British Journal of Mathematics & Computer Science
21(1): 1-20, 2017; Article no.BJMCS.31815
ISSN: 2231-0851

SCIENCEDOMAIN *international*
www.sciencedomain.org

The Behrens-Fisher Problem: A Z Distribution Approach

Ying-Ying Zhang[1*], Teng-Zhong Rong[1] and Man-Man Li[1]

[1] *Department of Statistics and Actuarial Science, College of Mathematics and Statistics, Chongqing University, Chongqing, China.*

Authors' contributions

This work was carried out in collaboration between all authors. Author YYZ defined the SIG distribution and the Z distribution, applied the Z distribution in the hypothesis testing of two normal means, and wrote the first draft of the manuscript. Author TZR did literature searches and revised the manuscript. Author MML revised the manuscript. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/BJMCS/2017/31815
Editor(s):
(1) Morteza Seddighin, Indiana University East Richmond, USA.
Reviewers:
(1) El-Nabulsi Ahmad Rami, Neijiang Normal University, Neijiang, Sichuan, China.
(2) H. Khalil, The University of Poonch Rawalakot, Azad Jammu and Kashmir, Pakistan.
(3) Weijing Zhao, Civil Aviation University of China, Tianjin, P.R. China.
Complete Peer review History: http://www.sciencedomain.org/review-history/18234

Received: 26th January 2017
Accepted: 7th March 2017
Published: 16th March 2017

Original Research Article

Abstract

We propose the Z distribution to tackle the Behrens-Fisher problem. First, we define the Z distribution which is a generalization of the t distribution, and then find the pdf and cdf of the Z distribution. After that, we apply the Z distribution in the hypothesis testing of two normal means, where three different assumptions of the variances are considered. The Z distribution is very flexible in the applications in which one statistics that obeys the Z distribution is applicable to all the three assumptions of the variances. Finally, we provide two groups of simulation studies for the hypothesis testing problems of two normal means.

Keywords: Behrens-Fisher problem; Z distribution; SIG distribution; two normal means; infinite series; exponential error structure.

**Corresponding author: E-mail: robertzhangyying@qq.com;*

2010 Mathematics Subject Classification: 62Exx, 62F03.

1 Introduction

In statistics, the Behrens-Fisher problem ([1]), named after Walter Ulrich Behrens and Ronald Fisher, is the problem of interval estimation and hypothesis testing concerning the difference between the means of two normally distributed populations when the variances of the two populations are not assumed to be equal, based on two independent samples.

There is a large literature dealing with the Behrens-Fisher problem, see e.g., [2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35]. Some articles are closely related to the Behrens-Fisher problem, see e.g., [36, 37, 38, 39]. Some people considered the generalized Behrens-Fisher problem (i.e., the k samples Behrens-Fisher problem), see e.g., [40, 41, 42, 43, 44, 22]. Others considered the multivariate Behrens-Fisher problem, see e.g., [45, 46, 47].

Our approach is different from the existing approaches. We introduce the Z distribution, which is very flexible in the applications of the hypothesis testing of two normal means. One statistics that obeys the Z distribution is applicable to all assumptions of σ_X^2 and σ_Y^2. Moreover, when the variances σ_X^2 and σ_Y^2 are unknown but the ratio of variances $R = \sigma_X^2/\sigma_Y^2 = 1/\rho^2$ is known, we find that the statistics given in [15] can be obtained through the Z distribution. Finally, we provide the simulation studies for the hypothesis testing problems in Remark 3.1. Two groups of simulation studies are considered. The simulation studies exemplify the power of our approach.

The rest of the paper is organized as follows. In the next Section 2, we define the SIG distribution and find its pdf, and then we define the Z distribution and find its pdf and cdf. In Section 3, we apply the Z distribution in the hypothesis testings of two normal means, three different assumptions: σ_X^2 and σ_Y^2 are known, $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ are unknown, and $\sigma_X^2 \neq \sigma_Y^2$ are unknown (the Behrens-Fisher problem) are considered. The Z distribution is very flexible in the applications in which one statistics that obeys the Z distribution is applicable to all assumptions of σ_X^2 and σ_Y^2. Section 4 provides two groups of simulation studies for the hypothesis testing problems in Remark 3.1. In particular, we assume two error structures for the l^∞ error: The polynomial error structure and the exponential error structure. Section 5 concludes.

2 SIG Distribution, Z Distribution, and the Cdf of the Z Distribution

In this section, we first define the SIG distribution, and then we utilize it to define the Z distribution. The two distributions are introduced by us. Finally, we analytically derive the cumulative distribution function (cdf) of the Z distribution which is useful in the applications in the hypothesis testing of two normal means.

Definition 2.1. Let $G_i \sim Gamma\left(\alpha_i, \beta_i\right),\ i = 1, 2$, G_1 and G_2 are independent. Let $W = G_1 + G_2$. Then W has a Sum of Independent Gamma (SIG) distribution, $SIG\left(\alpha_1, \beta_1, \alpha_2, \beta_2\right)$. Equivalently, a random variable W has an $SIG\left(\alpha_1, \beta_1, \alpha_2, \beta_2\right)$ distribution if it has a pdf

$$f_W\left(w|\alpha_1, \beta_1, \alpha_2, \beta_2\right) = \frac{1}{\Gamma\left(\alpha_1+\alpha_2\right)\beta_1^{\alpha_1}\beta_2^{\alpha_2}} w^{\alpha_1+\alpha_2-1} e^{-w/\beta_2} M_B\left(a\left(w\right)\right), \tag{2.1}$$
$$w > 0,\ \alpha_1, \beta_1, \alpha_2, \beta_2 > 0,$$

where

$$M_B\left(a\left(w\right)\right)=M_B\left(w\left(\frac{1}{\beta_2}-\frac{1}{\beta_1}\right)\right)=1+\sum_{k=1}^{\infty}\left(\prod_{r=0}^{k-1}\frac{\alpha_1+r}{\alpha_1+\alpha_2+r}\right)\frac{\left[w\left(\frac{1}{\beta_2}-\frac{1}{\beta_1}\right)\right]^k}{k!},$$

is the moment generating function (mgf) of $B\sim Beta\left(\alpha_1,\alpha_2\right)$ evaluated at $a\left(w\right)$.

The derivation of the pdf of the $SIG\left(\alpha_1,\beta_1,\alpha_2,\beta_2\right)$ distribution uses the convolution formula and the mgf of $Beta\left(\alpha_1,\alpha_2\right)$. The proof of (2.1) can be found in the supplement.

We make the following three remarks for the SIG distribution. The following remark is about the mgf of $B\sim Beta\left(\alpha_1,\alpha_2\right)$ evaluated at $a\left(w\right)$.

Remark 2.1. The mgf of $B\sim Beta\left(\alpha_1,\alpha_2\right)$ evaluated at $a\left(w\right)$ is

$$M_B\left(a\left(w\right)\right)=1+\sum_{k=1}^{\infty}a_kw^k=\sum_{k=0}^{\infty}a_kw^k,$$

where

$$a_0=1,\ a_k=\left(\prod_{r=0}^{k-1}\frac{\alpha_1+r}{\alpha_1+\alpha_2+r}\right)\frac{\left(\frac{1}{\beta_2}-\frac{1}{\beta_1}\right)^k}{k!},\ k\geq 1. \tag{2.2}$$

Moreover, if $\beta_1=\beta_2=\beta$, then $a_k=0$ for $k\geq 1$, and $M_B\left(a\left(w\right)\right)=1$.

The following remark is about the expectation, the variance, and the mgf of the SIG distribution.

Remark 2.2.

$$\begin{aligned}
EW&=EG_1+EG_2=\alpha_1\beta_1+\alpha_2\beta_2,\\
Var\left(W\right)&=Var\left(G_1+G_2\right)=Var\left(G_1\right)+Var\left(G_2\right)=\alpha_1\beta_1^2+\alpha_2\beta_2^2,\\
M_W\left(t\right)&=M_{G_1+G_2}\left(t\right)=M_{G_1}\left(t\right)M_{G_2}\left(t\right)=\left(1-\beta_1t\right)^{-\alpha_1}\left(1-\beta_2t\right)^{-\alpha_2},\ t<\min\left\{\frac{1}{\beta_1},\frac{1}{\beta_2}\right\}.
\end{aligned}$$

The following remark states that the SIG distribution generalizes the gamma distribution.

Remark 2.3. The SIG distribution is a generalization of the gamma distribution. More precisely,

$$\begin{aligned}
Gamma\left(\alpha,\beta\right)&=SIG\left(\alpha_1=\frac{\alpha}{2},\beta_1=\beta,\alpha_2=\frac{\alpha}{2},\beta_2=\beta\right)\\
&=Gamma\left(\frac{\alpha}{2},\beta\right)+Gamma\left(\frac{\alpha}{2},\beta\right).
\end{aligned}$$

We can verify it by writing out the pdfs of the two distributions and checking that they are equal. The proof can be found in the supplement.

With the SIG distribution, we are ready to define the Z distribution.

Definition 2.2. Let $Z=X/\sqrt{W}$, where $X\sim N\left(0,1\right)$ and $W\sim SIG\left(\alpha_1,\beta_1,\alpha_2,\beta_2\right)$ are independent. Then Z has a Z distribution, $Z\left(\alpha_1,\beta_1,\alpha_2,\beta_2\right)$. Equivalently, a random variable Z has a $Z\left(\alpha_1,\beta_1,\alpha_2,\beta_2\right)$ distribution if it has a pdf

$$\begin{aligned}
&f_Z\left(z|\alpha_1,\beta_1,\alpha_2,\beta_2\right)\\
&=\frac{1}{\sqrt{2\pi}\Gamma\left(\alpha_1+\alpha_2\right)\beta_1^{\alpha_1}\beta_2^{\alpha_2}}\sum_{k=0}^{\infty}a_k\Gamma\left(\alpha_1+\alpha_2+k+\frac{1}{2}\right)\left(\frac{1}{\frac{z^2}{2}+\frac{1}{\beta_2}}\right)^{\alpha_1+\alpha_2+k+\frac{1}{2}},\\
&z\in\mathbb{R},\ \alpha_1,\beta_1,\alpha_2,\beta_2>0,
\end{aligned} \tag{2.3}$$

where a_k $\left(k\geq 0\right)$ are given by (2.2).

The proof of (2.3) follows from the derivation of the t pdf. It is elementary but tedious, and thus we put it into the supplement. For the Z distribution, we have the following remark which states that the Z distribution generalizes the t distribution.

Remark 2.4. The $Z(\alpha_1, \beta_1, \alpha_2, \beta_2)$ distribution is a generalization of the $t(p)$ distribution. More precisely,

$$t(p) = Z\left(\alpha_1 = \frac{p}{4}, \beta_1 = \frac{2}{p}, \alpha_2 = \frac{p}{4}, \beta_2 = \frac{2}{p}\right).$$

We can verify it by writing out the pdfs of the two distributions and checking that they are equal. The proof can be found in the supplement.

For the cdf of the $Z(\alpha_1, \beta_1, \alpha_2, \beta_2)$ distribution, we have the following theorem whose proof can be found in the supplement.

Theorem 2.1. *The cdf of the $Z(\alpha_1, \beta_1, \alpha_2, \beta_2)$ distribution is given by*

$$F_Z(z|\alpha_1, \beta_1, \alpha_2, \beta_2) = \frac{1}{\sqrt{\pi}\Gamma(\alpha_1+\alpha_2)\beta_1^{\alpha_1}\beta_2^{\alpha_2}} \sum_{k=0}^{\infty} a_k \Gamma\left(\alpha_1+\alpha_2+k+\frac{1}{2}\right)\beta_2^{\alpha_1+\alpha_2+k} J_k,$$

where

$$J_k = \int_{-\infty}^{\sqrt{\frac{\beta_2}{2}}z} \frac{1}{(1+u^2)^{\alpha_1+\alpha_2+k+\frac{1}{2}}} du,$$

and a_k $(k \geq 0)$ are given by (2.2). In particular, when $p_0 = 2(\alpha_1+\alpha_2)$ is a positive integer,

$$\begin{aligned} &F_Z(z|\alpha_1, \beta_1, \alpha_2, \beta_2) \\ &= \frac{1}{\Gamma(\alpha_1+\alpha_2)\beta_1^{\alpha_1}\beta_2^{\alpha_2}} \sum_{k=0}^{\infty} a_k \beta_2^{\alpha_1+\alpha_2+k}\Gamma(\alpha_1+\alpha_2+k) F_{T_{p_k}}\left(\sqrt{(\alpha_1+\alpha_2+k)\beta_2}z\right), \end{aligned} \tag{2.4}$$

where $p_k = 2(\alpha_1+\alpha_2+k)$, $k = 0, 1, \ldots,$ and $F_{T_{p_k}}(\cdot)$ is the cdf of the t random variable with p_k degrees of freedom.

For Theorem 2.1, we have the following three remarks. The following remark states that the in particular part of Theorem 2.1 is usually applicable in the hypothesis testing of two normal means.

Remark 2.5. For the applications in Section 3, we have $\alpha_1 = (n-1)/2$ and $\alpha_2 = (m-1)/2$ or $\alpha_1 = \alpha_2 = (n+m-2)/4$, then $p_0 = 2(\alpha_1+\alpha_2) = n+m-2$ is a positive integer for $n+m \geq 3$. That is, the in particular part of Theorem 2.1 is usually applicable.

The following remark states that the l^∞ error of the approximation decays exponentially.

Remark 2.6. When $p_0 = 2(\alpha_1+\alpha_2)$ is a positive integer, the cdf of Z is usually approximated by its truncated sum

$$\begin{aligned} &F_Z^K(z|\alpha_1, \beta_1, \alpha_2, \beta_2) \\ &= \frac{1}{\Gamma(\alpha_1+\alpha_2)\beta_1^{\alpha_1}\beta_2^{\alpha_2}} \sum_{k=0}^{K} a_k \beta_2^{\alpha_1+\alpha_2+k}\Gamma(\alpha_1+\alpha_2+k) F_{T_{p_k}}\left(\sqrt{(\alpha_1+\alpha_2+k)\beta_2}z\right). \end{aligned}$$

Define the l^∞ error of the approximation $F_Z^K(z)$ by

$$l^\infty(K) = \max_{z\in\mathbb{R}} \left|F_Z(z) - F_Z^K(z)\right|.$$

Section 4 exemplifies that

$$l^\infty(K) = Cq^K,$$

for some positive constants C and $0 < q < 1$. That is, the $l^\infty(K)$ decays exponentially.

The following remark is useful in the programming of the pdf and the cdf of the Z distribution for some special parameterizations.

Remark 2.7. When $(\alpha_1, \beta_1, \alpha_2, \beta_2) = \left(\frac{p}{4}, \frac{2}{p}, \frac{p}{4}, \frac{2}{p}\right)$, then by Remark 2.4,

$$Z\left(\alpha_1 = \frac{p}{4}, \beta_1 = \frac{2}{p}, \alpha_2 = \frac{p}{4}, \beta_2 = \frac{2}{p}\right) = t\left(p\right).$$

Thus,

$$\begin{aligned} F_Z\left(z|\alpha_1 = \frac{p}{4}, \beta_1 = \frac{2}{p}, \alpha_2 = \frac{p}{4}, \beta_2 = \frac{2}{p}\right) &= F_{T_p}\left(z\right), \\ f_Z\left(z|\alpha_1 = \frac{p}{4}, \beta_1 = \frac{2}{p}, \alpha_2 = \frac{p}{4}, \beta_2 = \frac{2}{p}\right) &= f_{T_p}\left(z\right), \end{aligned}$$

where $T_p \sim t\left(p\right)$. When $\beta_1 = \beta_2$, then by Remark 2.1, $a_k = 0$ for $k \geq 1$. Thus,

$$F_Z\left(z|\alpha_1, \beta_1, \alpha_2, \beta_2\right) = F_{T_{p_0}}\left(\sqrt{\left(\alpha_1 + \alpha_2\right)\beta_2} z\right),$$

where $p_0 = 2\left(\alpha_1 + \alpha_2\right)$ and $T_{p_0} \sim t\left(p_0\right)$, and

$$\begin{aligned} f_Z\left(z|\alpha_1, \beta_1, \alpha_2, \beta_2\right) &= F'_Z\left(z|\alpha_1, \beta_1, \alpha_2, \beta_2\right) \\ &= \sqrt{\left(\alpha_1 + \alpha_2\right)\beta_2} f_{T_{p_0}}\left(\sqrt{\left(\alpha_1 + \alpha_2\right)\beta_2} z\right). \end{aligned}$$

3 Applications in the Hypothesis Testing of Two Normal Means

In this section, we apply the Z distribution in the hypothesis testing of two normal means under three different assumptions of the variances.

Let $X_1, \ldots, X_n$ be a random sample from a normal distribution $N\left(\mu_X, \sigma_X^2\right)$, and $Y_1, \ldots, Y_m$ be an independent random sample from another normal distribution $N\left(\mu_Y, \sigma_Y^2\right)$. We are interested in testing

$$H_0 : \mu_X = \mu_Y \quad \text{versus} \quad H_1 : \mu_X \neq \mu_Y$$

under three different assumptions of the variances.

Assumption 1. σ_X^2 and σ_Y^2 are known

Under H_0, the usual pivot is

$$N = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} \sim N\left(0, 1\right).$$

The pivot using the Z distribution is given by

$$Z_1 = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} = \frac{\left(\bar{X} - \bar{Y}\right)/\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}/\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} = \frac{N}{\sqrt{W_1}},$$

where

$$W_1 = \frac{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}.$$

Since

$$\frac{(n-1)\,S_X^2}{\sigma_X^2} = \chi^2_{n-1},\ \frac{(m-1)\,S_Y^2}{\sigma_Y^2} = \chi^2_{m-1},$$

we have

$$S_X^2 = \frac{\sigma_X^2}{n-1}\chi^2_{n-1},\ S_Y^2 = \frac{\sigma_Y^2}{m-1}\chi^2_{m-1}.$$

Thus W_1 can be rewritten as

$$W_1 = \frac{\frac{\sigma_X^2}{n(n-1)}\chi^2_{n-1} + \frac{\sigma_Y^2}{m(m-1)}\chi^2_{m-1}}{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}.$$

Let

$$a_1 = \frac{\sigma_X^2}{n\,(n-1)\left(\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right)},\ a_2 = \frac{\sigma_Y^2}{m\,(m-1)\left(\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}\right)}. \tag{3.1}$$

Therefore,

$$\begin{aligned} W_1 &= a_1\chi^2_{n-1} + a_2\chi^2_{m-1} \\ &= a_1 G\left(\frac{n-1}{2}, 2\right) + a_2 G\left(\frac{m-1}{2}, 2\right) \\ &= G\left(\frac{n-1}{2}, 2a_1\right) + G\left(\frac{m-1}{2}, 2a_2\right) \\ &\sim SIG\left(\alpha_1 = \frac{n-1}{2}, \beta_1 = 2a_1, \alpha_2 = \frac{m-1}{2}, \beta_2 = 2a_2\right), \end{aligned}$$

where a_1 and a_2 are given by (3.1). Consequently,

$$Z_1 = \frac{N}{\sqrt{W_1}} \sim Z\left(\alpha_1 = \frac{n-1}{2}, \beta_1 = 2a_1, \alpha_2 = \frac{m-1}{2}, \beta_2 = 2a_2\right).$$

Assumption 2. $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ are unknown

Under H_0, the usual pivot is

$$T = \frac{\bar{X} - \bar{Y}}{\sqrt{S_p^2\left(\frac{1}{n} + \frac{1}{m}\right)}} \sim t\,(n+m-2)\,,$$

where

$$S_p^2 = \frac{1}{n+m-2}\left[\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 + \sum_{i=1}^{m}\left(Y_i - \bar{Y}\right)^2\right] = \frac{(n-1)\,S_X^2 + (m-1)\,S_Y^2}{n+m-2}$$

is referred to as a pooled variance estimate. Now we derive the pivot using the Z distribution. By Remark 2.4 and $T \sim t\,(n+m-2)$, we have

$$T \sim Z\left(\alpha_1 = \frac{n+m-2}{4}, \beta_1 = \frac{2}{n+m-2}, \alpha_2 = \frac{n+m-2}{4}, \beta_2 = \frac{2}{n+m-2}\right).$$

That is, the usual pivot T is a Z distribution.

The random variable T has a Z distribution with other parameter values. In fact,

$$T = \frac{\left(\bar{X} - \bar{Y}\right)/\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}{\sqrt{S_p^2\left(\frac{1}{n} + \frac{1}{m}\right)}/\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} = \frac{N}{\sqrt{W_2'}},$$

where $N \sim N(0,1)$,

$$\begin{aligned}
W_2' &= \frac{S_p^2\left(\frac{1}{n}+\frac{1}{m}\right)}{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}} = \frac{S_p^2\left(\frac{1}{n}+\frac{1}{m}\right)}{\sigma^2\left(\frac{1}{n}+\frac{1}{m}\right)} = \frac{S_p^2}{\sigma^2} = \frac{(n+m-2)\,S_p^2}{\sigma^2\,(n+m-2)} \\
&= \frac{(n-1)\,S_X^2+(m-1)\,S_Y^2}{\sigma^2\,(n+m-2)} = \frac{\chi^2_{n-1}+\chi^2_{m-1}}{n+m-2} \\
&= \frac{1}{n+m-2} G\left(\frac{n-1}{2},2\right) + \frac{1}{n+m-2} G\left(\frac{m-1}{2},2\right) \\
&= G\left(\frac{n-1}{2},\frac{2}{n+m-2}\right) + G\left(\frac{m-1}{2},\frac{2}{n+m-2}\right) \\
&\sim SIG\left(\alpha_1=\frac{n-1}{2},\beta_1=\frac{2}{n+m-2},\alpha_2=\frac{m-1}{2},\beta_2=\frac{2}{n+m-2}\right).
\end{aligned}$$

Therefore,

$$T=\frac{N}{\sqrt{W_2'}} \sim Z\left(\alpha_1=\frac{n-1}{2},\beta_1=\frac{2}{n+m-2},\alpha_2=\frac{m-1}{2},\beta_2=\frac{2}{n+m-2}\right).$$

By (2.3), it is easy to check that

$$\begin{aligned}
&Z\left(\alpha_1=\frac{n+m-2}{4},\beta_1=\frac{2}{n+m-2},\alpha_2=\frac{n+m-2}{4},\beta_2=\frac{2}{n+m-2}\right) \\
&= Z\left(\alpha_1=\frac{n-1}{2},\beta_1=\frac{2}{n+m-2},\alpha_2=\frac{m-1}{2},\beta_2=\frac{2}{n+m-2}\right).
\end{aligned}$$

That is, the two Z distributions with different parameterizations are equal.

We can derive another pivot using the Z distribution. Let

$$Z_2=\frac{\bar X-\bar Y}{\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}} = \frac{\left(\bar X-\bar Y\right)/\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}{\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}/\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}} = \frac{N}{\sqrt{W_2}},$$

where $N \sim N(0,1)$,

$$W_2=\frac{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}} = a_1\chi^2_{n-1}+a_2\chi^2_{m-1},$$

$$a_1=\frac{\sigma_X^2}{n\,(n-1)\left(\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}\right)} = \frac{1}{n\,(n-1)\left(\frac{1}{n}+\frac{1}{m}\right)} = \frac{m}{(n-1)\,(m+n)}, \tag{3.2}$$

$$a_2=\frac{\sigma_Y^2}{m\,(m-1)\left(\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}\right)} = \frac{1}{m\,(m-1)\left(\frac{1}{n}+\frac{1}{m}\right)} = \frac{n}{(m-1)\,(m+n)}. \tag{3.3}$$

Note the derivation for W_2 is similar to that for W_1. Thus,

$$W_2 \sim SIG\left(\alpha_1=\frac{n-1}{2},\beta_1=2a_1,\alpha_2=\frac{m-1}{2},\beta_2=2a_2\right),$$

where a_1 and a_2 are given by (3.2) and (3.3). Consequently,

$$Z_2=\frac{N}{\sqrt{W_2}} \sim Z\left(\alpha_1=\frac{n-1}{2},\beta_1=2a_1,\alpha_2=\frac{m-1}{2},\beta_2=2a_2\right).$$

Assumption 3. $\sigma_X^2 \neq \sigma_Y^2$ are unknown

This is the Behrens-Fisher Problem. Under H_0, the usual pivot is

$$T' = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}} = \frac{\left(\bar{X} - \bar{Y}\right)/\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}}{\sqrt{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}/\sqrt{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}}} = \frac{N}{\sqrt{W_3}}.$$

However, the exact distribution of T' is not pleasant ([48]). The distribution of T' is usually approximated by using Satterthwaite's approximation ([7]).

$$W_3 = \frac{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} \approx \frac{\chi_\nu^2}{\nu},$$

where ν can be estimated with

$$\hat{\nu} = \frac{\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)^2}{\frac{S_X^4}{n^2(n-1)} + \frac{S_Y^4}{m^2(m-1)}}.$$

Therefore,

$$T' = \frac{N}{\sqrt{W_3}} \approx \frac{N}{\sqrt{\frac{\chi_{\hat{\nu}}^2}{\hat{\nu}}}} = t\left(\hat{\nu}\right).$$

By Remark 2.4,

$$T' \approx t\left(\hat{\nu}\right) = Z\left(\alpha_1 = \frac{\hat{\nu}}{4}, \beta_1 = \frac{2}{\hat{\nu}}, \alpha_2 = \frac{\hat{\nu}}{4}, \beta_2 = \frac{2}{\hat{\nu}}\right).$$

That is, the usual pivot T' is approximately a Z distribution.

We can approximate T' by another Z distribution. First,

$$\begin{aligned} W_3 &= \frac{\frac{S_X^2}{n} + \frac{S_Y^2}{m}}{\frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{m}} = a_1\chi_{n-1}^2 + a_2\chi_{m-1}^2 \\ &\sim SIG\left(\alpha_1 = \frac{n-1}{2}, \beta_1 = 2a_1, \alpha_2 = \frac{m-1}{2}, \beta_2 = 2a_2\right), \end{aligned}$$

where a_1 and a_2 are given by (3.1). Note that the derivation for W_3 is similar to that for W_1. Since σ_X^2 and σ_Y^2 are unknown and they are not equal, and a_1 and a_2 depend on σ_X^2 and σ_Y^2, the distribution of W_3 is not completely known. It is natural to use $S_X^2 \approx \sigma_X^2$ and $S_Y^2 \approx \sigma_Y^2$. Therefore,

$$\hat{a}_1 = \frac{S_X^2}{n\left(n-1\right)\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)},\ \hat{a}_2 = \frac{S_Y^2}{m\left(m-1\right)\left(\frac{S_X^2}{n} + \frac{S_Y^2}{m}\right)},$$

$$W_3 \approx SIG\left(\alpha_1 = \frac{n-1}{2}, \beta_1 = 2\hat{a}_1, \alpha_2 = \frac{m-1}{2}, \beta_2 = 2\hat{a}_2\right),$$

$$T' = \frac{N}{\sqrt{W_3}} \approx Z\left(\alpha_1 = \frac{n-1}{2}, \beta_1 = 2\hat{a}_1, \alpha_2 = \frac{m-1}{2}, \beta_2 = 2\hat{a}_2\right).$$

Note that the above Z distribution shares some characteristics of Fisher's solution using fiducial argument (see the supplemental file). Namely, the Z distribution assumes that $\hat{R} = S_X^2/S_Y^2 = s_X^2/s_Y^2$ is known.

In particular, when the ratio of variances $R = \sigma_X^2/\sigma_Y^2 = 1/\rho^2$ is known, [15] showed that under H_0,

$$\begin{aligned}\tilde{T} &= \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{1}{n}+\frac{\rho^2}{m}}\sqrt{\frac{(n-1)S_X^2+(m-1)S_Y^2/\rho^2}{n+m-2}}} \\ &\sim t\left(n+m-2\right) \\ &= Z\left(\alpha_1 = \frac{n+m-2}{4}, \beta_1 = \frac{2}{n+m-2}, \alpha_2 = \frac{n+m-2}{4}, \beta_2 = \frac{2}{n+m-2}\right).\end{aligned} \tag{3.4}$$

We can derive another pivot using the Z distribution.

$$Z_4 = \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}} = \frac{\left(\bar{X}-\bar{Y}\right)/\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}{\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}/\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}} = \frac{N}{\sqrt{W_4}},$$

where $N \sim N\left(0,1\right)$,

$$\begin{aligned}W_4 &= \frac{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}} = a_1\chi^2_{n-1} + a_2\chi^2_{m-1} \\ &\sim SIG\left(\alpha_1 = \frac{n-1}{2}, \beta_1 = 2a_1, \alpha_2 = \frac{m-1}{2}, \beta_2 = 2a_2\right),\end{aligned}$$

a_1 and a_2 are given by (3.1). But now $R = \sigma_X^2/\sigma_Y^2$ is known. Therefore,

$$a_1 = \frac{1}{n\left(n-1\right)\left(\frac{1}{n}+\frac{1/R}{m}\right)} = \frac{mR}{\left(n-1\right)\left(mR+n\right)},$$

$$a_2 = \frac{1}{m\left(m-1\right)\left(\frac{R}{n}+\frac{1}{m}\right)} = \frac{n}{\left(m-1\right)\left(mR+n\right)}.$$

Consequently,

$$Z_4 = \frac{N}{\sqrt{W_4}} \sim Z\left(\alpha_1 = \frac{n-1}{2}, \beta_1 = 2a_1, \alpha_2 = \frac{m-1}{2}, \beta_2 = 2a_2\right).$$

The Z distribution is very flexible in the applications of the hypothesis testing of two normal means. If a pivot is a $t\left(p\right)$ distribution, then by Remark 2.4, it is a $Z\left(\alpha_1 = \frac{p}{4}, \beta_1 = \frac{2}{p}, \alpha_2 = \frac{p}{4}, \beta_2 = \frac{2}{p}\right)$ distribution. The following remark states that one statistics that obeys the Z distribution is applicable to all assumptions of σ_X^2 and σ_Y^2.

Remark 3.1. Under H_0, the statistics

$$\begin{aligned}Z &= \frac{\bar{X}-\bar{Y}}{\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}} = \frac{\left(\bar{X}-\bar{Y}\right)/\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}{\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}/\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}} = \frac{N}{\sqrt{W}} \\ &\sim Z\left(\alpha_1 = \frac{n-1}{2}, \beta_1 = 2a_1, \alpha_2 = \frac{m-1}{2}, \beta_2 = 2a_2\right),\end{aligned}$$

where $N \sim N\left(0,1\right)$,

$$\begin{aligned}W &= \frac{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}} = a_1\chi^2_{n-1} + a_2\chi^2_{m-1} \\ &\sim SIG\left(\alpha_1 = \frac{n-1}{2}, \beta_1 = 2a_1, \alpha_2 = \frac{m-1}{2}, \beta_2 = 2a_2\right),\end{aligned}$$

$$a_1 = \frac{\sigma_X^2}{n(n-1)\left(\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}\right)} = \frac{1}{n(n-1)\left(\frac{1}{n}+\frac{1/R}{m}\right)} = \frac{mR}{(n-1)(mR+n)}, \tag{3.5}$$

$$a_2 = \frac{\sigma_Y^2}{m(m-1)\left(\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}\right)} = \frac{1}{m(m-1)\left(\frac{R}{n}+\frac{1}{m}\right)} = \frac{n}{(m-1)(mR+n)}, \tag{3.6}$$

$$R = \frac{\sigma_X^2}{\sigma_Y^2}.$$

- If σ_X^2 and σ_Y^2 are known, then R is known, so a_1 and a_2 are known, and the Z distribution is well defined.
- If $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ are unknown, but $R = 1$ is known, so a_1 and a_2 are known, and the Z distribution is well defined.
- If $\sigma_X^2 \neq \sigma_Y^2$ are unknown, but $R \neq 1$ is known, so a_1 and a_2 are known, and the Z distribution is well defined.
- If $\sigma_X^2 \neq \sigma_Y^2$ are unknown, and R is also unknown (the Behrens-Fisher Problem), then we can approximate R by $\hat{R} = S_X^2/S_Y^2$, so a_1 and a_2 are approximated by

$$\hat{a}_1 = \frac{m\hat{R}}{(n-1)\left(m\hat{R}+n\right)},\ \hat{a}_2 = \frac{n}{(m-1)\left(m\hat{R}+n\right)}.$$

Therefore, the Z distribution is well defined.

The P value of the Z distribution is computed as follows.

$$\begin{aligned} p &= P(|Z| \geq |z|) = 2P(Z \geq |z|) \\ &= 2(1-P(Z \leq |z|)) = 2(1-F_Z(|z|)), \end{aligned}$$

where $z = (\bar{x}-\bar{y})/\sqrt{s_X^2/n + s_Y^2/m}$ is the observed value of Z, $F_Z(\cdot)$ is the cdf of the random variable Z. If $p < \alpha$, the chosen nominal significance level of the test, then reject $H_0 : \mu_X = \mu_Y$, otherwise, accept H_0. A simulation study of the P value is given in the next section.

The $100 \times (1-\alpha)\,\%$ interval estimate of $\theta\ (= \mu_X - \mu_Y)$ based on the $Z(\theta)$ statistics,

$$Z(\theta) = \frac{\left(\bar{X}-\bar{Y}\right)-\theta}{\sqrt{\frac{S_X^2}{n}+\frac{S_Y^2}{m}}},$$

is defined as a set $\Theta_{1-\alpha} = \{\theta : P(|Z(\theta)| \geq |Z_{obs}(\theta)|) \geq \alpha\}$ and is given as

$$(\bar{x}-\bar{y}) \pm \gamma_{1-\frac{\alpha}{2}}\sqrt{\frac{s_X^2}{n}+\frac{s_Y^2}{m}},$$

where

$$Z_{obs}(\theta) = \frac{(\bar{x}-\bar{y})-\theta}{\sqrt{\frac{s_X^2}{n}+\frac{s_Y^2}{m}}}$$

is the observed value of $Z(\theta)$, and

$$\gamma_{1-\frac{\alpha}{2}} = F_Z^{-1}\left(1-\frac{\alpha}{2}\right)$$

is the lower $1-\alpha/2$ critical value (cut-off point or quantile) of the Z distribution. How to calculate or compute $\gamma_{1-\frac{\alpha}{2}}$ is a problem.

The following remark states that the $\tilde{T}$ statistics (3.4) given in [15] can be obtained through the Z distribution.

Remark 3.2. We can define the Z distribution differently. Let

$$\begin{aligned} Z &= \frac{\bar{X}-\bar{Y}}{\sqrt{f\left(S_X^2,S_Y^2\right)}} = \frac{\left(\bar{X}-\bar{Y}\right)/\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}}{\sqrt{f\left(S_X^2,S_Y^2\right)}/\sqrt{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}}} = \frac{N}{\sqrt{W}} \\ &\sim Z\left(\alpha_1 = \frac{n-1}{2}, \beta_1 = 2a_1, \alpha_2 = \frac{m-1}{2}, \beta_2 = 2a_2\right), \end{aligned}$$

where $N \sim N(0,1)$,

$$f\left(S_X^2,S_Y^2\right) = k_1(n,m,R)\,S_X^2 + k_2(n,m,R)\,S_Y^2,$$

$$\begin{aligned} W &= \frac{f\left(S_X^2,S_Y^2\right)}{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}} = \frac{k_1S_X^2+k_2S_Y^2}{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}} \\ &= \frac{k_1\frac{\sigma_X^2}{n-1}\chi_{n-1}^2 + k_2\frac{\sigma_Y^2}{m-1}\chi_{m-1}^2}{\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}} \\ &= a_1\chi_{n-1}^2 + a_2\chi_{m-1}^2 \\ &\sim SIG\left(\alpha_1 = \frac{n-1}{2}, \beta_1 = 2a_1, \alpha_2 = \frac{m-1}{2}, \beta_2 = 2a_2\right), \end{aligned}$$

$$\begin{aligned} a_1 &= \frac{k_1\sigma_X^2}{(n-1)\left(\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}\right)} = \frac{k_1}{(n-1)\left(\frac{1}{n}+\frac{1/R}{m}\right)} = \frac{k_1mnR}{(n-1)(mR+n)}, \\ a_2 &= \frac{k_2\sigma_Y^2}{(m-1)\left(\frac{\sigma_X^2}{n}+\frac{\sigma_Y^2}{m}\right)} = \frac{k_2}{(m-1)\left(\frac{R}{n}+\frac{1}{m}\right)} = \frac{k_2mn}{(m-1)(mR+n)}, \\ R &= \frac{\sigma_X^2}{\sigma_Y^2}. \end{aligned}$$

When R is known, we can choose

$$k_1 = k_1(n,m,R) \text{ and } k_2 = k_2(n,m,R)$$

such that the resulting Z distribution has a simple form. By Remark 2.1, we know that when $\beta_1 = \beta_2 = \beta$, then the pdf of the Z distribution is simplified. When we choose $a_1 = a_2$, that is,

$$\frac{k_1mnR}{(n-1)(mR+n)} = \frac{k_2mn}{(m-1)(mR+n)},$$

we obtain

$$\frac{k_1}{k_2} = \frac{n-1}{(m-1)R}.$$

In fact, we can choose k_1 and k_2, such that the resulting Z distribution is a $t(n+m-2)$ distribution. We have

$$\begin{aligned} W &= a_1\chi_{n-1}^2 + a_2\chi_{m-1}^2 = \frac{k_1mnR}{(n-1)(mR+n)}\left(\chi_{n-1}^2+\chi_{m-1}^2\right) \\ &= \frac{k_1mnR}{(n-1)(mR+n)}\chi_{n+m-2}^2 = \frac{\chi_{n+m-2}^2}{n+m-2}, \end{aligned}$$

which means that

$$\frac{k_1 mnR}{(n-1)(mR+n)} = \frac{1}{n+m-2}.$$

Thus,

$$k_1 = \frac{(n-1)(mR+n)}{mnR(n+m-2)}. \tag{3.7}$$

Therefore,

$$k_2 = \frac{(m-1)R}{n-1}k_1 = \frac{(m-1)(mR+n)}{mn(n+m-2)}. \tag{3.8}$$

Consequently, if we let k_1 and k_2 be defined by (3.7) and (3.8), then

$$Z = \frac{N}{\sqrt{W}} \sim t(n+m-2).$$

One can easily check that the Z statistics with k_1 and k_2 defined by (3.7) and (3.8) is the same as the $\tilde{T}$ statistics (3.4) given in [15]. In another words, the $\tilde{T}$ statistics (3.4) given in [15] can be obtained through the Z distribution.

It is also easy to check that when

$$k_1 = \frac{1}{n} \text{ and } k_2 = \frac{1}{m},$$

a_1 and a_2 reduces to (3.5) and (3.6).

4 Simulation Studies

In this section, we provide the simulation studies for the hypothesis testing problems in Remark 3.1. We consider two groups of simulation studies. The first group considers $\mu_X = 1$ and $\mu_Y = 1.1$, for this group we would probably accept $H_0 : \mu_X = \mu_Y$. The parameters for group 1 are given in Table 1. The second group considers $\mu_X = 1$ and $\mu_Y = 2$, for this group we would expect rejecting $H_0 : \mu_X = \mu_Y$ and accepting $H_1 : \mu_X \neq \mu_Y$. The parameters for group 2 are the same as those for group 1 in Table 4, except that $\mu_Y = 2$.

Table 1. Parameters for group 1.

	$n = 10, m = 20, \mu_X = 1, \mu_Y = 1.1$
Case 1	σ_X^2, σ_Y^2 are known, $R = 2/3$ $X \sim N(1,2)$ and $Y \sim N(1.1,3)$
Case 2	$\sigma_X^2 = \sigma_Y^2 = \sigma^2$ are unknown, $R = \sigma_X^2/\sigma_Y^2 = 1$ $X \sim N(1,1)$ and $Y \sim N(1.1,1)$
Case 3	$\sigma_X^2 \neq \sigma_Y^2$ are unknown, $R = \sigma_X^2/\sigma_Y^2 = 2$ is known $X \sim N(1,2)$ and $Y \sim N(1.1,1)$
Case 4	$\sigma_X^2 \neq \sigma_Y^2$ are unknown, $R = \sigma_X^2/\sigma_Y^2 = 1/2$ is known $X \sim N(1,1)$ and $Y \sim N(1.1,2)$
Case 5	$\sigma_X^2 \neq \sigma_Y^2$ are unknown, $R = \sigma_X^2/\sigma_Y^2$ is also unknown, $\hat{R} = S_X^2/S_Y^2 > 1$ $X \sim N(1,2)$ and $Y \sim N(1.1,1)$, $\hat{R} = S_X^2/S_Y^2 \approx \sigma_X^2/\sigma_Y^2 = 2 > 1$
Case 6	$\sigma_X^2 \neq \sigma_Y^2$ are unknown, $R = \sigma_X^2/\sigma_Y^2$ is also unknown, $\hat{R} = S_X^2/S_Y^2 < 1$ $X \sim N(1,1)$ and $Y \sim N(1.1,2)$, $\hat{R} = S_X^2/S_Y^2 \approx \sigma_X^2/\sigma_Y^2 = 1/2 < 1$

Before the simulation studies, we do some theoretical analysis of the error structure. By error, we refer to the l^∞ error (maximum absolute error) of the approximate solution (a vector) and the exact solution (a vector), which is not known since it is an infinite series and is replaced by the approximate solution with very large truncation number K of the series. We assume two error structures for the l^∞ error. The first one is the polynomial error, which is the one we tried first. The second one is the exponential error, which is the better error structure as the simulation studies will exemplify.

For polynomial error, we assume that there exist positive constants C and β_1, such that

$$l^\infty(K) = C\left(\frac{1}{K}\right)^{\beta_1}.$$

Taking base 10 logs (of course one can try base e) on both sides of the above equation, we obtain

$$\log_{10} l^\infty(K) = \log_{10} C + \beta_1 \log_{10}\frac{1}{K}.$$

Let

$$Y_K = \log_{10} l^\infty(K),\ \beta_0 = \log_{10} C,\ \text{and } X_K = \log_{10}\frac{1}{K}.$$

Then

$$Y_K = \beta_0 + \beta_1 X_K. \tag{4.1}$$

That is, Y_K and X_K have a linear relationship. We then use the R function lm() to compute the coefficients (β_0, β_1). Then $C = 10^{\beta_0}$. To see whether the polynomial error structure is good for the l^∞ error of our problem, we can compute and record the sum of squares of the residuals $S^2_{r,poly}$ (the smaller the better), the residual standard error $\hat{\sigma}_{poly}$ (the smaller the better), and the multiple R-squared R^2_{poly} (the larger the better) of the linear model. Given K (e.g., $K = 150$), if we want to know what is the predictive $l^\infty(K)$, we can simply calculate

$$l^\infty(K) = C\left(\frac{1}{K}\right)^{\beta_1} = 10^{\beta_0}\left(\frac{1}{K}\right)^{\beta_1} = \frac{10^{\beta_0}}{K^{\beta_1}}. \tag{4.2}$$

If we want to know for what value of K the predictive $l^\infty(K)$ is less than a specified value δ, e.g., 0.001, we should let

$$l^\infty(K) = \frac{10^{\beta_0}}{K^{\beta_1}} < \delta.$$

Therefore, by elementary calculations, we get

$$K > 10^{\frac{\beta_0 - \log_{10}\delta}{\beta_1}}. \tag{4.3}$$

For exponential error, we assume that there exist positive constants C and $0 < q < 1$, such that

$$l^\infty(K) = Cq^K.$$

Taking base 10 logs on both sides of the above equation, we obtain

$$\log_{10} l^\infty(K) = \log_{10} C + K\log_{10} q.$$

Let

$$Y_K = \log_{10} l^\infty(K),\ \beta_0 = \log_{10} C,\ \text{and } \beta_1 = \log_{10} q.$$

Then

$$Y_K = \beta_0 + \beta_1 K. \tag{4.4}$$

That is, Y_K and K have a linear relationship. We then use the R function lm() to compute the coefficients (β_0, β_1). Then

$$C = 10^{\beta_0} \text{ and } q = 10^{\beta_1}.$$

As in the polynomial error structure, we can compute and record $S^2_{r,\exp}$, $\hat{\sigma}_{\exp}$, and $R^2_{\exp}$ of the linear model to see whether the exponential error structure is good for the l^∞ error of our problem. Note that for the two error structures, $Y_K = \log_{10} l^\infty(K)$, and thus the sum of squares of the residuals ($S^2_{r,poly}$ and $S^2_{r,\exp}$) and the residual standard error ($\hat{\sigma}_{poly}$ and $\hat{\sigma}_{\exp}$) are comparable. Given K, the predictive $l^\infty(K)$ is

$$l^\infty(K) = Cq^K = 10^{\beta_0+\beta_1 K}. \tag{4.5}$$

If we want to know for what value of K the predictive $l^\infty(K)$ is less than a specified value δ, e.g., 0.001, we should let

$$l^\infty(K) = 10^{\beta_0+\beta_1 K} < \delta.$$

By elementary calculations, the above inequality reduces to

$$K > \frac{\log_{10}\delta - \beta_0}{\beta_1}. \tag{4.6}$$

Now we plot the figures and display the results. Take group 1 case 3 for example. See Figure 4. In Figure 4, the plot on the left of the first row is the cdf plot for various K. We see that as K increases the cdf increases and approaches to a limit, which is the infinite series given by (2.4), the line corresponding to $K = 150$ is very close to the limit and it is regarded as the true cdf. We also see that for the line corresponding to $K = 150$, it tends to 0 (1) as z tends to $-\infty$ (∞). The plot on the right of the first row is the error plot for various K, where the line corresponding to $K = 150$ is regarded as the true cdf and thus it is not shown in the plot. We see that as K increases, the error decreases to 0. For each K, the error is already 0 for the z values less than some negative constant. The plot on the left of the second row is the polynomial loglog error plot. It exemplifies a line given by (4.1). We see that the fitting of the line to the points is good. The red horizontal line corresponding to $\delta = 10^{-3}$. (4.3) gives $K > 104.2120$, which means that when $K \geq 105$, the $l^\infty(K)$ will be less than the prescribed $\delta = 10^{-3}$. The plot on the right of the second row is the polynomial error plot. It plots (4.2) for various K. We see that the predicted polynomial error fits the points well. The red horizontal line corresponding to $\delta = 10^{-3}$. The red line crosses the polynomial curve at $K = 104.2120$, which also means that when $K \geq 105$, the $l^\infty(K) < \delta = 10^{-3}$. The plot on the left of the third row is the exponential log error plot. It exemplifies a line given by (4.4). We see that the fitting of the line to the points is very good. The line fitting is much better than that for the polynomial loglog error plot. The red horizontal line corresponding to $\delta = 10^{-3}$. (4.6) gives $K > 106.5275$, which means that when $K \geq 107$, the $l^\infty(K) < \delta = 10^{-3}$. The plot on the right of the third row is the exponential error plot. It plots (4.5) for various K. We see that the predictive exponential error fits the points very well. The curve fitting is much better than that for the polynomial error plot, as reassured by the three statistics (S^2_r, $\hat{\sigma}$, and R^2) given later. The red horizontal line corresponding to $\delta = 10^{-3}$. The red line crosses the exponential curve at $K = 106.5275$, which also means that when $K \geq 107$, the $l^\infty(K) < \delta = 10^{-3}$.

The plots for other cases of group 1 are similar to Fig 1 and thus are omitted.

The P values, the K values corresponding to a $\delta = 10^{-3}$ error (K_{poly} and $K_{\exp}$), the sum of squares of the residuals ($S^2_{r,poly}$ and $S^2_{r,\exp}$), the residual standard error ($\hat{\sigma}_{poly}$ and $\hat{\sigma}_{\exp}$), the multiple R-squared (R^2_{poly} and $R^2_{\exp}$), the predictive l^∞ errors for $K = 150$ of the exponential error structure, and the q values of the exponential error structure for the 6 cases of group 1 are given in Table 2. From Table 2 we see that the P values are all greater than $0.9 >> 0.1 = \alpha$, therefore we accept $H_0 : \mu_X = \mu_Y$. The K values corresponding to a $\delta = 10^{-3}$ error for the two error structures are close. However, we should trust those K values calculated by the exponential error structure. The three statistics: The sum of squares of the residuals S^2_r (the smaller the better), the residual standard error $\hat{\sigma}$ (the smaller the better), and the multiple R-squared R^2 (the larger the better) for the exponential error structure are all better than those for the polynomial error structure. Therefore, we guess that the $l^\infty(K)$ decays exponentially for K. This exemplifies Remark 2.6. The

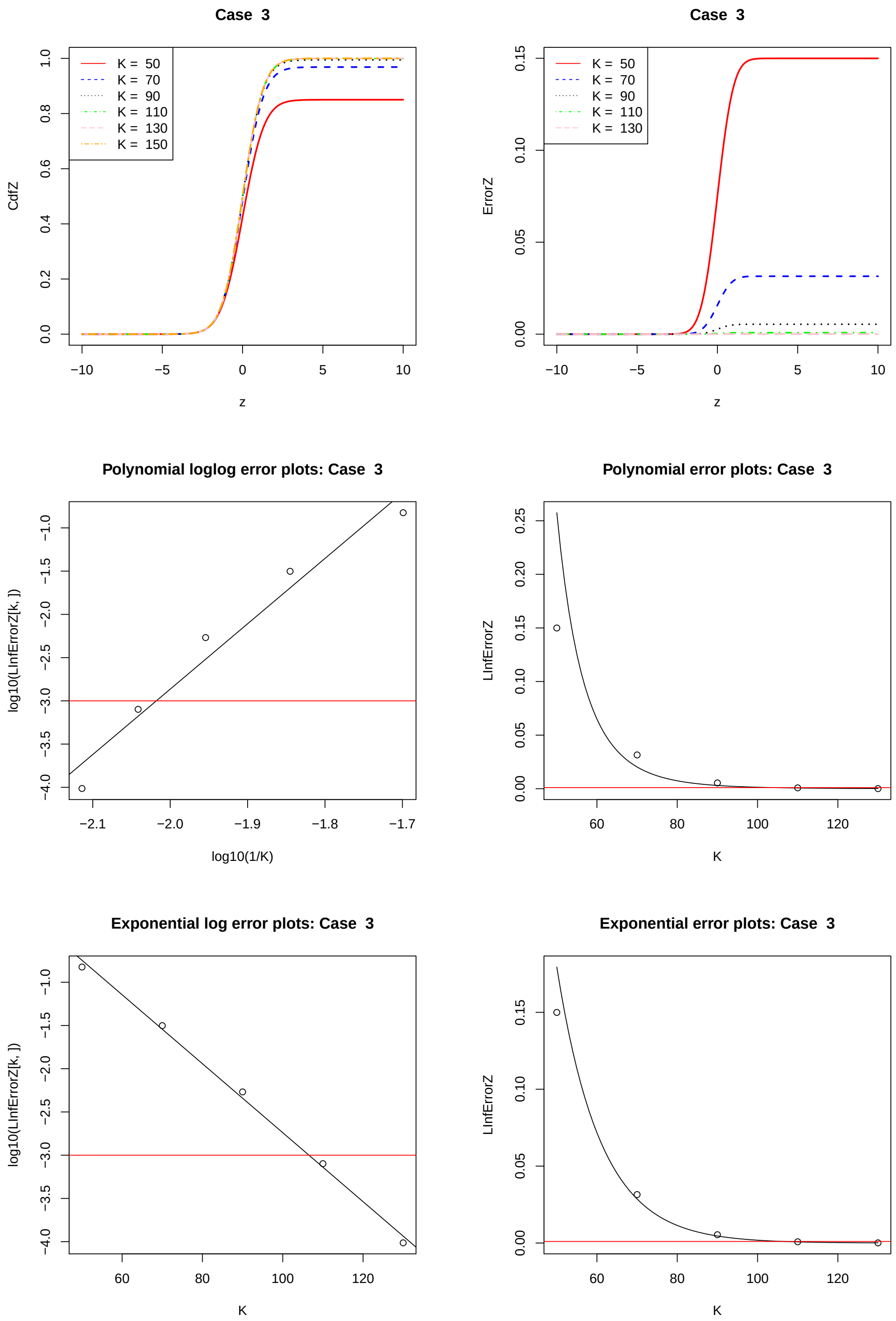

Fig. 1. The cdf plot, error plot, and l^∞ error plots for group 1 case 3

predictive l^∞ errors for $K = 150$ (which is regarded as the true solution) of the exponential error structure vary for different cases, from the most accurate case 6 (3.3×10^{-87}) to the least accurate case 3 (1.8×10^{-5}). Nevertheless, the predictive l^∞ errors for $K = 150$ is less than 2×10^{-5} for all 6 cases, and the solution corresponding to $K = 150$ can be reasonably regarded as the true solution. The q values of the exponential error structure are related to the K values corresponding to a $\delta = 10^{-3}$ error and the predictive l^∞ errors for $K = 150$. A small q corresponds to a small K and a small predictive l^∞ error, and vice versa.

Table 2. Statistics for 6 cases of group 1

	Polynomial	Exponential
P	(0.9807, 0.9402, 0.9446, 0.9037, 0.9444, 0.9036)	
K ($\delta = 10^{-3}$)	(25.9825, 52.4246, 104.2120, 13.6867, 59.8821, 6.4892)	(28.2943, 49.7654, 106.5275, 18.0592, 60.7933, 6.6181)
S_r^2	(1.7121, 1.0527, 0.2440, 3.9838, 0.6913, 5.5683)	(0.1101, 0.0186, 0.0208, 0.1402, 0.0183, 0.0645)
$\hat{\sigma}$	(0.7554, 0.5924, 0.2852, 1.1524, 0.4800, 1.3624)	(0.1916, 0.0787, 0.0833, 0.2162, 0.0782, 0.1467)
R^2	(0.8766, 0.9742, 0.9618, 0.8943, 0.9723, 0.9343)	(0.9921, 0.9995, 0.9967, 0.9963, 0.9993, 0.9992)
l^∞ (150)		(1.7e-19, 7.7e-14, 1.8e-05, 5.9e-32, 9.0e-11, 3.3e-87)
q		(0.7421, 0.7927, 0.9123, 0.6110, 0.8337, 0.2617)

The plots of the 6 cases of group 2 are similar to those of group 1 and thus are omitted.

The statistics for the 6 cases of group 2 are given in Table 3. From Table 3 we see that the P values are all smaller than $0.1 = \alpha$, therefore we reject $H_0 : \mu_X = \mu_Y$ and accept $H_1 : \mu_X \neq \mu_Y$, as expected. The explanations of the other statistics in Table 3 are similar to those in Table 2 and thus are omitted. Compare Table 2 and Table 3, we see that except the P values have been changed, the other statistics remain the same. Because we only change the Y samples from $Y_1^1, Y_2^1, \ldots, Y_m^1$ to $Y_1^2, Y_2^2, \ldots, Y_m^2$, it only changes the sample mean $\bar{Y}$, the sample variances remain unchanged due to set.seed(1). Therefore the R and $\hat{R} = S_X^2/S_Y^2$ values remain unchanged, and $\alpha_1, \beta_1, \alpha_2, \beta_2$ remain unchanged. Consequently, the cdf and its "derivatives" remain unchanged.

Table 3. Statistics for 6 cases of group 2

	Polynomial	Exponential
P	(0.0949, 0.0121, 0.0543, 0.0266, 0.0509, 0.0257)	
K ($\delta = 10^{-3}$)	(25.9825, 52.4246, 104.2120, 13.6867, 59.8821, 6.4892)	(28.2943, 49.7654, 106.5275, 18.0592, 60.7933, 6.6181)
S_r^2	(1.7121, 1.0527, 0.2440, 3.9838, 0.6913, 5.5683)	(0.1101, 0.0186, 0.0208, 0.1402, 0.0183, 0.0645)
$\hat{\sigma}$	(0.7554, 0.5924, 0.2852, 1.1524, 0.4800, 1.3624)	(0.1916, 0.0787, 0.0833, 0.2162, 0.0782, 0.1467)
R^2	(0.8766, 0.9742, 0.9618, 0.8943, 0.9723, 0.9343)	(0.9921, 0.9995, 0.9967, 0.9963, 0.9993, 0.9992)
l^∞ (150)		(1.7e-19, 7.7e-14, 1.8e-05, 5.9e-32, 9.0e-11, 3.3e-87)
q		(0.7421, 0.7927, 0.9123, 0.6110, 0.8337, 0.2617)

Finally, we give some discussions of the results obtained.

- The Z distribution is very flexible in the applications of the hypothesis testings of two normal means. One statistics that obeys the Z distribution is applicable to all assumptions of σ_X^2 and σ_Y^2.
- We carry out two groups of simulation studies in this section. The first group considers $\mu_X = 1$ and $\mu_Y = 1.1$. For the first group, we see that the P values are all greater than $0.9 >> 0.1 = \alpha$, and thus we accept $H_0 : \mu_X = \mu_Y$. The second group considers $\mu_X = 1$ and $\mu_Y = 2$. For the second group, we see that the P values are all smaller than $0.1 = \alpha$, and thus we reject $H_0 : \mu_X = \mu_Y$ and accept $H_1 : \mu_X \neq \mu_Y$, as expected.
- We assume two error structures for the l^∞ error: The polynomial error structure and the exponential error structure. The simulation studies exemplify that the exponential error structure is the better error structure for the l^∞ error.

5 Conclusions

By introducing the Z distribution, we provide a total new way to tackle the Behrens-Fisher problem. First, we define the SIG distribution and find its pdf, which is a generalization of the gamma distribution. Three remarks considering the properties of the SIG distribution are given. Second, we define the Z distribution and find its pdf, which is a generalization of the t distribution. Third, we find the cdf of the Z distribution in Theorem 2.1. Three remarks considering the applicability, the $l^{\infty}(K)$ error structure, and the programming of the cdf of the Z distribution are then given. Note that the pdf of the SIG distribution, the pdf and cdf of the Z distribution are all infinite series. Fourth, we apply the Z distribution in the hypothesis testings of two normal means. In three different assumptions of the variances, namely, σ_X^2 and σ_Y^2 are known, $\sigma_X^2 = \sigma_Y^2 = \sigma^2$ are unknown, and $\sigma_X^2 \neq \sigma_Y^2$ are unknown (the Behrens-Fisher problem), we find the pivots using the Z distribution. Remark 3.1 shows that the Z distribution is very flexible in the applications of the hypothesis testings of two normal means. One statistics that obeys the Z distribution is applicable to all assumptions of σ_X^2 and σ_Y^2. Remark 3.2 shows that we can define the Z distribution differently, such that the resulting Z distribution has a simple form. Moreover, the $\tilde{T}$ statistics (3.4) given in [15] can be obtained through the Z distribution. Finally, we provide the simulation studies for the hypothesis testing problems in Remark 3.1. Two groups of simulation studies are considered. The first group considers $\mu_X = 1$ and $\mu_Y = 1.1$. For the first group, we see that the P values are all greater than $0.9 >> 0.1 = \alpha$, and thus we accept $H_0 : \mu_X = \mu_Y$. The second group considers $\mu_X = 1$ and $\mu_Y = 2$. For the second group, we see that the P values are all smaller than $0.1 = \alpha$, and thus we reject $H_0 : \mu_X = \mu_Y$ and accept $H_1 : \mu_X \neq \mu_Y$, as expected. In the simulation studies, we assume two error structures for the l^{∞} error: The polynomial error structure and the exponential error structure. The latter one is the better error structure as the simulation studies exemplify.

Acknowledgement

The authors gratefully acknowledge the constructive comments offered by the referees. Their comments improve the quality of the paper significantly. The research was supported by the Fundamental Research Funds for the Central Universities (CQDXWL-2012-004 and CDJRC10100010), China Scholarship Council (201606055028), and the MOE project of Humanities and Social Sciences on the west and the border area (14XJC910001).

Competing Interests

Authors have declared that no competing interests exist.

References

[1] Wikipedia. Behrens-Fisher problem [OL]; 2017.

[2] Behrens WU. Ein beitrag zur fehlerberechnung bei wenigen beobachtungen (translation: A contribution to error estimation with few observations) [J]. Landwirtschaftliche Jahrbucher. 1929;68:807-837.

[3] Fisher RA. The fiducial argument in statistical inference [J]. Annals of Human Genetics. 1935;6(4):391-398.

[4] Welch BL. The significance of the difference between two means when the population variances are unequal [J]. Biometrika. 1938;29:350-362.

[5] Fisher RA. The asymptotic approach to Behrens's integral, with further tables for the d test of significance [J]. Annals of Human Genetics. 1941;11:141-172.

[6] Scheffe H. On solutions of the Behrens-Fisher problem, based on the t-distribution [J]. The Annals of Mathematical Statistics. 1943;14:35-44.

[7] Satterthwaite FE. An approximate distribution of estimates of variance components [J]. Biometrics Bulletin (now called Biometrics). 1946;2:110-114.

[8] Scheffe H. Practical solutions of the Behrens-Fisher problem [J]. Journal of the American Statistical Association. 1970;65:1501-1508.

[9] Davenport JM and Webster JT. The Behrens-Fisher problem, an old solution revisited [J]. Metrika. 1975;22(1):47-54.

[10] Lee AFS and Gurland J. Size and power of tests for equality of means of two normal populations with unequal variances [J]. Journal of the American Statistical Association. 1975;70:933-941.

[11] Robinson GK. Properties of student's t and of the Behrens-Fisher solution to the two means problem [J]. Annals of Statistics. 1976;4:963-971.

[12] Barnard GA. Comparing the means of two independent samples [J]. Journal of the Royal Statistical Society. Series C (Applied Statistics). 1984;33:266-271.

[13] Tsui KW and Weerahandi S. Generalized p values in significance testing of hypotheses in the presence of nuisance parameters [J]. Journal of the American Statistical Association. 1989;84:602-607.

[14] Bhoj DS. An approximate solution to the Behrens-Fisher problem [J]. Biometrical Journal. 1993;35(5):635-640.

[15] Sprott DA and Farewell VT. The difference between two normal means [J]. American Statistician. 1993;47:126-128.

[16] Meng XL. Posterior predictive p-values [J]. Annals of Statistics. 1994;22:1142-1160.

[17] Dudewicz EJ and Ahmed SU. New exact and asymptotically optimal solution to the Behrens-Fisher problem, with tables [J]. American Journal of Mathematical and Management Sciences. 1998;18:359-426.

[18] Kim SH and Cohen AS. On the Behrens-Fisher problem: A review [J]. Journal of Educational and Behavioral Statistics. 1998;23:356-377.

[19] Dudewicz EJ and Ahmed SU. New exact and asymptotically optimal heteroscedastic statistical procedures and tables, II [J]. American Journal of Mathematical and Management Sciences. 1999;19:157-187.

[20] Ruben H. A simple conservative and robust solution of the Behrens-Fisher problem [J]. Sankhya, Series A. 2002;64(1):139-155.

[21] Sawilowsky S. S. Fermat, Schubert Einstein, Behrens-Fisher. The probable difference between two means when $\sigma_1^2 \neq \sigma_2^2$ [J]. Journal of Modern Applied Statistical Methods. 2002;1(2):461-472.

[22] Witkovsky V. On the Behrens-Fisher distribution and its generalization to the pairwise comparisons [J]. Discussiones Mathematicae: Probability and Statistics. 2002;22(1-2):73-104.

[23] Reed JF. Solutions to the Behrens-Fisher problem [J]. Computer Methods and Programs in Biomedicine. 2003;70:259-263.

[24] Dong LB. The Behrens-Fisher problem: An empirical likelihood approach [R]. Technical report, Econometrics Working Paper EWP0404, University of Victoria; 2004.

[25] Huang P, Tilley BC, Woolson RF, Lipsitz S. Adjusting O'Brien's test to control type I error for the generalized nonparametric Behrens-Fisher problem [J]. Biometrics. 2005;61:532-539.

[26] Krishnamoorthy K, Xia Y. On selecting tests for equality of two normal mean vectors [J]. Multivariate Behavioral Research. 2006;41:533-548.

[27] Robinson GK. Encyclopedia of statistical sciences, chapter Behrens-Fisher Problem-I [G]. Wiley, New York. 2006;1.
DOI: 10.1002/0471667196.ess0105.pub2

[28] Dudewicz EJ, Ma Y, Mai SE, Su H. Exact solutions to the Behrens-Fisher problem: Asymptotically optimal and finite sample efficient choice among [J]. Journal of Statistical Planning and Inference. 2007;137(5):1584-1605.

[29] Pardo JA, Pardo MDC. A simulation study of a new family of test statistics for the Behrens-Fisher problem [J]. Kybernetes. 2007;36(5-6):806-816.

[30] Belloni A, Didier G. On the Behrens-Fisher problem: A globally convergent algorithm and a finite-sample study of the Wald, LR and LM tests [J]. Annals of Statistics. 2008;36(5):2377-2408.

[31] Chang CH and Pal N. A revisit to the Behrens-Fisher problem: Comparison of five test methods [J]. Communications in Statistics - Simulation and Computation. 2008;37(6):1064-1085.

[32] Drton M. Multiple solutions to the likelihood equations in the Behrens-Fisher problem [J]. Statistics & Probability Letters. 2008;78(18):3288-3293.

[33] Fraser DAS, Rousseau J. Studentization and deriving accurate p-values [J]. Biometrika. 2008;95(1):1-16.

[34] Liu XQ, Jiang HY, Rong JY. A new method for the Behrens-Fisher problem [R]; 2008.

[35] Zheng SR, Shi NZ, Ma WQ. Statistical inference on difference or ratio of means from heteroscedastic normal populations [J]. Journal of Statistical Planning and Inference. 2010;140(5):1236-1242.

[36] Walker GA, Saw JG. The distribution of linear combinations of t-variables [J]. Journal of the American Statistical Association. 1978;73:876-878.

[37] Witkovsky V. On the exact computation of the density and of the quantiles of linear combinations of t and f random variables [J]. Journal of Statistical Planning and Inference. 2001;94:1-13.

[38] Witkovsky V. Computing the distribution of a linear combination of inverted gamma variables [J]. Kybernetika. 2001;37:79-90.

[39] Witkovsky V. Exact distribution of positive linear combinations of inverted chi-square random variables with odd degrees of freedom [J]. Statistics & Probability Letters. 2002;56(1):45-50.

[40] Welch BL. The generalization of 'student's' problem when several different population variances are involved [J]. Biometrika. 1947;34:28-35.

[41] Dunnett CW. Pairwise multiple comparisons in the unequal variance case [J]. Journal of the American Statistical Association. 1980;75:796-800.

[42] Fligner MA, Rust SW. A modification of mood's median test for the generalized Behrens-Fisher problem [J]. Biometrika. 1982;69:221-226.

[43] Chen YI, Wolfe DA. Modifications of the mack-wolfe umbrella tests for a generalized Behrens-Fisher problem [J]. The Canadian Journal of Statistics / La Revue Canadienne de Statistique. 1990;18:245-253.

[44] Nikulin MS, Voinov VG. On the problem of the means of weighted normal populations [J]. Questiio. 1995;19(1-2-3):93-106.

[45] Kim SJ. A practical solution to the multivariate Behrens-Fisher problem [J]. Biometrika. 1992;79:171-176.

[46] Lix LM, Keselman HJ, Hinds AM. Robust tests for the multivariate Behrens-Fisher problem [J]. Computer Methods and Programs in Biomedicine. 2005;77:129-139.

[47] Yanagihara H, Yuan KH. Three approximate solutions to the multivariate Behrens-Fisher problem [J]. Communications in Statistics - Simulation and Computation. 2005;34:975-988.

[48] Casella G, Berger RL. Statistical Inference [M]. Duxbury, USA. 2nd edition; 2002.

Peer-review history:
The peer review history for this paper can be accessed here (Please copy paste the total link in your browser address bar)
http://sciencedomain.org/review-history/18234