



Research on Statistics in Large Sample Test

Yangjing Chong, Fangfang Guo, Wenyan Sun *

Mathematics Department, Yanbian University, Yanji, P. R. China

Email address:

513603970@qq.com (Yangjing Chong), 1024821072@qq.com (Fangfang Guo), 112277188@qq.com (Wenyan Sun)

*Corresponding author

To cite this article:

Yangjing Chong, Fangfang Guo, Wenyan Sun. Research on Statistics in Large Sample Test. *International Journal of Statistical Distributions and Applications*. Vol. 3, No. 3, 2017, pp. 61-63. doi: 10.11648/j.ijstd.20170303.16

Received: March 16, 2017; Accepted: April 7, 2017; Published: October 31, 2017

Abstract: Assuming that the test is one of the main contents of the statistical inference, it is widely used, and almost all of the major fields use hypothesis test, a statistical idea, and in the process of hypothesis testing, we need to construct a reasonable test statistic. Therefore, based on the hypothesis test and the knowledge of the central limit theorem, we construct a new test statistic when the sample size is sufficiently large, This statistic makes the calculation of the original test problem easier. This paper verifies the rationality of this test statistic by an example.

Keywords: Large Sample Test, Central Limit Theorem, Test Statistic

1. Introduction

One of the basic problems of mathematical statistics [1] is to take samples from the population to be observed or experimental, according to the limited information obtained on the whole to infer the problem, each inference must be accompanied by a certain probability to show the reliability of the inferred, This inference with a certain probability is called statistical inference, and one of the main aspects of statistical inference is the statistical hypothesis test. [1] Therefore, it is important to assume that the test is an important part of mathematical statistics, and that the test statistic is particularly important in this process. Hypothetical test [2] is the basic common statistical methods, it is an important teaching content in the course of "probability theory and mathematical statistics". Many domestic scholars on the hypothesis test conducted a more in-depth study, and achieved fruitful results. Yang Gang [3] first demonstrated the basic idea of hypothesis testing, then he given us a reasoning and derivation process for rejecting domain and accepted domain in the one sided test and two-sided test of Population parameter, and showing an application example. We know that the general steps of the hypothesis test [4] are: (1) according to practical problems to establish the original hypothesis and alternative hypothesis; (2) select the test statistic and give the denial domain; (3) select the significance level; (4) give the domain; (5) make judgments. Mao Shisong [5] puts forward the statistic

$u = \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sqrt{\sigma^2(\bar{\theta})}}$ for the large sample test construct in the course

of "probability theory and mathematical statistics". The existing literature examines how to conduct hypothetical tests from different perspectives, making it easier for readers to accept and understand hypothesis testing. Therefore, this paper attempts to start from the large sample test [6], to provide a new test statistics, broaden the reader's vision.

2. The Construction of Large Sample Test Statistic

In actual use, if the sample size is large, People often use asymptotic normal [7] structural inspection statistics to obtain large sample tests. The general idea is as follows: suppose x_1, \dots, x_n is a sample from an population distribution $F(x; \theta)$, and set the population mean is θ , the variance is a function of θ , denoted as $\sigma^2(\theta)$. For example: the distribution of two o'clock [8] $b(1, \theta)$, its variance $\sigma^2(\theta) = \theta(1 - \theta)$ is a function of the mean θ .

We are looking for a large sample test for the following three types of hypothesis testing: (θ_0 is known in practical problems)

$$I \quad H_0: \theta \leq \theta_0 \quad vs \quad H_1: \theta > \theta_0$$

$$\text{II } H_0: \theta \geq \theta_0 \quad \text{vs} \quad H_1: \theta < \theta_0$$

$$\text{III } H_0: \theta = \theta_0 \quad \text{vs} \quad H_1: \theta \neq \theta_0$$

When the sample size n is sufficiently large, we know from the central limit theorem [9] that $\bar{x} \sim N(\theta, \frac{\sigma^2(\theta)}{n})$. Therefore, If $\sigma^2(\theta)$ is the monotonic non-decreasing function of θ , the following test statistic can be used when $\theta = \theta_0$, the following test statistic can be used

$$u = \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sqrt{\sigma^2(\theta_0)}} \sim N(0, 1) \quad (1)$$

It can be approximately determined rejection region. The rejection region for the above three types of test questions [10] is

$$W_I = \{u \geq u_{1-\alpha}\},$$

$$W_{II} = \{u \leq u_\alpha\},$$

$$W_{III} = \{|u| \geq u_{1-\alpha/2}\}.$$

The approximate p-value of the test can also be calculated, exactly the same as the single-sample normal population, as omitted here. Mao Shisong puts forward the statistic $u = \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sqrt{\sigma^2(\hat{\theta})}}$ for the large sample test construct in the course

of "probability theory and mathematical statistics". Where $\hat{\theta}$ is the maximum likelihood estimate [11] of θ .

First, the rationality of the test statistic constructed in this paper is explained. Since the variance is a function of the mean value, the variance should be $\sigma^2(\theta_0)$ when the mean is θ_0 . Using the central limit theorem [9] and the relevant knowledge of the hypothesis test, we can see that the test statistic constructed in this paper is reasonable.

Compared with the statistic $u = \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sqrt{\sigma^2(\hat{\theta})}}$ proposed by Mao

Zesong, the statistic $u = \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sqrt{\sigma^2(\theta_0)}}$ has the advantage that the calculation can be simplified without the estimated value of the parameter θ and the monotonicity of $\sigma^2(\theta)$ is easy to determine.

All in all, the variance of the parameter itself is more reasonable and the calculation is more convenient than using the variance of the parameter estimation.

3. Derivation Process

For the type I hypothesis test,

When $\theta = \theta_0$, it is easy to derive that the rejection region is $W_I = \{u \geq u_{1-\alpha}\}$, which is

$$W_I = \left\{ \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sqrt{\sigma^2(\theta_0)}} \geq u_{1-\alpha} \right\}.$$

When $\theta < \theta_0$, it is known from the proof that if $u \geq u_{1-\alpha}$ is satisfied, it is necessary to ensure that

$$\frac{\sqrt{n}(\bar{x} - \theta_0)}{\sqrt{\sigma^2(\theta)}} \geq \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sqrt{\sigma^2(\theta_0)}},$$

Which is $\sigma^2(\theta) \leq \sigma^2(\theta_0)$, therefore it get evidence.

For the type II hypothesis test,

When $\theta = \theta_0$, it is easy to derive that the rejection region is $W_{II} = \{u \leq u_\alpha\}$, which is

$$W_{II} = \left\{ \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sqrt{\sigma^2(\theta_0)}} \leq u_\alpha \right\}.$$

When $\theta > \theta_0$, it is known from the proof that if $u \leq u_\alpha$ is satisfied, it is necessary to ensure that

$$\frac{\sqrt{n}(\bar{x} - \theta_0)}{\sqrt{\sigma^2(\theta)}} \leq \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sqrt{\sigma^2(\theta_0)}},$$

Which is $\sigma^2(\theta) \geq \sigma^2(\theta_0)$, therefore it get evidence.

For the type III hypothesis test,

When $\theta = \theta_0$, it is easy to derive that the rejection region is $W_{III} = \{|u| \geq u_{1-\alpha/2}\}$, which is

$$W_{III} = \left\{ \left| \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sqrt{\sigma^2(\theta_0)}} \right| \geq u_{1-\alpha/2} \right\}.$$

4. Example Analysis

The following is a concrete example to illustrate the rationality of the test statistic $u = \frac{\sqrt{n}(\bar{x} - \theta_0)}{\sqrt{\sigma^2(\theta_0)}}$ constructed in this paper.

Example: a construction company claims that its construction site under the average daily number of accidents does not exceed 0.6, now it is recorded under the command of the company's construction site 200 days of safe production, the number of accidents recorded as follows:

Table 1. 200 days of construction site safety production.

The number of accidents occurring one day	0	1	2	3	4	5	≥6	total
Days	102	59	30	8	0	1	0	200

Test the construction company's claim to be established (take $\alpha = 0.05$).

Solution: X to remember the construction site one day the number of accidents, it can be considered $X \sim P(\lambda)$, we should test the hypothesis which is:

$$H_0: \lambda \leq 0.6 \quad \text{vs} \quad H_1: \lambda > 0.6$$

Since $n = 200$ is large, a large sample test can be used. The mean and variance of the Poisson distribution are both λ , and the variance is the monotonic non-decreasing function of the mean. Calculate the mean $\bar{x} = \frac{0 \times 102 + 1 \times 59 + 2 \times 30 + 3 \times 8 + 4 \times 0 + 5 \times 1}{200} = 0.74$, meaning of the questions shows that $\lambda = 0.6$ and the test statistic is

$$u = \frac{\sqrt{n}(\bar{x} - \lambda)}{\sqrt{\lambda}} = \frac{\sqrt{200}(0.74 - 0.6)}{\sqrt{0.6}} = 2.556$$

if $\alpha = 0.05$, then $u_{0.95} = 1.645$, reject domain is:

$$W = \{u \geq 1.645\}$$

Now $u = 2.556$ has fallen into the domain, so we refused the original hypothesis and think the construction company's claim was clearly not established.

According to the method of Mao Shisong's version of the textbook, the test statistic is constructed

$$u = \frac{\sqrt{n}(\bar{x} - \lambda)}{\sqrt{\hat{\lambda}}}$$

then $\lambda = 0.6$, according to the moment estimation $\hat{\lambda} = \bar{x} = 0.74$, the test statistic is

$$u = \frac{\sqrt{200}(0.74 - 0.6)}{\sqrt{0.74}} = 2.302$$

If $\alpha = 0.05$, the same conclusion can be obtained as above.

5. Conclusion

It can be seen from the above analysis that the structure of the statistics in this paper allows us to get a new large sample test statistic, which does not need to calculate the maximum likelihood estimate of the parameter. By determining the monotonicity of the variance function, more directly to the rejection domain, making the hypothesis test becomes simple. At the same time to help you thoroughly understand the hypothesis test. In this paper, the binomial distribution is taken as an example to illustrate the rationality of the statistic structure. In addition, the construction of the statistic can be applied to the parameter checking of the Poisson distribution and so on. Since then, we will further study more convenient and simple statistics, broaden our horizons.

References

- [1] Wang Zhifu, Pan Xu, Jin Shu, Tian Feng, Wang Yan. Hypothesis test principle and its application [J]. Journal of Bohai University (Natural Science Edition), 2013, (02): 101-105.
- [2] Shi Ying. How to set hypothesis in parameter hypothesis test [J]. Journal of Mathematical Medicine, 2016, (12): 1886-1887.
- [3] Luo Ronghua, Wu Kun. A New Thinking of Hypothesis Testing [J]. Statistics & Decision, 2014, (08): 23-25.
- [4] Su Zaixing, WANG Zhi-fu, WANG Jun, HAN Dan-dan. On the relationship between hypothesis testing null hypothesis and alternative hypothesis [J]. Journal of Information Science and Technology, 2010, (21): 643 + 926.
- [5] Mao Shisong, Cheng Yiming, Pu Xiaolong. Probability Theory and Mathematical Statistics Course Second Edition. Higher Education Press.
- [6] Huo Jian. A new method of large sample normality test [J]. Statistics and Decision, 2013, (19): 67-68.
- [7] Wei Zongshu. Probability Theory and Mathematical Statistics Tutorial [M]. Beijing: Higher Education Press, 2008.
- [8] Zhu Songtao, Song Zixing. Probability Theory and Mathematical Statistics Tutorial [M]. Jinan: Shandong University Press.
- [9] Ding Jian, Li Hongju. Application of central limit theorem in statistical inference [J]. Journal of Changchun Normal University, 2015, (02): 12-14.
- [10] Yang Gang. Study on the Determination of Rejected Domain in Parameter Hypothesis Test [J]. Journal of Changchun Institute of Technology (Natural Science Edition), 2012, (02): 126-128.
- [11] Li Mingquan. On the maximum likelihood estimation method teaching [J]. Journal of Mudanjiang University, 2010, (07): 116-118.