J. Japan Statist. Soc.
Vol. 32 No. 2 2002 183–192

NEW CRITERIA FOR TESTS OF DIMENSIONALITY UNDER ELLIPTICAL POPULATIONS

Kiyotaka Yoshida*, Hideyuki Imai* and Yoshiharu Sato*

We consider tests of dimensionality in the multivariate analysis of variance (MANOVA). Three types of test criteria (Likelihood-Ratio-type, Lawley-Hotelling-type and Bartlett-Nanda-Pillai-type) are popular. As is well known, their null distributions depend on nuisance parameters. When a sample size is large, these criteria are distributed approximately according to chi-squared distributions. However, when the sample size is small, the effect of the nuisance parameters cannot be ignored. Under normal populations, other criteria that do not depend on nuisance parameters were proposed. These criteria are also upper limits for the null distributions of LR-type and LH-type. Under elliptical populations, modified test criteria with a better chi-squared approximation were proposed in the case of a large sample. In this paper, we generalize Schott's results under elliptical populations and obtain new test criteria that do not depend on nuisance parameters.

Key words and phrases: Elliptical distribution, dimensionality, multivariate analysis of variance, nuisance parameter.

1. Introduction

It is of interest to infer dimensionality in the one-way MANOVA model. Dimensionality means the number of discriminant functions necessary to describe group differences. This is determined as the dimension of a hyperplane formed by mean vectors.

The following is a formulation for tests of dimensionality (a fuller account can be found in Backhouse and McKay (1982)).

Let $\boldsymbol{y}_{ij}$ $(i = 1, \ldots, k; j = 1, \ldots, n_i; \sum_{i=1}^{k} n_i = n)$ be a p-variate random vector that expresses an observation of the j-th object in the i-th population, where k is the number of populations, n_i $(i = 1, \ldots, k)$ is the sample size from the i-th population, and n is the total sample size. Here, suppose that $E[\boldsymbol{y}_{ij}] = \boldsymbol{\mu}_i$, $V[\boldsymbol{y}_{ij}] = \boldsymbol{\Sigma}$, and these parameters are unknown. We define the following variation matrices

$$(1.1)\qquad \boldsymbol{E} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}}_i)(\boldsymbol{y}_{ij} - \bar{\boldsymbol{y}}_i)', \qquad \boldsymbol{H} = \sum_{i=1}^{k} n_i(\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{y}})(\bar{\boldsymbol{y}}_i - \bar{\boldsymbol{y}})',$$

$$\boldsymbol{\Delta} = \sum_{i=1}^{k} n_i(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})(\boldsymbol{\mu}_i - \bar{\boldsymbol{\mu}})',$$

where $\bar{\boldsymbol{y}}_i = \frac{1}{n_i}\sum_{j=1}^{n_i} \boldsymbol{y}_{ij}$, $\bar{\boldsymbol{y}} = \frac{1}{n}\sum_{i=1}^{k} n_i\bar{\boldsymbol{y}}_i$, $\bar{\boldsymbol{\mu}} = \frac{1}{n}\sum_{i=1}^{k} n_i\boldsymbol{\mu}_i$.

Received December 7, 2001. Revised June 13, 2002. Accepted August, 8, 2002.

*Division of Systems and Information Engineering, Graduate School of Engineering, Hokkaido University, Kita 13, Nishi 8, Kitaku, Sapporo, 060-8628, Japan.

Let $\delta_1 \geq \cdots \geq \delta_p (\geq 0)$ be eigenvalues of $\boldsymbol{\Delta}\boldsymbol{\Sigma}^{-1}$, then null hypotheses, 'dimensionality $= s$' can be expressed as

$$H_s : \delta_1 \geq \cdots \geq \delta_s > \delta_{s+1} = \cdots = \delta_p = 0. \tag{1.2}$$

The essential part in (1.2) is $\delta_{s+1} = \cdots = \delta_p = 0$, while $\delta_1, \ldots, \delta_s$ are nuisance parameters. Now let $l_1 > \cdots > l_p > 0$ be eigenvalues of $\boldsymbol{H}\boldsymbol{E}^{-1}$. Then test criteria for H_s are obtained as functions of the $p - s$ smallest eigenvalues as follows:

$$\begin{aligned} &\text{LR-type: } T_1 = f_1(l_{s+1}, \ldots, l_p) = \log \prod_{i=s+1}^{p} (1 + l_i), \\ &\text{LH-type: } T_2 = f_2(l_{s+1}, \ldots, l_p) = \sum_{i=s+1}^{p} l_i, \\ &\text{BNP-type: } T_3 = f_3(l_{s+1}, \ldots, l_p) = \sum_{i=s+1}^{p} l_i/(1 + l_i). \end{aligned} \tag{1.3}$$

As is well known, however, the null distributions of these criteria depend on the nuisance parameters, $\delta_1, \ldots, \delta_s$. When the sample size n is small, the effect of the nuisance parameters is large. To avoid this, Schott (1984) suggested upper limits for the null distributions of T_1 and T_2. These upper limits do not depend on nuisance parameters. We are interested in elliptical population cases. In the canonical correlation model, Muirhead and Waternaux (1980) showed that the asymptotic distributions as $n \to \infty$ of usual criteria are chi-squared distributions under elliptical populations. Furthermore, Seo *et al.* (1995) proposed modified test criteria with a better chi-squared approximation in the canonical correlation and MANOVA model.

In this paper, our purpose is to extend Schott's results to elliptical cases and obtain new criteria that do not depend on nuisance parameters. Consequently, we can test the hypothesis H_s without the effect of the nuisance parameters under elliptical populations. This paper consists of six sections. In section 2, a brief definition of elliptical distributions is given, and their properties are described. In section 3, with generalized Schott's results, the upper bounds for a null distribution of T_i $(i = 1, 2, 3)$ are obtained under elliptical populations. In section 4, it is shown that the upper bounds for T_i $(i = 1, 2)$ are also upper limits under contaminated normal populations. In section 5, some numerical examples are given. Conclusions are given in section 6.

2. Elliptical distributions

In this section, a definition of elliptical distributions is given, and their properties are described (a fuller account can be found in Muirhead (1982) and Fang and Zhang (1988)).

Let $\boldsymbol{x}$ be a p-variate random vector. The distribution of $\boldsymbol{x}$ is called an elliptical distribution, and we write $\boldsymbol{x} \sim EC_p(\boldsymbol{\mu}, \boldsymbol{\Omega}, \psi)$ when a characteristic function (CF) is expressed as

$$\phi(\boldsymbol{t}) = \exp(i\boldsymbol{t}'\boldsymbol{\mu})\psi(\boldsymbol{t}'\boldsymbol{\Omega}\boldsymbol{t}), \tag{2.1}$$

where $\boldsymbol{t}$ is a p-variate vector, $\boldsymbol{\mu}$ is a p-variate location parameter, $\boldsymbol{\Omega}$ is a positive semidefinite scale parameter matrix of order p, and ψ is some differentiable function. Then the mean vector and the covariance matrix are $E[\boldsymbol{x}] = \boldsymbol{\mu}$ and $V[\boldsymbol{x}] = -2\psi'(\mathbf{0})\boldsymbol{\Omega}$, respectively, where ψ' is a derivative of ψ. If its probability density function (PDF) exists, it is expressed as

$$f(\boldsymbol{x}) = C_p|\boldsymbol{\Omega}|^{-1/2}g\{(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\},$$

where C_p is a normalized constant and g is some function. In (2.1), if $\psi(x) = \exp(-x/2)$, it is the same as CF of a normal distribution so that elliptical distributions are generalizations of normal distributions.

It is easy to show that $\boldsymbol{x} \sim EC_p(\boldsymbol{\mu}, \boldsymbol{\Omega}, \psi)$ implies $\boldsymbol{Fx} \sim EC_q(\boldsymbol{F\mu}, \boldsymbol{F\Omega F}', \psi)$, where $\boldsymbol{F}$ is a $q \times p$ matrix. This means that the distribution family remains invariant under a linear transformation.

Two examples of elliptical distributions are given below.

(a) Contaminated normal distribution:

$$\begin{aligned} f(\boldsymbol{x}) &= (1-\varepsilon)(2\pi)^{-(p/2)}|\boldsymbol{\Omega}|^{-(1/2)}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} \\ &+\varepsilon(2\pi\sigma^2)^{-(p/2)}|\boldsymbol{\Omega}|^{-(1/2)}\exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\} \quad (0 \le \varepsilon \le 1). \end{aligned}$$

(b) Multivariate t distribution:

$$f(\boldsymbol{x}) = \frac{\nu^{\nu/2}\Gamma\left(\frac{\nu+p}{2}\right)}{\pi^{p/2}\Gamma\left(\frac{\nu}{2}\right)}|\boldsymbol{\Omega}|^{-(1/2)}\{\nu + (\boldsymbol{x}-\boldsymbol{\mu})'\boldsymbol{\Omega}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}^{-(\nu+p)/2} \qquad (\nu > 0).$$

3. Upper bounds under elliptical populations

In this section, we show upper bounds for null distributions of $T_i (i = 1, 2, 3)$ under elliptical populations.

Let $\boldsymbol{y} = (\boldsymbol{y}'_{11}, \ldots, \boldsymbol{y}'_{kn_k})'$ be an np-variate random vector. Suppose $\boldsymbol{y} \sim EC_{np}(\boldsymbol{\mu}, \boldsymbol{I}_n \otimes \boldsymbol{\Omega}, \psi)$, where $\boldsymbol{\mu} = (\overbrace{\boldsymbol{\mu}'_1, \ldots, \boldsymbol{\mu}'_1}^{n_1}, \ldots, \overbrace{\boldsymbol{\mu}'_k, \ldots, \boldsymbol{\mu}'_k}^{n_k})'$ is an unknown location parameter vector and $\boldsymbol{\Omega}$ is an unknown positive definite scale parameter matrix. Then we can transform $\boldsymbol{y}$ to a canonical form as $\tilde{\boldsymbol{y}} \sim EC_{np}(\tilde{\boldsymbol{m}}, \boldsymbol{I}_n \otimes \boldsymbol{\Omega}, \psi)$. Here $\tilde{\boldsymbol{y}} = (\tilde{\boldsymbol{y}}'_1, \ldots, \tilde{\boldsymbol{y}}'_n)'$ and $\tilde{\boldsymbol{m}} = (\sqrt{n}\bar{\boldsymbol{\mu}}', \tilde{\boldsymbol{m}}'_2, \ldots, \tilde{\boldsymbol{m}}'_k, \mathbf{0}', \ldots, \mathbf{0}')'$ satisfy that $\boldsymbol{\Delta} = \sum_{i=2}^{k} \tilde{\boldsymbol{m}}_i \tilde{\boldsymbol{m}}'_i$, $\boldsymbol{H} = \sum_{i=2}^{k} \tilde{\boldsymbol{y}}_i \tilde{\boldsymbol{y}}'_i$, $\boldsymbol{E} = \sum_{i=k+1}^{n} \tilde{\boldsymbol{y}}_i \tilde{\boldsymbol{y}}'_i$. Then upper bounds for null distributions of $T_i (i = 1, 2, 3)$ in (1.3) for hypothesis H_s are obtained by the following theorem.

THEOREM 1. *Let* $\boldsymbol{z} = (\boldsymbol{z}_1^{(1)'}, \ldots, \boldsymbol{z}_{k-1-s}^{(1)'}, \boldsymbol{z}_1^{(2)'}, \ldots, \boldsymbol{z}_{n-k}^{(2)'})'$ *be an* $(n-1-s)(p-s)$*-variate random vector distributed according to* $EC_{(n-1-s)(p-s)}(\mathbf{0}, \boldsymbol{I}_{n-1-s} \otimes \boldsymbol{I}_{p-s}, \psi)$. *Let* $w_1 > \cdots > w_{p-s}$ *be eigenvalues of* $\boldsymbol{W}_H \boldsymbol{W}_E^{-1}$, *where*

$\boldsymbol{W}_H = \sum_{i=1}^{k-1-s} \boldsymbol{z}_i^{(1)} \boldsymbol{z}_i^{(1)\prime}$, $\boldsymbol{W}_E = \sum_{i=1}^{n-k} \boldsymbol{z}_i^{(2)} \boldsymbol{z}_i^{(2)\prime}$, *and let*

$$T_1^* = f_1(w_1, \ldots, w_{p-s}) = \log \prod_{i=1}^{p-s} (1 + w_i),$$

$$\text{(3.1)} \qquad T_2^* = f_2(w_1, \ldots, w_{p-s}) = \sum_{i=1}^{p-s} w_i,$$

$$T_3^* = f_3(w_1, \ldots, w_{p-s}) = \sum_{i=1}^{p-s} w_i/(1 + w_i).$$

Then

$$\text{(3.2)} \qquad \Pr\{T_i > c\} \leq \Pr\{T_i^* > c\} \qquad (i = 1, 2, 3)$$

holds for any $\delta_1, \ldots, \delta_s$.

PROOF. Let

$$\tilde{\boldsymbol{Y}} = \begin{bmatrix} \tilde{\boldsymbol{y}}_1' \\ \tilde{\boldsymbol{Y}}_2 \\ \tilde{\boldsymbol{Y}}_3 \end{bmatrix}, \tilde{\boldsymbol{Y}}_2 = \begin{bmatrix} \tilde{\boldsymbol{y}}_2' \\ \vdots \\ \tilde{\boldsymbol{y}}_k' \end{bmatrix}, \tilde{\boldsymbol{Y}}_3 = \begin{bmatrix} \tilde{\boldsymbol{y}}_{k+1}' \\ \vdots \\ \tilde{\boldsymbol{y}}_n' \end{bmatrix}, \tilde{\boldsymbol{M}} = \begin{bmatrix} \sqrt{n}\bar{\boldsymbol{\mu}}' \\ \tilde{\boldsymbol{M}}_2 \\ \boldsymbol{O}_{(n-k)\times p} \end{bmatrix}, \tilde{\boldsymbol{M}}_2 = \begin{bmatrix} \tilde{\boldsymbol{m}}_2' \\ \vdots \\ \tilde{\boldsymbol{m}}_k' \end{bmatrix}.$$

Let $\boldsymbol{\Gamma}_2$ be an orthogonal matrix of order $k-1$ such that $\boldsymbol{\Gamma}_2 \tilde{\boldsymbol{M}}_2 = \begin{bmatrix} \boldsymbol{Q}_2 \\ \boldsymbol{O}_{s\times p} \end{bmatrix}$, where $\boldsymbol{Q}_2$ is a $(k-1-s) \times p$ matrix and $\boldsymbol{O}_{s\times p}$ is a $s \times p$ zero matrix. Similarly, using an orthogonal matrix $\boldsymbol{\Gamma} = \begin{bmatrix} 1 & & \boldsymbol{O} \\ & \boldsymbol{\Gamma}_2 & \\ \boldsymbol{O} & & \boldsymbol{I}_{n-k} \end{bmatrix}$, $\tilde{\boldsymbol{Y}}$ is transformed as $\boldsymbol{U} = \boldsymbol{\Gamma}\tilde{\boldsymbol{Y}}$, where $\boldsymbol{U} = [\boldsymbol{u}_1, \boldsymbol{U}_2', \boldsymbol{U}_3', \boldsymbol{U}_4']'$, $\boldsymbol{U}_2 = [\boldsymbol{u}_1^{(1)}, \ldots, \boldsymbol{u}_{k-s-1}^{(1)}]'$, $\boldsymbol{U}_3 = [\boldsymbol{u}_{k-s}^{(1)}, \ldots, \boldsymbol{u}_{k-1}^{(1)}]'$, $\boldsymbol{U}_4 = [\boldsymbol{u}_1^{(2)}, \ldots, \boldsymbol{u}_{n-k}^{(2)}]'$. Then the variation matrices are expressed as $\boldsymbol{H} = \boldsymbol{H}_1 + \boldsymbol{H}_2$ and $\boldsymbol{E} = \boldsymbol{U}_4'\boldsymbol{U}_4$, where $\boldsymbol{H}_1 = \boldsymbol{U}_2'\boldsymbol{U}_2$ and $\boldsymbol{H}_2 = \boldsymbol{U}_3'\boldsymbol{U}_3$. Here, let $r_1(\boldsymbol{U})$ and $r_2(\boldsymbol{U}_2, \boldsymbol{U}_4)$ be *PDFs* of $\boldsymbol{U}$ and $(\boldsymbol{U}_2, \boldsymbol{U}_4)$, respectively, and we define the following sets:

$$B_1 = \left\{ \boldsymbol{U} \mid \log \prod_{i=s+1}^{p} (1 + \mathrm{ch}_i(\boldsymbol{H}\boldsymbol{E}^{-1})) > c \right\},$$

$$B_2 = \{(\boldsymbol{U}_2, \boldsymbol{U}_4) \mid \log \prod_{i=s+1}^{p} (1 + \mathrm{ch}_i(\boldsymbol{H}_1\boldsymbol{E}^{-1})) > c\},$$

where c is a constant and $\mathrm{ch}_i(\boldsymbol{A})$ denotes the i-th largest eigenvalue of a matrix $\boldsymbol{A}$. Then

$$\text{(3.3)} \quad \Pr\{T_1 > c\} = \int_{B_1} r_1(\boldsymbol{U}) d\boldsymbol{U} \leq \int_{B_2} \left\{ \int \int r_1(\boldsymbol{U}) d\boldsymbol{U}_1 d\boldsymbol{U}_3 \right\} d\boldsymbol{U}_2 d\boldsymbol{U}_4 = \int_{B_2} r_2(\boldsymbol{U}_2, \boldsymbol{U}_4) d\boldsymbol{U}_2 d\boldsymbol{U}_4.$$

On the other hand, let $\boldsymbol{F}_1$ be a $(p-s) \times p$ matrix of rank $p-s$ that satisfies $\boldsymbol{F}_1\boldsymbol{Q}_2' = \boldsymbol{O}_{(p-s)\times(k-1-s)}$, $\boldsymbol{F}_1\boldsymbol{\Omega}\boldsymbol{F}_1' = \boldsymbol{I}_{p-s}$ and $\boldsymbol{Z}_1' = \boldsymbol{F}_1\boldsymbol{U}_2' = [\boldsymbol{z}_1^{(1)}, \ldots, \boldsymbol{z}_{k-1-s}^{(1)}]$,

$Z_2' = F_1 U_4' = [z_1^{(2)}, \dots, z_{n-k}^{(2)}]$.

Since $\mathrm{vec}\left(\begin{bmatrix} U_2 \\ U_4 \end{bmatrix}'\right) \sim EC_{(n-1-s)p}\left(\mathrm{vec}\left(\begin{bmatrix} Q_2 \\ O_{(n-k)\times p} \end{bmatrix}'\right), I_{n-1-s} \otimes \Omega, \psi\right)$ and $\mathrm{vec}\left(\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}'\right) = (I_{n-1-s} \otimes F_1)\mathrm{vec}\left(\begin{bmatrix} U_2 \\ U_4 \end{bmatrix}'\right)$, we find that

$$\mathrm{vec}\left(\begin{bmatrix} Z_1 \\ Z_2 \end{bmatrix}'\right) \sim EC_{(n-1-s)(p-s)}(\mathbf{0}, I_{n-1-s} \otimes I_{p-s}, \psi).$$

Here, let $T_1^* = \log \prod_{i=1}^{p-s}[1 + \mathrm{ch}_i\{(Z_1'Z_1)(Z_2'Z_2)^{-1}\}]$ and $B_3 = \{(U_2, U_4) \mid T_1^* > c\}$. Since $Z_1'Z_1 = F_1 H_1 F_1'$ and $Z_2'Z_2 = F_1 E F_1'$, we can see that $\mathrm{ch}_i\{(Z_1'Z_1)(Z_2'Z_2)^{-1}\} \geq \mathrm{ch}_{s+i}\{H_1 E^{-1}\}$ $(i = 1, \dots, p-s)$ from Olkin and Tomsky (1981). Then $B_2 \subset B_3$ holds. Hence,

$$\text{(3.4)} \qquad \int_{B_2} r_2(U_2, U_4) dU_2 dU_4 \leq \int_{B_3} r_2(U_2, U_4) dU_2 dU_4 = \Pr\{T_1^* > c\}.$$

From (3.3) and (3.4),

$$\Pr\{T_1 > c\} \leq \Pr\{T_1^* > c\}.$$

We can prove the theorem for T_2 and T_3 by a similar procedure. □

Using test criteria $T_i^* (i = 1, 2, 3)$ defined in Theorem 1, we can test H_s without the effects of the nuisance parameters. Furthermore, if we have a critical point c_α^* with the level of significance α based on T_i^*, then $\Pr\{T_i > c_\alpha^*\} \leq \alpha$ holds for any $\delta_1, \dots, \delta_s$. This means that $\Pr\{T_i > c_\alpha^*\}$ remains less than the level of significance even in the least favorable case.

4. Upper limits under contaminated normal populations

If the difference between $\Pr\{T_i > c\}$ and $\Pr\{T_i^* > c\}$ is large, T_i^* is not suitable as test criteria. Schott (1984) showed that under normal populations $\Pr\{T_i^* > c\} \to \Pr\{T_i > c\}$ if $\delta_1, \dots, \delta_s \to \infty$ for $i = 1$ and $i = 2$ (LR-type and LH-type). In this section, we show that T_1^* and T_2^* defined in Theorem 1 are upper limits under contaminated normal distributions.

First, the following theorem shows an equivalent condition of (3.2).

THEOREM 2. *Suppose that population distributions are contaminated normal distributions, that is, an np-variate random vector* $\tilde{y}$ *and an* $(n-1-s)(p-s)$*-variate random vector* z *have the PDFs*

$$\tilde{h}(\tilde{y}) = (1-\varepsilon)\tilde{h}_1(\tilde{y}) + \varepsilon \tilde{h}_2(\tilde{y}), \quad h(z) = (1-\varepsilon)h_1(z) + \varepsilon h_2(z), \quad \textit{respectively.}$$

Here, $\tilde{h}_1(\tilde{y})$, $\tilde{h}_2(\tilde{y})$, $h_1(z)$ *and* $h_2(z)$ *are PDFs of* $N_{np}(\tilde{m}, I_n \otimes \Omega)$, $N_{np}(\tilde{m}, I_n \otimes \sigma^2\Omega)$, $N_{(n-1-s)(p-s)}(\mathbf{0}, I_{n-1-s} \otimes I_{p-s})$, *and* $N_{(n-1-s)(p-s)}(\mathbf{0}, I_{n-1-s} \otimes \sigma^2 I_{p-s})$, *respectively, and* ε, σ^2 *are known. Then*

$$\text{(4.1)} \qquad \lim_{\delta_1, \dots, \delta_s \to \infty} \Pr\{T_i > c\} = \Pr\{T_i^* > c\}, \quad (i = 1, 2),$$

where T_i and T_i^* are defined in (1.3) *and* (3.1), *holds.*

In order to prove Theorem 2, the following lemma shall be obtained.

LEMMA 1. *Let $\boldsymbol{GDG}'$ be a spectral decomposition of $\boldsymbol{\Sigma}^{-1/2}\boldsymbol{\Delta}\boldsymbol{\Sigma}^{-1/2}$, where $\boldsymbol{\Sigma} = \kappa\boldsymbol{\Omega}$, $\kappa = 1-\varepsilon+\varepsilon\sigma^2$, $\boldsymbol{D} = \mathrm{diag}(\delta_1,\dots,\delta_s,0,\dots,0)$ and $\boldsymbol{G}$ is an orthogonal matrix of order p. Let $\boldsymbol{b} = (\boldsymbol{e}_1',\dots,\underset{i}{\boldsymbol{e}_s'},\boldsymbol{0}',\dots,\boldsymbol{0}')$ and $\boldsymbol{V} = \mathrm{diag}((\kappa\delta_1)^{-1},\dots,(\kappa\delta_s)^{-1},1,\dots,1)$, where $\boldsymbol{e}_i = (0,\dots,0,\overset{i}{1},0,\dots,0)' \in \boldsymbol{R}^p$.*
Suppose that $\boldsymbol{x} = (\boldsymbol{x}_1^{(1)'},\dots,\boldsymbol{x}_{k-1}^{(1)'},\boldsymbol{x}_1^{(2)'},\dots,\boldsymbol{x}_{n-k}^{(2)'})'$ is an $(n-1)p$-variate random vector that has the following PDF:

$$g(\boldsymbol{x}) = (1-\varepsilon)g_1(\boldsymbol{x}) + \varepsilon g_2(\boldsymbol{x}),$$

where $g_1(\boldsymbol{x})$ and $g_2(\boldsymbol{x})$ are PDFs of $N_{(n-1)p}(\boldsymbol{b},\boldsymbol{I}_{n-1}\otimes\boldsymbol{V})$ and $N_{(n-1)p}(\boldsymbol{b},\boldsymbol{I}_{n-1}\otimes\sigma^2\boldsymbol{V})$, respectively. Further, let $\rho_1 > \cdots > \rho_p$ be eigenvalues of $\boldsymbol{W}_1\boldsymbol{W}_2^{-1}$, where $\boldsymbol{W}_1 = \sum_{i=1}^{k-1}\boldsymbol{x}_i^{(1)}\boldsymbol{x}_i^{(1)'}$ and $\boldsymbol{W}_2 = \sum_{i=1}^{n-k}\boldsymbol{x}_i^{(2)}\boldsymbol{x}_i^{(2)'}$. Then the following statements hold.

(i) $(\boldsymbol{H},\boldsymbol{E}) \overset{d}{=} (\boldsymbol{F}_2\boldsymbol{W}_1\boldsymbol{F}_2',\boldsymbol{F}_2\boldsymbol{W}_2\boldsymbol{F}_2')$, *where $\boldsymbol{F}_2 = \boldsymbol{\Omega}^{1/2}\boldsymbol{G}\boldsymbol{V}^{-1/2}$ and $\overset{d}{=}$ denotes equivalence in distribution.*

(ii) *If $\delta_1,\dots,\delta_s \to \infty$, then $(\rho_{s+1},\dots,\rho_p) \overset{d}{\to} (w_1,\dots,w_{p-s})$, where $\overset{d}{\to}$ denotes convergence in distribution.*

PROOF OF LEMMA 1. (i) Let $\Phi(\Theta_1,\Theta_2)$ and $\Psi(\Theta_1,\Theta_2)$ be *CF*s of $(\boldsymbol{H},\boldsymbol{E})$ and $(\boldsymbol{F}_2\boldsymbol{W}_1\boldsymbol{F}_2',\boldsymbol{F}_2\boldsymbol{W}_2\boldsymbol{F}_2')$, respectively, that is, $\Phi(\Theta_1,\Theta_2) = (1-\varepsilon)\Phi_1(\Theta_1,\Theta_2) + \varepsilon\Phi_2(\Theta_1,\Theta_2)$, $\Psi(\Theta_1,\Theta_2) = (1-\varepsilon)\Psi_1(\Theta_1,\Theta_2) + \varepsilon\Psi_2(\Theta_1,\Theta_2)$, where $\Theta_l = ((1+\delta_{\alpha\beta})\theta_{\alpha\beta}^{(l)})$, $\theta_{\alpha\beta}^{(l)} = \theta_{\beta\alpha}^{(l)}$, $\Phi_l(\Theta_1,\Theta_2) = \int \tilde{h}_l(\tilde{\boldsymbol{y}})\mathrm{etr}(i\Theta_1\boldsymbol{H}+i\Theta_2\boldsymbol{E})d\tilde{\boldsymbol{y}}$, $\Psi_l(\Theta_1,\Theta_2) = \int g_l(\boldsymbol{x})\mathrm{etr}(i\Theta_1\boldsymbol{F}_2\boldsymbol{W}_1\boldsymbol{F}_2' + i\Theta_2\boldsymbol{F}_2\boldsymbol{W}_2\boldsymbol{F}_2')d\boldsymbol{x}(l=1,2)$, $\delta_{\alpha\beta}$ is Kronecker's delta symbol, and etr$\boldsymbol{A}$ denotes exp(trace$\boldsymbol{A}$) for a matrix $\boldsymbol{A}$.

Then we can easily calculate that

$$\begin{aligned}\Phi_1(\Theta_1,\Theta_2) &= |\boldsymbol{I}_p - i\Theta_1\boldsymbol{\Omega}|^{-(k-1)/2}|\boldsymbol{I}_p - i\Theta_2\boldsymbol{\Omega}|^{-(n-k)/2}\\ &\quad\times \mathrm{etr}\left(-\frac{1}{2}\boldsymbol{\Delta}\boldsymbol{\Omega}^{-1}\right)\mathrm{etr}\left(\frac{1}{2}\boldsymbol{\Delta}\boldsymbol{\Omega}^{-1}(\boldsymbol{I}_p - i\Theta_1\boldsymbol{\Omega})^{-1}\right),\end{aligned}$$

and

$$\begin{aligned}\Psi_1(\Theta_1,\Theta_2) &= |\boldsymbol{I}_p - i\Theta_1\boldsymbol{F}_2\boldsymbol{V}\boldsymbol{F}_2'|^{-(k-1)/2}|\boldsymbol{I}_p - i\Theta_2\boldsymbol{F}_2\boldsymbol{V}\boldsymbol{F}_2'|^{-(n-k)/2}\\ &\quad\times \mathrm{etr}\left[-\frac{1}{2}\left(\sum_{j=1}^{s}\boldsymbol{F}_2\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{F}_2'\right)(\boldsymbol{F}_2\boldsymbol{V}\boldsymbol{F}_2')^{-1}\right]\\ &\quad\times \mathrm{etr}\left[\frac{1}{2}\left(\sum_{j=1}^{s}\boldsymbol{F}_2\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{F}_2'\right)(\boldsymbol{F}_2\boldsymbol{V}\boldsymbol{F}_2')^{-1}(\boldsymbol{I}_p - i\Theta_1\boldsymbol{F}_2\boldsymbol{V}\boldsymbol{F}_2')^{-1}\right].\end{aligned}$$

Since $\boldsymbol{F}_2\boldsymbol{V}\boldsymbol{F}_2' = \boldsymbol{\Omega}$ and $\sum_{j=1}^{s}(\boldsymbol{F}_2\boldsymbol{e}_j\boldsymbol{e}_j'\boldsymbol{F}_2') = \boldsymbol{\Delta}$, we find that $\Psi_1(\Theta_1,\Theta_2) = \Phi_1(\Theta_1,\Theta_2)$. Similarly, $\Psi_2(\Theta_1,\Theta_2) = \Phi_2(\Theta_1,\Theta_2)$ so that $\Psi(\Theta_1,\Theta_2) = \Phi(\Theta_1,\Theta_2)$. Therefore,

$$(\boldsymbol{H},\boldsymbol{E}) \overset{d}{=} (\boldsymbol{F}_2\boldsymbol{W}_1\boldsymbol{F}_2',\boldsymbol{F}_2\boldsymbol{W}_2\boldsymbol{F}_2').$$

(ii) Divide $\boldsymbol{x}_i^{(1)}$, $\boldsymbol{x}_i^{(2)}$ as follows:

$$\boldsymbol{x}_i^{(1)} = \begin{bmatrix} \boldsymbol{x}_{1i}^{(1)} \\ \boldsymbol{x}_{2i}^{(1)} \end{bmatrix}, \qquad \boldsymbol{x}_i^{(2)} = \begin{bmatrix} \boldsymbol{x}_{1i}^{(2)} \\ \boldsymbol{x}_{2i}^{(2)} \end{bmatrix},$$

where $\boldsymbol{x}_{1i}^{(1)}$ and $\boldsymbol{x}_{1i}^{(2)}$ are s-variate vectors. Let

$$\boldsymbol{x}_{i*}^{(1)} = \begin{cases} \begin{bmatrix} \boldsymbol{j}_i \\ \boldsymbol{x}_{2i}^{(1)} \end{bmatrix} & (i = 1, \dots, s) \\ \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{x}_{2i}^{(1)} \end{bmatrix} & (i = s+1, \dots, k-1) \end{cases}, \quad \boldsymbol{x}_{i*}^{(2)} = \begin{bmatrix} \boldsymbol{0} \\ \boldsymbol{x}_{2i}^{(2)} \end{bmatrix} \quad (i = 1, \dots, n-k),$$

where $\boldsymbol{j}_i = (0, \dots, 0, \overset{i}{1}, 0, \dots, 0)' \in \boldsymbol{R}^s$. If $\delta_1, \dots, \delta_s \to \infty$, (i.e., variances of $\boldsymbol{x}_{1i}^{(1)}$ and $\boldsymbol{x}_{1i}^{(2)}$ are close to $\boldsymbol{0}$), then $\boldsymbol{x}_i^{(1)} \to \boldsymbol{x}_{i*}^{(1)}$, $\boldsymbol{x}_i^{(2)} \to \boldsymbol{x}_{i*}^{(2)}$. Define $\boldsymbol{W}_{1*} = \sum_{i=1}^{k-1} \boldsymbol{x}_{i*}^{(1)} \boldsymbol{x}_{i*}^{(1)'}$ and $\boldsymbol{W}_{2*} = \sum_{i=1}^{n-k} \boldsymbol{x}_{i*}^{(2)} \boldsymbol{x}_{i*}^{(2)'}$. We consider the following determinantal equation:

$$|\boldsymbol{W}_{1*} - \rho \boldsymbol{W}_{2*}| = 0. \tag{4.2}$$

Because $\operatorname{rank}(\boldsymbol{W}_{2*}) = s$ with probability one, the s largest roots of (4.2) are infinite, that is, $\rho_1, \dots, \rho_s \to \infty$. Then (4.2) can be expressed as

$$\begin{vmatrix} \boldsymbol{I}_s & \sum_{i=1}^{s} \boldsymbol{j}_i \boldsymbol{x}_{2i}^{(1)'} \\ \sum_{i=1}^{s} \boldsymbol{x}_{2i}^{(1)} \boldsymbol{j}_i' & \sum_{i=1}^{k-1} \boldsymbol{x}_{2i}^{(1)} \boldsymbol{x}_{2i}^{(1)'} - \rho \sum_{i=1}^{n-k} \boldsymbol{x}_{2i}^{(2)} \boldsymbol{x}_{2i}^{(2)'} \end{vmatrix} = \left| \sum_{i=s+1}^{k-1} \boldsymbol{x}_{2i}^{(1)} \boldsymbol{x}_{2i}^{(1)'} - \rho \sum_{i=1}^{n-k} \boldsymbol{x}_{2i}^{(2)} \boldsymbol{x}_{2i}^{(2)'} \right| = 0. \tag{4.3}$$

Roots of (4.3) are the $p - s$ smallest roots of (4.2), that is, $\rho_{s+1}, \dots, \rho_p$. It is easily seen that $(\sum_{i=s+1}^{k-1} \boldsymbol{x}_{2i}^{(1)} \boldsymbol{x}_{2i}^{(1)'}, \sum_{i=1}^{n-k} \boldsymbol{x}_{2i}^{(2)} \boldsymbol{x}_{2}^{(2)'}) \overset{d}{=} (\boldsymbol{W}_H, \boldsymbol{W}_E)$. Hence, if $\delta_1, \dots, \delta_s \to \infty$, then

$$(\rho_{s+1}, \dots, \rho_p) \xrightarrow{d} (w_1, \dots, w_{p-s})$$

holds. □

From Lemma 1, we can prove Theorem 2.

Proof of Theorem 2. By Lemma 1-(i), we can see

$$(l_1, \dots, l_p) \overset{d}{=} (\rho_1, \dots, \rho_p). \tag{4.4}$$

where $\rho_1, \dots, \rho_p$ are obtained in Lemma 1. By Lemma 1-(ii) and (4.4), we conclude that

$$(l_{s+1}, \dots, l_p) \xrightarrow{d} (w_1, \dots, w_{p-s}).$$

Since f_1, f_2 are Borel measurable functions, $T_1 \xrightarrow{d} T_1^*$ and $T_2 \xrightarrow{d} T_2^*$. Therefore,

$$\lim_{\delta_1,\ldots,\delta_s\to\infty} \Pr\{T_i > c\} = \Pr\{T_i^* > c\}, \quad (i = 1, 2),$$

holds. □

From Theorem 1 and Theorem 2, we obtain the following theorem.

THEOREM 3. *Let* $\boldsymbol{y}$ *be a random vector distributed according to a contaminated normal distribution defined in Theorem 2. Then*

$$\sup \Pr\{T_i > c\} = \Pr\{T_i^* > c\}, \quad (i = 1, 2),$$

holds.

If the difference between $\Pr\{T_i > c\}$ and $\Pr\{T_i^* > c\}$ is large, we cannot use T_i^* as test criteria in practice. However, under contaminated normal populations, Theorem 3 guarantees that $\Pr\{T_i > c\} = \Pr\{T_i^* > c\}$, $(i = 1, 2)$, for some $\delta_1, \ldots, \delta_s$. Therefore, we can use $T_i^*(i = 1, 2)$ as test criteria in practice.

5. Numerical examples

In the previous section, we mentioned that under contaminated normal populations our new criteria T_1^* and T_2^* are upper limits for null distributions and suitable as test criteria. In this section, we illustrate the fact by some numerical examples.

Example 1. We consider $\Pr\{T_i > c_{i,0.05}^*\}$ $(i = 1, 2)$ under some contaminated normal populations when $s = 1$. The procedure is as follows: (i) Set

Table 1. $\overline{\Pr}\{T_i > c_{i,0.05}^*\}$ $(s = 1)$.

	$\delta_1 = 20$	$\delta_1 = 40$	$\delta_1 = 80$	$\delta_1 = 120$	$\delta_1 = 160$	$\delta_1 = 200$
LR-type	0.0207	0.0305	0.0405	0.0445	0.0459	0.0481
LH-type	0.0186	0.0297	0.0401	0.0430	0.0452	0.0483

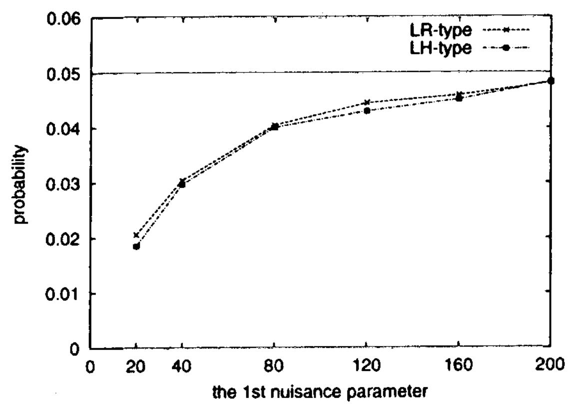

Figure 1. $\overline{\Pr}\{T_i > c_{i,0.05}^*\}$ $(s = 1)$.

Table 2. $\overline{\Pr}\{T_i > c^*_{i,0.05}\}$ $(s = 2)$.

		$\delta_2 = 20$	$\delta_2 = 40$	$\delta_2 = 80$	$\delta_2 = 120$	$\delta_2 = 160$	$\delta_2 = 200$
LR-type	$\delta_1 = 20$	0.0129					
	$\delta_1 = 40$	0.0182	0.0260				
	$\delta_1 = 80$	0.0218	0.0337	0.0389			
	$\delta_1 = 120$	0.0238	0.0346	0.0411	0.0429		
	$\delta_1 = 160$	0.0248	0.0357	0.0418	0.0433	0.0438	
	$\delta_1 = 200$	0.0251	0.0364	0.0429	0.0453	0.0456	0.0460
		$\delta_2 = 20$	$\delta_2 = 40$	$\delta_2 = 80$	$\delta_2 = 120$	$\delta_2 = 160$	$\delta_2 = 200$
LH-type	$\delta_1 = 20$	0.0123					
	$\delta_1 = 40$	0.0178	0.0257				
	$\delta_1 = 80$	0.0216	0.0335	0.0383			
	$\delta_1 = 120$	0.0236	0.0343	0.0413	0.0433		
	$\delta_1 = 160$	0.0243	0.0356	0.0414	0.0435	0.0436	
	$\delta_1 = 200$	0.0247	0.0360	0.0431	0.0454	0.0454	0.0459

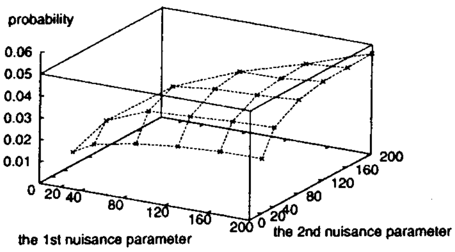

Figure 2. $\overline{\Pr}\{T_1 > c^*_{1,0.05}\}$ $(s = 2)$.

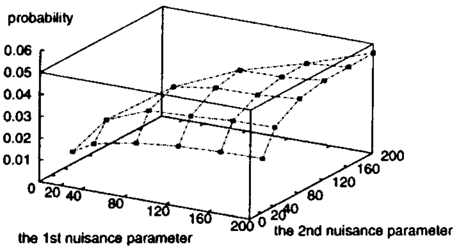

Figure 3. $\overline{\Pr}\{T_2 > c^*_{2,0.05}\}$ $(s = 2)$

$p = 4$, $k = 5$, $n_1 = \cdots = n_5 = 5$, $\varepsilon = 0.5$, $\sigma^2 = 100$, and $\boldsymbol{\Omega} = \boldsymbol{I}$. (ii) For $\delta_1 = 20, 40, 80, 120, 160, 200$, we generate 10000 sets of random vectors $\boldsymbol{y}$, $\boldsymbol{z}$. (iii) Calculate T_i and T_i^*. (iv) Set the level of significance, $\alpha = 0.05$ and calculate a critical point for T_i^*, that is, $c^*_{i,0.05}$. Further, calculate $\Pr\{T_i > c^*_{i,0.05}\}$. (v) Repeat steps (i)-(iv) 10 times, and then calculate an average of $\Pr\{T_i > c^*_{i,0.05}\}$, that is, $\overline{\Pr}\{T_i > c^*_{i,0.05}\}$. Table 1 and Figure 1 report the results.

Example 2. We consider $\Pr\{T_i > c^*_{i,0.05}\}$ $(i = 1, 2)$ under some contaminated normal populations when $s = 2$ with a similar procedure of Example 1. Table 2, Figure 2 and Figure 3 report the results.

These examples show that when all nuisance parameters are large $\Pr\{T_i > c^*_{i,0.05}\}$ close to 0.05, that is, $\Pr\{T_i^* > c^*_{i,0.05}\}$. This result gives agreement with Theorem 3, and illustrate that our new criteria are suitable.

6. Concluding remarks

Because the null distributions of the usual criteria $T_i (i = 1, 2, 3)$ for testing dimensionality depend on nuisance parameters, these parameters should be replaced by their estimators. When testing is based on T_i, it is probable that the probability of rejecting H_s is greater than the level of significance, α. However, our new criteria $T_i^*(i = 1, 2, 3)$ do not depend on these parameters, and it is therefore possible to test without the effects of nuisance parameters.

Furthermore, $\Pr\{T_i > c_\alpha^*\} \leq \Pr\{T_i^* > c_\alpha^*\} = \alpha$, where c_α^* is a critical point based on T_i^*, holds. This means that $\Pr\{T_i > c_\alpha^*\}$ remains less than the level of significance even in the least favorable case. Moreover, under contaminated normal populations, T_i^* is the upper limit for the null distribution of T_i, for $i = 1$ and $i = 2$ (LR-type and LH-type). Therefore, testing based on T_i^* is useful when the sample size n is small.

Future study is needed to obtain the upper limits for a null distribution of the BNP-type criterion under normal populations and to obtain the upper limits under other elliptical populations.

Acknowledgements

The authors would like to thank the referees for their valuable comments.

REFERENCES

Backhouse, A. R. and McKay, R. J. (1982). Tests of dimensionality in multivariate analysis of variance, *Communications in Statistics-Theory and Methods*, **11**, 1003–1027.

Fang, K. T., and Zhang, Y. T. (1988). *Generalized Multivariate Analysis*, Springer-Verlag, Berlin.

Muirhead, R. J. (1982). *Aspects of Multivariate Statistical Theory*, Wiley, New York.

Muirhead, R. J., and Waternaux, C. (1980). Asymptotic distributions in canonical correlation analysis and other multivariate procedures for nonnormal populations, *Biometrika*, **67**, 31–43.

Olkin, I. and Tomsky, J. L. (1981). A new class of multivariate tests based on the union-intersection principle, *Ann. Statist.*, **9**, 792–802.

Schott, J. R. (1984). Optimal bounds for the distributions of some test criteria for tests of dimensionality, *Biometrika*, **71**, 561–567.

Seo, T., Kanda, T., and Fujikoshi, Y. (1995). The effects of nonnormality on tests for dimensionality in canonical correlation and MANOVA models, *J. Multivariate Anal.*, **52**, 325–337.