distributed with means ξ and η, and variances a σ^2 and τ^2 respectively, and consider the hypothesis $H : \tau \leq \sigma$ a against $K : \sigma < \tau$.

 (i) If ξ and η are known, there exists a UMP test given by the rejection region $\sum(Y_j - \eta)^2 / \sum(X_i - \xi)^2 \geq C$.
(ii) No UMP test exists when ξ and η are unknown.

Problem 3.68 Suppose X is a $k \times 1$ random vector with $E(|X|^2) < \infty$ and covariance matrix Σ. Let A be an $m \times k$ (nonrandom) matrix and let $Y = AX$. Show Y has mean vector $AE(X)$ and covariance matrix $A\Sigma A^\top$.

Problem 3.69 Suppose $(X_1, \ldots, X_k)$ has the multivariate normal distribution with unknown mean vector $\xi = (\xi_1, \ldots, \xi_k)$ and known covariance matrix Σ. Suppose X_1 is independent of $(X_2, \ldots, X_k)$. Show that X_1 is partially sufficient for ξ_1 in the sense of Problem 3.65. Provide an alternative argument for Case 2 of Example 3.9.2.

Problem 3.70 In Example 3.9.2, Case 2, verify the claim for the least favorable distribution.

Problem 3.71 In Example 3.9.3, provide the details for Cases 3 and 4.

3.11   Notes

Hypothesis testing developed gradually, with early instances frequently being rather vague statements of the significance or nonsignificance of a set of observations. Isolated applications are found in the eighteenth century [Arbuthnot (1710), Daniel Bernoulli (1734), and Laplace (1773), for example] and centuries earlier in the Royal Mint's Trial of the Pyx [discussed by Stigler (1977)]. They became more frequent in the nineteenth century in the writings of such authors as Gavarret (1840), Lexis (1875, 1877), and Edgeworth (1885). A new stage began with the work of Karl Pearson, particularly his χ^2 paper of 1900, followed in the decade 1915–1925 by Fisher's normal theory and χ^2 tests. Fisher presented this work systematically in his enormously influential book *Statistical Methods for Research Workers* (1925b).

The first authors to recognize that the rational choice of a test must involve consideration not only of the hypothesis but also of the alternatives against which it is being tested were Neyman and Pearson (1928). They introduced the distinction between errors of the first and second kind, and thereby motivated their proposal of the likelihood ratio criterion as a general method of test construction. These considerations were carried to their logical conclusion by Neyman and Pearson in their paper of 1933, in which they developed the theory of UMP tests. Accounts of their collaboration can be found in Pearson's recollections (1966), and in the biography of Neyman by Reid (1982).

The Neyman–Pearson Lemma has been generalized in many directions, including the results in Sections 3.6, 3.8, and 3.9. Dantzig and Wald (1951) give necessary conditions including those of Theorem 3.6.1, for a critical function which maximizes an integral subject to a number of integral side conditions, to satisfy (3.30). The role of the Neyman–Pearson Lemma in hypothesis testing is surveyed in Lehmann (1985a).

An extension to a selection problem, proposed by Birnbaum and Chapman (1950), is sketched in Problem 3.51. Further developments in this area are reviewed in Gibbons (1986, 1988). Grenander (1981) applies the fundamental lemma to problems in stochastic processes.

Lemmas 3.4.1, 3.4.2, and 3.7.1 are due to Lehmann (1961).

Complete class results for simple null hypothesis testing problems are obtained in Brown and Marden (1989).

The earliest example of confidence intervals appears to occur in the work of Laplace (1812), who points out how an (approximate) probability statement concerning the difference between an observed frequency and a binomial probability p can be inverted to obtain an associated interval for p. Other examples can be found in the works of Gauss (1816), Fourier (1826), and Lexis (1875). However, in all these cases, although the statements made are formally correct, the authors appear to consider the parameter as the variable which with the stated probability falls in the fixed confidence interval. The proper interpretation seems to have been pointed out for the first time by E. B. Wilson (1927). About the same time two examples of exact confidence statements were given by Working and Hotelling (1929) and Hotelling (1931).

A general method for obtaining exact confidence bounds for a real-valued parameter was proposed by Fisher (1930), who however later disavowed this interpretation of his work. For a discussion of Fisher's controversial concept of fiducial probability, see Section 5.7. At about the same time,[16] a completely general theory of confidence statements was developed by Neyman and shown by him to be intimately related to the theory of hypothesis testing. A detailed account of this work, which underlies the treatment given here, was published by Neyman in his papers of 1937 and 1938.

The calculation of p-values was the standard approach to hypothesis testing throughout the nineteenth century and continues to be widely used today. For various questions of interpretation, extensions, and critiques, see Cox (1977), Berger and Sellke (1987), Marden (1991), Hwang et al. (1992), Lehmann (1993), Robert (1994), Berger et al. (1994), Meng (1994), Blyth and Staudte (1995, 1997), Liu and Singh (1997), Sackrowitz and Samuel-Cahn (1999), Marden (2000), Sellke et al. (2001), and Berger (2003).

Extensions of p-values to hypotheses with nuisance parameters are discussed by Berger and Boos (1994) and Bayarri and Berger (2000), and the large-sample behavior of p-values in Lambert and Hall (1982) and Robins et al. (2000). An optimality theory in terms of p-values is sketched by Schweder (1988), and p-values for the simultaneous testing of several hypotheses is treated by Schweder and Spjøtvoll (1982), Westfall and Young (1993), and by Dudoit et al. (2003).

An important use of p-values occurs in meta-analysis when one is dealing with the combination of results from independent experiments. The early literature on this topic is reviewed in Hedges and Olkin (1985, Chapter 3). Additional references are Marden (1982b, 1985), Scholz (1982), and a review article by Becker (1997). Associated confidence intervals are proposed by Littell and Louv (1981).

---

[16] Cf. Neyman (1941b).