Statistical Inference (MA862)

Lecture Slides

Topic 4: Hypothesis Testing

Testing of Hypothesis

- We have discussed point and interval estimation, where we try to find meaningful guess for unknown parameters.
- In testing of hypothesis, we do not guess the value. We try to check if some statement is true or not.

Example: Cherry Blossom Race

Example 4.1:

- 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.
- Question is: Were runners faster in 2012? Of course answer should be yes or no.
- We assume that it is not possible to have the running times of all the participants in 2012.
- How can we proceed?
- Take a random sample of size n from the 2012 runners, and denote the running time by $X_1, X_2, \ldots, X_n$.

Example: Cherry Blossom Race

- Let us also assume that the distribution of the running time is a normal.
- Let us also assume that the variance of the normal distribution is 373 (a value found by analysing the original data).
- We are given i.i.d. random variables $X_1, X_2, \ldots, X_n$ and we want to know if $X_1 \sim N(103.5, 373)$.
- This is a problem of testing of 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.

Example: Cherry Blossom Race

- 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.
- The only thing that is not fixed is $\mu = E(X_1)$.
- We want to test: Is $\mu = 103.5$ or $\mu < 103.5$?
- By modeling assumptions we have reduced the number of ways the hypothesis $X_1 \sim N(103.5, 373)$ may be rejected.
- The only way it can be rejected is if $X_1 \sim N(\mu, 373)$ for some $\mu < 103.5$.
- We compare an expected value to a fixed reference number (here 103.5).

Example: Cherry Blossom Race

- Simple heuristic would be: If $\overline{X} < 103.5$ then $\mu < 103.5$.
- 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. Of course to do that we need to take into account the size of fluctuations of $\overline{X}$.

Example: Clinical Trail

Example 4.2:

- Pharmaceutical companies use hypothesis testing to test if a new drug is efficient.
- To do so, they administer a drug to a group of patients (test group) and placebo to another group (control group).
- Assume that the drug is a cough syrup.
- Let μ_1 denotes the expected number of expectorations per hour after a patient has used placebo.
- Let μ_2 denotes the expected number of expectorations per hour after a patient has used the syrup.
- We want to know if $\mu_2 < \mu_1$.
- Two expectations are compared. No reference number.

Example: Clinical Trail

- Let $X_1, X_2, \ldots, X_{n_1}$ denote n_1 i.i.d. RVs with distribution $P(\mu_1)$.
- Let $Y_1, \ldots, Y_{n_2}$ denote n_2 i.i.d. RVs with distribution $P(\mu_2)$.
- We want to test if $\mu_2 = \mu_1$ or $\mu_2 < \mu_1$.
- Heuristic: We should compare $\overline{X}$ and $\overline{Y}$.

Example: Coin Toss

Example 4.3: A coin is tossed 80 times, and head are obtained 55 times. Can we conclude that the coin is significantly fair?

- Here $n = 80$, $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} Bernoulli(p)$.
- We want to test $p = 0.5$ or $p \neq 0.5$.
- $\overline{X} = 55/80 = 0.6875$.
- If p is actually equal to 0.5, using CLT we have

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

- The observed value of $T_n = 3.3541$.
- Conclusion: It seems quite reasonable to reject the hypothesis $p = 0.5$, as the observed value of T_n is too extreme with respect to a standard normal distribution.

Example: Coin Toss

Example 4.4: 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 $T_n = -1.1180$.
- Conclusion: 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 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

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

  What value of C should we choose?
- This will be considered as we proceed.

Some Definitions

Definition 4.1: A hypothesis is a statement about the unknown parameter(s).

Definition 4.2: Suppose that one wants to choose between two reasonable hypotheses $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$, where $\Theta_0 \subset \Theta$, $\Theta_1 \subset \Theta$ and $\Theta_1 \cap \Theta_2 = \emptyset$. We call H_0 and H_1 are null hypothesis and alternative hypothesis, respectively.

Remark 4.1: The aim here 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 careful about these two hypotheses.

Approach: As illustrated in the examples, we will consider a reasonable statistic and make the choice based on the statistic.

Some Definitions

Definition 4.3: Let R be a subset of χ^n (sample space of the corresponding random sample) such that we reject H_0 if $\boldsymbol{x} \in R$. Then R is called **rejection region** or **critical region**. R^c is called **acceptance region**.

Definition 4.4: The error committed by rejecting H_0 when it is actually true is called **Type-I Error**. Error committed by accepting H_0 when it is actually false is called **Type-II Error**.

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

Aim: To choose R such that probabilities of errors are as small as possible.

Example

Example 4.5: 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 $H_1 : \theta = 7.5$. Let use consider two critical regions $R_1 = \{\boldsymbol{x} \in \mathbb{R}^9 : \overline{x} > 6\}$ and $R_2 = \{\boldsymbol{x} \in \mathbb{R}^9 : \overline{x} > 7\}$. Let us compute the probability of errors. For R_1,

$$P\left(\text{Type-I Error}\right) = P_{\theta=5.5}\left(\overline{X} > 6\right) = 1 - \Phi\left(3(6 - 5.5)\right) = 0.06681.$$

$$P\left(\text{Type-II Error}\right) = P_{\theta=7.5}\left(\overline{X} \leq 6\right) = \Phi\left(3(6 - 7.5)\right) \sim 0.$$

Similarly the probabilities for R_2 can be computed and given in following table.

	R_1	R_2
P(Type-I)	0.06681	0
P(Type-II)	0	0.06681

Some remarks

Remark 4.2:

- Note that in the previous example $R_2 \subset R_1$.
- If we take $R = \emptyset$, then $P(\text{Type-I error}) = 0$ and $P\left(\text{Type-II error}\right) = 1$.
- If we take $R = \mathbb{R}^n$, then $P(\text{Type-I error}) = 1$ and $P\left(\text{Type-II error}\right) = 0$.
- If we try to reduce probability of one error, probability of the other one increases.
- In this type of optimization problem people can use some combination of two functions and then try to minimize the combination.
- However for hypothesis testing the **approach** is as follows: Put a bound on the probability of Type-I error and try to minimize the probability of Type-II error.

Power Function

Definition 4.5: 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 θ is the true value of the parameter, i.e.,

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

Remark 4.3:

- For $\theta \in \Theta_0$, $\beta(\cdot)$ is the probability of Type-I error.
- For $\theta \in \Theta_1$, $\beta(\cdot)$ is one minus probability of Type-II error.

Example 4.6: 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 $H_1 : \theta = 7.5$. Let us consider the critical region $R_2 = \{\boldsymbol{x} \in \mathbb{R}^9 : \overline{x} > 7\}$. The power function of the critical region R_2 is given by

$$\beta(\theta) = 1 - \Phi(21 - 3\theta) \text{ for } \theta = 5.5, 7.5.$$

Size and Level

Definition 4.6: Let $\alpha \in (0, 1)$ be a fixed real number. A test for $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$ with power function $\beta(\cdot)$ is called a size α test if $\sup_{\theta \in \Theta} \beta(\theta) = \alpha$.

Definition 4.7: A test is called level α test if $\beta(\theta) \leq \alpha$ for all $\theta \in \Theta_0$.

Remark 4.4:

- 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.

Test Function

Definition 4.8: A function $\psi : \mathscr{X}^n \to [0, 1]$ is called a critical function or test function, where $\psi(\boldsymbol{x})$ stands for the porbability 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 4.7: 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 $H_1 : \theta = 7.5$. Let use consider two critical regions $R_1 = \{\boldsymbol{x} \in \mathbb{R}^9 : \overline{x} > 6\}$ and $R_2 = \{\boldsymbol{x} \in \mathbb{R}^9 : \overline{x} > 7\}$. The critical regions R_1 and R_2, respectively, can be expressed as the test functions

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

Test Function

Remark 4.5:

- Last example shows that test function is an alternative way of writing critical region.
- Then what does we gain by defining test function? To discuss this, first note that if $\psi(\mathbf{x})$ is a probability, then why should we restrict only 0 and 1? We can consider other values between 0 and 1. And this is the gain we have. To illustrate it consider the next example.

Definition 4.9: The power function of a test function is defined by $\beta(\theta) = E_\theta\left(\psi(\mathbf{X})\right)$ for all $\theta \in \Theta_0 \cup \Theta_1$.

Randomized Test

Definition 4.10: A test is called randomized test if $\psi(\boldsymbol{x}) \in (0, 1)$ for some $\boldsymbol{x}$ in the sample space. Otherwise, it is called a non-randomized test.

Remark 4.6:

- Any test that is given by a critical region is a non-randomized test as the test function in this case is indicator function of the critical region.
- 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 occurs, we reject H_0, otherwise we accept it.

Example

Example 4.8: Let X be a sample of size one form a $Bin(3, p)$ distribution. We want to check if $H_0 : p = 1/4$ against $H_1 : p = 3/4$. The probabilities of occurring different values of 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? It is very easy to see that there is not critical region of size $\frac{1}{32}$.

Example

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

Most Powerful Test

Definition 4.11: Consider the collection $\mathscr{C}$ of all level α tests for $H_0 : \theta \in \Theta_0$ against $H_1 : \theta \in \Theta_1$. A test belonging to $\mathscr{C}$ with power function $\beta(\cdot)$ is called **uniformly most powerful** (UMP) level α test if $\beta(\theta) \geq \beta^*(\theta)$ for all $\theta \in \Theta_1$, where $\beta^*(\cdot)$ is the power function of any other test in $\mathscr{C}$. If the alternative hypothesis is simple (that means that Θ_1 is singleton), the test is called **most powerful** (MP) level α test.

Remark 4.7:

- 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.
- When $H_1 : \theta = \theta_1$ for some fixed θ_1, *i.e.*, H_1 is simple, it boils down to check if $\beta(\theta_1) \geq \beta^*(\theta_1)$. Hence the word 'uniformly' is removed.

Neyman-Pearson Lemma

Theorem 4.1: Let $\theta_0 \neq \theta_1$ be two fixed numbers in Θ. The MP level α test for $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ is given by

$$\psi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } L(\theta_1) > kL(\theta_0) \\ \gamma & \text{if } L(\theta_1) = kL(\theta_0) \\ 0 & \text{if } L(\theta_1) < kL(\theta_0), \end{cases}$$

where $k \geq 0$ and $\gamma \in [0, 1]$ such that $\beta(\theta_0) = E_{\theta_0}(\psi(\boldsymbol{X})) = \alpha$. Here $L(\cdot)$ is the likelihood function.

Remark 4.8:

- In the theorem, both null and alternative are simple.
- $L(\theta_1) > kL(\theta_0)$ can be expressed as $L(\theta_1)/L(\theta_0) > k$ if $L(\theta_0) > 0$. Hence, the MP test rejects the null hypothesis for large values of the ratio.

Examples

Example 4.9: Let $X_1, X_2, \ldots, X_n \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$, where σ is known. Let $\mu_0 < \mu_1$ be two real numbers. We are interested to test $H_0 : \mu = \mu_0$ against $H_1 : \mu = \mu_1$, where $\mu_0 \neq \mu_1$. Then, the MP level α test is given by

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

Remark 4.9:

- The test is quite intuitive in the sense that we reject H_0 if sample mean is large.
- The test is based on the sufficient statistic.
- Note the way of solving the problem. We try to simplify $L(\mu_1)/L(\mu_0) > k$ so that we can write it as a condition on a statistic whose distribution under H_0 is known or can be found. If this statistic is a continuous RV, we will have a non-randomized test. Otherwise we may need to consider a randomized test.

Example

Example 4.10: 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 $H_1 : \theta = \theta_1$. Then, the MP level α test is given by

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

where $K \in \{1, 2, \ldots, n\}$ satisfies

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

Example 4.11: Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} U(0, \theta)$. Let $\theta_0 > \theta_1 > 0$ be two real numbers. The MP level α test for testing $H_0 : \theta = \theta_0$ against $H_1 : \theta = \theta_1$ rejects H_0 if and only if $X_{(n)} < \theta_0 \alpha^{\frac{1}{n}}$.

Examples

Example 4.12: 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 $H_1 : \mu > \mu_0$. The UMP level α test for $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$ is

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

Example 4.13: 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. The UMP level α test for testing $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$ does not exist for all $\alpha \in (0, 1)$.

Remark 4.10: However the problem of hypotheses testing $H_0 : \mu = \mu_0$ against $\mu \neq \mu_0$ is practically quite meaningful. Hence, we need some alternative.

Likelihood Ratio Test: Algorithm

- We want to test $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_1$.
- Consider

$$\Lambda(\boldsymbol{x}) = \frac{\sup_{\theta \in \Theta_0} L(\theta, \boldsymbol{x})}{\sup_{\theta \in \Theta_0 \cup \Theta_1} L(\theta, \boldsymbol{x})}.$$

  $\Lambda(\boldsymbol{X})$ is called likelihood ratio test statistic.
- Likelihood ratio level α test is given by

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

  where γ and k are such that $E_\theta\left(\psi(\boldsymbol{X})\right) = \alpha$ for all $\theta \in \Theta_0$.

Likelihood Ratio Test: Discussion

- $\sup_{\theta \in \Theta_0} L(\theta, \boldsymbol{x})$ can be considered as the max value of the likelihood function (probability for DRVs) over Θ_0 when $\boldsymbol{X} = \boldsymbol{x}$ is observed.
- Similarly, $\sup_{\theta \in \Theta_0 \cup \Theta_1} L(\theta, \boldsymbol{x})$ can be considered as the max value of the likelihood function over $\Theta_0 \cup \Theta_1$ when $\boldsymbol{X} = \boldsymbol{x}$ is observed.
- Clearly $\Lambda(\boldsymbol{x}) \in [0, 1]$.
- We reject H_0 if Λ is small, as in this case the likelihood under Θ_0 is small compared to that over $\Theta_0 \cup \Theta_1$. This means that the observed values are more likely under Θ_1 than under Θ_0. Hence, we reject H_0.

Example

Example 4.14: 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 $H_1 : \mu \neq \mu_0$. The likelihood ratio level α test is given by

$$\psi(\mathbf{x}) = \begin{cases} 1 & \text{if } \frac{\sqrt{n}|\overline{x} - \mu_0|}{\sigma} > z_{\alpha/2} \\ 0 & \text{otherwise.} \end{cases}$$

Example 4.15: Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$. Let μ_0 be a real number. We are interested to test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$. The likelihood ratio level α test is given by

$$\psi(\mathbf{x}) = \begin{cases} 1 & \text{if } \frac{\sqrt{n}|\overline{x} - \mu_0|}{s} > t_{n-1;\alpha/2} \\ 0 & \text{otherwise.} \end{cases}$$

Example

Example 4.16: Let $X_1, X_2, \ldots, X_n \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$. Let σ_0 be a positive real number. We are interested to test $H_0 : \sigma^2 = \sigma_0^2$ vs. $H_1 : \sigma^2 \neq \sigma_0^2$. Hence the likelihood ratio level α test is given by

$$\psi(\boldsymbol{x}) = \begin{cases} 1 & \text{if } \frac{(n-1)s^2}{\sigma_0^2} < \chi^2_{n-1;1-\alpha/2} \text{ or } \frac{(n-1)s^2}{\sigma_0^2} > \chi^2_{n-1;\alpha/2} \\ 0 & \text{otherwise.} \end{cases}$$

p-value

Definition 4.12: For varying level α, assume that the test is a non-randomized test with critical region R_α. The test is called nested if

$$R_\alpha \subset R_{\alpha'} \quad \text{for all } \alpha < \alpha'.$$

Definition 4.13: The p-value of a nested test is defined by

$$\widehat{p} = \widehat{p}(\boldsymbol{X}) = \inf \left\{ \alpha \in [0, 1] : \boldsymbol{X} \in R_\alpha \right\}.$$

Remark 4.11:

- The p-value provides an idea of how strong the data contradict the null hypothesis.
- It also enables other to reach a verdict based on the level of their choice.
- If p-value is **smaller** than α, we **reject** the null hypothesis. Otherwise, we accept the null hypothesis.

Example

Example 4.17: Let $X_1, X_2, \ldots, X_n$ be a RS form a population having normal distribution with unknown mean μ and known variance σ^2. Consider $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$. The critical region of likelihood ratio level α test is given by

$$R_\alpha = \left\{ \boldsymbol{x} \in \mathbb{R}^n : \sqrt{n}\frac{|\overline{x} - \mu_0|}{\sigma} > z_{\frac{\alpha}{2}} \right\}.$$

The test is a nested test, and hence, we can talk about p-value. Using the fact that $\Phi(\cdot)$ is a strictly increasing function, one can show that the p-value of the test is

$$\widehat{p}(\boldsymbol{X}) = 2\left[1 - \Phi\left(\sqrt{n}\frac{|\overline{X} - \mu_0|}{\sigma}\right)\right].$$