Research Article

Testing Normal Means: The Reconcilability of the P Value and the Bayesian Evidence

Yuliang Yin[1] and Junlong Zhao[2]

[1] *School of Economics, Beijing Technology and Business University, Beijing 100048, China*
[2] *School of Mathematics and System Science, Beihang University, LMIB of the Ministry of Education, Beijing 100083, China*

Correspondence should be addressed to Yuliang Yin; imyyl@163.com

The problem of reconciling the frequentist and Bayesian evidence in testing statistical hypotheses has been extensively studied in the literature. Most of the existing work considers cases without the nuisance parameters which is not the frequently encountered situation since the presence of the nuisance parameters is very common in practice. In this paper, we consider the reconcilability of the Bayesian evidence against the null hypothesis H_0 in terms of the posterior probability of H_0 being true and the frequentist evidence against H_0 in terms of the P value in testing normal means where the nuisance parameters are present. The reconcilability of evidence can be obtained both for testing a normal mean and for the Behrens-Fisher problem.

1. Introduction

In the problem of testing a statistical hypothesis H_0, a frequentist may give evidence against H_0 by the observed significance level, the P value, while a Bayesian may give it by the posterior probability that H_0 is true. Lindley [1] illustrated the possible discrepancy between the Bayesian and the frequentist evidence. The relationship of these two measures of evidence is then extensively studied in the literature. Pratt [2] revealed that the P values are usually approximately equal to the posterior probabilities in the one-sided testing problems. Casella and Berger [3] considered testing the one-sided hypothesis for a location parameter and showed that the lower bounds of the posterior probability over some reasonable classes of priors are exactly equal to the corresponding P values in many cases. Some important papers which deal with the reconcilability of the Bayesian and frequentist evidence are Bartlett [4], Cox [5], Shafer [6], Berger and Delampady [7], and Berger and Sellke [8].

Although many researches have been carried out to deal with the problem of reconciling the Bayesian and frequentist evidence and some of them show that evidence is reconcilable in several specific situations, most of the existing work assumes that no other unknown parameters are present except the parameters of interest. In fact, we may be confronted with the nuisance parameters in various situations. In the location-scale settings, for example, when the location parameter is unknown, so is the scale parameter, in general.

However, in significance testing of hypotheses with the nuisance parameters, the classical P values are typically not available. Tsui and Weerahandi [9], considering testing the one-sided hypothesis of the form

$$H_0 : \theta \le c \quad \text{versus} \quad H_1 : \theta > c, \tag{1}$$

where θ is the parameter of interest and c is a fixed constant, introduced the concept of the generalized P value, which appears to be useful in situations where conventional frequentist approaches do not provide useful solutions.

Tsui and Weerahandi [9] and some later relevant works formulated the generalized P values for many specific examples. Hannig et al. [10] provided a general method for constructing the generalized P value via fiducial inference.

In this paper, for the one-sided testing situations about normal means where the nuisance parameters are present, we study the reconcilability of the Bayesian evidence and

the generalized P value. It is shown that, under the conjugate class of prior distributions, the Bayesian evidence and the generalized P value are reconcilable both for the problem of testing a normal mean and for the Behrens-Fisher problem.

This paper is organized as follows. In Section 2, we give the main results of the reconcilability of the P value and the Bayesian evidence in testing normal means. Some conclusions and discussions are given in Section 3.

2. Main Results

In this section, we consider two testing problems in which the nuisance parameters are present. When no efficient classical frequentist evidence is available because of the presence of the nuisance parameters, we formulate the frequentist evidence by the generalized P value.

2.1. One-Sample Normal Mean. Let $X_1, \ldots, X_n$ be a random sample from a normal population $N(\mu, \sigma^2)$, where both the mean μ and the variance σ^2 are unknown. Consider now the following problem of testing the mean of a normal distribution

$$H_0 : \mu \leq c \quad \text{versus} \quad H_1 : \mu > c, \tag{2}$$

where c is a fixed constant.

For this testing problem, where the nuisance parameter is present, we can still obtain the classical P value as

$$p(x) = P\left(T_{n-1} \leq \frac{\sqrt{n}(c - \overline{x})}{s}\right), \tag{3}$$

where T_{n-1} is a t-variable with $n - 1$ degrees of freedom and $\overline{x}$ and s^2 stand for the observed sample mean and sample variance, respectively.

To derive the Bayesian evidence, we need a prior for the parameters. One reasonable and conventional class of priors for μ and σ^2 is the following conjugate class of prior distributions G_{c_1}:

$$\mu \mid \sigma^2 \sim N\left(\mu_0, \frac{\sigma^2}{\kappa_0}\right), \qquad \frac{1}{\sigma^2} \sim \text{Gamma}\left(\frac{\nu_0}{2}, \frac{\nu_0\sigma_0^2}{2}\right), \tag{4}$$

where the prior parameters (μ_0, κ_0) can be interpreted as the mean and sample size of the normal prior observations and (σ_0^2, ν_0) the sample variance and sample size of the Gamma prior observations.

Under (4) we have

$$\mu \mid x, \sigma^2 \sim N\left(\mu_n(x), \frac{\sigma^2}{\kappa_n}\right),$$

$$\frac{1}{\sigma^2} \mid x \sim \text{Gamma}\left(\frac{\nu_n}{2}, \frac{\nu_n\sigma_n^2(x)}{2}\right), \tag{5}$$

where $x = (x_1, \ldots, x_n)'$, $\kappa_n = \kappa_0 + n$, $\mu_n(x) = (\kappa_0\mu_0 + n\overline{x})/\kappa_n$, $\nu_n = \nu_0 + n$, and $\sigma_n^2(x) = [\nu_0\sigma_0^2 + (n-1)s^2 + \kappa_0 n(\overline{x} - \mu_0)^2/\kappa_n]/\nu_n$. Therefore, we can give the posterior density for (μ, σ^2) as

$$\begin{aligned}
\pi\left(\mu, \sigma^2 \mid x\right) &= \frac{\sqrt{\kappa_n}}{\sqrt{2\pi}\sigma} \exp\left[-\frac{\kappa_n(\mu - \mu_n(x))^2}{2}\right] \\
&\quad \times \frac{\left((\nu_n/2)\sigma_n^2(x)\right)^{\nu_n/2}}{\Gamma(\nu_n/2)\sigma^{\nu_n+2}} \exp\left[-\frac{\nu_n\sigma_n^2(x)}{2\sigma^2}\right] \\
&= \frac{\sqrt{\kappa_n}\left((\nu_n/2)\sigma_n^2(x)\right)^{\nu_n/2}}{\sqrt{2\pi}\Gamma(\nu_n/2)\sigma^{\nu_n+3}} \\
&\quad \times \exp\left[-\frac{\kappa_n(\mu - \mu_n(x))^2 + \nu_n\sigma_n^2(x)}{2\sigma^2}\right].
\end{aligned} \tag{6}$$

Then the marginal posterior density for μ can be obtained by integrating out σ^2 as

$$\begin{aligned}
\pi\left(\mu \mid x\right) &= \frac{\sqrt{\kappa_n}\left((\nu_n/2)\sigma_n^2(x)\right)^{-1/2}\Gamma((\nu_n+1)/2)}{\Gamma(\nu_n/2)} \\
&\quad \times \left[1 + \frac{\kappa_n(\mu - \mu_n(x))^2}{\nu_n\sigma_n^2(x)}\right]^{-(\nu_n+1)/2},
\end{aligned} \tag{7}$$

from which we know that

$$\frac{\sqrt{\kappa_n}(\mu - \mu_n(x))}{\sigma_n(x)} \sim t(\nu_n). \tag{8}$$

Consequently, the posterior probability of H_0 being true is

$$P\left(H_0 \mid x\right) = P\left(T_{\nu_n} \leq \sqrt{\frac{\kappa_n}{\sigma_n^2(x)}}\left(c - \mu_n(x)\right)\right), \tag{9}$$

where T_{ν_n} is a t-variable with ν_n degrees of freedom. Notice that if $\mu_0 = c$, we have

$$\lim_{\kappa_0, \sigma_0 \to 0} P\left(H_0 \mid x\right) = P\left(T_{\nu_n} \leq \sqrt{\frac{\nu_n}{n-1}}\frac{\sqrt{n}(c - \overline{x})}{s}\right). \tag{10}$$

Lemma 1. *Let T be a t-variable with α degrees of freedom, where α is a positive real number. Then for $R = T/\sqrt{\alpha}$ and a fixed constant r, $P(R \leq r)$ is nonincreasing in α if $r \leq 0$ and is nondecreasing in α if $r > 0$.*

Proof. Suppose that X is a nonpositive random variable obtained by the negative part of T; that is, the density of X is

$$f(x, \alpha) = \frac{2\Gamma((\alpha+1)/2)}{\sqrt{\alpha\pi}\Gamma(\alpha/2)}\left(1 + \frac{x^2}{\alpha}\right)^{-(\alpha+1)/2}, \quad x \leq 0. \tag{11}$$

Then the density of $Y = X/\sqrt{\alpha}$ is

$$p(y, \alpha) = \frac{2\Gamma((\alpha+1)/2)}{\sqrt{\pi}\Gamma(\alpha/2)}\left(1 + y^2\right)^{-(\alpha+1)/2}, \quad y \le 0. \quad (12)$$

By Theorem 3.3.2 in Lehmann [11], for any fixed nonpositive constant r, we have that $P(Y \le r)$ is nonincreasing in α since it can be verified that the family of densities $p(y, \alpha)$ has monotone likelihood ratio in y. This implies that Lemma 1 holds for the case when $r \le 0$ since $P(R \le r) = P(X \le r)/2$. Since when $r > 0$, we have $P(R \le r) = 1/2 + P(0 \le R \le r)$, the proof for the latter case is completely analogous if we introduce a nonnegative random variable obtained by the positive part of T. □

Now take $r = \sqrt{n}(c-\overline{x})/(\sqrt{n-1}s)$. By Lemma 1, for $r \le 0$, we have

$$P\left(\frac{T_{\nu_n}}{\sqrt{\nu_n}} \le r\right) \le P\left(\frac{T_{n-1}}{\sqrt{n-1}} \le r\right)$$

$$= P\left(T_{n-1} \le \frac{\sqrt{n}(c-\overline{x})}{s}\right). \quad (13)$$

Then comparing (3) and (10), for $\mu_0 = c$ and any fixed nonnegative ν_0, we have

$$\lim_{\kappa_0, \sigma_0 \to 0} P(H_0 \mid x) < p(x), \quad \text{as } p(x) < \frac{1}{2}, \quad (14)$$

which implies that

$$\inf_{\pi \in G_{c_1}} P(H_0 \mid x) < p(x), \quad \text{as } p(x) < \frac{1}{2}. \quad (15)$$

The reconcilability of the Bayesian and frequentist evidence is therefore obtained in this testing problem. We summarize this as the following theorem.

Theorem 2. *For testing the hypothesis of the form* (2) *under a normal distribution* $N(\mu, \sigma^2)$ *with* σ^2 *unknown, the Bayesian and frequentist lines of evidence are reconcilable under the conjugate class of priors* (4).

2.2. Behrens-Fisher Problem. Now we turn to consider the Behrens-Fisher problem. It is a classical testing situation in which the nuisance parameters are present and no useful pivotal quantities are available. Suppose that $X_1, \ldots, X_m$ and $Y_1, \ldots, Y_n$ are two independent random samples from two normal populations $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively, where both σ_1^2 and σ_2^2 are completely unspecified. We are interested in testing the hypothesis of the form

$$H_0 : \mu_1 - \mu_2 \le c \quad \text{versus} \quad H_1 : \mu_1 - \mu_2 > c, \quad (16)$$

where c is a fixed constant.

In situations where the traditional frequentist approaches fail to provide useful solutions, the conception of the generalized P values introduced by Tsui and Weerahandi [9] appears to be helpful in deriving the frequentist evidence for testing

a statistical hypothesis. For this specific problem of testing hypothesis (16), we can give the generalized P value as

$$p(x) = P\left(\overline{x} - \overline{y} - \left(\frac{\sqrt{m-1}s_1 U}{\sqrt{m}\sqrt{\chi_{m-1}^2}} - \frac{\sqrt{n-1}s_2 V}{\sqrt{n}\sqrt{\chi_{n-1}^2}}\right) \le c\right)$$

$$= P\left(\frac{s_2}{\sqrt{n}}T_{n-1} - \frac{s_1}{\sqrt{m}}T_{m-1} \le c - (\overline{x} - \overline{y})\right), \quad (17)$$

where $U \sim N(0, 1)$, $V \sim N(0, 1)$, $\chi_{m-1}^2 \sim \chi^2(m-1)$, $\chi_{n-1}^2 \sim \chi^2(n-1)$, $T_{m-1} \sim t(m-1)$, and $T_{n-1} \sim t(n-1)$.

In this problem, we consider the reconcilability of evidence under the following conjugate class of prior distributions G_{c_2}:

$$\mu_1 \mid \sigma_1^2 \sim N\left(\mu_{01}, \frac{\sigma_{01}^2}{\kappa_{01}}\right), \qquad \frac{1}{\sigma_1^2} \sim \text{Gamma}\left(\frac{\nu_{01}}{2}, \frac{\nu_{01}\sigma_{01}^2}{2}\right);$$

$$\mu_2 \mid \sigma_2^2 \sim N\left(\mu_{02}, \frac{\sigma_{02}^2}{\kappa_{02}}\right), \qquad \frac{1}{\sigma_2^2} \sim \text{Gamma}\left(\frac{\nu_{02}}{2}, \frac{\nu_{02}\sigma_{02}^2}{2}\right). \quad (18)$$

Under G_{c_2}, the posterior density of $(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$ is

$$\pi\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2 \mid x, y\right)$$

$$= \frac{f\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, x, y\right)}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_0^{\infty}\int_0^{\infty} f\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, x, y\right) d\sigma_1^2 d\sigma_2^2 d\mu_1 d\mu_2}, \quad (19)$$

where

$$f\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, x, y\right)$$

$$= \left(\exp\left[-\frac{\Sigma_{i=1}^m(x_i - \mu_1)^2 + \kappa_{01}(\mu_1 - \mu_{01})^2 + \nu_{01}\sigma_{01}^2}{2\sigma_1^2}\right.\right.$$

$$\left.\left. - \frac{\Sigma_{i=1}^n(y_i - \mu_2)^2 + \kappa_{02}(\mu_2 - \mu_{02})^2 + \nu_{02}\sigma_{02}^2}{2\sigma_2^2}\right]\right)$$

$$\times \left((\sigma_1^2)^{(m+\nu_{01}+4)/2}(\sigma_2^2)^{(n+\nu_{02}+4)/2}\right)^{-1}. \quad (20)$$

Let $\theta = \mu_1 - \mu_2$. Then the posterior density of $(\theta, \mu_2, \sigma_1^2, \sigma_2^2)$ is

$$\pi\left(\theta, \mu_2, \sigma_1^2, \sigma_2^2 \mid x, y\right)$$

$$= \frac{f\left(\theta, \mu_2, \sigma_1^2, \sigma_2^2, x, y\right)}{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_0^{\infty}\int_0^{\infty} f\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, x, y\right) d\sigma_1^2 d\sigma_2^2 d\mu_1 d\mu_2}, \quad (21)$$

where

$$
\begin{aligned}
& f\left(\theta, \mu_2, \sigma_1^2, \sigma_2^2, x, y\right) \\
& = \left(\exp\left[-\frac{\sum_{i=1}^m (x_i - \theta - \mu_2)^2 + \kappa_{01}(\theta + \mu_2 - \mu_{01})^2 + \nu_{01}\sigma_{01}^2}{2\sigma_1^2}\right.\right. \\
& \left.\left. -\frac{\sum_{i=1}^n (y_i - \mu_2)^2 + \kappa_{02}(\mu_2 - \mu_{02})^2 + \nu_{02}\sigma_{02}^2}{2\sigma_2^2}\right]\right) \\
& \times \left((\sigma_1^2)^{(m+\nu_{01}+4)/2}(\sigma_2^2)^{(n+\nu_{02}+4)/2}\right)^{-1}.
\end{aligned}
\tag{22}
$$

So that the posterior probability of H_0 is

$$
\begin{aligned}
& P\left(H_0 \mid x\right) \\
& = \frac{\int_{-\infty}^c \int_{-\infty}^\infty \int_0^\infty \int_0^\infty f\left(\theta, \mu_2, \sigma_1^2, \sigma_2^2, x, y\right) d\sigma_1^2 d\sigma_2^2 d\mu_2 d\theta}{\int_{-\infty}^\infty \int_{-\infty}^\infty \int_0^\infty \int_0^\infty f\left(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, x, y\right) d\sigma_1^2 d\sigma_2^2 d\mu_1 d\mu_2}.
\end{aligned}
\tag{23}
$$

It is straightforward to check that

$$
\begin{aligned}
& \lim_{\kappa_{01},\kappa_{02},\sigma_{01},\sigma_{02}\to 0} P\left(H_0 \mid x\right) \\
& = \frac{\Gamma\left((m+\nu_{01}+2)/2\right)\Gamma\left((n+\nu_{02}+2)/2\right)}{\pi\sqrt{(m-1)(n-1)}\Gamma\left((m+\nu_{01}+1)/2\right)\Gamma\left((n+\nu_{02}+1)/2\right)} \\
& \times \int_{-\infty}^\infty \int_{\sqrt{m}(\bar{x}-\bar{y}-c)/s_1+\sqrt{m}s_2 r/s_1}^\infty \left(1+\frac{t^2}{m-1}\right)^{(-m-\nu_{01}-2)/2} \\
& \times \left(1+\frac{r^2}{n-1}\right)^{(-n-\nu_{02}-2)/2} dt\, dr, \\
& = P\left(\bar{x}-\bar{y}-\left(\frac{\sqrt{m+\nu_{01}+1}s_1 U}{\sqrt{m+\nu_{02}+2}\sqrt{\chi_{m+\nu_{01}+1}^2}}\right.\right. \\
& \left.\left. -\frac{\sqrt{n+\nu_{02}+1}s_2 V}{\sqrt{n+\nu_{02}+2}\sqrt{\chi_{m+\nu_{01}+1}^2}}\right) \le c\right) \\
& = P\left(\frac{s_2}{\sqrt{n+\nu_{02}+2}}T_{n+\nu_{02}+1} - \frac{s_1}{\sqrt{m+\nu_{01}+2}}T_{m+\nu_{01}+1}\right. \\
& \left. \le c-(\bar{x}-\bar{y})\right),
\end{aligned}
\tag{24}
$$

where $\bar{x}$ and $\bar{y}$ are the observation of the sample mean $\bar{X}$ and $\bar{Y}$, respectively, s_1^2 and s_2^2 are that of the sample variance S_1^2 and S_2^2 respectively, $U \sim N(0,1)$, $V \sim N(0,1)$, $\chi_{m+\nu_{01}+1}^2 \sim \chi^2(m+\nu_{01}+1)$ and $\chi_{n+\nu_{02}+1}^2 \sim \chi^2(n+\nu_{02}+1)$.

Now we prove an interesting result that, when m and n are sufficiently large, the frequentist and Bayesian lines of evidence given respectively by (17) and (23) are reconcilable for any fixed ν_{01} and ν_{02} under the prior class of G_{c_2}.

Theorem 3. *As $\min\{m, n\} \to +\infty$, for any fixed $0 < \nu_{01}, \nu_{02} < \infty$, we have*

$$
\lim_{\kappa_{01},\kappa_{02},\sigma_{01},\sigma_{02}\to 0} P\left(H_0 \mid x\right) < p(x), \quad \text{as } p(x) < \frac{1}{2}, \tag{25}
$$

which implies that

$$
\inf_{\pi \in G_{c_2}} P\left(H_0 \mid x\right) < p(x), \quad \text{as } p(x) < \frac{1}{2}. \tag{26}
$$

Proof. Let

$$
\begin{aligned}
A_1 &= \frac{(n-1)s_2^2}{n\chi_{n-1}^2}, \\
B_1 &= \frac{(n+\nu_{02}+1)s_2^2}{(n+\nu_{02}+2)\chi_{n+\nu_{02}+1}^2}.
\end{aligned}
\tag{27}
$$

(I) We first prove that, given s_2, as n is sufficiently large, we have

$$
B_1 \overset{d}{<} A_1. \tag{28}
$$

In fact, for any $\gamma > 0$, as n is sufficiently large,

$$
\begin{aligned}
P\left(B_1 < \gamma\right) &= P\left(\chi_{n+\nu_{02}+1}^2 > \frac{(n+\nu_{02}+1)s_2^2}{(n+\nu_{02}+2)\gamma}\right), \\
&= P\left(\chi_{n+\nu_{02}+1}^2 > \frac{(n-1)s_2^2}{n\gamma} + \epsilon(n,\gamma)\right),
\end{aligned}
\tag{29}
$$

where $\epsilon(n,\gamma) = (\nu_{02}+2)s_2^2/[(n+\nu_{02}+2)n\gamma] \to 0$, as $n \to \infty$. On the other hand, we have

$$
P\left(A_1 < \gamma\right) = P\left(\chi_{n-1}^2 > \frac{(n-1)s_2^2}{n\gamma}\right). \tag{30}
$$

Since $\chi_{n+\nu_{02}+1}^2 \overset{d}{>} \chi_{n-1}^2$ holds for any n, it follows that, as $n \to \infty$,

$$
P\left(B_1 < \gamma\right) > P\left(A_1 < \gamma\right). \tag{31}
$$

That is,

$$
B_1 \overset{d}{<} A_1. \tag{32}
$$

Similarly, as $m \to +\infty$, we have

$$
B_2 := \frac{(m+\nu_{01}+1)s_1^2}{(m+\nu_{01}+2)\chi_{m+\nu_{01}+1}^2} \overset{d}{<} \frac{(m-1)s_1^2}{m\chi_{m-1}^2} := A_2. \tag{33}
$$

Therefore, we have

$$
B_1 + B_2 \overset{d}{<} A_1 + A_2. \tag{34}
$$

Consequently, we have

$$
\sqrt{B_1 + B_2} \overset{d}{<} \sqrt{A_1 + A_2}. \tag{35}
$$

(II) We now show that the final conclusion holds. In fact, if we let $C(x, y) = c - (\overline{x} - \overline{y})$, then

$$
\begin{aligned}
p(x) &= P\left(\sqrt{A_1}V - \sqrt{A_2}U \le C(x, y)\right) \\
&= E\left(P\left(\sqrt{A_1}V - \sqrt{A_2}U \le C(x, y) \mid A_1, A_2\right)\right) \\
&= E\left(\Phi\left(\frac{C(x, y)}{\sqrt{A_1 + A_2}}\right)\right),
\end{aligned}
\tag{36}
$$

where $\Phi(\cdot)$ stands for the cumulative distribution function of a standard normal distribution and the last equation is due to the fact that U and V are independent normal distributions.

Similarly, for (24), we have

$$
\begin{aligned}
\lim_{\kappa_{01}, \kappa_{02}, \sigma_{01}, \sigma_{02} \to 0} &P(H_0 \mid x, y) \\
&= E\left(P\left(\sqrt{B_1}V - \sqrt{B_2}U \le C(x, y) \mid B_1, B_2\right)\right) \\
&= E\left(\Phi\left(\frac{C(x, y)}{\sqrt{B_1 + B_2}}\right)\right).
\end{aligned}
\tag{37}
$$

Note that for each $C(x, y)$ in $(-\infty, 0)$, $\Phi(C(x, y)/R)$ is increasing in $R \in (0, \infty)$. Therefore, by (35), we have

$$
\Phi\left(\frac{C(x, y)}{\sqrt{B_1 + B_2}}\right) \overset{d}{<} \Phi\left(\frac{C(x, y)}{\sqrt{A_1 + A_2}}\right), \quad \text{as } C(x, y) < 0.
\tag{38}
$$

Consequently, we have

$$
E\left(\Phi\left(\frac{C(x, y)}{\sqrt{B_1 + B_2}}\right)\right) < E\left(\Phi\left(\frac{C(x, y)}{\sqrt{A_1 + A_2}}\right)\right),
\tag{39}
$$
$$
\text{as } C(x, y) < 0.
$$

In addition, by the symmetry of the t-distribution, it follows that $C(x, y) < 0$ is equivalent to $p(x) < 1/2$. Therefore, by (36), (37), and (39), the conclusion of Theorem 3 holds. $\qquad\square$

The following theorem shows that, even for fixed m and n with $2 < m, n < \infty$, we still obtain the reconcilability of the frequentist and Bayesian evidence.

Theorem 4. *As $\min\{\nu_{01}, \nu_{02}\} \to \infty$, the conclusion of Theorem 3 holds for any fixed m and n with $2 < m, n < \infty$; that is,*

$$
\inf_{\pi \in G_{c_2}} P(H_0 \mid x) < p(x), \quad \text{as } p(x) < \frac{1}{2}.
\tag{40}
$$

Proof. We still adopt the notations of Theorem 3. We first prove that, as $\nu_{02} \to \infty$,

$$
B_1 \overset{d}{<} A_1.
\tag{41}
$$

By the proof of Theorem 3, we have

$$
P(B_1 < \gamma) = P\left(\chi^2_{n+\nu_{02}+1} > \frac{(n-1)s_2^2}{n\gamma} + \epsilon(n, \gamma)\right),
\tag{42}
$$

where

$$
\epsilon(n, \gamma) = \frac{(\nu_{02} + 2)s_2^2}{(n + \nu_{02} + 2)n\gamma}.
\tag{43}
$$

It is obvious that

$$
\sup_{\nu_{02}} \epsilon(n, \gamma) = \lim_{\nu_{02} \to \infty} \epsilon(n, \gamma) = \frac{s_2^2}{n\gamma}.
\tag{44}
$$

Let $f_n(t)$ denote the density function of χ^2_n. Then it is easy to see that $f_n(t)$ reaches the maximum at $t = n - 2$ and that $\max_t f_n(t) = f_n(n - 2) \to 0$, as $n \to \infty$. Therefore, as ν_{02} is sufficiently large, it holds that

$$
\begin{aligned}
P(B_1 < \gamma) &= P\left(\chi^2_{n+\nu_{02}+1} > \frac{s_2^2}{\gamma}\right) + \epsilon(\nu_{02}) \\
&= P(\widetilde{B}_1 < \gamma) + \epsilon(\nu_{02})
\end{aligned}
\tag{45}
$$

for some $\epsilon(\nu_{02}) > 0$, where $\epsilon(\nu_{02}) \to 0$, as $\nu_{02} \to \infty$, and $\widetilde{B}_1 = s_2^2/\chi^2_{n+\nu_{02}+1}$.

Therefore, for any fixed n $(2 < n < \infty)$, as $\nu_{02} \to \infty$, we have

$$
\begin{aligned}
P(\widetilde{B}_1 < \gamma) &= P\left(\chi^2_{n+\nu_{02}+1} > \frac{s_2^2}{\gamma}\right) \\
&= P\left(\frac{\chi^2_{n+\nu_{02}+1} - (n + \nu_{02} + 1)}{\sqrt{2(n + \nu_{02} + 1)}}\right. \\
&\qquad \left. > \frac{s_2^2/\gamma - (n + \nu_{02} + 1)}{\sqrt{2(n + \nu_{02} + 1)}}\right) \\
&= 1 - \Phi\left(\frac{s_2^2/\gamma - (n + \nu_{02} + 1)}{\sqrt{2(n + \nu_{02} + 1)}}\right) \\
&\quad - \epsilon_1(\nu_{02}) \longrightarrow 1,
\end{aligned}
\tag{46}
$$

where $\epsilon_1(\nu_{02}) \to 0$, as $\nu_{02} \to \infty$.

On the other hand, for any fixed n $(2 < n < \infty)$ and $\gamma \neq 0$, we have

$$
P(A_1 < \gamma) = P\left(\chi^2_{n-1} > \frac{(n-1)s_2^2}{n\gamma}\right) \le 1 - \epsilon_0,
\tag{47}
$$

for some $\epsilon_0 > 0$. Therefore, as $\nu_{02} \to \infty$, by (46) and (47), we have $P(\widetilde{B}_1 < \gamma) - P(A_1 < \gamma) > 0$. Furthermore, by (45), we have

$$
P(B_1 < \gamma) > P(A_1 < \gamma).
\tag{48}
$$

TABLE 1: P value and $\lim P(H_0 \mid x)$ for testing the Behrens-Fisher problem for different m and n.

				$m = 2, n = 2$					
$c - (\overline{x} - \overline{y})$	−2.5000	−2.1000	−1.8000	−1.6000	−1.3000	−0.9000	−0.5000	−0.3000	−0.1000
$p(x)$	0.1165	0.1405	0.1785	0.1880	0.2330	0.3110	0.3850	0.4155	0.4745
$\lim P(H_0 \mid x)$	0.0195	0.0420	0.0505	0.0705	0.1280	0.1945	0.3140	0.3770	0.4675
				$m = 5, n = 5$					
$c - (\overline{x} - \overline{y})$	−2.5000	−2.1000	−1.8000	−1.6000	−1.3000	−0.9000	−0.5000	−0.3000	−0.1000
$p(x)$	0.0385	0.0520	0.0760	0.0995	0.1405	0.2250	0.3460	0.3845	0.4705
$\lim P(H_0 \mid x)$	0.0060	0.0195	0.0400	0.0465	0.0765	0.1630	0.2840	0.3640	0.4630
				$m = 1, n = 8$					
$c - (\overline{x} - \overline{y})$	−2.5000	−2.1000	−1.8000	−1.6000	−1.3000	−0.9000	−0.5000	−0.3000	−0.1000
$p(x)$	0.1165	0.1405	0.1785	0.1880	0.2330	0.3110	0.3850	0.4155	0.4745
$\lim P(H_0 \mid x)$	0.0195	0.0420	0.0505	0.0705	0.1280	0.1945	0.3140	0.3770	0.4675
				$m = 3, n = 5$					
$c - (\overline{x} - \overline{y})$	−2.5000	−2.1000	−1.8000	−1.6000	−1.3000	−0.9000	−0.5000	−0.3000	−0.1000
$p(x)$	0.0600	0.0755	0.1030	0.1360	0.1765	0.2615	0.3495	0.4355	0.4595
$\lim P(H_0 \mid x)$	0.0070	0.0205	0.0405	0.0545	0.0845	0.1715	0.3005	0.3760	0.4670
				$m = 12, n = 3$					
$c - (\overline{x} - \overline{y})$	−2.5000	−2.1000	−1.8000	−1.6000	−1.3000	−0.9000	−0.5000	−0.3000	−0.1000
$p(x)$	0.0835	0.1130	0.1260	0.1585	0.2075	0.2975	0.3700	0.4245	0.4995
$\lim P(H_0 \mid x)$	0.0185	0.0315	0.0515	0.0730	0.1025	0.1810	0.2905	0.3635	0.4585

TABLE 2: P value and $\lim P(H_0 \mid x)$ for testing the Behrens-Fisher problem for different ν_{01} and ν_{02}.

				$\nu_{01} = 0.5, \nu_{02} = 0.5$					
$c - (\overline{x} - \overline{y})$	−2.5000	−2.1000	−1.8000	−1.6000	−1.3000	−0.9000	−0.5000	−0.3000	−0.1000
$p(x)$	0.0030	0.0080	0.0130	0.0240	0.0490	0.1310	0.2720	0.3495	0.4565
$\lim P(H_0 \mid x)$	0.0010	0.0035	0.0060	0.0135	0.0280	0.1030	0.2245	0.3375	0.4410
				$\nu_{01} = 2, \nu_{02} = 2$					
$c - (\overline{x} - \overline{y})$	−2.5000	−2.1000	−1.8000	−1.6000	−1.3000	−0.9000	−0.5000	−0.3000	−0.1000
$p(x)$	0.0025	0.0045	0.0165	0.0270	0.0545	0.1130	0.2550	0.3480	0.4595
$\lim P(H_0 \mid x)$	0.0005	0.0015	0.0050	0.0080	0.0245	0.1000	0.2125	0.3155	0.4345
				$\nu_{01} = 0.2, \nu_{02} = 0.5$					
$c - (\overline{x} - \overline{y})$	−2.5000	−2.1000	−1.8000	−1.6000	−1.3000	−0.9000	−0.5000	−0.3000	−0.1000
$p(x)$	0.0025	0.0085	0.0145	0.0295	0.0470	0.1225	0.2550	0.3400	0.4610
$\lim P(H_0 \mid x)$	0.0015	0.0030	0.0045	0.0130	0.0275	0.1055	0.2215	0.3245	0.4455
				$\nu_{01} = 0.5, \nu_{02} = 0.2$					
$c - (\overline{x} - \overline{y})$	−2.5000	−2.1000	−1.8000	−1.6000	−1.3000	−0.9000	−0.5000	−0.3000	−0.1000
$p(x)$	0.0055	0.0070	0.0170	0.0265	0.0545	0.1215	0.2485	0.3535	0.4650
$\lim P(H_0 \mid x)$	0.0010	0.0055	0.0070	0.0125	0.0440	0.1085	0.2500	0.3285	0.4300
				$\nu_{01} = 5, \nu_{02} = 0.5$					
$c - (\overline{x} - \overline{y})$	−2.5000	−2.1000	−1.8000	−1.6000	−1.3000	−0.9000	−0.5000	−0.3000	−0.1000
$p(x)$	0.0025	0.0060	0.0150	0.0230	0.0570	0.1145	0.2645	0.3685	0.4460
$\lim P(H_0 \mid x)$	0.0020	0.0025	0.0065	0.0095	0.0330	0.0890	0.2205	0.3105	0.4305

That is,

$$B_1 \overset{d}{<} A_1. \tag{49}$$

Similarly, for any fixed m ($2 < m < \infty$), as $\nu_{01} \rightarrow \infty$, we have

$$B_2 \overset{d}{<} A_2. \tag{50}$$

Therefore, similar to Theorem 3, as $\min\{\nu_{01}, \nu_{02}\} \rightarrow \infty$, for any fixed m and n with $2 < m, n < \infty$, we have

$$\sqrt{B_1 + B_2} \overset{d}{<} \sqrt{A_1 + A_2}. \tag{51}$$

The rest part of the proof is similar to that of (II) of Theorem 3. □

The following simulation results show that even for small and fixed values of m and n or ν_{01} and ν_{02}, the generalized P value and Bayesian evidence for testing the Behrens-Fisher problem are still reconcilable.

For fixed $\nu_{01} = 0.5$ and $\nu_{02} = 1$ and for $s_1^2 = 1$ and $s_2^2 = 4$, taking different values of m and n, some results of comparing the P value and $\lim_{\kappa_{01}, \kappa_{02}, \sigma_{01}, \sigma_{02} \to 0} P(H_0 \mid x)$ are listed in Table 1.

For fixed $m = 8$ and $n = 10$ and for $s_1^2 = 1$ and $s_2^2 = 4$, taking different values of ν_{01} and ν_{02}, we list some results of comparing P value and $\lim_{\kappa_{01}, \kappa_{02}, \sigma_{01}, \sigma_{02} \to 0} P(H_0 \mid x)$ in Table 2.

3. Conclusions

In the presence of the nuisance parameters, we study the reconcilability of the P value and the Bayesian evidence in the one-sided hypothesis testing problem about normal means. For the problem of testing a normal mean where the nuisance parameter is present, it is shown that the Bayesian and frequentist lines of evidence are reconcilable. For the Behrens-Fisher problem, it is illustrated that if the sample sizes m and n tend to infinity, then for fixed prior parameters ν_{01} and ν_{02}, both lines of evidence are reconcilable. Furthermore, it is illustrated that if the prior parameters ν_{01} and ν_{02} tend to infinity, then for any fixed sample sizes m and n, lines of evidence are reconcilable. Simulation results show that even for small and fixed values of sample sizes m and n or for small values of prior parameters ν_{01} and ν_{02}, the reconcilable conclusion of the Bayesian and frequentist evidence still holds.

This provides another illustration of testing situation where the Bayesian and frequentist evidence can be reconciled and may therefore to some extent prevent people from debasing or even dismissing P values as evidence in hypothesis testing problems. Furthermore, our results of the reconcilability in the one-sided testing situations may help us to come to the idea that maybe it is arbitrary to assert the irreconcilability of the evidence in the two-sided (point or interval) hypothesis testing problems and perhaps we should be concerned more about the appropriateness of the methods we employ to tackle a two-sided hypothesis in both the frequentist and the Bayesian frameworks.

Acknowledgments

References

[1] D. V. Lindley, "A statistical paradox," *Biometrika*, vol. 44, pp. 187–192, 1957.

[2] J. W. Pratt, "Bayesian interpretation of standard inference statements," *Journal of the Royal Statistical Society B*, vol. 27, pp. 169–203, 1965.

[3] G. Casella and R. L. Berger, "Reconciling Bayesian and frequentist evidence in the one-sided testing problems," *Journal of the American Statistical Association*, vol. 82, no. 397, pp. 106–111, 1987.

[4] M. S. Bartlett, "A comment on D. V. Lindley's statistical paradox," *Biometrika*, vol. 44, pp. 533–534, 1957.

[5] D. R. Cox, "The role of significance tests," *Scandinavian Journal of Statistics*, vol. 4, pp. 49–70, 1977.

[6] G. Shafer, "Lindley's paradox," *Journal of the American Statistical Association*, vol. 77, no. 378, pp. 325–351, 1982.

[7] J. O. Berger and M. Delampady, "Testing precise hypotheses," *Statistical Science*, vol. 2, pp. 317–352, 1987.

[8] J. O. Berger and T. Sellke, "Testing a point null hypothesis: the irreconcilability of p-values and evidence," *Journal of the American Statistical Association*, vol. 82, pp. 112–122, 1987.

[9] K. W. Tsui and S. Weerahandi, "Generalized *p*-values in significance testing of hypotheses in the presence of nuisance parameters," *Journal of the American Statistical Association*, vol. 84, pp. 602–607, 1989.

[10] J. Hannig, H. Iyer, and P. Patterson, "Fiducial generalized confidence intervals," *Journal of the American Statistical Association*, vol. 473, pp. 254–269, 2006.

[11] E. L. Lehmann, *Testing Statistical Hypotheses*, John Wiley & Sons, New York, NY, USA, 2nd edition, 1986.