Journal of Statistical Research *ISSN 0256 - 422 X*
2008, Vol. 42, No. 1, pp. 17-35
Bangladesh

LIKELIHOOD ANALYSIS FOR THE DIFFERENCE IN MEANS OF TWO INDEPENDENT NORMAL DISTRIBUTIONS WITH ONE VARIANCE UNKNOWN

A. C. M. Wong

Department of Mathematics and Statistics, York University
4700 Keele Street, Toronto, Ontario, Canada. M3J 1P3

Email: august@mathstat.yorku.ca

Y. Y. Wu

Department of Mathematics and Statistics, York University
4700 Keele Street, Toronto, Ontario, Canada. M3J 1P3

Email: yanyanwu@live.ca

SUMMARY

Maity and Sherman (2006) considered the situation in two-sample testing for the difference in means when one variance is assumed to be known while the other variance is treated as unknown. This problem arises in many real life situations, for example, when one is interested in comparing a standard treatment with a new treatment in medical studies. The variance for the standard treatment is assumed to be known from historical data, and the variance for the new treatment is unknown. Following the argument in Satterthwaite (1941, 1946), Maity and Sherman (2006) obtained the confidence interval for the difference in means based on an approximate t-distribution. In this paper, a likelihood-based third order asymptotic method is introduced to obtain the confidence intervals for the difference in means. Simulations are used to show that the proposed method has better coverage property than Maity and Sherman's t-method, especially when the sample sizes are small.

Keywords and phrases: Behrens-Fisher problem; Canonical parameter; Modified signed log-likelihood ratio statistics; Satterthwaite.

AMS Classification: Place Classification here. Leave as is, if there is no classification

1 Introduction

Let $X_1, \ldots, X_m \sim \text{Normal}(\mu_1, \sigma_1^2)$ and $Y_1, \ldots, Y_n \sim \text{Normal}(\mu_2, \sigma_2^2)$ where the notation $\sim$ means *independent and identically distributed as*. Moreover, the two populations are assumed to be independent. The problem of obtaining statistical inference for the difference in means, $\psi = \mu_1 - \mu_2$, is a mainstay in statistical practice and is introduced in most introductory courses. When the two variances are known, this leads directly to a test statistic

$$\frac{\bar{X} - \bar{Y} - \psi}{\sqrt{\sigma_1^2/m + \sigma_2^2/n}}$$

with a standard normal distribution, where

$$\bar{X} = \frac{\sum_{i=1}^m X_i}{m} \qquad \text{and} \qquad \bar{Y} = \frac{\sum_{j=1}^n Y_j}{n}.$$

When the two variances are unknown but equal, the two-sample test uses a pooled estimate of variance and the resulting test statistic

$$\frac{\bar{X} - \bar{Y} - \psi}{\sqrt{S_p^2 \left(\frac{1}{m} + \frac{1}{n}\right)}}$$

has an exact t-distribution with $(m + n - 2)$ degrees of freedom, where

$$S_p^2 = \frac{(m-1)S_1^2 + (n-1)S_2^2}{m+n-2}, \; S_1^2 = \frac{\sum_{i=1}^m (X_i - \bar{X})^2}{m-1}, \; S_2^2 = \frac{\sum_{j=1}^n (Y_j - \bar{Y})^2}{n-1}.$$

In the case where the two variances are unknown and unequal, we have the Behrens-Fisher problem. This problem has received considerable attention in statistical literatures because the resulting test statistic does not have an exact t-distribution. Many authors have studied this problem and various solutions have been proposed. See Scheffé (1970) for more background on this problem.

Maity and Sherman (2006) considered the case when one of the variances is assumed to be known and the other is treated as unknown. This problem is rarely discussed in statistical literature, however it arises naturally in both biostatistics and engineering studies. For example, in clinical trials, comparing a standard treatment to a new one, the variance for the standard treatment may be considered to be known from historical data, while the variance for the new treatment may not be the same as the old one. Similarly, in engineering statistics, it happens when the production process changes.

In Section 2, we briefly review the method given in Maity and Sherman (2006). In Section 3, a general likelihood-based third order method is proposed. In Section 4, the proposed method is applied to the problem discussed in Maity and Sherman (2006) and in Section 5, we compare results obtained by Maity and Sherman (2006) and our proposed method in a real-life example and in simulation studies.

2 Result from Maity and Sherman

Consider the Behrens-Fisher problem. It is well known that the test statistic:

$$T_1 = \frac{\bar{X} - \bar{Y} - \psi}{\sqrt{\frac{S_1^2}{m} + \frac{S_2^2}{n}}}$$

does not have an exact t-distribution because

$$\left(\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}\right)^{-1} \left(\frac{S_1^2}{m} + \frac{S_2^2}{n}\right)$$

does not have an appropriate χ^2 distribution. It is shown in Satterthwaite (1941,1946) that

$$\gamma_1 \left(\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}\right)^{-1} \left(\frac{S_1^2}{m} + \frac{S_2^2}{n}\right) \tag{2.1}$$

is approximately distributed as χ^2 distribution with γ_1 degrees of freedom. The value of γ_1 is approximated by matching the variances of (2.1) and the $\chi^2_{\gamma_1}$ distribution. Hence

$$\widehat{\gamma}_1 = \left(\frac{S_1^2}{m} + \frac{S_2^2}{n}\right)^2 \left(\frac{\left(S_1^2/m\right)^2}{m-1} + \frac{\left(S_2^2/n\right)^2}{n-1}\right)^{-1}.$$

Thus T_1 is approximately distributed as t-distribution with $\widehat{\gamma}_1$ degrees of freedom.

Maity and Sherman (2006) considered the case when one variance, σ_1^2, is known and the other variance, σ_2^2, is unknown. The test statistics is:

$$T_2 = \frac{\bar{X} - \bar{Y} - \psi}{\sqrt{\frac{\sigma_1^2}{m} + \frac{S_2^2}{n}}}. \tag{2.2}$$

However, T_2 does not have an exact t-distribution because

$$\left(\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}\right)^{-1} \left(\frac{\sigma_1^2}{m} + \frac{S_2^2}{n}\right)$$

does not have the appropriate χ^2 distribution. Following the argument in Satterthwaite (1941, 1946), Maity and Sherman (2006) approximated the distribution of T_2 by a t-distribution with $\widehat{\gamma}_2$ degrees of freedom where

$$\widehat{\gamma}_2 = \left(\frac{\sigma_1^2}{m} + \frac{S_2^2}{n}\right)^2 \left(\frac{\left(S_2^2/n\right)^2}{n-1}\right)^{-1}.$$

Hence, an approximate $100(1-\alpha)\%$ confidence interval for ψ is:

$$(\bar{X} - \bar{Y}) \pm t_{\alpha/2,\hat{\gamma}_2}\sqrt{\frac{\sigma_1^2}{m} + \frac{S_2^2}{n}} \tag{2.3}$$

where $t_{\alpha/2,\hat{\gamma}_2}$ is the $(1-\alpha/2)100$th percentile of the t-distribution with $\hat{\gamma}_2$ degrees of freedom.

3 Proposed Third Order Likelihood-Based Method

Consider a sample $y = (y_1, \ldots, y_n)$ from a statistical model with log-likelihood function $\ell(\theta) = \ell(\theta; y)$, where θ is a vector of parameters with length p. Let $\psi = \psi(\theta)$ be the scalar parameter of interest.

From the standard likelihood theories, we denote $\hat{\theta}$ as the overall maximum likelihood estimator of θ, which is obtained by solving $\frac{\partial \ell(\theta)}{\partial \theta}|_{\hat{\theta}} = 0$, and $j_{\theta\theta'}(\hat{\theta}) = -\left.\frac{\partial^2 \ell(\theta)}{\partial\theta\partial\theta'}\right|_{\hat{\theta}}$, which is the observed information matrix evaluated at $\hat{\theta}$. Moreover, let $\hat{\theta}_\psi$ be the constrained maximum likelihood estimator of θ for a given $\psi(\theta) = \psi$, which can be obtained by maximizing $\ell(\theta)$ subject to the constraint $\psi(\theta) = \psi$. Lagrange multiplier technique can be applied to solve the constrained maximization problem. In other words, all we need is to maximize

$$H(\theta, \lambda) = \ell(\theta) + \lambda[\psi(\theta) - \psi].$$

Then $\hat{\theta}_\psi$ and $\tilde{\lambda}$ must satisfy

$$\left.\frac{\partial H(\theta,\lambda)}{\partial \theta}\right|_{(\hat{\theta}_\psi, \tilde{\lambda})} = 0 \qquad \text{and} \qquad \left.\frac{\partial H(\theta,\lambda)}{\partial \lambda}\right|_{(\hat{\theta}_\psi, \tilde{\lambda})} = 0.$$

We define the tilted log-likelihood function is $\tilde{\ell}(\theta) = \ell(\theta) + \tilde{\lambda}[\psi(\theta) - \psi]$. Note that $\tilde{\ell}(\hat{\theta}_\psi) = \ell(\hat{\theta}_\psi)$. And $\tilde{j}_{\theta\theta'}(\hat{\theta}_\psi) = -\left.\frac{\partial^2 \tilde{\ell}(\theta)}{\partial\theta\partial\theta'}\right|_{\hat{\theta}_\psi}$ is the observed information matrix evaluated at $\hat{\theta}_\psi$ based on $\tilde{\ell}(\theta)$.

With the regularity conditions stated in Cox and Hinkley (1974), the signed log-likelihood ratio statistics,

$$\begin{aligned} r \equiv r(\psi) &= sgn(\hat{\psi} - \psi)\left\{2\left[\ell(\hat{\theta}) - \ell(\hat{\theta}_\psi)\right]\right\}^{1/2} \\ &= sgn(\hat{\psi} - \psi)\left\{2\left[\ell(\hat{\theta}) - \tilde{\ell}(\hat{\theta}_\psi)\right]\right\}^{1/2} \end{aligned} \tag{3.1}$$

is asymptotically distributed as the standard normal distribution. Hence a $100(1-\alpha)\%$ confidence interval for ψ based on $r(\psi)$ is

$$\{\psi : |r(\psi)| \leq z_{\alpha/2}\}$$

where $z_{\alpha/2}$ is the $(1-\alpha/2)100$th percentile of the standard normal distribution. Note that, this method has accuracy $O(n^{-1/2})$.

In recent years, various adjustments to $r(\psi)$ have been proposed to improve the accuracy of the signed log-likelihood ratio method. In this paper we consider the modified signed log-likelihood ratio statistic, r^*, introduced by Barndorff-Nielsen (1986,1991), and it has the form

$$r^* \equiv r^*(\psi) = r(\psi) + \frac{1}{r(\psi)} \log\left\{\frac{Q(\psi)}{r(\psi)}\right\} \tag{3.2}$$

where $r(\psi)$ is the signed log-likelihood ratio statistic as defined in (3.1), and $Q(\psi)$ is a standardized maximum likelihood departure in an appropriate parameter scaling. It is shown in Barndorff-Nielsen (1986,1991) that $r^*(\psi)$ is asymptotically distribution as the standard normal distribution with accuracy $O(n^{-3/2})$. Hence a $100(1-\alpha)\%$ confidence interval for ψ based on $r^*(\psi)$ is

$$\{\psi : |r^*(\psi)| \leq z_{\alpha/2}\}.$$

Fraser and Reid (1995) showed that the appropriate parameterization for $Q(\psi)$ is the canonical parameterization of the exponential family model. Let $\varphi(\theta)$ be the canonical parameter of the exponential family model. Then

$$\chi(\theta) = \psi_\theta(\hat{\theta}_\psi)\varphi_\theta^{-1}(\hat{\theta}_\psi)\varphi(\theta)$$

is the parameter of interest recalibrated in the $\varphi(\theta)$ scale and $\varphi_\theta(\theta) = \partial\varphi(\theta)/\partial\theta$. Furthermore, by the chain rule in differentiation, the determinant of the observed information matrix obtained from $\ell(\theta)$ evaluated at $\hat{\theta}$ and the determinant of the observed information obtained from the tilted likelihood evaluated at $\hat{\theta}_\psi$ in the $\varphi(\theta)$ scale are:

$$\begin{aligned} |j_{(\theta\theta')}(\hat{\theta})| &= |j_{\theta\theta'}(\hat{\theta})||\varphi_\theta(\hat{\theta})|^{-2} \\ |\tilde{j}_{(\theta\theta')}(\hat{\theta}_\psi)| &= |\tilde{j}_{\theta\theta'}(\hat{\theta}_\psi)||\varphi_\theta(\hat{\theta}_\psi)|^{-2} \end{aligned}$$

respectively. Hence an estimate of the asymptotic variance of $\chi(\hat{\theta})$ is

$$\widehat{var}(\chi(\hat{\theta})) = \frac{\psi_\theta(\hat{\theta}_\psi)\tilde{j}_{\theta\theta'}^{-1}(\hat{\theta}_\psi)\psi'_\theta(\hat{\theta}_\psi)|\tilde{j}_{(\theta\theta')}(\hat{\theta}_\psi)|}{|j_{(\theta\theta')}(\hat{\theta})|}.$$

Thus $Q(\psi)$ expressed in $\varphi(\theta)$ scale is

$$Q(\psi) = sgn(\hat{\psi} - \psi)\frac{|\chi(\hat{\theta}) - \chi(\hat{\theta}_\psi)|}{\sqrt{\widehat{var}(\chi(\hat{\theta}))}}. \tag{3.3}$$

Fraser and Reid (1995) provided a general methodology to obtain $Q(\psi)$ when $\varphi(\theta)$ is not available explicitly. This methodology is summarized in Appendix A.

Note that to obtain confidence interval for ψ from $r(\psi)$ or $r^*(\psi)$, numerical methods are required.

4 The Two-Sample Test with One Variance Unknown Problem

Let $(x_1, \ldots, x_m)$ and $(y_1, \ldots, y_n)$ be the observed data for the model discussed by Maity and Sherman (2006). The log-likelihood function can be written as

$$\ell(\theta) = -\frac{1}{2\sigma_1^2}\sum_{i=1}^{m}(x_i - \mu_1)^2 - \frac{n}{2}\log\sigma_2^2 - \frac{1}{2\sigma_2^2}\sum_{j=1}^{n}(y_j - \mu_2)^2. \tag{4.1}$$

The overall maximum likelihood estimator of θ, $\hat{\theta}$, and the observed information matrix evaluated at $\hat{\theta}$ are

$$\hat{\theta} = (\hat{\mu}_1, \hat{\mu}_2, \hat{\sigma}_2^2)' = \left(\bar{x}, \bar{y}, \frac{1}{n} \sum_{j=1}^{n} (y_j - \bar{y})^2 \right)'$$

and

$$j_{\theta\theta'}(\hat{\theta}) = \begin{pmatrix} \frac{m}{\sigma_1^2} & 0 & 0 \\ 0 & \frac{n}{\hat{\sigma}_2^2} & 0 \\ 0 & 0 & \frac{n}{2\hat{\sigma}_2^4} \end{pmatrix}.$$

To obtain the constrained maximum likelihood estimator for a given ψ, $\hat{\theta}_\psi$, we need to maximize $\ell(\theta)$ subject to $\psi(\theta) = \psi$. By applying the Lagrange multiplier technique, $(\hat{\theta}_\psi, \tilde{\lambda}) = (\tilde{\mu}_1, \tilde{\mu}_2, \tilde{\sigma}_2^2, \tilde{\lambda})$ can be obtained by solving

$$\begin{aligned} \frac{1}{\sigma_1^2} \sum_{i=1}^{m} (x_i - \tilde{\mu}_1) + \tilde{\lambda} &= 0 \\ \frac{1}{\tilde{\sigma}_2^2} \sum_{j=1}^{n} (y_j - \tilde{\mu}_2) - \tilde{\lambda} &= 0 \\ -\frac{n}{2\tilde{\sigma}_2^2} + \frac{1}{2\tilde{\sigma}_2^4} \sum_{j=1}^{n} (y_j - \tilde{\mu}_2)^2 &= 0 \\ \tilde{\mu}_1 - \tilde{\mu}_2 - \psi &= 0 \end{aligned}$$

simultaneously. Note that the explicit form of the solution for $\hat{\theta}_\psi$ is not available, but it can be obtained numerically. Hence the tilted log-likelihood function can be obtained and the observed information matrix evaluated at $\hat{\theta}_\psi$ from $\tilde{\ell}(\theta)$ is

$$\tilde{j}_{\theta\theta'}(\hat{\theta}_\psi) = \begin{pmatrix} \frac{m}{\sigma_1^2} & 0 & 0 \\ 0 & \frac{n}{\tilde{\sigma}_2^2} & \frac{n}{\tilde{\sigma}_2^4}(\bar{y} - \tilde{\mu}_2) \\ 0 & \frac{n}{\tilde{\sigma}_2^4}(\bar{y} - \tilde{\mu}_2) & \frac{n}{2\tilde{\sigma}_2^4} \end{pmatrix}.$$

Thus with $r(\psi)$ as defined in (3.1), the $100(1-\alpha)\%$ confidence interval for ψ can be obtained based on $r(\psi)$.

From the log-likelihood function in (4.1), the canonical parameter, $\varphi(\theta) = \left(\mu_1, \frac{1}{\sigma_2^2}, \frac{\mu_2}{\sigma_2^2}\right)'$, is explicitly available. Hence

$$\varphi_\theta(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & -\frac{1}{\sigma_2^4} \\ 0 & \frac{1}{\sigma_2^2} & -\frac{\mu_2}{\sigma_2^4} \end{pmatrix}.$$

Since $\psi(\theta) = \mu_1 - \mu_2$, we have $\psi_\theta(\theta) = (1 \;\; -1 \quad 0)$. Therefore the recalibrated parameter of interest and its variance in the $\varphi(\theta)$ scale are

$$\begin{aligned} \chi(\theta) &= (\bar{x} - \tilde{\mu}_1) + \frac{\tilde{\sigma}_2^2}{\hat{\sigma}_2^2}(\tilde{\mu}_2 - \bar{y}) \\ \widehat{var}(\chi(\hat{\theta})) &= \frac{\tilde{\sigma}_2^6}{\hat{\sigma}_2^6}\left(\frac{\sigma_1^2}{m} + \frac{\tilde{\sigma}_2^2}{n} - \frac{\sigma_1^2}{m}\frac{2(\bar{y} - \tilde{\mu}_2)^2}{\tilde{\sigma}_2^2}\right). \end{aligned}$$

Then $Q(\psi)$, as defined in (3.3), can be written as

$$Q(\psi) = sgn(\bar{x} - \bar{y} - \psi)\left|(\bar{x} - \tilde{\mu}_1) + \frac{\tilde{\sigma}_2^2}{\hat{\sigma}_2^2}(\tilde{\mu}_2 - \bar{y})\right|\frac{\hat{\sigma}_2^3}{\tilde{\sigma}_2^3}\left(\frac{\sigma_1^2}{m} + \frac{\tilde{\sigma}_2^2}{n} - \frac{\sigma_1^2}{m}\frac{2(\bar{y} - \tilde{\mu}_2)^2}{\tilde{\sigma}_2^2}\right)^{-\frac{1}{2}}.$$

Thus with $r^*(\psi)$, as defined in (3.2), the $100(1-\alpha)\%$ confidence interval for ψ can be obtained based on $r^*(\psi)$.

5 Examples

Polymer is manufactured in a batch chemical process. Viscosity measurements are normally made on each batch, and long experience with the process has indicated that the variability of the process is fairly stable with $\sigma_1 = 20$. A sample of fifteen batch viscosity measurements are given as follows:

$$724, 718, 776, 760, 745, 759, 795, 756, 742, 740, 761, 749, 739, 747, 742.$$

A process change is made which involves switching the type of catalyst involved in the process. Following the process change, a sample of eight batch viscosity measurements are taken:

$$735, 775, 729, 755, 783, 760, 738, 780.$$

Maity and Sherman (2006) analysed this data set and reported the 90% confidence interval for ψ, the difference in mean batch viscosity resulting from the process change, to be $(-22.6638, 9.3138)$. Applying the method discussed in Section 4, the 90% confidence interval for ψ obtained based on the signed log-likelihood ratio statistic $r(\psi)$ and the modified signed log-likelihood ratio statistic $r^*(\psi)$ are $(-21.6239, 8.2739)$ and $(-22.7046, 9.3546)$ respectively.

It is important to note that Maity and Sherman's t-method of obtaining the confidence interval for the difference of means has a closed form and is easy to calculate. The proposed method requires numerical methods for obtaining the confidence interval for the difference of means. However the implementation of the numerical method is relatively easy. To illustrate the simplicity of implementing the proposed method into standard statistical softwares, a R source code which produced the 90% confidence intervals for the above example is given in Appendix B.

Since both $r(\psi)$ and $r^*(\psi)$ are asymptotically distributed as the standard normal distribution, we can obtain the probability to the left of the data point for various ψ values, $\Phi(r(\psi))$ and $\Phi(r^*(\psi))$ respectively where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal distribution. Figure 1 plots the left tail probability for ψ based on Maity and Sherman's t-method, signed log-likelihood ratio method and modified signed log-likelihood ratio method. The results from Maity and Sherman's t-method, and the proposed method are almost indistinguishable, however results from the signed log-likelihood ratio statistic method are different from the other two methods.

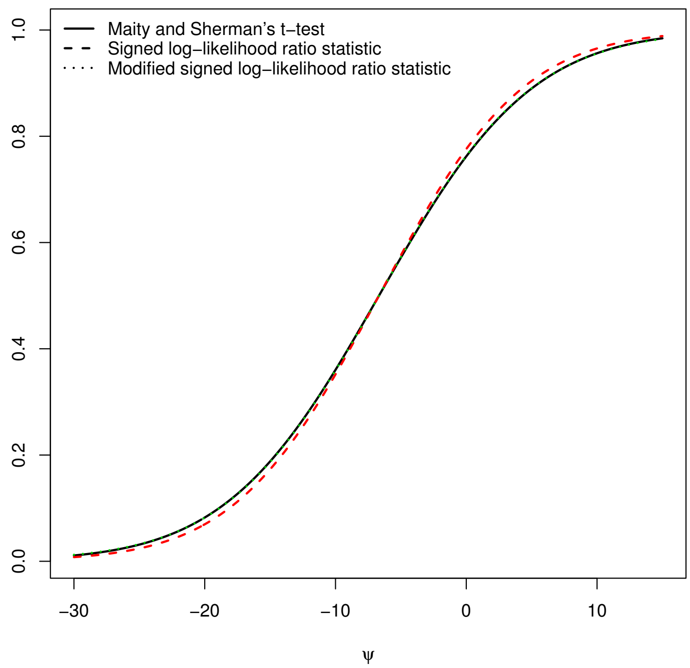

Figure 1: Tail probability curves

In order for us to have a better understanding on the difference of accuracy among the three methods, simulation studies were conducted. Table 1 gives the sample sizes and the parameter configurations that we will consider in our studies. For each design, we have generated 10,000 samples from two independent normal distributions. For each sample, we calculated the 95% confidence interval for $\psi = \mu_1 - \mu_2$ with equal tail probabilities from the three methods examined in this paper. Table 2 records the proportion of intervals that contains the true ψ (coverage probability), the proportion of true ψ that falls outside the lower bound of the confidence interval (lower error probability) and the proportion of true ψ that falls outside the upper bound of the confidence interval (upper error probability). The theoretic values for the coverage probability, and the lower and upper errors probabilities are $0.95, 0.025$, and 0.025 respectively.

From Table 2, it is evidence that the signed log-likelihood ratio method gives the worst coverage probability. Both the Maity and Sherman's t-method and the proposed method give good results when the sample sizes are moderately large. However, when the sample sizes are small, the proposed method out-performed the other methods.

Table 1: Sample sizes and parameter configurations for the simulation study

Design	m	n	μ_1	σ_1^2	μ_2	σ_2^2	Design	m	n	μ_1	σ_1^2	μ_2	σ_2^2
1	1	2	2	4	4	16	25	8	2	2	4	4	16
2	1	2	2	16	4	4	26	8	2	2	16	4	4
3	1	2	4	4	2	16	27	8	2	4	4	2	16
4	1	2	4	16	2	4	28	8	2	4	16	2	4
5	1	8	2	4	4	16	29	8	8	2	4	4	16
6	1	8	2	16	4	4	30	8	8	2	16	4	4
7	1	8	4	4	2	16	31	8	8	4	4	2	16
8	1	8	4	16	2	4	32	8	8	4	16	2	4
9	1	15	2	4	4	16	33	8	15	2	4	4	16
10	1	15	2	16	4	4	34	8	15	2	16	4	4
11	1	15	4	4	2	16	35	8	15	4	4	2	16
12	1	15	4	16	2	4	36	8	15	4	16	2	4
13	2	2	2	4	4	16	37	15	2	2	4	4	16
14	2	2	2	16	4	4	38	15	2	2	16	4	4
15	2	2	4	4	2	16	39	15	2	4	4	2	16
16	2	2	4	16	2	4	40	15	2	4	16	2	4
17	2	8	2	4	4	16	41	15	8	2	4	4	16
18	2	8	2	16	4	4	42	15	8	2	16	4	4
19	2	8	4	4	2	16	43	15	8	4	4	2	16
20	2	8	4	16	2	4	44	15	8	4	16	2	4
21	2	15	2	4	4	16	45	15	15	2	4	4	16
22	2	15	2	16	4	4	46	15	15	2	16	4	4
23	2	15	4	4	2	16	47	15	15	4	4	2	16
24	2	15	4	16	2	4	48	15	15	4	16	2	4

Table 2: Coverage probabilities and error probabilities for the by Maity and Sherman's t-method, the signed log-likelihood ratio method (r), and the proposed method (r^*). The nominal values are 0.95, 0.025 and 0.025 respectively.

Design	Method	Coverage probability	Upper error probability	Lower error probability
1	t	0.9080	0.0443	0.0477
	r	0.8712	0.0608	0.0680
	r^*	0.9304	0.0325	0.0371
2	t	0.9528	0.0245	0.0227
	r	0.9459	0.0286	0.0264
	r^*	0.9517	0.0254	0.0229
3	t	0.9115	0.0432	0.0453
	r	0.8750	0.0620	0.0630
	r^*	0.9333	0.0322	0.0345
4	t	0.9531	0.0232	0.0237
	r	0.9454	0.0275	0.0271
	r^*	0.9515	0.0248	0.0237
5	t	0.9508	0.0248	0.0244
	r	0.9453	0.0274	0.0273
	r^*	0.9502	0.0251	0.0247
6	t	0.9496	0.0276	0.0228
	r	0.9492	0.0277	0.0231
	r^*	0.9495	0.0276	0.0229
7	t	0.9505	0.0247	0.0248
	r	0.9441	0.0279	0.0280
	r^*	0.9505	0.0247	0.0248
8	t	0.9475	0.0255	0.0270
	r	0.9473	0.0255	0.0272
	r^*	0.9475	0.0255	0.0270

Table 2: (continued)

Design	Method	Coverage probability	Upper error probability	Lower error probability
9	t	0.9504	0.0262	0.0234
	r	0.9489	0.0268	0.0243
	r^*	0.9503	0.0262	0.0235
10	t	0.9503	0.0250	0.0247
	r	0.9499	0.0253	0.0248
	r^*	0.9501	0.0251	0.0248
11	t	0.9512	0.0234	0.0254
	r	0.9495	0.0239	0.0266
	r^*	0.9511	0.0235	0.0254
12	t	0.9503	0.0252	0.0245
	r	0.9503	0.0252	0.0245
	r^*	0.9503	0.0252	0.0245
13	t	0.8906	0.0544	0.0550
	r	0.8406	0.0814	0.0780
	r^*	0.9254	0.0384	0.0362
14	t	0.9492	0.0274	0.0234
	r	0.9380	0.0322	0.0298
	r^*	0.9502	0.0274	0.0224
15	t	0.8965	0.0531	0.0504
	r	0.8448	0.0806	0.0746
	r^*	0.9285	0.0365	0.0350
16	t	0.9516	0.0222	0.0262
	r	0.9391	0.0277	0.0332
	r^*	0.9507	0.0221	0.0272

Table 2: (continued)

Design	Method	Coverage probability	Upper error probability	Lower error probability
17	t	0.9451	0.0279	0.0270
	r	0.9359	0.0319	0.0322
	r^*	0.9455	0.0276	0.0269
18	t	0.9508	0.0235	0.0257
	r	0.9502	0.0240	0.0258
	r^*	0.9508	0.0235	0.0257
19	t	0.9476	0.0284	0.0240
	r	0.9383	0.0332	0.0285
	r^*	0.9480	0.0280	0.0240
20	t	0.9507	0.0268	0.0225
	r	0.9499	0.0269	0.0232
	r^*	0.9506	0.0268	0.0226
21	t	0.9501	0.0251	0.0248
	r	0.9479	0.0259	0.0262
	r^*	0.9500	0.0252	0.0248
22	t	0.9491	0.0242	0.0267
	r	0.9488	0.0242	0.0270
	r^*	0.9491	0.0242	0.0267
23	t	0.9506	0.0241	0.0253
	r	0.9479	0.0256	0.0265
	r^*	0.9505	0.0242	0.0253
24	t	0.9514	0.0249	0.0237
	r	0.9512	0.0250	0.0238
	r^*	0.9513	0.0249	0.0238

Table 2: (continued)

Design	Method	Coverage probability	Upper error probability	Lower error probability
25	t	0.8719	0.0615	0.0666
	r	0.7868	0.1033	0.1099
	r^*	0.9239	0.0347	0.0414
26	t	0.9274	0.0360	0.0366
	r	0.9059	0.0475	0.0466
	r^*	0.9385	0.0305	0.0310
27	t	0.8783	0.0593	0.0624
	r	0.7944	0.1029	0.1027
	r^*	0.9275	0.0360	0.0365
28	t	0.9282	0.0363	0.0355
	r	0.9036	0.0491	0.0473
	r^*	0.9377	0.0322	0.0301
29	t	0.9460	0.0265	0.0275
	r	0.9292	0.0349	0.0359
	r^*	0.9479	0.0256	0.0265
30	t	0.9515	0.0232	0.0253
	r	0.9483	0.0253	0.0264
	r^*	0.9508	0.0236	0.0256
31	t	0.9458	0.0280	0.0262
	r	0.9292	0.0366	0.0342
	r^*	0.9477	0.0266	0.0257
32	t	0.9502	0.0253	0.0245
	r	0.9477	0.0269	0.0254
	r^*	0.9497	0.0256	0.0247

Table 2: (continued)

Design	Method	Coverage probability	Upper error probability	Lower error probability
33	t	0.9496	0.0253	0.0251
	r	0.9433	0.0291	0.0276
	r^*	0.9499	0.0250	0.0251
34	t	0.9511	0.0229	0.0260
	r	0.9501	0.0235	0.0264
	r^*	0.9510	0.0229	0.0261
35	t	0.9494	0.0257	0.0249
	r	0.9422	0.0302	0.0276
	r^*	0.9499	0.0255	0.0246
36	t	0.9501	0.0264	0.0235
	r	0.9494	0.0267	0.0239
	r^*	0.9501	0.0264	0.0235
37	t	0.8751	0.0651	0.0598
	r	0.7723	0.1155	0.1122
	r^*	0.9268	0.0411	0.0321
38	t	0.9097	0.0459	0.0444
	r	0.8771	0.0610	0.0619
	r^*	0.9323	0.0327	0.0350
39	t	0.8789	0.0592	0.0619
	r	0.7760	0.1092	0.1148
	r^*	0.9270	0.0346	0.0384
40	t	0.9071	0.0443	0.0486
	r	0.8771	0.0592	0.0637
	r^*	0.9303	0.0322	0.0375

Table 2: (continued)

Design	Method	Coverage probability	Upper error probability	Lower error probability
41	t	0.9439	0.0288	0.0273
	r	0.9252	0.0373	0.0375
	r^*	0.9451	0.0285	0.0264
42	t	0.9511	0.0249	0.0240
	r	0.9460	0.0277	0.0263
	r^*	0.9507	0.0251	0.0242
43	t	0.9452	0.0264	0.0284
	r	0.9272	0.0347	0.0381
	r^*	0.9469	0.0254	0.0277
44	t	0.9502	0.0243	0.0255
	r	0.9454	0.0261	0.0285
	r^*	0.9496	0.0246	0.0258
45	t	0.9457	0.0278	0.0265
	r	0.9368	0.0327	0.0205
	r^*	0.9464	0.0273	0.0263
46	t	0.9501	0.0251	0.0248
	r	0.9486	0.0257	0.0257
	r^*	0.9499	0.0252	0.0249
47	t	0.9457	0.0266	0.0277
	r	0.9365	0.0311	0.0324
	r^*	0.9461	0.0263	0.0276
48	t	0.9516	0.0241	0.0243
	r	0.9507	0.0244	0.0249
	r^*	0.9514	0.0241	0.0245

6 CONCLUSION

A likelihood-based third order method is proposed to obtain confidence interval for the difference in means problem when the two populations are independently normally distributed with one variance assumed to be known as the other variance remained unknown. Simulation results illustrated the supreme accuracy of the proposed method in terms of both coverage probability and the symmetry of error rate especially when the sample sizes are small. It is also important to note that the proposed method can be extended to the Behrens-Fisher problem.

Appendix A. General Methodology for Obtaining $Q(\psi)$

To derive the general formula for $Q(\psi)$, we need to first reduce the dimension of the variable to the dimension of the parameter. This dimension reduction can be achieved by conditioning on an implicit ancillary statistic. However, it is not easy to obtain the ancillary statistic. Moreover, the exact ancillary statistic may not exist or even if exists, it may not be unique. Fraser and Reid (1995) showed that only tangent directions, V, to this ancillary statistic are necessary. In fact, V can be obtained by using a pivotal quantity $k(y, \theta)$ and differentiating the data y with respect to the parameter θ while holding the pivotal quantity fixed. In other words,

$$V = \left.\frac{\partial y}{\partial \theta}\right|_{\hat{\theta}} = \left\{\frac{\partial k(y,\theta)}{\partial y}\right\}^{-1}\left.\left\{\frac{\partial k(y,\theta)}{\partial \theta}\right\}\right|_{\hat{\theta}}.$$

Fraser and Reid (1995) also showed that the resulting model can then be approximated by a tangent exponential model with the locally defined canonical parameter

$$\varphi(\theta) = \frac{\partial \ell(\theta)}{\partial y}V.$$

This locally defined canonical parameter gives the relevant parameterization for likelihood inference. Given this new parameterization, and without explicitly specifying the nuisance parameter, Fraser, Reid and Wu (1999) developed a marginalization procedure that gives the recalibrated parameter of interest $\chi(\theta)$ as given in Section 3. Hence $Q(\psi)$ can be obtained as in Section 3.

Appendix B. R Code for the Proposed Method

```
# function to calculate mle standard deviation
msd<-function(data) sd(data) * sqrt((length(data)-1)/length(data))

# function to calculate left tail probability for t, r, rstar respectively.
# "x", "y" are vector values for sample X and Y
# "s1" is the standard deviation for X, "psi" is value of interest.
```

```
tailprob <- function(x,y,s1,psi) {
    xbar <- mean(x)
    ybar <- mean(y)
    msd2 <- msd(y)
    m <- length(x)
    n <- length(y)
    A <- m
    B <- m*psi-2*m*ybar-m*xbar
    C <- (-2)*m*psi*ybar+m*(msd2^2)+m*(ybar^2)+2*m*xbar*ybar+n*(s1^2)
    n <- length(y)
    D <- m*psi*(msd2^2)+m*psi*(ybar^2)-m*xbar*(ybar^2)-n*ybar*(s1^2)
            -m*xbar*(msd2^2)
    v <- 40       # initial value to solve below function "fx"
    fx  <- A*(v^3) + B*(v^2)+ C* v +D
    dfx <- A*(v^2) + B*v + C
    d <- 1
    while ( d!=0 )  {
       v <- v - fx/dfx
       fx <- A*(v^3) + B*(v^2)+ C* v +D
       d <- abs(round(fx,5))
       }
    u20 <- v
    msdc2 <- sqrt(msd2^2 + (ybar-u20)^2 ) # Constrained MLE SD
    lamda <- n/(msdc2^2)*(ybar-u20)
    u10 <- ((s1^2)/m)*lamda+xbar

    lkh.mle <- sum(dnorm(x,xbar,s1,log=TRUE))+sum(dnorm(y,ybar,msd2,log=T))
    lkh.psi <- sum(dnorm(x, u10,s1,log=TRUE))+sum(dnorm(y,u20,msdc2,log=T))
    r <- sign(xbar-ybar-psi)*sqrt(2*(lkh.mle-lkh.psi))
    if (r==0)   rstar <- 0     else  {
    Q <- sign(xbar-ybar-psi)*abs(xbar-u10+(u20-ybar)*(msdc2^2)
          /(msd2^2))*sqrt(((msd2^6)/(msdc2^6))/(s1^2/m + msdc2^2/n
            - (s1^2/m)*(2*(ybar-u20)^2)/(msdc2^2)))
    rstar <- r-(1/r)*log(r/Q)
    }
    df <- (s1^2/m+sd(y)^2/n)^2/((sd(y)^2/n)^2/(n-1))
    t  <- (xbar-ybar-psi)/sqrt(s1^2/m+sd(y)^2/n)
    c(1-pt(t,df),1-pnorm(r), 1-pnorm(rstar))
}
# function to calculate the (1-2*alpha)100% CIs for rstar and r. "ini" is
# the initial value of lower bound the CI. The left tail probability for
```

```
# initial value needs to be less than alpha.
rstarCI <- function(x,y,s1,ini,alpha){
    prstar <-tailprob(x,y,s1,ini)[3]
    while ( prstar < alpha) {
    ini <- ini+0.0001
    prstar <-tailprob(x,y,s1,ini)[3]
    }
    if ( abs(tailprob(x,y,s1,ini-0.0001)[3]-alpha) < abs(prstar-alpha)) {
    prstar <- tailprob(x,y,s1,ini-0.0001)[3]
    ini <- ini - 0.0001
    c(prstar,ini,2*(mean(x)-mean(y))-ini)
    }
    c(ini,2*(mean(x)-mean(y))-ini)
}
#
rCI <- function(x,y,s1,ini,alpha){
    pr <-tailprob(x,y,s1,ini)[2]
    while ( pr < alpha) {
    ini <- ini+0.0001
    pr <-tailprob(x,y,s1,ini)[2]
    }
    if ( abs(tailprob(x,y,s1,ini-0.0001)[2]-alpha) < abs(pr-alpha)) {
    pr <- tailprob(x,y,s1,ini-0.0001)[2]
    ini <- ini - 0.0001
    }
    c(ini,2*(mean(x)-mean(y))-ini)
 }
# numerical example in Section 5
x<-c(724,718,776,760,745,759,795,756,742,740,761,749,739,747,742)
y<-c(735,775,729,755,783,760,738,780)
s1<-20
rstarCI(x,y,s1,-22.7050,0.05) # 90% CI for rstar
rCI(x,y,s1,-21.6245,0.05)     # 90% CI for r
######################
```

References

[1] Barndorff–Nielsen, O. E. (1986). Inference on full and partial parameters, based on the standardized signed log-likelihood ratio. *Biometrika*, **73**, 307–322.

[2] Barndorff–Nielsen, O. E. (1991). Modified signed log-likelihood ratio statistic. *Biometrika*, **78**, 557–563.

[3] Barndorff–Nielsen, O. E. and Cox, D. R. (1994). *Inference and Asymptotics*. Chapman and Hall, London.

[4] Cox, D. R. and Hinkley, D. V. (1974). *Theoretical Statistics*. Chapman and Hall, London.

[5] Fraser, D. A. S. and Reid, N. (1995). Ancillaries and third-order significance. *Utilitas Math*, **47**, 33–53.

[6] Fraser, D. A. S., Reid, N. and Wu, J. (1999). A simple general formula for tail probabilities for frequentist and Bayesian inference. *Biometrika*, **86**, 249–264.

[7] Maity, A. and Sherman, M. (2006). The two sample t–test with one variance unknown. *The American Statistician*, **60**, 163–166.

[8] Satterthwaite, F. E. (1941). Synthesis of variance. *Psychometrika*, **6**, 309–316.

[9] Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, **6**, 110–114.

[10] Scheffé, H. (1970). Practical solutions of the Behrens-Fisher problem. *Journal of the American Statistical Association*, **65**, 1501–1508.