Chapter 9 Testing Hypothesis and Assesing Goodness of Fit

9.1-9.3 The Neyman-Pearson Paradigm and Neyman Pearson Lemma

We begin by reviewing some basic terms in hypothesis testing.
Let H_0 denote the null hypothesis to be tested against the alternative hypothesis H_A.

Definition: Let C be the subset of the sample space which leads to rejection of the hypothesis under consideration. C is called the **critical region** or **rejection region** of the test.

Based on our analysis of the data and resulting conclusion, there are two types of errors that can be made:

• **type I error** is where the null hypothesis is rejected when it is, in fact, true. This is usually denoted by α and called the significance level of the test.

• **type II error** is where the null hypothesis is accepted (or not rejected) when it is, in fact, false.

The probability tha H_0 is rejected when it is false is called the power of the test and is denoted by $1 - \beta$.

We will first be concerned with testing a simple null hypothesis $H_0 : \theta = \theta'$ against a simple alternative hypothesis $H_A : \theta = \theta''$.

Definition: Let C denote a subset of the sample space. Then C is called a **best critical region** of size α for testing $H_0 : \theta = \theta'$ against $H_A : \theta = \theta''$ if, for every subset A of the sample space for which $P[A : H_0] = \alpha$, the following are true

(a) $P[C; H_0] = \alpha$

and

(b) $P[C; H_1] \geq P[A; H_1]$.

Example 1 Consider a random variable X that has a binomial distribution with $n = 5$ and $p = \theta$. We wish to test $H_0 : \theta = 1/2$ versus $H_A : \theta = 3/4$, at significance level $\alpha = 1/32$.

X	f(x;1/2)	f(x;3/4)	f(x;12)/f(x;34)
0	1/32	1/1024	32
1	5/32	15/1024	32/3
2	10/32	90/1024	32/9
3	10/32	270/1024	32/27
4	5/32	405/1024	32/81
5	1/32	243/1024	32/243

To satisfy the significance level $\alpha = 1/32$ we select the critical region as either $A_1 = \{x : x = 0\}$ or $A_2 = \{x : x = 5\}$. In fact, these are the only two possible ways of selecting a critical region that has probability no greater than 1/32 under the null hypothesis $H_0 : \theta = 1/2$. Thus, at least one of them must be a best critical region.

However, note that $P[A_1; H_A] = \frac{1}{1024} < \frac{243}{1024} = P[A_2; H_A]$

Thus, A_2 is the unique best critical region.

Lemma (Neyman-Pearson): Let $X_1, X_2, ..., X_n$, where n is a fixed positive integer, denote a random sample from a distribution with pdf $f(x;\theta)$. Then the joint pdf of $X_1, X_2, ..., X_n$ is

$$L(\theta; x_1, ..., x_n) = f(x_1;\theta)f(x_2;\theta)\cdots f(x_n;\theta).$$

Let θ' and θ'' be distinct values of the parameter space $\Omega = \{\theta : \theta = \theta', \theta''\}$, and let k be a positive number.

Define C as the subset of the sample space such that

(a) $\frac{L(\theta';x_1,...,x_n)}{L(\theta'';x_1,...,x_n)} \leq k$, for $(x_1, ..., x_n) \in C$.

(b) $\frac{L(\theta';x_1,...,x_n)}{L(\theta'';x_1,...,x_n)} > k$, for $(x_1, ..., x_n) \in C^*$.

(c) $\alpha = P[C; H_0]$.

Then C is a best critical region of size α for testing the simple null hypothesis $H_0 : \theta = \theta'$ against the simple alternative hypothesis $H_A : \theta = \theta''$.

Example 2: Let $X_1, X_2, ..., X_n$ denote a random sample from a distribution with pdf

$$f(x;\theta) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{(x-\theta)^2}{2}\right]$$

$-\infty < x < \infty$. We wish to test $H_0 : \theta = 0$ versus $H_1 : \theta = 1$.

$$\frac{L(0; x_1, ..., x_n)}{L(1; x_1, ..., x_n)} = \frac{(1/\sqrt{2\pi})^n \exp\left[-\frac{\sum_{i=1}^n x_i^2}{2}\right]}{(1/\sqrt{2\pi})^n \exp\left[-\frac{\sum_{i=1}^n (x_i-1)^2}{2}\right]} = \exp\left[\frac{n}{2} - \sum_{i=1}^n x_i\right]$$

For $k > 0$ the set of points $(x_1, ..., x_n)$ satisfying

$$\exp\left[\frac{n}{2} - \sum_{i=1}^n x_i\right] \le k$$

is a best critical region. Notice that this inequality is equivalent to

$$\frac{n}{2} - \sum_{i=1}^n x_i \le ln(k)$$

or

$$\sum_{i=1}^{n} x_i \geq \frac{n}{2} - ln(k) = c.$$

Given a particular significance level α, we can find c. Note that under H_0, $\sum X_i$ has a $N(0, n)$ distribution. Under H_0,

$$\alpha = P[\sum X_i \geq c] = P[Z \geq c/\sqrt{n}] = 1 - P[Z \leq c/\sqrt{n}]$$

where Z is a standard normal random variable.

Thus,

$$P[Z \leq c/\sqrt{n}] = 1 - \alpha$$

and

$$c = \sqrt{n}\Phi^{-1}(1 - \alpha)$$

where Φ denotes the cdf of a standard normal distribution.

Now, we will extend the notion of best critical regions (most powerful tests), to the case where the alternative hypothesis is a composite hypothesis.

Example 3: Suppose we know that the distribution of a random variable$X$ has the form

$$f(x;\theta) = \frac{1}{\theta}e^{-x/\theta}$$

for $0 < x < \infty$, where the parameter space $\Omega = \{\theta : \theta \geq 2\}$. We wish to test

$H_0 : \theta = 2$ versus $H_A : \theta > 2$.

We can see that X has an exponential distribution or equivalently a Gamma distribution with $\alpha = 1$ and $\beta = \theta$. Thus, $E[X;\theta] = \theta$.

Based on this, we would have evidence in favor of H_A for large values of the sample mean or $X_1 + X_2$.

We will take a random sample X_1, X_2 of size 2 and test H_0 versus H_A according to the critical region $C = \{(x_1, x_2) : 9.5 \leq x_1 + x_2 < \infty\}$.

Find the significance level of this test.

significance level (size)$=P[C; \theta = 2]$.

When H_0 is true, the joint pdf of X_1 and X_2 is

$$f(x_1, x_2; \theta = 2) = f(x_1; 2) f(x_2; 2) = \frac{1}{4} e^{-(x_1+x_2)/2}$$

and

$$P[C; \theta = 2] = 1 - P[C^*; \theta = 2]$$

$$= 1 - \int_0^{9.5} \int_0^{9.5-x_2} \frac{1}{4} e^{-(x_1+x_2)/2} dx_1 dx_2 \approx 0.05.$$

In fact, using the previous theory, we can see that for any $\theta'' > 2$, C is a best critical region of size 0.05 for testing the simple hypothesis $H_0 : \theta = \theta' = 2$, versus the simple alternative $H_A : \theta = \theta''$.

$$\frac{L(2; x_1, x_2)}{L(\theta''; x_1, x_2)} + \frac{(1/2)^2 \exp\left[-\frac{x_1+x_2}{2}\right]}{(1/\theta'')^2 \exp\left[-\frac{x_1+x_2}{\theta''}\right]} \leq k$$

This implies

$$2\ln(1/2) - 2\ln(1/\theta'') + (x_1 + x_2)(-1/2 + 1/\theta'')$$

$$\leq \ln(k)$$

Equivalently,

$$(x_1 + x_2) \geq$$

$$[\ln(k) - 2\ln(1/2) + 2\ln(1/\theta'')]/(-1/2 + 1/\theta'') = c$$

Thus, we would reject H_0 when $x_1 + x_2$ is greater than c. For a given significance level, we solve for c to find the best critical region C.

For example if, $\alpha = 0.05$, we see that $c = 9.5$ is the correct choice, no matter the value of $\theta'' > 2$.

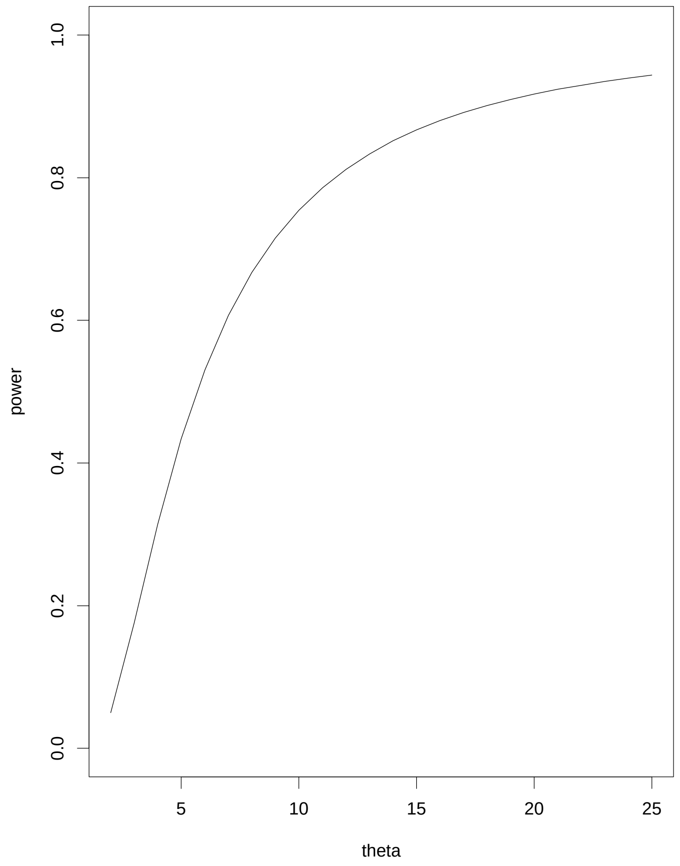
Because this is true for all θ'', it is also the best critical region of size 0.05, when the alternative hypothesis is a composite hypothesis $H_A : \theta > 2$.

Let $K(\theta) = P[C;\theta]$ denote the power function.

$$K(\theta) = 1 - \int_0^{9.5} \int_0^{9.5-x_2} \frac{1}{\theta^2} e^{-(x_1+x_2)/\theta} dx_1 dx_2$$

$$= \left(\frac{\theta + 9.5}{\theta}\right) e^{-9.5/\theta}$$

Figure 1: Plot of $K(\theta)$ versus θ



Definition: The critical region C is a **uniformly most powerful critical region** of size α for testing a simple null hypothesis H_0 against a composite alternative hypothesis H_A if C is a best critical region of size α for testing H_0 against each simple hypothesis in H_A.

Definition: A test defined by a uniformly most powerful critical region is called a **uniformly most powerful test**, with significance level α, for testing the simple null hypothesis H_0 against the composite alternative hypothesis H_A.

Uniformly most powerful tests do not always exist. However, when they do, using the Neyman-Pearson Lemma as in the previous example is a useful method for finding them.