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.