Math 408 - Mathematical Statistics

Lecture 29-30. Testing Hypotheses: The Neyman-Pearson Paradigm

April 12-15, 2013

Agenda

- Example: Two Coins Tossing
- General Framework
- Type I Error and Type II Error
- Significance Level
- Power
- Neyman-Pearson Lemma
- Example: the likelihood ratio test for Gaussian variables
- The Concept of p-value
- Summary

Example: Two Coins Tossing

Suppose Bob has two coins:

- Coin "0" has probability of heads $p_0 = 0.5$
- Coin "1" has probability of heads $p_1 = 0.7$

Bob chooses one of the coins, tosses it $n = 10$ times and tells Alice the number of heads, but does not tell her whether it was coin 0 or coin 1.

On the basis of the number of heads, Alice has to decide which coin it was.
How should her decision rule be?

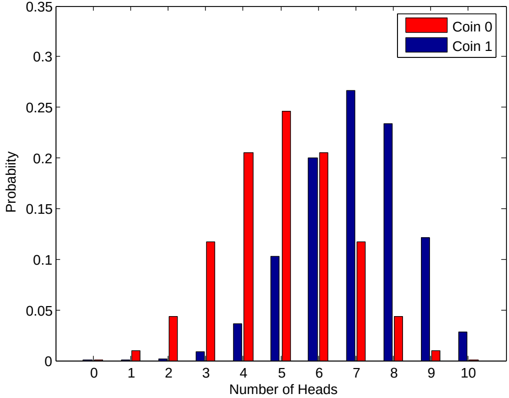
Let X denote the number of heads.

$$X \in \mathcal{X} = \{0, 1, 2, \ldots, 10\}$$

Then for each coin we can compute the probability that Bob got exactly x heads:

$$\mathbb{P}_i(X = x) = \binom{n}{x} p_i^x (1 - p_i)^{n-x}, \qquad i = 0, 1.$$

Example: Two Coins Tossing



Suppose that Bob observed 2 heads. Then $\frac{\mathbb{P}_0(X=2)}{\mathbb{P}_1(X=2)} \approx 30$, and, therefore, coin 0 was about 30 times more likely to produce this result than was coin 1.

On the other hand, if there were 8 heads, then $\frac{\mathbb{P}_0(X=8)}{\mathbb{P}_1(X=8)} \approx 0.19$, which would favor coin 1.

Hypothesis Testing

The example with two coins is an example of **hypothesis testing**:

- The Null Hypothesis H_0: Bob tossed coin 0
- The Alternative Hypothesis H_1: Bob tossed coin 1

Alice would accept H_0 if the likelihood ratio

$$\frac{\mathcal{L}(\text{Data}|\text{Coin } 0)}{\mathcal{L}(\text{Data}|\text{Coin } 1)} = \frac{\mathbb{P}_0(X = x)}{\mathbb{P}_1(X = x)} > 1$$

and she would reject H_0 if the likelihood ratio

$$\frac{\mathcal{L}(\text{Data}|\text{Coin } 0)}{\mathcal{L}(\text{Data}|\text{Coin } 1)} = \frac{\mathbb{P}_0(X = x)}{\mathbb{P}_1(X = x)} < 1$$

In this example, Alice would accept H_0 if

$$x \leq 6$$

and she would reject H_0 if

$$x > 6$$

Hypothesis Testing: General Framework

More formally, suppose that we partition the parameter space Θ into two disjoint sets Θ_0 and Θ_1 and that we wish to test

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

We call H_0 the **null hypothesis** and H_1 the **alternative hypothesis**.

Let X be data and let $\mathcal{X}$ be the range of X. We test a hypothesis by finding an appropriate subset of outcomes $\mathcal{R} \subset \mathcal{X}$ called the **rejection region**. If $X \in \mathcal{R}$ we reject the null hypothesis, otherwise, we do not reject the null hypothesis:

$$X \in \mathcal{R} \Rightarrow \text{ reject } H_0$$
$$X \notin \mathcal{R} \Rightarrow \text{ accept } H_0$$

In the Two Coins Example,

- X is the number of heads
- $\mathcal{X}$ is $\{0, 1, 2, \dots, 10\}$
- $\mathcal{R}$ is $\{7, 8, 9, 10\}$

Hypothesis Testing: General Framework

Usually the rejection region $\mathcal{R}$ is of the form

$$\mathcal{R} = \{x \in \mathcal{X} : T(x) < c\}$$

where T is a **test statistic** and c is a **critical value**. The main problem in hypothesis testing is

to find an appropriate test statistic T and an appropriate cutoff value c

In the Two Coins Example,

- $T(x) = \frac{\mathbb{P}_0(X=x)}{\mathbb{P}_1(X=x)}$ is the likelihood ratio
- $c = 1$

Main Definitions

In hypothesis testing, there are two types of errors we can make:

- Rejecting H_0 when H_0 is true is called a **type I error**
- Accepting H_0 when H_1 is true is called a **type II error**

Definition

- The probability of a type I error is called the **significance level** of the test and is denoted by α

$$\alpha = \mathbb{P}(\text{type I error}) = \mathbb{P}(\text{Reject } H_0 | H_0)$$

- The probability of a type II error is denoted by β

$$\beta = \mathbb{P}(\text{type II error}) = \mathbb{P}(\text{Accept } H_0 | H_1)$$

- $(1 - \beta)$ is called the **power** of the test

$$\textbf{power} = 1 - \beta = 1 - \mathbb{P}(\text{Accept } H_0 | H_1) = \mathbb{P}(\text{Reject } H_0 | H_1)$$

Thus, the **power** of the test is the probability of rejecting H_0 when it is false.

Neyman-Pearson Lemma

Definition

- A hypothesis of the form $\theta = \theta_0$ is called a **simple hypothesis**.
- A hypothesis of the form $\theta > \theta_0$ or $\theta < \theta_0$ is called a **composite hypothesis**.

The Neyman-Pearson Lemma shows that the test that is based on the likelihood ratio (as in the Two Coins Example) is optimal for simple hypotheses:

Neyman-Pearson Lemma

Suppose that H_0 and H_1 are simple hypotheses, $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$. Suppose that the **likelihood ratio test** *that rejects H_0 whenever the likelihood ratio is less than c,*

$$\text{Reject } H_0 \quad \Leftrightarrow \quad \frac{\mathcal{L}(Data|\theta_0)}{\mathcal{L}(Data|\theta_1)} < c$$

has significance level α_{LR}. Then any other test for which the significance level $\alpha \leq \alpha_{\mathrm{LR}}$ has power less than or equal to that of the likelihood ratio test

$$1 - \beta \leq 1 - \beta_{LR}$$

Example

Example

Let $X_1, \ldots, X_n \sim N(\mu, \sigma^2)$, where σ^2 is known. Consider two simple hypotheses:

$$H_0 : \mu = \mu_0$$
$$H_1 : \mu = \mu_1 > \mu_0$$

Construct the likelihood ratio test with significance level α.

Answer:

$$\text{Reject } H_0 \quad \Leftrightarrow \quad \overline{X}_n > \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}$$

- Neyman-Pearson: this test is the most powerful test among all tests with significance level α.

The Concept of p-value

Reporting "reject H_0" or "accept H_0" is not very informative.
For example, if the test just reposts to reject H_0, this does not tell us how strong the evidence against H_0 is. This evidence is summarized in terms of **p-value**.

Definition

Suppose for every $\alpha \in (0,1)$ we have a test of significance level α with rejection region $\mathcal{R}_\alpha$. Then, the p-value is the smallest significance level at which we can reject H_0:

$$p\text{-value} = \inf\{\alpha : X \in \mathcal{R}_\alpha\}$$

Informally, the p-value is a measure of the evidence against H_0:
the smaller the p-value, the stronger the evidence against H_0
Typically, researchers use the following evidence scale:

- p-value < 0.01: very string evidence against H_0
- $0.01 < p$-value < 0.05: strong evidence against H_0
- $0.05 < p$-value < 0.10: weak evidence against H_0
- p-value > 0.10: little or no evidence against H_0

Summary

- In general, we partition the parameter space Θ into two disjoint sets Θ_0 and Θ_1 and test

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

  - H_0 is called the null hypothesis
  - H_1 is called the alternative hypothesis
  - If $H_i : \theta = \theta_i$, then the hypothesis is called simple
- If X is data and $\mathcal{X}$ is the range of X, then we reject $H_0 \Leftrightarrow X \in \mathcal{R} \subset \mathcal{X}$.
  - Rejection region $\mathcal{R} = \{x : T(x) < c\}$
  - For the likelihood ratio test, $T(x) = \frac{\mathbb{P}(X=x|H_0)}{\mathbb{P}(X=x|H_1)}$
- Type I Error: Rejecting H_0 when H_0 is true
  - $\alpha = \mathbb{P}(\text{Reject } H_0|H_0)$ is called significance level (small α is good)
- Type II Error: Accepting H_0 when H_1 is true
  - $1 - \beta = 1 - \mathbb{P}(\text{Accept } H_0|H_1)$ is called power (large power is good)
- Neyman-Pearson Lemma: basing the test on the likelihood ratio is optimal.
- p-value summarizes the evidence against the null hypothesis, $p\text{-value} = \inf\{\alpha : X \in \mathcal{R}_\alpha\}$.