Lecture 5

Xiaoping Shi
Department of Statistics, University of Toronto
xpshi@utstat.toronto.edu

July 26, 2013

- Hypothesis test
- The Neyman-Pearson Paradigm
- p-value
- Uniformly Most Powerful Tests

In Bayesian approach, we introduced some concepts about hypothesis test. Now we continue the introduction of the following terminology.

- **Null hypothesis** H_0 v.s. **alternative hypothesis** H_1 or H_a.

  For example, given a sample data, we concern about the value of parameter and then make null hypothesis and alternative hypothesis about the possible values of the parameter. The decision will be made by sample data to support H_0 (accept H_0 or reject H_1) or support H_1 (accept H_1 or reject H_0).

- Since in practice, we do not know H_0 is true or H_1 is true. After our decision, we may make two errors: **Type I error** and **Type II error**.

- Type I error: when H_0 is true, our decision is rejecting H_0 or accepting H_1.
  The probability of a type I error is called the **significance level** of the test and is usually denoted by α.

$$\alpha = P(\text{rejecting } H_0 | H_0)$$

- Type II error: when H_1 is true, our decision is rejecting H_1 or accepting H_0.
  The probability of a type II error is usually denoted by β

$$\beta = P(\text{accepting } H_0 | H_1)$$

  The **power** of the test is equal to $1 - \beta$, i.e.

$$P(\text{rejecting } H_0 | H_1)$$

To handle the two errors, usually we control the probability of Type I error α to be 0.05, e.g.. For the power, the bigger means the statistic is better.

Our decision is based on likelihood ratio: find a **test statistic** based on likelihood ratio.
The set of values of the test statistic that leads to rejection of H_0 is called the **rejection region**, and the set of values that leads to acceptance is called the **acceptance region**.

The probability distribution of the test statistic when the null hypothesis is true is called the **null distribution**.

Example 1 Let $X_1, X_2, \cdots, X_n$ denote a random sample from $N(\theta, 1)$. It is desired to test the hypothesis

$$H_0 : \theta = \theta' = 0 \quad v.s. \quad H_1 : \theta = \theta'' = 1$$

Show the statistic based on likelihood ratio, rejection region and power.

The likelihood ratio is

$$\begin{aligned} \frac{L(\theta')}{L(\theta'')} &= \frac{(1/\sqrt{2\pi})^n \exp[-\sum_{i=1}^n X_i^2/2]}{(1/\sqrt{2\pi})^n \exp[-\sum_{i=1}^n (X_i - 1)^2/2]} \\ &= \exp\left[-\sum_{i=1}^n X_i + \frac{n}{2}\right] \end{aligned}$$

The rejection region is for $k > 0$,

$$\frac{L(\theta')}{L(\theta'')} = \exp\left[-\sum_{i=1}^n X_i + \frac{n}{2}\right] \leq k$$

This inequality holds if and only if

$$-\sum_{i=1}^{n} X_i + \frac{n}{2} \leq \log k$$

or, equivalently,

$$\sum_{i=1}^{n} X_i \geq \frac{n}{2} - \log k = c$$

In this case, the rejection region is the set

$$\sum_{i=1}^{n} X_i \geq c$$

where c is a constant that can be determined so that the probability of Type I error is significance level α.

$$
\begin{aligned}
\alpha &= P(\text{rejecting} \quad H_0|H_0) \\
&= P(\sum_{i=1}^{n} X_i \geq c|\theta = \theta' = 0) \\
&= P(\sum_{i=1}^{n} X_i/\sqrt{n} \geq c/\sqrt{n}|\theta = \theta' = 0) \\
&= \int_{c/\sqrt{n}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-x^2/2\right] dx
\end{aligned}
$$

$$
c/\sqrt{n} = z_\alpha, c = z_\alpha\sqrt{n}
$$

For example $\alpha = 0.05$, $z_\alpha = 1.645$ and $n = 25$, then $c = 1.65\sqrt{25} = 0.329$.

The **power** of the test is

$$\begin{aligned}
\beta &= P(\text{rejecting} \quad H_0|H_1) \\
&= P(\sum_{i=1}^{n} X_i \geq z_\alpha \sqrt{n}|\theta = \theta'' = 1) \\
&= P(\sum_{i=1}^{n}(X_i - 1)/\sqrt{n} \geq z_\alpha - \sqrt{n}|\theta = \theta'' = 1) \\
&= \int_{z_\alpha - \sqrt{n}}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-x^2/2\right] dx
\end{aligned}$$

For example $\alpha = 0.05$, $z_\alpha = 1.645$ and $n = 25$, then $\beta = 0.9996$.

Question: Suppose the random sample is from $N(\theta, 2)$. How to find the statistic based on likelihood ratio, rejection region and power?

Neyman-Pearson Lemma

Suppose that H_0 and H_1 are simple hypotheses and that the test that rejects H_0 whenever the likelihood ratio is less than c and significance level α. Then **any other test** for which the significance level is less than or equal to α has power less than or equal to that of the likelihood ratio test.

$$H_0 : f(\mathbf{X}) = f_0(\mathbf{X}) \text{ v.s. } H_1 : f(\mathbf{X}) = f_1(\mathbf{X})$$

A decision function

$$d(\mathbf{X}) = \begin{cases} 1, & \text{if } f_0(\mathbf{X})/f_1(\mathbf{X}) < c, \text{ reject } H_0 \\ 0, & \text{if } f_0(\mathbf{X})/f_1(\mathbf{X}) \geq c, \text{ accept } H_0 \end{cases}$$

Let $d^*(\mathbf{X})$ be the decision function of another test satisfying

$$E_0(d^*(\mathbf{X})) \leq E_0(d(\mathbf{X})) = \alpha$$

Using the key inequality

$$d^*(\mathbf{X})[c - f_0(\mathbf{X})/f_1(\mathbf{X})] \leq d(\mathbf{X})[c - f_0(\mathbf{X})/f_1(\mathbf{X})]$$

multiplying $f_1(\mathbf{X})$ and integrating both sides w.r.t. $\mathbf{X}$ give

$$E_1(d^*(\mathbf{X})) \leq E_1(d(\mathbf{X}))$$

Example 2 Let $X_1, X_2, \cdots, X_n$ be a random sample from a normal distribution having known variance σ^2. Consider two simple hypothesis:

$$H_0 : \mu = \mu_0 \ \text{ v.s. } \ H_1 : \mu = \mu_1$$

where $\mu_1 > \mu_0$. Let the significance level α be prescribed. Find the most powerful test using Neyman-Pearson Lemma.

The Neyman-Pearson Lemma states that among all tests with significance level α, the test that rejects for small values of the **likelihood ratio** is most powerful. We thus calculate the likelihood ratio statistic, which is

$$\frac{L(\theta')}{L(\theta'')} = \frac{\exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu_0)^2\right]}{\exp\left[\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(X_i - \mu_1)^2\right]} \leq k$$

This inequality holds if and only if

$$\sum_{i=1}^{n} X_i \geq \left[2\sigma^2 \log k - n(\mu_1^2 - \mu_0^2))\right] / (\mu_0 - \mu_1)$$

In this case, the rejection region is the set

$$\sum_{i=1}^{n} X_i \geq c$$

where c is a constant that can be determined so that the probability of Type I error is significance level α.

$$
\begin{aligned}
\alpha &= P(\text{rejecting} \quad H_0 | H_0) \\
&= P(\sum_{i=1}^{n} X_i \geq c | \mu = \mu_0) \\
&= P(\sum_{i=1}^{n} (X_i - \mu_0)/(\sqrt{n}\sigma) \geq (c - n\mu_0)/(\sqrt{n}\sigma) | \mu = \mu_0) \\
&= \int_{(c-n\mu_0)/(\sqrt{n}\sigma)}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left[-x^2/2\right] dx
\end{aligned}
$$

$$
(c - n\mu_0)/(\sqrt{n}\sigma) = z_\alpha, c = z_\alpha \sqrt{n}\sigma + n\mu_0
$$

If $\sum_{i=1}^{n} X_i > z_\alpha \sqrt{n}\sigma + n\mu_0$, then reject H_0, otherwise accept H_0.

Exercise: Consider the case $\mu_1 < \mu_0$.

The theory requires the specification of the significance level, α, in advance of analyzing the data, but gives no guidance about how to make this choice.

In practice it is almost always the case that the choice of α is essentially arbitrary, such as 0.01 and 0.05.
p value is defined to be the probability

$$P(T \geq T_{\text{stat}} | H_0)$$

where T_{stat} is the statistic with some distribution and r.v. T is from the distribution.
If p value is less than or equal to the significance level α

$$P(T \geq T_{\text{stat}} | H_0) \leq \alpha = P(T \geq t_\alpha | H_0)$$

$$T_{\text{stat}} \geq t_\alpha$$

then we reject H_0.

In Example 2,

$$H_0 : \mu = \mu_0 \ \text{ v.s. } \ H_1 : \mu = \mu_1$$

where $\mu_1 > \mu_0$.
The most powerful test is

$$\sum_{i=1}^{n} X_i > z_\alpha \sqrt{n}\sigma + n\mu_0$$

where rejection region depends on μ_0, σ, n, but **not** on μ_1. So it is uniformly most powerful for the testing

$$H_0 : \mu = \mu_0 \ \text{ v.s. } \ H_1 : \mu > \mu_0$$

It should be noted that had the alternative hypothesis been either

$$H_1 : \mu < \mu_0 \quad \text{or} \quad H_1 : \mu > \mu_0$$

called **one-sided alternative**, a uniformly most powerful test would exist in each instance.

There is usually no uniformly most powerful test in the alternative

$$H_1 : \mu \neq \mu_0$$

called **two-sided alternative**.