科研費 基盤研究B 代表者 杉山 高一
「高次元データの推測理論の開発と応用」
2009 年 12 月 5 日〜6 日

RECENT DEVELOPMENT OF EIGENVALUES AND EIGENVECTORS FOR COVARIANCE MATRIX

Takakazu Sugiyama, Shin-ichi Tsukada, Yuichi Takeda & Hidetoshi Murakami

Key words and phrases: eigenvalue, eigenvector, covariance matrix, hypothesis testing

0.1. Introduction

We denotes $\Pi_g(\boldsymbol{\mu}_g, \Sigma_g)$ as the population distribution of group g, where $\boldsymbol{\mu}_g$ is a mean vector and Σ_g is a covariance matrix. Let $\lambda_j^{(g)}$ be the j-th largest eigenvalue of Σ_g where $\lambda_1^{(g)} > \cdots > \lambda_p^{(g)}$, and let $\boldsymbol{\eta}_j^{(g)}$ be the eigenvector corresponding to $\lambda_j^{(g)}$.

From these populations, we obtain a p-variate random sample $\{\boldsymbol{X}_1^{(g)}, \ldots, \boldsymbol{X}_{N_g}^{(g)}\}$. Let $\bar{\boldsymbol{X}}_g$ be the sample mean vector. The sample covariance matrix is denoted as

$$S_g = \frac{1}{N_g - 1} \sum_{i=1}^{N_g} \left(\boldsymbol{X}_i^{(g)} - \bar{\boldsymbol{X}}_g \right) \left(\boldsymbol{X}_i^{(g)} - \bar{\boldsymbol{X}}_g \right)'.$$

We denote $l_j^{(g)}$ and $\boldsymbol{h}_j^{(g)}$ as the j-th largest sample eigenvalue of S_g where $l_1^{(g)} > \cdots > l_p^{(g)}$ and the eigenvector corresponding to $l_j^{(g)}$, respectively.

There are two sections, hypothesis testing for eigenvalue and for eigenvector. In Section 2, the hypothesis testing for eigenvalue is dealt with the test for equality of variances under 2-population and multi-population. For eigenvector, we describe in Section 3.

0.2. Nonparametric test for eigenvalue

0.2.1. Two-population case

In this section, we consider testing the hypothesis

$$H_0 : \lambda_j^{(1)} = \lambda_j^{(2)},$$
$$H_1 : \text{not } H_0.$$

For two-population, Sugiyama and Ushizawa (1998) propose the nonparametric procedure which is the Ansari-Bradley test by using the principal component scores. Their procedure is concretely as follows:

The principal component scores are

$$Y_1 = \left\{ y_{j1}^{(1)}, y_{j2}^{(1)}, y_{j3}^{(1)}, \ldots, y_{jN_1}^{(1)} \right\}$$

and

$$Y_2 = \left\{ y_{j1}^{(2)}, y_{j2}^{(2)}, y_{j3}^{(2)}, \ldots, y_{jN_2}^{(2)} \right\},$$

where $y_{ji}^{(g)} = \boldsymbol{h}_j^{(g)\prime} \left(\boldsymbol{X}_i^{(g)} - \bar{\boldsymbol{X}}_g \right)$.

(1) Two samples Y_1 and Y_2 are combined and ordered, and then ranks are assigned as follows:

$$\{1, 2, \ldots, (N_1 + N_2)/2, (N_1 + N_2)/2, \ldots, 2, 1\}$$

if $N_1 + N_2$ is even and

$$\{1, 2, \ldots, (N_1 + N_2 - 1)/2, (N_1 + N_2 + 1)/2, (N_1 + N_2 - 1)/2, \ldots, 2, 1\}$$

if $N_1 + N_2$ is odd.

(2) Calculate the sum of the above ranks, W, associated with the sample Y_1. The following criterion based W

$$Z = \frac{W - E(W)}{\sqrt{V(W)}}$$

can be approximated by a standard normal distribution $N(0,1)$ according to Ansari-Bradley (1960) under a few assumptions. One of the assumptions for the Ansari-Bradley test is that the sample values are pairwise independent. But, when we use the sample latent vector $\boldsymbol{h}_j^{(k)}$ to calculate the $y_{ji}^{(k)}$, they are no longer independent, though the degree of dependence is very weak. What kind of nonparametric test is available depends on the population situation.

The distribution of the sample eigenvalue may be approximated by chi-square. They attempt to apply the F test for dispersion to their testing problem. Their method is more reliable than the F test in their situation.

Takeda (2001) propose the criterion

$$T_1 = l_j^{(2)} / l_j^{(1)}$$

and apply the permutation test using it under the multivariate normal population and multivariate contaminated normal population. The exact distribution of T_1 is derived as follows:

$$\begin{aligned} h(t_1) = {} & C(\Sigma_1, n_1) C(\Sigma_2, n_2) n_1^{\tilde{n}_1} n_2^{\tilde{n}_2} \\ & \sum_{k=0}^{\infty} \sum_{k'=0}^{\infty} \sum_{\kappa} \sum_{\kappa'} \frac{\left(\frac{p+1}{2}\right)_\kappa}{\left(\frac{n_1+p+1}{2}\right)_\kappa} \frac{\left(\frac{p+1}{2}\right)_{\kappa'}}{\left(\frac{n_2+p+1}{2}\right)_{\kappa'}} \frac{C_\kappa\left(n_1 \tilde{\Sigma}_1^{-1}\right)}{k!} \frac{C_{\kappa'}\left(n_2 \tilde{\Sigma}_2^{-1}\right)}{k'!} \\ & \left\{ (\tilde{n}_1 + k)(\tilde{n}_2 + k') \Gamma\left(\tilde{n}_1 + \tilde{n}_2 + k + k'\right) \Delta_0^{-1} \right. \\ & - \left(\operatorname{tr}\, n_1 \tilde{\Sigma}_1\right) (\tilde{n}_2 + k') \Gamma\left(\tilde{n}_1 + \tilde{n}_2 + k + k' + 1\right) \Delta_1^{-1} \\ & - (\tilde{n}_1 + k) \left(\operatorname{tr}\, n_2 \tilde{\Sigma}_2\right) \Gamma\left(\tilde{n}_1 + \tilde{n}_2 + k + k' + 1\right) \Delta_1^{0} \\ & \left. + \left(\operatorname{tr}\, n_1 \tilde{\Sigma}_1\right) \left(\operatorname{tr}\, n_2 \tilde{\Sigma}_2\right) \Gamma\left(\tilde{n}_1 + \tilde{n}_2 + k + k' + 2\right) \Delta_2^{0} \right\}, \end{aligned} \tag{0.1}$$

where

$$C(\mathbf{\Sigma}, n) = |\mathbf{\Sigma}|^{-\frac{1}{2}n}\Gamma_p\left(\frac{p+1}{2}\right)\Big/2^{\frac{1}{2}np}\Gamma_p\left(\frac{n+p+1}{2}\right),$$

$$\Gamma_p(n) = \pi^{\frac{p(p-1)}{4}}\prod_{i=1}^{p}\Gamma\left(n-\frac{1}{2}(i-1)\right),$$

$$\Delta_j^i = \frac{t_1^{\tilde{n}_2+k'+i}}{\left(\text{tr } n_1\tilde{\Sigma}_1^{-1} + t_1 \text{ tr } n_2\tilde{\Sigma}_2^{-1}\right)^{\tilde{n}_1+\tilde{n}_2+k+k'+j}},$$

$\tilde{n}_g = pn_g/2$ and $\tilde{\Sigma}_g^{-1} = \Sigma_g^{-1}/2$.

Hino *et al.* (2009) suggest that the criterion T_2 is a more useful version as

$$T_2 = \sqrt{N}\log(l_j^{(2)}/l_j^{(1)})$$

where $N = N_1 + N_2$ in high dimension cases. The statistical properties of the criterion T_2 are the same as those of the criterion T_1. However, the limiting distribution of T_2 is a standard normal distribution. Therefore, the criterion T_2 is more useful than T_1 for testing the equality of the j-th largest eigenvalue because it is easy to find the critical value of the limiting distribution of T_2. They derive Edgeworth expansion for the distribution of T_2 as follows:

$$P\left(\frac{T_2}{\sigma} \le x\right) = \Phi(x) - \frac{1}{\sqrt{N}}\phi(x)\{a_1\sigma^{-1} + a_3\sigma^{-3}h_2(x)\} + o(N^{-\frac{1}{2}}), \tag{0.2}$$

where

$$\sigma^2 = \frac{1}{\lambda_j^{(2)2}}\rho_2^{-2}\kappa_4^{j(2)} + 2\rho_2^{-2} + \frac{1}{\lambda_j^{(1)2}}\rho_1^{-2}\kappa_4^{j(1)} + 2\rho_1^{-2},$$

$$\begin{aligned} a_1 = {} & \frac{1}{\lambda_j^{(2)}}\rho_2^{-2}\sum_{\alpha\neq j}\lambda_{j\alpha}^{(2)}\kappa_{22}^{j\alpha(2)} + \rho_2^{-2}\sum_{\alpha\neq j}\lambda_{j\alpha}^{(2)}\lambda_{\alpha}^{(2)} - \frac{1}{2\lambda_j^{(2)2}}\rho_2^{-2}\kappa_4^{j(2)} - \rho_2^{-2} \\ & - \frac{1}{\sqrt{N}}\left\{\frac{1}{\lambda_j^{(1)}}\rho_1^{-2}\sum_{\alpha\neq j}\lambda_{j\alpha}^{(1)}\kappa_{22}^{j\alpha(1)} - \rho_1^{-2}\sum_{\alpha\neq j}\lambda_{j\alpha}^{(1)}\lambda_{\alpha}^{(1)}\right. \\ & \left. + \frac{1}{2\lambda_j^{(1)2}}\rho_1^{-2}\kappa_4^{j(1)} + \rho_1^{-2}\right\}, \end{aligned}$$

$$\begin{aligned} a_3 = {} & \frac{1}{6\lambda_j^{(2)3}}\rho_2^{-3}\kappa_6^{j(2)} - \frac{1}{2\lambda_j^{(2)4}}\rho_2^{-4}\kappa_4^{j(2)2} + \frac{1}{\lambda_j^{(2)3}}\rho_2^{-4}\sum_{\alpha\neq j}\lambda_{j\alpha}^{(2)}\kappa_{31}^{j\alpha(2)2} \\ & - \frac{1}{6\lambda_j^{(1)3}}\rho_1^{-3}\kappa_6^{j(1)} + \frac{1}{2\lambda_j^{(1)4}}\rho_1^{-4}\kappa_4^{j(1)2} - \frac{1}{\lambda_j^{(1)3}}\rho_1^{-4}\sum_{\alpha\neq j}\lambda_{j\alpha}^{(1)}\kappa_{31}^{j\alpha(1)2} \\ & - \frac{2}{\lambda_j^{(2)2}}\rho_2^{-4}\kappa_4^{j(2)} + \frac{2}{\lambda_j^{(2)2}}\rho_2^{-3}\kappa_4^{j(2)} + \frac{2}{3\lambda_j^{(2)3}}\rho_2^{-3}\kappa_3^{j(2)2} - 2\rho_2^{-4} \\ & + \frac{2}{\lambda_j^{(1)2}}\rho_1^{-4}\kappa_4^{j(1)} - \frac{2}{\lambda_j^{(1)2}}\rho_1^{-3}\kappa_4^{j(1)} - \frac{2}{3\lambda_j^{(1)3}}\rho_1^{-3}\kappa_3^{j(1)2} + 2\rho_1^{-4} \\ & + \frac{4}{3}\rho_2^{-3} - \frac{4}{3}\rho_1^{-3}, \end{aligned}$$

where $\rho_g = \sqrt{N_g/N}$, $\kappa_3^{j(g)}$ denotes the third cumulants of $x_j^{(g)}$, $\kappa_4^{j(g)}$ does the fourth cumulants of $x_j^{(g)}$, $\kappa_{22}^{j\alpha(g)}$ does the cumulants of $x_j^{(g)2}x_\alpha^{(g)2}$, $\kappa_{31}^{j\alpha(g)}$ does the cumulants of $x_j^{(g)3}x_\alpha^{(g)}$ and $\kappa_6^{j(g)}$ does the sixth cumulants of $x_j^{(g)}$.

By simulation they investigate the accuracy for their test procedures under normal population and contaminated normal population. The criterion T_2, the Ansari-Bradley test, and the Mood test show convergence for the significance level. The power is greatest for the criterion T_2. Therefore, they expect that using the criterion T_2 is suitable for two populations.

0.2.2. Multi-sample case

In this section, we consider testing the hypothesis

$$H_0 : \lambda_j^{(1)} = \cdots = \lambda_j^{(k)},$$
$$H_1 : \text{not } H_0.$$

We deal with testing the equality of the j-th largest eigenvalues in the k-population using the principal component scores

$$\begin{aligned} Y_1 &= \left\{ y_{j1}^{(1)}, y_{j2}^{(1)}, \ldots, y_{jN_1}^{(1)} \right\}, \\ Y_2 &= \left\{ y_{j1}^{(2)}, y_{j2}^{(2)}, \ldots, y_{jN_2}^{(2)} \right\}, \\ &\vdots \\ Y_k &= \left\{ y_{j1}^{(k)}, y_{j2}^{(k)}, \ldots, y_{jN_k}^{(k)} \right\}. \end{aligned}$$

Under the null hypothesis, we may test the equality of eigenvalues as testing the sample variance of Y_g under the multivariate normal distribution. The variance of Y_g are unequal under the alternative hypothesis. Since the expected values of $l_j^{(g)}$ and the variance of $y_{j\alpha}^{(g)}$ are given by

$$\mathrm{E}\left[l_j^{(g)} \right] = \lambda_j^{(g)} + \frac{\lambda_j^{(g)}}{N_g} \sum_{i \neq j} \frac{\lambda_i^{(g)}}{\lambda_j^{(g)} - \lambda_i^{(g)}} + O(N_g^{-2}), \tag{0.3}$$

$$\mathrm{Var}\left[y_{j\alpha}^{(g)} \right] = \lambda_j^{(g)} - \frac{2}{N_g - 1} \sum_{i \neq j}^{p} \frac{\mathrm{E}\left[x_{i\alpha}^{(g)2} x_{j\alpha}^{(g)2} \right]}{\lambda_i^{(g)} - \lambda_j^{(g)}} + O(N_g^{-2}), \tag{0.4}$$

respectively, we may treat that the variances of Y_g are approximately equal when the sample size N_g are sufficiently large. In addition, we need larger sample sizes when the eigenvalues are close.

Ansari-Bradley test for k-population The Ansari-Bradley test is known as a method of testing the variances. One of the assumptions for the Ansari-Bradley test is that the sample values are independent. However, there exist weak correlations between each principal component scores. Sugiyama and Ushizawa (1998) prove that the degree of

dependence between each principal component score is weak when the total sample size N is sufficiently large under the multivariate normal distribution. Then they show the Ansari-Bradley test could be applicable to test the equality for the variance of the principal component Y_1 and Y_2 (cf. Ansari and Bradley (1960)).

When the total sample size $N = N_1 + \cdots + N_k$ is even, we give the criterion AB_e for the Ansari-Bradley test under k-population as follows:

$$AB_e = \frac{48(N-1)}{N(N^2-4)} \sum_{g=1}^{k} N_g \left(\bar{A}_j^{(g)} - \frac{N+2}{4} \right)^2 .$$

If N is odd, we give the criterion AB_o as follows:

$$AB_o = \frac{48N^2}{N(N+1)(N^2+3)} \sum_{g=1}^{k} N_g \left(\bar{A}_j^{(g)} - \frac{(N+1)^2}{4N} \right)^2 .$$

Here, $\bar{A}_j^{(g)}$ denotes

$$\bar{A}_j^{(g)} = \frac{1}{N_g} \sum_{i=1}^{N_g} \left(\frac{N+1}{2} - \left| R_{ji}^{(g)} - \frac{N+1}{2} \right| \right),$$

where $R_{ji}^{(g)}$ be the increasing order rank of $y_{ji}^{(g)}$ in the combined $N = N_1 + \cdots + N_k$ observations. The limiting distribution of the criterion for the Ansari-Bradley test is the χ^2 distribution with $(k-1)$ degrees of freedom (Tsai *et al.*, 1975).

Mood test for k-population It is well known that the asymptotic relative efficiency of the Mood test is higher than that of the Ansari-Bradley test (Gibbons and Chakraborti(2003)).

The criterion of the Mood test is as follows:

$$M = \frac{180}{N(N+1)(N^2-4)} \sum_{g=1}^{k} N_g \left(\bar{M}_j^{(g)} - \frac{N^2-1}{12} \right)^2 ,$$

where

$$\bar{M}_j^{(g)} = \frac{1}{N_g} \sum_{i=1}^{N_g} \left(R_{ji}^{(g)} - \frac{N+1}{2} \right)^2 .$$

The limiting distribution of the criterion M is also the χ^2 distribution with $(k-1)$ degrees of freedom under the null hypothesis (Tsai *et al.*, 1975).

Other nonparametric tests Murakami *at el.*(2008) propose a criterion T_M as follows:

$$T_M = \frac{N}{2k} \sum\sum_{1 \le g_1 < g_2 \le k} \left(\log \frac{l_j^{(g_2)}}{l_j^{(g_1)}} \right)^2 ,$$

where

$$N = \sum_{g=1}^{k} N_g, \quad l_j^{(g)} = \frac{1}{N_g} \sum_{i=1}^{N_g} \left(y_{ji}^{(g)} - \bar{y}_j^{(g)} \right)^2 \quad \text{and} \quad \bar{y}_j^{(g)} = \frac{1}{N_g} \sum_{i=1}^{N_g} y_{ji}^{(g)}.$$

Under the normal population, they derive that the limiting null distribution of T_M is the chi-square distribution with $(k-1)$ degrees of freedom. But the limiting distribution of T_M may not be a simple expression. Therefore, they adopt a permutation test for T_M.

They investigate the type I error and the power of tests at a significance level of 5%. Simulations are carried out for normal populations, contaminated normal populations as an example of symmetric non-normal populations and skew normal populations as an example of asymmetric non-normal populations. Let the number of populations be three. The simulations are repeated 100,000 times with 10,000 permutations for the criterion T_M. Therefore, the type I error and the power maintain precision down to two decimal places. They set the critical value of the Ansari-Bradley test and the Mood test as 5.991 for $k = 3$.

Gibbons and Chakraborti (2003) indicate that the asymptotic relative efficiency of the Mood test is higher than that of the Ansari-Bradley test. They simulate for the criteria AB_j and M_j, and investigate the type I error and the power of the test using the criteria T_M, AB_j and M_j. They set the populations as normal populations $N_3(\mathbf{0}, \Sigma_g)$ and contaminated normal populations $0.95 \times N_3(\mathbf{0}, \Sigma_g) + 0.05 \times N_3(\mathbf{0}, 3\Sigma_g)$, and the sample sizes are $N_1 = N_2 = N_3 = 50$, $N_1 = N_2 = N_3 = 100$ and $N_1 = 200$, $N_2 = 150$, $N_3 = 100$. Since the distribution of the eigenvalues may not depend on the eigenvector, it is sufficient to examine only the case that the population covariance matrices are diagonal without loss of generality. They set the population covariance matrices as

$$\Sigma_g = \mathrm{diag}\left(\lambda_1^{(g)}, \lambda_2^{(g)}, \lambda_3^{(g)} \right)$$

and treat the following four cases.

Case 1:

$$\lambda_1^{(1)} = 6,\ \lambda_2^{(1)} = 3,\ \lambda_3^{(1)} = 1; \quad \lambda_1^{(2)} = 6,\ \lambda_2^{(2)} = 3,\ \lambda_3^{(2)} = 1; \quad \lambda_1^{(3)} = 6,\ \lambda_2^{(3)} = 3,\ \lambda_3^{(3)} = 1,$$

Case 2:

$$\lambda_1^{(1)} = 10,\ \lambda_2^{(1)} = 3,\ \lambda_3^{(1)} = 1; \quad \lambda_1^{(2)} = 8,\ \lambda_2^{(2)} = 3,\ \lambda_3^{(2)} = 1; \quad \lambda_1^{(3)} = 6,\ \lambda_2^{(3)} = 3,\ \lambda_3^{(3)} = 1,$$

Case 3:

$$\lambda_1^{(1)} = 9,\ \lambda_2^{(1)} = 5,\ \lambda_3^{(1)} = 2; \quad \lambda_1^{(2)} = 6,\ \lambda_2^{(2)} = 3,\ \lambda_3^{(2)} = 1; \quad \lambda_1^{(3)} = 6,\ \lambda_2^{(3)} = 3,\ \lambda_3^{(3)} = 1,$$

Case 4:

$$\lambda_1^{(1)} = 9,\ \lambda_2^{(1)} = 5,\ \lambda_3^{(1)} = 2;\ \lambda_1^{(2)} = 7.5,\ \lambda_2^{(2)} = 4,\ \lambda_3^{(2)} = 1.5; \lambda_1^{(3)} = 6,\ \lambda_2^{(3)} = 3,\ \lambda_3^{(3)} = 1.$$

Case 1 is under the null hypothesis for all eigenvalues. Case 2 is that only the largest eigenvalues are different and other eigenvalues are same. Case 3 and Case 4 is that all eigenvalues are different. In the all of Tables, the results on the left of the dividing lines shows the type I error of tests. And, the results on the right reveal the power of tests. Table 1 and Table 2 represent the results for the normal population.

Table 1. Normal populations

		$N_1 = N_2 = N_3 = 50$				$N_1 = N_2 = N_3 = 100$			
		Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
$j = 1$	M_{1k}	0.041	0.242	0.224	0.169	0.046	0.478	0.428	0.327
	AB_{1k}	0.042	0.201	0.184	0.144	0.047	0.395	0.348	0.269
	T_M	**0.040**	**0.302**	**0.269**	**0.206**	**0.045**	**0.593**	**0.531**	**0.415**
$j = 2$	M_{2k}	0.048	0.049	0.335	0.253	0.049	0.050	0.615	0.488
	AB_{2k}	0.048	0.049	0.271	0.211	0.049	0.050	0.511	0.403
	T_M	**0.047**	**0.049**	**0.408**	**0.315**	**0.048**	**0.050**	**0.733**	**0.601**
$j = 3$	M_{3k}	0.056	0.056	0.564	0.445	0.053	0.053	0.867	0.763
	AB_{3k}	0.055	0.055	0.465	0.370	0.052	0.052	0.776	0.664
	T_M	**0.057**	**0.056**	**0.666**	**0.550**	**0.054**	**0.054**	**0.942**	**0.871**

Table 2. Normal populations

		$N_1 = 150, N_2 = 100, N_3 = 50$				$N_1 = 200, N_2 = 150, N_3 = 100$			
		Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
$j = 1$	M_{1k}	0.045	0.367	0.453	0.241	0.047	0.613	0.641	0.420
	AB_{1k}	0.046	0.309	0.375	0.207	0.048	0.518	0.538	0.349
	T_M	**0.043**	**0.458**	**0.408**	**0.307**	**0.046**	**0.737**	**0.687**	**0.529**
$j = 2$	M_{2k}	0.049	0.050	0.695	0.428	0.049	0.050	0.852	0.641
	AB_{2k}	0.049	0.050	0.594	0.361	0.049	0.050	0.761	0.545
	T_M	**0.049**	**0.050**	**0.680**	**0.534**	**0.049**	**0.050**	**0.898**	**0.768**
$j = 3$	M_{3k}	0.055	0.055	0.928	0.703	0.052	0.052	0.985	0.898
	AB_{3k}	0.055	0.055	0.861	0.613	0.051	0.052	0.956	0.823
	T_M	**0.060**	**0.060**	**0.933**	**0.815**	**0.055**	**0.055**	**0.994**	**0.962**

The simulation results show the validity of the permutation test and nonparametric tests. Every test may be conservative for $j = 1, 2$ under the null hypothesis. The nonparametric tests are more robust than the test using the criterion T_M. However, the power of the test using the criterion T_M is the greatest in both cases regardless of whether the sample sizes are equal or unequal, except only for Case 3 in Table 2. The power of the Mood test is greatest for $j = 1$ and 2 of Case 3 on the left of Table 2. The total sample sizes are equal in the right of Table 1 and the left of Table 2. The power of the test using the criterion T_M tends to decrease when the sample sizes are unequal.

Further when they apply the permutation test for the criteria M_j and AB_j, the powers are almost equivalent to the power of the Mood test and the Ansari-Bradley test.

In the study of environmental problems, the contaminated normal distribution is used a lot. Therefore they treat the contaminated normal population as a symmetric non-normal population in Table 3 and Table 4. Table 3 indicates the results for the cases of equal sample sizes. They show the results for unequal sample sizes in Table 4.

Table 3. Contaminated normal populations

		$N_1 = N_2 = N_3 = 50$				$N_1 = N_2 = N_3 = 100$			
		Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
$j = 1$	M_{1k}	0.040	0.233	0.215	0.163	0.044	0.462	0.411	0.311
	AB_{1k}	0.042	0.198	0.178	0.141	0.045	0.386	0.337	0.260
	T_M	**0.033**	**0.261**	**0.228**	**0.174**	**0.037**	**0.512**	**0.452**	**0.344**
$j = 2$	M_{2k}	0.048	0.048	0.329	0.248	0.047	0.048	0.601	0.473
	AB_{2k}	0.048	0.048	0.268	0.208	0.048	0.048	0.501	0.395
	T_M	**0.046**	**0.047**	**0.379**	**0.293**	**0.043**	**0.044**	**0.672**	**0.541**
$j = 3$	M_{3k}	0.055	0.055	0.555	0.438	0.051	0.050	0.858	0.751
	AB_{3k}	0.054	0.054	0.460	0.366	0.051	0.050	0.769	0.655
	T_M	**0.056**	**0.058**	**0.639**	**0.524**	**0.049**	**0.050**	**0.913**	**0.824**

Table 4. Contaminated normal populations

		$N_1 = 150, N_2 = 100, N_3 = 50$				$N_1 = 200, N_2 = 150, N_3 = 100$			
		Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
$j = 1$	M_{1k}	0.042	0.373	0.452	0.247	0.045	0.593	0.617	0.401
	AB_{1k}	0.044	0.318	0.377	0.214	0.046	0.505	0.521	0.338
	T_M	**0.034**	**0.425**	**0.355**	**0.279**	**0.038**	**0.653**	**0.587**	**0.447**
$j = 2$	M_{2k}	0.048	0.049	0.691	0.431	0.047	0.047	0.836	0.621
	AB_{2k}	0.050	0.050	0.593	0.365	0.048	0.048	0.745	0.531
	T_M	**0.046**	**0.047**	**0.638**	**0.514**	**0.045**	**0.043**	**0.841**	**0.700**
$j = 3$	M_{3k}	0.055	0.055	0.926	0.705	0.050	0.050	0.982	0.887
	AB_{3k}	0.055	0.055	0.860	0.616	0.050	0.050	0.951	0.812
	T_M	**0.058**	**0.058**	**0.911**	**0.798**	**0.049**	**0.049**	**0.986**	**0.935**

The simulation results permit that we may use the test using the criterion T_M for the contaminated normal population. All tests may be conservative for $j = 1, 2$ under the null hypothesis. If any N_i is greater than 100, all tests are also conservative for $j = 3$. In addition, the nonparametric tests are more robust than the test using the criterion T_M, especially $j = 1$. But the power of the test using the criterion T_M is the greatest in these cases regardless of whether the sample sizes are equal. For Case 3 on unequal sample sizes, the Mood test is more suitable than the test using the criterion T_M. However, the power of the test using the criterion T_M is higher than that of the

nonparametric tests for Case 2 and Case 4. They further applied the permutation test for the criteria M_{jk} and AB_{jk}. Their powers are almost equivalent to the power of the original tests. The total sample sizes are equal in the right of Table 3 and the left of Table 4, but the power of the test using the criterion T_M tends to decrease when the sample sizes are unequal.

Finally, they simulate the skew normal populations $SN_3(\mathbf{0}, \Sigma_g, \boldsymbol{\nu}^{(g)})$ for an asymmetric non-normal population, where $\boldsymbol{\nu}^{(g)} = (\nu_1, \nu_2, \nu_3)'$ denotes the shape parameter (cf. Genton(2004), *Chap.* 1, *Sec.* 5). They set $\nu_1 = \nu_2 = \nu_3 = 0.5$. The reason why they treat this population is that they sometimes assume this population in financial markets, property markets, labor markets *etc.*

Table 5 to Table 8 represents the results for the skew normal populations. In Table 5 and Table 6, they give the type I error and the power for the case of equal sample sizes. For the case of the unequal sample sizes, we show the type I error and the power in Table 7 and Table 8. Let the correlation matrix be

$$\begin{pmatrix} 1 & r_1 & r_2 \\ r_1 & 1 & r_3 \\ r_2 & r_3 & 1 \end{pmatrix}.$$

In the case of $r = r_1 = r_2 = r_3$, the results are listed in the left of Table 5 - Table 8. When $r_1 \neq r_2 \neq r_3$, they show the results in the right of Table 5 - Table 8.

Table 5. Skew normal populations ($N_1 = N_2 = N_3 = 50$)

		$r = 0.5$				$r_1 = 0.5, r_2 = 0.3, r_3 = 0.2$			
		Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
$j = 1$	M_{1k}	0.049	0.180	0.270	0.207	0.049	0.190	0.258	0.198
	AB_{1k}	0.049	0.155	0.226	0.174	0.049	0.161	0.211	0.167
	T_M	**0.049**	**0.220**	**0.326**	**0.254**	**0.049**	**0.232**	**0.312**	**0.243**
$j = 2$	M_{2k}	0.050	0.057	0.319	0.243	0.045	0.051	0.321	0.241
	AB_{2k}	0.050	0.055	0.259	0.203	0.046	0.050	0.261	0.201
	T_M	**0.051**	**0.060**	**0.389**	**0.302**	**0.045**	**0.051**	**0.391**	**0.299**
$j = 3$	M_{3k}	0.056	0.056	0.545	0.429	0.054	0.054	0.548	0.428
	AB_{3k}	0.055	0.055	0.449	0.356	0.053	0.053	0.450	0.356
	T_M	**0.058**	**0.058**	**0.645**	**0.529**	**0.054**	**0.055**	**0.650**	**0.530**

For equal sample sizes, the simulation results reveal that all tests maintain a significant level for Case 1. The nonparametric tests are more robust than the test using the criterion T_M. However, the power of the test using the criterion T_M is higher than the power of the Ansari-Bradley test or the Mood test under the alternative hypothesis. Next Table 7 and Table 8 represent the results for unequal sample sizes.

For the cases of unequal sample sizes, the simulation results reveal that all tests

Table 6. Skew normal populations ($N_1 = N_2 = N_3 = 100$)

		$r = 0.5$				$r_1 = 0.5, r_2 = 0.3, r_3 = 0.2$			
		Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
$j = 1$	M_{1k}	0.050	0.336	0.500	0.387	0.050	0.356	0.479	0.371
	AB_{1k}	0.050	0.277	0.409	0.319	0.050	0.293	0.391	0.305
	T_M	**0.050**	**0.423**	**0.609**	**0.486**	**0.049**	**0.448**	**0.586**	**0.466**
$j = 2$	M_{2k}	0.050	0.063	0.578	0.456	0.048	0.058	0.585	0.459
	AB_{2k}	0.050	0.061	0.478	0.376	0.049	0.056	0.482	0.379
	T_M	**0.050**	**0.066**	**0.695**	**0.566**	**0.048**	**0.059**	**0.703**	**0.570**
$j = 3$	M_{3k}	0.053	0.053	0.851	0.743	0.052	0.053	0.859	0.753
	AB_{3k}	0.052	0.053	0.757	0.643	0.052	0.052	0.766	0.653
	T_M	**0.054**	**0.054**	**0.930**	**0.851**	**0.053**	**0.053**	**0.936**	**0.861**

Table 7. Skew normal populations ($N_1 = 150, N_2 = 100, N_3 = 50$)

		$r = 0.5$				$r_1 = 0.5, r_2 = 0.3, r_3 = 0.2$			
		Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
$j = 1$	M_{1k}	0.049	0.268	0.545	0.306	0.049	0.282	0.522	0.291
	AB_{1k}	0.049	0.229	0.455	0.260	0.049	0.241	0.435	0.248
	T_M	**0.049**	**0.336**	**0.511**	**0.385**	**0.048**	**0.354**	**0.486**	**0.367**
$j = 2$	M_{2k}	0.051	0.062	0.647	0.385	0.047	0.054	0.645	0.377
	AB_{2k}	0.051	0.061	0.548	0.326	0.048	0.055	0.545	0.318
	T_M	**0.052**	**0.073**	**0.625**	**0.484**	**0.047**	**0.062**	**0.621**	**0.476**
$j = 3$	M_{3k}	0.055	0.056	0.915	0.679	0.055	0.055	0.925	0.700
	AB_{3k}	0.055	0.055	0.843	0.589	0.054	0.055	0.856	0.609
	T_M	**0.060**	**0.062**	**0.919**	**0.793**	**0.060**	**0.062**	**0.930**	**0.814**

Table 8. Skew normal populations ($N_1 = 200, N_2 = 150, N_3 = 100$)

		$r = 0.5$				$r_1 = 0.5, r_2 = 0.3, r_3 = 0.2$			
		Case 1	Case 2	Case 3	Case 4	Case 1	Case 2	Case 3	Case 4
$j = 1$	M_{1k}	0.050	0.443	0.731	0.505	0.050	0.468	0.708	0.483
	AB_{1k}	0.050	0.369	0.628	0.422	0.050	0.390	0.605	0.403
	T_M	**0.049**	**0.552**	**0.781**	**0.624**	**0.049**	**0.582**	**0.757**	**0.599**
$j = 2$	M_{2k}	0.050	0.068	0.814	0.594	0.049	0.061	0.816	0.593
	AB_{2k}	0.050	0.065	0.716	0.502	0.049	0.059	0.718	0.500
	T_M	**0.050**	**0.077**	**0.863**	**0.719**	**0.048**	**0.068**	**0.865**	**0.719**
$j = 3$	M_{3k}	0.053	0.053	0.981	0.882	0.052	0.053	0.984	0.894
	AB_{3k}	0.052	0.053	0.947	0.802	0.052	0.053	0.953	0.812
	T_M	**0.055**	**0.055**	**0.992**	**0.952**	**0.054**	**0.055**	**0.993**	**0.959**

maintain the significance level for $j = 1$ and 2 in Case 1. However, in Case 2 for $j = 2$, the Mood test, the Ansari-Bradley test and their test don't keep the significance level well. For $j = 3$ in Case 1 and Case 2, the nonparametric tests maintain the significance level. Therefore the Mood test and the Ansari-Bradley test are more robust than their test. In Case 3 of Table 7, the Mood test is more powerful than their test, but the power of their test is higher than that of nonparametric tests under other alternative hypothesis.

Through the simulation studies, every test may be conservative for $j = 1, 2$ under the null hypothesis and the nonparametric tests are more robust than the test using the criterion T_M. However, the power of the test using the criterion T_M is the greatest in both cases regardless of whether the sample sizes are equal or unequal. Therefore, they expect that the proposed test procedure is suitable for a multipopulation and this is shown by simulation.

Tsukada and Murakami(2008) propose the criterion as follows:

$$\sum_{g_1<g_2}^{k} \frac{N r_{g_1} r_{g_2}}{2r} \left(\log l_{\alpha}^{(g_1)} - \log l_{\alpha}^{(g_2)} \right)^2, \tag{0.5}$$

where $r_g = N_g/N$, $r = \sum_{g=1}^{k} r_g$ and $N = \sum_{g=1}^{k} N_g$.

They compare a permutation test using above criterion, the Mood test, test by chi-squared criterion using sample mean in Fligner and Killeen(1976), bootstrap test by Boos and Brownie(1989), randomization test by Wludyka and Sa(2004), Levene test by Levene(1960), Brown-Forsythe test using 10% trimmed mean by Brown and Forsythe(1974) and O'Brien's Test(1979) as tests for the equivalence of variances. They compare the actual significance level and the power of tests by simulation.

The test proposed by them is superior compared with the rank test as the Ansari-Bradley test and the Mood test. But the test applying bootstrap method is superior to their method under the alternative hypothesis which they adopt. In the case that the sample size is unbalanced, the procedure applying O'Brien test is superior to the test using bootstrap method.

0.2.3. Covariance for principal components

Nonparametric test requires the independence of each sample. But $y_{\alpha i}$ and $y_{\alpha k}$ are no longer independent. Now we evaluate the degree of dependence. We omit the suffix representing the population and let $E(\boldsymbol{x}_i) = \mathbf{0}$ without less of generality. When $\Sigma = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$ and λ_k is simple, the covariance of $y_{\alpha i}$ and $y_{\alpha k}$ is as follows:

$$\begin{aligned} E[y_{\alpha i} y_{\alpha k}] &= E\left[\sum_{u=1}^{p} \sum_{v=1}^{p} h_{u\alpha} h_{v\alpha} x_{ui} x_{vk} \right] \\ &= -\frac{2}{n^2} \sum_{l \neq \alpha}^{p} \lambda_{l\alpha}^2 m_{\alpha l}^{21} m_{\alpha l}^{21} - \frac{1}{n^2} \sum_{u \neq \alpha}^{p} \lambda_{u\alpha}^2 (m_{\alpha u}^{21} m_{\alpha u}^{21} + m_{\alpha}^{3} m_{\alpha u}^{12}) \end{aligned}$$

$$+\frac{1}{n^2}\sum_{\substack{l,u\neq\alpha\\ u\neq l}}^{p}\lambda_{u\alpha}\lambda_{l\alpha}(m^{21}_{ul}m^{21}_{\alpha l}+m^{111}_{ul\alpha}m^{111}_{ul\alpha})-\frac{1}{n^2}\sum_{v\neq\alpha}^{p}\lambda^2_{v\alpha}(m^3_{\alpha}m^{12}_{\alpha v}+m^{21}_{\alpha v}m^{21}_{\alpha v})$$
$$+\frac{1}{n^2}\sum_{\substack{v,l\neq\alpha\\ v\neq l}}^{p}\lambda_{v\alpha}\lambda_{l\alpha}(m^{111}_{vl\alpha}m^{111}_{vl\alpha}+m^{21}_{vl}m^{21}_{\alpha l})$$
$$+\frac{1}{n^2}\sum_{u,v\neq\alpha}^{p}\lambda_{u\alpha}\lambda_{v\alpha}(m^{12}_{\alpha u}m^{12}_{\alpha v}+m^{111}_{\alpha uv}m^{111}_{\alpha uv})+O(n^{-3}) \tag{0.6}$$

where $\lambda_{\alpha\beta}=(\lambda_\alpha-\lambda_\beta)^{-1}$, the third moments denote $m^3_i=E(x_ix_ix_i)$, $m^{21}_{ik}=E(x_ix_ix_k)$, $m^{12}_{ik}=E(x_ix_kx_k)$ and $m^{111}_{ikt}=E(x_ix_kx_t)$. Though we do not express terms of higher order for the above expansion, the expansion consists of odd-order moments. For a symmetric population,

$$E[y_{\alpha i}y_{\alpha k}]=0,$$

the degree of dependence may be very weak. This expansion shows that the degree of dependence is weak when the sample size is sufficiently large. Therefore, for large sample we may ignore the influence of dependence. But the degree of dependence may be influenced by the third moments for an asymmetric population.

0.3. Test for eigenvector

0.3.1. Test for a specified vector

Tsukada(1997,1998) and Tsukada and Sugiyama(1997) investigate a hypothesis testing for eigenvector. He considers the following hypothesis

$$H_0 : \boldsymbol{\eta}_j=\boldsymbol{\eta}_0,$$
$$H_1 : \boldsymbol{\eta}_j\neq\boldsymbol{\eta}_0,$$

where $\boldsymbol{\eta}_0$ is a specified vector. Concerning this hypothesis testing, he compare the power of tests using the inner product of the sample eigenvector and the specified vector, the ratio of the principal score and that of $\boldsymbol{\eta}_0'(\boldsymbol{X}_i-\bar{\boldsymbol{X}})$, the criterion by Anderson, the following T_{v1} and T_{v2} obtained by the diference of vectors,

$$T_{v1} = n\left(\boldsymbol{\eta}_0'S^2\boldsymbol{\eta}_0-2l_j\boldsymbol{\eta}_0'S\boldsymbol{\eta}_0+l_j^2\right),$$
$$T_{v2} = n\left(\frac{1}{l_j}\boldsymbol{\eta}_0'S^2\boldsymbol{\eta}_0-2\boldsymbol{\eta}_0'S\boldsymbol{\eta}_0+l_j\right),$$

where $n=N-1$. Since the asymptotic distribution of these criteria are complicate, bootstrap test is proposed.

(1) Calculate the sample covariance matrix S_X, the orthogonal matrix H which diagonalize S_X, and a criterion $T(\boldsymbol{X})$ from the original sample.

(2) Calculate the following orthogonal matrix V based on the specified vector $\boldsymbol{\eta}_0$

$$V=[\boldsymbol{\xi}_1,\ldots,\boldsymbol{\xi}_{j-1},\boldsymbol{\eta}_0,\boldsymbol{\xi}_{j+1},\ldots,\boldsymbol{\xi}_p]$$

by Gram-Schmidt procedure ($\boldsymbol{\xi}_i$ is not the eigenvector of S_X).

(3) Transform the original sample $\boldsymbol{X}_i$ as follows:

$$\boldsymbol{Y}_i = VH'\boldsymbol{X}_i \qquad (i = 1, \ldots, N).$$

(4) We make the bootstrap sample $\boldsymbol{Y}^* = \{\boldsymbol{Y}_1^*, \ldots, \boldsymbol{Y}_N^*\}$ from the transformed original sample $\boldsymbol{Y} = \{\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_N\}$ and calculate $T(\boldsymbol{Y}^*)$. We generate B bootstrap values $\{T(\boldsymbol{Y}_1^*), \ldots, T(\boldsymbol{Y}_B^*)\}$ of $T(\boldsymbol{Y}^*)$ and estimate the achieved significance level by

$$\sharp\{T(\boldsymbol{Y}^*) \geq T(\boldsymbol{X})\}/B = \sum_{i=1}^{B} I\{T(\boldsymbol{Y}_i^*) \geq T(\boldsymbol{X})\}/B,$$

where the indicator function $I(\cdot)$ is 1 when $T(\boldsymbol{Y}_i^*) \geq T(\boldsymbol{X})$ and 0 when $T(\boldsymbol{Y}_i^*) < T(\boldsymbol{X})$.

(5) Formally we choose a small probability α, like .05 or .01, and reject H_0 if the achieved significance level is less than α.

As long as using the criterion by Anderson, T_{v1} and T_{v2}, this procedure does not depend on the choice of the vectors $\{\boldsymbol{\xi}_1, \ldots, \boldsymbol{\xi}_{j-1}, \boldsymbol{\xi}_{j+1}, \ldots, \boldsymbol{\xi}_p\}$. Because when we calculate the factor $\boldsymbol{\eta}_0' S_{Y^*} \boldsymbol{\eta}_0$ included in the criteria for the data $\boldsymbol{Y}^*$, we have

$$\boldsymbol{\eta}_0' S_{Y^*} \boldsymbol{\eta}_0 = \boldsymbol{\eta}_0' VH' S_{X^*} HV' \boldsymbol{\eta}_0 = \boldsymbol{e}_j' H' S_{X^*} H \boldsymbol{e}_j,$$

where $\boldsymbol{e}_j$ is the vector that j-th element is 1 and other elements are 0.

The following is the reason why we transform the original sample $\boldsymbol{X}$ into the sample $\boldsymbol{Y}$ in (3). Calculating the covariance matrix of the sample $\boldsymbol{Y}$, we have

$$\begin{aligned} S_Y &= \sum_{i=1}^{N} \left(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}}\right)\left(\boldsymbol{Y}_i - \bar{\boldsymbol{Y}}\right)'/n = VH' \sum_{i=1}^{N} \left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}\right)\left(\boldsymbol{X}_i - \bar{\boldsymbol{X}}\right)' HV'/n \\ &= VH' S_X HV' = V\mathrm{diag}(l_1, \ldots, l_j, \ldots, l_p)V'. \end{aligned}$$

This is nothing but the eigenvalue decomposition of S_Y, the eigenvector corresponding to the j-th largest eigenvalue of the covariance matrix of the sample $\boldsymbol{Y}$ is the specified vector $\boldsymbol{\eta}_0$. Hence the sample $\boldsymbol{Y}$ follows the distribution under the hypothesis H_0.

0.3.2. Two-populations case

We assumed that the existence of the finite fourth-order moments of population distribution and the covariance matrix are nonsingular and without multiple eigenvalues. Kollo and Neudecker(1995) obtain that the eigenvector $\sqrt{n}\left(\boldsymbol{h}_j - \boldsymbol{\eta}_j\right)$ is asymptotically distributed as

$$N(\boldsymbol{0}, V_j),$$

where

$$\begin{aligned} V_j &= \{\boldsymbol{\eta}_j' \otimes \Gamma(\lambda_j I - \Lambda)^{+}\Gamma'\} M_4(x) \{\boldsymbol{\eta}_j \otimes \Gamma(\lambda_j I - \Lambda)^{+}\Gamma'\}, \\ M_4(x) &= E[(\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})' \otimes (\boldsymbol{x} - \boldsymbol{\mu})(\boldsymbol{x} - \boldsymbol{\mu})'], \end{aligned}$$

where A^+ is its Moore-Penrose inverse, $\Gamma = [\boldsymbol{\eta}_1, \ldots, \boldsymbol{\eta}_p]$ and $\Lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$.

By applying this result, we can test the following hypothesis

$$H_0 : \boldsymbol{\eta}_j^{(1)} = \boldsymbol{\eta}_j^{(2)}$$
$$H_1 : \text{not } H_0.$$

Tsukada(2007) investigate Wald criterion for testing the hypothesis. He simulate the accuracy for criteria to the nominal significance level. He set the hypothesis as

$$H_0 : \boldsymbol{\eta}_1^{(1)} = \boldsymbol{\eta}_1^{(2)},$$
$$H_1 : \boldsymbol{\eta}_1^{(1)} \neq \boldsymbol{\eta}_1^{(2)},$$

and let the sample size be 100, 200, 500 and 1000. For each situation the significance level is estimated from a million simulations. The nominal significance level used is 0.05. He sets the populations as follows:

Normal Population

We set $\Lambda_{31} = \mathrm{diag}(7, 2, 1)$, $\Lambda_{51} = \mathrm{diag}(52, 16, 4, 2, 1)$, $\Lambda_{71} = \mathrm{diag}(160, 50, 8, 6, 4, 2, 1)$, $\Lambda_{32} = \mathrm{diag}(12, 2, 1)$, $\Lambda_{52} = \mathrm{diag}(84, 11, 6, 3, 1)$ and $\Lambda_{72} = \mathrm{diag}(240, 30, 16, 8, 4, 2, 1)$. The variation explained for the first and second principal component are about 0.7 and 0.2 in the case of Λ_{p1} which is called Case 5. In the case of Λ_{p2}, the variation explained for the first and second principal component are about 0.8 and 0.1 and we call this Case 6.

Let $\Sigma_{31} = \Lambda_{3g}$ and $\Sigma_{32} = \Gamma_3 \Lambda_{3g} \Gamma_3'$ $(g = 1, 2)$ in the case of the tri-variate distribution$(p = 3)$, where

$$\Gamma_3 = \begin{pmatrix} 1.000 & 0.000 & 0.000 \\ 0.000 & 0.985 & 0.174 \\ 0.000 & -0.174 & 0.985 \end{pmatrix}.$$

In the case of penta-variate distribution$(p = 5)$, let $\Sigma_{51} = \Lambda_{5g}$ and $\Sigma_{52} = \Gamma_5 \Lambda_{5g} \Gamma_5'$ $(g = 1, 2)$, where

$$\Gamma_5 = \begin{pmatrix} 1.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.989 & 0.071 & 0.086 & 0.101 \\ 0.000 & -0.086 & 0.990 & 0.072 & 0.086 \\ 0.000 & -0.087 & -0.086 & 0.990 & 0.071 \\ 0.000 & -0.087 & -0.087 & -0.086 & 0.989 \end{pmatrix}.$$

In the case of hepta-variate distribution$(p = 7)$, let $\Sigma_{71} = \Lambda_{7g}$ and $\Sigma_{72} = \Gamma_7 \Lambda_{7g} \Gamma_7'$ $(g = 1, 2)$, where

$$\Gamma_7 = \begin{pmatrix} 1.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 & 0.000 \\ 0.000 & 0.981 & -0.086 & -0.086 & -0.086 & -0.087 & -0.087 \\ 0.000 & 0.056 & 0.984 & -0.084 & -0.085 & -0.086 & -0.087 \\ 0.000 & 0.069 & 0.058 & 0.985 & -0.083 & -0.085 & -0.086 \\ 0.000 & 0.083 & 0.070 & 0.058 & 0.985 & -0.084 & -0.086 \\ 0.000 & 0.098 & 0.084 & 0.070 & 0.058 & 0.984 & -0.086 \\ 0.000 & 0.114 & 0.098 & 0.083 & 0.069 & 0.056 & 0.981 \end{pmatrix}.$$

We set the populations as follows:

$$\Pi_1 : N_p(\mathbf{0}, \Sigma_{p1}),\ \Pi_2 : N_p(\mathbf{0}, \Sigma_{p2}).$$

Contaminated Normal Population

As the elliptical population, we set the population as follows:

$$\Pi_1 : 0.1N_p(\mathbf{0}, \Sigma_{p1}) + 0.9N_p(\mathbf{0}, 4^2\Sigma_{p1}),\ \Pi_2 : 0.1N_p(\mathbf{0}, \Sigma_{p2}) + 0.9N_p(\mathbf{0}, 4^2\Sigma_{p2}),$$

where Σ_{31}, Σ_{32}, Σ_{51}, Σ_{52}, Σ_{71} and Σ_{72} are used in the above.

Skew Normal Population

As the skew population, we choice the skew normal population (cf. Genton(2004)) as follows:

$$\Pi_1 : SN_p(\mathbf{0}, \Omega_{p1}, \boldsymbol{\alpha}_{p1}),\ \Pi_2 : SN_p(\mathbf{0}, \Omega_{p2}, \boldsymbol{\alpha}_{p2}),$$

where

$$\begin{aligned}
\boldsymbol{\alpha}_{31} &= (-1.063, 6.426, 41.26)'/100,\ \boldsymbol{\alpha}_{32} = (-1.197, 8.723, 39.46)'/100,\\
\boldsymbol{\alpha}_{51} &= (-3.211, -8.285, 4.187, 10.79, 61.40)'/100,\\
\boldsymbol{\alpha}_{52} &= (-3.364, 3.228, 10.38, 7.783, 42.48)'/100,\\
\boldsymbol{\alpha}_{71} &= (-0.694, -2.069, -5.664, -3.705, 5.984, 81.21, 439.4)'/1000,\\
\boldsymbol{\alpha}_{72} &= (-0.752, 1.741, -0.406, -0.912, 2.090, 39.14, 209.8)'/1000,
\end{aligned}$$

$$\Omega_{31} = \begin{pmatrix} 49.00 & 3.500 & 1.750 \\ 3.500 & 4.000 & 0.500 \\ 1.750 & 0.500 & 1.000 \end{pmatrix}, \Omega_{32} = \begin{pmatrix} 49.00 & 3.447 & 1.803 \\ 3.447 & 3.880 & 0.325 \\ 1.803 & 0.325 & 1.061 \end{pmatrix},$$

$$\Omega_{51} = \begin{pmatrix} 2704.0 & 208.0 & 52.00 & 26.00 & 13.00 \\ 208.0 & 256.0 & 16.00 & 8.000 & 4.000 \\ 52.00 & 16.00 & 16.00 & 2.000 & 1.000 \\ 26.00 & 8.000 & 2.000 & 4.000 & 0.500 \\ 13.00 & 4.000 & 1.000 & 0.500 & 1.000 \end{pmatrix}, \Omega_{52} = \begin{pmatrix} 2704.0 & 203.88 & 52.75 & 27.50 & 14.87 \\ 203.88 & 245.97 & 9.540 & 3.018 & 0.297 \\ 52.75 & 9.540 & 16.46 & 1.990 & 0.916 \\ 27.50 & 3.018 & 1.990 & 4.474 & 0.664 \\ 14.87 & 0.297 & 0.916 & 0.664 & 1.309 \end{pmatrix},$$

$$\Omega_{71} = \begin{pmatrix} 25600.0 & 2000.0 & 320.0 & 240.0 & 160.0 & 80.00 & 40.00 \\ 2000.0 & 2500.0 & 100.0 & 75.00 & 50.00 & 25.00 & 12.50 \\ 320.0 & 100.0 & 64.00 & 12.00 & 8.00 & 4.00 & 2.00 \\ 240.0 & 75.00 & 12.00 & 36.00 & 6.00 & 3.00 & 1.50 \\ 160.0 & 50.00 & 8.00 & 6.00 & 16.00 & 2.00 & 1.00 \\ 80.00 & 25.00 & 4.00 & 3.00 & 2.00 & 4.00 & 0.50 \\ 40.00 & 12.50 & 2.00 & 1.50 & 1.00 & 0.50 & 1.00 \end{pmatrix},$$

$$\Omega_{72} = \begin{pmatrix} 25600.0 & 1929.7 & 326.9 & 251.2 & 174.6 & 97.83 & 59.68 \\ 1929.7 & 2327.4 & 43.54 & 25.95 & 9.454 & -4.404 & -9.263 \\ 326.9 & 43.54 & 66.81 & 13.31 & 8.778 & 4.375 & 2.311 \\ 251.2 & 25.95 & 13.31 & 39.43 & 7.569 & 3.997 & 2.278 \\ 174.6 & 9.454 & 8.778 & 7.569 & 19.06 & 3.315 & 2.006 \\ 97.83 & -4.404 & 4.375 & 3.997 & 3.315 & 5.982 & 1.481 \\ 59.68 & -9.263 & 2.311 & 2.278 & 2.006 & 1.481 & 2.226 \end{pmatrix}.$$

Table 9. Estimated Significance Level for each criterion (Case 5)

	$p=3$				$p=5$				$p=7$			
Normal population												
N_1, N_2	C_N	C_E	C_G	C_F	C_N	C_E	C_G	C_F	C_N	C_E	C_G	C_F
100	.0659	.0738	.0739	.0528	.0744	.0849	.1143	.0649	.0951	.1090	.2006	.0966
200	.0577	.0612	.0637	.0515	.0638	.0685	.0966	.0612	.0803	.0866	.1683	.0898
500	.0536	.0548	.0586	.0514	.0583	.0599	.0865	.0594	.0721	.0741	.1500	.0851
1000	.0523	.0528	.0571	.0513	.0565	.0571	.0832	.0587	.0696	.0705	.1448	.0840
Contaminated Normal population												
N_1, N_2	C_N	C_E	C_G	C_F	C_N	C_E	C_G	C_F	C_N	C_E	C_G	C_F
100	.0874	.0763	.0759	.0526	.1061	.0885	.1189	.0649	.1380	.1139	.2070	.0960
200	.0777	.0629	.0651	.0517	.0925	.0705	.0991	.0616	.1181	.0881	.1708	.0888
500	.0719	.0551	.0590	.0512	.0851	.0606	.0875	.0597	.1074	.0747	.1508	.0849
1000	.0702	.0529	.0572	.0513	.0814	.0570	.0834	.0586	.1042	.0707	.1454	.0843
Skew Normal population												
N_1, N_2	C_N	C_E	C_G	C_F	C_N	C_E	C_G	C_F	C_N	C_E	C_G	C_F
100	.0667	.0740	.0620	.0535	.0740	.0837	.0680	.0634	.0863	.0987	.0809	.0791
200	.0588	.0617	.0537	.0527	.0644	.0681	.0570	.0603	.0728	.0775	.0625	.0728
500	.0551	.0555	.0499	.0528	.0597	.0603	.0536	.0593	.0666	.0675	.0582	.0703
1000	.0545	.0542	.0493	.0533	.0594	.0590	.0534	.0601	.0660	.0657	.0592	.0710

The population eigenvectors are same as the normal population and the contaminated normal population in the skew normal population.

Table 9 and Table 10 represent the results in Case 5 and Case 6, respectively. In Tables, C_F represents the results adapted partial common principal component model (see Flury(1987)), C_N does the results in the case that the asymptotic covariance matrix V_i of the eigenvector is rewritten under the normal population, C_E does the results in the case that V_i is rewritten under the elliptical population, and C_G does the results in the case that V_i is rewritten under the general population which has finite fourth moments. An underlined value represents the nearest value to the nominal significance level among four criteria.

In the normal population, the estimated significance level for C_F or C_N is close to the nominal significance level. Because they are derived under the normal population, this is a natural result. When the dimension is high, the accuracy for C_N is better than that of C_F. Since kurtosis are estimated by force though the population is normal, the accuracy for C_G is not good. The estimated significance level for C_E converges nearly to 0.05 when the sample size is large.

Under the contaminated normal population, the estimated significance level for C_E is near the nominal significance level when the dimension is high. The accuracy for C_F is also good when the dimension is low ($p = 3, 5$). The accuracy for C_N and C_G are not good, but that for C_G improves in the case that the sample size is considerably large compared with the dimension.

In the skew normal population, the accuracy for C_G is good. The accuracy for C_F is also good when the dimension is low ($p = 3$). When the dimension is high, the accuracy for C_G is better than that for C_F.

The hypothesis testing using these criteria is liberal. In all case, the accuracy is

Table 10. Estimated Significance Level for each criterion (Case 6)

	$p=3$				$p=5$				$p=7$			
Normal population												
N_1, N_2	C_N	C_E	C_G	C_F	C_N	C_E	C_G	C_F	C_N	C_E	C_G	C_F
100	.0616	.0693	.0693	.0534	.0669	.0769	.1006	.0613	.0758	.0885	.1696	.0809
200	.0555	.0590	.0614	.0517	.0587	.0631	.0856	.0574	.0645	.0701	.1413	.0737
500	.0527	.0539	.0576	.0513	.0543	.0558	.0773	.0555	.0584	.0602	.1258	.0698
1000	.0517	.0522	.0565	.0512	.0528	.0534	.0744	.0547	.0563	.0571	.1209	.0683
Contaminated Normal population												
N_1, N_2	C_N	C_E	C_G	C_F	C_N	C_E	C_G	C_F	C_N	C_E	C_G	C_F
100	.0821	.0712	.0707	.0532	.0965	.0797	.1041	.0608	.1138	.0921	.1750	.0804
200	.0750	.0606	.0626	.0520	.0859	.0649	.0879	.0579	.0981	.0713	.1434	.0732
500	.0707	.0540	.0579	.0512	.0799	.0566	.0784	.0556	.0895	.0606	.1264	.0692
1000	.0694	.0523	.0565	.0511	.0768	.0532	.0745	.0546	.0867	.0573	.1214	.0685
Skew Normal population												
N_1, N_2	C_N	C_E	C_G	C_F	C_N	C_E	C_G	C_F	C_N	C_E	C_G	C_F
100	.0626	.0696	.0576	.0539	.0679	.0773	.0605	.0609	.0759	.0875	.0676	.0728
200	.0567	.0597	.0515	.0529	.0600	.0637	.0523	.0575	.0647	.0692	.0541	.0662
500	.0540	.0545	.0487	.0526	.0556	.0562	.0500	.0555	.0597	.0605	.0528	.0637
1000	.0534	.0532	.0481	.0528	.0543	.0541	.0493	.0550	.0592	.0589	.0537	.0641

much better when the variance explained for the first principal component is large. The adjustment for the kurtosis parameter and the fourth moments which is included in the asymptotic covariance is reason why the accuracy for C_E or C_G is good under the contaminated normal population or the skew normal population.

References

[1] Ansari, A. R. and Bradley, R. A. (1960). Rank-sum tests for dispersions, *Ann. Math. Stat.*, **31**, 1174-1189.

[2] Boos, D. D. and Brownie, C. (1989). Bootstrap methods for testing homogeneity of variances, *Technometrics*, **31**, 69-82.

[3] Brown, M.B. and Forsythe, A.B. (1974). Robust tests for the equality of variances, *J. Am. Stat. Assoc.*, **69**, 364-367.

[4] Flinger, M.A. and Killeen, T.J. (1976). Distribution-free two-sample tests for scale, *J. Am. Stat. Assoc.*, **71**, 210-213.

[5] Flury, B.K.(1987). Two generalizations of the common principal component model, *Biometrika*, **74**(1), 59-69.

[6] Genton, M. G. (2004). *Skew-elliptical Distributions and Their Applications*, New York: Chapman & Hall.

[7] Gibbons, J. D. and Chakraborti, S. (2003). *Nonparametric Statistical Inference*, 4th ed., Dekker.

[8] Hino, E., Murakami, H. and Sugiyama, T. (2009). Permutation test for equality

of individual an eigenvalue from a covariance matrix in high-dimension, *Commun. Stat., Simul. and Comput.*, **38**, 1675-1689.

[9] Kollo, T. and Neudecker, H.(1995). Asymptotic of eigenvalues and unit-length eigenvectors of sample variance and correlation matrices, *J. Multivar. Anal.*, **47**(2), 283-300.

[10] Levene, H. (1960). Robust tests for equality of variances, in *Contributions to Probability and Statistics*, ed. I. Olkin, Palo Alto, Calif: Stanford University Press, 278-292.

[11] Mood, A. M. (1954). On the asymptotic efficiency of certain nonparametric two sample tests, *Ann. Math. Stat.*, **25**, 514-522.

[12] Murakami, H., Tsukada, S. and Takeda, Y. (2008). A statistic for testing the equality of eigenvalue of covariance matrix on multipopulation, *J. Jpn. Soc. Compt. Stat.*, **21**, No. 1, 21-30.

[13] O'Brien, R.G. (1979). A general ANOVA method for robust tests of additive models for variances, *J. Am. Stat. Assoc.*, **74**, 877-880.

[14] Sugiyama, T. and Ushizawa, K. (1998). A non-parametric method to test equality of intermediate latent roots of two populations in a principal component analysis, *J. Jpn Stat. Soc.*, **28**, No. 2, 227-235.

[15] Takeda, Y. (2001). Permutation test for equality of each characteristic root in two populations, *J. Jpn. Soc. Compt. Stat.*, **14**, 1-10.

[16] Tsai, W. S., Duran, B. S. and Lewis, T. O. (1975). Small-sample behavior of some multisample nonparametric tests for scale. *J. Am. Stat. Assoc.*, **70**, 791-796.

[17] Tsukada, S. (1997). Power comparison of hypothesis testing for an intermediate latent vector of covariance matrix, *J. Jpn. Soc. Compt. Stat.*, **10**, 73-88.

[18] Tsukada, S. (1998). Three statistics for hypothesis testing of intermediate latent vector of covariance matrix and their bootstrap tests, *J. Jpn Stat. Soc.*, **28**, 163-174.

[19] Tsukada, S. (2007). Wald criteria for equality of latent vector of covariance matrix in two populations, *Proc. Ninth Jpn-China Symp. Stat.*, p.307-312.

[20] Tsukada, S. and Sugiyama, T. (1997). On asymptotic expansion of null distributions and power comparison of tests using three statistics for hypothesis testing of intermediate latent vector of covariance matrix, *Bull. Comput. Stat. Jpn.*, **10**, 19-35. (In Japanese).

[21] Tsukada, S. and Murakami, H. (2008). Nonparametric test for latent root of covariance matrix in multi-population, *Proc. Comput. Stat.*, A Springer Company, 565-573 (On CD).

[22] Wludyka, P. and Sa, P. (2004). A robust I-sample analysis of means type randomization test for variances for unbalanced designs, *J. Stat. Comput. Simul.*, **74**, 701-726.