Chapter 5 Hypothesis Testing

BY YUEJIAN MO

June 23, 2018

Here are two important areas of **statistical inference**. The first one is the estimation of **parameters**, and the second is the **testing of hypothesis**. Based on observations from a random sample, statisticians follow a formal process to determine whether or not to reject a null hypothesis(H_0). This process is called **hypothesis testing**.

Here are four steps of hypothesis testing:

- *State the hypotheses.* This involes stating the null and alternative by hypotheses.

- *Formulate an analysis plan.* The analysis plan describle how to use sample data to evaluate the null hypothesis.

- *Analyze sample data.* Find the value of the test statistic(mean score, proportion, t-statistic, z-score, etc) desribed in the analysis plan.

- *Interpret results.* Apply the decision rule decribed in the anylysis plan. If the value of the test statistic is unlikely, based on the null hypothesis, reject the null hypothesis.

A *statistical hypothesis* is an assumption about a population parameter. This assumption may or may not be true. A research can conduct a statistical experiment to test the validity of this hypothesis. If the statistical hypothesis specifies the population distribution, it is called *simple hypothesis*. Otherwise, it is called a *composite hypothesis.*

The first, the hypothesis being testing, is called the *null hypothesis*(原假设), denoted by H_0. The second is called the *alternative hypothesis*(备择假设), denoted by H_1 or H_a. Two hypotheses H_0 and H_1 divide the parameter space Θ into two subsets, concluding rejection region and acceptance region. One of thinking is that if the null hypothesis is false, then the alternative hypothesis is true, and vice versa.(But here are other thinking is 不接受不同于拒绝).

1 Type I error and Type II error, Power function

We then define rejection region and acceptance region. Let $\mathbb{S}$ be the set of all possible values of $\mathbf{x} = (X_1, ..., X_n)^T$. A test partitions $\mathbb{S}$ into two subsets: $\mathbb{C}$ and its complement $\mathbb{C}^{'}$.(other thinking: 不相交即可，无需张满完全空间).

- We reject H_0 or accept H_1 if $\boldsymbol{x} = (x_1, ..., x_n)^T \in \mathbb{C}$.

- We accept H_0 or reject H_1 if $\boldsymbol{x} = (x_1, ..., x_n)^T \in \mathbb{C}'$.

Being adjudged to be	The man is crimeless	The man is crime
Guilty(yes, 有罪假设)	Type I Error(弃真)	
Guiltess(no, 无罪假设)		Type II Error(取伪)

Table 1. 法官判决

Any decision/action wil have two outcomes: correct decision or wroing decision. We should enhance the probability of making correct decision and reduce the probility of making a wrong decision. Here we define type I error and type II error.

Rejeciton of the null hypothesis H_0 when it is ture is called *Type I error*. The probalibity of making a Type I error is denoted by

$$\begin{aligned}\alpha(\theta)=\Pr(\text{Type}\,I\,\text{error}) &=\Pr(\text{rejecting}\,H_0|H_o\,\text{is true})\\ &=\Pr(\boldsymbol{x}\in C|\theta\in\Theta_0)\end{aligned}$$

which is a function of θ defined in Θ_0. $\alpha(\theta)$ is called the *Type I error function*. When $\Theta_0=\{\theta_0\}, \alpha(\theta)=\alpha(\theta_0)\hat{=}\alpha$ is called the *Type I error rate*.

Similaritily, the probability of making a Type II error is denoted by

$$\begin{aligned}\beta(\theta) \;=\; \Pr(\text{Type}\,\mathbb{I}\,\text{error}) &\;=\; \Pr(\text{accepting}\,H_0|H_0\,\text{is false})\\ &\;=\; \Pr(\mathbf{x}\in\mathbb{C}'|\theta\in\Theta_1)\end{aligned}$$

Typically, the Type I error is more serious thant the Tyoep] II error. But we can adjust to H_0 and H_1.

	H_0 is true($\theta\in\Theta_0$)	H_0 is false($\theta\in\Theta_1$)
Reject $H_0(\boldsymbol{x}\in\mathbb{C})$	$\alpha(\theta)$(弃真)	Correct
Accept $H_0(\boldsymbol{x}\in\mathbb{C}')$	Correct	$\beta(\theta)$(取伪)

Table 2. Use type I and I error function to measure error

Let $\mathbb{C}$ be the critical region of a test for testing H_0 against H_1, then the function

$$p(\theta)=\Pr(\text{reject}\,H_0|\theta)=\Pr(\mathbf{x}\in\mathbb{C}|\theta)$$

is the *power function of the test*. The values of the power function are the probilities of rejecting the null hypothesis H_0 for variance values of the parameter θ. The power function plays the same role in hypothesis testing as that (MSE) played in estimation. The power function is golden standard in assessing the goodness of a test T or in comparing two competing tests T_1 and T_2(???). The power function can be simplied as

$$p(\theta)=\begin{cases}\alpha(\theta) & ,\theta\in\Theta_0\\ 1-\beta(\theta) & ,\theta\in\Theta_1\\ \Pr(\mathbf{x}\in\mathbb{C}|\theta) & ,\theta\notin\Theta_0\cup\Theta_1\end{cases}$$

Here we try to coose a good test. In practice, we fix the probalility of Type I error at preassigned (small) level $a^*(0 < a^* < 1)$, then minimize the probability of Type II error. That is , consider the test with

$$\sup_{\theta \in \Theta_0} p(\theta) = \sup_{\theta \in \Theta_0} \alpha(\theta) \leqslant \alpha^*$$

and choose the one with the probability of Type II error $\beta(\theta)$ being minimized. For the comparison of two tests T_1 and T_2:

- if $\alpha_{T_1}(\theta)$, $\alpha_{T_2}(\theta) \leqslant \alpha^*$ and $\beta_{T_1}(\theta) \leqslant \beta_{T_2}(\theta)$, then T_1 is better than T_2.

2 The Neyman-Pearson Lemma

If there are a set of $\{T_j\}_{j=1}^{\infty}$ to test H_0 against H_1, we would like to identify the *most powerful test*(MPT) for the case where both H_0 and H_1 are simple, and identify *the uniformly most powerful test*(UMPT) for the case where both H_0 and H_1 are composite.

Consider to test $H_0\colon \theta \in \Theta_0$ against $H_1\colon \theta \in \Theta_1 = \Theta - \Theta_0$. A test φ with critical region $\mathbb{C}$ is said to have *size* α(显著性水平) if

$$\sup_{\theta \in \Theta_0} p_{\varphi}(\theta) = \sup_{\theta \in \Theta_0} \Pr(\mathbf{x} \in \mathbb{C} | \theta) = \sup_{\theta \in \Theta_0} \alpha_{\varphi}(\theta) = \alpha$$

where $\mathbf{x} = (X_1, ..., X_n)^T$.

When H_0 is a simple null hypothesis; i.e., when $\Theta_0 = \{\theta_0\}$, from above formula, we have

$$\sup_{\theta \in \Theta_0} p_{\varphi}(\theta) = \alpha_{\varphi}(\theta_0) = \alpha$$

Here we introduce ***the most powerful test***. A test φ with critical region $\mathbb{C}$ is said to be the most powerful test with size α for testing $H_0\colon \theta = \theta_0$ against $H_1\colon \theta = \theta_1$, if

$$p_{\varphi}(\theta_0) = \alpha \qquad \text{and} \qquad p_{\varphi}(\theta_1) \geqslant p_{\psi}(\theta_1)$$

for any other test ψ with

$$p_{\psi}(\theta_0) \leqslant \alpha,$$

where θ_0 and θ_1 are two given values.

Lemma 1. *(Neyman-Pearson Lemma). Assume that* $X_1, ..., X_n \overset{\text{iid}}{\sim} f(x;\theta)$. *Let the likehood function be* $L(\theta) = L(\theta; \boldsymbol{x})$. *Then a test* φ *with critical region*

$$\mathbb{C} = \left\{ \boldsymbol{x} = (x_1, ..., x_n)^T \colon \frac{L(\theta_0)}{L(\theta_1)} \leqslant k \right\}$$

and size α *is the most powerful test of size* α *for testing* $H_0\colon \theta = \theta_0$ *against* $H_1\colon \theta = \theta_1$, *where* k *is a value determined by size* α.

the most powerful test(MPT)

the uniformly most powerful test(UMPT) 一致最优检验

$p(\theta) = \Pr(\text{rejecting } H_0|\theta) = \Pr()$

We find the UMPT as follow

Setp 1. Given two composite hypotheses H_0: $\theta \in \Theta_0$ and H_1: θ: Θ_1, we first consider two simple hypotheses H_0: $\theta = \theta_0 \in \Theta_0$ and H_1: $\theta = \theta_1 \in \Theta_1$.

Setp 2.

Setp 3.

3 Likelihood Ratio Test

The Ney-man-Pearson Lemma provides a means of constructing the most powerful crtical region for testing a simple H_0 against H_1, but it does not always apply to composite hypothesis. When a UMPT of size α does not exist, we may employ the *likehood ratio test*(LRT). The LRT is a general method/tool for finding a test statistic or constructing the critital region of a test, it can be applied to any H_0 and H_1, but the resulting test may not be optimal; while the MPT or UMPT emphasies theat the derived test of size α has the *highest power* among a class of tests with size less than or equal to α, but UMPT may not exist.

Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} f(x;\theta)$ and $x_1, \ldots, x_n$ be their realizations. Then define

The ratio

$$\lambda(\boldsymbol{x}) = \frac{\sup_{\theta \in \Theta} L(\theta)}{\sup_{\theta \in \Theta} L(\theta)}$$

is referred to as a value of the LR statistic

$$\lambda(\boldsymbol{x}) = \frac{L\big(\hat{\theta}^R\big)}{L(\hat{\theta})}$$

where $\hat{\theta}^R$ denotes the restricted MLE of θ in Θ_0.

When H_0: $\theta \in \Theta_0$ is true, since θ is not uniformly distributed in Θ_0, we wonder where is the most possible place that θ located inside Θ_0? Therefore, when $\hat{\theta}^R$ should be the global maximum. When H_0 is false, $\hat{\theta}^R$ is the local maximum; where $\hat{\theta}$ is the global maximum.

If $\lambda(\boldsymbol{x})$ is very small, we may suspect the null hypothesis. Thereforce, the cirtical region that H_0 is rejected is

$$\mathbb{C} = \{\boldsymbol{x}: \lambda(\boldsymbol{x}) \leqslant \lambda_\alpha\}, 0 < \lambda_\alpha < 1, \tag{1}$$

where $\lambda(\boldsymbol{x}) \neq 1$. The LRT of size α is test with critical region $\mathbb{C}$ given by (1), and λ_0 is determined by

$$\sup_{\theta \in \Theta_0} \Pr(\lambda(\boldsymbol{x}) \leqslant \lambda_a | \theta\} = \alpha$$

When testing a simle null hypothesis against a simpel alternatie hypothesis, the LRT will lead to the same test as that given by the Neyman-Pearson Lemma.

Calculate the LR statistic $\lambda(\boldsymbol{x}) = h(T)$ with $T = T(\mathbb{x})$ being a sufficient statistic fo θ.

Find tghe critical region $\mathbb{C}$

Check if or not $h(t)$ is monotone or long-concave;

Find an equivalent $\mathbb{C}$ involving the sufficient statistic T and a constant k.

Find the constant k via the definitionof size α by noting that a pivotal quantity $P = P(T = \theta)$ follows a certain standard distribution.

4 Test on Normal Means

Step 1: To find a test statistic

Since $\hat{X}$ is a sufficient statistic of μ and the distribution of

$$Z = \frac{\bar{X} - \mu}{\sigma_0 / \sqrt{n}}$$

is $N(0,1)$ that doest not depend on the unknown parameter μ, we know that Z is a pivotal quantity. The test statistic is

$$Z_0 \hat{=} \frac{\bar{X} - \mu_0}{\sigma_0 / n} = \frac{(\bar{X} - \mu) + (\mu + \mu_0)}{\sigma_0 / \sqrt{n}} = Z + \frac{\mu - \mu_0}{\sigma_0 / \sqrt{n}}$$

where H_0 is true, i.e., $\mu = \mu_0$, we obtain

$$Z_0 = Z \sim N(0,1).$$

Step 2: To determine a critical region of size α

Since

$$\alpha = \Pr(|Z| \geqslant z_{\alpha/2}), \alpha = \Pr(Z \geqslant z_\alpha), \alpha = \Pr(Z \leqslant -z_\alpha),$$

the critical regions of size α for the corresponding alternatives $\mu \neq \mu_0$, $\mu > \mu_0$, $\mu < \mu_0$ are given by

$$\mathbb{C}_1 = \{\boldsymbol{x}: |z_0| \geqslant z_{\alpha/2}\}, \mathbb{C}_2 = \{\boldsymbol{x}: z_0 \geqslant z_\alpha\}, \mathbb{C}_3 = \{\boldsymbol{x}: z_0 \leqslant -z_a\},$$

respectively, where

$$z_0 = \frac{\bar{x} - \mu_0}{\sigma_0 / \sqrt{n}}$$

.

4.1 p-Value

$$\begin{aligned} p-\text{value} &= 2\Pr(Z \geqslant |z_0|) & \text{if } H_1\colon \mu \neq \mu \\ &= \Pr(Z^2 \geqslant z_0^2) & \\ &= \Pr\{\chi^2(1) \geqslant z_0^2\} & \\ p-\text{value} &= \Pr(Z \geqslant z_0) & \text{if } H_1\colon \mu > \mu_0 \\ p-\text{value} &= \Pr(Z \leqslant z_0) & \text{if } H_1\colon \mu < \mu_0 \end{aligned}$$

where Z is specified by (5.29) and z_0 given by (5.30) denotes the observed value of Z_0.

Two approaches should be equivalent. The

4.2 One-sample t test

4.3 Two-sample t test

1. Find a test statisti

2. To determine a critical region

3.

5 Goodness of Fit Test(拟合优度检验)

Let $X_1, ..., X_n \overset{\text{iid}}{\sim} F(x; \boldsymbol{\theta})$ and $x_1, ..., x_n$ denote their realizations, where $\boldsymbol{\theta} = (\theta_1, ..., \theta_q)^T$ is the parameter vector.

When we obtain the data，but how to find out and check the CDF of population.

Let n bve a positive integer. If a random vector $(Y_1, ..., Y_m)^T$ has the following joint density

$$f(y_1, ..., y_m: p_1, ..., p_m) = ()$$

Theorem 5.1. Let $(N_1, ...N_m)^T \sim \text{Mulitinomia}(n; p_1, ..., p_m)$, where $n = \sum_{j=1}^m p_j = 1$. Define

$$Q_n = \sum_{j=1}^{m} \frac{(N_j - np_j)^2}{np_j}$$

Then Q_n has a limiting distribution, as n approaches infinity, the chi-square distribution with $m-1$ degrees of freedom, i.e.

$$Q_n \xrightarrow{L} \chi^2(m-1) \text{ as } \quad n \to \infty$$