Hypothesis tests

Aldo Solari
Statistical Inference II
PhD in Economics, Statistics and Data Science
University of Milano-Bicocca

XXXVII cycle

Outline

Simple null hypotheses

Composite null hypotheses

Main references

- Cox and Hinkley (1976) Theoretical Statistics. Chapman and Hall/CRC, §4, §5

The Neyman–Pearson formulation of hypothesis testing requires to fix the probability of rejecting H_0 when it is true, denoted by α, aiming to maximize the probability of rejecting H_1 when false.

This approach demands the explicit formulation of the *alternative hypothesis* H_1.

The decision procedure, i.e. rejecting or not H_0, is called the *test* of H_0 against H_1.

Suppose Y has distribution $f_Y(y;\theta)$ for $\theta \in \Theta$

Formulate a null hypothesis $H_0 : \theta \in \Theta_0$ and an alternative hypothesis $H_1 : \theta \in \Theta_1$ with $\Theta_0 \cup \Theta_1 = \Theta$ and $\Theta_0 \cap \Theta_1 = \emptyset$

A *test* or *critical function* $\phi = \phi(Y)$ assigns to each possible value y one of these two decisions

$$\phi : \mathcal{Y} \mapsto \{0, 1\}$$

where 1 denotes the decision of rejecting H_0 and 0 denotes the decision of not rejecting H_0, and thereby partition the sample space $\mathcal{Y}$ into two complementary regions $\mathcal{Y}_0$ and $\mathcal{Y}_1$

When performing a test one may arrive at the correct decision, or one may commit one of two errors: rejecting H_0 when it is true (*type I error*) or not rejecting it when it is false (*type II error*).

Critical region

Unfortunately, the probabilities of the two types of error cannot be controlled simultaneously

Choose the *level of significance* $\alpha \in (0, 1)$, and control the probability of type I error at α, i.e.

$$\mathrm{pr}_\theta(Y \in \mathcal{Y}_1) \leq \alpha \quad \forall\ \theta \in \Theta_0$$

The *size* of the test is

$$\sup_{\theta \in \Theta_0} \mathrm{pr}_\theta(Y \in \mathcal{Y}_1)$$

If, for all α, the size of the test is α, we call $\mathcal{Y}_1$ a *critical region of size* α, denoted by $\mathcal{Y}_\alpha$

Power function

Subject to

$$\sup_{\theta \in \Theta_0} \mathrm{pr}_\theta(Y \in \mathcal{Y}_\alpha) = \alpha$$

it is desired to maximize

$$\mathrm{pr}_\theta(Y \in \mathcal{Y}_\alpha) \quad \forall\ \theta \in \Theta_1$$

Considered as a function of θ, this probability is called the *power function* of the test

$$\mathrm{pow}(\theta; \alpha) = \mathrm{pr}_\theta(Y \in \mathcal{Y}_\alpha; \theta)$$

p-value

If we require that the rejection regions $\mathcal{Y}_\alpha$ and $\mathcal{Y}_{\tilde{\alpha}}$ are *nested* in the sense that

$$\mathcal{Y}_\alpha \subset \mathcal{Y}_{\tilde{\alpha}} \quad \text{if } \alpha < \tilde{\alpha}$$

the p-value is defined as the smallest significance level at which the null hypothesis would be rejected for the given observation:

$$p_{\text{obs}} = \inf\{\alpha : y \in \mathcal{Y}_\alpha\}$$

Table of Contents

In the present section we consider only the case where H_0 is a simple hypothesis

It is best to begin with a simple alternative hypothesis H_1

$$H_0 : Y \sim f_0(y), \quad H_1 : Y \sim f_1(y)$$

Let $\mathcal{Y}_\alpha$ and $\mathcal{Y}'_\alpha$ be two critical region of size α, i.e.

$$\mathrm{pr}_0(Y \in \mathcal{Y}_\alpha) \quad = \quad \mathrm{pr}_0(Y \in \mathcal{Y}'_\alpha) \tag{1}$$

We regard $\mathcal{Y}_\alpha$ as preferable to $\mathcal{Y}'_\alpha$ for the alternative H_1 if

$$\mathrm{pr}_1(Y \in \mathcal{Y}_\alpha) \quad > \quad \mathrm{pr}_1(Y \in \mathcal{Y}'_\alpha) \tag{2}$$

The region $\mathcal{Y}_\alpha$ is called the *critical region* of size α if (2) is satisfied for all other $\mathcal{Y}'_\alpha$ satisfying the size condition (1).

We call $\mathrm{pr}_1(Y \in \mathcal{Y}_\alpha)$ the *size α power* of the test against H_1

Neyman-Pearson lemma

For simplicity, suppose that the likelihood ratio $\mathrm{lr}(Y) = f_1(Y)/f_0(Y)$ is, under H_0, a continuous random variable such that for all α, there exists a unique c_α such that

$$\mathrm{pr}_0(\mathrm{lr}(Y) \geq c_\alpha) = \alpha$$

We call the region defined by

$$\mathrm{lr}(y) \geq c_\alpha$$

the size α *likelihood ratio critical region*

A fundamental result, called the Neyman-Pearson lemma, is that, *for any size α, the likelihood ratio critical region is the best critical region.*

Let $\mathcal{Y}_\alpha$ be the likelihood ratio critical region and let $\mathcal{Y}_1$ be any other critical region, both being of size α. Then

$$\alpha = \int_{\mathcal{Y}_\alpha} f_0(y)dy = \int_{\mathcal{Y}_1} f_0(y)dy$$

so that

$$\int_{\mathcal{Y}_\alpha \setminus \mathcal{Y}_1} f_0(y)dy = \int_{\mathcal{Y}_1 \setminus \mathcal{Y}_\alpha} f_0(y)dy$$

since $\int_{\mathcal{Y}_\alpha} f_0(y)dy = \int_{\mathcal{Y}_\alpha \setminus \mathcal{Y}_1} f_0(y)dy + \int_{\mathcal{Y}_\alpha \cap \mathcal{Y}_1} f_0(y)dy$

Now, if $y \in \mathcal{Y}_\alpha \setminus \mathcal{Y}_1$, which is inside $\mathcal{Y}_\alpha$, $f_1(y) \geq c_\alpha f_0(y)$, while if $y \in \mathcal{Y}_1 \setminus \mathcal{Y}_\alpha$, which is outside $\mathcal{Y}_\alpha$, $c_\alpha f_0(y) > f_1(y)$.

We have that

$$\int_{\mathcal{Y}_\alpha \setminus \mathcal{Y}_1} f_1(y)dy \geq c_\alpha \int_{\mathcal{Y}_\alpha \setminus \mathcal{Y}_1} f_0(y)dy = c_\alpha \int_{\mathcal{Y}_1 \setminus \mathcal{Y}_\alpha} f_0(y)dy \geq \int_{\mathcal{Y}_1 \setminus \mathcal{Y}_\alpha} f_1(y)dy$$

with strict inequality unless the regions are equivalent

Then

$$\int_{\mathcal{Y}_\alpha} f_1(y)dy \geq \int_{\mathcal{Y}_1} f_1(y)dy$$

thus the power of $\mathcal{Y}_\alpha$ is at least that of $\mathcal{Y}_1$

Note that if $\mathcal{Y}_1$ had been of size less than α the final inequality holds

Let $Y_1, \ldots, Y_n$ be i.i.d. $N(\mu, 1)$. Consider

$$H_0 : \mu = \mu_0, \quad H_1 : \mu = \mu_1$$

with $\mu_1 > \mu_0$.

$$\mathrm{lr}(y) = \exp\left\{n\bar{y}(\mu_1 - \mu_0) - \frac{1}{2}n\mu_1^2 + \frac{1}{2}n\mu_0^2\right\}$$

Because all quantities, except for $\bar{y}$, are fixed constants, and because $\mu_1 - \mu_0 > 0$, a critical region of the form $\mathrm{lr}(y) \geq c_\alpha$ is equivalent to one of the form $\bar{y} \geq d_\alpha$. Since $\bar{Y} \overset{H_0}{\sim} N(\mu_0, 1/n)$

$$d_\alpha = \mu_0 + \frac{z_\alpha}{\sqrt{n}}$$

where z_α is the $1 - \alpha$ quantile of $N(0, 1)$, and

$$\mathcal{Y}_\alpha^+ = \{y_1, \ldots, y_n : \sqrt{n}(\bar{y} - \mu_0) \geq z_\alpha\}$$

Let $Y_1, \ldots, Y_n$ be i.i.d. in the single parameter exponential family

$$\exp\{a(\theta)b(y) + c(\theta) + d(y)\}$$

among them the normal, gamma, binomial and Poisson distribution, and that the hypotheses are $H_0 : \theta = \theta_0$ and $H_1 : \theta = \theta_1$ Then the likelihood ratio involves the data only through the sufficient statistic $S = \sum b(Y_j)$ and the best critical region has the form

$$\exp\{a(\theta_1) - a(\theta_0)\}s \geq e_\alpha$$

If $a(\theta_1) - a(\theta_0) > 0$, this is equivalent to $s \geq \tilde{e}_\alpha$, the critical region being the same for all such θ_1

Let $Y \sim \text{Poisson}(\lambda)$. Consider

$$H_0 : \lambda = 1, \quad H_1 : \lambda = \lambda_1 > 1$$

The likelihood critical regions have the form $y \geq d_\alpha$

However, because Y is discrete, the only critical regions are of the form $y \geq r$, where r is an integer

r	0	1	2	3	4	5	6
$\text{pr}_0(Y \geq r)$	1	0.632	0.264	0.08	0.0189	0.0037	0.0006

If α is one of the values above, a likelihood ratio region of the required size does exist.

By a mathematical artifice, it is, however, possible to achieve likelihood ratio critical regions with other values of α

Suppose that $\alpha = 0.05$. The region $y \geq 4$ is too small, whereas the region $y \geq 3$ is too large. All values $y \geq 4$ are put in the critical region, whereas if $y = 3$ we regard the data as in the critical region with probability π such that

$$\mathrm{pr}_0(Y \geq 4) + \pi \cdot \mathrm{pr}_0(Y = 3) = 0.05$$

leading to $\pi = 0.51$. This is a *randomized critical region* of size 0.05.

The randomized definition of p_{obs} corresponding to $Y = y$ is

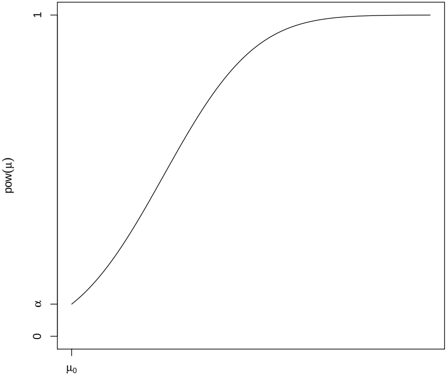
$$\mathrm{pr}_0(Y > y) + U \cdot \mathrm{pr}_0(Y = y)$$

where $U \sim \mathrm{Uniform}(0, 1)$, independently of Y. The corresponding random variable is, under H_0, $\mathrm{Uniform}(0, 1)$.

Composite alternatives

Suppose $H_0 : \theta = \theta_0$ and $H_1 : \theta \in \Theta_1$. Two cases now arise.

- We get the same size α best critical region for all $\theta \in \Theta_1$. Then we say that the region is *uniformly most powerful* size α region. If this holds for each α, then the test itself is called uniformly most powerful (UMP).
- The best critical region depends on the particular $\theta \in \Theta_1$. Then no uniformly most powerful exists. One possibility is to take $\theta \in \Theta_1$ very close to θ_0, to maximize the power locally near the null hypothesis.



$Y_1, \ldots, Y_n$ be i.i.d. $N(\mu, 1)$. Test $H_0 : \mu = \mu_0$ against $H_1 : \mu > \mu_0$ with critical region $\mathcal{Y}_\alpha^+ = \{y_1, \ldots, y_n : \sqrt{n}(\bar{y} - \mu_0) \geq z_\alpha\}$

$$\mathrm{pow}(\mu; \alpha) = \Phi(\sqrt{n}(\mu - \mu_0) - z_\alpha)$$

Two-sided tests

Let $Y_1, \ldots, Y_n$ be i.i.d. $N(\mu, 1)$. Test $H_0 : \mu = \mu_0$ against $H_1 : \mu < \mu_0$ with critical region $\mathcal{Y}_\alpha^- = \{y_1, \ldots, y_n : \sqrt{n}(\bar{y} - \mu_0) \leq -z_\alpha\}$

Suppose now that we wish to test $H_0 : \mu = \mu_0$ against $H_1 : \mu \neq \mu_0$.

The critical region

$$\mathcal{Y}_\alpha = \mathcal{Y}_\alpha^- \cup \mathcal{Y}_\alpha^+$$

has size 2α, and no uniformly more powerful test exists for the two-sided alternative.

A critical region $\mathcal{Y}_\alpha$ of size α is called *unbiased* if

$$\mathrm{pr}(Y \in \mathcal{Y}_\alpha; \theta) \geq \mathrm{pr}(Y \in \mathcal{Y}_\alpha; \theta_0) \quad \forall\ \theta \in \Theta_1$$

One may restrict attention to unbiased regions and among these look for the one with maximum power

A test which is uniformly most powerful amongst the class of all unbiased tests is *uniformly most powerful unbiased* (UMPU)

Suppose $Y = (Y_1, \ldots, Y_m)^t$ is multivariate normal with mean vector $\mu = (\mu_1, \ldots, \mu_m)^t \geq 0$ and known nonsingular covariance matrix Σ

For testing $H_0 : \mu = 0$ against $H_1 : \mu = \mu_1$, the most powerful test rejects for large values of

$$\mu_1^t \Sigma^{-1} Y$$

In particular, no UMP test exists

For testing $H_0 : \mu = 0$ against $H_1 : \mu = (k, \ldots, k)^t$ for $k > 0$, a UMP test exists and rejects for large values of the sum of the components of $\Sigma^{-1} Y$. If, in particular, Σ has diagonal elements 1 and off-diagonal elements ρ, then the test rejects when

$$\sum_i Y_i \geq z_\alpha (m + m(m-1)\rho)^{1/2}$$

Locally most powerful tests

Let $f_0(y) = f(y; \theta_0)$ and $f_1(y) = f(y; \theta_1)$ with $\theta_1 = \theta_0 + \epsilon$ for small $\epsilon > 0$

$$\begin{aligned}
\log \mathrm{lr}(y) &= \log \frac{f(y; \theta_0 + \epsilon)}{f(y; \theta_0)} \\
&= \log f(y; \theta_0 + \epsilon) - \log f(y; \theta_0) \\
&= \epsilon \frac{\partial \log f(y; \theta_0)}{\partial \theta_0} + \dots
\end{aligned}$$

Then the appropriate test statistic is the score

$$U(\theta_0) = \frac{\partial \log f(Y; \theta_0)}{\partial \theta_0}$$

Its null distribution has mean 0 and variance $i(\theta_0)$, the Fisher information

Let $Y_1, \ldots, Y_n$ be i.i.d. in the Cauchy distribution

$$\frac{1}{\pi[1 + (y - \theta)^2]}$$

For the null hypothesis $H_0 : \theta = \theta_0$ the score from Y_1 is

$$U_1(\theta_0) = \frac{2(Y_1 - \theta_0)}{1 + (Y_1 - \theta_0)^2}$$

and the information from a single observation is

$$i_1(\theta_0) = \frac{1}{2}$$

The test statistic is thus

$$U(\theta_0) = 2 \sum_{i=1}^{n} \frac{(Y_i - \theta_0)}{1 + (Y_i - \theta_0)^2}$$

Its null distribution has mean o and variance $n/2$

Table of Contents