Communications in Statistics—Theory and Methods, 35: 1361–1374, 2006

ISSN: 0361-0926 print/1532-415X online
DOI: 10.1080/03610920600694611

Taylor & Francis
Taylor & Francis Group

Inference

Modified *p*-Value of Two-Sided Test for Normal Distribution with Restricted Parameter Space

HSIUYING WANG

Institute of Statistical Science, Academia Sinica, Taipei, Taiwan

This article proposes a modified p-value for the two-sided test of the location of the normal distribution when the parameter space is restricted. A commonly used test for the two-sided test of the normal distribution is the uniformly most powerful unbiased (UMPU) test, which is also the likelihood ratio test. The p-value of the test is used as evidence against the null hypothesis. Note that the usual p-value does not depend on the parameter space but only on the observation and the assumption of the null hypothesis. When the parameter space is known to be restricted, the usual p-value cannot sufficiently utilize this information to make a more accurate decision. In this paper, a modified p-value (also called the rp-value) dependent on the parameter space is proposed, and the test derived from the modified p-value is also shown to be the UMPU test.

Keywords *p*-Value; *rp*-Value; Two-sided test; Uniformly most powerful unbiased test.

AMS Mathematics Classifications Primary 62F03; Secondary 62F30.

1. Introduction

Traditionally, statistical theory is established in natural parameter space. However, in many real applications, the parameter space is restricted, and the methodology established from the natural parameter space does not sufficiently utilize the important information regarding the restriction. This bounded parameter space problem has been discussed in recent literature. Mandelkern (2002) gave the example that the class Neyman procedure is not satisfactory to many scientists where the parameter is known to be bounded. This problem occurs frequently in analyzing the data from physics experiments. Feldman and Cousins (1998) and Roe and Woodroofe (2001) proposed several alternatives for setting confidence bounds in this situation. In this paper, we mainly focus on the bounded parameter space in hypothesis testing.

Received February 19, 2005; Accepted November 23, 2005
Address correspondence to Hsiuying Wang, Institute of Statistical Science, Academia Sinica, Taipei 115, Taiwan; E-mail: hywang@stat.sinica.edu.tw

Let $X_1, \ldots, X_n$ be a random sample from a $N(\theta, \sigma^2)$ population. The location parameter θ is the parameter of interest. Assume that the parameter space Ω is restricted and the bounds of the parameter space are known. For example, if we are interested in testing if the average weight of newborn infants for mothers in a certain group is less than a certain value, the mean of the weight is certainly bound, and the bound information can be obtained from empirical experience. Another example is to test if the mean of the length of products produced by a machine is less than or greater than a standard length. In this case, we can measure some products first, then obtain rough upper and lower bounds for the mean. From these examples, we can conclude that for most applications, it is not difficult to find rough bounds for the parameter of interest, and we can assume that Ω has a lower bound, an upper bound, or both bounds. In this paper, we consider the two-sided hypothesis

$$H_0 : \theta = \theta_0 \quad \text{versus} \quad H_1 : \theta \neq \theta_0. \tag{1}$$

Situations of σ^2 both known and unknown will be discussed. For the two-sided hypothesis (1), there does not exist a uniformly most powerful test (see Lehmann, 1983). A uniformly most powerful unbiased (UMPU) test is usually used for testing (1). When σ^2 is known, without loss of generality, σ^2 is assumed to be 1. The rejection region of a level α UMPU test is $\{x : |\bar{x} - \theta_0| \geq c/\sqrt{n}\}$, where c is the $\alpha/2$ upper cutoff point of a standard normal distribution. For an observation $\bar{x}$, the corresponding p-value of the test is $P(|Z| \geq \sqrt{n}|\bar{x} - \theta_0|)$ where Z has a standard normal distribution. When σ^2 is unknown, the rejection region of a level α UMPU test is $\{x : |\bar{x} - \theta_0|/\sqrt{s_n^2/n} \geq c\}$ where $s_n^2 = \sum_{i=1}^n (x_i - \bar{x})^2/(n-1)$ and c is the $\alpha/2$ upper cutoff point of the t distribution with degrees of freedom $n-1$. The p-value corresponding to observations $\bar{x}$ and s_n^2 is $P(|T| \geq |\bar{x} - \theta_0|/\sqrt{s_n^2/n})$, where T has a t distribution with degrees of freedom $n-1$. When the p-value is small, the null hypothesis H_0 tends to be rejected, otherwise, H_0 is not rejected. Thus the p-value is implicitly used as a measure of evidence against the null hypothesis. However, using only the p-value as evidence might lead to a wrong decision (see Berger and Wolpert, 1984). Hwang et al. (1992) also point out the need for evidence evaluation.

Note that the usual p-values of the UMPU tests do not depend on the parameter space of θ. That is, even if we have more information about the parameter space, the usual p-value cannot reflect the merit of having the information, with the result that the restriction information cannot help us to estimate θ more accurately if we use the p-value as evidence against the null hypothesis. There are also other criticisms leveled at the p-value, e.g., Lindley (1957) and Berger and Delampady (1987), etc. Most criticisms of the p-value in the literature are not from the point of view of restriction of the parameter space. In this paper, we propose a modified p-value from the point of view of bounded parameter space.

A measure of evidence against the null hypothesis is proposed in this paper. Since the p-value is widely used, we may call the proposed evidence a modified p-value. However, since the modified p-value is not the usual probability, to avoid confusion, the proposed evidence is also called the rp-value, which means a modified measure of evidence from the usual p-value based on the restricted parameter space. For the other testing problems regarding restricted parameter space, Woodroofe and Wang (2000) provide a modified p-value for the one-sided testing problem of the Poisson distribution. For the simple hypothesis versus the simple alternative hypothesis testing problem and the one-sided testing problem, Wang (2004, 2005) has proposed modified *p*-values better than the usual *p*-value for some distributions.

This paper is organized as follows. A modified measure of evidence depending on the bound of the parameter space is proposed in Sec. 2. In Sec. 3, the proposed method is illustrated using a real data example, and simulation results comparing the *p*-value and the *rp*-value are presented. The advantage of the *rp*-value from a testing point of view is demonstrated in Sec. 4. Section 5 examines the *p*-value and the *rp*-value according to a criterion provided in the literature.

2. The *rp*-Values

A measure of evidence against the null hypothesis dependent on the parameter space is proposed in this section. Suppose that the parameter space has a lower bound a and an upper bound b. Let $X_1, \ldots, X_n$ be a random sample from an $N(\theta, \sigma^2)$. First, we consider the case of σ^2 known and assume $\sigma^2 = 1$. The usual p-value of the UMPU test of (1) with respect to an observation $\bar{x}$ is

$$p_{\theta_0}(\bar{x}) = P\left(|Z| > \frac{|\bar{x} - \theta_0|}{\sqrt{1/n}}\right), \tag{2}$$

where Z denotes a random variable of standard normal distribution.

For a fixed observation $\bar{x}$, we propose

$$r_{\theta_0}(\bar{x}) = \frac{P\left(|Z| > \frac{|\bar{x}-\theta_0|}{\sqrt{1/n}}\right) - \min_{\theta\in(a,b)} P\left(|Z| > \frac{|\bar{x}-\theta|}{\sqrt{1/n}}\right)}{\max_{\theta\in(a,b)} P\left(|Z| > \frac{|\bar{x}-\theta|}{\sqrt{1/n}}\right) - \min_{\theta\in(a,b)} P\left(|Z| > \frac{|\bar{x}-\theta|}{\sqrt{1/n}}\right)} \tag{3}$$

as a modified measure of evidence against the null hypothesis. The motivation for proposing (3) is as follows.

The range of (2) is

$$\left(\min_{\theta_0\in(a,b)} P\left(|Z| > \frac{|\bar{x} - \theta_0|}{\sqrt{1/n}}\right), \max_{\theta_0\in(a,b)} P\left(|Z| > \frac{|\bar{x} - \theta_0|}{\sqrt{1/n}}\right)\right)$$

when θ_0 belongs to (a, b). No matter what value of $\theta_0 \in (a, b)$ is chosen to be the null hypothesis $\theta = \theta_0$, the p-value $p_{\theta_0}(\bar{x})$ is always greater than $\min_{\theta_0\in(a,b)} P\big(|Z| > |\bar{x} - \theta_0|/\sqrt{1/n}\big)$. Thus the magnitude $\min_{\theta_0\in(a,b)} P\big(|Z| > |\bar{x} - \theta_0|/\sqrt{1/n}\big)$ should not be included in a measure of evidence against the null hypothesis. Hence first we suggest that a reasonable measure of evidence should be the usual p-value minus this magnitude, which is

$$p_{\theta_0}(\bar{x}) - \min_{\theta_0\in(a,b)} P\left(|Z| > \frac{|\bar{x} - \theta_0|}{\sqrt{1/n}}\right). \tag{4}$$

Moreover, usually we would decide to reject the null hypothesis or accept the null hypothesis by comparing the p-value with a value α, where $0 < \alpha < 1$. If we can transform (4) so that the range is $(0, 1)$, then it is more reasonable to compare it with a value α between 0 and 1. Thus, for fixed $\bar{x}$, we divide (4) by

$$\left(\max_{\theta_0\in(a,b)} P\left(|Z| > \frac{|\bar{x}-\theta_0|}{\sqrt{1/n}}\right) - \min_{\theta_0\in(a,b)} P\left(|Z| > \frac{|\bar{x}-\theta_0|}{\sqrt{1/n}}\right)\right)$$

so that its range is $(0, 1)$, which leads to (3). Note that $r_{\theta_0}(\bar{x})$ is an *x-dependent* transformation of $p_{\theta_0}(\bar{x})$, because $\max_{\theta_0\in(-\infty,\infty)} P(|Z| > |\bar{x}-\theta_0|/\sqrt{1/n}) = 1$ and $\min_{\theta_0\in(a,b)} P(|Z| > |\bar{x}-\theta_0|/\sqrt{1/n}) = 0$.

Under this transformation, the range of (3) is $(0, 1)$. Note that when the parameter space is the natural parameter space, (3) is exactly equal to the usual p-value (2). As mentioned in Sec. 1, since (3) is a modification of the p-value, we call it an *rp*-value, because it is a modification of the p-value based on restricted parameter space.

For the case of σ^2 unknown, the usual p-value is

$$p'_{\theta_0}(\bar{x}) = P_{\theta_0}\left(|T| > \frac{|\bar{x}-\theta_0|}{\sqrt{s_n^2/n}}\right).$$

We propose

$$r'_{\theta_0}(\bar{x}) = \frac{P\left(|T| > \frac{|\bar{x}-\theta_0|}{\sqrt{s_n^2/n}}\right) - \min_{\theta\in(a,b)} P\left(|T| > \frac{|\bar{x}-\theta|}{\sqrt{s_n^2/n}}\right)}{\max_{\theta\in(a,b)} P\left(|T| > \frac{|\bar{x}-\theta|}{\sqrt{s_n^2/n}}\right) - \min_{\theta\in(a,b)} P\left(|T| > \frac{|\bar{x}-\theta|}{\sqrt{s_n^2/n}}\right)} \qquad (5)$$

as the *rp*-value for the case where T has a t distribution with degrees of freedom $n-1$. Note that $r'_{\theta_0}(\bar{x})$ is also equal to the usual p-value if the parameter space is the natural parameter space.

When using the *rp*-value, confidence intervals of θ can be constructed. A $1-\alpha$ confidence interval based on the *rp*-value is

$$\left\{\theta_0 : \frac{P\left(|T| > \frac{|\bar{x}-\theta_0|}{\sqrt{s_n^2/n}}\right) - \min_{\theta\in(a,b)} P\left(|T| > \frac{|\bar{x}-\theta|}{\sqrt{s_n^2/n}}\right)}{\max_{\theta\in(a,b)} P\left(|T| > \frac{|\bar{x}-\theta|}{\sqrt{s_n^2/n}}\right) - \min_{\theta\in(a,b)} P\left(|T| > \frac{|\bar{x}-\theta|}{\sqrt{s_n^2/n}}\right)} > \alpha\right\}, \qquad (6)$$

which can be rewritten as

$$\left\{\theta : \frac{|\overline{X}-\theta|}{\sqrt{s_n^2/n}} < t_{\beta/2}\right\},$$

where

$$\beta = \alpha\left(\max_{\theta\in(a,b)} P\left(|T| > \frac{|\bar{x}-\theta|}{\sqrt{s_n^2/n}}\right) - \min_{\theta\in(a,b)} P\left(|T| > \frac{|\bar{x}-\theta|}{\sqrt{s_n^2/n}}\right)\right) + \min_{\theta\in(a,b)} P\left(|T| > \frac{|\bar{x}-\theta|}{\sqrt{s_n^2/n}}\right)$$

and $t_{\beta/2}$ is the $\beta/2$ upper cutoff point of the t distribution with degrees of freedom $n-1$. Note that β depends on the observation $\bar{x}$. Since this paper focuses on hypothesis testing, the properties of confidence interval (6) remain under investigation.

3. Examples

In this section, the proposed method is illustrated using a real data example. Data on 189 infant birth weights was collected at Baystate Medical Center, Springfield, Massachussets, in 1986. In this example, the population is the entire data set, 189 infant birth weights. The mean and variance of these weights are 2944 g and 729^2 g, respectively. We assume that infant birth weight follows a normal distribution $N(\theta, \sigma^2)$. Suppose that the researchers do not know all the data. A sample of size n is chosen from the data, and they will use the sample to test (1), where θ is the mean of the data. In this example, we assume that according to empirical information, the mean of baby weights has a lower bound a and an upper bound b. If $a = 2700$ and $b = 3200$ and we have a sample of five weights: 2600, 2055, 3062, 3232, 1970, the mean of the sample is 2583.8. Let θ_0 be 2944. Then the usual p-value of the UMPU test is

$$P\left(|Z| > \frac{2944 - 2583.8}{\sqrt{729^2/5}}\right) = 0.269.$$

Before calculating the rp-value, we need to derive the values of $\max_{\theta\in\Omega} P(|Z| \geq |2583.8 - \theta|/\sqrt{729^2/5})$ and $\min_{\theta\in\Omega} P(|Z| \geq |2583.8 - \theta|/\sqrt{729^2/5})$, where $\Omega = (2700, 3200)$. It is very easy to calculate these values because $\max_{\theta\in\Omega} P(|Z| \geq |2583.8 - \theta|/\sqrt{729^2/5})$ happens at $\theta = 2700$, which is

$$P\left(|Z| > \frac{2700 - 2583.8}{\sqrt{729^2/5}}\right) = 0.721,$$

and $\min_{\theta\in\Omega} P(|Z| \geq |2583.8 - \theta|/\sqrt{729^2/5})$ happens at $\theta = 3200$, which is

$$P\left(|Z| > \frac{3200 - 2583.8}{\sqrt{729^2/5}}\right) = 0.059,$$

By definition, the rp-value is $(0.269 - 0.059)/(0.721 - 0.059) = 0.318$. In this case, H_0, which is true, will not be rejected by the usual p-value or the rp-value if the significance level is 0.1. If θ_0 is 3100, then the usual p-value is

$$P\left(|Z| > \frac{3100 - 2583.8}{\sqrt{729^2/5}}\right) = 0.113.$$

The rp-value is $(0.113 - 0.059)/(0.721 - 0.059) = 0.082$. The rp-value leads to rejection of the untrue H_0, but the usual p-value cannot reject H_0 when the significance level is 0.1. In this case, making a decision based on the rp-value is more appropriate.

We also conduct a simulation to compare both measures of evidence. The two measures of evidence used in the simulation are $p'_{\theta_0}(\bar{x})$ and $r'_{\theta_0}(\bar{x})$. Tables 2.1 and

Table 2.1
The ratios of the p-value and the rp-value less than 0.05 based on 1000 replicates. The true value of θ is 2944. The parameter space is (2700, 3200)

θ_0	H_0 true or false	n	p-value ≤ 0.05	rp-value ≤ 0.05
2700	false	10	0.173	0.516
2800	false	10	0.081	0.104
2944	true	10	0.049	0.01
3000	false	10	0.053	0.013
3100	false	10	0.07	0.089
3200	false	10	0.151	0.532

2.2 show the ratios of the p-values and the rp-values less than 0.05, in 1000 samples corresponding to some θ_0 and n when the lower bound and the upper bound are chosen to be 2700 and 3200, respectively.

Table 2.3 contains the simulation results when the lower bound and the upper bound are chosen to be 2600 and 3100, respectively.

From Tables 2.1–2.3, when θ_0 is equal to 2944, which means H_0 is true, the number of p-values less than 0.05 is greater than the number of rp-values less than 0.05. And when θ_0 is not equal to 2944, in most situations, the number of p-values less than 0.05 is less than the number of rp-values less than 0.05. From the simulation results, the performance of the rp-value, which sufficiently utilizes the bound information, is better than the usual p-value.

4. Testing Viewpoint

In this section, we will show that the test derived from the rp-value is the UMPU test under the criterion introduced below. Note that here the test derived from the rp-value is not a new test. It is shown that the test is exactly equal to the UMPU test under some conditions.

For any measure of evidence $e(X)$ used against the null hypothesis, where $e(X)$ denotes a function of $X = (X_1, \ldots, X_n)$, the null hypothesis is rejected if $e(X)$ is small. For example, the null hypothesis is rejected if the p-value is small. Thus a test

Table 2.2
The ratios of p-value and rp-value less than 0.05 based on 1000 replicates. The true value of θ is 2944. The parameter space is (2700, 3200)

θ_0	H_0 true or false	n	p-value ≤ 0.05	rp-value ≤ 0.05
2700	false	20	0.294	0.537
2800	false	20	0.135	0.152
2944	true	20	0.041	0.023
3000	false	20	0.062	0.045
3100	false	20	0.105	0.151
3200	false	20	0.301	0.557

Table 2.3
The ratios of the p-value and the rp-value less than 0.05 based on 1000 replicates. The true value of θ is 2944. The parameter space is (2600, 3100)

θ_0	H_0 true or false	n	p-value ≤ 0.05	rp-value ≤ 0.05
2600	false	30	0.72	0.834
2700	false	30	0.442	0.482
2800	false	30	0.155	0.157
2944	true	30	0.042	0.021
3000	false	30	0.042	0.044
3100	false	30	0.167	0.289

derived from $e(X)$ should take the form

$$\phi(X) = \begin{cases} 1 & \text{if } e(X) < k \\ 0 & \text{otherwise,} \end{cases} \tag{7}$$

where k is some constant between 0 and 1, and $\phi(X) = 1$ denotes rejecting the null hypothesis. If a test derived from a measure of evidence against null distribution has a good property, we say that the measure of evidence is good from a testing point of view.

Theorem 1. *Let $X_1, \ldots, X_n$ be a normal random variable $N(\theta, 1)$. Assume that the parameter space of θ is (a, b) and k is a constant satisfying*

$$k < \text{Min}\left(\frac{\int_{\frac{\theta_0 - a}{\sqrt{1/n}}}^{\frac{b-a}{\sqrt{1/n}}} e^{-\frac{t^2}{2}}\,dt}{\int_0^{\frac{b-a}{\sqrt{1/n}}} e^{-\frac{t^2}{2}}\,dt},\ \frac{\int_{\frac{b-\theta_0}{\sqrt{1/n}}}^{\frac{b-a}{\sqrt{1/n}}} e^{-\frac{t^2}{2}}\,dt}{\int_0^{\frac{b-a}{\sqrt{1/n}}} e^{-\frac{t^2}{2}}\,dt} \right). \tag{8}$$

For testing

$$H_0 : \theta = \theta_0 \quad vs. \quad H_1 : \theta \neq \theta_0,$$

the test derived from the rp-value $r_{\theta_0}(\overline{X})$ is

$$\phi(X) = \begin{cases} 1 & \text{if } r_{\theta_0}(\overline{X}) < k \\ 0 & \text{otherwise,} \end{cases}$$

where k satisfies $E_{\theta_0}\phi(X) = \alpha$. Then $\phi(X)$ is a level α UMPU test.

Proof. For any observation $\bar{x}$, at least one of the following three cases is true:

(i) $\bar{x} \leq a$, then

$$r_{\theta_0}(\bar{x}) = \frac{P\left(|Z| > \frac{\theta_0 - \bar{x}}{\sqrt{1/n}}\right) - P\left(|Z| > \frac{b-\bar{x}}{\sqrt{1/n}}\right)}{P\left(|Z| > \frac{a-\bar{x}}{\sqrt{1/n}}\right) - P\left(|Z| > \frac{b-\bar{x}}{\sqrt{1/n}}\right)};$$

(ii) $\bar{x} \geq b$, then

$$r_{\theta_0}(\bar{x}) = \frac{P\left(|Z| > \frac{\bar{x}-\theta_0}{\sqrt{1/n}}\right) - P\left(|Z| > \frac{\bar{x}-a}{\sqrt{1/n}}\right)}{P\left(|Z| > \frac{\bar{x}-b}{\sqrt{1/n}}\right) - P\left(|Z| > \frac{\bar{x}-a}{\sqrt{1/n}}\right)};$$

(iii) $a \leq \bar{x} \leq b$, then

$$r_{\theta_0}(\bar{x}) = \frac{P\left(|Z| > \frac{|\bar{x}-\theta_0|}{\sqrt{1/n}}\right) - \min\left[P\left(|Z| > \frac{b-\bar{x}}{\sqrt{1/n}}\right), P\left(|Z| > \frac{\bar{x}-a}{\sqrt{1/n}}\right)\right]}{1 - \min\left[P\left(|Z| > \frac{b-\bar{x}}{\sqrt{1/n}}\right), P\left(|Z| > \frac{\bar{x}-a}{\sqrt{1/n}}\right)\right]}.$$

Note that $r_{\theta_0}(\bar{x})$ is continuous at $\bar{x} = a$ and $\bar{x} = b$.

Now we will show that $\phi(X)$ is the UMPU test. If $\bar{x}$ belongs to the first case, by straightforward calculation, we have

$$r_{\theta_0}(\bar{x}) = \frac{\int_{(\theta_0-\bar{x})/\sqrt{1/n}}^{(b-\bar{x})/\sqrt{1/n}} e^{-\frac{t^2}{2}}\,dt}{\int_{(a-\bar{x})/\sqrt{1/n}}^{(b-\bar{x})/\sqrt{1/n}} e^{-\frac{t^2}{2}}\,dt} = \frac{\int_0^{(b-\theta_0)/\sqrt{1/n}} e^{-\frac{(t+(\theta_0-\bar{x})/\sqrt{1/n})^2}{2}}\,dt}{\int_{(a-\theta_0)/\sqrt{1/n}}^{(b-\theta_0)/\sqrt{1/n}} e^{-\frac{(t+(\theta_0-\bar{x})/\sqrt{1/n})^2}{2}}\,dt}. \tag{9}$$

Thus the rejection region $r_{\theta_0}(\bar{x}) < k$ is equivalent to

$$\frac{\int_0^{(b-\theta_0)/\sqrt{1/n}} e^{-\frac{(t+(\theta_0-\bar{x})/\sqrt{1/n})^2}{2}}\,dt}{\int_{(a-\theta_0)/\sqrt{1/n}}^{0} e^{-\frac{(t+(\theta_0-\bar{x})/\sqrt{1/n})^2}{2}}\,dt} < \frac{k}{1-k}. \tag{10}$$

Since the normal distribution has a monotone likelihood ratio,

$$\frac{e^{-\frac{(t_1+(\theta_0-\bar{x})/\sqrt{1/n})^2}{2}}}{e^{-\frac{(t_2+(\theta_0-\bar{x})/\sqrt{1/n})^2}{2}}}$$

is a nondecreasing function of $\bar{x} - \theta_0$ for any $t_1 > t_2$. Thus the set

$$\left\{x : \frac{e^{-\frac{(t_1+(\theta_0-\bar{x})/\sqrt{1/n})^2}{2}}}{e^{-\frac{(t_2+(\theta_0-\bar{x})/\sqrt{1/n})^2}{2}}} < \frac{k}{1-k}\right\}$$

is equivalent to the set $\{x : \bar{x} - \theta_0 < -c_1\}$, where c_1 is a constant depending on t_1 and t_2. Since any point in the set $\{t : t \in (0, (b-\theta_0)/\sqrt{1/n})\}$ is not less than any point in the set $\{t : t \in ((a-\theta_0)/\sqrt{1/n}, 0)\}$, (10) is equivalent to $\{x : \bar{x} - \theta_0 < -v\}$, where v is a positive constant.

If $\bar{x}$ belongs to the second case, by a similar argument to the above, $r_{\theta_0}(\bar{x}) < k$ is equivalent to

$$\frac{\int_{(a-\theta_0)/\sqrt{1/n}}^{0} e^{-\frac{(t+(\theta_0-\bar{x})/\sqrt{1/n})^2}{2}}\,dt}{\int_0^{(b-\theta_0)/\sqrt{1/n}} e^{-\frac{(t+(\theta_0-\bar{x})/\sqrt{1/n})^2}{2}}\,dt} < \frac{k}{1-k}. \tag{11}$$

Note that by straightforward calculation,

$$\left\{x : \frac{e^{-\frac{(t+(\theta_0-\bar{x})/\sqrt{1/n})^2}{2}}}{e^{-\frac{(t+(\theta_0-\bar{x})/\sqrt{1/n})^2}{2}}} > \frac{1-k}{k}\right\}$$

is equivalent to the set $\{x : \bar{x} - \theta_0 > c_1\}$. Thus (11) is equal to $\{x : \bar{x} - \theta_0 > v\}$.

If $\bar{x}$ belongs to the third case, then first consider the case of $a \leq \bar{x} \leq \theta_0$ and $\bar{x} - a < b - \bar{x}$. Then by a similar argument to the above, $r_{\theta_0}(\bar{x}) < k$ is equivalent to

$$\frac{\int_0^{(b-\theta_0)/\sqrt{1/n}} e^{-\frac{(t+(\theta_0-\bar{x})/\sqrt{1/n})^2}{2}} dt}{\int_{(\bar{x}-\theta_0)/\sqrt{1/n}}^0 e^{-\frac{(t+(\theta_0-\bar{x})/\sqrt{1/n})^2}{2}} dt} < \frac{k}{1-k}. \tag{12}$$

which is equivalent to $\{x : \bar{x} - \theta_0 < -v_1\}$ by the monotone likelihood ratio property, where v_1 is a positive constant.

If $\bar{x}$ belongs to the third case, $a \leq \bar{x} \leq \theta_0$ and $\bar{x} - a \geq b - \bar{x}$. Then by a similar argument to the above, $r_{\theta_0}(\bar{x}) < k$ is equivalent to

$$\frac{\int_0^{(2\bar{x}-a-\theta_0)/\sqrt{1/n}} e^{-\frac{(t+(\theta_0-\bar{x})/\sqrt{1/n})^2}{2}} dt}{\int_{(\bar{x}-\theta_0)/\sqrt{1/n}}^0 e^{-\frac{(t+(\theta_0-\bar{x})/\sqrt{1/n})^2}{2}} dt} < \frac{k}{1-k}. \tag{13}$$

By the monotone likelihood ratio property, (13) is equal to $\{x : \bar{x} - \theta_0 < -v_2\}$, where v_2 is a positive constant. Note that $2\bar{x} - a - \theta_0 > b - \theta_0$ in this case, otherwise $\bar{x} - a > b - \bar{x}$ does not hold. Comparing (12) and (13), we have $v_2 > v_1$.

By a similar argument, when $\theta_0 \leq \bar{x} \leq b$ and $\bar{x} - a > b - \bar{x}$, $r_{\theta_0}(\bar{x}) < k$ is equivalent to

$$\frac{\int_{(a-\theta_0)/\sqrt{1/n}}^0 e^{-\frac{(t+(\theta_0-\bar{x})/\sqrt{1/n})^2}{2}} dt}{\int_0^{(\bar{x}-\theta_0)/\sqrt{1/n}} e^{-\frac{(t+(\theta_0-\bar{x})/\sqrt{1/n})^2}{2}} dt} < \frac{k}{1-k}, \tag{14}$$

which is equal to $\{x : \bar{x} - \theta_0 > v_3\}$, where v_3 is a positive constant. When $\theta_0 \leq \bar{x} \leq b$ and $\bar{x} - a < b - \bar{x}$, $r_{\theta_0}(\bar{x}) < k$ is equivalent to

$$\frac{\int_{(2\bar{x}-b-\theta_0)/\sqrt{1/n}}^0 e^{-\frac{(t+(\theta_0-\bar{x})/\sqrt{1/n})^2}{2}} dt}{\int_0^{(\bar{x}-\theta_0)/\sqrt{1/n}} e^{-\frac{(t+(\theta_0-\bar{x})/\sqrt{1/n})^2}{2}} dt} < \frac{k}{1-k}, \tag{15}$$

which is equal to $\{x : \bar{x} - \theta_0 > v_4\}$, where v_4 is a positive constant. Note that $2\bar{x} - \theta_0 - b < a - \theta_0$ because of the condition $\bar{x} - a < b - \bar{x}$, so comparing the left-hand sides of (14) and (15), we have $v_4 > v_3$.

Moreover, comparing the left-hand sides of (10) and (12), we have $v_1 > v$; comparing the left-hand sides of (11) and (14), we have $v_3 > v$. Hence if v is greater than the maximum value of $b - \theta_0$ and $\theta_0 - a$, then the rejection region is empty when an observation belongs to the third case, which implies that the rejection region is $\{x : |\bar{x} - \theta_0| > v\}$. A necessary and sufficient condition for v greater than the maximum value of $b - \theta_0$ and $\theta_0 - a$ is that k is smaller than the minimum of

$r_{\theta_0}(a)$ and $r_{\theta_0}(b)$, because $r_{\theta_0}(a)$ and $r_{\theta_0}(b)$ need to belong to the accept region, which leads to k less than

$$\text{Min}\left(\frac{\int_{(\theta_0-a)/\sqrt{1/n}}^{(b-a)/\sqrt{1/n}} e^{-\frac{t^2}{2}}dt}{\int_0^{(b-a)/\sqrt{1/n}} e^{-\frac{t^2}{2}}dt}, \frac{\int_{(b-\theta_0)/\sqrt{1/n}}^{(b-a)/\sqrt{1/n}} e^{-\frac{t^2}{2}}dt}{\int_0^{(b-a)/\sqrt{1/n}} e^{-\frac{t^2}{2}}dt}\right). \qquad \square$$

Note that the noncentral t distribution also has a monotone likelihood ratio in the noncentrality parameter. Therefore, for the variance unknown case, we can follow a similar argument to that in Theorem 1 to reach Theorem 2.

Theorem 2. *Let the assumption be the same as in Theorem* 1 *except that* σ^2 *is unknown. Then the test derived from the rp-value* $r_1(\bar{x})$ *is also the UMPU test if* k *is less than*

$$\text{Min}\left(\frac{\int_{(\theta_0-a)/\sqrt{1/n}}^{(b-a)/\sqrt{1/n}} f(t)dt}{\int_0^{(b-a)/\sqrt{1/n}} f(t)dt}, \frac{\int_{(b-\theta_0)/\sqrt{1/n}}^{(b-a)/\sqrt{1/n}} f(t)dt}{\int_0^{(b-a)/\sqrt{1/n}} f(t)dt}\right),$$

where

$$f(t) = \frac{1}{\sqrt{\pi(n-1)}}\frac{\Gamma(n/2)}{\Gamma((n-1)/2)}\frac{1}{(1+t^2/(n-1))^{(n/2)}}. \tag{16}$$

Proof. (16) is the density function of a t distribution. By Lehmann (1986), the t distribution has a monotone likelihood ratio in the noncentrality parameter. The proof of Theorem 2 can be followed by a similar argument to that of the proof of Theorem 1. $\square$

Corollary 1. *In Theorems* 1 *and* 2, *if the bounds satisfy* $\theta_0 - a = b - \theta_0$, *the results hold directly for any* $0 < k < 1$.

Proof. In the case of $\theta_0 - a = b - \theta_0$, we have $v_1 = v_2$ in the proof of Theorem 1. The cases of (i) $a \le \bar{x} \le \theta_0$ and $\bar{x} - a > b - \bar{x}$, and (ii) $\theta_0 < \bar{x} < b$ and $\bar{x} - a < b - \bar{x}$, will not happen. This leads to the rejection region being $\{x : |\bar{x} - \theta_0| \ge v^*\}$, where v^* is a positive constant. $\square$

From the above result, the test based on the *rp*-value is the UMPU test. The test based on the usual *p*-value under criterion (7) is also the UMPU test. Therefore, from the criterion in this section, both the *rp*-value and the usual *p*-value are good measures of evidence from the testing point of view. For the same k, the significance levels of the two tests based on the *p*-value and the *rp*-value are different. The significance level of the test based on the *rp*-value cannot be directly calculated from the argument in the proof of Theorem 1. It can be derived from numerical calculation. Let $\bar{x} = \theta_0 + h$. Then the minimum of c satisfying $r_{\theta_0}(\theta_0 + h)$ less than k is the critical value of the test corresponding to k.

5. Estimation Criterion

Section 3 gives examples where the *rp*-value appears to perform better than the usual *p*-value. In this section, a more definitive analysis of its performance characteristics is provided and compared to the usual *p*-value. The criterion proposed in Hwang et al. (1992) can be used to evaluate the usual *p*-value and the *rp*-value. Hwang et al. (1992) and Robert (2001) pointed out that it is necessary to evaluate *p*-values under an adapted loss.

Let $I(\theta \in \Theta_0)$ denote the indicator function

$$I(\theta \in \Theta_0) = \begin{cases} 1 & \text{if } \theta \in \Theta_0 \\ 0 & \text{if otherwise,} \end{cases}$$

where Θ_0 is the parameter space of the null hypothesis. Hwang et al. (1992) suggested evaluating *p*-values under the squared error loss function

$$L(e(x), \theta) = (e(x) - I(\theta \in \Theta_0))^2. \tag{17}$$

$I(\theta \in \Theta_0)$ is used to measure the accuracy of the test. This loss function was first suggested by Schaarfsma et al. (1989). In Woodroofe and Wang (2000) and Wang (2004), the evaluation of evidence for testing a hypothesis is based on this loss function.

In the previous sections, we established some advantages of the *rp*-value $r_{\theta_0}(\bar{x})$. In this section, we apply the stricter criterion (17). Under this criterion, it will be shown that $\omega r_{\theta_0}(\bar{x})$ has a smaller mean squared error than the usual *p*-value for all θ belonging to the restricted parameter space, where ω is a positive constant belonging to an interval.

Theorem 3 gives the interval of ω such that $\omega r_\theta(x)$ has better performance than the usual *p*-value under criterion (17). In the numerical analysis, ω can be chosen as a constant near 1. Therefore we can say that the *rp*-value has good performance under this criterion.

Theorem 3. *Assume that $\overline{X}$ has $N(\theta, 1)$ distribution, and the parameter space of θ is (a, b). Under the loss function* (17),

(i) *For $\theta = \theta_0$, $\omega r_\theta(\bar{x})$ has a smaller mean squared error than the usual p-value when ω belongs to the interval*

$$\frac{E_{\theta_0} r_{\theta_0}(\bar{x}) \pm \sqrt{(E_{\theta_0} r_{\theta_0}(\bar{x}))^2 - E_{\theta_0} r^2_{\theta_0}(\overline{X})(1 - E_{\theta_0}(p_{\theta_0}(\overline{X}) - 1)^2)}}{E_{\theta_0} r^2_{\theta_0}(\overline{X})}. \tag{18}$$

(ii) *For $\theta \neq \theta_0$, $\omega r_\theta(\bar{x})$ has a smaller mean squared error than the usual p-value when ω belongs to the interval*

$$\left(0, \min_{\theta \in (a,b), \theta \neq \theta_0} \left(\frac{E_\theta p^2_{\theta_0}(\overline{X})}{E_\theta r^2_{\theta_0}(\overline{X})}\right)^{1/2}\right). \tag{19}$$

Proof. (i) When θ is equal to θ_0, the indicator function $I(\theta = \theta_0)$ is 1. If ω satisfies

$$E_{\theta_0}(p_{\theta_0}(\overline{X}) - 1)^2 - E_{\theta_0}(\omega r_{\theta_0}(\bar{x}) - 1)^2 > 0, \tag{20}$$

Table 5.1

Assume that X has a $N(0, 1)$ distribution and the parameter space of θ is $(-2, 2)$. For testing $H_0 : \theta_0 = 0$ versus $H_1 : \theta_0 \neq 0$, the mean squared errors E_θ $(p\ value$ $-I(\theta = 0))^2$ of the usual p-value and c-value with $\omega = 1.0049$ corresponding to θ in the parameter space are listed. The results are based on 5000 replications

θ_0	MSE of p-value	MSE of rp-value
0	0.33333	0.33332
0.1	0.33199	0.33173
0.3	0.32144	0.32119
0.6	0.28829	0.28805
1	0.22288	0.22269
1.5	0.13519	0.13508
2	0.06752	0.06750

$\omega r_{\theta_0}(\bar{x})$ is better than the usual p-value. By straightforward calculation, (20) is equivalent to

$$\omega^2 E_{\theta_0} r^2_{\theta_0}(\overline{X}) - 2\omega E_{\theta_0} r_{\theta_0}(\bar{x}) + 1 - E_{\theta_0}(p_{\theta_0}(\overline{X}) - 1)^2 < 0.$$

Hence if ω satisfies condition (18), (20) holds directly.

(ii) When $\theta \neq \theta_0$, the indicator function $I(\theta = \theta_0)$ is 0. If ω satisfies

$$E_\theta p^2_{\theta_0}(\overline{X}) - \omega^2 E_\theta r^2_{\theta_0}(\overline{X}) > 0 \quad \text{for } \theta \neq \theta_0, \tag{21}$$

then $r_{\theta_0}(\bar{x})$ is better than the usual p-value. By a straightforward calculation, the set of ω satisfying (21) is equivalent to (19), where ω is a positive number. □

We can also have a similar result for the variance unknown case if $r_{\theta_0}(\bar{x})$ and $p_{\theta_0}(\overline{X})$ in Theorem 3 are replaced by $r'_{\theta_0}(\overline{X})$ and $p'_{\theta_0}(\overline{X})$.

Table 5.2

The assumption is the same as in Table 5.1 except that the parameter space is $(-1.5, 1.5)$ and $\omega = 1.015$. The results are also based on 5000 replications

θ_0	MSE of p-value	MSE of rp-value
0	0.33333	0.33209
0.3	0.32144	0.30932
0.6	0.28829	0.27747
0.9	0.24056	0.23163
1.2	0.18684	0.18010

Table 5.3
The assumption is the same as in Table 5.1 except that the parameter space is $(-1, 1)$ and $\omega = 1.03$. The results are also based on 5000 replications

θ_0	MSE of the usual p-value	MSE of the modified p-value
0	0.33333	0.32154
0.3	0.32144	0.28807
0.6	0.28829	0.25947
0.9	0.24056	0.21833

Theorem 4. *Assume that* $\overline{X}$ *has* $N(\theta, \sigma^2)$ *distribution, where* σ^2 *is unknown, and the parameter space of* θ *is* (a, b). *Then the results in Theorem* 1 *hold when* $r_{\theta_0}(\bar{x})$ *and* $p_{\theta_0}(\overline{X})$ *are replaced by* $r'_{\theta_0}(\overline{X})$ *and* $p'_{\theta_0}(\overline{X})$.

Corollary 2. *From Theorem* 3, *if there exists an* ω *such that both conditions* (i) *and* (ii) *in Theorem* 3 *are satisfied, then the* $\omega r_{\theta_0}(\bar{x})$ *is better than the usual p-value for all* θ *in the parameter space* (a, b).

From Theorems 3 and 4 and Corollary 2, we do numerical analysis for the interval of ω as follows. Note that the endpoints of (18) and (19) can be calculated by software such as Mathematica. Here we discuss cases when $\theta_0 - a = b - \theta_0$. For general cases, the lower bound a can be chosen to be smaller and the upper bound b can be chosen to be larger, so that $\theta_0 - a = b - \theta$. Let $m = \theta_0 - a = b - \theta_0$. We found that the minimal value of (17) always happens at b. In Theorem 3, if there exists an ω such that (18) and (19) hold simultaneously, $\omega r_{\theta_0}(\bar{x})$ dominates $p_{\theta_0}(\overline{X})$ for estimating $I(\theta = \theta_0)$ under the squared error loss function (17). Consider the case of $m = 1$ in Theorem 3; then the two intervals (18) and (19) of ω are (0.9974, 2.2406) and (0, 1.048). Thus, if ω belongs to (0.9974, 1.048), $\omega r_{\theta_0}(\bar{x})$ is better than the usual p-value for all θ from the estimation point of view. If $m = 2$, the two intervals of ω are (1.0049, 2.0114) and (0, 1.00501); hence $r_{\theta_0}(\bar{x})$ is better than the usual p-value for $\omega \in (1.0049, 1.0051)$. For the case of σ^2 unknown, by Theorem 4, ω can be chosen by a similar argument to that in the case of σ^2 known. Tables 5.1–5.3 show the mean squared errors of the usual p-value and the rp-value for the cases of $m = 1, 1.5$, and 2. As can be seen from the tables, the simulation results are consistent with the theoretical results that the rp-value is better than the usual p-value for all θ in the parameter space.

References

Berger, J. O., Wolpert, R. W. (1984). *The Likelihood Principle*. 2nd ed. Hayward, Calif: IMS.

Berger, J. O., Sellke, T. (1987). Testing a point null hypothesis: the irreconcilability of p values and evidence (with discussion). *J. Amer. Statist. Assoc.* 82:112–139.

Feldman, G. J., Cousins, R. D. (1998). Unified approach to the classical statistical analysis of small signals. *Phys. Rev. D* 57:3873–3889.

Hwang, J. T., Casella, G., Robert, C., Wells, M. T., Farrell, R. H. (1992). Estimation of accuracy in testing. *Ann. Statist.* 20:490–509.

Lehmann, E. L. (1983). *Theory of Point Estimation*. New York: Springer-Verlag.

Lehmann, E. L. (1986). *Testing Statistical Hypotheses*. New York: John Wiley.
Lindley, D. V. (1957). A statistical paradox. *Biometrika* 44:187–192.
Mandelkern, M. (2002). Setting confidence intervals for bounded parameters. *Statistical Science* 17:149–172.
Robert, C. P. (2001). *The Bayesian Choice*. Springer.
Roe, B. P., Woodroofe, M. B. (2001). Setting confidence belts. *Phys. Rev. D* 60:3009–3015.
Schaarfsma, W., Tobloom, J., Van der Menlen, B. (1989). Discussing truth or falsity by computing a q-value. In: Dodge, Y., ed. *Statistical Data Analysis and Inference*. Amsterdam: North-Holland, pp. 85–100.
Wang, H. (2004). Improved estimation of accuracy in simple hypothesis versus simple alternative hypothesis. *J. Multivariate Anal.* 90:269–281.
Wang, H. (2005). Modified p-value for one-sided testing in restricted parameter space. Technical report.
Woodroofe, M., Wang, H. (2000). The problem of low counts in a single plus noise model. *Ann. Statist.* 28:1561–1569.