A Hypothesis Testing Method
Based on Sample and Two Types of Errors

Yuepeng Yang

Advisor: Wentao Zhang

Shenzhen Middle School

Abstract

Hypothesis Testing is an important statistical tool. It is used to judge whether a hypothesis should be rejected or accepted. In hypothesis testing, two types of errors are possible: type I error which indicates rejecting null hypothesis H_0 when it's true, and type II error which refers to accepting H_0 when it's false. Traditional hypothesis testing controls the possibility of Type I error by setting α. While in researches and industries, we have many cases where a well consideration of both types of errors is necessary, such as controlling the economical lost in quality testing and decision making. Thus, we raise a new approach to parametric tests to take the risks caused by both types of error into account and apply the method on the test of population mean.

First, we provide a method to calculate the possibility of Type I error α and Type II error β. Previous researchers calculate β by giving the distribution parameter θ a prior distribution $\pi(\theta)$. We, on the other hand, choose use empirical Bayesian method to give the empirical prior distribution $\pi_S(\theta)$ of parameter θ with the obtained Sample S. Thereby, we get the formulas to calculate α and β:

$$\alpha = \int_{H_0} \alpha_\Omega(\theta)\pi_S(\theta)d\theta \, , \beta = \int_{H_1} \beta_\Omega(\theta)\pi_S(\theta)d\theta$$

To take the impact of both types of errors into consideration, we introduce target function T to measure the risks caused by the two types of errors. T can be calculated with the sample S and rejection region Ω. Then, we optimize T to obtain an ideal rejection region Ω_S that leads to a minimal T. After getting Ω_S from the sample S and target function T, we are able to judge directly whether we should reject null hypothesis H_0 or not.

We also apply this general method to the specific test of population mean μ with known variance. We discuss the numerical relation of α, β, T, μ_S in this case, and solve the rejection region under different target functions.

Overall, we give a new hypothesis testing method base on sample and Type I and II errors to minimize the risks caused by these errors. We also implement the method numerically on a common test.

Key Words: Hypothesis Testing, Type I & II errors, Target Function, Optimization, Empirical Bayes

1 Introduction and Notation

1.1 Introduction

Hypothesis Test is an important tool in Statistics which is widely used in researches and industry. It is used to test if there is enough evidence to reject certain hypothesis. Usual way to conduct a Hypothesis Test is to provide a pair of null hypothesis H_0 and alternative hypothesis H_1 as well as a significance level α and to find a rejection region that when H_0 is true, the probability to reject H_0 is α.

This tradition hypothesis testing method is clear and direct, but it pays little attention to Type II error. As we know[1], Type I error is rejecting the null hypothesis when it is true, while Type II error is failing to reject the null hypothesis when it is false. In some cases, Type II error is also significantly important. For example, in quality tests, economical loss are produce both when qualified products are rejected and when disqualified ones fail to be rejected. Thus, minimizing the loss brought by two types of errors come to be crucial.

Researchers have noticed that in the same hypothesis test with the same sample, β decreases when α increases. In parameter test, β can be calculated by assuming the parameter of population distribution. Wu[2](1990) calculates β by assuming the actual population mean μ.

Concurrently, there's people consider about synthesizing α and β in hypothesis test. Degroot[3](1975) designed a special pair of hypothesis and determine the rejection region with the weighted average of α and β. Pericchi 和 Pereira[4](2013) used Bayesian methods to expand Degroot's method in general hypothesis and calculates α and β by granting distribution parameter θ a set of prior distributions.

Our goal is to design a method that both synthesize Type I and II method that does not require people to give a prior distribution of parameter θ.

1.2 Notation

H_0	Null hypothesis
H_1	Alternative hypothesis
α	Probability of Type I error, Significance level
β	Probability of Type II error
Ω	Rejection region
S	Sample
θ	Distribution parameter
θ_S	Sample estimator of distribution parameter
θ_Ω	The boundary of rejection region in test of distribution parameter
$\pi(\theta)$	Prior distribution of distribution parameter
μ	Population
μ_S	Sample mean
μ_Ω	The boundary of rejection region in test of population mean
L，$L(\alpha, \beta)$	Target Function

2 Calculating the probability of Type I and Type II errors

To synthesize of two types of errors, the first thing to clarify is how to calculate their probability. Unlike α, the probability of Type I error, calculating β the probability of Type II error is not direct.

2.1 Existing Calculating Methods

In conventional hypothesis test, the probability of Type I error α is significance level, which is the probability when sample mean is in the rejection region when the null hypothesis is true.

Degroot provided the following method in 1975. He points out that, in test of在 population mean, when null hypothesis and alternative hypothesis looks like

$$H_0: \mu = \mu_0 \text{ VS } H_1: \mu = \mu_1 \tag{1}$$

where $\mu_0 \leq \mu_1$, the rejection region is

$$\Omega = \left\{ \mu_S \middle| \mu_S \geq \mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}} \right\} \tag{2}$$

Where α is the significance level and n is the sample size.

Since null hypothesis has given μ_0, when σ is known, the rejection region depends on α, so the rejection region can be written in the form of $\Omega(\alpha)$. Then, the probability of Type II error is

$$\beta = P\{\text{fail to reject } H_0 | H_1\} = \int_{-\infty}^{\mu_0 + z_\alpha \frac{\sigma}{\sqrt{n}}} f(x) dx \tag{3}$$

in which

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(\mu_S - \mu_1)^2}{2\sigma^2}} \tag{4}$$

However, when hypothesis test has the form of

$$H_0: \mu = \mu_0 \text{ VS } H_1: \mu > \mu_0 \tag{5}$$

β depends on the actual value of μ. So when μ is unknown, β is not easy to calculate. One approach[2] is to assume $\mu = \mu_1$, and then calculate β with formula (3)

and (4). Similarly, Pericchi 和 Pereira[3] adopted Bayesian method to discuss in a broader situation. They set the null hypothesis and alternative hypothesis as

$$H_0: \theta \in \Theta_0 \quad \text{VS} \quad H_1: \theta \in \Theta_1 \tag{6}$$

Where θ is distribution parameter, Θ is the set where θ is hypothesized to be in. They give θ in H_0 and H_1 prior distribution $\pi_0(\theta)$ and $\pi_1(\theta)$ respectively and guarantee that

$$\int_{\Theta_0} \pi_0(\theta)d\theta + \int_{\Theta_1} \pi_1(\theta)d\theta = 1$$

$$\tag{7}$$

. Then they use

$$\alpha(\delta) = \int_{\Theta_0} \alpha(\delta)\pi_0(\theta)d\theta$$

$$\tag{8}$$

$$\beta(\delta) = \int_{\Theta_1} \beta(\delta)\pi_1(\theta)d\theta$$

$$\tag{9}$$

to obtain the probability of Type I and II error in test δ respectively.

2.2 The method of calculating α and β based on sample

To avoid reliance on the prior distribution mentioned in 2.1, we developed a new method with the help of Pericchi and Pereira's method. We used empirical Bayesian method so that it is not necessary to give a prior distribution beforehand. We treat distribution θ as a random variable and used sample S to obtain the empirical prior distribution of θ: $\pi_S(\theta)$, this distribution expresses the probability density of θ in its possible range when S is collected.

$$\int_A \pi_S(\theta) \, d\theta = P(\theta \in A|S)$$

$$\tag{10}$$

This calculation method is similar to calculating confidence interval. It is also worth-noted that as we are using sample to estimate the distribution of population distribution parameter, we do not have to give prior distribution $\pi_0(\theta)$ and $\pi_1(\theta)$ to the null hypothesis and alternative hypothesis as what Pericchi and Pereira does.

Now we are considering the cases when parameter θ equals to its possible candidates. For parameter $\theta = \theta_0$, where $\theta_0 \in H_0$

$$\alpha_\Omega(\theta_0) = P(\theta_S \in \Omega | \theta = \theta_0) \tag{11}$$

$$\beta_\Omega(\theta_0) = 0 \tag{12}$$

For $\theta = \theta_1$, where $\theta_1 \in H_1$

$$\alpha_\Omega(\theta_1) = 0 \tag{13}$$

$$\beta_\Omega(\theta_1) = P(\theta_S \notin \Omega | \theta = \theta_1) \tag{14}$$

Now we have obtained the calculation method of α and β when we have known θ and Ω. After obtaining the empirical prior distribution of θ, $\pi_S(\theta)$, we include it in our calculation. For a rejection region Ω, the probability of Type I and Type II error is respectively

$$\alpha_\Omega(\theta) = P(\theta_S \in \Omega | H_0) = \int_{H_0} \alpha_\Omega(\theta)\pi_S(\theta)d\theta$$

$$\tag{15}$$

$$\beta_\Omega(\theta) = P(\theta_S \notin \Omega | H_1) = \int_{H_1} \beta_\Omega(\theta)\pi_S(\theta)d\theta$$

$$\tag{16}$$

In the formula, $\alpha_\Omega(\theta)$ and $\beta_\Omega(\theta)$ are the probability of Type I and Type II error when θ equals a certain value. With Formula (15) and (16), it's possible to calculate α and β directly without having to give a prior distribution beforehand.

2.4 Comparison of the calculating methods

Existing methods calculate with two approaches: the first adopts special alternative hypothesis $H_1: \theta = \theta_1$; the second grants θ prior distribution. When the hypothesis or prior distribution is accurate, these two methods have good performance. Yet this indicates that these methods heavily rely on the accuracy of hypothesis or prior distribution.

Our method, however, is based on actual sample. We use sample to derive empirical prior distribution so that we do not have to provide a prior distribution

beforehand. The empirical prior distribution can also fit actual situation better and shows a better robustness than prior distribution provided manually.

3 Derive Rejection Region with Target Function

Our purpose is to find an index that would evaluate the risk of Type I and Type II error. Thus, we constructed a target function $L(\alpha, \beta)$ and derive an ideal rejection region Ω by optimizing $L(\alpha, \beta)$.

3.1 Establishing Target Function

To synthesize α and β, Pericchi and Pereira[3] use simple weighted addition to establish their target function

$$SERRORS(\delta) = a \cdot \alpha(\delta) + b \cdot \beta(\delta) \tag{17}$$

However, we believe that the target function, which is desired to evaluate the risk when Type I and Type II errors take place, would not necessarily be a concise linear addition. It is necessary to consider the non-linear relation between the risk and α or β. It is also possible that α and β would have interactive effect on the risk. Therefore, we construct the target function in a more general term:

$$L(\alpha, \beta) \tag{18}$$

When the two type of error have only linear relation with the target function and no interactive effect, we can write the target function in the form of
$$L(\alpha, \beta) = a\alpha + (1 - a)\alpha$$
in which $0 < a < 1$. This expression can standardize the linear addition by making the sum of weight equal 1 and help us to do comparison among target functions in linear form.

3.2 Optimizing Target Function

In 2.2, we have already provided a method to calculate α and β with Sample S and rejection region Ω. Therefore, we can write $L(\alpha, \beta)$ in the form of $L(\Omega, S)$, by optimizing function

$$L(\Omega, S) \tag{19}$$

We can obtain a rejection region Ω_S so that when we have Sample S, the target

function $L(\Omega, S)$, in other words the risk of having two types of error, is smallest when we use Ω_S as the rejection region.

After obtaining rejection region Ω, we can use Ω and S to decide whether we should rejection null hypothesis or not. When

$$S \in \Omega_S \tag{20}$$

, we reject null hypothesis. One worth-noted point is that the rejection region Ω_S here relies on Sample S. Therefore, we can have the rejection region in general

$$\Omega = \{S | S \in \Omega_S\} \tag{21}$$

that includes all Samples that we should reject.

As long as we get Sample S and provide target function L, we are able to tell whether we should reject null hypothesis or not. At same time, we have to obtain S and provide α in traditional hypothesis testing. The important difference between α and L is that α shows the probability of Type I error. In real applications, it is usually given by convention, such as 0.1 or 0.05. On the other hand, target function has a very specific meaning in reality: the total risk of having Type I and II errors. The method can optimize the loss produced by Type I and Type II error which brings it a good prospect in applied field.

4 Case Study: Hypothesis Test of Population Mean

In section 2 and 3, we have covered the general method in use Sample S and target function L to minimize the total risk of Type I and Type II error. Now we are going to use it in a common test: test of population mean μ. Though the numerical calculation in this specific case, we can have a more direct understanding of the method we have designed.

According to central limit theorem[1], for independent Sample following same 对 distribution, when sample number is large enough sample mean $\mu_S \sim N(\mu, \sigma^2/n)$.

Now we have this random variable X whose population mean is μ (unknown) and population variance is σ^2; assume we know σ, we test the following null hypothesis:

$$H_0: \mu \le \mu_0 \tag{22}$$

and alternative hypothesis:

$$H_1: \mu > \mu_0 \tag{23}$$

Also, we collect Sample S whose sample size is n; the sample mean is μ_S, whereas the variance of μ_S is σ^2/n. To simplify, we make a linear transformation $X' = \frac{X - \mu_0}{\sigma/\sqrt{n}}$ so that $\mu_0 = 0$ and $\sigma^2/n = 1$.

4.1 The relation between β and α and μ_Ω and μ_S

First, we treat μ as a random variable. For distinction, we call it x here. According to central limit theorem[1], sample mean $\mu_S \sim N(x, 1)$, so that when we have sample S, the probability density function of x is

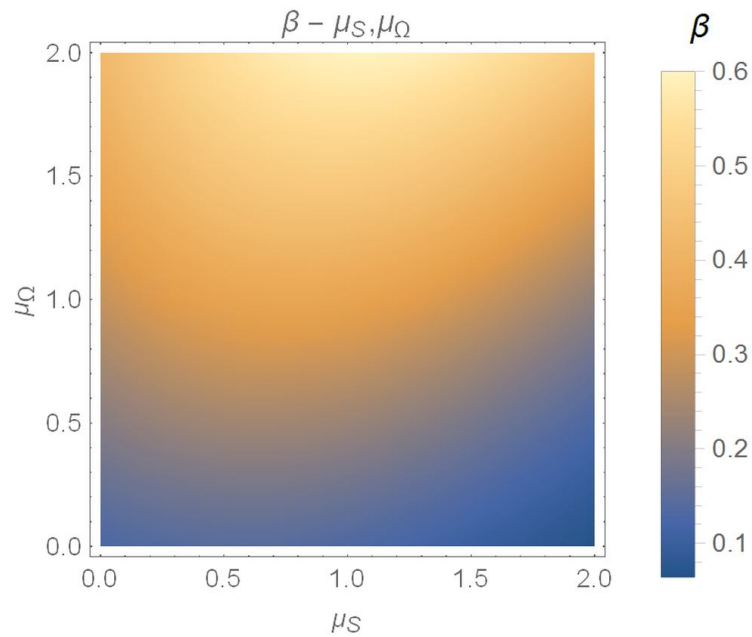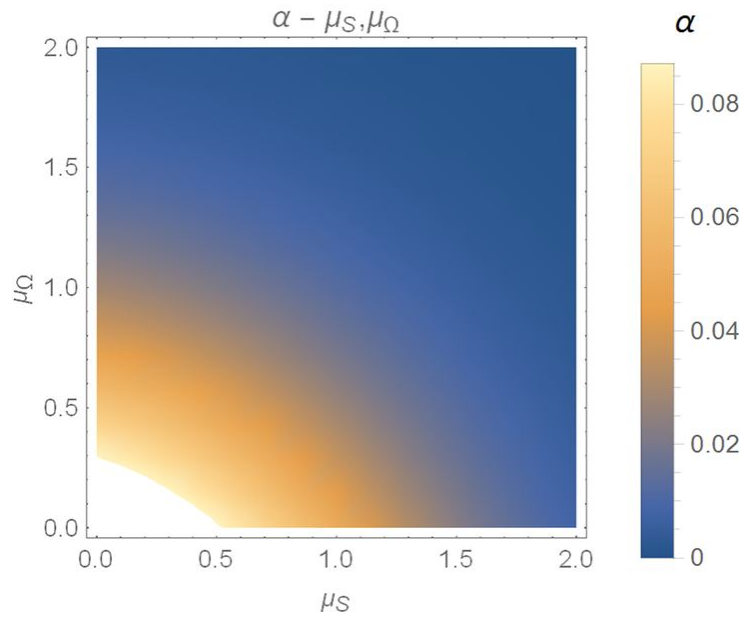$$\pi_S(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x - \mu_S)^2}{2}} \tag{24}$$

which is our empirical distribution.

Using formula (15) and (16), and rejection region $\Omega = \{\mu_S > \mu_\Omega\}$, α and β equals:

$$\alpha_\Omega(\mu) = P(\mu_S \in \Omega | H_0) = \int_{H_0} \alpha_\Omega(\mu) \pi_S(\mu) d\mu$$

$$= \int_{-\infty}^{0} \int_{\mu_\Omega}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(\mu - \mu_S)^2}{2}} d\mu \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{(x - \mu_S)^2}{2}} dx$$

$$\tag{25}$$

$$\beta_\Omega(\mu) = P(\mu_S \notin \Omega | H_1) = \int_{H_1} \beta_\Omega(\theta) \pi_S(\mu) d\theta$$

$$= \int_{0}^{\infty} \int_{-\infty}^{\mu_\Omega} \frac{1}{\sqrt{2\pi}} e^{-\frac{(\mu - \mu_S)^2}{2}} d\mu \frac{1}{\sqrt{2\pi}} e^{-\frac{(x - \mu_S)^2}{2}} dx$$

$$\tag{26}$$

For different μ_Ω and μ_S, the α and β changes as in graph 1 and 2

Graph 1 $\quad \alpha$ with different μ_Ω and μ_S

Graph 2 $\quad \beta$ with different μ_Ω and μ_S

In the graph, color indicated the probability of having the type of error. As we see, when μ_S increases, α gradual decrease while β increase first and then decrease; when μ_Ω, the boundary of rejection region increases, α decreases gradually while β increases.

4.2 Establish and Optimize Target Function to Derive the Rejection Region

Using formula (18) in section 3.1, we can evaluate the loss cause by Type I and Type II error with target function $L(\alpha, \beta)$. Specifically, when the loss have linear relation with α and β independently, we can use formula (19) to get

$$L_\Omega(S) = a\alpha_\Omega(\mu) + (1-a)\beta_\Omega(\mu) \tag{27}$$

where $0 < a < 1$. For each target function, we can find the right rejection region Ω_S to minimize $L_\Omega(S)$. Now we use $a=0.9$ to analysis how $L_\Omega(S)$ changes with μ_Ω with fixed μ_S.
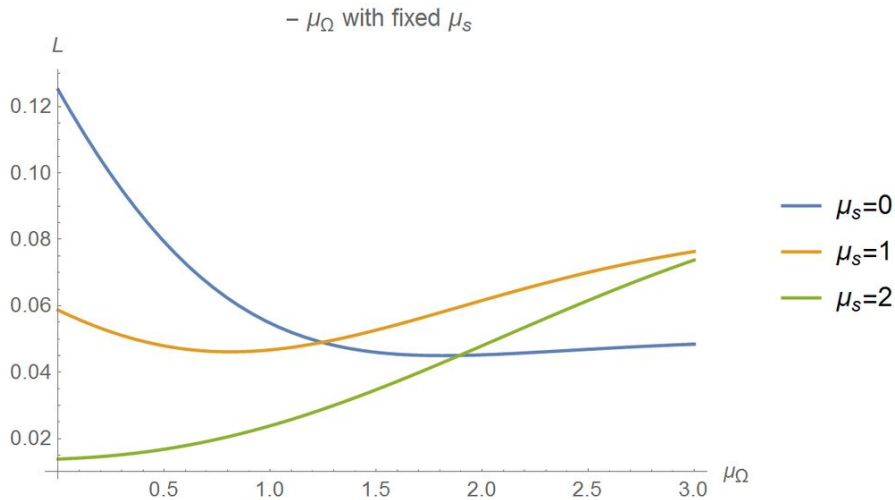
Graph 3 L changing with μ_Ω when $\mu_S = 0, 1, 2$

Now we should try the right μ_Ω to get the smallest $L_\Omega(S)$. It is observed that when μ_S increases, the μ_Ω we should choose is smaller.

We also want to know the variation of μ_Ω when we adopt different target function L. When a equals different number, $L_\Omega(S)$ changes with μ_Ω as Graph 4 shows.
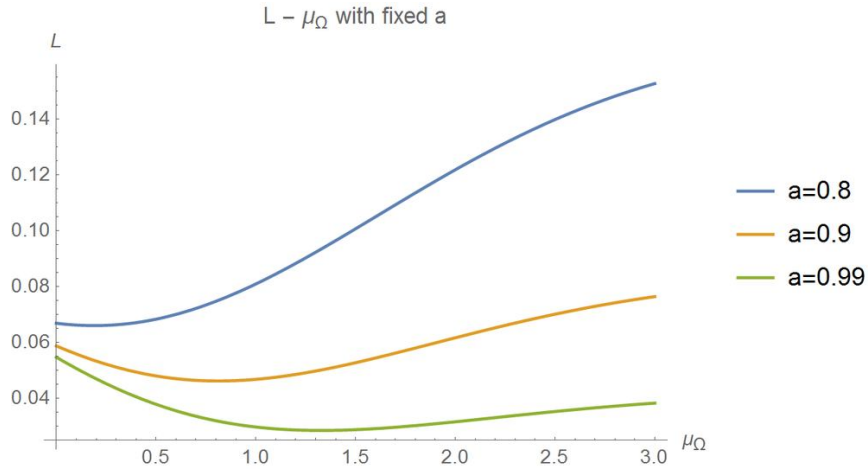
Graph 4 L changing with μ_Ω when $a = 0.8, 0.9, 0.99$

where $\mu_S = 1$. We see that as a increases, the μ_Ω we should choose decreases.

4.3 Deriving Rejection Region

For each sample S, we can have a rejection region Ω_S. Using this formula

$$\Omega = \{S | S \in \Omega_S\} \tag{28}$$

we can have the rejection region for the test.

We choose a few target function in the form of $L_\Omega(S) = a\alpha_\Omega(\mu) + (1 - a)\beta_\Omega(\mu)$, to calculate its rejection region Ω, the results are listed:

Table 1 μ_Ω with different a

a	0.5	0.8	0.9	0.99	0.999
μ_Ω	0	0.59	0.91	1.64	2.18

After we give L, we can use sample S to derive rejection region Ω. Although 尽 this calculation method may not be direct, it minimizes the risk of having Type I and II error. We can give different target function for effective usage in different cases.

We can also see that when $a = 0.99$, $\mu_\Omega = 1.64$. Whereas in traditional hypothesis test, when $\mu_\Omega = 1.645$ when significance level is 0.05. When α has much larger important β, the test is close to the traditional test.

5 Conclusion

In many real applications, we care about the total risk brought by Type I and Type II errors. Consequently, we need to design a hypothesis testing method to minimize their risk. Our hypothesis testing method starts with this basic consideration. Previously, researchers have granted distribution parameter a prior distribution $\pi(\theta)$ to calculate α, the probability of Type I error, and β, the probability of Type II error. However, instead of using this method depending on prior distribution given by people, we use empirical Bayesian method to determine the empirical prior distribution of parameter θ to calculate α and β. After that, we give a target function $L(\alpha, \beta)$ instead of giving significance level as in traditional hypothesis test. With the obtained sample S, we optimize $T_{\Omega}(S)$ to get the ideal Ω_S and then decide whether to reject the null hypothesis or not by testing if $S \in \Omega_S$.

Our hypothesis testing method is not direct comparing to traditional method. However, we successful take Type II error into consideration. In some real life applications such as quality test of industrial products, both types of error lead to certain loss. Traditional method control the probability of Type I error under the significance level while the method we established consider both types of error to minimize the total loss. This feature gives our method good prospect in applied areas.

Due to the limited time, we still have ideas not analyzed yet. We are curious about worth-thinking question such as the application of this method in non-parameter test and how non-linear target function affect the rejection region.

Reference

[1] Sheng, Z., & Pan, Y. (2010). Probability Theory and Statistcis (Fourth Edition). Higher Education Publication.

[2] Wu, G. (1990) The Determination of Two Types of Error and Rejection Region in Statistical Test. Xi'an Geology Academy Journal, 12(1).

[3] DeGroot, M. (1975) Probability and Statistics, 2nd Edition. New York. Addison-Wesley.

[4] Pericchi, L. and Pereira C. (2013) Changing the paradigm of fixed significance levels: Testing Hypothesis by Minimizing Sum of Errors Type I and Type II

[5] Shen, Y. (2006) Discussion and Application of Empirical Bayesian Method in multi-parameter estimation problems.