KYBERNETIKA — VOLUME 37 (2001), NUMBER 1, PAGES 61-78

TESTS OF SOME HYPOTHESES ON CHARACTERISTIC ROOTS OF COVARIANCE MATRICES NOT REQUIRING NORMALITY ASSUMPTIONS[1]

FRANTIŠEK RUBLÍK

Test statistics for testing some hypotheses on characteristic roots of covariance matrices are presented, their asymptotic distribution is derived and a confidence interval for the proportional sum of the characteristic roots is constructed. The resulting procedures are robust against violation of the normality assumptions in the sense that they asymptotically possess chosen significance level provided that the population characteristic roots are distinct and the covariance matrices of certain quadratic functions of the random vectors are regular. The null hypotheses considered include hypotheses on proportional sums of characteristic roots, hypotheses on equality of characteristic roots of covariance matrices of the underlying populations or on equality of their sums.

1. INTRODUCTION

Asymptotic properties of distributions of the characteristic roots of the sample covariance matrices have been used for constructing tests of hypotheses on the eigenvalues of the sampled populations provided that the underlying distributions are gaussian. However, if the normality assumption does not hold, then the asymptotic results on distribution of such test statistics derived under the normality assumption may no longer be valid. This follows from the results of [12] where the asymptotic distribution of the characteristic roots was obtained without the normality assumption provided that the population characteristic roots are distinct, and from the results of [1]. This situation, i.e., the case when the underlying population need not be gaussian, is from a general point of view handled in [11], where assertions on the asymptotic distribution of the characteristic roots of random matrices converging to a matrix with multiple characteristic roots are derived. Recent papers dealing with matters related to principal components and characteristic roots focus their attention on testing hypotheses concerning the characteristic vectors. Testing hypotheses on the linear subspace spanned by the eigenvectors corresponding to the d

[1]The research was supported by the grant VEGA 1/7295/20 from the Scientific Grant Agency of the Slovak Republic.

largest characteristic roots of the correlation matrix is in the gaussian setting studied in [10], testing of the multipopulation hypothesis of the equality of the subspaces spanned by the first d principal components of several gaussian populations is the topic of [9]. In this paper a similar multipopulation hypothesis of the equality of the d largest characteristic roots is treated, but without the normality assumption. In a way, this paper continues in the approach set up in [1], [2] and [12], because the crucial assumption of the following text is the distinctness of the eigenvalues. However, while the papers [1], [2] and [12] concentrate their effort on investigating the effect of departures from normality on asymptotic distribution of statistics involving characteristic roots, the aim of this paper is to propose new confidence intervals for the ratio (2.1) and new test statistics for testing some hypotheses on characteristic roots, which yield rules asymptotically robust against violation of the normality assumptions. The main feature of the presented procedures is that they are asymptotically valid provided that the covariance matrices of certain quadratic functions of the random vectors are regular and the characteristic roots of the population covariance matrices are distinct.

Throughout the paper for any symmetric $k \times k$ matrix S the symbol $\lambda_j(S)$ denotes its jth characteristic root, i.e.,

$$\lambda_1(S) \geq \ldots \geq \lambda_k(S), \tag{1.1}$$

and the whole set of the characteristic roots will be denoted by the vector

$$\lambda(S) = (\lambda_1(S), \ldots, \lambda_k(S))'. \tag{1.2}$$

The second section contains assertions concerning the asymptotic distribution of sample characteristic roots and statistical inference on characteristic roots of one statistical population, the topic of the third section is testing of some hypotheses involving characteristic roots of several populations, the proofs are given in the fourth section of the paper.

2. ONE–SAMPLE STATISTICAL INFERENCE ON PROPORTIONAL SUM OF CHARACTERISTIC ROOTS

Let $\lambda_1 \geq \ldots \geq \lambda_k$ denote the characteristic roots of the covariance matrix of a random vector under consideration and $1 \leq d < k$ be a chosen integer. In the principal component analysis the ratio (which is in this paper termed as the proportional sum of characteristic roots)

$$R_d(\lambda) = \frac{\lambda_1 + \ldots + \lambda_d}{\lambda_1 + \ldots + \lambda_k} \tag{2.1}$$

serves as an indicator, whether the first d principal components yield a sufficient amount of information on the random vector. If this is the case, then the experimenter may simplify the process of evaluation by neglecting the remaining $k - d$ principal components and considering only the first d ones. In this section asymptotically valid procedures on $R_d(\lambda)$ are proposed. Theorem 2.2 forms a basis for asymptotically valid tests on the magnitude of the proportional sum, the asymptotically valid confidence intervals for this quantity are presented in Theorem 2.3.

The performance of the confidence intervals is illustrated by the simulation results in Table 1, the section is concluded with Theorem 2.4 and the Conjecture, concerning the validity of the involved assumptions and the existence of the proposed test statistics.

Throughout this section Σ denotes a symmetric $k \times k$ matrix and its spectral decomposition

$$\Sigma = P \operatorname{diag}(\lambda_1, \dots, \lambda_k)\, P',$$

where $\lambda_1 \geq \dots \geq \lambda_k$ and P is an orthogonal matrix with the columns

$$p_j = \begin{pmatrix} p_{1j} \\ \vdots \\ p_{kj} \end{pmatrix}, \qquad j = 1, \dots, k,\ \ k > 1. \tag{2.2}$$

In the assertions of this paper the following assumption will be used.

(A I) The characteristic roots of the matrix Σ are distinct, i.e.,

$$\lambda_1 > \dots > \lambda_k .$$

The asymptotic distribution of sample characteristic roots will be derived by means of the following assertion.

Theorem 2.1. Let $\{S_n\}_{n=1}^{\infty}$ be symmetric $k \times k$ random matrices such that the weak convergence of distributions

$$\mathcal{L}\left[n^{1/2}(S_n - \Sigma)\right] \longrightarrow \mathcal{L}^* \tag{2.3}$$

holds and the matrix Σ fulfills (A I). If $U \subset R^k$ is an open set containing $\lambda(\Sigma) = (\lambda_1, \dots, \lambda_k)'$ and $g : U \to R^z$ is a mapping whose coordinates possess all partial derivatives of the first order and this derivatives are continuous on U, then

$$g\left(\lambda(S_n)\right) - g\left(\lambda(\Sigma)\right) = D \begin{pmatrix} p_1{}'(S_n - \Sigma)p_1 \\ \vdots \\ p_k{}'(S_n - \Sigma)p_k \end{pmatrix} + o_P(n^{-1/2}), \tag{2.4}$$

where

$$D = \begin{pmatrix} \frac{\partial g_1(\lambda)}{\partial \lambda_1}, & \dots & , \frac{\partial g_1(\lambda)}{\partial \lambda_k} \\ \vdots & & \vdots \\ \frac{\partial g_z(\lambda)}{\partial \lambda_1}, & \dots & , \frac{\partial g_z(\lambda)}{\partial \lambda_k} \end{pmatrix}.$$

Especially,

$$\lambda(S_n) = \begin{pmatrix} p_1{}'S_n p_1 \\ \vdots \\ p_k{}'S_n p_k \end{pmatrix} + o_P(n^{-1/2}). \tag{2.5}$$

Throughout the rest of the section suppose that $\{X_m\}_{m=1}^{\infty}$ are independent identically distributed random vectors. The sampled distribution of the random vector X_1 will be subjected to the following conditions.

(A II) X_1 has all fourth order moment finite.

(A III) The assumption (A II) holds and the $k \times k$ symmetric matrix V with the elements

$$V_{rt} = E\left((p_r{}'(X_1 - \mu))^2 (p_t{}'(X_1 - \mu))^2 \right) - \lambda_r \lambda_t\,, \quad r, t = 1, \dots, k \tag{2.6}$$

is regular. In this notation $\mu = E(X_1)$, $\lambda = \lambda(\Sigma)$, Σ is the covariance matrix of X_1 and $p_1, \dots, p_k$ are the vectors (2.2).

Finally, let $\overline{X} = \frac{1}{n}\sum_{m=1}^{n} X_m$ and

$$S_n = \frac{1}{n}\sum_{m=1}^{n}(X_m - \overline{X})(X_m - \overline{X})' \tag{2.7}$$

denote the arithmetic mean and the sample covariance matrix, respectively. In this setting Theorem 2.1 implies validity of the following assertion.

Corollary 2.1. If X_1 fulfills (A II) and its covariance matrix Σ fulfills (A I), then (cf. (2.7), (2.6))

$$\mathcal{L}\left[n^{1/2}(\lambda(S_n) - \lambda(\Sigma)) \right] \longrightarrow N_k(0, V) \tag{2.8}$$

as $n \to \infty$.

We remark that another form of the weak convergence result (2.8) has already been established in Theorem on p. 640 of [12] and in Theorem 5 on p. 290 of [2]. A result, similar to (2.8), can also be found in Theorem 3.1 of [8], which is proved under the assumption that the components of the vector $\Sigma^{-1/2}X$ are independent.

The formula (2.4) can be useful in the case when the asymptotic covariance of $g(\lambda(S_n))$ has to be found. In this paper mainly the relations (2.8) and (2.6) will be employed.

In the rest of this section in accordance with (1.1) and (1.2) the symbols

$$\hat{\lambda}_j = \lambda_j(S_n)\,, \qquad \hat{\lambda} = (\hat{\lambda}_1, \dots, \hat{\lambda}_k)' \tag{2.9}$$

will denote the characteristic roots of the sample covariance matrix (2.7).

Lemma 2.1. Let $\hat{p}_j = \hat{p}_j(X_1, \dots, X_n)$ be random vectors such that

$$S_n \hat{p}_j = \hat{\lambda}_j \hat{p}_j\,, \quad \|\hat{p}_j\| = 1\,, \ \ j = 1, \dots, k\,.$$

Further, let $\hat{V}$ denote the $k \times k$ symmetric random matrix with the elements

$$\hat{V}_{rt} = \frac{1}{n}\sum_{m=1}^{n} \left(\hat{p}_r{}'(X_m - \overline{X})\right)^2 \left(\hat{p}_t{}'(X_m - \overline{X})\right)^2 - \hat{\lambda}_r \hat{\lambda}_t\,. \tag{2.10}$$

(I) If (A II) holds then

$$\hat{V} = V + o_P(1)\,, \tag{2.11}$$

where V is the $k \times k$ symmetric matrix defined by (2.6).

(II) If both (A II) and (A III) hold, then $\hat{V}$ is regular with probability tending to 1 as $n \to \infty$.

The consistency (2.11) will be used in proving the assertions of the following theorems.

Theorem 2.2. Suppose that the random vector X_1 and its covariance matrix Σ fulfill the assumptions (A I), (A II), (A III) and assume that $d < k$ is a positive integer. Fix a number $\gamma \in (0,1)$ and consider the test statistic

$$T_n = T_n(\hat{\lambda}) = \sqrt{n}\frac{a'\hat{\lambda}}{\sqrt{a'\hat{V}a}}\,, \tag{2.12}$$

where $a = (1-\gamma,\ldots,1-\gamma,-\gamma,\ldots,-\gamma)'$ is the vector from R^k, which has on the first d positions the number $1-\gamma$, on the remaining $k-d$ positions $-\gamma$, and $\hat{V}$ is the $k \times k$ symmetric matrix defined by (2.10).

(I) The statistic T_n is well-defined with probability tending to 1 as $n \to \infty$.

(II) For the proportional sum (2.1) and for every real number t

$$\lim_{n\to\infty} P(T_n > t) = \begin{cases} 1 & R_d(\lambda) > \gamma\,, \\ 1-\Phi(t) & R_d(\lambda) = \gamma\,, \\ 0 & R_d(\lambda) < \gamma\,, \end{cases} \tag{2.13}$$

and

$$\lim_{n\to\infty} P(T_n < t) = \begin{cases} 0 & R_d(\lambda) > \gamma\,, \\ \Phi(t) & R_d(\lambda) = \gamma\,, \\ 1 & R_d(\lambda) < \gamma\,, \end{cases} \tag{2.14}$$

where Φ denotes the distribution function of the standard normal $N(0,1)$ distribution.

The purpose of (2.13) is to provide a basis for construction of the test of the hypothesis

$$H_0: \quad R_d(\lambda) \le \gamma.$$

Indeed, if the critical region consists of the sample points for which $T_n > u_{1-\alpha}$, where T_n is the statistic (2.12) and $u_{1-\alpha}$ denotes the $(1-\alpha)$th quantile of the $N(0,1)$ distribution, then in the setting used in the previous theorem owing to (2.13) this test is a consistent test of the hypothesis $R_d(\lambda) \le \gamma$ at the asymptotic significance level α. Construction of such a test of the hypothesis $R_d(\lambda) \ge \gamma$ by means of (2.14) can be obviously carried out in an analogous way.

The confidence intervals constructed in the next theorem will be defined by means of the statistic

$$\hat{\sigma}^2 = \hat{\delta}'\hat{V}\hat{\delta}\,, \tag{2.15}$$

where $\hat{\delta} = (\hat{\delta}_1, \ldots, \hat{\delta}_k)'$ is the random vector with the coordinates

$$\hat{\delta}_j = \begin{cases} \dfrac{\hat{\lambda}_{d+1} + \ldots + \hat{\lambda}_k}{(\hat{\lambda}_1 + \ldots + \hat{\lambda}_k)^2} & j = 1, \ldots, d, \\ -\dfrac{(\hat{\lambda}_1 + \ldots + \hat{\lambda}_d)}{(\hat{\lambda}_1 + \ldots + \hat{\lambda}_k)^2} & j = d+1, \ldots, k, \end{cases} \tag{2.16}$$

and $\hat{V}$ is the $k \times k$ symmetric matrix defined by (2.10).

Theorem 2.3. If the random vector X_1 and its covariance matrix Σ fulfill the assumptions (A I), (A II), (A III) and $d < k$ is a positive integer, then (cf. (2.15), (2.1) and (2.9))

$$\mathcal{L}\left[n^{1/2}\, \frac{R_d(\hat{\lambda}) - R_d(\lambda)}{\hat{\sigma}} \right] \longrightarrow N(0,1) \tag{2.17}$$

as $n \to \infty$. Hence if Φ denotes distribution function of the $N(0,1)$ distribution and $\Phi(u_\beta) = \beta$, then in this setting

$$I_1 = \left\langle R_d(\hat{\lambda}) - \hat{\sigma}\, \frac{u_{1-\alpha/2}}{\sqrt{n}},\, R_d(\hat{\lambda}) + \hat{\sigma}\, \frac{u_{1-\alpha/2}}{\sqrt{n}} \right\rangle, \tag{2.18}$$

$$I_2 = \left\langle R_d(\hat{\lambda}) - \hat{\sigma}\, \frac{u_{1-\alpha}}{\sqrt{n}},\, +\infty \right), \qquad I_3 = \left(-\infty\,,\, R_d(\hat{\lambda}) + \hat{\sigma}\, \frac{u_{1-\alpha}}{\sqrt{n}} \right\rangle,$$

are confidence intervals for the proportional sum (2.1) with asymptotic confidence coefficient $1-\alpha$, i.e., for $j = 1,2,3$

$$\lim_{n\to\infty} P(\, R_d(\lambda) \in I_j\,) = 1 - \alpha\,.$$

The intervals (2.18) are competitors of the classical confidence intervals for $R_d(\lambda)$, which are based on normality assumptions. If the sampling is made from the normal distribution with distinct eigenvalues, then according to pp. 233–234 of [4] the interval

$$J_1 = \left\langle R_d(\hat{\lambda}) - \hat{\tau}\frac{u_{1-\alpha/2}}{\sqrt{n-1}}\,, R_d(\hat{\lambda}) + \hat{\tau}\frac{u_{1-\alpha/2}}{\sqrt{n-1}} \right\rangle, \tag{2.19}$$

where

$$\hat{\tau}^2 = \frac{2\,\mathrm{tr}(S_n^{*\,2})}{(\mathrm{tr}(S_n^*))^2}(R_d(\hat{\lambda})^2 - 2R_d(\hat{\lambda})\hat{a} + \hat{a})\,,$$

$$S_n^* = \frac{1}{n-1}\sum_{m=1}^{n}(X_m - \overline{X})(X_m - \overline{X})'\,, \quad \hat{a} = \frac{\hat{\lambda}_1^2 + \ldots + \hat{\lambda}_d^2}{\hat{\lambda}_1^2 + \ldots + \hat{\lambda}_k^2}\,,$$

is a confidence interval for $R_d(\lambda)$ with the asymptotic confidence coefficient $1-\alpha$ (here $\hat{\lambda}_1 \geq \ldots \geq \hat{\lambda}_k$ are the characteristic roots of S_n^*).

The effect of the asymptotics in the case of (2.18) and (2.19) is illustrated by the following table. The presented simulation estimates of the characteristics of the confidence intervals are obtained from $N = 5000$ trials for each considered distribution and sample size. The sampled distribution is 8-dimensional with zero mean and the covariance matrix

$$\Sigma = \mathrm{diag}(\ 9,\ 16,\ 25,\ 36,\ 64,\ 81,\ 100,\ 169\ )\,, \tag{2.20}$$

and in the confidence intervals the values

$$d = 3, \qquad R_d(\lambda) = 0.7, \qquad \alpha = 0.05$$

are considered. In accordance with (2.20), in the following table the symbol D_1 denotes the $N_8(0, \Sigma)$ distribution and D_2 stands for the distribution with the covariance matrix (2.20), where all coordinates are independent and exponentially distributed. All the simulations considered in this paper were carried out by means of MATLAB, version 4.2c.1.

Table 1. Limits of the intervals (2.18) and (2.19) and their probabilities of covering $R_d(\lambda)$.

		n	15	30	50	100	150	300	500
D_1	I_1	average lower limit	0.725	0.688	0.678	0.675	0.675	0.679	0.683
		average upper limit	0.896	0.833	0.798	0.764	0.750	0.733	0.725
		probability of covering R_d	0.31	0.61	0.76	0.86	0.89	0.92	0.94
D_1	J_1	average lower limit	0.714	0.683	0.675	0.674	0.675	0.679	0.683
		average upper limit	0.907	0.837	0.800	0.765	0.750	0.733	0.725
		probability of covering R_d	0.39	0.66	0.79	0.87	0.90	0.93	0.94

		n	15	30	50	100	150	300	500
D_2	I_1	average lower limit	0.728	0.676	0.661	0.653	0.654	0.660	0.667
		average upper limit	0.931	0.875	0.839	0.799	0.781	0.757	0.744
		probability of covering R_d	0.34	0.65	0.77	0.88	0.91	0.94	0.93
D_2	J_1	average lower limit	0.740	0.702	0.690	0.681	0.680	0.682	0.685
		average upper limit	0.919	0.849	0.810	0.771	0.755	0.735	0.726
		probability of covering R_d	0.29	0.49	0.58	0.66	0.68	0.71	0.69

Since the distribution of $R_d(\hat{\lambda})$ depends both on the dimensionality, type and on characteristic roots of the sampled distribution, it is difficult to characterize the precision of these confidence intervals in general. Nevertheless, the values given in Table 1 suggest that under the normality assumption the probability of covering $R_d(\lambda)$ is for the interval (2.18) similar as for (2.19), and if the shape of the sampled distribution is strikingly different from the gaussian case, then (2.18) may yield probability of covering $R_d(\lambda)$ better than the interval (2.19).

The matrix $\hat{V}$, defined by means of (2.10), may depend on the choice of characteristic vectors and when this choice would be made absolutely arbitrarily, then the matrix mapping $\hat{V}$ could be not measurable. For this reason the measurability of $\hat{p}_j$ is postulated in the previous theorems. Since the set of k-dimensional orthogonal matrices is compact and the characteristic roots depend on the symmetric $k \times k$

matrix in a continuous way, such a measurable choice of the characteristic vectors is possible (and can be proved, e. g., by means of the Lemma 2.1, p. 54 of [7]). This assumption of measurability should not be perceived as some kind of restriction on practical computation of the matrix $\hat{V}$. Indeed, if the characteristic roots of S_n are distinct, then the characteristic vectors corresponding to the rth characteristic root form a one-dimensional linear space, the only possible choice of the rth characteristic vector is either $\hat{p}_r$ or the $-\hat{p}_r$ and the value of (2.10) remains unchanged regardless of this choice. Thus under the validity of (A I) – (A III), imposed in the previous theorems, the values of presented statistics are not influenced by the choice of the possible values of the characteristic vectors with probability tending to 1 as the sample size tends to infinity.

The asymptotic assertions of Theorems 2.2 and 2.3 are derived under the assumption that the matrix V is regular. The following theorem shows that this is true in a wide range of cases.

Theorem 2.4. Suppose that X_1 possesses a density with respect to the Lebesgue measure on $(R^k, \mathcal{B}^k)$. If X_1 fulfills (A II), then the matrix V with the elements (2.6) is well-defined and is regular.

The regularity of the matrix with the elements (2.10) plays an essential role in the construction of test statistics, proposed in this paper. The usual property of the almost sure existence of the concerned test statistics for the sample sizes exceeding k will hold, if the following statement is true.

Conjecture. Suppose that $X_1, \dots, X_n$ are independent identically distributed random vectors and X_1 possesses a density with respect to the Lebesgue measure on $(R^k, \mathcal{B}^k)$. If $n > k$, then the matrix $\hat{V}$ with the elements (2.10) is positive definite with probability 1.

Even though simulations support validity of the previous statement, no exact proof of this assertion is available.

3. MULTISAMPLE TESTS OF SOME HOMOGENEITY HYPOTHESES ON CHARACTERISTIC ROOTS

The variability of a random vector is usually expressed by its covariance matrix and equal variability (stochastic stability) of several populations is verified by testing equality of their covariance matrices. Since one of the goals of the principal component analysis is reducing the dimensionality of the examined data, from this point of view the equal variability of q statistical populations can be perceived also as the equality of their d largest characteristic roots, or as the equality of their sums, or as the equality of the proportional sums (2.1). The test statistics for these hypotheses are proposed in Theorem 3.1, 3.2 and 3.3, respectively, where also their asymptotic distribution is derived and F_j denotes the distribution function of the chi-square distribution with j degrees of freedom. The asymptotic behaviour of the proposed test for testing the equality of the d largest characteristic roots is illustrated by the simulation results, presented in Table 2.

Let $q > 1$ be a fixed integer and for $i = 1, \dots, q$

$$X_1^{(i)}, \dots, X_{n_i}^{(i)}$$

denote random sample of size n_i from a distribution with the $k \times k$ covariance matrix Σ_i. Assume throughout this section that the random samples $(X_1^{(1)}, \dots, X_{n_1}^{(1)}), \dots$ $\dots, (X_1^{(q)}, \dots, X_{n_q}^{(q)})$ are independent, and for every $i = 1, \dots, q$ the random vector $X_1^{(i)}$ and its covariance matrix fulfill the assumptions (A I), (A II) and (A III), defined in the previous section. Thus for $i = 1, \dots, q$ the covariance matrix Σ_i has distinct characteristic roots

$$\lambda_1^{(i)} > \dots > \lambda_k^{(i)} ,$$

the vector $X_1^{(i)}$ has all fourth order moments finite and the $k \times k$ matrix $V^{(i)}$ with the elements

$$V_{rt}^{(i)} = E\left((p_r^{(i)\,\prime}(X_1^{(i)} - \mu_i))^2 (p_t^{(i)\,\prime}(X_1^{(i)} - \mu_i))^2 \right) - \lambda_r^{(i)} \lambda_t^{(i)} \tag{3.1}$$

is regular. Here $\mu_i = E(X_1^{(i)})$ and $p_r^{(i)}$ is the rth column of the orthogonal matrix P_i, fulfilling the equality $\Sigma_i = P_i \operatorname{diag}(\lambda_1^{(i)}, \dots, \lambda_k^{(i)})\, P_i{}'$.

The sample characteristics used in this section are those from the previous section, except for the notation identifying the sampled distribution. The arithmetic mean $\overline{X}_i$ and the sample covariance matrix $S_{n_i}^{(i)}$ of the ith sample are defined by means of the formulas

$$\overline{X}_i = \frac{1}{n_i} \sum_{m=1}^{n_i} X_m^{(i)} , \qquad S_{n_i}^{(i)} = \frac{1}{n_i} \sum_{m=1}^{n_i} (X_m^{(i)} - \overline{X}_i)(X_m^{(i)} - \overline{X}_i)' ,$$

and the characteristic roots of the matrix $S_{n_i}^{(i)}$ will be denoted by the symbols

$$\hat{\lambda}^{(i)} = (\hat{\lambda}_1^{(i)}, \dots, \hat{\lambda}_k^{(i)})' , \qquad \hat{\lambda}_1^{(i)} \geq \dots \geq \hat{\lambda}_k^{(i)} .$$

To construct test statistics for testing hypotheses mentioned in the introduction of this section, suppose that $\hat{p}_j^{(i)} = \hat{p}_j^{(i)}(X_1^{(i)}, \dots, X_{n_i}^{(i)})$ are random vectors such that

$$S_{n_i}^{(i)} \hat{p}_j^{(i)} = \hat{\lambda}_j^{(i)} \hat{p}_j^{(i)} , \qquad \|\hat{p}_j^{(i)}\| = 1 , \qquad i = 1, \dots, q , \; j = 1, \dots, k ,$$

d is a fixed integer such that $1 \leq d \leq k$ and $\hat{V}_d^{(i)}$ is the $d \times d$ symmetric matrix consisting of the elements

$$\hat{V}_{rt}^{(i)} = \frac{1}{n_i} \sum_{m=1}^{n_i} (\hat{p}_r^{(i)\,\prime}(X_m^{(i)} - \overline{X}_i))^2 (\hat{p}_t^{(i)\,\prime}(X_m^{(i)} - \overline{X}_i))^2 - \hat{\lambda}_r^{(i)} \hat{\lambda}_t^{(i)} , \qquad r, t = 1, \dots, d .$$

First we pay attention to the hypothesis

$$\lambda_j^{(1)} = \dots = \lambda_j^{(q)} , \; 1 \leq j \leq d , \tag{3.2}$$

that the d largest characteristic roots of the sampled populations are equal.

Theorem 3.1. In the case that the matrices $\hat{V}_d^{(1)}, \ldots, \hat{V}_d^{(q)}$ are regular, put

$$\tilde{W} = \sum_{i=1}^{q} n_i (\hat{V}_d^{(i)})^{-1}, \quad \tilde{\lambda} = \tilde{W}^{-1} \sum_{i=1}^{q} n_i (\hat{V}_d^{(i)})^{-1} \lambda_i^*, \quad \lambda_i^* = (\hat{\lambda}_1^{(i)}, \ldots, \hat{\lambda}_d^{(i)})'$$

and

$$T_{n_1,\ldots,n_q} = \sum_{i=1}^{q} n_i (\lambda_i^* - \tilde{\lambda})' (\hat{V}_d^{(i)})^{-1} (\lambda_i^* - \tilde{\lambda}). \tag{3.3}$$

The statistics (3.3) are well-defined with probability tending to 1 as $n_1 \to \infty, \ldots$ $\ldots, n_q \to \infty$, and for every $t > 0$

$$\lim_{n_1 \to \infty, \ldots, n_q \to \infty} P(T_{n_1,\ldots,n_q} > t) = \begin{cases} 1 - F_{(q-1)d}(t) & \text{if (3.2) holds,} \\ 1 & \text{otherwise.} \end{cases} \tag{3.4}$$

The next theorem is aimed at constructing a test of the hypothesis

$$\sum_{j=1}^{d} \lambda_j^{(1)} = \ldots = \sum_{j=1}^{d} \lambda_j^{(q)}, \tag{3.5}$$

that the sums of the d largest characteristic roots of the sampled populations are equal.

Theorem 3.2. Put

$$\hat{\sigma}_i^2 = \sum_{r=1}^{d} \sum_{t=1}^{d} \hat{V}_{rt}^{(i)}, \quad \tilde{\sigma}^2 = \sum_{i=1}^{q} \frac{n_i}{\hat{\sigma}_i^2}.$$

$$\hat{\kappa}_i = \sum_{j=1}^{d} \hat{\lambda}_j^{(i)}, \quad \tilde{\kappa} = \frac{1}{\tilde{\sigma}^2} \sum_{i=1}^{q} \frac{n_i}{\hat{\sigma}_i^2} \hat{\kappa}_i$$

and

$$T_{n_1,\ldots,n_q} = \sum_{i=1}^{q} n_i \frac{(\hat{\kappa}_i - \tilde{\kappa})^2}{\hat{\sigma}_i^2}. \tag{3.6}$$

The statistics (3.6) are well-defined with probability tending to 1 as $n_1 \to \infty, \ldots$ $\ldots, n_q \to \infty$, and for every $t > 0$

$$\lim_{n_1 \to \infty, \ldots, n_q \to \infty} P(T_{n_1,\ldots,n_q} > t) = \begin{cases} 1 - F_{q-1}(t) & \text{if (3.5) holds,} \\ & \text{otherwise.} \end{cases} \tag{3.7}$$

In accordance with (2.1) the proportional sum of characteristic roots of the ith population is defined by the formula

$$R_d^{(i)} = \frac{\sum_{z=1}^{d} \lambda_z^{(i)}}{\sum_{w=1}^{k} \lambda_w^{(i)}}. \tag{3.8}$$

Thus the equality

$$R_d^{(1)} = \ldots = R_d^{(q)} \tag{3.9}$$

means that the proportional sums of the characteristic roots of the sampled populations are equal; the last theorem of this section is aimed at testing this hypothesis. Here the sample counterparts

$$\hat{R}_d^{(i)} = \frac{\sum_{z=1}^{d} \hat{\lambda}_z^{(i)}}{\sum_{w=1}^{k} \hat{\lambda}_w^{(i)}} \tag{3.10}$$

of (3.8) will be used. The estimator of the asymptotic variance appearing in the weak-convergence result (2.17), when the sampling is made from the ith population, is now

$$\hat{\sigma}_i^2 = \hat{\delta}_i{}' \hat{V}_k^{(i)} \hat{\delta}_i \,, \tag{3.11}$$

where $\hat{\delta}_i = (\,\hat{\delta}_i(1), \ldots, \hat{\delta}_i(k)\,)'$ and

$$\hat{\delta}_i(j) = \begin{cases} \dfrac{\sum_{z=d+1}^{k} \hat{\lambda}_z^{(i)}}{\left(\sum_{w=1}^{k} \hat{\lambda}_w^{(i)}\right)^2} & j = 1, \ldots, d, \\[2ex] -\dfrac{\left(\sum_{z=1}^{d} \hat{\lambda}_z^{(i)}\right)}{\left(\sum_{w=1}^{k} \hat{\lambda}_w^{(i)}\right)^2} & j = d+1, \ldots, k, \end{cases} \tag{3.12}$$

i. e., the index i serves for identifying the sampled population.

Theorem 3.3. Suppose that the positive integer $d < k$ and define the test statistic by the formula (cf. (3.10), (3.11))

$$T_{n_1,\ldots,n_q} = \sum_{i=1}^{q} n_i \frac{(\hat{R}_d^{(i)} - \tilde{R}_d)^2}{\hat{\sigma}_i^2} \,, \tag{3.13}$$

where

$$\tilde{R}_d = \frac{1}{\tilde{\sigma}^2} \sum_{i=1}^{q} \frac{n_i}{\hat{\sigma}_i^2} \hat{R}_d^{(i)} \,, \qquad \tilde{\sigma}^2 = \sum_{i=1}^{q} \frac{n_i}{\hat{\sigma}_i^2} \,.$$

The statistics (3.13) are well-defined with probability tending to 1 as $n_1 \to \infty, \ldots, n_q \to \infty$, and for every $t > 0$

$$\lim_{n_1 \to \infty, \ldots, n_q \to \infty} P(\,T_{n_1,\ldots,n_q} > t\,) = \begin{cases} 1 - F_{q-1}(t) & \text{if (3.9) holds,} \\ 1 & \text{otherwise.} \end{cases}$$

Similarly as in the case of (2.13) and (2.14), the purpose of the previous theorems is to provide a basis for construction of asymptotic tests of the hypotheses (3.2), (3.5) or (3.9). The test rule rejects the underlying null hypothesis if the concerned test statistic $T_{n_1,\dots,n_q}$ exceeds the $(1-\alpha)$th quantile of the chi-square distribution having the degrees of freedom described in the particular theorem.

The possible problems with the measurability and the uniqueness of the presented test statistics are similar to those in the one-sample case and their aspects have been discussed in the previous section.

To illustrate the effect of the asymptotics in the case of testing the hypothesis that the underlying distributions do not differ in their d largest characteristic roots, consider the sampling from the distributions $N_8(0,\Sigma_1)$, $N_8(0,\Sigma_2)$ and $N_8(0,\Sigma_3)$, where

$$\begin{aligned}
\Sigma_1 &= \mathrm{diag}(\ 9,\ 16,\ 25,\ 36,\ 64,\ 81,\ 100,\ 169\ ),\\
\Sigma_2 &= \mathbf{P}\,\mathrm{diag}(\ 1,\ 4,\ 2.6^2,\ 3.7^2,\ 49,\ 81,\ 100,\ 169\ )\,\mathbf{P}',\\
\Sigma_3 &= \mathbf{P}\,\mathrm{diag}(\ 1,\ 4,\ 2.6^2,\ 3.7^2,\ 49,\ 100,\ 144,\ 256\ )\,\mathbf{P}',
\end{aligned}$$

and

$$\mathbf{P} = \begin{pmatrix}
0.5 & 0.5 & 0.5 & 0.5 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & -0.5 & 0.5 & 0.5 & -0.5\\
-0.5 & -0.5 & 0.5 & 0.5 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & 0.5 & 0.5 & 0.5 & 0.5\\
-0.5 & 0.5 & -0.5 & 0.5 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & -0.5 & -0.5 & 0.5 & 0.5\\
-0.5 & 0.5 & 0.5 & -0.5 & 0 & 0 & 0 & 0\\
0 & 0 & 0 & 0 & -0.5 & 0.5 & -0.5 & 0.5
\end{pmatrix}$$

is an orthogonal matrix. In the following table n_j denotes the size of sample from the $N_8(0,\Sigma_j)$ distribution, T_{n_i,n_j} is the statistic (3.3) for testing the null hypothesis that the sampled populations do not differ in the largest $d = 3$ characteristic roots (i.e., the equality $\lambda_s^{(i)} = \lambda_s^{(j)}$ holds for $s = 1,2,3$), and $P(T_{n_i,n_j} > 7.8147)$ denotes the simulation estimate, based on $N = 5000$ trials, of the probability of rejection this null hypothesis, corresponding to the asymptotic significance level $\alpha = 0.05$ (7.8147 is the 0.95th quantile of the chi-square distribution with $(q-1)d = 1.3 = 3$ degrees freedom).

Table 2. Probability $P(T_{n_i,n_j} > 7.8147)$.

	n_2				
n_1	30	50	100	200	500
30	0.02	0.01	0.02	0.03	0.05
50	0.03	0.02	0.02	0.02	0.03
100	0.06	0.04	0.02	0.02	0.03
200	0.10	0.06	0.04	0.03	0.03
500	0.13	0.09	0.06	0.05	0.04
	n_3				
n_1	30	50	100	200	500
30	0.11	0.19	0.37	0.49	0.58
50	0.13	0.24	0.44	0.63	0.74
100	0.17	0.31	0.59	0.82	0.93
200	0.20	0.37	0.70	0.93	0.99
500	0.23	0.41	0.79	0.99	1.00

When the sampling is made from the first and the second normal distribution, then the null hypothesis is fulfilled, and when the samples are drawn from the first and the third normal distribution, then the null hypothesis does not hold. Thus for these particular distributions the upper part of the table illustrates the magnitude of the probability of the error of the type I and the lower part the power of the test. Similarly as in the cases considered in Table 1, it is difficult to draw generally valid conclusions on the distribution of the test statistic (3.3) in the finite sample cases, because this distribution depends on the dimensionality, type and on characteristic roots of the sampled distribution. Nevertheless, the results from Table 2 suggest that the more balanced sampling schemes are preferable to the less ones, because in the unbalanced cases the increase of the sample size may not result in a corresponding increase of the power (as it is indeed the case, when in the previous table $n_3 = 30$ is fixed and n_1 is allowed to increase till 500).

4. PROOFS

Proof of Theorem 2.1. For the purpose of this proof assume that $e(\Sigma^*) = (\Sigma^*(11), \Sigma^*(12), \ldots, \Sigma^*(1k), \Sigma^*(22), \ldots, \Sigma^*(2k), \ldots, \Sigma^*(kk))' \in R^{k(k+1)/2}$ denotes elements of the symmetric $k \times k$ matrix Σ^* and $\lambda(e(\Sigma^*))$ are its characteristic roots arranged in decreasing order. According to Note 2 on p. 160 of [3] there exist a neighbourhood $N \subset R^{k(k+1)/2}$ of $e(\Sigma)$ and a mapping $p = (p_1, \ldots, p_k)$ defined on N and taking values in the set of the orthogonal $k \times k$ matrices such that all coordinates of λ and p possess on N continuous partial derivatives and if a symmetric matrix Σ^* is such that $e(\Sigma^*) \in N$, then for the ith characteristic root $\lambda_i(e(\Sigma^*))$ of Σ^*

$$\frac{\partial \lambda_i(e(\Sigma^*))}{\partial \Sigma^*(r,w)} = \begin{cases} p_{ri}^{*2} & r = w\,, \\ 2p_{ri}^* p_{wi}^* & r < w\,, \end{cases}$$

where $P^* = (p_{st}^*)$ denotes the orthogonal matrix of the eigenvectors $p(e(\Sigma^*)) = (p_1^*, \ldots, p_k^*)$ of Σ^*. These partial derivatives are continuous on the set N and therefore under the assumptions of this theorem the use of the Taylor theorem yields the equality

$$\begin{aligned} \lambda_i(S_n) - \lambda_i(\Sigma) &= \sum_{r=1}^{k} \sum_{w=r}^{k} \frac{\partial \lambda_i(e(\Sigma^*))}{\partial \Sigma^*(r,w)} (S_n(r,w) - \Sigma(r,w)) = p_i^{*\prime}(S_n - \Sigma)p_i^* \\ &= p_i{}'(S_n - \Sigma)p_i + o_P(n^{-1/2})\,, \end{aligned} \tag{4.1}$$

and (2.5) is proved.

To prove (2.4) assume that D is a neighbourhood of $\lambda(\Sigma)$ such that $D \subset U$. If $i \in \{1, \ldots, z\}$ is fixed, g_i denotes the ith component of g and $\lambda(S_n) \in D$, then by means of the Taylor formula

$$g_i(\lambda(S_n)) - g_i(\lambda(\Sigma)) = \sum_{j=1}^{k} \frac{\partial g_i(\lambda^*)}{\partial \lambda_j^*} (\lambda_j(S_n) - \lambda_j(\Sigma))\,,$$

where the vector λ^* belongs to the segment with the endpoints $\lambda(S_n)$, $\lambda(\Sigma)$. Thus

$$g_i(\lambda(S_n)) - g_i(\lambda(\Sigma)) = \sum_{j=1}^{k} \frac{\partial g_i(\lambda(\Sigma))}{\partial \lambda_j(\Sigma)}(\lambda_j(S_n) - \lambda_j(\Sigma)) + o_P(n^{-1/2}), \qquad (4.2)$$

because the derivatives of g_i are continuous and (2.3), (4.1) yield that $\lambda_j(S_n) - \lambda_j(\Sigma) = O_P(n^{-1/2})$. The relations (4.2) and (4.1) together with the fact, that $\lambda(S_n) \in D$ with probability tending to 1 as $n \to \infty$, imply (2.4). □

Proof of Corollary 2.1. Put

$$\eta_m = X_m - \mu, \quad \overline{\eta} = \frac{1}{n}\sum_{j=1}^{n} \eta_j, \quad M_n = \frac{1}{n}\sum_{m=1}^{n} \eta_m \eta_m{}'.$$

It follows from the central limit theorem that $\overline{\eta} = O_P(n^{-1/2})$ and therefore

$$S_n = \frac{1}{n}\sum_{m=1}^{n} \eta_m \eta_m{}' - \overline{\eta}\,\overline{\eta}' = M_n + o_P(n^{-1/2}). \qquad (4.3)$$

Since the fourth order moments of η_1 are finite, the components of the random matrix $\eta_1\eta_1{}'$ have finite variances, which together with $E(\eta_1\eta_1{}') = \Sigma$ and the central limit theorem means, that the weak convergence $\mathcal{L}(\sqrt{n}(M_n - \Sigma)) \longrightarrow \mathcal{L}^*$ holds, hence (4.3) yields validity of the condition (2.3). This together with (A I) means that the conditions of Theorem 2.1 are fulfilled and therefore (2.5) holds. An application of (2.5) and (4.3) leads to the equality

$$\sqrt{n}\Big[\lambda(S_n) - \lambda(\Sigma)\Big] = \sqrt{n}\begin{bmatrix} p_1{}'(M_n - \Sigma)\,p_1 \\ \vdots \\ p_k{}'(M_n - \Sigma)\,p_k \end{bmatrix} + o_P(1),$$

which together with the central limit theorem implies (2.8). □

Proof of Lemma 2.1. Without the loss of generality assume that $\mu = 0$. This by the law of large numbers means that $\overline{X} = o_P(1)$. But

$$\hat{p}_r'(X_m - \overline{X})(X_m - \overline{X})'\hat{p}_r = \hat{p}_r'X_m X_m{}'\hat{p}_r - 2\hat{p}_r'X_m\overline{X}{}'\hat{p}_r + \hat{p}_r'\overline{X}\,\overline{X}{}'\hat{p}_r$$

and therefore

$$(\hat{p}_r'(X_m - \overline{X}))^2(\hat{p}_t'(X_m - \overline{X}))^2 = \hat{p}_r'X_m X_m{}'\hat{p}_r\hat{p}_t'X_m X_m{}'\hat{p}_t + \Delta_m, \qquad (4.4)$$

where by means of the Cauchy-Schwarz inequality after some calculation

$$|\Delta_m| \le 4\|X_m\|^3\|\overline{X}\| + 6\|X_m\|^2\|\overline{X}\|^2 + 4\|X_m\|\,\|\overline{X}\|^3 + \|\overline{X}\|^4. \qquad (4.5)$$

Thus by the relation $\overline{X} = o_P(1)$ and the law of large numbers

$$\frac{1}{n}\sum_{m=1}^{n}(\hat{p}_r'(X_m - \overline{X}))^2(\hat{p}_t'(X_m - \overline{X}))^2 = \frac{1}{n}\sum_{m=1}^{n}(\hat{p}_r'X_m)^2(\hat{p}_t'X_m)^2 + o_P(1). \qquad (4.6)$$

Since the characteristic roots depend on the symmetric matrix in a continuous way, with probability tending to 1 as $n \to \infty$ the inequalities $\hat{\lambda}_1 > \ldots > \hat{\lambda}_k > 0$ hold, the value of $(\hat{p}_r' X_m)^2$ does not depend on the choice of the eigenvector and in accordance with Theorem 7 of [3], p.158, we may assume that the differences $\alpha_r = \hat{p}_r - p_r$ tend to zero in probability as $n \to \infty$. Hence similarly as in (4.4) – (4.6)

$$\frac{1}{n}\sum_{m=1}^{n}(\hat{p}_r' X_m)^2(\hat{p}_t' X_m)^2 = \frac{1}{n}\sum_{m=1}^{n}(p_r' X_m)^2(p_t' X_m)^2 + o_P(1)$$

which together with (4.6), the law of large numbers and the continuity of characteristic roots yields (2.11); the validity of (II) can be easily proved by means of (2.11) and the continuity of determinant. □

P r o o f o f T h e o r e m 2.2. Since the vector $a \neq 0$, the statistic (2.12) is well-defined on the set A_n of the sample points, for which the matrix $\hat{V}$ is regular. This together with the Lemma 2.1 (II) means that the assertion (I) holds.

Since

$$a'\lambda = \left(\sum_{i=1}^{k}\lambda_i\right)(R_d(\lambda) - \gamma)\,,$$

the rest of the proof follows from (2.8) and (2.11). □

P r o o f o f T h e o r e m 2.3. Obviously

$$\frac{\partial R_d(\lambda)}{\partial \lambda_j} = \delta_j\,,$$

where δ is the vector defined by (2.16) and $\hat{\lambda}$ is replaced with λ. Thus by the delta method and (2.8)

$$\mathcal{L}[\sqrt{n}(\,R_d(\hat{\lambda}) - R_d(\lambda)\,)] \longrightarrow N_1(0,\sigma^2)\,,$$

where $\sigma^2 = \delta' V \delta$. Hence (2.17) holds, because (2.11) and the consistency of $\hat{\lambda}$ imply that $\hat{\sigma} = \sigma + o_P(1)$. The rest of the proof is obvious. □

P r o o f o f T h e o r e m 2.4. Put

$$Y = (\,(p_1'(X_1 - \mu))^2, \ldots, (p_k'(X_1 - \mu))^2\,)' \tag{4.7}$$

and suppose that $a \in R^k$ is a non-zero vector. Obviously

$$a' V a = E[g(Y)]\,,\ g(Y) = (a'Y - d)^2\,,\ d = a'E(Y)\,.$$

Since $g(Y)$ is a polynomial in coordinates of the random sample $X_1, \ldots, X_n$, according to the Lemma of [5] it is sufficient to show that $g(Y) > 0$ for a suitable chosen value of X_1.

Since a is a non-zero vector, there exists an index j such that $a_j \neq 0$. For

$$X_1 = Kp_j + \mu\,, \qquad K = (|d| + 1)/\sqrt{|a_j|}$$

one obtains that

$$|a'Y| = |K^2 a_j| > |d|\,.$$

Hence for this value of X_1 the inequality $g(Y) > 0$ holds. □

Proof of Theorem 3.1. According to the assumptions the matrices $V^{(i)}$, $i = 1, \dots, q$, are positive definite. Hence also the matrices $V_d^{(i)}$, consisting for $r, t = 1, \dots d$ of the elements (3.1), are positive definite and therefore regular. This together with the Lemma 2.1 means, that the matrices $\hat{V}_d^{(i)}$, $i = 1, \dots, q$ are regular with probability tending to 1, as $n_1 \to \infty, \dots, n_q \to \infty$. Thus it remains to prove (3.4).

Let $n = n_1 + \dots + n_q$, $\hat{a}_i = n_i/n$, denote the total and the relative sample size, respectively. Since every bounded sequence of real numbers contains a convergent subsequence, we may assume without the loss of generality that

$$\hat{a}_j \longrightarrow a_j\,, \quad j = 1, \dots, q\,. \tag{4.8}$$

To prove the first line in (3.4) suppose that $\lambda_j^{(1)} = \dots = \lambda_j^{(q)}$ for $j = 1, \dots, d$ and put $\xi_i = n_i^{1/2}(\lambda_i^* - \lambda)$, where λ denotes the joint value of characteristic roots $(\lambda_1^{(i)}, \dots, \lambda_d^{(i)})'$. Since

$$\sum_{i=1}^{q} n_i(\lambda_i^* - \lambda)'(\hat{V}_d^{(i)})^{-1} = (\tilde{\lambda} - \lambda)'\tilde{W}\,,$$

the equality

$$\begin{aligned} T_{n_1,\dots,n_q} &= \sum_{i=1}^{q} n_i\Big((\lambda_i^* - \lambda) + (\lambda - \tilde{\lambda})\Big)'(\hat{V}_d^{(i)})^{-1}\Big((\lambda_i^* - \lambda) + (\lambda - \tilde{\lambda})\Big) \\ &= \sum_{i=1}^{q} \xi_i'(\hat{V}_d^{(i)})^{-1}\xi_i - \sum_{i=1}^{q}\sum_{j=1}^{q} \sqrt{n_i n_j}\,\xi_i'(\hat{V}_d^{(i)})^{-1}\tilde{W}^{-1}(\hat{V}_d^{(j)})^{-1}\xi_j \end{aligned}$$

holds. Hence

$$T_{n_1,\dots,n_q} = \xi'\hat{A}\xi\,, \quad \xi = (\xi_1', \dots, \xi_q')'\,, \;\; \hat{A} = \hat{B} - \hat{C}\,, \tag{4.9}$$

where the block-diagonal matrix $\hat{B} = \operatorname{diag}((\hat{V}_d^{(1)})^{-1}, \dots, (\hat{V}_d^{(q)})^{-1})$ and the (i,j)th block of the matrix $\hat{C}$ equals

$$\hat{a}_i^{1/2}\,(\hat{V}_d^{(i)})^{-1}\Big(\sum_{z=1}^{q} \hat{a}_z(\hat{V}_d^{(z)})^{-1}\Big)^{-1}\hat{a}_j^{1/2}\,(\hat{V}_d^{(j)})^{-1}\,.$$

The relations (4.8) and (2.11) imply that

$$\hat{A} = A + o_P(1)\,, \tag{4.10}$$

where $A = B - C$ and B, C are the matrices which one obtains by replacing $\hat{V}_d^{(i)}$, $\hat{a}_i$ in $\hat{B}$ and $\hat{C}$ with their limits $V_d^{(i)}$, a_i, respectively. Since $\mathcal{L}(\xi) \to N_{qd}(0, B^{-1})$, from (4.9), (4.10) one obtains that the first line of (3.4) is true if

$$\mathcal{L}(x'Ax \,|\, N_{qd}(0, B^{-1})) = \chi^2_{(q-1)d} \,. \tag{4.11}$$

The matrix CB^{-1} is idempotent and therefore AB^{-1} also has this property. But for

$$W = \sum_{z=1}^{q} a_z (V_d^{(z)})^{-1}$$

the equalities

$$\operatorname{tr}(AB^{-1}) = qd - \operatorname{tr}(CB^{-1}) = qd - \sum_{i=1}^{q} \operatorname{tr}\left(a_i (V_d^{(i)})^{-1} W^{-1}\right) = qd - \operatorname{tr}(WW^{-1}) = (q-1)d$$

hold and therefore Theorem 9.2.1 of [6] implies the validity of (4.11).

To prove the second line in (3.4) assume that

$$\overline{\lambda}_i = (\lambda_1^{(i)}, \ldots, \lambda_d^{(i)})', \quad \overline{\overline{\lambda}} = W^{-1} \sum_{i=1}^{q} a_i (V_d^{(i)})^{-1} \overline{\lambda}_i \,,$$

where the matrices W, $V_d^{(i)}$ are defined in the previous part of the proof. It follows from (2.8) and (2.11) that

$$\tilde{\lambda} = \overline{\overline{\lambda}} + o_P(1) \,.$$

Since now the identity $\overline{\lambda}_1 = \ldots = \overline{\lambda}_q$ does not hold, $\overline{\lambda}_i \neq \overline{\overline{\lambda}}$ for some i. For this index

$$(\lambda_i^* - \tilde{\lambda})'(\hat{V}_d^{(i)})^{-1}(\lambda_i^* - \tilde{\lambda}) \longrightarrow (\overline{\lambda}_i - \overline{\overline{\lambda}})'(V_d^{(i)})^{-1}(\overline{\lambda}_i - \overline{\overline{\lambda}}) > 0$$

in probability. Hence

$$T_{n_1,\ldots,n_q} \geq n_i (\lambda_i^* - \tilde{\lambda})'(\hat{V}_d^{(i)})^{-1}(\lambda_i^* - \tilde{\lambda}) \longrightarrow +\infty$$

in probability. □

Proof of Theorem 3.2. Since $\hat{\sigma}_i^2 = b'\hat{V}_d^{(i)} b$, where $b \in R^d$ is the vector with all coordinates equal to 1, and according to the proof of the previous theorem the matrices $\hat{V}_d^{(1)}, \ldots, \hat{V}_d^{(q)}$ are regular with probability tending to 1 as $n_1 \to \infty, \ldots, n_q \to \infty$, the statistics (3.6) are well-defined with probability tending to 1 as $n_1 \to \infty, \ldots, n_q \to \infty$.

If $\sum_{j=1}^d \lambda_j^{(1)} = \ldots = \sum_{j=1}^d \lambda_j^{(q)} = \kappa$, then for the statistic (3.6) the equality (4.9) holds with $\xi_i = n_i^{1/2}(\hat{\kappa}_i - \kappa)$, $\hat{B} = \operatorname{diag}(\hat{\sigma}_1^{-2}, \ldots, \hat{\sigma}_q^{-2})$ and the matrix $\hat{C}$ consisting of the elements $(\hat{a}_i \hat{a}_j)^{1/2} (\hat{\sigma}_i \hat{\sigma}_j)^{-2} (\sum_{z=1}^d \hat{a}_z \hat{\sigma}_z^{-2})^{-1}$. Hence the assertion (3.7) can be proved analogously as (3.4). □

Proof of Theorem 3.3. Suppose that $R_d^{(1)} = \ldots = R_d^{(q)} = R_d$ and put $\sigma_i^2 = \delta_i{}'V^{(i)}\delta_i$, where $\delta_i = (\delta_i(1), \ldots, \delta_i(k))'$ is defined by means of the formula (3.12) in which $\hat{\lambda}_z^{(i)}$, $\hat{\lambda}_w^{(i)}$ are replaced with $\lambda_z^{(i)}$, $\lambda_w^{(i)}$. Then (2.11) and (2.8) imply that $\hat{\sigma}_i^2 = \sigma_i^2 + o_P(1)$ and since for $\xi_i = \sqrt{n_i}(\hat{R}_d^{(i)} - R_d)$ according to (2.17) the weak convergence of distributions

$$\mathcal{L}(\xi_i) \longrightarrow N_1(0, \sigma_i^2)$$

holds, the rest of the proof can be carried out similarly as in the case of Theorems 3.1 and 3.2, and is therefore omitted. □

(Received December 22, 1999.)

REFERENCES

[1] A. W. Davis: Asymptotic theory for principal components analysis: non-normal case. Austral. J. Statist. *19* (1977), 206–212.
[2] T. Kollo and H. Neudecker: Asymptotics of eigenvalues and unit-length eigenvectors of sample variance and correlation matrices. J. Multivariate Anal. *47* (1993), 283–300.
[3] J. R. Magnus and H. Neudecker: Matrix Differential Calculus with Applications in Statistics and Econometrics. Wiley, New York 1988.
[4] K. V. Mardia, J. T. Kent, and J. M. Bibby: Multivariate Analysis. Academic Press, New York 1979.
[5] M. Okamoto: Distinctness of the eigenvalues of a quadratic form in a multivariate sample. Ann. Statist. *1* (1973), 763–765.
[6] C. R. Rao and S. K. Mitra: Generalized Inverse of Matrices and its Applications. Wiley, New York 1971.
[7] F. Rublík: On consistency of the MLE. Kybernetika *31* (1995), 45–64.
[8] F. H. Ruymgaart and S. Yang: Some applications of Watson's perturbation approach to random matrices. J. Multivariate Anal. *60* (1997), 48–60.
[9] J. R. Schott: Some tests for common principal component subspaces in several groups. Biometrika *78* (1991), 771–777.
[10] J. R. Schott: Asymptotics of eigenprojections of correlation matrices with some applications in principal components analysis. Biometrika *84* (1997), 327–337.
[11] D. E. Tyler: The asymptotic distribution of principal component roots under local alternatives to multiple roots. Ann. Statist. *11* (1983), 1232–1242.
[12] C. M. Waternaux: Asymptotic distribution of the sample roots for a nonnormal population. Biometrika *63* (1976), 639–645.

Doc. RNDr. František Rublík, CSc., Institute of Measurement Science, Slovak Academy of Sciences, Dúbravská cesta 9, 842 19 Bratislava. Slovakia.
e-mail: umerrubl@savba.sk