Tests and Their Power

Ling Kiong Doong
Department of Mathematics
National University of Singapore

1. Introduction

In Statistical Inference, the two main areas of study are estimation and testing of hypotheses. In this workshop, we confine ourselves to the area of testing of hypotheses.

Basic concepts needed in testing of hypotheses will be introduced in the next section. Section 3 considers the problems of testing a simple null hypothesis against a simple alternative hypothesis. The well-known Neyman Pearson Lemma will also be discussed. Applications of the Neyman Pearson Lemma to some test problems which involve composite hypotheses are given in Section 4. Finally, in Section 5, we introduce the likelihood ratio method of finding a test.

2. Basic Concepts

A standard form of problems in testing of hypotheses is

Test a null hypothesis, usually denoted by H_0, against an alternative hypothesis, usually denoted by H_1,

where H_0 and H_1 are hypotheses concerning the distribution of a random variable X.

A hypothesis is said to be *simple* if it completely determines the distribution of X. Hypotheses which are not simple are called *composite* hypotheses. For example, if X has a normal distribution with mean μ and variable σ^2, i.e., $X \sim N(\mu, \sigma^2)$, then the hypothesis that $\mu = 0$ and $\sigma^2 = 1$ is simple, while the hypothesis that $\mu = 5$ is composite.

Lecture given at the Workshop on Hypothesis Testing and Non-Parametric Statistics for school teachers, 7 September 1987

In many practical situation, the *probability density function* (pdf) of a random variable X is partly known in form, but involves some unknown *characteristics* or *parameters.* A hypothesis which is concerned with the parameters of the distribution of X is referred to as a *parametric hypothesis.* Hypotheses that are not parametric are said to be *non-parametric.* For example, the two hypotheses given in the preceding paragraph are parametric. The hypothesis that X is normally distributed and the hypothesis that X has a distribution which is symmetric about the origin are examples of non-parametric hypotheses.

In this talk we shall deal with only those test problems which involve parametric hypotheses.

Now let $\mathbf{X} = (X_1, X_2, \cdots, X_n)$ be a random sample of size n drawn on a random variable X whose distribution is indexed by parameter(s) $\theta \in \Omega$, the *parameter space.* Suppose that we wish to test the null hypothesis $H_0 : \theta \in \omega_0$ against the alternative hypothesis $H_1 : \theta \in \omega_1 = \Omega \setminus \omega_0$. That is, when we are given sample observations $\mathbf{x} = (x_1, x_2, \cdots, x_n)$, we want to take one of the following two actions :

(i) to accept H_0, or equivalently to reject H_1; or

(ii) to accept H_1, or equivalently to reject H_0.

Let us assume that a certain test procedure or rule is used so that to each sample point $\mathbf{x}$ in $\mathcal{X}_{\mathbf{X}}$ the sample space of $\mathbf{X}$ which is usually taken to be the n-dimensional Euclidean space R^n, one and only one of the above two actions is taken. Denote by C the set of those sample points which leads to a rejection of H_0 and $A = \mathcal{X}_{\mathbf{X}} \setminus C$. The regions C and A are called the *critical* (or *rejection*) and the *acceptance* region of the test precedure respectively. Therefore finding a test procedure is equivalent to partitioning $\mathcal{X}_{\mathbf{X}}$ into two non-overlapping regions C and A.

Once a decision is made, then we may commit either a *Type I error* or a *Type II* error which are best illustrated by the following table.

Actual State \ Decision	Accept H_0	Reject H_0
H_0 is true	a correct decision	Type I Error
H_1 is true	Type II Error	a correct decision

Let us consider now the case where both H_0 and H_1 are simple. Define :

$$\alpha = P(\text{Committing a Type I error}),$$

and

$$\beta^* = P(\text{Committing a Type II error}).$$

Thus α is the probability of wrongly rejecting H_0 and β^* is the probability of wrongly accepting H_0. Ideally, it is natural to seek a test procedure which will give the values of α and β^* as small as possible. Unfortunately that is not possible in most cases. Thus, instead, we look for a test procedure which has a given α and a small β^* or equivalently a large $\beta = 1 - \beta^*$. (β gives the probability of correctly rejecting H_0.) In the literature, α is called the ***size*** or the ***significance level*** and β is called the ***power*** of the test procedure. The above consideration leads to the following definitions of a test.

Definition 1 : *A non-randomized test is a function*

$$\phi : \mathcal{X}_{\mathbf{X}} \longrightarrow \{0,1\}$$

such that

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in C \\ 0 & \text{if } \mathbf{x} \in A. \end{cases}$$

Definition 2 : *A randomized test is a function*

$$\phi : \mathcal{X}_{\mathbf{X}} \longrightarrow [0,1].$$

For a given randomized test ϕ, if $0 < \phi(\mathbf{x}) = p < 1$, then the probability of rejecting H_0 is p. That is, if $\mathbf{x}_0$ is a sample point with $\phi(\mathbf{x}_0) = p$, then we flip a biased coin with $P(\text{head}) = p$. If a head appears then we reject H_0 and accept H_0 if a tail appears. (Sometimes, we consult a table of random digits instead of flipping a biased coin.)

Also we note that

$$\alpha = E(\phi(\mathbf{X}) \mid H_0) \quad \text{and} \quad \beta = E(\phi(\mathbf{X}) \mid H_1).$$

When H_0 is composite, we define

$$\alpha = \sup_{\theta \in \omega_0} (E(\phi(\mathbf{X}) \mid \theta))$$

and when H_1 is composite we define

$$\beta(\theta) = E(\phi(\mathbf{X}) \mid \theta), \quad \text{for } \theta \in \omega_1 .$$

For composite H_0, α is the *size* of the test ϕ and for composite H_1, $\beta(\theta)$ is called the *power function* of the test ϕ. The functional value $\beta(\theta_1)$ at a point $\theta_1 \in \omega_1$ is called the *power* of the test at $\theta = \theta_1$.

3. Neyman-Pearson Lemma

This section deals with a test problem whose null hypothesis and alternative hypothesis are simple. Let us begin with a definition.

Definition 3 : *A test ϕ is said to be a most powerful (MP) test of size α in testing a simple H_0 against a simple H_1 if*

(i) $E(\phi(\mathbf{X}) \mid H_0) = \alpha$; *and*

(ii) $E(\phi(\mathbf{X}) \mid H_1) \geq E(\phi^*(\mathbf{X}) \mid H_1)$ *for any other test ϕ^* satisfying* (i).

To every test ϕ, we can associate with it a pair (α,β) where α and β are respectively the size and the power of ϕ. Now let N be the set of points (α, β) such that there exists a test whose associated pair is (α, β). Then it is not difficult to see that

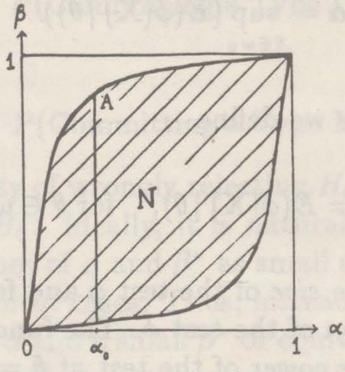
(i) $(0,0) \in N$ and $(1,1) \in N$;

(ii) $(\alpha_1,\beta_1) \in N$ and $(\alpha_2,\beta_2) \in N$, and $0 \leq t \leq 1$, then $t(\alpha_1,\beta_1) + (1-t)(\alpha_2,\beta_2) \in N$, (i.e., N is convex.);

(iii) $(\alpha,\beta) \in N$ implies $(1-\alpha, 1-\beta) \in N$. (i.e., N is symmetric with respect to the point $(\frac{1}{2}, \frac{1}{2})$.); and

(iv) N is closed.

Graphically, N has the shape of the following diagram.



The point A (in the preceding diagram) corresponds to a MP test of size α_0. From the diagram it is also clear that if ϕ is a MP test of size α and power β, then $\alpha < \beta$.

For a given significance level α, our main objective is to determine a MP test of size α. The solution to this problem is given by the following

Neyman-Pearson Lemma (see page 118 of Zack [5].) *Let f_0 and f_1 denote the pdf of* $\mathbf{X}$ *under H_0 and H_1 respectively. Then for testing H_0 against H_1 we have the following:*

(a) *Any test of the form*

$$\phi(\mathbf{X}) = \begin{cases} 1 & \text{if } f_1(\mathbf{x}) > k f_0(\mathbf{x}) \\ \gamma & \text{if } f_1(\mathbf{x}) = k f_0(\mathbf{x}) \\ 0 & \text{if } f_1(\mathbf{x}) < k f_0(\mathbf{x}). \end{cases}$$

for some $0 \le k < \infty$ and $0 \le \gamma \le 1$ is MP relative to all tests of its sizes.

(b) (Existence) *For testing H_0 against H_1 at level of significance α, there exists $k_\alpha, 0 \le k_\alpha < \infty$ and $\gamma_\alpha, 0 \le \gamma_\alpha \le 1$ such that the corresponding test of the form in* (a) *is MP of size α.*

(c) (Uniqueness) *If a test ϕ^* is MP of size α, then it is of the form in* (a), *except perhaps on the set $\{\mathbf{x} \mid f_1(\mathbf{x}) = k f_0(\mathbf{x})\}$; unless there exists a test of size smaller than α and power 1.*

For a proof, the readers are referred to pages 119-121 of Zack [5]. When X is absolutely continuous, the MP test of size α is a non-randomized one. However, when X is discrete, the MP test of size α is a randomized test for most values of α. The Neyman Pearson Lemma is useful not only for testing simple H_0 against simple H_1, it is also very helpful in finding a uniformly most powerful (UMP) test (a term to be defined in the next section) for testing a simple H_0 against a one-sided composite H_1. Many examples of finding MP tests using the Neyman Pearson Lemma may be found in many books on Elementary Mathematical Statistics. To end this section, we give the following

Example 1 : Let the pdf of a random variable X be given by

$$f(x \mid \theta) = \frac{1}{\theta} exp\{-x/\theta\}, \quad x \geq 0,$$

where $\theta > 0$, is the mean of X. On the basis of a random sample $\mathbf{X}$, obtain the MP test for testing $H_0 : \theta = 1$ against $H_1 : \theta = 2$ at level of significance $\alpha, (0 < \alpha < 1)$.

Solution : Here

$$f_0(\mathbf{x}) = exp\{-\Sigma x_i\}, \quad x_i \geq 0, i = 1, 2, \cdots, n;$$

and

$$f_1(\mathbf{x}) = \left(\frac{1}{2}\right)^n exp\{-\Sigma x_i/2\}, \quad x_i \geq 0, i = 1, 2, \cdots, n.$$

Hence, by applying the Neyman Pearson Lemma, the MP test ϕ of size α is given by

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } (f_1(\mathbf{x})/f_0(\mathbf{x})) \geq k \\ 0 & \text{if } (f_1(\mathbf{x})/f_0(\mathbf{x})) < k. \end{cases}$$

which is equivalent to

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } \sum_{i=1}^n x_i \geq k^* \\ 0 & \text{if } \sum_{i=1}^n x_i < k^*, \end{cases}$$

where k^* is a constant to be determined so that

$$E(\phi(\mathbf{X}) \mid \theta = 1) = P(\sum_{i=1}^{n} X_i \geq k^*) = \alpha.$$

Now we note when $\theta = 1$, then $2\sum_{i=1}^{n} X_i$ has chi-square distribution with $2n$ degrees of freedom. Therefore $k^* = \frac{1}{2}\chi^2_\alpha(2n)$, a value which can be obtained from Statistical Tables [2]. For example when $\alpha = 0.05, n = 5$, we have $\chi^2_{0.05}(10) = 18.307$.

4. Uniformly Most Powerful Tests

When the alternative hypothesis H_1 is composite, the notion of a MP test does not apply. Instead, we introduce the concept of a uniformly most powerful (UMP) test.

Definition 4: *A test ϕ is called a UMP test of size α if*

(i) $\sup_{\theta \in \omega_0} E(\phi(\mathbf{X}) \mid \theta) = \alpha$; *and*

(ii) *for each $\theta \in \omega_1, E(\phi(\mathbf{X}) \mid \theta) \geq E(\phi^*(\mathbf{X}) \mid \theta)$ for any test ϕ^* satisfying* (i).

When H_0 is simple and H_1 is composite, we can regard H_1 as a union of simple hypotheses. In some cases we can obtain a UMP test by applying the Neyman Pearson Lemma. To do this, we consider the problem of testing the simple H_0 against a typical simple $H_1^* : \theta = \theta_1$ where θ_1 is a typical member of ω_1. If the MP test of size α for testing H_0 against H_1^* does not depend on the value of θ_1 in ω_1, then this test is MP for testing H_0 against every simple H_1' in H_1. Hence this MP test of size α becomes a UMP test of size α for testing H_0 against H_1.

Example 2 : Let X be the random variable considered in Example 1. Here, we wish to establish a UMP test of size α for testing

$$H_0 : \theta = 1 \quad \text{against} \quad H_1 : \theta > 1.$$

Solution : Note that $\omega_0 = \{1\}$ and $\omega_1 = \{\theta \mid \theta > 1\}$. Let θ_1 be a fixed element in ω_1. If we consider the problem of testing

$$H_0 \; : \; \theta = 1 \quad \text{against} \quad H_1^* \; : \; \theta = \theta_1,$$

(here H_1^* is simple), then we can apply the Neyman Pearson Lemma. It leads to the following MP test of size α

$$\phi(\mathbf{x}) = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} x_i \geq \frac{1}{2}\chi^2_\alpha(2n) \\ 0 & \text{if } \sum_{i=1}^{n} x_i < \frac{1}{2}\chi^2_\alpha(2n). \end{cases}$$

(See Example 1 for the derivation.) This test ϕ is independent of the value of θ_1 and hence it is the UMP test for testing H_0 against H_1.

Remark :

(1) If H_1 in Example 2 is replaced by $H_2 \ : \theta \in \omega_2 = \{\theta \mid \theta > \theta_2\}$, where θ_2 is a given real number > 1, then the UMP test of size α for testing H_0 against H_2 is the one given in Example 2.

(2) If H_1 is replaced by $H_3 : \theta \in \omega_3 = \{\theta \mid \theta < \theta_3\}$, where θ_3 is a given positive number < 1, then the UMP test of size α for testing H_0 against H_3 is

$$\phi(x) = \begin{cases} 1 & \text{if } \sum_{i=1}^{n} x_i \leq \frac{1}{2}\chi^2_{1-\alpha}(2n) \\ 0 & \text{if } \sum_{i=1}^{n} x_i > \frac{1}{2}\chi^2_{1-\alpha}(2n). \end{cases}$$

(3) When H_1 is replaced by $H_4 : \theta \neq 1$ (H_4 is a two-sided hypothesis), then a UMP test of size α for testing H_0 against H_4 does not exist.

Consider now the case in which both $H_0 : \theta \in \omega_0$ and $H_1 : \theta \in \omega_1$ are one-sided composite hypotheses. Assume that for any $\theta_0 \in \omega_0$ and any $\theta_1 \in \omega_1$ we have $\theta_0 < \theta_1$. Let $\theta_0^* = \sup_{\theta \in \omega_0} \theta$. (In the case $\theta_0 > \theta_1$, for every $\theta_0 \in \omega_0$ and every $\theta_1 \in \omega_1$, we take $\theta_0^* = \inf_{\theta \in \omega_0} \theta$.) If ϕ is the UMP test of size α for testing $H_0^* : \theta \in \theta_0^*$ against H_1, and if

$$\sup_{\theta \in \omega_0} E(\phi(\mathbf{X}) \mid \theta) = \alpha$$

then ϕ is the UMP test of size α for testing H_0 against H_1.

For additional reading material concerning UMP tests, readers are referred to the books by Roussas [3] and Zacks [5].

5. Likelihood Ratio Method

The Neyman Pearson Lemma provides us a method of finding the MP test for testing a simple H_0 against a simple H_1. In some cases an application of the Lemma will also lead us to obtain the UMP tests. In order to solve test problems which do not admit UMP tests, we introduce in this section the likelihood ratio method of constructing tests. The method uses a test statistic which is analogous to likelihood ratio $f_1(\mathbf{x})/f_0(\mathbf{x})$ used in the Neyman Pearson Lemma. For this reason, some authors refer to this method as the generalized likelihood ratio method. When both null and alternative hypotheses are simple, this method of constructing tests should not be applied, for the resulting test obtained may not be MP. Instead, we should apply the Neyman Pearson Lemma.

Consider the problem of testing $H_0 : \theta \in \omega_0$ against $H_1 : \theta \in \omega_1$. Suppose that $\Omega = \omega_0 \cup \omega_1$ contains three or more elements, i.e., at least one of H_0 and H_1 is composite. The likelihood ratio method uses the test statistic, $\lambda(\mathbf{X})$, usually known as the ***likelihood ratio statistic***, which is defined by

$$\lambda(\mathbf{x}) = \frac{\sup_{\theta \in \omega_0} f(\mathbf{x} \mid \theta)}{\sup_{\theta \in \Omega} f(\mathbf{x} \mid \theta)}.$$

Since $\omega_0 \subset \Omega$, it is clear that $0 \leq \lambda(\mathbf{x}) \leq 1$ for every sample point $\mathbf{x} \in \mathcal{X}_{\mathbf{X}}$. Note that

$$\sup_{\theta \in \omega_0} f(\mathbf{x} \mid \theta) = f(\mathbf{x} \mid \hat{\theta}_\omega)$$

and

$$\sup_{\theta \in \Omega} f(\mathbf{x} \mid \theta) = f(\mathbf{x} \mid \hat{\theta})$$

where $\hat{\theta}_\omega$ and $\hat{\theta}$ are the maximum likelihood estimates of θ under ω_0 and Ω respectively. When H_0 is not true, the denominator, $f(\mathbf{x} \mid \hat{\theta})$, in the definition of $\lambda(\mathbf{x})$ tends to be much larger than the numerator, $f(\mathbf{x} \mid \hat{\theta}_\omega)$, and so the value of $\lambda(\mathbf{x})$ tends to be small. Thus the method suggests the following test:

$$\phi(\mathbf{x}) = \begin{cases} 1, & \text{if } \lambda(\mathbf{x}) \leq \lambda_0 \\ 0, & \text{if } \lambda(\mathbf{x}) > \lambda_0 \end{cases}$$

where λ_0, $0 < \lambda_0 < 1$, is a constant to be determined so that

$$\sup_{\theta \in \omega_0} E(\phi(\mathbf{X}) \mid \theta) = \alpha.$$

This resultant test ϕ is called ***the likelihood ratio test***. The value of λ_0 depends on the size, α. Generally, for any given α, we require the knowledge of the distribution of the likelihood ratio statistic under H_0 to determine λ_0. In most cases the inequality $\lambda(\mathbf{x}) \leq \lambda_0$ used in defining the likelihood ratio test ϕ may be reduced to a simpler and equivalent inequality, say

$T(\mathbf{x}) \leq T_0$, and the distribution of $T(\mathbf{X})$ may be easier to derive than that of $\lambda(\mathbf{X})$ when H_0 is true.

Likelihood ratio tests are not necessarily UMP. But they have some desirable properties. (We will not discuss them here.) A possible drawback of the likelihood ratio method is, perhaps, that the maximum likelihood estimates $\hat{\theta}_\omega$ and $\hat{\theta}$ may be difficult to obtain. Another drawback is that the distribution of the likelihood ratio statistic $\lambda(\mathbf{X})$ or its equivalent statistic $T(\mathbf{X})$, (needed for the determination of λ_0 or T_0), may not be available in the literature and may be difficult to derive. Fortunately, in most situations, the asymptotic distribution (i.e., when n tends to infinity) of $-2ln\lambda(\mathbf{X})$ under H_0 has been proven to follow a chi-square distribution. (This result will be stated without proof as a theorem.) Therefore for sufficiently large samples, an approximate value of the cut point λ_0 can be easily obtained.

Theorem : *For testing $H_0 : \theta \in \omega_0$ against $H_1 : \theta \in \omega_1$, let the dimension of $\Omega = \omega_0 \cup \omega_1$ be r and the dimension of ω_0 be $r_1 (< r)$. When H_0 is true, then under certain regularity conditions, the asymptotic distribution of $-2ln\lambda(\mathbf{X})$, as n tends to infinity, is distributed like a chi-square distribution with $r - r_1$ degrees of freedom.*

Note : For the regularity conditions about the distribution of the random variable X and a proof to the above theorem, the readers are referred to Wilks [4].

To end the talk we give an example of the determination of the degrees of freedom of the asymptotic distribution of $-2ln\lambda(\mathbf{X})$.

Example 3 : Let X_i, $i = 1, 2, \cdots, k(\geq 2)$ be k independent random variables. The pdf of each X_i is known but depends on the unknown mean μ_i and variance σ_i^2. Consider now the null hypothesis

$$H_0 : \sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$$

and the alternative hypothesis

$$H_1 : \text{not all } \sigma_i^2\text{'s are equal.}$$

Here we have

$$\Omega = \{(\mu_1, \cdots, \mu_k, \sigma_1^2, \cdots, \sigma_k^2) \mid -\infty < \mu_i < +\infty,\ \sigma_i^2 > 0, i = 1, \cdots, k\}$$

and

$$\omega_0 = \{(\mu_1, \cdots, \mu_k, \sigma^2, \cdots, \sigma^2) \mid -\infty < \mu_i < +\infty,\ \sigma^2 > 0, i = 1, \cdots, k\}.$$

Clearly the dimension of Ω is $2k$ (i.e. $r = 2k$) and the dimension of ω_0 is $k+1$, i.e., $r_1 = k+1$. Therefore for large samples sizes, the distribution of $-2ln\lambda(\mathrm{X})$ is approximately distributed as the chi-square distribution with $k-1$ degrees of freedom.

Remark : H_0 in Example 3 may be rewritten as

$$H_0' : \gamma_2 = \gamma_3 = \cdots = \gamma_k = 1,$$

where $\gamma_j = \sigma_j^2 / \sigma_1^2$. Then

$$\Omega' = \{(\mu_1, \cdots, \mu_k, \sigma_1^2, \gamma_2, \cdots, \gamma_k) \mid -\infty < \mu_i < +\infty,$$
$$\sigma_1^2 > 0, \gamma_j > 0, i = 1, \cdots, k; j = 2, \cdots, k\}.$$

Let $\theta = (\mu_1, \cdots, \mu_k, \sigma_1^2, \gamma_2, \cdots, \gamma_k)$. Then to describe H_0', we require $k-1$ components of θ, i.e., $\gamma_2, \cdots, \gamma_k$, to take fixed values. In general, the number of components that take fixed values in describing H_0 is the number of degrees of freedom of the asymptotic distribution of $-2ln\lambda(\mathrm{X})$.

References

[1] E. L. Lehmann, *Testing Statistical Hypotheses*, John Wiley, 1959.

[2] J. Murdoch and J. A. Barnes, *Statistical Tables*, (second edition), Macmillan, 1970.

[3] G. G. Roussas, *A First Course in Mathematical Statistics*, Addison-Wesley, 1973.

[4] S. Wilks, *Mathematical Statistics*, John Wiley, 1962.

[5] S. Zacks, *Parametric Statistical Inference*, Pergamon, 1981.