Mathematical Statistics and Data Analysis

Lecture 9: Hypothesis Testing

Lyu Ni

DaSE@ECNU
(lni@dase.ecnu.edu.cn)

December 2, 2019

Outlines

Reading Material

Textbook:

- Rice: Chapter 9;
- Mao: Chapter 7.1, 7.2, 7.3, 7.4, 6.6;

Hypothesis Testing

Example: The Lady Tasting Tea

- Lady Ottoline claimed that she was able to point out that the server had poured milk first or tea first. This means she could distinguish
  - TM: Tea first and then Milk;
  - MT: Milk first and then Tea.
- There is a hypothesis:

$$H : \text{This lady was not able to distinguish TM and MT.}$$

- The experiment was designed as follows:
  - Prepared 8 cups: 4 cups for TM and 4 cups for MT;
- Result: Lady Ottoline identified 8 out of 8 correctly.
- What can you conclude from this experiment?

Example: The Lady Tasting Tea (Con'd)

- Fisher's Idea:
  - Suppose the hypothesis is correct. The probability that Lady Ottoline correctly identified 8 out of 8 is

$$\binom{8}{4}^{-1} = \frac{1}{70} \approx 0.014$$

    which is a small probability.
  - A small probability event is considered to an event that cannot be actually occurred in an experiment.
  - However, this small probability event occurred.
  - This means the hypothesis is **not correct** and we need to **reject the hypothesis**.
  - Therefore, Lady Ottoline was deemed to be able to distinguish TM and MT.

Example: Normal Distribution

- A plant casts a type of alloy.
- The alloy intensity is thought to be distributed as $N(\theta, 16)$, where θ is required to be not less than 110 Pa.
- To guarantee the alloy quality, the plant needs to examine whether the manufacturing process goes wrong, that is, the intensity of the alloy is less than 110 Pa.
- The plant randomly selects 25 pieces of alloy and measures their intensity: $x_1, x_2, \cdots, x_{25}$.
- The sample mean is $\bar{x} = 108.2$ Pa.
- **Problem: Does the manufacturing process go wrong?**

Example: Normal Distribution (Con'd)

Let's analyze this problem as follows:

- It is a **(statistical) hypothesis testing problem**.
  - For example, we are interested in the proposition whether the alloy intensity is less than 110 Pa?
  - It is not a parameter estimation problem.
  - We need to make a decision, that is, the answer is "Yes" or "No".
- Define the hypothesis.
  - For example, the involved parameter spaces are, respectively,

$$\Theta_0 = \{\theta : \theta \geq 110\}, \quad \Theta_1 = \{\theta : \theta < 110\}.$$

  - If the hypothesis is correct, $\theta \in \Theta_0$; otherwise, $\theta \in \Theta_1$.

Example: Normal Distribution (Con'd)

Let's analyze this problem as follows:

- It is a **(statistical) hypothesis testing problem**.
  - For example, we are interested in the proposition whether the alloy intensity is less than 110 Pa?
  - It is not a parameter estimation problem.
  - We need to make a decision, that is, the answer is "Yes" or "No".
- Define the hypothesis.
  - For example, the involved parameter spaces are, respectively,

$$\Theta_0 = \{\theta : \theta \geq 110\}, \quad \Theta_1 = \{\theta : \theta < 110\}.$$

  - If the hypothesis is correct, $\theta \in \Theta_0$; otherwise, $\theta \in \Theta_1$.

Example: Normal Distribution (Con'd)

- Conduct a test via a statistic.
  - For example, $x_1, x_2, \cdots, x_{25} \overset{\text{iid}}{\sim} N(\theta, 16)$ and $\bar{x} = 108.2$;
  - The sample mean $\bar{x}$ is a reasonable statistic since $\bar{x}$ is a complete and sufficient statistic for θ.
  - It is known that $\bar{x} \sim N(\theta, 16/25)$.
  - If $\bar{x}$ is smaller, θ is thought to be smaller, and thus we are more likely to reject the null hypothesis: $H_0 : \theta \geq 110$.
  - Our decision is to reject H_0 if $\bar{x} \leq c$, where c is a constant.
  - Since the sample is random, the decision may be wrong.
  - We would like to minimize the probability that we reject H_0 when H_0 is true.

Hypothesis Testing

Example: Normal Distribution (Con'd)

- Conduct a test via a statistic.
  - Under H_0, the probability that we reject H_0 is

$$\begin{aligned} P(\bar{x} \leq c | \theta \geq 110) &= P\left(\frac{\bar{x}-\theta}{\sqrt{16/25}} \leq \frac{c-\theta}{\sqrt{16/25}} \middle| \theta \geq 110 \right) \\ &= \Phi(1.25 * (c-\theta) | \theta \geq 110) \\ &\leq \Phi(1.25 * (c-110)) \end{aligned}$$

  - Let $\Phi(1.25 * (c-110)) = 0.05$. Then, we obtain $c = \Phi^{-1}(0.05) * 0.8 + 110 \approx 108.684$.
- Make a conclusion: we will reject H_0 since $\bar{x} = 108.2 < 108.684$.

Hypothesis Testing

Remark

- this is a **parametric hypothesis testing** if the parameters are involved in the hypotheses.
- otherwise it is a **nonparametric hypothesis testing**.
  - For example, we would like to test a hypothesis that **the population is a normal distribution**.

Hypothesis Testing

Basic Step 1: Construct hypotheses

Suppose that there is a parametric distribution $\{F(x, \theta), \theta \in \Theta\}$ and the sample is $x_1, x_2, \cdots, x_n$, where Θ is a parameter space.

- Suppose that $\Theta_0 \in \Theta$ and $\Theta_0 \neq \emptyset$. The **null hypothesis** is defined as a proposition $H_0 : \theta \in \Theta_0$.
- Suppose that $\Theta_1 \in \Theta$ and $\Theta_1 \cap \Theta_0 = \emptyset$.
    - The most common choice: $\Theta_1 = \Theta - \Theta_0$.

  The **alternative hypothesis** is defined as a proposition $H_1 : \theta \in \Theta_1$.

Thus, we are interested in a pair of hypotheses that

$$H_0 : \theta \in \Theta_0 \quad \text{vs} \quad H_1 : \theta \in \Theta_1$$

Basic Step 1: Construct hypotheses (Con'd)

- **Simple & Composite**:
  - If $\Theta_0 = \{\theta : \theta = \theta_0\}$, a null hypothesis is called a **simple** null hypothesis; otherwise, a null hypothesis is called a **composite** null hypothesis.
  - The simple null hypothesis could be written as

$$H_0 : \theta = \theta_0$$

- **Two-sided & One-sided**: When $H_0 : \theta = \theta_0$,
  - H_0 vs $H_1' : \theta \neq \theta_0$ is called **two-sided** hypothesis.
  - H_0 vs $H_1'' : \theta < \theta_0$ and H_0 vs $H_1''' : \theta > \theta_0$ are called **one-sided** hypothesis.

Hypothesis Testing

Basic Step 2: Find a test statistic and give a rejection region

- Given the sample $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$, the possible outcomes of a test:
  - Reject the null hypothesis H_0;
  - Fail to reject the null hypothesis H_0;
- The sample space is divided into two disjoint parts:
  - The **rejection region** W: Reject H_0 if the sample
  $$\boldsymbol{x} = (x_1, x_2, \cdots, x_n) \in W;$$
  - The **acceptance region** $\overline{W}$: Fail to reject H_0 if
  $$\boldsymbol{x} = (x_1, x_2, \cdots, x_n) \in \overline{W};$$
- Find a **test statistic** and give a rejection region.

Hypothesis Testing

Basic Step 3: Choose a significance level

Since the sample is random, we may make a right or wrong decision. Two types of error are defined as follows:

- **Type I error**: $\boldsymbol{x} \in W$ when $\theta \in \Theta_0$;
- **Type II error**: $\boldsymbol{x} \in \overline{W}$ when $\theta \in \Theta_1$;

We give two notations for the probabilities:

- The probability of type I error: $\alpha = P\{\boldsymbol{x} \in W | H_0\}$;
- The probability of type II error: $\beta = P\{\boldsymbol{x} \in \overline{W} | H_1\}$;

Basic Step 3: Choose a significance level (Con'd)

Definition

Suppose that there is a testing problem

$$H_0 : \theta \in \Theta_0 \quad \text{vs} \quad H_1 : \theta \in \Theta_1$$

and the rejection region is W. The **power function** is defined as the probability that $\boldsymbol{x} \in W$, that is,

$$g(\theta) = P_\theta(\boldsymbol{x} \in W), \theta \in \Theta = \Theta_0 \cup \Theta_1.$$

Thus, the power function is defined on the parameter space Θ:

- $g(\theta) = \alpha = \alpha(\theta), \theta \in \Theta_0$;
- $g(\theta) = 1 - \beta = 1 - \beta(\theta), \theta \in \Theta_1$;

Basic Step 3: Choose a significance level (Con'd)

Obviously, α and β is a function of θ, that is

$$\begin{cases} \alpha(\theta) = g(\theta), & \theta \in \Theta_0, \\ \beta(\theta) = 1 - g(\theta), & \theta \in \Theta_1. \end{cases}$$

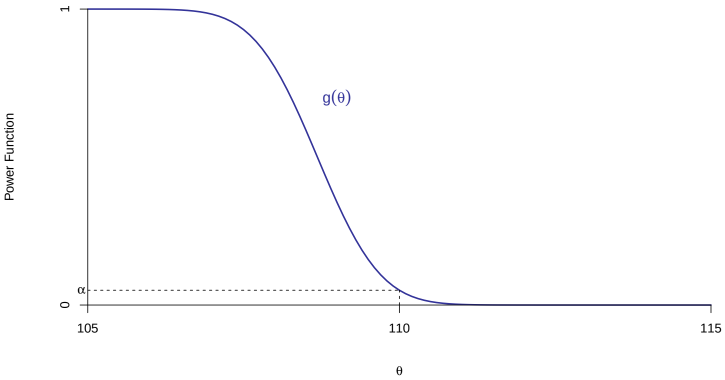
Revisit example: Normal Distribution

The rejection region is defined as $W = \{\bar{x} \le c\}$. The power function is

$$\begin{aligned} g(\theta) &= P_\theta(\bar{x} \le c) = P\left(\frac{\bar{x} - \theta}{4/5} \le \frac{c - \theta}{4/5}\right) \\ &= \Phi\left(\frac{c - \theta}{4/5}\right). \end{aligned}$$

Revisit example: Normal Distribution (Con'd)

The power function is decreasing in θ shown as follows:



Revisit example: Normal Distribution (Con'd)

The probability of Type I error and Type II error are defined as follows:

$$\begin{aligned}\alpha(\theta) &= \Phi\left(\frac{c-\theta}{4/5}\right), \theta \in \Theta_0 \\ \beta(\theta) &= 1-\Phi\left(\frac{c-\theta}{4/5}\right), \theta \in \Theta_1\end{aligned}$$

Remark

- $\alpha\downarrow \Rightarrow c\downarrow \Rightarrow \beta\uparrow$;
- $\beta\downarrow \Rightarrow c\uparrow \Rightarrow \alpha\uparrow$
- There is a tradeoff between α and β.

Hypothesis Testing

Definition

Consider a testing problem

$$H_0 : \theta \in \Theta_0 \quad \text{vs} \quad H_1 : \theta \in \Theta_1.$$

If a test satisfies

$$g(\theta) \leq \alpha$$

for every $\theta \in \Theta_0$, then the test is said to be a **significance test of (significance) level** α

Thumb rule

- $\alpha = 0.05$ is the most common choice;
- Sometimes, $\alpha = 0.1$ or $\alpha = 0.01$ is also useful.

Basic step 4: Give a rejection region

After the significance level α is determined, we can give a rejection region W for the test. For example,

- Given a significance level α, for $\theta \geq 110$,

$$g(\theta) = \Phi\left(\frac{5(c-\theta)}{4}\right) \leq \alpha.$$

- $g(\theta)$ is a decreasing function of θ.
- Just let $g(110) = \alpha$, that is,

$$\Phi\left(\frac{5(c-110)}{4}\right) = \alpha$$

- The rejection region is $W = \{\bar{x} \leq 110 + 0.8 * \Phi^{-1}(\alpha)\}$.

Basic step 5: Make a decision

After the rejection region W is determined, we can make a decision. For example,

- When $\bar{x} \leq 110 + 0.8 * \Phi^{-1}(\alpha)$, we reject H_0;
- When $\bar{x} > 110 + 0.8 * \Phi^{-1}(\alpha)$, we fail to reject H_0.

Summary

Find a significance test in following steps:

- Construct a statistical hypothesis H_0 vs H_1;
- Find an appropriate test statistic $T(\boldsymbol{x})$ of which the distribution is known under H_0;
- Given a significance level α, derive the rejection region W;
- Calculate $T(\boldsymbol{x})$ from the sample $\boldsymbol{x} = (x_1, \cdots, x_n)$ and make a decision by judging whether $T(\boldsymbol{x}) \in W$.

Hypothesis Testing

p value

By determining different significance levels, we may make different conclusion:

Significance level α	Rejection Region W	Conclusion
$\alpha = 0.1$	$\bar{x} \leq 108.975$	Reject H_0
$\alpha = 0.05$	$\bar{x} \leq 108.684$	Reject H_0
$\alpha = 0.025$	$\bar{x} \leq 108.432$	Reject H_0
$\alpha = 0.01$	$\bar{x} \leq 108.139$	Not reject H_0
$\alpha = 0.005$	$\bar{x} \leq 107.939$	Not reject H_0

- If $\alpha = 0.05$ is chosen, H_0 could be rejected;
- If $\alpha = 0.01$ is chosen, H_0 could not be rejected.

Hypothesis Testing

p value

From a different perspective, when $\theta = 110$, the test statistic

$$u = \frac{\bar{x} - \theta}{4/5} \sim N(0, 1).$$

It is calculated that $u_0 = \theta + 0.8 * \Phi^{-1}(\alpha) = -2.25$ from the sample if the significance level $\alpha = 0.05$. The probability is

$$P(u \leq u_0) = P(u \leq -2.25) = \Phi(-2.25) \approx 0.0122$$

- When $\alpha \geq 0.0122$ and $u_\alpha \geq -2.25$, H_0 could be rejected since the rejection region is $W = \{u \leq u_\alpha\}$;
- When $\alpha < 0.0122$ and $u_\alpha < -2.25$, H_0 could be not rejected since the rejection region is $W = \{u \leq u_\alpha\}$;

Hypothesis Testing

Definition

p **value** is defined as the probability under the null hypothesis of a result as or more extreme than that actually observed.

- If $\alpha \geq p$, then reject H_0 at the significance level α;
- If $\alpha < p$, then do not reject H_0 at the significance level α.