Theory of Hypothesis Testing: Part I
BIOSTAT 802, Winter 2018

1 Preliminaries

This section is devoted to some basic concepts relevant to the classical Neyman-Pearson school of hypothesis testing.

1.1 Hypothesis and Hypothesis Testing

For a sample (or a random variable or a random vector) X, denoted its sample space by $(\mathcal{X}, \mathcal{B}_x)$ and its distribution family by $\mathcal{P} = \{P_\theta, \theta \in \Theta\}$ with parameter space Θ.

Definition 1 *Let Θ_H be a non-empty subset of Θ. A hypothesis on the distribution of X refers to a statement "*for a certain $\theta \in \Theta_H$ the distribution of X is P_θ.*"*

Generally, one may use a short-handed notation for the hypothesis as "$\theta \in \Theta_H$", even as simple as a letter "H". For the latter case, we refer it to as "hypothesis H". If the distribution of X is P_{θ_0} for a parameter value $\theta_0 \in \Theta_H$, then we say that the hypothsis H is true; otherwise, if $\theta_0 \notin \Theta_H$ then we say that the hypothesis H is false.

Let $\Theta_K = \Theta - \Theta_H = \Theta_H^c$, or $\Theta_H \cup \Theta_K = \Theta$. A statement "$\theta \in \Theta_K$" is referred to as the *alternative hypothesis*. To present a complete statement of hypothesis, most of time we write the following contrast:

$$H : \theta \in \Theta_H \text{ versus (or against or } \leftrightarrow) \ K : \theta \in \Theta_K. \tag{1}$$

Note that Θ_K may be just a subset of Θ. For example, to test $H : \theta = 0$ in a normal distribution $N(\theta, 1)$, and an altherative hypothesis may be $\theta \neq 0$ or $\theta > 0$. In the latter case, we have the following statement:

$$H : \theta = 0 \text{ against } K : \theta > 0.$$

The statement opposite to the alternative hypothesis is called the *null hypothesis*.

Testing hypothsis may be regarded as a two-action problem. That is, for a given hypothesis statement $H \leftrightarrow K$, there are two actions to take: Accept the null hypothesis H (or reject the alternative hypothesis K), denoted by action d_0; or reject the null hypothesis H (or accept the alternative hypothesis K), denoted by action d_1.

A statistical solution to this problem is to establish a rule such that for a sample $x \in \mathcal{X}$ this rule enables to make a decision between d_0 and d_1. Such rule is termed as *test*. Clearly, in this two-action problem we may partition the sample space $\mathcal{X}$ into two parts: $\mathcal{A}$ or $\mathcal{A}^c = \mathcal{X} - \mathcal{A}$, so that when $x \in \mathcal{A}$ the hypothesis H is rejected while $x \in \mathcal{A}^c$ the hypothesis H is not rejected. In other words, $\mathcal{A}$ defintes the region of rejection or the so-called *critical region* or *rejection region*.

From the decision theory point of view, a test is in fact a decision function δ, namely a mapping from the sample space $\mathcal{X}$ to the action space $A = \{d_0, d_1\}$. The rejection region is then given by

$$\mathcal{A} = \{x : \delta(x) = d_1\}.$$

Along with the idea of randomized decision rule, here we may also propose a concept of randomized test. That is, with a given x, one may allocate a probability $\phi(x)$ to reject the hypothesis H, or equivalently a probability $(1 - \phi(x))$ to accept the hypothesis H. This procedure involves both randomness from the data x and from decision-making according to the Bernoulli distribution with probability of rejection, $\phi(x) \in [0, 1]$. That is, $P_\theta(\delta(x) = d_1) = \phi(x)$ and $P_\theta(\delta(x) = d_0) = 1 - \phi(x)$.

Definition 2 *Let $\phi(x)$ be $\mathcal{B}_x$-measurable function satisfying $0 \le \phi(x) \le 1$. Function $\phi(x)$ is said to be a critical function if for a sample x, with probability $\phi(x)$ the null hypothesis H is rejected; or with probability $1 - \phi(x)$ the null hypothesis H is accepted. When $\phi(x)$ is a two-value function, i.e. $\phi(x) \in \{0, 1\}$, it reduces to the case of nonrandomized test. The set of points x for which $\phi(x) = 1$ then defines the region of rejection, namely $\mathcal{A} = \{x : \phi(x) = 1\}$.*

Often, critical function $\phi(x)$ is also simply called as *a test*.

1.2 Two Types of Errors and Power Function

To evaluate the performance of a test, a loss function is needed. Because the action space contains only two elements d_0 and d_1, the resulting loss function may be simply denoted by $L_0(\theta)$ and $L_1(\theta)$, respectively; that is,

$$L_0(\theta) = L(\theta, d_0), \text{ and } L_1(\theta) = L(\theta, d_1), \ \theta \in \Theta.$$

For the null hypothesis $H : \theta \in \Theta_H$, decision d_0 is correct, while for the alternative hypothesis $K : \theta \in \Theta_K$ decision d_1 is correct. Thus, zero loss is assumed in both scenarios:

$$L_0(\theta) = 0, \forall\theta \in \Theta_H; \text{ and } L_1(\theta) = 0, \forall\theta \in \Theta_K.$$

Type I error occurs if decision d_1 is made when the null hypothesis H is true; and *type II error* is committed if decision d_0 is adopted when the alternative hypothesis K is true. In both scenarios,

$$L_0(\theta) > 0, \forall \theta \in \Theta_K; \text{ and } L_1(\theta) > 0, \forall \theta \in \Theta_H.$$

These loss functions L_0 and L_1 clearly do not represent in general the true loss, and there are no definitive guidelines for the specification of loss function. Most of time, for mathematical convenience, in the literature the loss function is simply specified as follows:

$$L_0(\theta) = a > 0, \forall \theta \in \Theta_K; \text{ and } L_1(\theta) = b > 0, \forall \theta \in \Theta_H, \tag{2}$$

where a and b are two suitable positive constants. In this case, the penalty pertains only to making a correct decision or not, and does not carry any weight on the severity of incorrect decision.

Let $\phi(x)$ be a test for a hypothesis testing problem given in (1). We may write

$$L(\theta, \phi(x)) = (1 - \phi(x))L_0(\theta) + \phi(x)L_1(\theta), \theta \in \Theta, x \in \mathcal{X}. \tag{3}$$

With the above loss function (3), it is easy to see that the risk function of $\phi(x)$ is given by

$$R(\theta, \phi) = \begin{cases} bE_\theta\{\phi(X)\}, & \text{if } \theta \in \Theta_H \\ a[1 - E_\theta\{\phi(X)\}], & \text{if } \theta \in \Theta_K. \end{cases}$$

It is ineresting to note that the above risk function is completely determined by a term $E_\theta\{\phi(X)\}$, denoted by

$$\beta_\phi(\theta) = E_\theta\{\phi(X)\} = \int_{\mathcal{X}} \phi(x) dP_\theta(x), \ \theta \in \Theta. \tag{4}$$

Definition 3 *Let $\phi(x)$ be a test for the hypothesis (1). Function $\beta_\phi(\theta)$ defined in (4) is called the power function of test ϕ.*

In the case of ϕ being a nonrandomized test, namely $\phi(x) = 1 \Leftrightarrow \delta(x) = d_1, \phi(x) = 0 \Leftrightarrow \delta(x) = d_0$, for a given rejection region $\mathcal{A}$,

$$\beta_\phi(\theta) = \int_{\mathcal{A}} + \int_{\mathcal{A}^c} \phi(x) dP_\theta(x) = \int_{\mathcal{A}} dP_\theta(x) = P_\theta(X \in \mathcal{A}), \theta \in \Theta.$$

So, $\beta_\phi(\theta)$ is the probability of rejecting the null hypothesis H when $X \sim P_\theta$. Therefore, when $\theta \in \Theta_H$, $\beta_\phi(\theta)$ gives the probability of type I error; and when $\theta \in \Theta_K$, $1 - \beta_\phi(\theta)$ presents the probability of committing type II error. Because of the above probability expression, sometimes the probability of type I error is also termed as *size* (of the rejection region).

1.3 Type I Error Control and Significance Level

Ideally, to choose a good test, we hope that its power function is small over $\theta \in \Theta_H$ and large over $\theta \in \Theta_K$. However, when the sample size is fixed, these two requirements contradict each other. To overcome this, the most popoular approach in the statstical literature is given as follows. First, we specify a *significance level* α in advance, namely to require

$$\beta_\phi(\theta) \leq \alpha, \ \forall \theta \in \Theta_H. \tag{5}$$

This requirement means that the probability of type I error cannot exceed a prespecified level α. Under this constraint, we hope that $\beta_\phi(\theta)$ can be as large as possible for $\theta \in \Theta_K$. This strategy is the so-called principle of type I error control.

Definition 4 *Let $\phi(x)$ be a test for the hypothesis (1). For some α, $0 \leq \alpha \leq 1$, ϕ is said to be a test with significance level α (or a level-α test) if its power function $\beta_\phi(\theta)$ satisfies (5).*

Clearly, a level-α test ϕ implies that

$$\sup_{\theta \in \Theta_H} \beta_\phi(\theta) \leq \alpha.$$

When α is prefixed, one value of the power function $\beta_\phi(\theta)$ at a value $\theta \in \Theta_K$ is named as *power* of test ϕ at θ.

2 Uniformly Most Powerful (UMP) Test

Definition 5 *Test ϕ is said to be a level-α uniformly most powerful (UMP) test for the hypothesis (1), if for any other level-α test $\tilde{\phi}$ there exists*

$$\beta_\phi(\theta) \geq \beta_{\tilde{\phi}}(\theta), \ \forall \theta \in \Theta_K.$$

Obviously the UMP test is the optimal one that should be used in a hypothesis testing problem. However, such test barely exists in reality, except for the case where the alternative hypothesis is simple. A class of distribution is called *simple* if it contains only a single distribution (i.e. Θ_K being a singleton), and otherwise it is said to be *composite*. In the case when both the null and alternative hypotheses are simple, the UMP test can be constructed explicitly with a closed-form expression, according to the famous Neyman-Pearson Lemma.

Suppose $(\mathcal{X}, \mathcal{B}_x)$ denotes the sample space of X. Consider the following hypothesis:

$$H : X \text{ follows a distribution } P_0 \ \leftrightarrow \ K : X \text{ follows a distribution } P_1. \tag{6}$$

Without loss of generality, suppose that there is a σ-finite measure μ on $\mathcal{B}_x$ that dominates both P_0 and P_1. In fact, we can simply set $\mu = P_0 + P_1$. Denote $p_0(x) = dP_0(x)/d\mu, p_1(x) = dP_1(x)/d\mu$. We suppose that there is no such $x \in \mathcal{X}$ at which both $p_0(x)$ and $p_1(x)$ are 0 simultaneously; otherwise, we can always reduce $\mathcal{X}$ to make this condition satisfied. Also, in the following discussion, by convention $\frac{a}{0}$ means ∞ for some $a > 0$.

Theorem 1 *(Neyman-Pearson Lemma) Set* $0 \leq \alpha \leq 1$.

(i) Existence. For the hypothesis (6), there exists a test ϕ *and a constant* k *such that*

$$\phi(x) = \begin{cases} 1, & \text{when } p_1(x)/p_0(x) > k; \\ 0, & \text{when } p_1(x)/p_0(x) < k; \end{cases} \tag{7}$$

$$E_0\{\phi(X)\} = \int_{\mathcal{X}} \phi(x)p_0(x)d\mu(x) = \alpha. \tag{8}$$

(ii) Sufficient condition for a most powerful test. If a test satisfies (7) and (8) for some k*, then it is most powerful for the hypothesis (6) at significance level* α.

(iii) Necessary condition for a most powerful test. If ϕ *is a size-*α *UMP test for the hypothesis (6), then for some* k *it satisfies (7) a.e.*μ *(almost everywhere with respect to* μ*). If the following condition holds,*

$$E_1\{\phi(X)\} = \int_{\mathcal{X}} \phi(x)p_1(x)d\mu(x) < 1,$$

then ϕ *also satisifies (8).*

Proof See the presentation on the board in class. The details are provided in a separate file. □

Corollary 1 *Let* β *be the power of the size-*α *UMP test for testing the hypothesis (6). Then* $\beta \geq \alpha$. *If* $0 < \alpha < 1$ *and* $P_0 \neq P_1$ *(i.e. there exists a set* $A \in \mathcal{B}_x$ *with* $\mu(A) > 0$ *such that* $P_0(A) \neq P_1(A)$*), then* $\beta > \alpha$.

Proof See the presentation on the board in class. Details are included in a separate file. □

3 Distributions with Monotone Likelihood Ratio

The case that both the null and alternative hypotheses are simple is mainly of theoretical interest, since problems arising in applications typically involve a parametric family of distributions depending one or more parameters. Neyman-Pearson Lemma may be applied to construct UMP tests in more general settings.

Definition 6 *The real-parameter family of densities* $f_\theta(x)$, $\{f_\theta(x)d\mu(x), \theta \in \Theta\}$, *is said to have monotone likelihood ratio (MLR) if there exists a real-valued function (i.e. a statistic)* $T(x)$ *such that for any* $\theta_1 < \theta_2$ *the distributions* P_{θ_1} *and* P_{θ_2} *are distinct, and the ratio* $f_{\theta_2}(x)/f_{\theta_1}(x)$ *is a nondecreasing function of a statistic* $T(x)$.

Theorem 2 *Suppose that* $\{f_\theta(x), \theta \in \Theta\}$ *has MLR in* $T(x)$. *For a given* $\theta_0 \in \Theta$, $\{\theta > \theta_0\} \cap \Theta$ *is not an empty set. Set* $0 < \alpha < 1$.

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

$$\phi(x) = \begin{cases} 1, & \text{when } T(x) > c, \\ r, & \text{when } T(x) = c, \\ 0, & \text{when } T(x) < c, \end{cases} \tag{9}$$

where c *and* $r, 0 \leq r \leq 1$ *are determined by*

$$E_{\theta_0}\{\phi(X)\} = \alpha. \tag{10}$$

(ii) The power function $\beta(\theta) = E_\theta\{\phi(X)\}$ *of this test in (9) is nondecreasing on* Θ *and is strictly increasing on* $\{\theta : \theta \in \Theta, 0 < \beta(\theta) < 1\}$.

(iii) For any $\theta < \theta_0$ *the test minimizes* $\beta(\theta)$ *(the probability of type I error) among all tests satisfying (10).*

Proof See the presentation on the board in class. Details are included in a separate file. □

Corollary 2 *Consider the exponential family*

$$f_\theta(x)d\mu(x) = C(\theta)\exp\{Q(\theta)T(x)\}\,d\mu(x), \;\; \theta \in \Theta,$$

where $\Theta \subset R$ *is the natural parameter space, and* $Q(\theta)$ *is strictly monotone on* Θ. *Suppose* $\theta_0 \in$ *the interior of* Θ.

(i) If $Q(\theta)$ *is strictly increasing, then a size-*α *UMP test for* $H : \theta \leq \theta_0$ *aganist* $K : \theta > \theta_0$ *is given by (9) and (10).*

(ii) if $Q(\theta)$ *is strictly decreasing, then a size-*α *UMP test for* $H : \theta \leq \theta_0$ *aganist* $K : \theta > \theta_0$ *is given by*

$$\phi(x) = \begin{cases} 1, & \text{when } T(x) < c, \\ r, & \text{when } T(x) = c, \\ 0, & \text{when } T(x) > c, \end{cases}$$

where c *and* $r, 0 \leq r \leq 1$ *are determined by* $E_{\theta_0}\phi(X) = \alpha$.

Proof The proof is done by simply checking the exponential family has MLR in statistic $T(x)$. $\square$

Theorem 3 *Suppose that $\{f_\theta(x), \theta \in \Theta\}$ has MLR in $T(x)$. For a given $\theta_0 \in \Theta$, $\{\theta < \theta_0\} \cap \Theta$ is not an empty set. Set $0 < \alpha < 1$.*

(i) For testing $H : \theta \geq \theta_0$ aganist $K : \theta < \theta_0$, there exists a UMP test, which is given by

$$\phi(x) = \begin{cases} 1, & \text{when } T(x) < c, \\ r, & \text{when } T(x) = c, \\ 0, & \text{when } T(x) > c, \end{cases} \tag{11}$$

where c and $r, 0 \leq r \leq 1$ are determined by

$$E_{\theta_0}\{\phi(X)\} = \alpha. \tag{12}$$

(ii) The power function $\beta(\theta) = E_\theta \phi(X)$ of this test in (11) is nonincreasing on Θ and is strictly decreasing on $\{\theta : \theta \in \Theta, 0 < \beta(\theta) < 1\}$.

(iii) For any $\theta > \theta_0$ the test minimizes $\beta(\theta)$ (the probability of type I error) among all tests satisfying (12).

Proof The proof can be done trivially by following the steps in the proof of Theorem 2. $\square$

Example 1 Suppose $X_1, \ldots, X_n$ are iid Bernoulli random variables with probability $\theta, 0 < \theta < 1$. For a given $\theta_0 \in (0, 1)$, consider the following hypothesis:

$$H_0 : \theta \leq \theta_0 \text{ against } K : \theta > \theta_0.$$

Since

$$\prod_{i=1}^{n} f_\theta(x_i) = \theta^{\sum_{i=1}^n x_i}(1-\theta)^{n-\sum_{i=1}^n x_i},$$

it is easy to see that the above distribution is an exponential family distribution with $Q(\theta) = \log(\theta/(1-\theta))$ and $T(x) = \sum_{i=1}^n x_i$. Because the function $Q(\theta)$ is strictly increasing in θ, according to Corollary 2, a level-α UMP test is given by

$$\phi(x) = \begin{cases} 1, & \text{when } T(x) > c, \\ r, & \text{when } T(x) = c, \\ 0, & \text{when } T(x) < c. \end{cases}$$

Noting that $T(x) \sim \text{Binomial}(n, \theta)$, we can determine the cutoff c and constant r $(0 \leq r \leq 1)$ as follows;

$$\sum_{i=c+1}^{n} C_n^i \theta_0^i (1-\theta_0)^{n-i} + r C_n^c \theta_0^c (1-\theta_0)^{n-c} = \alpha.$$

Example 2 Suppose $X_1, \ldots, X_n$ iid $N(\theta, 1)$ random variables. To test the following hypothesis:

$$H_0 : \theta \leq \theta_0 \text{ against } K : \theta > \theta_0,$$

we first write the joint density of the sample by

$$\prod_{i=1}^{n} f_\theta(x_i) = (2\pi)^{-n/2} \exp\left\{-n\theta^2/2 - \sum_{i=1}^{n} x_i^2/2\right\} \exp\left\{\theta \sum_{i=1}^{n} x_i\right\}.$$

So, it is an exponential family distribution with $Q(\theta) = \theta$ and $T(x) = \sum_{i=1}^{n} x_i$. Clearly, this $Q(\theta)$ is strictly increasing in θ. According to Corollary 2, a level-α UMP test is given by

$$\phi(x) = \begin{cases} 1, & \text{when } T(x) > c, \\ r, & \text{when } T(x) = c, \\ 0, & \text{when } T(x) < c, \end{cases}$$

Noting that $T(x) \sim N(n\theta, n)$ is a continuous distribution, $P_\theta(T = c) = 0$ for any constant c. Thus, the above test can be simplified as

$$\phi(x) = \begin{cases} 1, & \text{when } T(x) > c, \\ 0, & \text{when } T(x) < c, \end{cases}$$

and the cutoff c can be easily determined by

$$E_{\theta_0}\{I(T > c)\} = P_{\theta_0}(T > c) = \alpha,$$

which results in

$$c = n\theta_0 + \sqrt{n} Z_\alpha,$$

where Z_α is the α-upper quantile of the standard normal $N(0, 1)$.

4 Uniformly Most Powerful Unbiased (UMPU) Test

4.1 Definition

From the above sections, we have seen that it is a nontrivial task to establish UMP test for a given hypothesis test problem, especially when the alternative hypothesis is composite and complex. For example, it is very hard to find UMP test for a two sided alternative $K : \theta \neq \theta_0$. One solution is to lower the bar of optimality; instead of UMP among all eligible tests, consider a smaller class of tests satisfying, say, the property of unbiasedness. Before formally introduce the concept of unbiased test, let us look at an example.

Example 3 Let $X_1, \ldots, X_n$ be an iid sample from $N(\theta, 1)$. Consider the following hypothesis test problem:

$$H_0 : \theta = 0 \text{ against } K : \theta \neq 0,$$

Now, for any given $\theta_1 > 0$, according to N-P lemma, there exists a size-α UMP test ϕ for $H_0 : \theta = 0$ against $K : \theta = \theta_1$, and the rejection region of ϕ is

$$\mathcal{A} = \{(x_1, \ldots, x_n) : \bar{x} \geq Z_\alpha/\sqrt{n}\}.$$

Furthermore, because the rejection region $\mathcal{A}$ is independent of the choice of θ_1, the above test ϕ is a UMP test for $H_0 : \theta = 0$ against $K : \theta > 0$. Also see Example 2 above for a rigorous derivation. It is easy to yield the power function of the test ϕ as follows:

$$\beta_\phi(\theta) = 1 - \Phi(Z_\alpha - \sqrt{n}\theta),$$

where $\Phi(\cdot)$ is the CDF of $N(0, 1)$.

Can we extend the above test ϕ for a broader alternative $K : \theta \neq 0$?

The answer is NO! Why?

This is because for any $\theta < 0$, power function $\beta_\phi(\theta) < \beta_\phi(0) = \alpha$; in other words, when $\theta < 0$, namely $H_0 : \theta = 0$ does not hold, there is very low power to detect such non-null situation. In fact, it means that the chance of rejecting the null when the null is false is smaller than the chance of rejecting the null when the null is true – a really unpleasant situation against to the hope for a good test, and thus this is not reasonable. Although such test is partially good for $\theta > 0$, its overall performance on $\theta \neq 0$ is unacceptable. and thus it should not be used.

Definition 7 *Let ϕ be a test for $H : \theta \in \Theta_H$ against $K : \theta \in \Theta_K$. Denote its power function by $\beta_\phi(\theta)$. If*

$$\beta_\phi(\theta_1) \leq \beta_\phi(\theta_2), \text{ for all } \theta_1 \in \Theta_H \text{ and } \theta_2 \in \Theta_K.$$

Then, test ϕ is said to be unbiased.

Clearly, the test given in Example 3 is not unbiased.

Definition 8 *Let ϕ be an unbiased level-α test for $H : \theta \in \Theta_H$ against $K : \theta \in \Theta_K$. Test ϕ is said to be a leval-α uniformly most powerful unbiased (UMPU) test if for any level-α unibased test $\tilde{\phi}$,*

$$\beta_\phi(\theta) \geq \beta_{\tilde{\phi}}(\theta), \text{ for all } \theta \in \Theta_K$$

where $\beta_\phi(\theta)$ and $\phi_{\tilde{\phi}}(\theta)$ are, respectively, the power functions of tests ϕ and $\tilde{\phi}$.

In Example 3, if we modify the rejection region to define a new test as follows:

$$\mathcal{A} = \{(x_1, \ldots, x_n) : |\bar{x}| \geq Z_{\alpha/2}/\sqrt{n}\},$$

then later we can show that this new test is a level-α UMPU test.

From the definition of UMPU test, we have the following properties:

1. For a level-α UMPU test ϕ for $H : \theta \in \Theta_H$ against $K : \theta \in \Theta_K$, its power $\beta_\phi(\theta) \geq \alpha, \theta \in \Theta_K$.

2. A UMP test must be a UMPU test.

3. Suppose Θ is Euclidean. Let ω be the common boundary of Θ_H and Θ_K (ie. for $\theta \in \omega$ if and only if for any $\rho > 0$ a ball with center of θ and radius ρ has no-empty intersection with either Θ_H and Θ_K). For a level-α UMPU test, if its power function $\beta_\phi(\theta)$ is continuous in θ, then

$$\beta_\phi(\theta) = \alpha, \text{ for all } \theta \in \omega.$$

4.2 UMPU Test in One-dimensional Exponential Families

Consider a probability space $\{\mathcal{X}, \mathcal{B}_x, \mathcal{P}_\theta, \theta \in \Theta\}$ where $\mathcal{P}_\theta = \{dP_\theta(x) = C(\theta)e^{\theta T(x)}d\mu(x) = f(x, \theta)d\mu(x), \theta \in \Theta\}$. For such exponential family, with the availability of MLR, we already know how to construct UMP test for $H : \theta \leq \theta_0$ against $K : \theta > \theta_0$ or $H : \theta \geq \theta_0$ against $K : \theta < \theta_0$. Now we consider two additional cases:

$$H : \theta_1 \leq \theta \leq \theta_2 \text{ against } K : \theta < \theta_1 \text{ or } \theta > \theta_2 \tag{13}$$

$$H : \theta = \theta_0 \text{ against } K : \theta \neq \theta_0 \tag{14}$$

For both cases of hypotheses, there do not exist UMP tests, seen arguments in Example 3. However, we can establish UMPU in both scenarios.

Consider the first scenario (13).

Theorem 4 *Define a test ϕ of the following form:*

$$\phi(x) = \begin{cases} 1, & \text{when } T(x) < c_1 \text{ or } T(x) > c_2, \\ r_i, & \text{when } T(x) = c_i, i = 1, 2, \\ 0, & \text{when } c_1 < T(x) < c_2. \end{cases}$$

where $0 \leq r_1, r_2 \leq 1$. If the above test ϕ satisfies $E_{\theta_1}\{\phi(X)\} = E_{\theta_2}\{\phi(X)\} = \alpha$, then test ϕ is a level-α UMPU test for hypothesis (13).

Proof Skip. $\square$

Consider the first scenario (14).

Theorem 5 *Define a test ϕ of the following form:*

$$\phi(x) = \begin{cases} 1, & \textit{when } T(x) < c_1 \textit{ or } T(x) > c_2, \\ r_i, & \textit{when } T(x) = c_i, i = 1, 2, \\ 0, & \textit{when } c_1 < T(x) < c_2. \end{cases}$$

where $0 \le r_1, r_2 \le 1$. If the above test ϕ satisfies $E_{\theta_0}\{\phi(X)\} = \alpha$ and $E_{\theta_0}\{T(X)\phi(X)\} = \alpha E_{\theta_0}\{T(X)\}$, then test ϕ is a level-α UMPU test for hypothesis (14).

Proof Skip. $\square$

Example 4 Let $X_1, \ldots, X_n$ be an iid sample from $N(\theta, \sigma^2)$ with σ^2 known. Consider the following hypothesis test problem:

$$H_0 : \theta = \theta_0 \text{ against } K : \theta \neq \theta_0.$$

It is known that the joint density of the sample is a member in the one-dimensional exponential family with $T(x) = \sum_{i=1}^n x_i$. Since this is a continuous distribution, $P(T(X) = c_i) = 0, i = 1, 2$. Thus, according to Theorem 5, a level α UMPU test for the above hypothesis takes the following form:

$$\phi(x) = \begin{cases} 1, & \text{when } T(x) < c_1 \text{ or } T(x) > c_2, \\ 0, & \text{when } c_1 < T(x) < c_2, \end{cases}$$

where constants c_1, c_2 are chosen to satisfy

$$\begin{aligned} P_{\theta_0}(c_1 < T(x) < c_2) &= 1 - \alpha \\ E_{\theta_0}\{T(X)I(c_1 < T(x) < c_2)\} &= (1 - \alpha)E_{\theta_0}\{T(X)\}. \end{aligned}$$

Both may be rewritten as follows:

$$\begin{aligned} \Phi\left(\frac{c_2 - n\theta_0}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{c_1 - n\theta_0}{\sqrt{n}\sigma}\right) &= 1 - \alpha \\ \int_{\frac{c_1 - n\theta_0}{\sqrt{n}\sigma}}^{\frac{c_2 - n\theta_0}{\sqrt{n}\sigma}} (n\theta_0 + \sqrt{n}\sigma z)\frac{1}{\sqrt{2\pi}}e^{-z^2/2}dz &= (1 - \alpha)n\theta_0. \end{aligned}$$

It is easy to show that $c_1 = n\theta_0 - \sqrt{n}\sigma Z_{\alpha/2}$ and $c_2 = n\theta_0 + \sqrt{n}\sigma Z_{\alpha/2}$ are a pair of constants that satisfy the above equations. Thus, we have the rejection region given by

$$\mathcal{A} = \left\{(x_1, \ldots, x_n) : \left|\frac{\sum_{i=1}^n x_i - n\theta_0}{\sigma}\right| > Z_{\alpha/2}\right\},$$

that is, when $|(\sum_{i=1}^n x_i - n\theta_0)/\sigma| > Z_{\alpha/2}$, we reject the null $H_0 : \theta = \theta_0$.

Example 5 Under the same notation of Example 4, we now consider the following hypothesis testing problem:

$$H_0 : \theta_1 \le \theta \le \theta_2 \text{ against } K : \theta < \theta_1 \text{ or } \theta > \theta_2.$$

According to Theorem 4, a level-α UMPU test for the above hypothesis takes the same form as that given in Example 4, that is,

$$\phi(x) = \begin{cases} 1, & \text{when } T(x) < c_1 \text{ or } T(x) > c_2, \\ 0, & \text{when } c_1 < T(x) < c_2, \end{cases}$$

where constants c_1, c_2 are chosen to satisfy

$$P_{\theta_1}(c_1 < T(x) < c_2) \;=\; P_{\theta_2}(c_1 < T(x) < c_2) = 1 - \alpha,$$

which may be rewritten as follows:

$$\Phi\left(\frac{c_2 - n\theta_1}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{c_1 - n\theta_1}{\sqrt{n}\sigma}\right) \;=\; \Phi\left(\frac{c_2 - n\theta_2}{\sqrt{n}\sigma}\right) - \Phi\left(\frac{c_1 - n\theta_2}{\sqrt{n}\sigma}\right) = 1 - \alpha$$

The solution pair (c_1, c_2) can be found by using a certain numerical search algorithm. Note that in a special case of testing for bioequivalence, $H_0 : |\theta| \le \theta_0$, where $\theta_2 = \theta_0, \theta_1 = -\theta_0$ or $\theta_2 = -\theta_1$, we have $c_2 = -c_1$. Then, a closed form expression for (c_2, c_1) can be obtained (which is left as an exercise).

4.3 Uniformly Most Powerful Invariant Test

Skip. This requires some basic knowledge of abstract algebra, in particlar the concept of *group*.