Statistical Inference and Multivariate Analysis (MA 324)

Class Notes
January – May, 2021

Instructor
Ayon Ganguly
Department of Mathematics
IIT Guwahati

Contents

Chapter 3

Tests of Hypotheses

3.1 Introduction

In the previous section, we have discussed point estimation, where we try to find meaningful guesses for unknown parameters or parametric functions. In testing of hypothesis, we do not guess the value of the parametric function. We try to check if a given statement about parameters is true or not. Let us start with examples.

Example 3.1. The Cherry Blossom Run is a 10 mile race that takes place every year in D.C. In 2009, there were 14974 participants and average running time of all the participants was 103.5 minutes. Now, in the year 2010, one may want to ask the question: Were runners faster in 2010 compared to 2009? Of course, if we have running times of all the participant, we can find the average running time on the year 2010 and compare it with average running time of the year 2009. However, assume that it is not possible to have the running times of all the participants in 2010. In this case, how should we proceed and answer the question? Again, we will rely on a RS, say of size n, drawn from the 2010 runners, and denote the running time by X_1, X_2, ..., X_n. Based on historic data, it my be plausible to assume that the distribution of the running time is a normal with variance 373. Hence, we are given a RS X_1, X_2, ..., X_n and we want to know if $X_1 \sim N(103.5,\ 373)$. This is a hypothesis about the distribution of running time and we want to test the hypothesis. There are many ways this hypothesis could be false:

- $E(X_1) \neq 103.5$
- $Var(X_1) \neq 373$
- X_1 is not normal.

From the analysis of the past data, it is found that the last two assumptions are reasonable and hence we put them as model assumptions. Thus, the only way in which the hypothesis $X_1 \sim N(103.5,\ 373)$ could be false is $\mu = E(X_1) \neq 103, 5$. Using modeling assumptions, we have reduced the number of ways the hypothesis $X_1 \sim N(103.5, 373)$ may be false. Now, we want to test: Is $\mu = 103.5$ or $\mu < 103.5$? Note the we have written $\mu < 103.5$, as our initial question was: Is the runner faster in the year 2010? Thus, μ must be less than or equal to 103.5. Now, the only way the hypothesis can be false is if $X_1 \sim N(\mu, 373)$ for some $\mu < 103.5$. This is an example, where we compare an expected value to a fixed reference number (here 103.5).

Simple heuristic would be: If $\overline{X} < 103.5$, then $\mu < 103.5$, where $\overline{X}$ is the sample mean based on a RS of size n drawn form runners of the year 2010. It is easy to understand that it can go wrong if we select, by chance, the fast runners in the sample. Better heuristic could be: If $\overline{X} < 103.5 - a$ then $\mu < 103.5$ for some a. We will try to make this intuitions more precise as we proceed. ||

Example 3.2. Pharmaceutical companies use hypothesis testing to test if a new drug is efficient. To do so, a group of patients are randomly divided into two groups. One of the groups is administered with the drug and the other is administered with placebo. The first and the second groups are called test group and control group, respectively. Assume that the drug is a cough syrup. Let μ_1 denote the expected number of expectorations per hour after a patient has used placebo and μ_2 denote the expected number of expectorations per hour after a patient has used the syrup. We want to know if $\mu_2 < \mu_1$. In this case, two expectations are compared. One of the method to attack this problem is to draw RS from both the groups. Let $X_1, X_2, \ldots, X_{n_1}$ denote a RS of size n_1 from the control group. Let $Y_1, \ldots, Y_{n_2}$ denote a RS of size n_2 drawn from test group. We want to test if $\mu_2 = \mu_1$ or $\mu_2 < \mu_1$. If $\mu_1 = \mu_2$, then the new drug is not efficient. If $\mu_2 < \mu_1$, then the new drug has some effect. Note that we have taken $\mu_2 < \mu_1$ and we have not considered if $\mu_2 > \mu_1$. The reason for the same is as follows: As we are trying to check the efficiency of the new drug, we implicitly assume that $\mu_2 \leq \mu_1$. Heuristically, we should compare $\overline{X}$ and $\overline{Y}$. ||

Example 3.3. Let a coin is tossed 80 times, and head are obtained 55 times. Can we conclude that the coin is fair based on this data? Let, for $i = 1, 2, \ldots, 80$, X_i be an indicator RV, which takes value 1 if ith toss is a head and takes value zero if the ith toss is a tail. Then, we have a RS $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} Bernoulli(p)$, where p is the probability of getting a head in a toss. We want to test $p = 0.5$ or $p \neq 0.5$. Intuitively, it makes sense to use $\overline{X}$ to check if $p = 0.5$ or not. Here, the observed value of $\overline{X}$ is $\bar{x} = 55/80 = 0.6875$. If p is actually equal to 0.5, then, using CLT, we have

$$T_n = \frac{\sqrt{n}\left(\overline{X}_n - 0.5\right)}{\sqrt{0.5 \times (1 - 0.5)}} \approx N(0, 1).$$

Now, if the number of heads is too small or too large (*i.e.*, the value of $\bar{x}$ is not close to 0.5), we should go for biased coin. If the number of head is moderate (*i.e.*, the value of $\bar{x}$ is close to 0.5), we should choose that the coin is fair. In the first case ($\bar{x}$ not close to 0.5), the absolute observed value of T_n will be large. The absolute observed value of T_n will be close to zero in the second case. This discussion suggests that we should reject the fact that the coin is fair if $|T_n| > C$ for some appropriate real constant C.

Here, the observed value of T_n is 3.3541, which is too extreme with respect to a standard normal distribution as $P(|Z| > 3.35) \approx 0.0008$, where $Z \sim N(0, 1)$. Therefore, it is quite reasonable to reject the hypothesis $p = 0.5$ based on the data. ||

Example 3.4. A coin is tossed 80 times, and head are obtained 35 times. Can we conclude that the coin is significantly fair? Here, the observed value of T_n is -1.1180. Data do not suggest to reject the fact that the coin is fair, as the observed value of T_n is not extreme with respect to a standard normal distribution. Note that $P(|Z| > 1.11) = 0.267$. ||

In the last two examples, we have talked about extreme or not extreme. The question is: Which values are considered as extreme and which are not? More precisely, we are rejecting $p = 0.5$ if the observation belong to the set

$$\{\boldsymbol{x} : |T_n(\boldsymbol{x})| > C\}.$$

What value of C should we choose so that we can make correct decision? This issues will be discussed as we proceed in the current chapter.

Definition 3.1 (Statistical Hypothesis). *A statistical hypothesis or simply hypothesis is a statement about the unknown parameters.*

Definition 3.2 (Null and Alternative Hypotheses). *Suppose that one wants to choose between two reasonable hypotheses $H_0 : \boldsymbol{\theta} \in \Theta_0$ against $H_1 : \boldsymbol{\theta} \in \Theta_1$, where $\Theta_0 \subset \Theta$, $\Theta_1 \subset \Theta$ and $\Theta_1 \cap \Theta_2 = \emptyset$. We call $H_0 : \boldsymbol{\theta} \in \Theta_0$ and $H_1 : \boldsymbol{\theta} \in \Theta_1$ as null hypothesis and alternative hypothesis, respectively.*

Definition 3.3 (Simple and Composite Hypothesis). *A hypothesis is called simple if it specifies the underlying distribution. Otherwise, it is called composite hypothesis.*

Our aim is to choice one hypothesis among null and alternative hypotheses. As we will see that the roles of these two hypotheses are asymmetric, we need to be careful while labeling these two hypotheses.

3.2 Errors and Errors Probabilities

As illustrated in Examples 3.3 and 3.4, the decision to accept or reject null hypothesis will be taken based on a RS. We will consider a reasonable statistic and make the choice based on the statistic. If the observed value of the statistic belongs to an appropriate set, we reject null hypothesis and if the value of the statistic does not belong to the set, we accept null hypothesis. Now, the statistic belongs to a set can be alternatively written as the sample observation belongs to a subset of $\mathbb{R}^n$. In Examples 3.3 and 3.4, suppose that we reject the null hypothesis if and only if $|T_n| > 1.96$. The set $\{|T_n| > 1.96\}$ is the set of all sample points $\boldsymbol{x}$ such that $T_n(\boldsymbol{x}) > 1.96$ or $T_n(\boldsymbol{x}) < -1.96$. Thus, the condition $|T_n| > 1.96$ actually divides $\mathbb{R}^n$ into two parts, *viz.*, $\{\boldsymbol{x} \in \mathbb{R}^n : |T_n(\boldsymbol{x})| > 1.96\}$ and $\{\boldsymbol{x} \in \mathbb{R}^n : |T_n(\boldsymbol{x})| \leq 1.96\}$. If the sample observation $\boldsymbol{x}$ belongs to the first partition, we reject the null hypothesis. If the sample observation $\boldsymbol{x}$ belongs the second partition, we accept the null hypothesis. Of course, we need to find a "meaningful" partition so that we can make correct decision.

Definition 3.4 (Critical Region and Acceptance Region). *Suppose that we want to test $H_0 : \boldsymbol{\theta} \in \Theta_0$ against $H_1 : \boldsymbol{\theta} \in \Theta_1$ based on a RS of size n drawn from a population having PMF/PDF $f(\cdot, \boldsymbol{\theta})$, where $\boldsymbol{\theta} \in \Theta_0 \cup \Theta_1$. Let R be a subset of $\mathbb{R}^n$ such that we reject H_0 if and only if $\boldsymbol{x} \in R$, where $\boldsymbol{x}$ denotes a realization of the RS. Then, R is called rejection region or critical region, and R^c is called acceptance region.*

In this process, there are four possibility as described in the following table. There are two cases, where we do not commit any error. These cases are accepting null hypothesis when it is actually true and rejecting null hypothesis when it is actually false. The green ticks in the following table signify that there are no errors. We commit errors in other two cases, *viz.*, accepting null hypothesis when it is actually false or rejecting null hypothesis when it is actually true.

	H_0 true	H_1 true
Accept H_0	✓	Type-II Error
Reject H_0	Type-I Error	✓

Definition 3.5 (Type-I and Type-II Errors)**.** *The error committed by rejecting H_0 when it is actually true is called Type-I Error. The error committed by accepting H_0 when it is actually false is called Type-II Error.*

Example 3.5. Let $X_1, X_2, \ldots, X_9 \overset{i.i.d.}{\sim} N(\theta, 1)$. Suppose that we are want to test $H_0 : \theta = 5.5$ against $H_1 : \theta = 7.5$. For comparison purpose, let use consider four critical regions:

$$R_1 = \emptyset, \quad R_2 = \left\{\boldsymbol{x} \in \mathbb{R}^9 : \bar{x} > 7\right\}, \quad R_3 = \left\{\boldsymbol{x} \in \mathbb{R}^9 : \bar{x} > 6\right\}, \quad R_4 = \mathbb{R}^9.$$

As we are trying to test hypotheses regarding population mean, it is intuitively make sense to use sample mean. That is the reason to take R_2 and R_3 in terms of sample mean $\bar{x}$. On the other hand, R_1 and R_4 are two extremes. The critical regions R_1 accepts null hypothesis irrespective of the realization of the RS. Similarly, we always reject the null hypothesis if the critical region is R_4.

Note that Type-I or Type-II errors are events. Thus, we may talk about the probabilities of these errors. For critical region R_3,

$$\begin{aligned} P\left(\text{Type-I Error}\right) &= P_{\theta=5.5}\left(\overline{X} > 6\right) = 1 - \Phi\left(3(6-5.5)\right) = 0.06681 \\ P\left(\text{Type-II Error}\right) &= P_{\theta=7.5}\left(\overline{X} \le 6\right) = \Phi\left(3(6-7.5)\right) \sim 0. \end{aligned}$$

Similarly, the probabilities of errors for other critical regions can be calculated and given in following table.

	R_1	R_2	R_3	R_4
P(Type-I)	0	~ 0	0.06681	1
P(Type-II)	1	0.06681	~ 0	0

In this example, $R_1 \subset R_2 \subset R_3 \subset R_4$. Notice that as we increase the size of the critical region, probability of Type-I error increases and that of Type-II error decreases. In other words, if we try to reduce probability of one error, probability of the other one increases. ||

Definition 3.6 (Power Function)**.** *The power function of a critical region, denoted by $\beta : \Theta_1 \cup \Theta_0 \to [0, 1]$, is the probability of rejecting the null hypothesis H_0 when $\boldsymbol{\theta}$ is the true value of the parameter, i.e.,*

$$\beta(\boldsymbol{\theta}) = P_{\boldsymbol{\theta}}(\boldsymbol{X} \in R).$$

It is clear that $\beta(\cdot)$ is the probability of Type-I error if $\boldsymbol{\theta} \in \Theta_0$. For $\boldsymbol{\theta} \in \Theta_1$, $\beta(\cdot)$ is one minus probability of Type-II error.

Example 3.6. Let $X_1, X_2, \ldots, X_9 \overset{i.i.d.}{\sim} N(\theta, 1)$. Suppose that we are want to test $H_0 : \theta = 5.5$ against $H_1 : \theta = 7.5$. The power function of the critical region $R_2 = \{\boldsymbol{x} \in \mathbb{R}^9 : \bar{x} > 7\}$ is

$$\beta(\theta) = P_\theta\left(\boldsymbol{X} \in R\right) = P_\theta\left(\overline{X} > 7\right) = P_\theta\left(3\left(\overline{X} - \theta\right) > 21 - 3\theta\right) = 1 - \Phi(21 - 3\theta).$$

for $\theta = 5.5$ and 7.5. ||

3.3 Best Test

Note that we want to find a "meaningful" partition. As mentioned, "meaningful" means that we take correct decisions by rejecting (accepting) the null hypothesis when it is actually false (true). That means that we want $\boldsymbol{x}$ to be in R ($\boldsymbol{x} \notin R$) when null hypothesis is false (true). A critical region R which minimizes the probabilities of both the errors could be a "meaningful" choice. Unfortunately, as shown in the previous example, the reduction of probability of one type of error forces to increase the probability of other type of error, in general. Optimization in such a situation can be done in several ways. For tests of hypotheses, the method is as follows: Put an upper bound on the probability of Type-I error and try to minimize the probability of Type-II error subject to the upper bound of the probability of Type-I error.

Definition 3.7 (Size of a Test). *Let $\alpha \in (0,\, 1)$ be a fixed real number. A test for $H_0 : \boldsymbol{\theta} \in \Theta_0$ against $H_1 : \boldsymbol{\theta} \in \Theta_1$ with power function $\beta(\cdot)$ is called a size α test if*

$$\sup_{\boldsymbol{\theta} \in \Theta_0} \beta(\boldsymbol{\theta}) = \alpha.$$

Definition 3.8 (Level of a Test). *A test is called level α if $\beta(\boldsymbol{\theta}) \leq \alpha$ for all $\boldsymbol{\theta} \in \Theta_0$.*

Size of a test can be considered as worst possible probability of Type-I error. If a test is of size α, then it is of level α.

Example 3.7. Let $X_1,\, X_2,\, \ldots,\, X_9 \overset{i.i.d.}{\sim} N(\theta,\, 1)$. Suppose that we are want to test $H_0 : \theta = 5.5$ against $H_1 : \theta = 7.5$. We have seen that the power function of the critical region $R_2 = \{\boldsymbol{x} \in \mathbb{R}^9 : \bar{x} > 7\}$ is

$$\beta(\theta) = 1 - \Phi(21 - 3\theta).$$

for $\theta = 5.5$ and 7.5. In this case, $\Theta_0 = \{5.5\}$ is single-tone. Therefore, the size of the test is

$$\sup_{\theta \in \Theta_0} \beta\,(\theta) = \beta\,(5.5) = 1 - \Phi(4.5) \simeq 3.4 \times 10^{-6}.$$

This test is of level α for any $\alpha \in [1 - \Phi(4.5),\, 1]$. ||

Definition 3.9 (Critical or Test Function). *A function $\psi : \mathscr{X}^n \to [0,\, 1]$ is called a critical function or test function, where $\psi(\boldsymbol{x})$ stands for the probability of rejecting H_0 when $\boldsymbol{X} = \boldsymbol{x}$ is observed. Here, $\mathscr{X}^n$ is the sample space of the random sample of size n.*

Example 3.8. Let $X_1,\, X_2,\, \ldots,\, X_9 \overset{i.i.d.}{\sim} N(\theta,\, 1)$. Suppose that we are want to test $H_0 : \theta = 5.5$ against $H_1 : \theta = 7.5$. Let us consider two critical regions $R_1 = \{\boldsymbol{x} \in \mathbb{R}^9 : \bar{x} > 6\}$ and $R_2 = \{\boldsymbol{x} \in \mathbb{R}^9 : \bar{x} > 7\}$. The critical regions R_1 and R_2, respectively, can be expressed as the test functions

$$\psi_1(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \bar{x} > 6 \\ 0 & \text{if } \bar{x} \leq 6, \end{cases} \qquad \text{and} \qquad \psi_2(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \bar{x} > 7 \\ 0 & \text{if } \bar{x} \leq 7. \end{cases}$$

Note that the power function for R_1 is

$$\beta_1(\theta) = P_\theta(\overline{X} > 6) = 1 - \Phi(18 - 3\theta) \text{ for } \theta = 5.5,\, 7.5,$$

which can be expressed as $E_\theta\,(\psi_1(\boldsymbol{X}))$. ||

Thus, the last example shows that the test function is an alternative way to write a critical region. Then, what does we gain by defining test function? First note that $\psi(\boldsymbol{x})$ being a probability, can take any value between zero and one (not only values 0 and 1). This is the gain. We will illustrate it in Example 3.9. However, before going into the example, let us discuss couple of definitions in terms of test function.

Definition 3.10 (Power Function)**.** *The power function of a test function is defined by*

$$\beta(\boldsymbol{\theta}) = E_{\boldsymbol{\theta}}\left(\psi(\boldsymbol{X})\right) \text{ for all } \boldsymbol{\theta} \in \Theta_0 \cup \Theta_1.$$

Once the power function is defined, we can now define size or level of a test using the power function as given in Definitions 3.7 and 3.8.

Definition 3.11 (Randomized and Non-randomized Tests)**.** *A test is called randomized test if $\psi(\boldsymbol{x}) \in (0,\, 1)$ for some $\boldsymbol{x}$. Otherwise, it is called a non-randomized test.*

Example 3.9. Let X be a sample of size one form a $Bin(3,\, p)$ distribution. We want to test $H_0 : p = \frac{1}{4}$ against $H_1 : p = \frac{3}{4}$. The probabilities of $X = x$ under H_0 is given in the table below:

x	Prob. under H_0
0	27/64
1	27/64
2	9/64
3	1/64

Do we have a critical region of size $\alpha_1 = \frac{5}{32}$? The answer is yes, and the critical region is given by $\{2,\, 3\}$ as $P(X = 2 \text{ or } 3) = \frac{5}{32}$ under H_0.

Does a critical region of size $\alpha_2 = \frac{1}{32}$ exist? The answer is no, there is no critical region of size $\frac{1}{32}$. However, we have a randomized test of of size $\frac{1}{32}$, and it is given by

$$\psi(x) = \begin{cases} 1 & \text{if } x = 3 \\ \frac{1}{9} & \text{if } x = 2 \\ 0 & \text{otherwise,} \end{cases}$$

as $E_{p=\frac{1}{4}}\left(\psi(X)\right) = 1 \times \frac{1}{64} + \frac{1}{9} \times \frac{9}{64} = \frac{1}{32}$. Hence, in this case though a critical region of size 1/32 does not exist, a randomized test function of the same size exists. This is the gain of defining a test function over critical region. ||

Remark 3.1. Test functions are more general in the sense that all critical regions can be represented as a test function, but the converse is not true. †

Remark 3.2. Let for a fixed $\boldsymbol{x}_0$, $\psi(\boldsymbol{x}_0) = 0.6$. If $\boldsymbol{X} = \boldsymbol{x}_0$ is observed, how should we accept or reject H_0? We will perform a random experiment with two outcomes (toss of a coin), with one (say head) has probability 0.4, and other (say tail) has probability 0.6. If tail occur, we reject H_0, otherwise we accept it. †

Definition 3.12 (Uniformly Most Powerful Test)**.** *Consider the collection $\mathscr{C}_\alpha$ of all level α tests for $H_0 : \boldsymbol{\theta} \in \Theta_0$ against $H_1 : \boldsymbol{\theta} \in \Theta_1$. A test belonging to $\mathscr{C}_\alpha$ with power function $\beta(\cdot)$ is called uniformly most powerful (UMP) level α test if $\beta(\boldsymbol{\theta}) \geq \beta^*(\boldsymbol{\theta})$ for all $\boldsymbol{\theta} \in \Theta_1$, where $\beta^*(\cdot)$ is the power function of any other test in $\mathscr{C}_\alpha$. If the alternative hypothesis is simple (that means that Θ_1 is singleton), the test is called most powerful (MP) level α test.*

Remark 3.3. Note that here we are putting a bound on probability of type one error. The bound is α. Among all the tests whose probability of Type-I error is bounded by α, we are trying to find one for which probability of Type-II error is minimum. A test satisfies this criterion is called a UMP level α test. †

Remark 3.4. When $H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_1$ for some fixed $\boldsymbol{\theta}_1$, *i.e.*, H_1 is simple, it boils down to check if $\beta(\boldsymbol{\theta}_1) \geq \beta^*(\boldsymbol{\theta}_1)$. Hence, the word 'uniformly' is removed. †

3.4 Simple Null Vs. Simple Alternative

Theorem 3.1 (Neyman-Pearson Lemma). *Let $\boldsymbol{\theta}_0 \neq \boldsymbol{\theta}_1$ be two fixed numbers in Θ. The MP level α test for $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $H_1 : \boldsymbol{\theta} = \boldsymbol{\theta}_1$ is given by*

$$\psi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } L(\boldsymbol{\theta}_1) > kL(\boldsymbol{\theta}_0) \\ \gamma & \text{if } L(\boldsymbol{\theta}_1) = kL(\boldsymbol{\theta}_0) \\ 0 & \text{if } L(\boldsymbol{\theta}_1) < kL(\boldsymbol{\theta}_0), \end{cases}$$

where $k \geq 0$ and $\gamma \in [0, 1]$ such that $\beta(\boldsymbol{\theta}_0) = E_{\boldsymbol{\theta}_0}(\psi(\boldsymbol{X})) = \alpha$. Here, $L(\cdot)$ is the likelihood function.

Proof: Let $\psi^*(\cdot) \in \mathscr{C}_\alpha$ and the power function of $\psi^*(\cdot)$ be $\beta^*(\cdot)$. We need to show that $\beta(\boldsymbol{\theta}_1) - \beta^*(\boldsymbol{\theta}_1)$. Let us consider the quantity

$$Q(\boldsymbol{x}) = (\psi(\boldsymbol{x}) - \psi^*(\boldsymbol{x}))(L(\boldsymbol{\theta}_1) - kL(\boldsymbol{\theta}_0)).$$

First, we will prove that $Q(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in \mathscr{X}^n$. Let us define

$$\begin{aligned} \mathscr{X}_1^n &= \{\boldsymbol{x} \in \mathscr{X}^n : L(\boldsymbol{\theta}_1) - kL(\boldsymbol{\theta}_0) > 0\}, \\ \mathscr{X}_2^n &= \{\boldsymbol{x} \in \mathscr{X}^n : L(\boldsymbol{\theta}_1) - kL(\boldsymbol{\theta}_0) = 0\}, \\ \mathscr{X}_3^n &= \{\boldsymbol{x} \in \mathscr{X}^n : L(\boldsymbol{\theta}_1) - kL(\boldsymbol{\theta}_0) < 0\}. \end{aligned}$$

First notice that $\mathscr{X}^n = \mathscr{X}_1^n \cup \mathscr{X}_2^n \cup \mathscr{X}_3^n$. Now, if $\boldsymbol{x} \in \mathscr{X}_1^n$, $\psi(\boldsymbol{x}) \geq \psi^*(\boldsymbol{x})$ as $\psi(\boldsymbol{x}) = 1$ and $0 \leq \psi^*(\boldsymbol{x}) \leq 1$. Therefore, $Q(\boldsymbol{x}) \geq 0$ for all $\boldsymbol{x} \in \mathscr{X}_1^n$. For $\boldsymbol{x} \in \mathscr{X}_2^n$, $Q(\boldsymbol{x}) = 0$. For $\boldsymbol{x} \in \mathscr{X}_3^n$, $\psi(\boldsymbol{x}) = 0 \leq \psi^*(\boldsymbol{x})$. Therefore, $Q(\boldsymbol{x}) \geq 0$ for $\boldsymbol{x} \in \mathscr{X}_3^n$. Thus,

$$\begin{aligned} & \int_{\mathscr{X}^n} Q(\boldsymbol{x})\, d\boldsymbol{x} \geq 0 \\ \implies & \int_{\mathscr{X}^n} (\psi(\boldsymbol{x}) - \psi^*(\boldsymbol{x}))\, L(\boldsymbol{\theta}_1)\, d\boldsymbol{x} + k \int_{\mathscr{X}^n} (\psi(\boldsymbol{x}) - \psi^*(\boldsymbol{x}))\, L(\boldsymbol{\theta}_0)\, d\boldsymbol{x} \geq 0 \\ \implies & \beta(\boldsymbol{\theta}_1) - \beta^*(\boldsymbol{\theta}_1) - k(\beta(\boldsymbol{\theta}_0) - \beta^*(\boldsymbol{\theta}_0)) \geq 0 \\ \implies & \beta(\boldsymbol{\theta}_1) - \beta^*(\boldsymbol{\theta}_1) \geq k(\beta(\boldsymbol{\theta}_0) - \beta^*(\boldsymbol{\theta}_0)) \geq 0, \end{aligned}$$

as $\beta(\boldsymbol{\theta}_0) = \alpha$ and $\beta^*(\boldsymbol{\theta}_0) \leq \alpha$. This completes the proof. □

Note that $L(\theta_1) > kL(\theta_0)$ can be expressed as $\dfrac{L(\theta_1)}{L(\theta_0)} > k$ if $L(\theta_0) > 0$. Hence, loosely speaking, the MP test rejects the null hypothesis for large values of the ratio.

Example 3.10. Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$, where σ is known. Let $\mu_0 < \mu_1$ be two fixed real numbers. We are interested to test $H_0 : \mu = \mu_0$ against $H_1 : \mu = \mu_1$. Here, the likelihood function is

$$\begin{aligned} L(\mu) &= \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2}\sum_{i=1}^{n}(x_i - \mu)^2\right] \\ &= \frac{1}{\sigma^n (2\pi)^{\frac{n}{2}}} \exp\left[-\frac{1}{2\sigma^2}\left\{\sum_{i=1}^{n} x_i^2 - 2n\mu\bar{x} + n\mu^2\right\}\right]. \end{aligned}$$

Therefore,

$$\frac{L(\mu_1)}{L(\mu_0)} = \exp\left[\frac{1}{2\sigma^2}\left\{2n\bar{x}(\mu_1 - \mu_0) + n\left(\mu_0^2 - \mu_1^2\right)\right\}\right].$$

Now.

$$\begin{aligned} \frac{L(\mu_1)}{L(\mu_0)} > k &\iff \exp\left[\frac{1}{2\sigma^2}\left\{2n\bar{x}(\mu_1 - \mu_0) + n\left(\mu_0^2 - \mu_1^2\right)\right\}\right] > k \\ &\iff \frac{1}{2\sigma^2}\left\{2n\bar{x}(\mu_1 - \mu_0) + n\left(\mu_0^2 - \mu_1^2\right)\right\} > \ln k \\ &\iff 2n(\mu_1 - \mu_0)\bar(x) > 2\sigma^2 \ln k - n\left(\mu_0^2 - \mu_1^2\right) \\ &\iff \bar{x} > \frac{2\sigma^2 \ln k - n\left(\mu_0^2 - \mu_1^2\right)}{2n(\mu_1 - \mu_0)} = \kappa \text{ (say), as } \mu_1 > \mu_0. \end{aligned}$$

Thus, $\dfrac{L(\mu_1)}{L(\mu_0)} > k$ if and only if $\bar{x} > \kappa$ for some constant κ. Hence, using Neymann-Pearson lemma, the test function of the MP level α test is given by

$$\psi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \bar{x} > \kappa \\ \gamma & \text{if } \bar{x} = \kappa \\ 0 & \text{if } \bar{x} < \kappa, \end{cases}$$

where κ_1 and γ are such that

$$\begin{aligned} E_{\mu_0}(\psi(\boldsymbol{X})) = \alpha &\iff P_{\mu_0}\left(\overline{X} > \kappa\right) + \gamma P_{\mu_0}\left(\overline{X} = \kappa\right) = \alpha \\ &\iff P_{\mu_0}\left(\overline{X} > \kappa\right) = \alpha \\ &\iff P_{\mu_0}\left(\sqrt{n}\,\frac{\overline{X} - \mu_0}{\sigma} > \sqrt{n}\,\frac{\kappa - \mu_0}{\sigma}\right) = \alpha \\ &\iff \sqrt{n}\,\frac{\kappa - \mu_0}{\sigma} = z_\alpha, \end{aligned}$$

where z_α is a real number such that $P(Z > z_\alpha) = \alpha$ for a RV $Z \sim N(0, 1)$ (please see the Figure 3.1). Here, we can take any value of $\gamma \in [0, 1]$ as $P_{\mu_0}(\overline{X} = \kappa) = 0$. In this situation γ is taken to be zero and that makes it a non-randomized test. Hence, the MP level α test is given by

$$\psi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \frac{\sqrt{n}}{\sigma}(\bar{x} - \mu_0) > z_\alpha \\ 0 & \text{otherwise.} \end{cases}$$

||

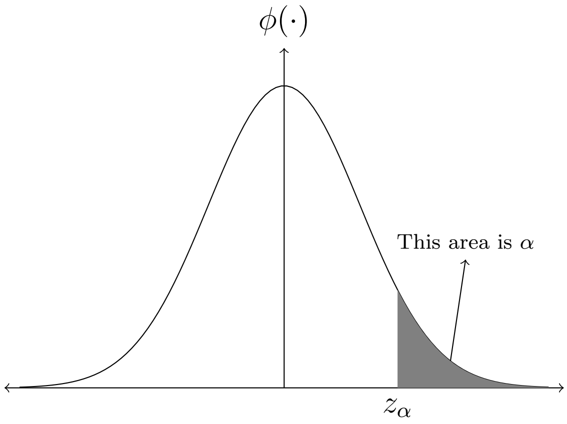

Figure 3.1: Upper α-point for a standard normal distribution

Remark 3.5. Note the way of solving the problem. We try to simplify $\dfrac{L(\mu_1)}{L(\mu_0)} > k$ so that we can write an equivalent condition on a statistic whose distribution under H_0 is known or can be found. If this statistic is a continuous random variable, we will have a non-randomized test. Otherwise we may need to consider $\gamma \in (0,\, 1)$ making the test a randomized one. †

Example 3.11. Let $X_1,\, X_2,\, \ldots,\, X_n \overset{i.i.d.}{\sim} Bernoulli(\theta)$. Let $0 < \theta_1 < \theta_0 < 1$ be two real numbers. We are interested to test $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$. Let $T = \sum_{i=1}^{n} X_i \sim Bin(n,\, \theta_0)$ under H_0. Now,

$$\frac{L(\theta_1)}{L(\theta_0)} = \left(\frac{\theta_1}{\theta_0} \times \frac{1-\theta_0}{1-\theta_1}\right)^t > k \iff t < k_1,$$

for some constant k_1. It see it, notice that

$$g(t) = \left(\frac{\theta_1}{\theta_0} \times \frac{1-\theta_0}{1-\theta_1}\right)^t$$

is a decreasing function of t for $\theta_0 > \theta_1$. Hence, the MP level α test is given by

$$\psi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } t < k_1 \\ \gamma & \text{if } t = k_1 \\ 0 & \text{if } t > k_1, \end{cases}$$

where $E_{\theta_0}(\psi(\boldsymbol{X})) = P_{\theta_0}(T < k_1) + \gamma P_{\theta_0}(T = k_1) = \alpha$. Let $\widetilde{K} \in \{1,\, 2,\, \ldots,\, n\}$ be such that

$$P_{\theta_0}\left(T < \widetilde{K}\right) \leq \alpha < P_{\theta_0}\left(T \leq \widetilde{K}\right).$$

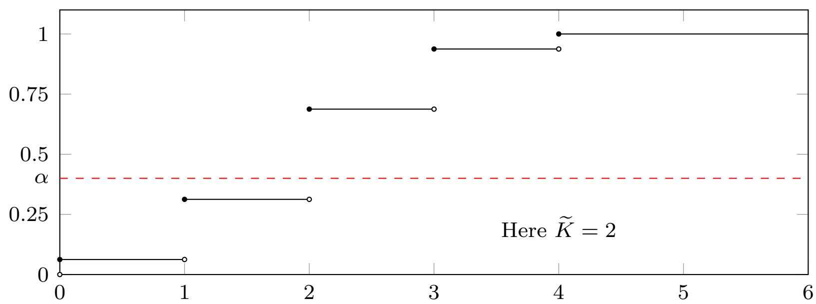

Figure 3.2: CDF of Bin(4,0.5), $\alpha = 0.4$

Take $k_1 = \widetilde{K}$ and $\gamma = \dfrac{\alpha - P_{\theta_0}\left(T < \widetilde{K}\right)}{P_{\theta_0}\left(T = \widetilde{K}\right)}$. The MP level α test is given by

$$\psi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } t < \widetilde{K} \\ \frac{\alpha - P_{\theta_0}\left(T<\widetilde{K}\right)}{P_{\theta_0}\left(T=\widetilde{K}\right)} & \text{if } t = \widetilde{K} \\ 0 & \text{if } t > \widetilde{K}. \end{cases} \qquad ||$$

Example 3.12. Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} U(0, \theta)$. Let $\theta_0 > \theta_1 > 0$ be two fixed real numbers. We are interested to test $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$. Now,

$$\frac{L(\theta_1)}{L(\theta_0)} = \begin{cases} \left(\frac{\theta_0}{\theta_1}\right)^n & \text{if } x_{(n)} < \theta_1 \\ 0 & \text{if } \theta_1 \leq x_{(n)} < \theta_0. \end{cases}$$

It shows that the ratio $\dfrac{L(\theta_1)}{L(\theta_0)}$ is a non-increasing function of $x_{(n)}$. Hence, $\dfrac{L(\theta_1)}{L(\theta_0)} > k$ if and only if $x_{(n)} < k_1$, for some constant k_1. Therefore, the MP level α test is given by

$$\psi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } x_{(n)} < k_1 \\ 0 & \text{otherwise,} \end{cases}$$

where k_1 is such that

$$E_{\theta_0}\left(\psi(\boldsymbol{X})\right) = \alpha \iff P\left(X_{(n)} > k_1\right) = \alpha \iff k_1 = \theta_0 \alpha^{\frac{1}{n}}. \qquad ||$$

3.5 One-sided Composite Alternative

Suppose that the null hypothesis $H_0 : \theta = \theta_0$ is simple, but the alternative hypothesis is of the form $H_1 : \theta > \theta_0$ or $H_1 : \theta < \theta_0$. In such cases, the alternative hypothesis is called one-sided. The alternative hypothesis $H_1 : \theta > \theta_0$ is upper-sided composite hypothesis, and $H_1 : \theta < \theta_0$ is lower-sided composite hypothesis. In this section, we will discuss discuss mainly two methods to construct UMP test for such cases.

3.5.1 UMP Test via Neyman-Pearson Lemma

Though the Neyman-Pearson lemma is for simple null vs. simple alternative, it can be used to find the UMP test for simple null vs. one-sided alternative in some situations. We will explain this method with the help of the following example.

Example 3.13. Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$, where σ is known. Let μ_0 be a real number. We are interested to test $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$. Note that here the alternative is not a simple. However, we can use NP lemma to find UMP level α test. Let $\mu_1 > \mu_0$ be a real number. Then, we know that the MP level α test for $H_0 : \mu = \mu_0$ vs. $H_1 : \mu = \mu_1$ is given by

$$\psi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \frac{\sqrt{n}}{\sigma}\left(\overline{x} - \mu_0\right) > z_\alpha \\ 0 & \text{otherwise.} \end{cases}$$

Please revisit Example 3.10. Note that this test does not depend on μ_1 as long as $\mu_1 > \mu_0$. Hence, same test is UMP level α test for $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$. $||$

3.5.2 UMP Test via Monotone Likelihood Ratio Property

Definition 3.13 (Monotone Likelihood Ratio)**.** *A family of distributions* $\{f(x, \theta) : \theta \in \Theta\}$ *with* $\Theta \subseteq \mathbb{R}$ *is said to have a monotone likelihood ratio (MLR) property in a real valued statistic* $T(\boldsymbol{X})$ *if for any* $\theta < \theta^*$*, the ratio* $\dfrac{L(\theta^*, \boldsymbol{x})}{L(\theta, \boldsymbol{x})}$ *is a nondecreasing function of* $T(\boldsymbol{x})$*.*

Remark 3.6. In the previous definition, $\dfrac{L(\theta^*, \boldsymbol{x})}{L(\theta, \boldsymbol{x})}$ should be nondecreasing or nonincreasing. If the likelihood ratio $\dfrac{L(\theta^*, \boldsymbol{x})}{L(\theta, \boldsymbol{x})}$ is nonincreasing instead of nondecreasing, the effect would be felt in the placement of rejection region. †

Example 3.14. Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$, where $\mu \in \mathbb{R}$ is unknown, but $\sigma > 0$ is known. Then for arbitrary real numbers $\mu^* > \mu$,

$$\frac{L(\mu^*, \boldsymbol{x})}{L(\mu, \boldsymbol{x})} = \exp\left[(\mu^* - \mu)\frac{T(\boldsymbol{x})}{\sigma^2} + \frac{n}{2\sigma^2}\left(\mu^2 - \mu^{*2}\right)\right],$$

where $T(\boldsymbol{x}) = \sum_{i=1}^n x_i$. Now, it is easy to see that the likelihood ratio is am increasing function in T. Therefore, we have a MLR (increasing) property in T. ||

Example 3.15. Let $X_1, X_2, \ldots, X_n$ be a RS from a exponential distribution with mean $\frac{1}{\lambda}$. Then, for arbitrary positive constants $\lambda^* > \lambda$,

$$\frac{L(\lambda^*, \boldsymbol{x})}{L(\lambda, \boldsymbol{x})} = \left(\frac{\lambda^*}{\lambda}\right)^n \exp\left[-(\lambda^* - \lambda)T(\boldsymbol{x})\right],$$

where $T(\boldsymbol{x}) = \sum_{i=1}^n x_i$. Now, as the likelihood ratio is a decreasing function in T, we have MLR (decreasing) property in T. ||

Theorem 3.2. *Let* $X_1, X_2, \ldots, X_n$ *be a RS from a PMF/PDF* $f(x, \theta)$ *with MLR (nondecreasing) property in* $T(\boldsymbol{x})$*, where* $\theta \in \Theta \subseteq \mathbb{R}$*. For testing* $H_0 : \theta \le \theta_0$ *against* $H_1 : \theta > \theta_0$*, there exists a UMP level* α *test, which is given by*

$$\psi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } T(\boldsymbol{x}) > k \\ \gamma & \text{if } T(\boldsymbol{x}) = k \\ 0 & \text{if } T(\boldsymbol{x}) < k, \end{cases}$$

where k *and* γ *are such that* $E_{\theta_0}(\psi(\boldsymbol{X})) = \alpha$*.*

Proof: The proof is skipped. □

Example 3.16. Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$, where μ is unknown, but $\sigma > 0$ is known. Let us consider the testing of $H_0 : \mu \le \mu_0$ against $H_1 : \mu > \mu_0$. In this case, we have MLR (nondecreasing) property in $T = \sum_{i=1}^n X_i$. The UMP level α test is given by

$$\psi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } T(\boldsymbol{x}) > k \\ 0 & \text{otherwise,} \end{cases}$$

where k is such that $P_{\mu_0}(T(\boldsymbol{X}) > k) = \alpha$. Solving for k, we have

$$\psi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \frac{\sqrt{n}}{\sigma}(\bar{x} - \mu_0) > z_\alpha \\ 0 & \text{otherwise.} \end{cases}$$

||

Notice that we have following four different combinations of MLR (nondecreasing or nonincreasing) property and null hypothesis ($H_0 : \theta \leq \theta_0$ or $H_0 : \theta \geq \theta_0$).

	$H_0 : \theta \leq \theta_0$	$H_0 : \theta \geq \theta_0$
MLR (nondecreasing)	Case I	Case II
MLR (nonincreasing)	Case III	Case IV

Of course, for each null hypothesis, the alternative is $\theta \in \mathbb{R} - \Theta_0$. The previous theorem provides the UMP test for Case I. For other three combinations, the UMP tests can be obtained in an almost similar fashion and are mentioned below.

Theorem 3.3. *Let $X_1, X_2, \ldots, X_n$ be a RS from a PMF/PDF $f(x, \theta)$ with MLR (nondecreasing) property in $T(\boldsymbol{x})$, where $\theta \in \Theta \subseteq \mathbb{R}$. For testing $H_0 : \theta \geq \theta_0$ against $H_1 : \theta < \theta_0$, there exists a UMP level α test, which is given by*

$$\psi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } T(\boldsymbol{x}) < k \\ \gamma & \text{if } T(\boldsymbol{x}) = k \\ 0 & \text{if } T(\boldsymbol{x}) > k, \end{cases}$$

where k and γ are such that $E_{\theta_0}(\psi(\boldsymbol{X})) = \alpha$.

Theorem 3.4. *Let $X_1, X_2, \ldots, X_n$ be a RS from a PMF/PDF $f(x, \theta)$ with MLR (nonincreasing) property in $T(\boldsymbol{x})$, where $\theta \in \Theta \subseteq \mathbb{R}$. For testing $H_0 : \theta \leq \theta_0$ against $H_1 : \theta > \theta_0$, there exists a UMP level α test, which is given by*

$$\psi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } T(\boldsymbol{x}) < k \\ \gamma & \text{if } T(\boldsymbol{x}) = k \\ 0 & \text{if } T(\boldsymbol{x}) > k, \end{cases}$$

where k and γ are such that $E_{\theta_0}(\psi(\boldsymbol{X})) = \alpha$.

Theorem 3.5. *Let $X_1, X_2, \ldots, X_n$ be a RS from a PMF/PDF $f(x, \theta)$ with MLR (nonincreasing) property in $T(\boldsymbol{x})$, where $\theta \in \Theta \subseteq \mathbb{R}$. For testing $H_0 : \theta \geq \theta_0$ against $H_1 : \theta < \theta_0$, there exists a UMP level α test, which is given by*

$$\psi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } T(\boldsymbol{x}) > k \\ \gamma & \text{if } T(\boldsymbol{x}) = k \\ 0 & \text{if } T(\boldsymbol{x}) < k, \end{cases}$$

where k and γ are such that $E_{\theta_0}(\psi(\boldsymbol{X})) = \alpha$.