Module 6: Statistical inference (III)

Yaqi Shi

July 19, 2024

Outline

This module we will review

- Basics of hypothesis testing
- Central Limit Theorem
- The Wald test
- The score test
- The likelihood ratio test

Hypothesis testing

Definition (Hypothesis testing)

Suppose that we partition the parameter space Θ into two disjoint sets Θ_0 and Θ_1 and that we wish to test

$$H_0 : \theta \in \Theta_0 \quad \text{versus} \quad H_1 : \theta \in \Theta_1$$

We call H_0 the null hypothesis and H_1 the alternative hypothesis.

Rejection region

Let X be a random variable and let $\mathcal{X}$ be the range of X. Rejection region is a subset of outcomes $R \in \mathcal{X}$

$$\begin{aligned} X \in R &\implies \text{reject } H_0 \\ X \notin R &\implies \text{retain (do not reject) } H_0 \end{aligned}$$

Usually, the rejection region is

$$R = \{x : T(x) > c\}$$

where T is a test statistic and c is a critical value.

Type I error and type II error

- Type I error, also known as a "false positive": the error of rejecting a null hypothesis when it is actually true. $P(X \in R|H_0)$.
- Type II error, also known as a "false negative": the error of not rejecting a null hypothesis when the alternative hypothesis is the true state of nature. $P(X \notin R|H_0)$.

Power function

Definition (Power function)

In a test of hypothesis about a parameter θ, let the null hypothesis be $H_0 : \theta = \theta_0$. The power function $\beta(\theta)$ is a function that gives, for any θ, the probability of rejecting the null hypothesis when the true parameter is equal to θ.

$$P(X \in R | \theta \text{ is the true parameter})$$

Note that the power function depends on the null hypothesis: if we change θ_0, also the power function changes.

Size of a test

Definition (The size of a test)

The size of a test is defined to be

$$\alpha = \sup_{\theta \in \Theta_0} \beta(\theta).$$

A test is said to have level α if its size is less than or equal to α.

Intuitively, we consider all the cases in which the null is true ($\theta \in \Theta_0$). For each case, we compute the probability of (incorrect) rejection. The size is equal to the largest value we find (worst-case scenario).

Exercise

Let $X_1, \ldots, X_n \sim N(\mu, \sigma)$ where σ is known. We want to test $H_0 : \mu \leq 0$ versus $H_1 : \mu > 0$. Hence, $\Theta_0 = (-\infty, 0]$ and $\Theta_1 = (0, \infty)$.

Consider the test:

$$\text{reject } H_0 \text{ if } T > c$$

where $T = \bar{X}$. The rejection region is

$$R = \{(x_1, \ldots, x_n) : T(x_1, \ldots, x_n) > c\}$$

What is the power function? What is the size of the test?

Exercise (cont'd)

Let Z denote a standard Normal random variable. The power function is

$$\begin{aligned}\beta(\mu) &= \mathbb{P}_\mu(\bar{X} > c) \\ &= \mathbb{P}_\mu\left(\frac{\sqrt{n}(\bar{X}-\mu)}{\sigma} > \frac{\sqrt{n}(c-\mu)}{\sigma}\right) \\ &= \mathbb{P}\left(Z > \frac{\sqrt{n}(c-\mu)}{\sigma}\right) \\ &= 1 - \Phi\left(\frac{\sqrt{n}(c-\mu)}{\sigma}\right)\end{aligned}$$

Exercise (cont'd)

$$\text{size} = \sup_{\mu \leq 0} \beta(\mu) = \beta(0) = 1 - \Phi\left(\frac{\sqrt{n}c}{\sigma}\right)$$

For a size α test, we set this equal to α and solve for c to get

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

We reject when $\bar{X} > \sigma\Phi^{-1}(1-\alpha)/\sqrt{n}$. Equivalently, we reject when

$$\frac{\sqrt{n}(\bar{X}-0)}{\sigma} > z_\alpha$$

where $z_\alpha = \Phi^{-1}(1-\alpha)$

Asymptotically normality

Definition

We say that an estimator is asymptotically normal if:

$$\frac{\hat{\theta}-\theta}{\sqrt{\mathrm{Var}(\hat{\theta})}} \xrightarrow{d} N(0,1)$$

Theorem

If an estimator is asymptotically normal and the scaled squared standard error $\sqrt{n\widehat{\mathrm{Var}}(\hat{\theta})} \xrightarrow{P} \sqrt{n\,\mathrm{Var}(\hat{\theta})}$ then

$$\frac{\hat{\theta}-\theta}{\sqrt{\widehat{\mathrm{Var}}(\hat{\theta})}} \xrightarrow{d} N(0,1)$$

Central Limit Thorem

Let $X_1, \ldots, X_n$ be a sequence of independent and identically distributed (i.i.d.) random variables drawn from a distribution with mean μ and variance σ^2. Then

$$\frac{\bar{X}_n - \mu}{\sqrt{\text{Var}\left(\bar{X}_n\right)}} = \frac{\sqrt{n}\left(\bar{X}_n - \mu\right)}{\sigma} \xrightarrow{d} N(0,1) \text{ as } n \to \infty$$

Proof: Omitted. By characteristic functions.

Example

When $X_1, \ldots, X_n \overset{\text{i.i.d}}{\sim} N\left(\mu, \sigma^2\right)$, then $\bar{X}_n$ satisfies that $\frac{\bar{X}_n-\mu}{\sqrt{\operatorname{Var}\left(\bar{X}_n\right)}} \xrightarrow{d} N(0,1)$ and $\sqrt{n \operatorname{var}\left(\bar{X}_n\right)} = \sqrt{n \frac{s^2}{n}} = s \xrightarrow{P} \sigma = \sqrt{n \operatorname{Var}\left(\bar{X}_n\right)}$.
Then we can use the theorem above to conclude that

$$\frac{\bar{X}_n - \mu}{\sqrt{\widehat{\operatorname{Var}}\left(\bar{X}_n\right)}} \xrightarrow{d} N(0,1).$$

The Wald test

We are interested in testing the hypotheses in a parametric model:

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

The Wald test most generally is based on an asymptotically normal estimator, i.e. we suppose that we have access to an estimator $\widehat{\theta}$ which under the null satisfies the property that:

$$\widehat{\theta} \xrightarrow{d} N\left(\theta_0, \sigma_0^2\right)$$

where σ_0^2 is the variance of the estimator under the null. The canonical example is when $\widehat{\theta}$ is taken to be the MLE.

The Wald statistic

If the hypothesis involves only a single parameter restriction, then the Wald statistic takes the following form:

$$W_n = \frac{\left(\hat{\theta} - \theta_0\right)^2}{\text{var}(\hat{\theta}_0)},$$

which under the null hypothesis follows an asymptotic χ_1^2-distribution with one degree of freedom.

Example

Suppose we considered the problem of testing the parameter of a Bernoulli, i.e. we observe $X_1, \dots, X_n \sim \text{Ber}(p)$, and the null is that $p = p_0$. Defining $\widehat{p} = \frac{1}{n}\sum_{i=1}^n X_i$. A Wald test could be constructed based on the statistic:

$$T_n = \frac{(\widehat{p} - p_0)^2}{\frac{p_0(1-p_0)}{n}},$$

which has an asymptotic χ^2_1 distribution. An alternative would be to use a slightly different estimated standard deviation, i.e. to define,

$$T_n = \frac{(\widehat{p} - p_0)^2}{\frac{\widehat{p}(1-\widehat{p})}{n}},$$

Observe that this alternative test statistic also has an asymptotically χ^2_1 distribution under the null.

The score test

Score test is based on the value of the score function $U(\theta)$ under the null hypothesis H_0.

Reminder: $U(\theta) = \ell'(\theta)$.

The score test statistic

$$S_n = \frac{U(\theta_0)^2}{\text{var}[U(\theta_0)]},$$

which has an asymptotic distribution of χ_1^2 under the null.

Reminder: the variance of the score function is the Fisher information.

The score test

Similary, we have

$$\widehat{I}(\theta_0) = -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \log f(x_i \mid \theta)}{\partial\theta^2}\bigg|_{\theta_0} \xrightarrow{\mathrm{P}} I(\theta_0)$$

so that

$$S_n = \frac{U(\theta_0)^2}{\widehat{I}(\theta_0)}$$

also has an asymptotic distribution of χ^2_1 under the null.

The likelihood ratio test

$$\Delta_n = \ell\left(\widehat{\theta}_n\right) - \ell\left(\theta_0\right) = \log\left(\frac{\sup_{\theta\in\Theta}(\theta \mid \mathbf{x})}{L\left(\theta_0 \mid \mathbf{x}\right)}\right) \geq 0$$

Under H_0,

$$2\Delta_n \xrightarrow{\mathrm{D}} \chi_1^2$$

Example

Let $X_1, \ldots, X_n \in \{0, 1\}$ be the results of n flips of a coin, and consider the following null and alternative hypotheses:

$$H_0 : X_1, \ldots, X_n \overset{IID}{\sim} \text{Bernoulli}\left(\frac{1}{2}\right)$$
$$H_1 : X_1, \ldots, X_n \overset{IID}{\sim} \text{Bernoulli}(p).$$

The joint PMF of $(X_1, \ldots, X_n)$ under H_0 and H_1 are, respectively,

$$f_0\left(x_1, \ldots, x_n\right) = \prod_{i=1}^{n} \frac{1}{2} = \frac{1}{2^n}$$
$$f_1\left(x_1, \ldots, x_n\right) = (1-p)^n \left(\frac{p}{1-p}\right)^{x_1+\ldots+x_n}.$$

Thus, the ratio is

$$L\left(X_1, \ldots, X_n\right) = \frac{f_0\left(X_1, \ldots, X_n\right)}{f_1\left(X_1, \ldots, X_n\right)} = \frac{1}{2^n(1-p)^n} \left(\frac{1-p}{p}\right)^{X_1+\ldots+X_n}$$

Uniformly most powerful test

Definition:

In statistical hypothesis testing, a uniformly most powerful (UMP) test is a hypothesis test which has the greatest power among all possible tests of a given size α.

Theorem (Neyman-Pearson lemma):

Let H_0 and H_1 be simple hypotheses. For a constant $c > 0$, suppose that the likelihood ratio test which rejects H_0 when $L(\mathbf{x}) < c$ has significance level α. Then for any other test of H_0 with significance level at most α, its power against H_1 is at most the power of this likelihood ratio test.

The Wald test, score test, and likelihood ratio test

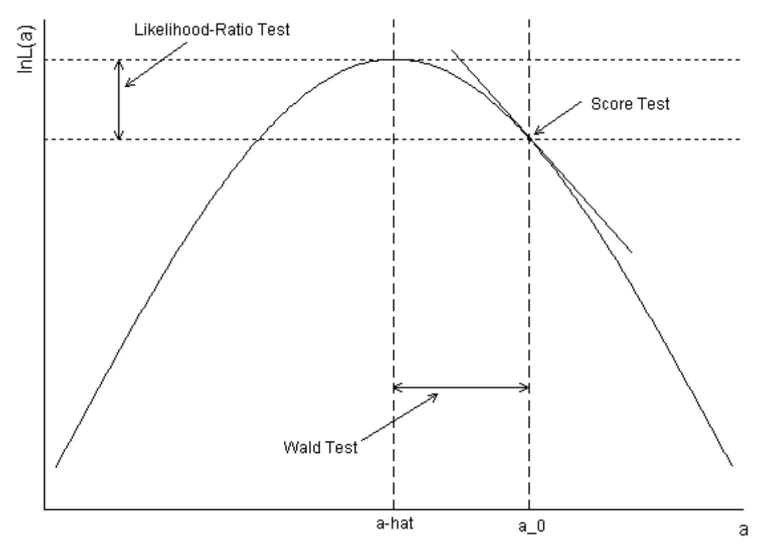

Figure 1: Fox, J. (1997) Applied regression analysis, linear models, and related methods. Thousand Oaks, CA: Sage Publications. P. 570.

Test equivalence

We can show that (when there is no misspecification)

- The tests are asymptotically equivalent in the sense that under H_0 they reach the same decision with probability 1 as $n \to \infty$.
- For a finite sample size n, they have some relative advantages and disadvantages with respect to one another.

Discussion

$$W_n = \frac{\left(\hat{\theta} - \theta_0\right)^2}{\text{var}(\hat{\theta}_0)} \xrightarrow{\mathrm{D}} \chi_1^2$$

$$S_n = \frac{U(\theta_0)^2}{\widehat{I}(\theta_0)} \xrightarrow{\mathrm{D}} \chi_1^2$$

$$2\Delta_n = 2\left\{\ell\left(\widehat{\theta}_n\right) - \ell\left(\theta_0\right)\right\} \xrightarrow{\mathrm{D}} \chi_1^2$$

- It is easy to create one-sided Wald and score tests.
- The score test does not require $\widehat{\theta}_n$ whereas the other two tests do.
- The Wald test is most easily interpretable and yields immediate confidence intervals.
- The score test and LR test are invariant under reparametrization, whereas the Wald test is not.

Resources

This tutorial is based on

- "All of statistics" Chapter 10 by Larry A. Wasserman.
- Arnaud Doucet's STA 461 Lecture notes [links].