Statistical Inference and Multivariate Analysis (MA324)

Lecture Slides
Lecture 18

Best Test

Indian Institute of Technology Guwahati

Jan-May 2023

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.

Size and Level of a Test

Def: **[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

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

Def: **[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 1: Let $X_1, X_2, \ldots, X_9 \overset{i.i.d.}{\sim} N(\theta, 1)$. Suppose that we 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 $R_2 = \left\{\boldsymbol{x} \in \mathbb{R}^9 : \bar{x} > 7\right\}$ 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

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

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

Test Function

Def: [**Critical or Test Function**] A function $\psi : \mathcal{X}^n \to [0,\, 1]$ 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, $\mathcal{X}^n$ is the sample space of the random sample of size n.

Example 2: Let $X_1, X_2, \ldots, X_9 \overset{i.i.d.}{\sim} N(\theta, 1)$. Suppose that we want to test $H_0 : \theta = 5.5$ against $H_1 : \theta = 7.5$. Let us consider two critical regions $R_1 = \left\{\boldsymbol{x} \in \mathbb{R}^9 : \bar{x} > 6\right\}$ and $R_2 = \left\{\boldsymbol{x} \in \mathbb{R}^9 : \bar{x} > 7\right\}$. 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

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

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

Power Function

- 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.

Def: [**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.

Def: [**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: 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 $P(X = 2$ 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 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 $E_{p=\frac{1}{4}}(\psi(X)) = 1 \times \frac{1}{64} + \frac{1}{9} \times \frac{9}{64} = \frac{1}{32}$. 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: 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: 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.

Uniformly Most Powerful Test

Def: [**Uniformly Most Powerful Test**] Consider the collection $\mathcal{C}_\alpha$ 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 $\mathcal{C}_\alpha$ 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 $\mathcal{C}_\alpha$.

If the alternative hypothesis is simple (that means that Θ_1 is singleton), the test is called **most powerful (MP) level α test**.

Remark: Note that here we are **putting a bound on probability of type-I 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: 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.