CHAPTER 3. Theory of Hypothesis Testing

1 Introduction

1. Remark. Recall the notation used for the test procedure in MT130/MT230. We have:

a null hypothesis H_0 and an alternative hypothesis H_1 (two 'conjectures' about the distribution of the sample variables);

a test statistic TS (a function of the sample random variables);

a chosen level of significance (l.o.s.) α with $0 < \alpha < 1$;

a critical region R, a subset of $\mathbb{R}^1$, such that $\Pr(TS \in R \,|\, H_0) = \alpha$;

a decision method which rejects the null hypothesis H_0 at the $100\alpha\%$ l.o.s if the observed value of the test statistic TS lies in the critical region R.

Associated with this test method we have:

Type I error = rejecting H_0 when, in fact, H_0 is true - with the probability of committing a Type I error being simply $\Pr(TS \in R \,|\, H_0) = \alpha =$ the chosen l.o.s.; and,

Type II error = accepting H_0 when, in fact, H_1 is true - with the probability of committing a Type II error being $\Pr(TS \notin R \,|\, H_1)$.

Recall, also, that the only role played by the alternative hypothesis was in determining the 'shape' of the critical region - a left- , right- or two-tail region/test.

In the earlier courses in statistics, the approach to a test of hypothesis has been to conjure up a test statistic and critical region, and to rely on intuition as the motivation and justification of the method. The student was not expected to inquire whether an alternative choice of test statistic and critical region might, in some sense, have been better, or whether there existed a best test procedure. We now provide a response to the question in terms of the Neyman-Pearson theory of hypothesis testing. The criterion used in this theory for distinguishing between two test statistics (and related critical regions) is the size of the probability of Type II error - the test statistic having the smaller probability of Type II error is preferred.

In the following introductory example, note that the calculation of the probabilities is effected by integrating the joint p.d.f. of the sample variables over appropriate regions in two dimensions.

Change of notation: $f(x \,|\, \theta) = f(x, \theta)$.

2. Example. Let X_1, X_2 be a sample of size 2 from a population whose distribution has the p.d.f. $f(x \,|\, \theta) = \theta x^{\theta-1}$ for $0 \le x \le 1$. We compare two tests of the null hypothesis H_0 $(\theta = 2)$ against the alternative hypothesis $H_1(\theta = 1)$, using the level of significance α:

(i) test statistic $S_1 = X_1 + X_2$ and critical region $R_1 = \{s : 0 \le s \le a\}$; and

(ii) test statistic $S_2 = \max(X_1, X_2)$ and critical region $R_2 = \{s : 0 \le s \le b\}$.

By the independence of the sample random variables, the joint p.d.f. of X_1 and X_2 is $f(x, y \,|\theta) = \theta^2\, x^{\theta-1} y^{\theta-1}$ for $0 \le x \le 1$, and $0 \le y \le 1$, reducing to $f(x, y \,|\, 2) = 4xy$ under

H_0 and $f(x, y \mid 1) = 1$ under H_1.

3. Remark. The above method of calculating the probabilities anticipates the change of emphasis in hypothesis testing as described in the Neyman-Pearson theory. This change involves moving away from the test statistic and one-dimensional critical region to a critical region as a subset of n-dimensional space ($n = 2$ in the example).

Writing $C_1 = \{(x_1, x_2) : 0 \leq x_1 + x_2 \leq a\}$ and $C_2 = \{(x_1, x_2)) : 0 \leq \max\ (x_1, x_2) \leq b\}$ we have:

$$S_1 \in R_1 \text{ is equivalent to } (X_1, X_2) \in C_1; \text{ and } S_2 \in R_2 \text{ is equivalent to } (X_1, X_2) \in C_2$$

It follows that the two tests in the Example could be defined purely in terms of the critical regions C_1 and C_2, with the null hypothesis rejected if the observed value of (X_1, X_2) lies in the respective critical region. The terms a *test* and *a critical region* can, in this sense, be used interchangeably. In this case, a critical region is a subset of $\mathbb{R}^2$. In general, when a sample $X_1, X_2, \ldots, X_n$ of size n is given, a critical region (test) is a subset of $\mathbb{R}^n$.

This provides a more general view of a statistical test than that provided by specifying a test statistic and a one-dimensional critical region - **however**, in practice, the general procedure usually reduces to the one-dimensional model.

4. Notation. We write $\mathbf{x} = (x_1, x_2, \ldots, x_n)$ for a general point in n-dimensional space $\mathbb{R}^n$ and, for the sample random variables $X_1, X_2, ..., X_n$, let $\mathbf{X} = (X_1, X_2, ..., X_n)$ denote the random point defined by these variables.

If, for example, the set $C = \{\mathbf{x} : \Sigma_{i=1}^n x_i \leq a\}$ then $\mathbf{x} \in C$ simply means $\Sigma_{i=1}^n x_i \leq a$, and $\mathbf{X} \in C$ signifies that $\Sigma_{i=1}^n X_i \leq a$, or that $\bar{X} \leq a/n$.

Further, when $X_1, X_2, ..., X_n$ have the p.d.f. $f(\mathbf{x})$ we have

$$\Pr(\mathbf{X} \in C) = \int \ldots \int_C f(\mathbf{x}) d\mathbf{x},$$

where $d\mathbf{x} = dx_1 dx_2 ... dx_n$ and, when $X_1, X_2, ..., X_n$ have the p.f. $f(\mathbf{x})$ we have

$$\Pr(\mathbf{X} \in C) = \sum \ldots \sum_{\mathbf{x} \in C} f(\mathbf{x}).$$

Although we need the general formulae for our theoretical considerations we do not need to calculate such integrals or sums directly - fortunately, all our examples reduce to one-dimensional integrals or sums!

2. The Neyman-Pearson theory

1. Notation. Throughout this section, let $X_1, X_2, ..., X_n$ denote the sample random variables, with their joint p.f./p.d.f. $f(\mathbf{x} \mid \theta)$ showing the unknown parameter θ. For most of the applications we consider, the random variables will be independent with a common p.f./p.d.f. $f(x \mid \theta)$and, hence, $f(\mathbf{x} \mid \theta) = f(x_1 \mid \theta) f(x_2 \mid \theta) \cdots f(x_n \mid \theta)$. The parameter θ will usually be identified with a point in some k-dimensional space, and we will be concerned with statistical hypotheses which relate to the numerical value of the coordinates of θ.

In each test problem, specified by the null and alternative hypotheses H_0 and H_1, we write Θ as the set of values of θ under discussion - Θ is termed the set of **admissible** values for the problem. The null and alternative hypotheses, H_0 and H_1, define a decomposition of $\Theta = \Theta_0 \cup \Theta_1$ into two disjoint sets, where Θ_0 is the set of values of θ specified in H_0 and Θ_1 the values corresponding to H_1 - this is shown by writing $\mathrm{H}_0(\theta \in \Theta_0)$ and $\mathrm{H}_1(\theta \in \Theta_1) = \mathrm{H}_1(\theta \in \Theta \backslash \Theta_0)$. If $\Theta_0 = \{\theta_0\}$consists of a single element we write $\mathrm{H}_0(\theta = \theta_0)$ for $\mathrm{H}_0(\theta \in \Theta_0)$, and, similarly, if $\Theta_1 = \{\theta_1\}$ we write $\mathrm{H}_1(\theta = \theta_1)$ for $\mathrm{H}_1(\theta \in \Theta_1)$.

2. Definition. A hypothesis is termed **simple** if it is specified by a single element of Θ, and **composite** otherwise.

3. Examples (i) Let $X_1, X_2, ..., X_n$ have the binomial distribution $\mathrm{B}(1, \theta)$.

(a) For the simple null hypothesis $\mathrm{H}_0(\theta = \frac{1}{2})$ against the simple alternative hypothesis $\mathrm{H}_1(\theta = \frac{2}{3})$, the above notation sets $\Theta = \{\frac{1}{2}, \frac{2}{3}\}$, $\Theta_0 = \{\frac{1}{2}\}$ and $\Theta_1 = \{\frac{2}{3}\}$.

(b) For the simple null hypothesis $\mathrm{H}_0\ (\theta = \frac{1}{3})$ against the composite hypothesis $\mathrm{H}_1(\theta \geq \frac{2}{3})$, we write $\Theta = \{\theta : \theta = \frac{1}{3} \text{ or } \theta \geq \frac{2}{3}\}$ with $\Theta_0 = \{\frac{1}{3}\}$ and $\Theta_1 = \{\theta : \theta \geq \frac{2}{3}\}$.

(ii) Let $X_1, X_2, ..., X_n$ have the normal distribution $\mathrm{N}(\mu, \sigma^2)$ with μ and σ unknown. To test the composite null hypothesis $\mathrm{H}_0(\mu = 0)$ against the composite alternative $\mathrm{H}_1(\mu \neq 0)$, we set

$$\Theta = \{\theta = (\mu, \sigma) : -\infty < \mu < \infty, \sigma > 0\}, \quad \Theta_0 = \{\theta = (0, \sigma) : \sigma > 0\},$$
$$\Theta_1 = \{\theta = (\mu, \sigma) : \mu \neq 0, \sigma > 0\}.$$

4. Structure. (i) In the Neyman-Pearson structure, a test of a null hypothesis H_0 against an alternative hypothesis H_1 consists of:

sample random variables $X_1, X_2, ..., X_n$ with a functional form for their joint p.d.f./p.f. $f(\mathbf{x}, \theta)$ for $\theta \in \Theta$ and observed sample data values $x_1, x_2, ..., x_n$;

a null hypothesis $\mathrm{H}_0(\theta \in \Theta_0)$ and an alternative hypothesis $\mathrm{H}_1(\theta \in \Theta_1) = \mathrm{H}_1(\theta \in \Theta \backslash \Theta_0)$;

a test, that is, an n-dimensional critical region C (we identify a test by its critical region C and speak of a test C);

the decision procedure that the null hypothesis is rejected if and only if, the observed value $(x_1, x_2, ..., x_n) \in C$.

(ii) *We say that C is a* ***test of size*** α *$(0 \leq \alpha \leq 1)$ if*

$$\sup_{\theta \in \Theta_0} \Pr(\boldsymbol{X} \in C | \theta) = \alpha$$

We say that C is a ***test of significance level*** α *$(0 \leq \alpha \leq 1)$ if*

$$\sup_{\theta \in \Theta_0} \Pr(\boldsymbol{X} \in C | \theta) \leq \alpha$$

(iii) Note that some authors use the terms *size* and *level* defined in (ii) interchangeably. The distinction between these two becomes important in some models where it is computationally impossible to construct a size α test (see Note 7(i) and Example 8 (ii) below). Note also that the set of level α tests contains the set of size α tests.

(iv) The ***power function*** of the test C is the function of θ defined by

$$\beta(\theta) = \Pr(\mathbf{X} \in C|\theta)$$

for $\theta \in \Theta$.

Thus the probability of Type I error is $\beta(\theta)$ for $\theta \in \Theta_0$ (bounded above by α) and the probability of Type II error is $1 - \beta(\theta)$ for $\theta \in \Theta_1$.

(v) Given two tests C_1 and C_2 of size α, we say C_1 is **uniformly more powerful** than C_2 if

$$\Pr(\mathbf{X} \notin C_1 \mid \theta) \leq \Pr(\mathbf{X} \notin C_2 \mid \theta) \quad \text{for all} \quad \theta \in \Theta_1.$$

Note, that if C_1 is uniformly more powerful than C_2, by complementation we have

$$\beta_{c_1}(\theta) \geq \beta_{c_2}(\theta) \quad \text{for all} \quad \theta \in \Theta_1.$$

(vi) If the test C is uniformly more powerful than any other test of size α then we say that C is **uniformly most powerful (UMP).**

If H_0 and H_1 are both simple hypotheses we omit the word uniformly and describe a UMP test simply as MP (most powerful).

(vii) We require one further piece of terminology to describe a desirable property for a test C. It seems reasonable to suggest that the test should be less likely to reject H_0 when it is true than when it is false.

A test C with power function $\beta(\theta)$ is **unbiased** *if $\beta(\theta') \leq \beta(\theta'')$ for every $\theta' \in \Theta_0$ and $\theta'' \in \Theta_1$.*

Clearly, if C is a test of size α, that is, $\sup_{\theta\in\Theta_0} \beta(\theta) = \alpha$, then unbiasedness implies that $\beta(\theta) \geq \alpha$, for every $\theta \in \Theta_1$,

$$\sup_{\theta\in\Theta_0} \beta(\theta) \leq \inf_{\theta\in\Theta_1} \beta(\theta).$$

5. Introducing the Neyman-Pearson fundamental lemma.

Suppose $\Theta = \{\theta_0,\theta_1\}$ with $\mathrm{H}_0(\theta = \theta_0)$ and $\mathrm{H}_1(\theta = \theta_1)$ as the simple null and alternative hypotheses respectively. In the Theorem below we consider a critical region C of the form

$$(*) \qquad C = \{\mathbf{x} : f(\mathbf{x} \mid \theta_1) \geq k f(\mathbf{x} \mid \theta_0)\} \qquad \text{where} \quad k > 0 \quad \text{is a constant.}$$

Example Suppose the sample variables $X_1, X_2, ..., X_n$ are independent and have the distribution $\mathrm{N}(\theta, 1)$.

For convenience of notation, we write the following discussion for the case when the sample variables have a continuous distribution with joint p.d.f. $f(\mathbf{x} \mid \theta)$ (all calculations given in terms of integration translate, somewhat cumbersomely, into summation notation for a discrete joint distribution).

6. Theorem. The Neyman-Pearson fundamental lemma

If there exists a positive constant k such that the region

$$(\%) \qquad C = \{\mathbf{x} : f(\mathbf{x} \mid \theta_1) \geq kf(\mathbf{x} \mid \theta_0)\} \quad \text{satisfies} \quad \Pr(\mathbf{X} \in C \mid \theta_0) = \alpha,$$

then C is a most powerful (MP) test of size α for the simple null hypothesis $\mathrm{H}_0(\theta = \theta_0)$ against the simple alternative hypothesis $\mathrm{H}_1(\theta = \theta_1)$. Further, the test C is unbiased.

7. Remarks. (i) Note the conditional 'if' in the statement of the fundamental Lemma. There is no guarentee that we can find C and k as required. This is particularly obvious in the case of discrete random variables when there may be no region D of any shape with $\Pr(\mathbf{X}\in D|\theta_0) = \alpha$, let alone one of the required shape specified in the lemma.

(ii) Note that the proof makes no use of the independence of the sample variables $X_1, X_2, ..., X_n$ - all that is used is knowledge of the joint p.d.f./p.f.

(iii) The MP region (if it exists) is essentially unique - we can modify it only by 'including or deleting' a set N with $\Pr(\mathbf{X}\in N|\theta_i) = 0$ for $i = 1, 2$.

(iv) One further observation follows from an examination of the proof. The requirement that the region D provides a test of size α is not required - the proof carries through provided D is a test at significance level α (i.e., with size $\alpha_1 \leq \alpha$). Hence, the test C may be described as being more powerful than any test of significance level α (i.e., than any test of size $\leq \alpha$).

(v) The result that the N-P region is a MP test of significance level α of θ_0 against θ_1 has a considerable intuitive appeal. Consider the ratio $f(\mathbf{x}|\theta_1)/f(\mathbf{x}|\theta_0)$. Any $\mathbf{x}$ for which this ratio is large provides evidence that θ_1 rather than θ_0 is true (this is obvious in the discrete case). If we must choose a subset of possible observations which indicate that θ_1 is the true value of the parameter, then it seems sensible to put into this subset those $\mathbf{x}$'s for which the ratio $f(\mathbf{x}|\theta_1)/f(\mathbf{x}|\theta_0)$ is large - in other words to choose a subset of the form $\{\mathbf{x} : f(\mathbf{x}|\theta_1) \geq kf(\mathbf{x}|\theta_0)\}$. The N-P analysis, based on probabilities of error, now gives us a basis to choose k so that $\Pr\{f(\mathbf{X}|\theta_1) \geq kf(\mathbf{X}|\theta_0)|\theta = \theta_0\} = \alpha$

8. Examples. (i) Suppose the sample variables $X_1, X_2, ..., X_n$ are independent with the normal distribution $\mathrm{N}(\theta, 1)$. We apply the Neyman-Pearson lemma to test the null hypothesis $\mathrm{H}_0(\theta = \theta_0)$ against the alternative hypothesis $\mathrm{H}_1(\theta = \theta_1)$, where $\theta_1 > \theta_0$.

Notes. (a) The critical region provided by the Neyman-Pearson lemma is equivalent to the test statistic $\overline{X}$ and one-dimensional critical region $\{x : x \geq k_1/n\}$ used in M130.

(b) For the null hypotheses $\mathrm{H}_0(\theta = \theta_0)$ against the alternative hypothesis $\mathrm{H}_1(\theta = \theta_1)$, where, now, $\theta_1 < \theta_0$, the inequality (%) is equivalent to $\sum_{i=1}^n x_i \leq k_1$ (since $\theta_1 - \theta_0 < 0$).

(ii) Suppose that the sample random variables $X_1, X_2, ..., X_n$ are independent with the binomial distribution $\mathrm{B}(1, \theta)$, and consider the hypotheses $\mathrm{H}_0(\theta = \frac{1}{2})$ and $\mathrm{H}_1(\theta = \frac{3}{4})$.

Now, the common p.f. of the sample variables is $f(x \,|\, \theta) = \theta^x(1 - \theta)^{1-x}$ for $x = 0$ or 1, and the joint p.f. of $X_1, X_2, ..., X_n$ is therefore $f(\mathbf{x}|\theta) = \theta^t(1 - \theta)^{n-t}$, where $t = \Sigma_{i=1}^n x_i$.

(iii) Suppose that the sample random variables $X_1, X_2, ..., X_n$ are independent with the common exponential p.d.f. $f(x, \theta) = \theta e^{-\theta x}$, where $x \geq 0$ and $\theta > 0$. We seek a test of the null hypothesis $\mathrm{H}_0(\theta = 1)$ against the alternative hypothesis $\mathrm{H}_1(\theta = 2)$.

9. Simple null hypothesis, composite alternative hypothesis. Suppose the set Θ of admissible values of the parameter is written as $\Theta = \Theta_0 \cup \Theta_1$ where $\Theta_0 = \{\theta_0\}$ consists of a single element, and Θ_1 consists of more than one element. We extend the Neyman-Pearson lemma idea to construct (where possible) a UMP test with an n-dimensional critical region C for the simple null hypothesis $\mathrm{H}_0(\theta = \theta_0)$ against the composite alternative $\mathrm{H}_1(\theta \in \Theta_1)$.

A UMP test C of size α, if such a region exists, must, by definition, provide a MP test for the simple hypothesis $\mathrm{H}_0(\theta = \theta_0)$ against each alternative hypothesis $\mathrm{H}_1(\theta = \theta_1)$ for every choice of $\theta_1 \in \Theta_1$. This suggests that, in the search for a UMP region, we apply the fundamental lemma to these simple hypotheses, obtaining a critical region $C(\theta_1)$ which, in principle, depends on the choice of $\theta_1 \in \Theta_1$. Should it happen that $C = C(\theta_1)$ is the same for all $\theta_1 \in \Theta_1$, then C must be a UMP test for the null hypothesis $\mathrm{H}_0(\theta = \theta_0)$ against the alternative hypothesis $\mathrm{H}_1(\theta \in \Theta_1)$.

10. Example. Suppose that the sample variables $X_1, X_2, ..., X_n$ are independent with the normal distribution $\mathrm{N}(\theta, 1)$. Let $\Theta = \{\theta : \theta \geq \theta_0\}$, $\Theta_0 = \{\theta_0\}$ and $\Theta_1 = \{\theta : \theta > \theta_0\}$, then we require a test of the null hypothesis $\mathrm{H}_0(\theta = \theta_0)$ against the alternative hypothesis $\mathrm{H}_1(\theta > \theta_0)$.

Referring to Example 8(i) for a test of $\mathrm{H}_0(\theta = \theta_0)$ against $\mathrm{H}_1(\theta = \theta_1)$, where $\theta_1 > \theta_0$, the critical region $C(\theta_1)$ is given by $C(\theta_1) = \{\mathbf{x}\colon \sum_{i=1}^n x_i \geq k_1\}$ where $k_1 = \frac{\log k + n(\theta_1^2 - \theta_0^2)/2)}{\theta_1 - \theta_0}$ is a function of θ_1 (θ_0 is known!). However, using $\Pr(\mathbf{X} \in C(\theta_1) \mid \theta_0) = \alpha$, the value of k_1 is, apparently, obtained from $1 - \Phi(\sqrt{n}[\frac{k_1}{n} - \theta_0]) = \alpha$ which shows that k_1, and hence the critical region $C(\theta_1)$, is independent of θ_1.

The critical region $C = \{\mathbf{x}\colon \sum_{i=1}^n x_i \geq k_1\}$ therefore provides a UMP test of size α for the simple null hypothesis $\mathrm{H}_0(\theta = \theta_0)$ against the composite alternative hypothesis $\mathrm{H}_1(\theta > \theta_0)$.

We may also read-off, from Example 8(i), the power function

$$\beta(\theta) = \Pr(\mathbf{X} \in C|\theta) = 1 - \Phi\big(\sqrt{n}\,\big(\tfrac{k_1}{n} - \theta\big)\big) = 1 - \Phi(z - \sqrt{n}(\theta - \theta_0))$$

where $z = \sqrt{n}(\frac{k_1}{n} - \theta_0)$ is fixed by $\Phi(z) = 1 - \alpha$.

Since $\Phi(x)$ increases, as x increases we see that $\Phi(z - \sqrt{n}(\theta - \theta_0)) \leq \Phi(z)$ for $\theta > \theta_0$, and, therefore $\beta(\theta) \geq \alpha$ for all $\theta \in \Theta$. Further, since $\Phi(x) \to 0$ as x decreases to $-\infty$, we note that:

$$\beta(\theta_1) \to 1 \text{ for each } \theta_1 > \theta_0 \text{ as } n \to +\infty; \text{ and } \beta(\theta_1) \to 1 \text{ as } \theta_1 \text{ increases to } +\infty,$$

confirming that the test is unbiased, and showing that it has large power if either the sample size is large, or if θ_1 is much greater than θ_0.

Similarly, the critical region $\{\mathbf{x}\colon \sum_{i=1}^n x_i \leq k_1\}$ will provide a UMP test for the null hypothesis $\mathrm{H}_0(\theta = \theta_0)$ against the alternative hypothesis $\mathrm{H}_1(\theta < \theta_0)$.

11. Remarks. **(i)** The procedure that is successful in Example 10, has, however, a limited range of application. For example, suppose the sample variables $X_1, X_2, ..., X_n$ are independent with the normal distribution $\mathrm{N}(\theta, 1)$, and we wish to test the null hypothesis $\mathrm{H}_0(\theta = \theta_0)$ against the alternative hypothesis $\mathrm{H}_1(\theta \neq \theta_0)$.

Following the method of Example 10, we have for all $\theta_1 > \theta_0$ the common critical region C_1 with shape $\{\mathbf{x}\colon \sum_{i=1}^{n} x_i \geq b\}$. Similarly for all $\theta_1 < \theta_0$, we obtain the common critical region C_2 with shape $\{\mathbf{x}\colon \sum_{i=1}^{n} x_i \leq a\}$. It follows that there is no UMP critical region for these hypotheses H_0 and H_1. For, if C is a UMP critical region, we must have

$\Pr(\mathbf{X} \in C \mid \theta_1) \geq \Pr(\mathbf{X} \in C_1 \mid \theta_1)$ for $\theta_1 > \theta_0$, and,

$\Pr(\mathbf{X} \in C \mid \theta_1) \geq \Pr(\mathbf{X} \in C_2 \mid \theta_1)$ for $\theta_1 < \theta_0$.

However, we have shown that C_1 and C_2 are each UMP for the alternatives $\theta > \theta_0$ and $\theta < \theta_0$ respectively, and this implies the inequalities are reversed. These contradictions show that a UMP critical region cannot exist for $\mathrm{H}_0(\theta = \theta_0)$ against $\mathrm{H}_1(\theta \neq \theta_0)$.

(ii) In M130 the test statistic $\overline{X}$ and the one-dimensional critical region $R = \{x : |x - \theta_0| \geq c\}$ was used for the two-tail test of $\mathrm{H}_0(\theta = \theta_0)$ against $\mathrm{H}_1(\theta \neq \theta_0)$. Translating R into the n-dimensional framework we obtain the critical region $C = \{\mathbf{x}\colon |\overline{x} - \theta_0| \geq c\}$, which is the symmetric combination of $C_1 = \{\mathbf{x}\colon \overline{x} - \theta_0 \geq b\}$ and $C_2 = \{\mathbf{x}\colon \overline{x} - \theta_0 \leq a\}$ with $b = -a = c$. This symmetric choice is necessary if we wish to retain the property that the critical region is unbiased.

12. Remark. The examples in 11 provide contrasting views on the theory of hypothesis testing. Example 11(i) shows that the criterion of seeking a uniformly most powerful test is too demanding, while Example 11(ii) indicates that the property of unbiasedness holds for the equi-tail region. It can be shown that the equi-tail region above is a **uniformly most powerful unbiased** (**UMPU**) test - uniformly more powerful than any other unbiased test at the same level of significance.

13. Composite hypotheses. We close this section by showing how the Neyman-Pearson lemma may be applied to obtain a UMP test for certain types of composite null and alternative hypotheses.

Suppose the decomposition $\Theta = \Theta_0 \cup \Theta_1$ of the set of admissible values of the parameter produces a composite hypotheses $\mathrm{H}_0(\theta \in \Theta_0)$ and $\mathrm{H}_1(\theta \in \Theta_1)$.

For sets Θ_0 of the type $\{\theta : \theta \leq \theta_0\}$ or $\{\theta : \theta \geq \theta_0\}$, a UMP test C of $\mathrm{H}_0(\theta \in \Theta_0)$ against $\mathrm{H}_1(\theta \in \Theta_1)$ with $\Pr(\mathrm{X} \in \mathrm{C} \mid \theta_0) = \alpha$ may sometimes be found by the method described in 10.

Example Suppose the sample random variables $X_1, X_2, ..., X_n$ are independent with the common p.d.f. $f(x \mid \theta) = \theta \exp(-\theta x)$ for $x \geq 0$ and $\theta > 0$, and that the null and alternative hypotheses are $\mathrm{H}_0(\theta \leq 1)$ and $\mathrm{H}_1(\theta \geq 2)$ respectively. Choosing any values $\theta_0 \leq 1$ and $\theta_1 \geq 2$, the inequality of the fundamental lemma applied to the null hypothesis $\mathrm{H}_0(\theta = \theta_0)$ against the alternative hypothesis $\mathrm{H}_1(\theta = \theta_1)$ gives, with $t = \sum_{i=1}^{n} x_i$ - see Example 8(iii) -

$$f(\mathbf{x}|\theta_1) = \theta_1^n e^{-\theta_1 t} \geq k\theta_0^n e^{-\theta_0 t} = kf(\mathbf{x}|\theta_0)$$

producing, since $\theta_1 - \theta_0 > 0$, a critical region $C = \{\mathbf{x} : \sum_{i=1}^{n} x_i \leq k_1\}$.

Now we require, for the test to be at level α, $\sup\{\Pr(\sum_{i=1}^{n} X_i \leq k_1|\theta) : \theta \leq 1\} = \alpha$.

Writing $W = 2\theta \sum_{i=1}^{n} X_i$, we observe that W has the chi-squared distribution $\chi^2(2n)$ [simple rescaling of the variables in 8(iii)] and that we require $\sup\{\Pr(W \leq 2k_1\theta) : \theta \leq 1\} = \alpha$. Now, since

$$\Pr(W \le 2k_1\theta) \le \Pr(W \le 2k_1) \text{ for } \theta \le 1,$$

we obtain the required bound by choosing k_1 to satisfy $\Pr(W \le 2k_1) = \alpha$.

Now, it is not difficult to show that $C = \{\mathbf{x} : \sum_{i=1}^n x_i \le k_1\}$ is a UMP test for $\mathrm{H}_0(\theta \le 1)$ against $\mathrm{H}_1(\theta \ge 2)$.

14. Remark. Similarly, it may be shown that:

(i) when the sample variables $X_1, X_2, ..., X_n$ are independent with the normal distribution $\mathrm{N}(\theta, 1)$ the region $\{\mathbf{x} : \overline{x} \ge c\}$ is UMP for the hypotheses $\mathrm{H}_0(\theta \le \theta_0)$ and $\mathrm{H}_1(\theta \ge \theta_1)$ where $\theta_1 \ge \theta_0$, and region $\{\mathbf{x} : \overline{x} \le c\}$ is UMP for the hypotheses $\mathrm{H}_0(\theta \ge \theta_0)$ and $\mathrm{H}_1(\theta \le \theta_1)$ where $\theta_1 \le \theta_0$.

(ii) when the sample variables $X_1, X_2, ..., X_n$ are independent with the binomial distribution $\mathrm{B}(1, \theta)$ the region $\{\mathbf{x} : \overline{x} \ge c\}$ is UMP for the hypotheses $\mathrm{H}_0(\theta \le \theta_0)$ and $\mathrm{H}_1(\theta \ge \theta_1)$ where $0 < \theta_0 \le \theta_1 \le 1$.

3. The Likelihood Ratio Test

1. Notation. The notation of the previous section continues to apply: the sample random variables $X_1, X_2, ..., X_n$ have the joint p.f./p.d.f. $f(\mathbf{x}|\theta)$; the set Θ of admissible values of the parameter θ is decomposed as a disjoint union $\Theta_0 \sqcup \Theta_1$; and, the null and alternative hypotheses are $\mathrm{H}_0(\theta \in \Theta_0)$ and $\mathrm{H}_1(\theta \in \Theta_1)$ respectively.

2. Remarks. When both the null and alternative hypotheses are simple, the search for a best test has been dealt with, in an unequivocal way, by the fundamental lemma. Further, the examples in section 2 show how the idea of the lemma may be extended, in suitable cases, to provide best tests for some composite hypothesis situations. The likelihood ratio test (LRT) is a general method for defining a critical region should the hypotheses be such that the considerations of the last section fail to provide a best test. Although intuitively appealing, as a combination of the idea of the fundamental lemma and the principle of maximum likelihood, the critical region defined by the LRT is not supported by any general theorem outlining desirable properties of a test. However, as our examples will show, the LRT does produce what are regarded as natural tests in the familiar situations.

3. The likelihood ratio test (LRT). To introduce the formal definition of the LRT recall that:

(a) the fundamental lemma rejects $\theta = \theta_0$ in favour of $\theta = \theta_1$ if the ratio $f(\mathbf{x}|\theta_0) \,/\, f(\mathbf{x}|\theta_1)$ is too small; and

(b) regarding the values $x_1, x_2, ..., x_n$ as fixed, the principle of maximum likelihood estimates θ by maximising $f(\mathbf{x}|\theta)$.

Combining these notions, the LRT procedure is to compute first, regarding the values $x_1, x_2, ..., x_n$ as fixed, the maxima:

(i) $\mathrm{M} = \mathrm{M}(\mathbf{x}) = \sup\{f(\mathbf{x}|\theta) : \theta \in \Theta\}$ - this being the **general model maximum**;

(ii) $\mathrm{M}_0 = \mathrm{M}_0(\mathbf{x}) = \sup\{f(\mathbf{x}|\theta) : \theta \in \Theta_0\}$ - this being the **restricted model maximum**, restricted in the sense that $\mathrm{H}_0(\theta \in \Theta_0)$ is assumed to be true.

Then, define $\Lambda(\mathbf{x}) = \mathrm{M}_0(\mathbf{x})/\mathrm{M}(\mathbf{x})$ and the critical region C as having the shape

$$C = \{\mathbf{x} : \Lambda(\mathbf{x}) \le \lambda\}.$$

The value λ (necessarily ≤ 1) is determined, via $\sup_{\theta \in \Theta_0} \Pr(\Lambda(\mathbf{X}) \le \lambda | \theta) = \alpha$.

Note that the intuitive base for the LRT is that, when H_0 is false, the ratio $\Lambda(\mathbf{x})$ is likely to be small.

4. Remarks. (i) It is not difficult to see that the LRT procedure is equivalent to the fundamental lemma when $\Theta = \{\theta_0, \theta_1\}$ and $\Theta_0 = \{\theta_0\}$ - since, in this case,

$$\sup\ \{f(\mathbf{x}|\theta) : \theta \in \Theta\} = f(\mathbf{x}|\theta_0) \text{ or } f(\mathbf{x}|\theta_1) \ \text{ and } \ \sup\ \{f(\mathbf{x}|\theta) : \theta \in \Theta_0\} = f(\mathbf{x}\ |\theta_0),$$

so that $\Lambda(\mathbf{x}) \le \lambda \le 1$ implies that $f(\mathbf{x}|\theta_0) \le \lambda f(\mathbf{x}|\theta_1)$.

(ii) Alternative descriptions of the LRT are:

(a) to express the critical region in terms of $\mathrm{M}(\mathbf{x})/\mathrm{M}_0(\mathbf{x})$ being 'large' - a trivial inversion;

(b) to use the ratio $\mathrm{M}_1(\mathbf{x})/\mathrm{M}_0(\mathbf{x})$, where $\mathrm{M}_1(\mathbf{x}) = \sup\{f(\mathbf{x}|\theta) : \theta \in \Theta_1\}$, with 'large' values suggesting that H_0 is false.

(iii) Assuming $f(\mathbf{x}|\theta)$ is continuous as a function of θ, the maximum of $f(\mathbf{x}|\theta)$ for $\theta \in \Theta$ is just the value $f(\mathbf{x}|\widehat{\theta})$, where $\widehat{\theta}$ is the maximum likelihood estimate of θ. Similarly maximizing $f(\mathbf{x}|\theta)$ over $\theta \in \Theta_0$ will give the value $f(\mathbf{x}|\widehat{\theta}_0)$ where $\widehat{\theta}_0$ is the restricted (by $\theta \in \Theta_0$) maximum likelihood estimate of θ.

5. Examples *Testing the normal mean.* These examples illustrate the application of the LRT to the parameters of a normal distribution. Suppose $X_1, X_2, ..., X_n$ is a random sample from the distribution $\mathrm{N}(\mu, \sigma^2)$, so that the joint p.d.f. is

$$f(\mathbf{x}|\mu\ ,\ \sigma) = \tfrac{1}{(\sigma\sqrt{2\pi})^n} \exp(-\tfrac{1}{2\sigma^2} \textstyle\sum_{i=1}^n (x_i - \mu)^2).$$

We test the null hypothesis $\mathrm{H}_0(\mu = \mu_0)$ against the alternative hypothesis $\mathrm{H}_1(\mu \ne \mu_0)$ in the two cases, σ known and σ unknown.

(i) $\boldsymbol{\sigma} = \boldsymbol{\sigma}_0$ **known**. Here, $\Theta = \{\theta = (\mu, \sigma_0) : -\infty < \mu < \infty\}$ and $\Theta_0 = \{(\mu_0, \sigma_0)\}$. The numerator in $\Lambda(\mathbf{x})$ requires no maximizing and is $\mathrm{M}_0(\mathbf{x}) = f(\mathbf{x}\ |\mu_0,\ \sigma_0)$. Since $\widehat{\mu} = \overline{x}$ is the MLE of μ (see Section 2.2), the denominator is $f(\mathbf{x}|\mu,\ \sigma_0)$ − that is

$$\mathrm{M}(\mathbf{x}) = \sup\ \{f(\mathbf{x}|\mu,\ \sigma_0) : -\infty < \mu < \infty\} = \tfrac{1}{(\sigma_0\sqrt{2\pi})^n} \exp\left[-\tfrac{1}{2\sigma_0^2} \textstyle\sum_{i=1}^n (x_i - \overline{x})^2\right].$$

Thus,

$$\begin{aligned} \Lambda(\mathbf{x}) &= \frac{f(\mathbf{x}\ |\mu_0, \sigma_0)}{\mathrm{M}(\mathbf{x})} = \frac{\frac{1}{(\sigma_0\sqrt{2\pi})^n} \exp(-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \mu_0)^2)}{\frac{1}{(\sigma_0\sqrt{2\pi})^n} \exp\left[-\frac{1}{2\sigma_0^2} \sum_{i=1}^n (x_i - \overline{x})^2\right]} \\ &= \exp[-\frac{1}{2\sigma_0^2}\{\sum_{i=1}^n (x_i - \mu_0)^2 - \sum_{i=1}^n (x_i - \overline{x})^2\}]. \end{aligned}$$

Hence, $\Lambda(\mathbf{x}) \le \lambda$ is equivalent to

$$n(\overline{x}-\mu_0)^2 = \sum_{i=1}^n (x_i-\mu_0)^2 - \sum_{i=1}^n (x_i-\overline{x})^2 \geq \lambda_1 \qquad \text{or} \qquad \sqrt{n}|\overline{x}-\mu_0|/\sigma_0 \geq \lambda_2,$$

which reduces to the one-dimensional critical region $R = \{x : |x-\mu_0| \geq c\}$ and test statistic $\overline{X}$ with $\sqrt{n}|\overline{X}-\mu_0|/\sigma_0$ being distributed $\mathrm{N}(0,1)$.

(ii) $\boldsymbol{\sigma}$ unknown. Now we must write $\theta = (\mu, \sigma)$, showing both μ and σ, and put

$$\Theta = \{(\mu,\sigma) : -\infty < \mu < \infty,\, \sigma > 0\} \text{ and } \Theta_0 = \{(\mu_0, \sigma) : \sigma > 0\},$$

thereby expressing the null hypothesis in the form $\mathrm{H}_0(\theta \in \Theta_0)$.

From Section 2.2, the MLE of (μ,σ) is $(\hat{\mu},\hat{\sigma})$, where $\widehat{\mu} = \overline{x}$ and $\widehat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n (x_i-\overline{x})^2$. So, the denominator in $\Lambda(\mathbf{x})$ is $\mathrm{M}(\mathbf{x}) = f(\mathbf{x}\,|\widehat{\mu},\,\widehat{\sigma}) = \frac{1}{(\widehat{\sigma}\sqrt{2\pi})^n}\exp(-\frac{n}{2})$.

Similarly, with $\mu = \mu_0$ fixed, we maximize $f(\mathbf{x}|\mu,\sigma)$ over σ to obtain $\mathrm{M}_0(\mathbf{x}) = f(\mathbf{x}|\mu_0, \widehat{\sigma}_0)$, where $\widehat{\sigma}_0^2 = \frac{1}{n}\sum_{i=1}^n (x_i-\mu_0)^2$. Thus, $\mathrm{M}_0(\mathbf{x}) = \frac{1}{(\widehat{\sigma}_0\sqrt{2\pi})^n}\exp(-\frac{n}{2})$ and

$$\Lambda(\mathbf{x}) = \left(\frac{\widehat{\sigma}^2}{\widehat{\sigma}_0^2}\right)^{\frac{n}{2}} = \left(\frac{\sum_{i=1}^n (x_i-\overline{x})^2}{\sum_{i=1}^n (x_i-\mu_0)^2}\right)^{\frac{n}{2}} \leq \lambda$$

is equivalent to $\sqrt{n}|\overline{x}-\mu_0|/s \geq \lambda_1$ where $s^2 = \frac{1}{n-1}\sum_{i=1}^n (x_i-\overline{x})^2$.

(Note that $(\frac{a^2}{b^2})^{n/2} \leq \lambda$ is equivalent to $\sqrt{\frac{b^2-a^2}{a^2}} \geq \sqrt{\lambda^{-2/n}-1}$ and now make the obvious substitutions, using $n(\overline{x}-\mu_0)^2 = \sum_{i=1}^n (x_i-\mu_0)^2 - \sum_{i=1}^n (x_i-\overline{x})^2$.)

Thus the LRT procedure reduced to the one-dimensional critical region and test statistic, recovers the test structure used in the basic t-test.