Available at WWW.MATHEMATICSWEB.ORG
POWERED BY SCIENCE DIRECT
Journal of Multivariate Analysis 88 (2004) 177–189
Journal of Multivariate Analysis
http://www.elsevier.com/locate/jmva

Generalized p-values and generalized confidence regions for the multivariate Behrens–Fisher problem and MANOVA

Jinadasa Gamage,[a] Thomas Mathew,[b,*] and Samaradasa Weerahandi[c]

[a] *Illinois State University, IL, USA*
[b] *Department of Mathematics and Statistics, University of Maryland, 1000 Hilltop Circle, Baltimore, MD 21250, USA*
[c] *Telcordia Technologies, Morristown, NJ, USA*

Received 29 March 2001

Abstract

For two multivariate normal populations with unequal covariance matrices, a procedure is developed for testing the equality of the mean vectors based on the concept of *generalized p-values*. The generalized p-values we have developed are functions of the sufficient statistics. The computation of the generalized p-values is discussed and illustrated with an example. Numerical results show that one of our generalized p-value test has a type I error probability not exceeding the nominal level. A formula involving only a finite number of chi-square random variables is provided for computing this generalized p-value. The formula is useful in a Bayesian solution as well. The problem of constructing a confidence region for the difference between the mean vectors is also addressed using the concept of *generalized confidence regions*. Finally, using the generalized p-value approach, a solution is developed for the heteroscedastic MANOVA problem.

AMS 1991 subject classifications: 62H15; 62F25

Keywords: Generalized confidence region; Generalized p-value; Generalized test variable; Heteroscedasticity; MANOVA; Type I error

*Corresponding author. Fax: 410-455-1066.
E-mail address: mathew@math.umbc.edu (T. Mathew).

0047-259X/03/$ - see front matter © 2003 Elsevier Inc. All rights reserved.
doi:10.1016/S0047-259X(03)00065-4

1. Introduction

1.1. Background

The multivariate Behrens–Fisher problem deals with statistical inference concerning the difference between the mean vectors of two multivariate normal populations with unequal covariance matrices. In this context, several exact and approximate tests are available for testing the equality of the mean vectors. Some of the available frequentist solutions are reviewed and compared in the article by Christensen and Rencher [1]. An exact Bayesian solution to the problem is given in [4]. Non-Bayesian exact solutions to the problem are not based on sufficient statistics. In this paper we develop some non-Bayesian exact solutions based on sufficient statistics. The solutions are exact in that they are based on exact probability statements as required in the context of generalized inference.

In the case of hypothesis testing, multivariate Behrens–Fisher problem can be formulated as follows. Consider two p-variate normal populations with distributions $N(\mu_1, \Sigma_1)$ and $N(\mu_2, \Sigma_2)$, where μ_1 and μ_2 are unknown $p \times 1$ vectors and Σ_1 and Σ_2 are unknown $p \times p$ positive definite matrices. Let $X_{\alpha 1} \sim N(\mu_1, \Sigma_1)$, $\alpha = 1, 2, \ldots, n_1$, and $X_{\alpha 2} \sim N(\mu_2, \Sigma_2)$, $\alpha = 1, 2, \ldots, n_2$, denote random samples. We are interested in the testing problem

$$H_0 : \mu_1 = \mu_2 \text{ vs. } H_1 : \mu_1 \neq \mu_2. \tag{1.1}$$

For $i = 1, 2$, let

$$\bar{X}_i = \frac{1}{n_i} \sum_{\alpha=1}^{n_i} X_{\alpha i}, \quad A_i = \sum_{\alpha=1}^{n_i} (X_{\alpha i} - \bar{X}_i)(X_{\alpha i} - \bar{X}_i)'$$
$$\text{and } S_i = A_i/(n_i - 1), \; i = 1, 2. \tag{1.2}$$

Then $\bar{X}_1$, $\bar{X}_2$, A_1 and A_2 are independent random variables having the distributions

$$\bar{X}_i \sim N\left(\mu_i, \frac{\Sigma_i}{n_i}\right), \quad \text{and} \quad A_i \sim W_p(n_i - 1, \Sigma_i), \; i = 1, 2, \tag{1.3}$$

where $W_p(r, \Sigma)$ denotes the p-dimensional Wishart distribution with df $= r$ and scale matrix Σ. It is desired to get a solution to the testing problem (1.1) based on the sufficient statistics in (1.3). Based on numerical results concerning type I error and power, Christensen and Rencher [1] recommend an approximate solution due to Kim [5] for testing (1.1). The numerical results in [1] show that many approximate tests have type I errors exceeding the nominal level. For Kim's [1] test, type I errors were below the nominal level most of the time even though it exceeded the nominal level in a few cases.

The purpose of the present article is to explore the concept of the generalized p-value for testing the hypotheses in (1.1). We shall also consider the problem of deriving a confidence region for $\mu_1 - \mu_2$ using the concept of a generalized confidence region. More generally, to encourage much needed further research in this direction, we shall also apply the generalized p-value to address the MANOVA problem under heteroscedasticity, i.e., the problem of testing if several multivariate normal means are equal, when the covariance matrices are different. Since there is no unique solution to the MANOVA problem even in the homoscedastic case, the solution we provide is expected to indicate the approach that one needs to take in deriving other solutions which may be more suitable under alternative desired properties.

The concepts of generalized p-values and generalized confidence regions were introduced by Tsui and Weerahandi [9] and Weerahandi [10]. Since then the concepts have been applied to solve a number of problems where conventional methods are difficult to apply or fail to provide exact solutions. For a discussion of several applications, the reader is referred to the book by Weerahandi [11]. We shall now briefly describe the generalized p-value and the generalized confidence interval as they apply to univariate problems.

1.2. Generalized p-values and generalized confidence intervals

A set up where the generalized p-value maybe defined is as follows. Let X be a random variable whose distribution depends on the parameters (θ, δ), where θ is a scalar parameter of interest and δ represents nuisance parameters. Suppose we want to test $H_0 : \theta \leqslant \theta_0$ vs. $H_1 : \theta > \theta_0$, where θ_0 is a specified value.

Let x denote the observed value of X and consider the *generalized test variable* $T(X; x, \theta, \delta)$, which also depends on the observed value and the parameters, and satisfies the following conditions:

(a) The distribution of $T(X; x, \theta_0, \delta)$ is free of the nuisance parameter δ.

(b) The observed value of $T(X; x, \theta_0, \delta)$, i.e., $T(x; x, \theta_0, \delta)$, is free of δ.

(c) $P[T(X; x, \theta, \delta) \geqslant t]$ is nondecreasing in θ, for fixed x and δ. (1.4)

Under the above conditions, the generalized p-value is defined by

$$P[T(X; x, \theta_0, \delta) \geqslant t],$$

where $t = T(x; x, \theta_0, \delta)$.

In the same set up, suppose $T_1(X; x, \theta, \delta)$ satisfies the following conditions:

(a) The distribution of $T_1(X; x, \theta, \delta)$ does not depend on any unknown parameters,

(b) The observed value of $T_1(X; x, \theta, \delta)$ is free of the nuisance parameter δ. (1.5)

Now let t_1 and t_2 be such that

$$P[t_1 \leqslant T_1(X; x, \theta, \delta) \leqslant t_2] = \gamma.$$

Then $\{\theta : t_1 \leqslant T_1(x; x, \theta, \delta) \leqslant t_2\}$ is a $100\gamma\%$ generalized confidence interval for θ.

In applications, an important point to note is that type I error probability of a test based on the generalized p-value, and the coverage probability of a generalized confidence interval, may depend on the nuisance parameters. For further details and for applications, we refer to the book by Weerahandi [11]. For the univariate Behrens–Fisher problem, a test based on the generalized p-value is given in [9]. For the multivariate Behrens–Fisher problem, an upper bound for the generalized p-value is given in [2].

1.3. Overview

In the next section, we have constructed a generalized test variable for the testing problem (1.1). The observed value of the generalized test variable is a function of the sufficient statistics in (1.3). In fact, the observed value is $(\bar{x}_1 - \bar{x}_2)'(\frac{s_1}{n_1} + \frac{s_2}{n_2})^{-1}(\bar{x}_1 - \bar{x}_2)$, a very natural quantity for the testing problem (1.1). Here $\bar{x}_1$ and $\bar{x}_2$ denote the observed values of $\bar{X}_1$ and $\bar{X}_2$, respectively, and s_1 and s_2 denote the observed values of the sample covariance matrices S_1 and S_2 in (1.2). The computation of the generalized p-value is discussed in Section 3. Some numerical results are given in Section 4 on type I error probabilities of the test based on the generalized p-value. It turns out that type I error probabilities of the proposed test do not exceed the nominal level. In other words, the test based on the generalized p-value, which is an exact probability of a well-defined extreme region, also provides a conservative test in the classical sense. An example is taken up in Section 5. The construction of a generalized confidence region is outlined in Section 6. In Section 7, we provide a solution to the MANOVA problem under unequal covariance matrices. Some concluding remarks appear in Section 8.

2. The generalized p-value for the Behrens–Fisher problem

Let $\bar{X}_1$ and $\bar{X}_2$ be the sample means and S_1 and S_2 be the sample covariance matrices defined in (1.2). Furthermore, let $\bar{x}_1$, $\bar{x}_2$, s_1 and s_2 denote the corresponding observed values. Define

$$\begin{aligned}
Y_1 &= \left(\frac{s_1}{n_1} + \frac{s_2}{n_2}\right)^{-1/2} \bar{X}_1, \quad Y_2 = \left(\frac{s_1}{n_1} + \frac{s_2}{n_2}\right)^{-1/2} \bar{X}_2, \\
V_1 &= \left(\frac{n_1 - 1}{n_1}\right)\left(\frac{s_1}{n_1} + \frac{s_2}{n_2}\right)^{-1/2} S_1 \left(\frac{s_1}{n_1} + \frac{s_2}{n_2}\right)^{-1/2}, \\
V_2 &= \left(\frac{n_2 - 1}{n_2}\right)\left(\frac{s_1}{n_1} + \frac{s_2}{n_2}\right)^{-1/2} S_2 \left(\frac{s_1}{n_1} + \frac{s_2}{n_2}\right)^{-1/2}, \\
\theta_1 &= \left(\frac{s_1}{n_1} + \frac{s_2}{n_2}\right)^{-1/2} \mu_1, \quad \theta_2 = \left(\frac{s_1}{n_1} + \frac{s_2}{n_2}\right)^{-1/2} \mu_2, \\
\Lambda_1 &= \left(\frac{s_1}{n_1} + \frac{s_2}{n_2}\right)^{-1/2} \frac{\Sigma_1}{n_1} \left(\frac{s_1}{n_1} + \frac{s_2}{n_2}\right)^{-1/2}, \\
\Lambda_2 &= \left(\frac{s_1}{n_1} + \frac{s_2}{n_2}\right)^{-1/2} \frac{\Sigma_2}{n_2} \left(\frac{s_1}{n_1} + \frac{s_2}{n_2}\right)^{-1/2},
\end{aligned} \tag{2.1}$$

where by $A^{1/2}$ we mean the positive definite square root of the positive definite matrix A, and $A^{-1/2} = (A^{1/2})^{-1}$. Then

$$\begin{aligned} &Y_1 \sim N(\theta_1, \Lambda_1), \quad Y_2 \sim N(\theta_2, \Lambda_2), \quad V_1 \sim W_p(n_1 - 1, \Lambda_1), \\ &\quad V_2 \sim W_p(n_2 - 1, \Lambda_2), \end{aligned} \tag{2.2}$$

and the testing problem (1.1) can be expressed as

$$H_0 : \theta_1 = \theta_2, \text{ vs. } H_1 : \theta_1 \neq \theta_2. \tag{2.3}$$

Let y_1, y_2, v_1 and v_2 denote observed values of Y_1, Y_2, V_1 and V_2, respectively. Note that these observed values are obtained by replacing $\bar{X}_1$, $\bar{X}_2$, S_1 and S_2 by the corresponding observed values (namely, $\bar{x}_1$, $\bar{x}_2$, s_1 and s_2) in the expressions for Y_1, Y_2, V_1 and V_2 in (2.1). Then we have the distributions

$$\begin{aligned} &Z = (\Lambda_1 + \Lambda_2)^{-1/2}(Y_1 - Y_2) \sim N(0, I_p), \text{ under } H_0, \\ &R_1 = [v_1^{-1/2}\Lambda_1 v_1^{-1/2}]^{-1/2}[v_1^{-1/2}V_1 v_1^{-1/2}][v_1^{-1/2}\Lambda_1 v_1^{-1/2}]^{-1/2} \sim W_p(n_1 - 1, I_p), \\ &R_2 = [v_2^{-1/2}\Lambda_2 v_2^{-1/2}]^{-1/2}[v_2^{-1/2}V_2 v_2^{-1/2}][v_2^{-1/2}\Lambda_2 v_2^{-1/2}]^{-1/2} \sim W_p(n_2 - 1, I_p). \end{aligned} \tag{2.4}$$

Now define

$$T_1 = Z'[v_1^{1/2}R_1^{-1}v_1^{1/2} + v_2^{1/2}R_2^{-1}v_2^{1/2}]Z. \tag{2.5}$$

Note that given R_1 and R_2, T_1 is a positive definite quadratic form in $Y_1 - Y_2$ (where Y_1 and Y_2 have the distributions in (2.2)), and T_1 is stochastically larger under H_1 than under H_0. Also, since the distribution of Z, R_1 and R_2 are free of any unknown parameters (under H_0) and since these quantities are independent, it follows that the distribution of T_1 is free of any unknown parameters (under H_0). Using the definition of Z, R_1 and R_2, we conclude that

$$\begin{aligned} \text{the observed value of } T_1 &= (y_1 - y_2)'(y_1 - y_2) \\ &= (\bar{x}_1 - \bar{x}_2)'\left(\frac{s_1}{n_1} + \frac{s_2}{n_2}\right)^{-1}(\bar{x}_1 - \bar{x}_2) \\ &= t_1 \text{ (say)}, \end{aligned} \tag{2.6}$$

which also does not depend on any unknown parameters. In other words, T_1 is a generalized test variable satisfying conditions similar to those in (1.4). Hence the generalized p-value is given by

$$P(T_1 \geqslant t_1 | H_0), \tag{2.7}$$

where T_1 and t_1 are given in (2.5) and (2.6), respectively.

A natural requirement is that the generalized p-value be invariant under the transformation

$$\begin{aligned} &(\bar{X}_1, \bar{X}_2, \bar{x}_1, \bar{x}_2) \to (P\bar{X}_1, P\bar{X}_2, P\bar{x}_1, P\bar{x}_2), \\ &(S_1, S_2, s_1, s_2) \to (PS_1P', PS_2P', Ps_1P', Ps_2P'), \end{aligned} \tag{2.8}$$

where P is a $p \times p$ nonsingular matrix. We shall now verify that this invariance holds. Obviously, the observed value in (2.6) is invariant under the above nonsingular transformation. Thus, it remains to verify the invariance of the distribution of T_1, under H_0. For this, we shall use representation (3.9) given in the next section. Note that in (3.9), Q_1, Q_2 and the Z_{0i}^2's are all independent chi-square random variables (under H_0). Furthermore, the d_i's in (3.9) are the eigenvalues of v_1. Thus, in order to conclude the invariance of the distribution of T_1 under H_0, it is enough to show that the d_i's are invariant under the transformation in (2.8). Towards this, we note from (2.1) that the eigenvalues of v_1 are the same as the eigenvalues of $\frac{n_1-1}{n_1}s_1(\frac{s_1}{n_1}+\frac{s_2}{n_2})^{-1}$, and the eigenvalues of the latter matrix are clearly invariant under the transformation in (2.8). Thus we conclude that the generalized p-value is invariant under the transformation in (2.8).

By doing straightforward algebra, it can also be verified that in the univariate case, our generalized p-value will reduce to that in [9], for the univariate Behrens–Fisher problem.

3. Computation of the generalized p-value

We shall now give a suitable representation for T in (2.5), which can be used for the computation of the generalized p-value. Let

$$W_1 = v_1^{-1/2} R_1 v_1^{-1/2}, \quad W_2 = v_2^{-1/2} R_2 v_2^{-1/2}, \tag{3.1}$$

where v_1 and v_2 denote the observed values of V_1 and V_2 in (2.1), and R_1 and R_2 are defined in (2.4). Then

$$W_1 \sim W_p(n_1-1, v_1^{-1}), \quad W_2 \sim W_p(n_2-1, v_2^{-1}), \tag{3.2}$$

and we have the representation

$$T_1 = Z'(W_1^{-1} + W_2^{-1})Z. \tag{3.3}$$

3.1. Equivalent representations of T_1

Define

$$Q_1 = \frac{Z'v_1Z}{Z'W_1^{-1}Z}, \quad Q_2 = \frac{Z'v_2Z}{Z'W_2^{-1}Z}. \tag{3.4}$$

Since $W_1 \sim W_p(n_1-1, v_1^{-1})$, it follows that, conditionally given Z, $Q_1 \sim \chi^2$, df $= n_1 - p$, which is also its unconditional distribution. Similarly, $Q_2 \sim \chi^2$, df $= n_2 - p$. Thus,

$$T_1 = Z'(W_1^{-1} + W_2^{-1})Z = \frac{1}{Q_1} Z'v_1Z + \frac{1}{Q_2} Z'v_2Z, \tag{3.5}$$

where $Z \sim N(0, I_p)$, $Q_1 \sim \chi^2$ (df $= n_1 - p$), $Q_2 \sim \chi^2$ (df $= n_2 - p$), and Z, Q_1 and Q_2 are independent. We note that the above representation for T_1 is similar to the expression (3.4) in [4, p. 147]. Hence the procedures mentioned in [4] can be used for computing the generalized p-value also. In particular, we have an expression to compute the generalized p-value involving the distribution of only univariate random variables. The expression given in [4] leads to an infinite series, although it reduces to a finite summation when a limited accuracy is required, as always the case. Now we will give another representation for T_1 involving a finite summation which is valid regardless of the required accuracy. From (2.1) we note that $\frac{v_1}{n_1-1}+\frac{v_2}{n_2-1}=I_p$. Hence $v_2=(n_2-1)(I_p-\frac{v_1}{n_1-1})$ and T_1 in (3.5) can be expressed as

$$T_1=\frac{1}{Q_1}Z'v_1Z+\frac{n_2-1}{Q_2}Z'\left(I_p-\frac{v_1}{n_1-1}\right)Z. \tag{3.6}$$

Let F be a $p\times p$ orthogonal matrix and $D=\mathrm{diag}(d_1,d_2,\ldots,d_p)$ be a $p\times p$ diagonal matrix such that

$$v_1=FDF'. \tag{3.7}$$

The diagonal elements of D are clearly the eigenvalues of v_1. Then

$$Z_0=(Z_{01},Z_{02},\ldots,Z_{0p})'=F'Z\sim N(0,I_p),\ \text{ under } H_0, \tag{3.8}$$

where Z is defined in (2.4). Hence T_1 in (3.6) can be expressed as

$$T_1=\frac{1}{Q_1}\sum_{i=1}^{p}d_iZ_{0i}^2+\frac{n_2-1}{Q_2}\sum_{i=1}^{p}\left(1-\frac{d_i}{n_1-1}\right)Z_{0i}^2. \tag{3.9}$$

In (3.9), each Z_{0i}^2 $(i=1,2,\ldots,p)$ has a chi-square distribution with 1 df, under H_0, $Q_1\sim\chi^2$ (df $=n_1-p$), $Q_2\sim\chi^2$ (df $=n_2-p$), and Q_1, Q_2 and the Z_{0i}^2's are all independent. Representation (3.9) can be used for simulating the generalized p-value in (2.7).

4. A generalized confidence region

The definition of the generalized confidence region is given in Section 1.2 for the univariate case. We shall now use similar ideas to construct a generalized confidence region for $\mu_1-\mu_2$ based on the random variables in (1.3). Towards this, let Y_1, Y_2, V_1, V_2, θ_1, θ_2, Λ_1 and Λ_2 be as defined in (2.1), and let R_1 and R_2 be as defined in (2.4). Also let

$$\begin{aligned}Z_1&=(\Lambda_1+\Lambda_2)^{-1/2}[(Y_1-\theta_1)-(Y_2-\theta_2)]\sim N(0,I_p),\\ T&=Z_1'[v_1^{1/2}R_1^{-1}v_1^{1/2}+v_2^{1/2}R_2^{-1}v_2^{1/2}]Z_1.\end{aligned} \tag{4.1}$$

It is readily verified that the distribution of T is free of any unknown parameters and the observed value of T, say t is given by

$$t=[(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)]'\left(\frac{s_1}{n_1}+\frac{s_2}{n_2}\right)^{-1}[(\bar{x}_1-\bar{x}_2)-(\mu_1-\mu_2)]. \tag{4.2}$$

Let k_γ satisfy

$$P(T \leqslant k_\gamma) = \gamma. \tag{4.3}$$

Then

$$\left\{ \mu_1 - \mu_2 : [(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)]' \left(\frac{s_1}{n_1} + \frac{s_2}{n_2} \right)^{-1} [(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)] \leqslant k_\gamma \right\} \tag{4.4}$$

is a $100\gamma\%$ generalized confidence region for $\mu_1 - \mu_2$.

Note that T has a representation similar to that of T_1 in (3.9) and this can be used for the Monte Carlo estimation of k_γ satisfying (4.3). Also note that k_γ depends on the observed values s_1 and s_2. It can be shown that the generalized confidence region given by (4.4) is numerically equivalent to the Bayesian confidence region given by Johnson and Wearahandi [4], with a noninformative prior.

5. Manova under heteroscedasticity

Consider k independent multivariate normal populations $N(\mu_i, \Sigma_i)$, $i = 1, 2, \ldots, k$, and suppose we wish to test

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k. \tag{5.1}$$

Let $X_{\alpha i}$, $\alpha = 1, 2, \ldots, n_i$, denote a random sample of size n_i from $N(\mu_i, \Sigma_i)$. For $i = 1, 2, \ldots, k$, let

$$\bar{X}_i = \frac{1}{n_i} \sum_{\alpha=1}^{n_i} X_{\alpha i}, \quad A_i = \sum_{\alpha=1}^{n_i} (X_{\alpha i} - \bar{X}_i)(X_{\alpha i} - \bar{X}_i)' \quad \text{and} \quad S_i = A_i/(n_i - 1). \tag{5.2}$$

Then $\bar{X}_i$ and A_i are all independent random variables having the distributions

$$\bar{X}_i \sim N\left(\mu_i, \frac{\Sigma_i}{n_i} \right) \quad \text{and} \quad A_i \sim W_p(n_i - 1, \Sigma_i), \quad i = 1, 2, \ldots, k. \tag{5.3}$$

As before, $\bar{x}_i$ and s_i will denote the observed values of $\bar{X}_i$ and S_i, respectively. For $i = 1, 2, \ldots, k$, define

$$R_i^* = [s_i^{-1/2} \Sigma_i s_i^{-1/2}]^{-1/2} [s_i^{-1/2} S_i s_i^{-1/2}] [s_i^{-1/2} \Sigma_i s_i^{-1/2}]^{-1/2}, \tag{5.4}$$

so that $(n_i - 1) R_i^* \sim W_p(n_i - 1, I_p)$. Under H_0 given in (5.1), let μ denote the common value of the μ_i's. If the Σ_i's are known, then

$$\hat{\mu} = \left(\sum_{i=1}^{k} n_i \Sigma_i^{-1} \right)^{-1} \sum_{i=1}^{k} n_i \Sigma_i^{-1} \bar{X}_i$$

is the best linear unbiased estimator of μ and

$$\tilde{T}_0 = \tilde{T}_0(\bar{X}_1, \dots, \bar{X}_k; \Sigma_1, \dots, \Sigma_k) = \sum_{i=1}^{k} n_i(\bar{X}_i - \hat{\mu})'\Sigma_i^{-1}(\bar{X}_i - \hat{\mu})$$

has a chi-square distribution with $\mathrm{df} = p(k-1)$ under H_0. Now define the generalized test variable

$$T_0 = \frac{\tilde{T}_0(\bar{X}_1, \dots, \bar{X}_k; \Sigma_1, \dots, \Sigma_k)}{\tilde{T}_0(\bar{x}_1, \dots, \bar{x}_k; \frac{1}{n_1} s_1^{1/2} R_1^{*-1} s_1^{1/2}, \dots, \frac{1}{n_k} s_k^{1/2} R_k^{*-1} s_k^{1/2})}. \tag{5.5}$$

It is readily verified that T_0 satisfies all the properties required of a generalized test variable and the observed value of T_0 is one. Hence a generalized p-value can be computed as

$$P(T_0 \geqslant 1 | H_0). \tag{5.6}$$

For the case of two populations, the above generalized p-value does not reduce to that in Section 2. In fact, when $k = 2$, one can use the matrix identity

$$(n_1\Sigma_1^{-1} + n_2\Sigma_2^{-1})^{-1} = \frac{\Sigma_1}{n_1}\left(\frac{\Sigma_1}{n_1} + \frac{\Sigma_2}{n_2}\right)^{-1}\frac{\Sigma_2}{n_2} = \frac{\Sigma_2}{n_2}\left(\frac{\Sigma_1}{n_1} + \frac{\Sigma_2}{n_2}\right)^{-1}\frac{\Sigma_1}{n_1}$$

and conclude that T_0 given in (5.5) reduces to the generalized test variable T_2 given by

$$T_2 = \frac{(\bar{X}_1 - \bar{X}_2)'(\frac{\Sigma_1}{n_1} + \frac{\Sigma_2}{n_2})^{-1}(\bar{X}_1 - \bar{X}_2)}{(\bar{x}_1 - \bar{x}_2)'[\frac{1}{n_1} s_1^{1/2} R_1^{*-1} s_1^{1/2} + \frac{1}{n_2} s_2^{1/2} R_2^{*-1} s_2^{1/2}]^{-1}(\bar{x}_1 - \bar{x}_2)}. \tag{5.7}$$

This gives another solution to the Behrens–Fisher problem. The generalized p-value based on the generalized test variable T_2 in (5.7) is given by

$$P(T_2 \geqslant 1 | H_0). \tag{5.8}$$

An unfortunate feature of the generalized p-values in (5.6) and (5.7) is that they are not invariant under nonsingular transformations, unlike the generalized p-value in Section 2. For the MANOVA problem under heteroscedasticity, the derivation of a generalized p-value that also enjoys such an invariance property remains to be investigated. Since there are multiple solutions even in the case of homoscedastic MANOVA problem, there is a need for further research to obtain solutions that enjoy the invariance as well.

6. Numerical results

As pointed out in the introduction, any test for the Behrens–Fisher problem including the test based on the generalized p-value has type I error probability and power both depending on the nuisance parameters. For example, if we decide to reject the null hypothesis when the generalized p-value is less than 0.05, type I error probability for such a test procedure will not be exactly 0.05, since the generalized

p-value does not have a uniform distribution (under the null hypothesis) unlike the usual p-value. Hence, it is of interest to study the behavior of type I error probability of the test based on the generalized p-value. In particular, we would like to see if type I error probability of the test can exceed the nominal level.

For the tests based on the generalized p-values in (2.7) and (5.8), some simulated type I error probabilities are given in Tables 1 and 2 in the bivariate case ($p = 2$). Shown also in the tables are the size performance of two widely used tests. In many applications, practitioners simply ignore the heteroscedasticity and apply the classical Hotelling's T^2 test; this test we denote as HOTE. When the hetoroscedasticity is very serious some practitioners replace the unknown covariances by the sample covariances and apply the classical chi-squared test, which is valid when the covariances are known; this test we denote as CHI. A comparison of some other tests are given by Christensen and Rencher [1].

Each test is carried out at the nominal level of 5%. In comparing these alternative tests, we first chose $\Sigma_1 = I_2$ and Σ_2 to be diagonal for our simulations. This is no loss of generality for the test based on (2.7), since it is invariant under nonsingular

Table 1
Simulated type I error probabilities of tests: case $\Sigma_1 = I_2$, $\Sigma_2 = \frac{n_2}{n_1} a I_2$

a	GP-1	GP-2	CHI	HOTE
$n_1 = 10$, $n_2 = 5$				
3	0.016	0.031	0.176	0.066
9	0.024	0.062	0.231	0.150
15	0.026	0.072	0.253	0.197
25	0.034	0.075	0.272	0.244
50	0.039	0.083	0.284	0.289
100	0.043	0.083	0.293	0.320
500	0.049	0.086	0.302	0.349
$n_1 = 10$, $n_2 = 10$				
3	0.027	0.040	0.118	0.057
9	0.035	0.051	0.133	0.070
15	0.038	0.055	0.138	0.075
25	0.043	0.058	0.143	0.079
50	0.046	0.063	0.146	0.083
100	0.050	0.070	0.147	0.084
500	0.049	0.074	0.149	0.085
$n_1 = 10$, $n_2 = 20$				
3	0.033	0.042	0.081	0.010
9	0.040	0.051	0.085	0.007
15	0.036	0.055	0.087	0.007
25	0.036	0.058	0.089	0.007
50	0.036	0.060	0.090	0.007
100	0.036	0.059	0.091	0.007
500	0.036	0.059	0.092	0.007

Table 2
Simulated type I error probabilities of tests: case $\Sigma_1 = I_2$, $\Sigma_2 = \text{diag}(1, b)$

b	GP-1	GP-2	CHI	HOTE
$n_1 = 10$, $n_2 = 5$				
4	0.016	0.034	0.188	0.085
9	0.020	0.042	0.203	0.117
16	0.022	0.044	0.208	0.137
25	0.024	0.046	0.211	0.149
$n_1 = 10$, $n_2 = 10$				
4	0.027	0.040	0.116	0.056
9	0.029	0.042	0.119	0.060
16	0.027	0.042	0.121	0.060
25	0.029	0.040	0.121	0.060
$n_1 = 10$, $n_2 = 20$				
4	0.036	0.042	0.089	0.027
9	0.037	0.041	0.088	0.025
16	0.037	0.039	0.088	0.024
25	0.025	0.039	0.088	0.023

transformations. (There exists a nonsingular matrix P such that $P'\Sigma_1 P = I_p$ and $P'\Sigma_2 P$ is diagonal; see [6, p. 41]). We also made the choices $\Sigma_2 = \frac{n_2}{n_1} aI_2$ and $\Sigma_2 = \text{diag}(1, b)$, for various choices of the scalars a and b. For $\Sigma_1 = I_2$ and for such choices of Σ_2, type I error probability of the test based on (2.7) was simulated as follows. First we generated one set of values of $\bar{X}_1 - \bar{X}_2 \sim N(0, \frac{1}{n_1}\Sigma_1 + \frac{1}{n_2}\Sigma_2)$, and S_1 and S_2, following the definitions in (1.2) and (1.3). Let the generated values be denoted by $\bar{x}_1 - \bar{x}_2$, s_1 and s_2. The quantity $t = (\bar{x}_1 - \bar{x}_2)'(\frac{s_1}{n_1} + \frac{s_2}{n_2})^{-1}(\bar{x}_1 - \bar{x}_2)$ and the eigenvalues d_1 and d_2 of the 2×2 matrix $\frac{n_1 - 1}{n_1} s_1 (\frac{s_1}{n_1} + \frac{s_2}{n_2})^{-1}$, were then computed. Using 1000 simulations, we then computed the generalized p-value $P(T_1 \geqslant t_1)$ where representation (3.9) for T_1 was used for the simulation (with $p = 2$). Let $p(\bar{x}_1 - \bar{x}_2, s_1, s_2)$ denote the generalized p-value so obtained. Thousand observed values of $\bar{X}_1 - \bar{X}_2$, S_1 and S_2 were generated and the corresponding generalized p-values were similarly calculated. The proportion of such p-values below 0.05 gives the simulated type I error probability of the test based on (2.7). The simulated type I error probability of the test based on (5.8) was similarly obtained. In Tables 1 and 2, GP-1 and GP-2 refers to the tests based on the generalized p-values in (2.7) and (5.8), respectively.

From the numerical results in Tables 1 and 2, it appears that the invariant test based on the generalized p-value in (2.7) has a type I error probability not exceeding the nominal level. In a number of other problems where the generalized p-value has been used, a similar conclusion has been arrived at by Thursby [8], Zhou and Mathew [12] and Gamage and Weerahandi [3]. On the other hand, type I error of the noninvariant test based on the generalized p-value in (5.8) sometimes exceeds the intended level of 5%. This fact, along with its noninvariance makes this test somewhat unattractive for practical use. Notice, however, that the test is less conservative than the invariant test. It seems that the widely used Hotelling's T^2 test and the chi-squared tests have very poor size poor performance under heteroscedasticity. This is also noted in [1]. Notice that in the case $\Sigma_1 = I_2$, $\Sigma_2 = \frac{n_2}{n_1} a I_2$ and with small sample sizes, type I error of the former can be as large 35% and that of latter can be as large as 30%, both at unacceptable levels. In the case $\Sigma_1 = I_2$, $\Sigma_2 = \mathrm{diag}(1, b)$ the situation is less serious, but with small samples, type I error of the former is as large 20% and that of latter is as large as 15%. As a consequence of the assumption of known covariances, the chi-squared test has type I error exceeding the 5% level in all cases. In addition to having better size performance than these tests, both GP-1 and GP-2 have the additional advantage of being based on exact probabilities of well defined extreme regions of the sample space formed by sufficient statistics.

7. An example

We shall now compute the generalized p-value based on some bivariate data given in [7, p. 54]. The data are measurements of the thorax length (in microns) and elytra length (in 0.01 mm) on two flea beetle species. Based on the data from Table 3.3 of [7], we have $n_1 = 10$, $n_2 = 13$,

$$\bar{x}_1 = \begin{pmatrix} 194.9 \\ 263.4 \end{pmatrix}, \quad \bar{x}_2 = \begin{pmatrix} 178.46154 \\ 292.92308 \end{pmatrix},$$

$$s_1 = \begin{pmatrix} 330.32222 & 325.26667 \\ 325.26667 & 354.71111 \end{pmatrix}, \quad s_2 = \begin{pmatrix} 109.26923 & 189.78846 \\ 189.78846 & 505.41026 \end{pmatrix}.$$

Using the expression in (2.1), we have

$$v_1 = \begin{pmatrix} 6.97944 & 1.8825 \\ 1.8825 & 2.08142 \end{pmatrix}$$

and the eigenvalues of v_1 are $d_1 = 7.61938$ and $d_2 = 1.44148$. Also, from (2.6), $t = 108.56002$. The generalized p-value based on (2.7) computed using 20,000 simulations turned out to be less than 0.00001. The generalized p-value based on (5.8) computed using 20,000 simulations turned out to be around 0.00001. Kim's (1992) test gave a p-value of 6.13×10^{-10}. Thus there is strong evidence to reject the equality of the mean vectors.

8. Concluding remarks

The concepts of the generalized p-value and generalized confidence interval have been successfully applied to solve many univariate hypothesis testing problems and interval estimation problems, since they were introduced by Tsui and Weerahandi [9] and Weerahandi [11]. Typically, the generalized p-value and the generalized confidence interval were found to be fruitful for problems where conventional frequentist procedures were nonexistent or were difficult to obtain. For the classical multivariate Behrens–Fisher problem, we have succeeded in obtaining generalized p-values for testing the equality of the mean vectors, and generalized confidence regions for the difference between the mean vectors. The associated computational issues are also addressed. Our simulation results support the expected result that, as in the univariate case, the invariant test based on the generalized p-value approach assures the level of the test in all cases. We have succeeded in providing a solution to a natural extension of the Behrens–Fisher problem, namely to the MANOVA problem under unequal covariance matrices. The particular test we have derived is not invariant and hence there is a need for further research in this direction to derive solutions with various desirable properties. It is expected that, for the case of heteroscedastic MANOVA, the generalized p-value approach can help derive all counterparts of the solution of the classical MANOVA problem.

References

[1] W.F. Christensen, A.C. Rencher, A comparison of type I error rates and power levels for seven solutions to the multivariate Beherns–Fisher problem, Comm. Statist. Simulation Comput. 26 (1997) 1251–1273.

[2] J. Gamage, Generalized p-values and the multivariate Beherns–Fisher problem, Linear Algebra Appl. 253 (1997) 369–377.

[3] J. Gamage, S. Weerahandi, Size performance of some tests in one-way ANOVA, Comm. Statist. Simulation Comput. 27 (1998) 625–640.

[4] R.A. Johnson, S. Weerahandi, A Bayesian solution to the multivariate Beherns–Fisher problem, J. Amer. Statist. Assoc. 83 (1998) 145–149.

[5] S. Kim, A practical solution to the multivariate Beherns–Fisher problem, Biometrika 79 (1992) 171–176.

[6] C.R. Rao, Linear Statistical Inference and its Applications, Wiley, New York, 1973.

[7] M.S. Srivastava, E.M. Carter, An Introduction to Applied Multivariate Statistics, North-Holland, New York, 1983.

[8] J.G. Thursby, A comparison of several exact and approximate tests for structural shift under heteroscedasticity, J. Econometrics 53 (1992) 363–386.

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

[10] S. Weerahandi, Generalized confidence intervals, J. Amer. Statist. Assoc. 88 (1993) 899–905.

[11] S. Weerahandi, Exact Statistical Methods for Data Analysis, Springer, New York, 1995.

[12] L. Zhou, T. Mathew, Some tests for variance components using generalized p-values, Technometrics 36 (1994) 394–402.