Hypothesis Testing

Problem: choose, on basis of data X, between two alternatives.

Formally: choose between 2 *hypotheses*: $H_o : \theta \in \Theta_0$ or $H_1 : \theta \in \Theta_1$ where Θ_0 and Θ_1 are a partition of the model $P_\theta; \theta \in \Theta$. That is $\Theta_0 \cup \Theta_1 = \Theta$ and $\Theta_0 \cap \Theta_1 = \{\}$.

Make desired choice using *rejection* or *critical* region of test:

$$R = \{X : \text{we choose } \Theta_1 \text{ if we observe } X\}$$

Neyman Pearson approach to hypothesis testing: treat two hypotheses asymmetrically.

Hypothesis H_o is referred to as the *null* hypothesis (because traditionally it has been the hypothesis that some treatment has no effect).

Definition: *power function* of test with critical region R is

$$\pi(\theta) = P_\theta(X \in R)$$

Optimality theory: problem of finding best R.

Good R: $\pi(\theta)$ small for $\theta \in \Theta_0$ and large for $\theta \in \Theta_1$.

There is a trade off: can be made in many ways.

Jargon:

Type I error: error made when $\theta \in \Theta_0$ but we choose H_1, that is, $X \in R$.

The other kind of error, when $\theta \in \Theta_1$ but we choose H_0 is called a **Type II error**.

Defn: The *level* or *size* of a test is

$$\alpha \equiv \max_{\theta \in \Theta_o} \pi(\theta).$$

(Worst case probability of Type I error.)

The other error probability is denoted β and defined as

$$\beta(\theta) = P_\theta(X \not\in R) \quad \text{for } \theta \in \Theta_1$$

Notice: β will depend on θ.

Simple versus Simple testing

Finding best test is easiest when hypotheses very precise.

Definition: A hypothesis H_i is **simple** if Θ_i contains only a single value θ_i.

The simple versus simple testing problem arises when we test $\theta = \theta_0$ against $\theta = \theta_1$ so that Θ has only two points in it. This problem is of importance as a technical tool, not because it is a realistic situation.

Suppose that the model specifies that if $\theta = \theta_0$ then the density of X is $f_0(x)$ and if $\theta = \theta_1$ then the density of X is $f_1(x)$. How should we choose R?

Minimize $\alpha + \beta$, the total error probability:

$$P_{\theta_0}(X \in R) + P_{\theta_1}(X \notin R)$$

Write as integral:

$$\int [f_0(x)1(x \in R) + \{1 - 1(x \in R)\}f_1(x)]dx$$

For each x put x in R or not in such a way as to minimize integral.

But for each x the quantity

$$f_0(x)1(x \in R) + \{1 - 1(x \in R)\}f_1(x)$$

can be chosen either to be $f_0(x)$ or $f_1(x)$.

Solution: put $x \in R$ iff $f_1(x) > f_0(x)$. Note can rephrase condition in terms of **likelihood ratio** $f_1(x)/f_0(x)$.

Theorem: For each fixed λ the quantity $\beta+\lambda\alpha$ is minimized by R which has

$$R = \left\{ x : \frac{f_1(x)}{f_0(x)} > \lambda \right\}.$$

Neyman-Pearson: two kinds of errors might have unequal consequences.

So: pick the more serious kind of error, label it **Type I** and require rule to hold probability α of Type I error at or below prespecified level α_0.

Typically: $\alpha_0 = 0.05$, chiefly for historical reasons.

Neyman-Pearson solution: minimize β subject to constraint $\alpha \le \alpha_0$.

Usually equivalent to constraint $\alpha = \alpha_0$.

Most Powerful Level α_0 test maximizes $1-\beta$ subject to $\alpha \le \alpha_0$.

The Neyman Pearson Lemma

Theorem: In testing f_0 against f_1 the probability β of a type II error is minimized, subject to $\alpha \leq \alpha_0$ by the rejection region:

$$R = \left\{ x : \frac{f_1(x)}{f_0(x)} > \lambda \right\}$$

where λ is the largest constant such that

$$P_0 \left\{ \frac{f_1(x)}{f_0(x)} \geq \lambda \right\} = \alpha_0$$

Example: If $X_1, \ldots, X_n$ are iid $N(\mu, 1)$ and we have $\mu_0 = 0$ and $\mu_1 > 0$ then

$$\frac{f_1(X_1, \ldots, X_n)}{f_0(X_1, \ldots, X_n)} =$$
$$\exp\{\mu_1 \sum X_i - n\mu_1^2/2 - \mu_0 \sum X_i + n\mu_0^2/2\}$$

which simplifies to

$$\exp\{\mu_1 \sum X_i - n\mu_1^2/2\}$$

Now choose λ so that

$$P_0(\exp\{\mu_1 \sum X_i - n\mu_1^2/2\} > \lambda) = \alpha_0$$

Rewrite the probability as

$$P_0(\sum X_i > [\log(\lambda) + n\mu_1^2/2]/\mu_1) = \\ 1 - \Phi([\log(\lambda) + n\mu_1^2/2]/[n^{1/2}\mu_1])$$

Notation: z_α: upper α critical point of N(0,1) distribution.

Then

$$z_{\alpha_0} = [\log(\lambda) + n\mu_1^2/2]/[n^{1/2}\mu_1]$$

which you can solve to get a formula for λ in terms of z_{α_0}, n and μ_1.

Rejection region looks complicated: reject if a complicated statistic is larger than λ which has a complicated formula.

But re-expressed rejection region as

$$\frac{\sum X_i}{\sqrt{n}} > z_{\alpha_0}$$

Key point: rejection region same for any $\mu_1 > 0$.

Definition: In the general problem of testing Θ_0 against Θ_1 level of critical region R is

$$\alpha = \sup_{\theta \in \Theta_0} P_\theta(X \in R).$$

The power function is

$$\pi(\theta) = P_\theta(X \in R).$$

A test with rejection region R is Uniformly Most Powerful at level α_0 if

1. the test has level $\alpha \le \alpha_o$

2. If R^* is another rejection region with level $\alpha \le \alpha_0$ then for every $\theta \in \Theta_1$ we have

$$P_\theta(X \in R^*) \le P_\theta(X \in R).$$

Application of the NP lemma: In the $N(\mu, 1)$ model consider $\Theta_1 = \{\mu > 0\}$ and $\Theta_0 = \{0\}$ or $\Theta_0 = \{\mu \le 0\}$. The UMP level α_0 test of $H_0 : \mu \in \Theta_0$ against $H_1 : \mu \in \Theta_1$ is

$$R^* == \{x : n^{1/2}\bar{X} > z_{\alpha_0}\}$$

Proof: For either choice of Θ_0 this test has level α_0 because for $\mu \le 0$ we have

$$\begin{aligned} P_\mu(n^{1/2}\bar{X} > & z_{\alpha_0}) \\ &= P_\mu(n^{1/2}(\bar{X} - \mu) > z_{\alpha_0} - n^{1/2}\mu) \\ &= P(N(0,1) > z_{\alpha_0} - n^{1/2}\mu) \\ &\le P(N(0,1) > z_{\alpha_0}) \\ &= \alpha_0 \end{aligned}$$

(Notice use of $\mu \le 0$.

Key idea: critical point fixed by behaviour on edge of null hypothesis.

Now suppose R is any other level α_0 critical region:

$$P_0((X_1, \ldots, X_n) \in R) \le \alpha_0.$$

Fix a $\mu > 0$. According to the NP lemma

$$P_\mu\{(X_1, \ldots, X_n) \in R\} \le P_\mu\{(X_1, \ldots, X_n) \in R_\lambda\}$$

where

$$R_\lambda = \{x : f_\mu(x_1, \ldots, x_n)/f_0(x_1, \ldots, x_n) > \lambda\}$$

for a suitable λ.

But we just checked that this test had a rejection region of the form

$$R^* = n^{1/2}\bar{X} > z_{\alpha_0}$$

The NP lemma produces the same test for every $\mu > 0$ chosen as an alternative.

So this test is UMP level α_0.

Proof of the Neyman Pearson lemma:

Lagrange Multipliers

Suppose you want to minimize $f(x)$ subject to $g(x) = 0$. Consider first the function

$$h_\lambda(x) = f(x) + \lambda g(x)$$

If x_λ minimizes h_λ then for any other x

$$f(x_\lambda) \le f(x) + \lambda[g(x) - g(x_\lambda)]$$

Now suppose you can find a value of λ such that the solution x_λ has $g(x_\lambda) = 0$. Then for any x we have

$$f(x_\lambda) \le f(x) + \lambda g(x)$$

and for any x satisfying the constraint $g(x) = 0$ we have

$$f(x_\lambda) \le f(x)$$

This proves that for this special value of λ the quantity x_λ minimizes $f(x)$ subject to $g(x) = 0$.

Notice that to find x_λ you set the usual partial derivatives equal to 0; then to find the special x_λ you add in the condition $g(x\lambda) = 0$.

Proof of NP lemma

$R_\lambda = \{x : f_1(x)/f_0(x) \ge \lambda\}$ minimizes $\lambda\alpha + \beta$.

As λ increases from 0 to ∞ level of R_λ decreases from 1 to 0.

Ignore technical problem: $f_1(X)/f_0(X)$ might be discrete.

There is thus a value λ_0 where level $= \alpha_0$.

According to theorem above test minimizes $\alpha + \lambda_0\beta$. Suppose R^* is some other test with level $\alpha^* \le \alpha_0$. Then

$$\lambda_0\alpha + \beta \le \lambda_0\alpha_{R^*} + \beta_{R^*}$$

We can rearrange this as

$$\beta_{R^*} \ge \beta + (\alpha - \alpha_{R^*})\lambda_0$$

Since

$$\alpha_{R^*} \le \alpha_0 = \alpha$$

the second term is non-negative and

$$\beta_{R^*} \ge \beta$$

which proves the Neyman Pearson Lemma.

General phenomenon: for any $\mu > \mu_0$, likelihood ratio f_μ / f_0 is an increasing function of $\sum X_i$.

Rejection region of NP test is thus always a region of the form $\sum X_i > k$.

Value of constant k determined by requirement that test have level α_0; this depends on μ_0 not on μ_1.

Definition: The family $f_\theta; \theta \in \Theta \subset R$ has monotone likelikelood ratio with respect to a statistic $T(X)$ if for each $\theta_1 > \theta_0$ the likelihood ratio $f_{\theta_1}(X)/f_{\theta_0}(X)$ is a monotone increasing function of $T(X)$.

Theorem: For a monotone likelihood ratio family the Uniformly Most Powerful level α test of $\theta \le \theta_0$ (or of $\theta = \theta_0$) against the alternative $\theta > \theta_0$ is

$$R = \{xT(x) > t_\alpha\}$$

where $P_0(T(X) > t_\alpha) = \alpha_0$.

Usual application: one parameter exponential family.

Almost any other problem: method doesn't work and there is no uniformly most powerful test.

For instance: testing $\mu = \mu_0$ against the two sided alternative $\mu \neq \mu_0$ there is no UMP level α test.