Available at
www.ElsevierMathematics.com
POWERED BY SCIENCE DIRECT
Journal of Statistical Planning and Inference 123 (2004) 49–60
journal of statistical planning and inference
www.elsevier.com/locate/jspi

Generalized confidence intervals for the ratio of means of two normal populations

Jack C. Lee[a,*], Shu-Hui Lin[b]

[a] Institute of Statistics and Graduate Institute of Finance, National Chiao Tung University, Hsinchu, Taiwan
[b] National Taichung Institute of Technology, Taichung 404, Taiwan

Received 16 September 2001; accepted 16 January 2003

Abstract

Based on the generalized p-values and generalized confidence interval developed by Tsui and Weerahandi (J. Amer. Statist. Assoc. 84 (1989) 602), Weerahandi (J. Amer. Statist. Assoc. 88 (1993) 899), respectively, hypothesis testing and confidence intervals for the ratio of means of two normal populations are developed to solve Fieller's problems. We use two different procedures to find two potential generalized pivotal quantities. One procedure is to find the generalized pivotal quantity based directly on the ratio of means. The other is to treat the problem as a pseudo Behrens–Fisher problem through testing the two-sided hypothesis on θ, and then to construct the $1-\alpha$ generalized confidence interval as a counterpart of generalized p-values. Illustrative examples show that the two proposed methods are numerically equivalent for large sample sizes. Furthermore, our simulation study shows that confidence intervals based on generalized p-values without the assumption of identical variance are more efficient than two other methods, especially in the situation in which the heteroscedasticity of the two populations is serious.

Keywords: Fieller's theorem; Generalized confidence interval; Generalized p-values; Generalized pivotal quantity; Heteroscedasticity; Pseudo Behrens–Fisher problem; Ratio estimation

1. Introduction

Much attention has been paid to Fieller's problems, because they occurred frequently in many important research areas such as bioassay and bioequivalence. In bioassay problem, the relative potency of a test preparation as compared with a standard is estimated by (i) the ratio of two means for direct assays, (ii) the ratio of two independent

* Corresponding author. Tel.: +886-3-5728746; fax: +886-3-5728745.
E-mail address: jclee@stat.nctu.edu.tw (J.C. Lee).

0378-3758/$ - see front matter
doi:10.1016/S0378-3758(03)00141-1

normal random variables for parallel-line assays and (iii) the ratio of two slopes for slope-ratio assays. In biological assay problems (Fieller, 1954; Finney, 1978) and bioequivalence problems (Chow and Liu, 1992; Berger and Hsu, 1996), one is interested in the relative potency of two drugs or treatments. Traditionally, Fieller (1944, 1954) provides a widely used general procedure for the construction of confidence intervals (often called Fieller's theorem) for the ratio of means (also discussed by Rao, 1973; Finney, 1978; Koschat, 1987 and Hwang, 1995). Under homoscedasticity case, Koschat (1987) has also shown that within a large class of sensible procedures the Fieller solution is the only one that gives exact coverage probability for all parameters. However, the conventional procedures are often restricted to the assumption of a common variance or pairwise observations for mathematical tractability. Thus, the exact approaches to Fieller's problems under the unequal variances assumption have also been intensively investigated. Consider the following problem: Let $\boldsymbol{X}=(X_1,X_2,\ldots,X_{n_1})$ and $\boldsymbol{Y}=(Y_1,Y_2,\ldots,Y_{n_2})$ be two independent sets of observations for the potency of a standard drug and a new drug, respectively. Assume that X_i are independently and identically distributed as $N(\mu_1,\sigma_1^2)$, Y_i are independently and identically distributed as $N(\mu_2,\sigma_2^2)$, where μ_1 and μ_2 are the true potencies. The problem is to determine, with any desired probability, the range of values for the ratio of means $\theta=\mu_2/\mu_1$, which is the relative potency of the new drug to the standard.

Under the assumption of identical variance, Fieller (1954) constructed a confidence interval based on the statistic

$$T=\frac{(\bar{Y}-\theta\bar{X})}{\sqrt{(\frac{1}{n_2}+\frac{\theta^2}{n_1})S^2}}, \tag{1.1}$$

where $\bar{X}=(1/n_1)\sum_{i=1}^{n_1}X_i$, $\bar{Y}=(1/n_2)\sum_{j=1}^{n_2}Y_j$ and $S^2=(\sum_{i=1}^{n_1}(X_i-\bar{X})^2+\sum_{j=1}^{n_2}(Y_j-\bar{Y})^2)/(n_1+n_2)$. It is obvious that $\sqrt{(n_1+n_2-2)/(n_1+n_2)}T$ has the Student's t distribution with (n_1+n_2-2) degrees of freedom. Solving the inequality

$$\left\{\theta\colon \sqrt{\frac{(n_1+n_2-2)}{n_1+n_2}}|T|<t_{1-\frac{\alpha}{2}}\right\}$$

$$=\left\{\theta\colon |\bar{y}-\theta\bar{x}|\leqslant t_{1-\frac{\alpha}{2}}\sqrt{\frac{(n_1+n_2)(1/n_2+\theta^2/n_1)s^2}{n_1+n_2-2}}\right\}, \tag{1.2}$$

where $t_{1-\alpha/2}$ is the $(1-\alpha/2)$th quantile of the t distribution, the exact $1-\alpha$ confidence interval for θ will be obtained.

On the other hand, if variances are related to the means, such as $\sigma_i^2=(c+\mu_i)^k\sigma^2$ with $c+\mu_i>0$, $i=1,2$ and k is known, Cox (1985) provided a interval estimate based on the statistic

$$T^*=\frac{(N-2)l^2}{aS^*}, \tag{1.3}$$

with $l=(c+\mu_2)(c+\bar{x})-(c+\mu_1)(c+\bar{y})$, $a=(c+\mu_2)^2((c+\mu_1)^k/n_1)+(c+\mu_1)^2((c+\mu_2)^k/n_2)$ and $S^*/\sigma^2=\sum_{i=1}^{n_1}(X_i-\bar{X})^2/\sigma_1^2+\sum_{i=1}^{n_2}(Y_i-\bar{Y})^2/\sigma_2^2$. It is noted that T^* has the Fisher-Snedecor's F distribution with 1 and n_1+n_2-2 degrees of freedom. Define $\theta=(c+\mu_2)/(c+\mu_1)$, the $100(1-\alpha)\%$ confidence interval for θ is obtained by solving the inequality

$$\left\{\theta\colon \frac{(N-2)l^2}{aS^*} \leqslant F_{1-\alpha}(1, n_1+n_2-2)\right\}. \tag{1.4}$$

For $c=0$ and $k=2$, the $100(1-\alpha)\%$ confidence interval for $\theta=\mu_2/\mu_1$ is based on solving the quadratic inequality

$$\left\{\theta\colon (\bar{y}-\theta\bar{x})^2 \leqslant \frac{F_{1-\alpha}(1, n_1+n_2-2)}{n_1+n_2-2}\left[\frac{s_2^2}{n_1/(n_1+n_2)} + \theta^2\frac{s_1^2}{n_2/(n_1+n_2)}\right]\right\}, \tag{1.5}$$

with $s_1^2=(1/n_1)\sum_{i=1}^{n_1}(x_i-\bar{x})^2$ and $s_2^2=(1/n_2)\sum_{i=1}^{n_2}(y_i-\bar{y})^2$, respectively.

In this article, we propose two different exact approaches based on generalized p-values and generalized confidence intervals, as defined by Tsui and Weerahandi (1989), Weerahandi (1993), respectively, to construct confidence intervals for the ratio of means of two normal populations under heteroscedasticity. The lack of exact confidence intervals in many applications can be attributed to the statistical problems involving nuisance parameters. The possibility of exact confidence interval can be achieved by extending the definition of confidence interval. To generalize the definition of confidence intervals, first examine the properties of interval estimates obtained by the conventional definition. To fix ideas, consider a random sample $\boldsymbol{X}=(X_1, X_2, \ldots, X_n)$ from a distribution with an unknown parameter θ. Let $A(\boldsymbol{X})$ and $B(\boldsymbol{X})$ be two statistics satisfying the equation

$$P[A(\boldsymbol{X}) \leqslant \theta \leqslant B(\boldsymbol{X})] = \gamma, \tag{1.6}$$

where γ is a prespecified constant between 0 and 1. Let $a=A(\boldsymbol{x})$ and $b=B(\boldsymbol{x})$ be the observed values of the two statistics, then, in the commonly used terminology, $[a,b]$ is a confidence interval for θ with the confidence coefficient γ. For example, if $\gamma=0.95$, then the interval $[a,b]$ obtained in this manner is a 95% confidence interval. This approach to constructing interval estimates is conceptually simple and easy to implement, but in most applications involving nuisance parameters it is not easy or impossible to find $A(\boldsymbol{x})$ and $B(\boldsymbol{x})$ so as to satisfy (1.6) for all possible values of the nuisance parameters. The idea in generalized confidence intervals is to make this possible by making probability statements relative to the observed sample, as done in Bayesian and nonparametric methods. In other words, we allow the functions $A(\cdot)$ and $B(\cdot)$ to depend not only on the observable random vector $\boldsymbol{X}$ but also on the observed data $\boldsymbol{x}_{\text{obs}}$.

We will briefly introduce the general theory and provide our first procedure for finding the generalized pivotal quantity based directly on the ratio in Section 2. The equal-tail confidence intervals are included as well. In Section 3, we tackle the problem as a pseudo Behrens–Fisher problem through the testing of two-sided hypothesis on θ, with the interval construction treated as a counterpart of generalized p-values.

It is interesting to note that these two procedures are numerically equivalent for large sample sizes. Through the procedure presented in Section 3, the interval proposed by Cox (1985) with $c=0$, $k=2$ can be viewed as an approximation of our method when $n_1=n_2$. Both procedures developed in this article get more precise interval than those of Fieller's and Cox's when serious heteroscedasticity is present. Two numerical examples are illustrated in Section 4 comparing the proposed methods with other methods in the presence of heteroscedasticity. A simulation study is conducted to calculate the coverage probabilities in different scenarios in Section 5, and finally some concluding remarks are given in Section 6.

2. Generalized pivotal quantity based directly on the ratio of means

2.1. Notations and theory

Let $\boldsymbol{X}$ be an observable random vector having a density function $f(\boldsymbol{X}|\boldsymbol{\zeta})$, where $\boldsymbol{\zeta}=(\theta,\boldsymbol{\eta})$ is a vector of unknown parameters, θ is a parameter of interest, and $\boldsymbol{\eta}$ is a vector of nuisance parameters. Let χ be the sample space of possible values of $\boldsymbol{X}$ and Θ be the parameter space of θ. A possible observation from $\boldsymbol{X}$ is denoted by $\boldsymbol{x}$, where $\boldsymbol{x}\in\chi$, and the value of $\boldsymbol{X}$ actually observed is denoted by $\boldsymbol{x}_{\text{obs}}$.

We are interested in finding interval estimates of θ based on observed values of $\boldsymbol{X}$. The problem is to construct generalized confidence intervals of the form $[A(\boldsymbol{x}_{\text{obs}}), B(\boldsymbol{x}_{\text{obs}})]$, where $A(\boldsymbol{x}_{\text{obs}})$ and $B(\boldsymbol{x}_{\text{obs}})$ are functions of $\boldsymbol{x}_{\text{obs}}$.

The conventional approach to constructing confidence intervals is based on the notion of a pivotal quantity $R=r(\boldsymbol{X};\theta)$ with the property that for given γ we can find a region C such that

$$P_{\boldsymbol{\zeta}}(r(\boldsymbol{X};\theta)\in C)=\gamma \tag{2.1}$$

for all $\boldsymbol{\zeta}$. We then define $\Theta(\boldsymbol{X})=\{\theta\,|\,r(\boldsymbol{X};\theta)\in C\}$. Since $P_{\boldsymbol{\zeta}}\{\theta\in\Theta(\boldsymbol{X})\}=P_{\boldsymbol{\zeta}}\{r(\boldsymbol{X};\theta)\in C\}=\gamma$, $\Theta(\boldsymbol{X})$ is a γ-level confidence region and we are "100γ% confident" that θ is in the observed region $\Theta(\boldsymbol{x}_{\text{obs}})$.

Weerahandi (1993) extended the definition of a pivotal quantity as follows. Let $R=r(\boldsymbol{X};\boldsymbol{x},\boldsymbol{\zeta})$ be a function of $\boldsymbol{X}$ and possibly $\boldsymbol{x}$ and $\boldsymbol{\zeta}$ as well. Then R is said to be a generalized pivotal quantity if it has the following two properties:

(i) For any fixed $\boldsymbol{x}\in\chi$, R has a probability distribution $P_{\boldsymbol{x}}$ free of unknown parameters.
(ii) If $\boldsymbol{X}=\boldsymbol{x}$, then $r(\boldsymbol{x};\boldsymbol{x},\zeta)$ does not depend on $\boldsymbol{\eta}$, the vector of nuisance parameters.

Using (i), for given γ we can find, for any fixed $\boldsymbol{x}$, a computable region $C(\boldsymbol{x})$ such that

$$P_{\boldsymbol{x}}\{r(\boldsymbol{X};\boldsymbol{x},\boldsymbol{\zeta})\in C(\boldsymbol{x})\}=\gamma. \tag{2.2}$$

By (ii), for any $\boldsymbol{x}$ whether $r(\boldsymbol{x};\boldsymbol{x},\boldsymbol{\zeta})\in C(\boldsymbol{x})$ holds depends only on θ, and not on the nuisance parameters $\boldsymbol{\eta}$, so we may define the 100γ% generalized confidence set

$$\Theta(\boldsymbol{x}_{\rm obs}) = \{\theta \mid r(\boldsymbol{x}_{\rm obs}; \boldsymbol{x}_{\rm obs}, \boldsymbol{\zeta}) \in C(\boldsymbol{x}_{\rm obs})\}. \tag{2.3}$$

This region is a realization of a random subset

$$\Theta(\boldsymbol{X}) = \{\theta \mid r(\boldsymbol{X}; \boldsymbol{X}, \boldsymbol{\zeta}) \in C(\boldsymbol{X})\} \tag{2.4}$$

of θ values. From (2.2) we obviously do not have the analogue of (2.1), and hence cannot claim $P_{\zeta}\{\theta \in \Theta(\boldsymbol{X})\} = \gamma$. However, from (2.2) it does follow that $P\{r(\boldsymbol{X}_1; \boldsymbol{X}_2, \boldsymbol{\zeta}) \in C(\boldsymbol{X}_2)\} = \gamma$ if $\boldsymbol{X}_1$ and $\boldsymbol{X}_2$ are independent of one another, each distributed as $\boldsymbol{X}$. Thus the generalized confidence set (2.4) intuitively corresponds to using the same $\boldsymbol{X}$ twice: viewing it as a future, unobserved variable, and also conditioning on its observed value in the data.

It is noted that a generalized pivotal quantity in interval estimation is the counterpart of generalized test variables in significance testing of hypotheses proposed by Tsui and Weerahandi (1989). If the form of a p-value for a one-sided test is readily available, a generalized confidence interval for θ can be deduced directly from its power function.

2.2. The procedure

Let $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ be random samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively, where μ_i and σ_i^2 are unknown with $i = 1, 2$. For the univariate Fieller–Creasy problem, we want to construct intervals for the parameter $\theta = \mu_2/\mu_1$. First, we will find a generalized pivotal quantity, R, based on the following sufficient statistics

$$\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i, \quad \bar{Y} = \frac{1}{n_2}\sum_{j=1}^{n_2} Y_j,$$

$$S_1^2 = \frac{\sum_{i=1}^{n_1}(X_i - \bar{X})^2}{n_1}, \quad \text{and} \quad S_2^2 = \frac{\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2}{n_2}.$$

Since a generalized pivotal can be a function of all unknown parameters, we can construct R based on the random quantities $Z_1 = \sqrt{n_1}(\bar{X} - \mu_1)/\sigma_1 \sim N(0,1)$, $Z_2 = \sqrt{n_2}(\bar{Y} - \mu_2)/\sigma_2 \sim N(0,1)$, $U_1 = n_1 S_1^2/\sigma_1^2 \sim \chi^2_{n_1-1}$ and $U_2 = n_2 S_2^2/\sigma_2^2 \sim \chi^2_{n_2-1}$, whose distributions are free of unknown parameters. Using

$$\theta \equiv \frac{\bar{Y} - Z_2\sigma_2/\sqrt{n_2}}{\bar{X} - Z_1\sigma_1/\sqrt{n_1}} = \frac{\bar{Y} - Z_2 S_2/\sqrt{U_2}}{\bar{X} - Z_1 S_1/\sqrt{U_1}},$$

we can define the following potential generalized pivotal

$$R(X, Y; x, y, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = \frac{\bar{y} - Z_2 s_2/\sqrt{U_2}}{\bar{x} - Z_1 s_1/\sqrt{U_1}} = \frac{\bar{y} - T_2 s_2/\sqrt{n_2 - 1}}{\bar{x} - T_1 s_1/\sqrt{n_1 - 1}},$$

where $s_1, s_2, \bar{x}, \bar{y}$ are the observed values of $S_1, S_2, \bar{X}, \bar{Y}$, respectively. Note that $T_1 \sim t(n_1 - 1)$ is independent of $T_2 \sim t(n_2 - 1)$, and $R(x, y; x, y, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = \theta$.

Now consider the problem of constructing lower confidence bounds for θ. Since the observed value of R is θ, the following probability statement will lead to a right-sided $100(1-\alpha)\%$ confidence interval,

$$\begin{aligned}1-\alpha &= P\{R \geqslant c\}\\ &= P\left\{T_2 \geqslant \frac{\sqrt{n_2-1}}{s_2}\left[\bar{y}+c\left(T_1\frac{s_1}{\sqrt{n_1-1}}-\bar{x}\right)\right]\Bigg| T_1 > \frac{\sqrt{n_1-1}}{s_1}\,\bar{x}\right\}\\ &\quad\times P\left\{T_1 > \frac{\sqrt{n_1-1}}{s_1}\,\bar{x}\right\}\\ &\quad+P\left\{T_2 \leqslant \frac{\sqrt{n_2-1}}{s_2}\left[\bar{y}+c\left(T_1\frac{s_1}{\sqrt{n_1-1}}-\bar{x}\right)\right]\Bigg| T_1 < \frac{\sqrt{n_1-1}}{s_1}\,\bar{x}\right\}\\ &\quad\times P\left\{T_1 < \frac{\sqrt{n_1-1}}{s_1}\bar{x}\right\} \qquad (2.5)\end{aligned}$$

where $T_1 \sim t(n_1-1)$ and $T_2 \sim t(n_2-1)$. It is evident that $[c_{1-\alpha},\infty)$ is the desired $100(1-\alpha)\%$ generalized confidence interval for θ, where $c_{1-\alpha}$ is the value of c satisfying (2.5) for a specified value of $1-\alpha$.

Similarly, the $100(1-\alpha)\%$ upper confidence bound $c'_{1-\alpha}$ for θ can be obtained through

$$\begin{aligned}1-\alpha &= P\{R \leqslant c'_{1-\alpha}\}\\ &= P\left\{T_2 \leqslant \frac{\sqrt{n_2-1}}{s_2}\left[\bar{y}+c'_{1-\alpha}\left(T_1\frac{s_1}{\sqrt{n_1-1}}-\bar{x}\right)\right]\Bigg| T_1 > \frac{\sqrt{n_1-1}}{s_1}\,\bar{x}\right\}\\ &\quad\times P\left\{T_1 > \frac{\sqrt{n_1-1}}{s_1}\,\bar{x}\right\}\\ &\quad+P\left\{T_2 \geqslant \frac{\sqrt{n_2-1}}{s_2}\left[\bar{y}+c'_{1-\alpha}\left(T_1\frac{s_1}{\sqrt{n_1-1}}-\bar{x}\right)\right]\Bigg| T_1 < \frac{\sqrt{n_1-1}}{s_1}\,\bar{x}\right\}\\ &\quad\times P\left\{T_1 < \frac{\sqrt{n_1-1}}{s_1}\bar{x}\right\}. \qquad (2.6)\end{aligned}$$

Likewise, the $100(1-\alpha)\%$ equal tail confidence interval for θ can also be derived through finding $c_{1-\alpha/2}$ and $c'_{1-\alpha/2}$ in (2.5) and (2.6), respectively.

The underlying family of distributions is invariant under the common scale transformations

$$(\bar{X},\bar{Y},S_1,S_2) \to (k\bar{X},k\bar{Y},kS_1,kS_2) \quad \text{and} \quad (\mu_1,\mu_2) \to (k\mu_1,k\mu_2),$$

where k is a positive constant. Obviously, the parameter of interest is unaffected by any change of scale, and therefore, the statistical problem is invariant. Furthermore, the observed value of the statistic R does not depend on the data, any scale invariant generalized confidence region of θ can be constructed from R alone (see Weerahandi, 1993, Theorem 3.1).

3. Generalized pivotal quantity defined through the testing procedure

In this section, we solve the Fieller–Creasy problem through the testing procedure with the interval estimation obtained as a counterpart of generalized p-values. Suppose $X_1, X_2, \ldots, X_{n_1}$ and $Y_1, Y_2, \ldots, Y_{n_2}$ are random samples from $N(\mu_1, \sigma_1^2)$ and $N(\mu_2, \sigma_2^2)$, respectively, where μ_i and σ_i^2 are unknown with $i = 1, 2$. Consider the problem of significance testing of hypotheses concerning the parameter $\theta = \mu_2/\mu_1$. Since $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$ can be transformed to $H_0^* : \delta = 0$ versus $H_1^* : \delta \neq 0$ where $\delta = \mu_2 - \mu_1\theta_0$, that is, we can treat this testing problem as a pseudo Behrens–Fisher problem. In view of the fact that $\bar{Y} - \theta_0\bar{X}$ is distributed as $N(\delta, (\sigma_2^2/n_2) + \theta_0^2(\sigma_1^2/n_1))$ which depends on the parameter of interest δ and the nuisance parameter $(\sigma_2^2/n_2) + \theta_0^2(\sigma_1^2/n_1)$, the following potential pivotal quantity can be defined as in Tsui and Weerahandi (1989),

$$R^*(X, Y; x, y, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2) = \frac{(\bar{Y} - \theta_0\bar{X}) - \delta}{\sqrt{\sigma_2^2/n_2 + \theta_0^2\sigma_1^2/n_1}} \sqrt{\frac{s_2^2\sigma_2^2}{n_2 S_2^2} + \theta_0^2 \frac{s_1^2\sigma_1^2}{n_1 S_1^2}} + \delta, \tag{3.1}$$

where

$$\bar{X} = \frac{1}{n_1}\sum_{i=1}^{n_1} X_i, \quad \bar{Y} = \frac{1}{n_2}\sum_{j=1}^{n_2} Y_j,$$

$$S_1^2 = \frac{\sum_{i=1}^{n_1}(X_i - \bar{X})^2}{n_1} \quad \text{and} \quad S_2^2 = \frac{\sum_{i=1}^{n_2}(Y_i - \bar{Y})^2}{n_2}$$

are summary statistics, and s_1 and s_2 are the observed values of S_1 and S_2, respectively. It is noted that the observed value of R^* is $r_{\text{obs}} = \bar{y} - \theta_0\bar{x}$, where $\bar{x}$ and $\bar{y}$ are the observed values of $\bar{X}$ and $\bar{Y}$, respectively. Furthermore, the distribution of R^* is the same as $Z\sqrt{(s_2^2/U_2) + \theta_0^2(s_1^2/U_1)} + \delta$, where $Z \sim N(0, 1)$, $U_1 = n_1S_1^2/\sigma_1^2 \sim \chi^2_{n_1-1}$, $U_2 = n_2S_2^2/\sigma_2^2 \sim \chi^2_{n_2-1}$, and the random variables Z, U_1, U_2 are independent. Since R^* is stochastically increasing in δ, the geneneralized p-values appropriate for testing the left-sided null hypothesis of the form $H_0 : \delta \leqslant 0$ versus $H_1 : \delta > 0$ is

$$\begin{aligned} p &= P\{R^* \geqslant r_{\text{obs}} | \delta = 0\} \\ &= E_B\left\{ G_{n_1+n_2-2}\left[-\frac{\sqrt{n_1+n_2-2}(\bar{y} - \theta_0\bar{x})}{\sqrt{s_2^2/(1-B) + \theta_0^2 s_1^2/B}} \right] \right\}, \end{aligned}$$

where $G_{n_1+n_2-2}$ is the *cdf* of the Student's t distribution with $n_1 + n_2 - 2$ degrees of freedom and the expectation is taken with respect to the beta random variable,

$$B = \frac{U_1}{U_1 + U_2} \sim \text{beta}\left(\frac{n_1 - 1}{2}, \frac{n_2 - 1}{2}\right).$$

In particular, the p-value for testing point null hypotheses of the form $H_0 : \delta = 0$ is

$$p = 2E_B \left\{ G_{n_1+n_2-2} \left[-\frac{\sqrt{n_1 + n_2 - 2}|\bar{y} - \theta_0 \bar{x}|}{\sqrt{s_2^2/(1 - B) + \theta_0^2 s_1^2/B}} \right] \right\}. \tag{3.2}$$

Two sided confidence intervals can be deduced from (3.2). A generalized $100(1 - \alpha)\%$ confidence interval for θ can be derived by solving

$$1 - \alpha = E_B \left\{ P\left(|T_{n_1+n_2-2}| \leqslant \frac{\sqrt{n_1 + n_2 - 2}|\bar{y} - \theta \bar{x}|}{\sqrt{s_2^2/(1 - B) + \theta^2 s_1^2/B}} \right) \right\}, \tag{3.3}$$

where T has a Student's t distribution with $n_1 + n_2 - 2$ degrees of freedom and E_B denotes the expectation with respect to beta$((n_1 - 1)/2, (n_2 - 1)/2)$. Eq. (3.3) can also be expressed as

$$1 - \alpha = E_B \left\{ H_{F_{1,n_1+n_2-2}} \left[\frac{(n_1 + n_2 - 2)(\bar{y} - \theta \bar{x})^2}{s_2^2/(1 - B) + \theta^2 s_1^2/B} \right] \right\}, \tag{3.4}$$

where $H_{F_{1,n_1+n_2-2}}$ is the *cdf* of the F distribution with $1, n_1 + n_2 - 2$ degrees of freedom.

It is interesting to note that if $\sigma_i^2 = \mu_i^2 \sigma,\ i = 1, 2$, Cox's confidence interval in (1.5) can be obtained from (3.4) by replacing B with its expected value $\frac{1}{2}$ when $n_1 = n_2$. Thus, in a way, Cox's result can be treated as an approximation of our method when $n_1 = n_2$. According to Cox's procedure with $c = 0,\ k = 2$, the $1 - \alpha$ confidence interval is based on solving the quadratic inequality

$$\left\{ \theta\colon (\bar{y} - \theta \bar{x})^2 \leqslant \frac{F_{1-\alpha}(1, n_1 + n_2 - 2)}{n_1 + n_2 - 2} \left[\frac{s_2^2}{1 - \hat{B}} + \theta^2 \frac{s_1^2}{\hat{B}} \right] \right\},$$

with $\hat{B} = n_2/(n_1 + n_2)$, which is equal to $E(B)$ when $n_1 = n_2$. Comparing (3.4) with Fieller's confidence interval in (1.2), we see that separate estimates s_1^2 and s_2^2 are used for σ_1^2 and σ_2^2, respectively, rather than a pooled estimate s^2 for the common variance σ^2. Also, the confidence interval is obtained via evaluating the expectation with respect to a beta distribution. Consequently, the proposed procedure will be more general and useful in getting a decent interval when serious heteroscedasticity is present.

It is noted that for the Behrens–Fisher problem, Tsui and Weerahandi (1989) derived the generalized test variable, which is similar to R^*, by the methods of invariance and similarity. Consider the equivalent problem of constructing interval estimates based on the three random quantities, R^*, U_1, and U_2. Recall that the distributions of each of these random variables is free of unknown parameters. Moreover, the observed value of U_1 and U_2 depend on the nuisance parameter σ_1 and σ_2, respectively, but they are independent of the data when $\sigma_1 = s_1$ and $\sigma_2 = s_2$. Therefore, according to

Tsui and Weerahandi (1989), all confidence intervals similar in both σ_1 and σ_2 can be generated using R^* alone.

4. Illustrative examples

Two examples are given to illustrate the advantages of our proposed methods for setting limits on the ratio of means of two normal populations. The objective of these examples is to show how one will fail to get a decent confidence interval, such as Fieller's method, when the variances are unequal.

4.1. Example 1

The data in Table 1 is taken from Jarvis et al. (1987) and Pagano and Gauvreau (1993, p. 254) to measure the relative level of carboxyhemoglobin for a group of nonsmokers and a group of cigarette smokers. The purpose of this example is to analyze the relative carboxyhemoglobin level for two large groups of nonsmokers and cigarette smokers, where μ_1 and μ_2 are the true means of carboxyhemoglobin levels for nonsmokers and cigarette smokers, respectively. The summary data is provided in Table 1 and the interval limits as well as interval widths for four procedures are demonstrated in Table 2.

It is found that the procedures without the assumption of equal variance is better than Fiellers. Thus, the procedures based on a common variance will be given at the cost of wider interval estimates when the population variances are different. Moreover, our procedures developed in Sections 2 and 3 are numerically equivalent and both procedures yield shorter intervals than the other methods.

4.2. Example 2

The second example is to construct confidence intervals for the ratio of two proportions. Let $\boldsymbol{x} = (x_1, \ldots, x_{n_1})$ and $\boldsymbol{y} = (y_1, \ldots, y_{n_2})$ be two independent sets of observations for the potency of a standard drug and a new drug, respectively. Suppose $X_i's$ are independently and identically distributed as $B(1, \eta)$ and $Y_i's$ are independently and identically distributed as $B(1, \phi)$. We are interested in the interval estimation for the ratio of proportions $\theta = \phi/\eta$. In this case, we assume the sample sizes are large enough, so the binomial distribution can be adequately approximated by the normal distribution. Suppose $\bar{x} = 0.38$, $\bar{y} = 0.52$, then the alternates for s_1^2 and s_2^2 are $((n_1 - 1)/n_1)\bar{x}(1 - \bar{x})$ and $((n_2 - 1)/n_2)\bar{y}(1 - \bar{y})$, respectively. It is noted that the MLE for the ratio of two proportions is $\hat{\theta} = \bar{y}/\bar{x} = 1.37$. We will compare four different procedures with different pairs of sample sizes $(n_1, n_2) = (50, 50), (80, 50), (100, 30)$. The 95% confidence intervals and the corresponding confidence widths for θ are shown in Table 3. It is noted that in this mild heteroscedasticity example, Fieller's and Cox's methods perform well in the case of equal sample size, but their performances deteriorate as the difference of sample sizes increases. Again, for large sample sizes, our proposed methods are

Table 1
Carboxyhemoglobin for nonsmokers and smokers groups, percent

Group	n_i	$\bar{x}_i$	s_i^2
Nonsmokers	121	1.3	1.704
Smokers	75	4.1	4.054

Table 2
95% confidence interval for μ_2/μ_1

Procedure	Interval limits	Interval width
Section 2 Eqs. (2.5) and (2.6)	(2.57, 3.95)	1.38
Section 3 Eq. (3.4)	(2.57, 3.95)	1.38
Fieller (1954)	(2.44, 4.40)	1.97
Cox (1985)	(2.52, 4.14)	1.62

Table 3
95% confidence interval for ϕ/η

Procedure	$n_1 = 50,\ n_2 = 50$		$n_1 = 80,\ n_2 = 50$		$n_1 = 100,\ n_2 = 30$	
	Limits	Width	Limits	Width	Limits	Width
Eqs. (2.5) and (2.6)	(0.88, 2.27)	1.39	(0.91, 2.06)	1.15	(0.83, 2.08)	1.25
Eq. (3.4)	(0.88, 2.27)	1.39	(0.91, 2.06)	1.15	(0.83, 2.08)	1.25
Fieller (1954)	(0.88, 2.27)	1.39	(0.93, 2.22)	1.29	(0.83, 2.61)	1.73
Cox (1985)	(0.88, 2.26)	1.38	(0.93, 2.21)	1.28	(0.88, 2.60)	1.72

numerically equivalent and they perform reasonably well comparing with the other methods in all cases.

5. A simulation study

A simulation study is conducted for calculating the coverage probabilities in different combinations of sample sizes and population variances. Two sets of normal data are generated with $\mu_1 = \mu_2 = 2$ and 95% coverage probabilities are calculated based on 1000 replicates. The results are demonstrated in Table 4. We find that Fieller's procedure has good coverage probabilities when the data are generated from two normal populations with identical variance, but its performance deteriorates as the degree of heteroscedasticity increases. Cox's method perform poorly in the situations in which the smaller sample sizes are associated with larger variances. On the other hand, our procedures perform quite well even when the population variances are different.

Table 4
95% Comparison of coverage probabilities for μ_2/μ_1 $(1-\alpha=0.95)$[a]

$n1:n2$	$\sigma_1:\sigma_2$	Fieller	Cox	Eq. (3.4)	Eqs. (2.5) and (2.6)
10:10	1:1	0.947	0.951	0.954	0.936
10:10	1:2	0.944	0.947	0.956	0.954
10:10	1:3	0.938	0.946	0.950	0.956
10:10	1:4	0.836	0.940	0.948	0.954
10:10	1:5	0.696	0.939	0.950	0.946
10:5	1:1	0.956	0.953	0.956	0.956
10:5	1:2	0.874	0.871	0.960	0.962
10:5	1:3	0.764	0.849	0.958	0.958
10:5	1:4	0.672	0.836	0.953	0.962
10:5	1:5	0.580	0.799	0.954	0.952
5:10	1:1	0.951	0.945	0.958	0.938
5:10	1:2	0.971	0.976	0.959	0.960
5:10	1:3	0.888	0.982	0.948	0.966
5:10	1:4	0.636	0.979	0.933	0.958
5:10	1:5	0.404	0.981	0.928	0.956

[a]Based on 1000 replicates in each combination.

6. Concluding remarks

In this article, we propose two different exact generalized approaches based on generalized p-values and generalized confidence intervals to solve the well-known Fieller-Creasy problem, which is widely used in many important research areas such as bioassay and bioequivalence. Under homogeneous case, Fieller's solution gives exact coverage probability for all parameters. Unfortunately, in the presence of serious heteroscedasticity, the methods under the restriction of identical variance cannot yield decent confidence intervals. Through the proposed methods in this article, an exact $1-\alpha$ generalized confidence intervals for the ratio of two means can be obtained under unequal variances and unequal sample sizes. According to our findings, the existing procedures ignoring the mild heteroscedasticity will perform well. However, they will perform very poorly in the situation in which serious heteroscedasticity is present. Thus our proposed methods are very valuable in practice, especially when the two variances are quite different.

Acknowledgements

We would like to thank the Editorial Board Member (EOBM) and two referees for their kindly help and constructive comments which led to a substantial improvement of the paper.

References

Berger, R.L., Hsu, J.C., 1996. Bioequivalence trials, intersection union tests and equivalence confidence sets. Statist. Sci. 11 (4), 283–319.

Chow, S.C., Liu, J.P., 1992. Design and Analysis of Bioavailability and Bioequivalence Studies. Marcel Dekker, Inc, New York.

Cox, C.P., 1985. Interval estimates for the ratio of the means of two normal populations with variances related to means. Biometrics 41, 261–265.

Fieller, E.C., 1944. A fundamental formula in the statistics of biological assay, and some applications. Q. J. Pharm. Pharmacol. 17, 117–123.

Fieller, E.C., 1954. Some problems in interval estimation. J. Roy. Statist. Soc. Ser. B 16, 175–185.

Finney, D.J., 1978. Statistical method in biological assay, 3rd Edition. Griffin, London.

Hwang, J.T., 1995. Fieller's problems and resampling techniques. Statistica Sinica 5 (1), 161–171.

Jarvis, M.J., Tunstall-Pedoe, H., Feyerabend, C., Vesey, C., Saloojee, Y., 1987. Comparison of tests used to distinguish smokers from nonsmokers. Amer. J. Public Health 77, 1435–1438.

Koschat, M.A., 1987. A characterization of the Fieller solution. Ann. Statist. 15 (1), 462–468.

Pagano, M., Gauvreau, K., 1993. Principles of Biostatistics. Cuxbury, California.

Rao, C.R., 1973. Linear Statistical Inference and its Applications, Second Edition. Wiley, New York.

Tsui, K.W., Weerahandi, S., 1989. Generalized p-values in significance testing of hypotheses in the presence of nuisance parameters. J. Amer. Statist. Assoc. 84, 602–607.

Weerahandi, S., 1993. Generalized confidence intervals. J. Amer. Statist. Assoc. 88, 899–905.