4. Hypothesis Testing

1. Introduction

Let $X = (X_1, \ldots, X_n)'$ be a random sample, and suppose the distribution of X is given by a parametric family $\mathcal{P} = \{\mathrm{P}_\theta | \, \theta \in \Theta\}$. We partition the parameter set Θ as

$$\Theta = \Theta_0 \cup \Theta_1$$

The *null hypothesis* is given by

$$H_0 : \; \theta \in \Theta_0$$

and is tested against the *alternative hypothesis*

$$H_1 : \; \theta \in \Theta_1$$

The null hypothesis H_0 is maintained unless it is rejected in favor of the alternative hypothesis H_1. When Θ_0 and Θ_1 are singleton sets, we say that the hypotheses are *simple*. Otherwise, they are *composite*.

The statistical hypothesis testing is usually based on a *test statistic* τ. According to the value of τ, the state space $\mathcal{X}$ is partitioned as the disjoint union of the *critical region* C and *acceptance region* A, i.e.,

$$\mathcal{X} = C \cup A$$

If $x \in C$, then H_0 is rejected in favor of H_1. If, on the other hand, $x \in A$, then H_0 is continued to be maintainted. A 'test' is thus completely synonymous to a 'critical region'. We will therefore refer to a test with its critical region. The critical region is usually given by

$$C = \{x | \, \tau(x) \geq c\}$$

for some constant c. The constant c is often referred to as the *critical value*.

The *power function* $\pi(\theta)$ of the test C is defined by

$$\pi(\theta) = \mathrm{P}_\theta(C)$$

Moreover,

$$\max_{\theta \in \Theta_0} \pi(\theta)$$

is called the *size* of the test, while the values of π at $\theta \in \Theta_1$ are called the *power* of the test. We define

(a) If $\mathrm{P}_\theta(C_1) > \mathrm{P}_\theta(C_2)$ at $\theta \in \Theta_1$ for tests C_1 and C_2 of the same size, we say that C_1 is *more powerful* than C_2.

(b) If C_* is such that $\mathrm{P}_\theta(C_*) \geq \mathrm{P}_\theta(C)$ at $\theta \in \Theta_1$ for any test C of the same size, the test C_* is said to be *most powerful.*

(c) The test C_* is said to be *uniformly most powerful* (UMP) if for all $\theta \in \Theta_1$ $\mathrm{P}_\theta(C_*) \geq \mathrm{P}_\theta(C)$ for any test C of the same size.

When both the null and alternative hypotheses are simple, we write $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$. Since $\Theta = \{\theta_0, \theta_1\}$ in this case, $\mathcal{P}$ consists of two distributions, which we write simply as

$$\mathrm{P}_{\theta_0} = \mathrm{P}_0 \quad \text{and} \quad \mathrm{P}_{\theta_1} = \mathrm{P}_1$$

for the null and alternative distributions, respectively. Clearly, $\mathrm{P}_0(C)$ and $\mathrm{P}_1(C)$ are the size and power of the test. Notice that $\mathrm{P}_0(C)$ is the probability of rejecting H_0 when it is true. On the other hand, $\mathrm{P}_1(A)$ is the probability of accepting H_0 when H_0 is false (and H_1 is true). Both of $\mathrm{P}_0(C)$ and $\mathrm{P}_1(A)$ are the probabilities of making errors, which we refer to as the *type* I and *type* II errors, respectively.

2. Likelihood Ratio Tests

Assume that both the null and alternative hypotheses are simple, and the distributions P_0 and P_1 are given by the densities (or likelihood functions) $p(x, \theta_0)$ and $p(x, \theta_1)$.

Lemma 1 (Neyman-Pearson) *The test which rejects H_0 when*

$$\lambda(x) = \frac{p(x, \theta_1)}{p(x, \theta_0)} \geq c$$

for a constant c is most powerful.

Proof Let $C_* = \{x|\,\lambda(x) \geq c\}$ for some constant c, and suppose C is any test with the same size as C_*, i.e. $\mathrm{P}_0(C) = \mathrm{P}_0(C_*)$. We need to show $\mathrm{P}_1(C) \leq \mathrm{P}_1(C_*)$. Assume without loss of generality that C and C_* are disjoint. It follows that

$$\begin{aligned} p(x,\theta_1) &\geq c\,p(x,\theta_0) \quad \text{on} \quad C_* \\ p(x,\theta_1) &< c\,p(x,\theta_0) \quad \text{on} \quad C \end{aligned}$$

We therefore have

$$\int_{C_*} p(x,\theta_1)dx \geq c\int_{C_*} p(x,\theta_0)dx \;\text{ i.e., } \mathrm{P}_1(C_*) \geq c\,\mathrm{P}_0(C_*)$$

and

$$\int_{C} p(x,\theta_1)dx \leq c\int_{C} p(x,\theta_0)dx \;\text{ i.e., } \mathrm{P}_1(C) \leq c\,\mathrm{P}_0(C)$$

from which the stated result follows immediately. ■

The test given by Neyman-Pearson lemma

$$C = \{x|\,\lambda(x) \geq c\}$$

for some constant c is called the *likelihood ratio* (LR) test.

Remarks:

(a) We may view $p(x,\theta_1)$ and $p(x,\theta_0)$ as marginal increases of power and size, respectively, when the point x is added to the critical region C. Neyman-Pearson lemma suggests that those points with large power increase per unit size increase be included in C for an optimal test.

(b) Neyman-Pearson lemma tells us only about the *form* of the optimal critical region $C = \{x|\lambda(x) \geq c\}$. The critical value c is to be determined so that $\mathrm{P}_0(C)$ becomes a prescribed size.

(c) For any monotone increasing function f, the test based on $\tau = f \circ \lambda$ is identical to that based on λ. Hence, it is the LR test. Usually, the LR tests are based on monotone increasing transformations of λ whose null distributions are known.

(d) Neyman-Pearson lemma yields the optimal LR test for simple null and simple alternative. For the composite hypotheses, the *generalized* LR test based on the ratio

$$\lambda(x) = \frac{\sup_{\theta \in \Theta_1} p(x, \theta)}{\sup_{\theta \in \Theta_0} p(x, \theta)}$$

is commonly used. Neyman-Pearson lemma does not apply to the generalized LR test. However, the generalized LR test is known to perform well in many different contexts.

Examples

(a) Let $X_1, \ldots, X_n$ be i.i.d. $\mathbf{N}(\mu, 1)$, and consider

$$H_0 : \ \mu = 0 \quad \text{against} \quad H_1 : \ \mu = 1$$

Since both null and alternative hypotheses are simple, we can apply the Neyman-Pearson lemma to obtain a best test. The likelihood ratio is

$$\begin{aligned} \lambda(x) &= \frac{p(x,1)}{p(x,0)} \\ &= \frac{\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}\sum (x_i - 1)^2\right)}{\left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left(-\frac{1}{2}\sum x_i^2\right)} \\ &= \exp\left(\sum x_i - \frac{n}{2}\right) \end{aligned}$$

Define

$$\tau(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i$$

Note that we may write $\tau = f \circ \lambda$ with $f(z) = (\log z + n/2)/\sqrt{n}$. Since f is monotone increasing, the test

$$C = \{x | \, \tau(x) \geq c\}$$

based on τ is the LR test.

We now determine the value of the constant c. Given a prescribed size, say 5%, we can easily find c such that $\mathrm{P}_0(C) = 0.05$. Note that

$$\tau(X) \sim \mathbf{N}(0, 1)$$

under H_0. We know from the $\mathbf{N}(0,1)$ table that

$$\mathrm{P}_0\{x|\ \tau(x) \geq 1.645\} = 0.05$$

and this gives the value of $c = 1.645$.

Next we consider

$$H_0:\ \mu = 0 \quad \text{against} \quad H_1:\ \mu > 0$$

with the composite alternative hypothesis. Note, however, that for any value of $\mu_1 > 0$, the most powerful test for

$$H_0 : \mu = 0 \quad \text{against} \quad H_1:\ \mu = \mu_1$$

is given by $C = \{x|\,\tau(x) \geq c\}$ with $\tau(x) = \sum x_i/\sqrt{n}$ as given above. Therefore, C is the uniformly most powerful test.

(b) Let $X_1, \ldots, X_n$ be i.i.d. $\mathbf{N}(0, \sigma^2)$, and consider

$$H_0 : \sigma^2 = \sigma_0^2 \quad \text{against} \quad H_1:\ \sigma^2 = \sigma_1^2 > \sigma_0^2$$

The likelihood ratio is

$$\begin{aligned}
\lambda(x) &= \frac{\left(\dfrac{1}{2\pi\sigma_1^2}\right)^{\frac{n}{2}} \exp\left(-\dfrac{1}{2\sigma_1^2}\sum x_i^2\right)}{\left(\dfrac{1}{2\pi\sigma_0^2}\right)^{\frac{n}{2}} \exp\left(-\dfrac{1}{2\sigma_0^2}\sum x_i^2\right)} \\
&= \left(\frac{\sigma_0^2}{\sigma_1^2}\right)^{\frac{n}{2}} \exp\left(\frac{1}{2}\left(\frac{1}{\sigma_0^2} - \frac{1}{\sigma_1^2}\right)\sum x_i^2\right)
\end{aligned}$$

Define

$$\tau(x) = \frac{1}{\sigma_0^2}\sum_{i=1}^n x_i^2$$

Note that we may write $\tau = f \circ \lambda$ with a monotone increasing f. Therefore the LR test is given by $C = \{x|\,\tau(x) \geq c\}$.

Now we find the constant c for any prescribed size α. Under H_0,

$$\tau(X) \sim \chi_n^2$$

and therefore, we may easily find the value c such that

$$\mathrm{P}_0\{x|\,\tau(x) \geq c\} = \alpha$$

from the χ^2 table. Such constant c is often denoted by $\chi_n^2(\alpha)$.

We note that for any H_1 with $\sigma_1^2 > \sigma_0^2$, the above LR test C continues to be the test given by the Neyman-Pearson lemma. Hence, C is UMP test for the composite alternative hypothesis

$$H_0 : \ \sigma^2 = \sigma_0^2 \quad \text{against} \quad H_1 : \ \sigma_1^2 > \sigma_0^2$$

Similarly, you can show that for the composite hypothesis

$$H_0 : \ \sigma^2 < \sigma_0^2 \quad \text{against} \quad H_1 : \ \sigma_1^2 > \sigma_0^2$$

the above test C is also UMP test.

(c) Let $X_1, \ldots, X_n$ be i.i.d. $\mathbf{N}(\mu, \sigma^2)$. We first consider

$$H_0 : \ \mu = \mu_0 \quad \text{against} \quad H_1 : \ \mu \neq \mu_0$$

where the null and alternative hypotheses are, respectively, one-way and two-way composite. Thus, we must consider generalized LR test with the following general likelihood ratio

$$\lambda(x) = \frac{\sup\limits_{\mu,\sigma^2} \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2\right)}{\sup\limits_{\sigma^2} \left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2}\sum(x_i-\mu_0)^2\right)}$$

We showed earlier that

$$\hat{\mu}_{\mathrm{ML}} = \bar{x} \quad \text{and} \quad \hat{\sigma}_{\mathrm{ML}}^2 = \frac{1}{n}\sum(x_i-\bar{x})^2$$

maximize the likelihood function. Therefore, we can compute the generalized LR using the above estimates as

$$\lambda(x) \;=\; \frac{\left(\frac{n}{\sum(x_i-\bar{x})^2}\right)^{\frac{n}{2}} \exp\left(-\frac{n}{2\sum(x_i-\bar{x})^2}\sum(x_i-\bar{x})^2\right)}{\left(\frac{n}{\sum(x_i-\mu_0)^2}\right)^{\frac{n}{2}} \exp\left(-\frac{n}{2\sum(x_i-\mu_0)^2}\sum(x_i-\mu_0)^2\right)}$$

$$= \left(\frac{\sum(x_i-\mu_0)^2}{\sum(x_i-\bar{x})^2}\right)^{\frac{n}{2}}$$
$$= \left(1+\frac{n(\bar{x}-\mu_0)^2}{\sum(x_i-\bar{x})^2}\right)^{\frac{n}{2}}$$

Define a monotone increasing transformation τ of λ by

$$\tau(x) = (n-1)\frac{n(\bar{x}-\mu_0)^2}{\sum(x_i-\bar{x})^2}$$

Since τ is a monotone increasing transformation of λ, the generalized LR test is given by $C = \{x|\ \tau(x) \geq c\}$ for a constant c. Let

$$Z = \frac{\sqrt{n}(\bar{X}-\mu_0)}{\sigma} \quad \text{and} \quad W = \frac{\sum(X_i-\bar{X})^2}{\sigma^2}$$

We have under H_0 that

$$Z \sim \mathbf{N}(0,1) \quad \text{and} \quad W \sim \chi^2_{n-1}$$

and they are independent. It therefore follows that

$$\tau(X) = \frac{Z^2}{\dfrac{W}{n-1}} \sim t^2_{n-1} \equiv F_{1,n-1}$$

In order to find the critical value c for a size α test, we find the constants $F_{1,n-1}(1-\alpha/2)$ and $F_{1,n-1}(\alpha/2)$ satisfying

$$\mathrm{P}_0\{x|\ F_{1,n-1}(1-\alpha/2) \leq \tau(x) \leq F_{1,n-1}(\alpha/2)\} = 1-\alpha$$

from the F-distribution table.

Next we consider

$$H_0:\ \sigma^2 = \sigma_0^2 \quad \text{against} \quad H_1:\ \sigma^2 \neq \sigma_0^2$$

where both hypothesis are again composite. The likelihood ratio is written as

$$\lambda(x) = \frac{\sup\limits_{\mu,\sigma^2}\left(\frac{1}{2\pi\sigma^2}\right)^{\frac{n}{2}}\exp\left(-\frac{1}{2\sigma^2}\sum(x_i-\mu)^2\right)}{\sup\limits_{\mu}\left(\frac{1}{2\pi\sigma_0^2}\right)^{\frac{n}{2}}\exp\left(-\frac{1}{2\sigma_0^2}\sum(x_i-\mu)^2\right)}$$

$$= \frac{\left(\frac{n}{\sum(x_i-\bar{x})^2}\right)^{\frac{n}{2}} \exp\left(-\frac{n}{2}\right)}{\left(\frac{1}{\sigma_0^2}\right)^{\frac{n}{2}} \exp\left(-\frac{1}{2\sigma_0^2}\sum(x_i-\bar{x})^2\right)}$$

$$= n^{\frac{n}{2}} e^{\frac{1-n}{2}} \left(\frac{\sum(x_i-\bar{x})^2}{\sigma_0^2}\right)^{-\frac{n}{2}} \exp\left(\frac{\sum(x_i-\bar{x})^2}{\sigma_0^2}\right)$$

Unfortunately, it is not easy to implement the generalized LR test. The null distribution of $\lambda(X)$ is difficult to derive. Moreover, no monotone increasing transformation of λ with known null distributions exists.

In practice, the test is usually based on the statistic

$$\tau(x) = \frac{\sum(x_i-\bar{x})^2}{\sigma_0^2}$$

Since $\tau(X) \sim \chi^2_{n-1}$ under H_0, the critical region is actually given by

$$C = \{x \mid \tau(x) \le \chi^2_{n-1}(1-\alpha/2) \ \text{ or } \ \tau(x) \ge \chi^2_{n-1}(\alpha/2)\}$$

for a size α test. To see how it is related to the generalized LR test C_* based on λ, write λ as $\lambda = f \circ \tau$ with $f(z) = z^{-\frac{n}{2}} e^z$. It can be easily seen that

$$C_* = \{x \mid \lambda(x) \ge c\} = \{x \mid \tau(x) \le c_1 \ \text{ or } \ \tau(x) \ge c_2\}$$

where c_1 and c_2 are the constants satisfying $c = f(c_1) = f(c_2)$ for any given c. In general, $f(\chi^2_{n-1}(1-\alpha/2)) \ne f(\chi^2_{n-1}(\alpha/2))$, and thus the test C given above is not the generalized LR test.

3. Exercises

1. Let a random variable X have the density $p(x,\theta) = \theta x^{\theta-1}, 0 < x < 1$ and 0, otherwise, where $\theta > 0$. Answer the following:

(a) In testing $H_0 : \theta \le 1$ versus $H_1 : \theta > 1$, find the power function and size of the test given by $C = \{x \mid x \ge 1/2\}$.

(b) Find a most powerful size-α test of $H_0 : \theta = 2$ versus $H_1 : \theta = 1$.

(c) Is there a uniformly most powerful size-α test of $H_0 : \theta = 2$ versus $\theta < 2$? If so, what is it?
(d) Find the generalized likelihood-ratio test of size-α of $H_0 : \theta = 1$ versus $H_1 : \theta \neq 1$.

2. Let X_1 and X_2 be independent and uniformly distributed on $(0, \theta)$. Consider the two tests with critical regions C_1 and C_2 given by

$$\begin{aligned} C_1 &= \{(x_1, x_2) |\, x_1 + x_2 \geq 1\} \\ C_2 &= \{(x_1, x_2) |\, \max\{x_1, x_2\} \geq 1/\sqrt{2}\} \end{aligned}$$

to test $H_0 : \theta \leq 1$ versus $H_1 : \theta > 1$.
(a) Find the distribution of $X_1 + X_2$ and $\max\{X_1, X_2\}$.
(b) Find the power functions of C_1 and C_2, and their sizes.
(c) Compare C_1 and C_2.

3. Let X_1 and X_2 be independent and uniformly distributed on $(\theta, \theta + 1)$. Consider the two tests with critical regions C_1 and C_2 given by

$$\begin{aligned} C_1 &= \{(x_1, x_2) |\, x_1 \geq .95\} \\ C_2 &= \{(x_1, x_2) |\, x_1 + x_2 \geq c\} \end{aligned}$$

to test $H_0 : \theta = 0$ versus $H_1 : \theta = 1/2$.
(a) Find the value of c so that C_2 has the same size as C_1.
(b) Find and compare the powers of C_1 and C_2.
(c) Show how to get a test that has the same size, but is more powerful than C_2.