Testing Hypotheses

March 11-13, 2008

1 Simple Hypotheses

In the simplest set-up for a **statistical hypothesis**, we consider two values $\theta_0, \theta_1 \in \Theta$, the parameter space. We write the test as

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta = \theta_1.$$

H_0 is called the **null hypothesis**. H_1 is called the **alternative hypothesis**. For this hypothesis test, the action space A has two points 0 and 1. As before, the decision function

$$d : \text{data} \to \{0, 1\}.$$

The loss function $\mathcal{L}$ must satisfy

$$\mathcal{L}(\theta_0, 0) \leq \mathcal{L}(\theta_0, 1) \quad \text{and} \quad \mathcal{L}(\theta_1, 1) \leq \mathcal{L}(\theta_1, 0).$$

Without loss of generality, we can take the $0 - 1 - c$ loss function

$$\mathcal{L}(\theta_0, 0) = \mathcal{L}(\theta_1, 1) = 0, \quad \mathcal{L}(\theta_1, 0) = 1 \quad \text{and} \quad \mathcal{L}(\theta_0, 1) = c.$$

This gives a risk function

$$\mathcal{R}(\theta_0, d) = P_{\theta_0}\{d(X) = 1\}, \quad \mathcal{R}(\theta_1, d) = cP_{\theta_1}\{d(X) = 1\}.$$

Typically, we shall choose $c = 1$.

- The action $a = 1$ is called **rejecting the hypothesis**. Rejecting the hypothesis when it is true is called a **type I error**. Its probability $\alpha = P_{\theta_0}\{d(X) = 1\}$ is called the **size of the test**.
- The action $a = 0$ is called **failing to reject the hypothesis**. Failiing to reject the hypothesis when if is false, called a **type II error**, has probability $\beta = P_{\theta_1}\{d(X) = 0\}$. The **power of the test** $1 - \beta = P_{\theta_1}\{d(X) = 1\}$.

Given observations X, the rejection of the hypothesis is based on whether or not X lands in a **critical region** C. Thus,

$$d(X) = 1 \quad \text{if and only if} \quad X \in C.$$

Given a choice α for the size of the test, the choice of decision function d or equivalently, critical region C is called **best** or **most powerful** if for any choice of critical region C^* and corresponding decision function,

$$d^*(\mathbf{x}) = I_{C^*}(\mathbf{x})$$

for a size α test,

$$\beta = P_{\theta_1}\{d(X) = 0\} \le P_{\theta_1}\{d^*(X) = 0\} = \beta^*$$

or in terms of the critical regions.

$$\beta = 1 - P_{\theta_1}\{X \in C\}, \quad \beta^* = 1 - P_{\theta_1}\{X \in C^*\}. \tag{1}$$

and $\beta \le \beta^*$.

2 The Neyman-Pearson Lemma

The Neyman-Pearson lemma tell us that the best test for a simple hypothesis is a **likelihood ratio test**.

Theorem 1 (Neyman-Pearson Lemma)**.** *Let $L(\theta|\mathbf{x})$ denote the likelihood function for the random variable X corresponding to the probability measure $P_\theta, \theta \in \Theta$. If there exists a critical region C of size α and a nonnegative constant k such that*

$$\frac{\mathbf{L}(\theta_1|\mathbf{x})}{\mathbf{L}(\theta_0|\mathbf{x})} \ge k \quad \textit{for } \mathbf{x} \in C$$

and

$$\frac{\mathbf{L}(\theta_1|\mathbf{x})}{\mathbf{L}(\theta_0|\mathbf{x})} \le k \quad \textit{for } \mathbf{x} \notin C,$$

then C is the most powerful critical region of size α.

Proof. Let C^* be a critical region of size less than or equal to α. Let β and β^* denote, respectively, the probability of type II error for the critical regions C and C^* respectively. The theorem is to show that $\beta^* \ge \beta$.

Write C and C^* as the disjoint union.

$$C = (C\backslash C^*) \cup (C \cap C^*), \quad \text{and} \quad C^* = (C^*\backslash C) \cup (C \cap C^*).$$

Thus,

$$\alpha = P_{\theta_0}\{X \in C\} = P_{\theta_0}\{X \in C\backslash C^*\} + P_{\theta_0}\{X \in C \cap C^*\}.$$

and

$$\alpha \ge P_{\theta_0}\{X \in C^*\} = P_{\theta_0}\{X \in C^*\backslash C\} + P_{\theta_0}\{X \in C \cap C^*\}.$$

Consequently,

$$P_{\theta_0}\{X \in C\backslash C^*\} = \alpha - P_{\theta_0}\{X \in C \cap C^*\} \ge P_{\theta_0}\{X \in C\backslash C^*\}. \tag{2}$$

From equation (1), we obtain

$$\beta^* - \beta = P_{\theta_1}\{X \in C\} - P_{\theta_1}\{X \in C^*\} = \int_C \mathbf{L}(\theta_1|\mathbf{x})\ d\mathbf{x} - \int_{C^*} \mathbf{L}(\theta_1|\mathbf{x})\ d\mathbf{x}.$$

Now subtract from both of the integrals the quantity

$$P_{theta_1}\{X \in C \cap C^*\} = \int_{C \cap C^*} \mathbf{L}(\theta_0|\mathbf{x})\ d\mathbf{x}$$

to find that

$$\beta^* - \beta = P_{\theta_1}\{X \in C\backslash C^*\} - P_{\theta_1}\{X \in C^*\backslash C\} = \int_{C\backslash C^*} \mathbf{L}(\theta_1|\mathbf{x})\ d\mathbf{x} - \int_{C^*\backslash C} \mathbf{L}(\theta_1|\mathbf{x})\ d\mathbf{x}. \tag{3}$$

For $\mathbf{x} \in C\backslash C^* \subset C$, $\mathbf{L}(\theta_1|\mathbf{x}) \geq k\mathbf{L}(\theta_0|\mathbf{x})$ and

$$\int_{C\backslash C^*} \mathbf{L}(\theta_1|\mathbf{x})\ d\mathbf{x} \geq k \int_{C\backslash C^*} \mathbf{L}(\theta_0|\mathbf{x})\ d\mathbf{x}. \tag{4}$$

For $\mathbf{x} \in C^*\backslash C \subset C^*$, $\mathbf{L}(\theta_1|\mathbf{x}) \leq k\mathbf{L}(\theta_0|\mathbf{x})$ and

$$\int_{C^*\backslash C} \mathbf{L}(\theta_1|\mathbf{x})\ d\mathbf{x} \leq k \int_{C^*\backslash C} \mathbf{L}(\theta_0|\mathbf{x})\ d\mathbf{x}. \tag{5}$$

Apply inequalities (4) and (5) to inequality (3).

$$\beta^* - \beta \geq k \int_{C\backslash C^*} \mathbf{L}(\theta_0|\mathbf{x})\ d\mathbf{x} - k \int_{C^*\backslash C} \mathbf{L}(\theta_0|\mathbf{x})\ d\mathbf{x} = k \left(\int_{C\backslash C^*} \mathbf{L}(\theta_0|\mathbf{x})\ d\mathbf{x} - \int_{C^*\backslash C} \mathbf{L}(\theta_0|\mathbf{x})\ d\mathbf{x} \right) \geq 0$$

by inquality (2). □

If T is a sufficient statistic, then the likelihood ratio

$$\frac{\mathbf{L}(\theta_1|\mathbf{x})}{\mathbf{L}(\theta_0|\mathbf{x})} = \frac{h(\mathbf{x})g(\theta_1, T(x))}{h(\mathbf{x})g(\theta_0, T(x))} = \frac{g(\theta_1, T(x))}{g(\theta_0, T(x))}$$

depends only on the value of the sufficient statistic and the parameter values.

3 Examples

Example 2. *Let $X = (X_1, \ldots, X_n)$ be independent normal observations with unknown mean and known variance σ^2. The hypothesis is*

$$H_0 : \mu = \mu_0 \quad versus \quad H_1 : \mu = \mu_1.$$

The likelihood ratio

$$\begin{aligned}
\frac{\mathbf{L}(\mu_1|\mathbf{x})}{\mathbf{L}(\mu_0|\mathbf{x})} &= \frac{\frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(x_1-\mu_1)^2}{2\sigma^2} \cdots \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(x_n-\mu_1)^2}{2\sigma^2}}{\frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(x_1-\mu_0)^2}{2\sigma^2} \cdots \frac{1}{\sqrt{2\pi\sigma^2}} \exp -\frac{(x_n-\mu_1)^2}{2\sigma^2}} \\
&= \frac{\exp -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_1)^2}{\exp -\frac{1}{2\sigma^2} \sum_{i=1}^n (x_i - \mu_0)^2} \\
&= \exp -\frac{1}{2\sigma^2} \sum_{i=1}^n \left((x_i - \mu_1)^2 - (x_i - \mu_0)^2 \right) \\
&= \exp -\frac{\mu_0 - \mu_1}{2\sigma^2} \sum_{i=1}^n (2x_i - \mu_1 - \mu_0)
\end{aligned}$$

The likelihood test is equivalent to

$$-(\mu_0 - \mu_1)\sum_{i=1}^{n} x_i \geq k_1,$$

or for some k_α

$$\bar{x} \geq k_\alpha \text{ when } \mu_0 < \mu_1 \quad \text{or} \quad \bar{x} \leq k_\alpha \text{ when } \mu_0 > \mu_1.$$

To determine k_α, note that under the null hypothesis, $\bar{X}$ is $N(\mu_0, \sigma^2/n)$ and

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

is a standard normal. Set z_α so that $P\{Z \geq z_\alpha\} = \alpha$. Then, for $\mu_0 < \mu_1$,

$$\bar{X} \geq \mu_0 + \frac{\sigma}{\sqrt{n}} z_\alpha = k_\alpha.$$

For $\mu_1 < \mu_0$, we have $\bar{X} \leq -k_\alpha$.

The power should

- increase *as a function of $|\mu_1 - \mu_0|$,*
- decrease *as a function of σ^2, and*
- increase *as a function of n.*

In this situation, the type II error probability,

$$\begin{aligned} \beta &= P_{\mu_1}\{X \notin C\} = P_{\mu_1}\{\bar{X} < \mu_0 + \frac{\sigma_0}{\sqrt{n}} z_\alpha\} \\ &= P_{\mu_1}\left\{\frac{\bar{X} - \mu_1}{\sigma_0/\sqrt{n}} < z_\alpha - \frac{|\mu_1 - \mu_0|}{\sigma_0/\sqrt{n}}\right\} = \Phi\left(z_\alpha - \frac{|\mu_1 - \mu_0|}{\sigma_0/\sqrt{n}}\right) \end{aligned}$$

For $\mu_0 = 10$ and $\mu_1 = 5$ and $\sigma = 3$. Consider the 16 observations and choose a level $\alpha = 0.05$ test, then

```
> x
 [1]  8.887753  2.353184 12.123175 10.020566  9.247956  3.711350
 [7] 13.907150  9.079790  8.826202  6.288765 12.120783 10.994228
[13] 12.522522  4.529421  8.191806 10.195854
> mean(x)
[1] 8.937532
```

Then

$$Z = \frac{8.937 - 10}{3/\sqrt{16}} = -1.417.$$

```
> qnorm(0.05)
[1] -1.644854
```

$k_\alpha = -1.645 > -1.417$ and we fail to reject the null hypothesis.

To compute the probability of a type II error, note that for $\alpha = 0.05$,

$$z_\alpha - \frac{|\mu_1 - \mu_0|}{\sigma_0/\sqrt{n}} = 1.645 - \frac{5}{3/\sqrt{16}} = -5.022$$

```
>> pnorm(-5.022)
[1] 2.556809e-07
```

This is called the z-test. If n is sufficiently large, the even if the data are not normally distributed, $\bar{X}$ is well approximated by a normal distribution and, as long as the variance σ^2 is known, the z-test is used in this case.

Example 3 (Bernoulli trials)**.** *Here $X = (X_1, \ldots, X_n)$ is a sequence of Bernoulli trials with unknown success probability θ, the likelihood*

$$\mathbf{L}(\theta|\mathbf{x}) = (1-\theta)^n \left(\frac{\theta}{1-\theta}\right)^{x_1+\cdots+x_n}.$$

For the test

$$H_0 : \theta = \theta_0 \quad \textit{versus} \quad H_1 : \theta = \theta_1$$

the likelihood ratio

$$\frac{\mathbf{L}(\theta_1|\mathbf{x})}{\mathbf{L}(\theta_0|\mathbf{x})} = \left(\frac{1-\theta_1}{1-\theta_0}\right)^n \left(\left(\frac{\theta_1}{1-\theta_1}\right) \Big/ \left(\frac{\theta_0}{1-\theta_0}\right)\right)^{x_1+\cdots+x_n}$$

Consequently, the test is to reject H_0 whenever

$$\sum_{i=1}^n x_i \ge k_\alpha \text{ when } \theta_0 < \theta_1 \quad \text{or} \quad \sum_{i=1}^n x_i \le k_\alpha \text{ when } \theta_0 > \theta_1.$$

Note that under H_0, $\sum_{i=1}^n X_i$ has a $Bin(n, \theta)$ distribution. Thus, in the case $\theta_0 < \theta_1$, we choose k_α so that

$$\sum_{k=k_\alpha}^n \binom{n}{k} \theta_0^k (1-\theta_0)^{n-k} \le \alpha. \tag{6}$$

In genreal, we cannot choose k_α to obtain the sum α. Thus, we take the minimum value of k_α to achieve the inequality in (6).

If $n\theta_0$ is sufficiently large, then, by the central limit theorem, $\sum_{i=1}^n X_i$ has a normal distribution. If we standardize

$$Z = \frac{\bar{X} - \theta_0}{\sqrt{\theta_0(1-\theta_0)/n}}$$

is approximately a standard normal random variable and we perform the z-test as in the previous exercise.

For example, if we take $\theta_0 = 1/2$ and $\theta_1 > 1/2$ and $\alpha = 0.05$, then with 60 heads in 100 coin tosses

$$Z = \frac{0.60 - 0.50}{0.05} = 2.$$

```
> qnorm(0.95)
[1] 1.644854
```

Thus, $k_{0.05} = 1.645 < 2$ and we reject the null hypothesis.