Mathematica Slovaca

DOI: 10.2478/s12175-009-0149-9
Math. Slovaca **59** (2009), No. 5, 579–592

A TEST OF THE HYPOTHESIS OF PARTIAL COMMON PRINCIPAL COMPONENTS

František Rublík

Dedicated to Professor Andrej Pázman on the occasion of his 70th birthday

(*Communicated by Gejza Wimmer*)

ABSTRACT. A test of the equality of the first h eigenvectors of covariance matrices of several populations is constructed without the assumption that the sampled distributions are Gaussian. It is proved that the test statistic is asymptotically chi-square distributed. In this general setting, an explicit formula for column space of the asymptotic covariance matrix of the sample eigenvectors is derived and the rank of this matrix is computed. An essential assumption in deriving the asymptotic distribution of the presented test statistic is the existence of the finite fourth moments and the simplicity of the h largest eigenvalues of population covariance matrices, which makes possible to use the formulas for derivatives of eigenvectors of symmetric matrices.

1. Introduction

Principal component analysis looks for small number h of linear combinations $\hat{p}_i'(x - \overline{x})$ of the observed data vector x, which can be used to summarize observed values. Since the sample vector x can be difficult to interpret, these linear combinations may help in discovering some overall properties of the original variables. For example, if in $x = (x_1, \dots, x_n)$ the coordinates $x_1, \dots, x_h$ denote some physical properties and $x_{h+1}, \dots, x_n$ some accompanying variables (time, space position, etc.), then one could ask whether $\hat{p}_1, \dots, \hat{p}_h$ lie in the space consisting of $(x_1, \dots, x_h, 0, \dots, 0)'$, because this would mean the relevant part of the variability of the data can be attributed to the physical quantities.

2000 Mathematics Subject Classification: Primary 62H15, 62H25.
Keywords: asymptotic distribution of eigenvectors, rank of asymptotic covariance matrix, equality of eigenvectors of several populations.
This research was supported by Grant No. 1/0077/09 from VEGA.

Springer

The problem of testing of null hypotheses on the set inclusion between a fixed linear space and the linear space spanned by m eigenvectors of a limiting matrix (including the covariance matrix as a special case) is studied in [16] and [17]. Asymptotic distribution of the sample total eigenprojection $\hat{P} = \sum \hat{p}_j \hat{p}_j'$ is in the case of the correlation matrix investigated by Schott in [14], who presents in this setting also tests of the null hypotheses H_0: $P = P_0$ and H_0: $P_0 P = P_0$ on the population total eigenprojection $P = \sum p_j p_j'$ for sampling from the normal population.

Suppose that the same variables $x_1, \dots, x_k$ are being measured in $q > 1$ statistical populations. Reduction of dimensionality simultaneously in these q groups can be viewed as a situation when the subspace spanned by the first h principal components is the same for all the groups. Thus if $P^{(i)}$ denotes the total eigenprojection corresponding to the h largest eigenvalues of the covariance matrix of the ith population, verification of this property amounts to testing the hypothesis H_0: $P^{(1)} = \dots = P^{(q)}$, which is investigated in the normal setting in [13], for the case $h = k$ is this problem investigated by Flury in [4] and [2].

Another problem associated with the overall picture of these q populations is the determination of the dimension r of the reduced space which retains most of the variability within all groups. In this context a test of the hypothesis H_0: $\mathrm{rank}(P^{(1)} + \dots + P^{(q)}) = s$ is under normality assumptions presented in [15].

If the observations from the q populations are summarized by linear combinations, it is reasonable to use the same number h of the combinations because of their coordinate-wise comparison from the multivariate point of view. If the data from the ith population are summarized by linear combinations $\hat{p}_j^{(i)'}(X - \overline{X}^{(i)})$, $j = 1, \dots, h$, then the equalities $\hat{p}_j^{(1)'} = \dots = \hat{p}_j^{(q)'}$, $j = 1, \dots, h$ mean that the hidden mechanism producing an essential part of the variability of the data is the same for all the q populations. This can be verified by testing that the violation of these equalities is not statistically significant, the null hypothesis in this case is that the eigenvectors of the population covariance matrices corresponding to their h largest eigenvalues, are common to all the q populations. This hypothesis, denoted here by (2.13), is in [3] called the partial common principal components model, its testing is investigated ibidem in the Gaussian case.

One of the important aspects of the principal component analysis is its role in the multivariate description of the data. This in the multisample case includes the task of finding out how similar the q groups are with respect to their overall features. This could mean that the principal axes of the concentration ellipsoid, corresponding to the h largest eigenvalues, are common to the all q populations. In terms of eigenvectors this property means that the eigenvectors of the population covariance matrices corresponding to their h largest eigenvalues, are common to all the q populations, which is again the hypothesis of partial common principal components. The aim of this paper is to construct an asymptotic test of this hypothesis without the normality assumptions. Similarly as in [12], an essential condition imposed in this paper is the distinctness of eigenvalues of covariance matrices, which is used in establishing the asymptotic distribution of the proposed test statistic.

2. Assumptions and the main results

It is supposed throughout the paper that $1 \leq h \leq k$ are fixed integers. Let X be a k-dimensional random vector and Σ denote its covariance matrix. Its spectral decomposition

$$\Sigma = P \operatorname{diag}(\lambda_1, \dots, \lambda_k) P' , \qquad \lambda_1 \geq \cdots \geq \lambda_k , \quad PP' = I_k , \tag{2.1}$$

$$P = \Big(p_1, \dots, p_k\Big) , \qquad p_j = (p_{1j}, \dots, p_{kj})' , \quad j = 1, \dots, k, \tag{2.2}$$

where I_k is the $k \times k$ identity matrix. The following conditions will be used.

(RC I) *The first h eigenvalues of the matrix Σ are simple. Thus if $h < k$, then $\lambda_1 > \cdots > \lambda_{h+1} \geq 0$ and if $h = k$, then $\lambda_1 > \cdots > \lambda_k > 0$.*

(RC II) *All mixed moments of the fourth order of coordinates of the random vector X are finite.*

For $k \times k$ matrix A with elements a_{ij}, the columns of A stacked one underneath the other will be denoted by $\operatorname{vec}(A)$, and $v(A)$ denotes its elements not above diagonal, i.e.,

$$v(A) = (a_{11}, a_{21}, a_{31}, \dots, a_{k1}, a_{22}, a_{32}, \dots, a_{k2}, \dots, a_{kk})' . \tag{2.3}$$

Throughout the paper $e_j = (0, \dots, 1, \dots, 0)'$ denotes the jth basis vector from $\mathbb{R}^k$, $\otimes$ denotes the Kronecker product and the superscript $^+$ denotes the Moore-Penrose inverse of the matrix. For column vectors $X_1, \dots, X_n$ let

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

denote the sample mean and the sample covariance matrix, respectively.

Theorem 2.1. *Suppose that* (RC I) *and* (RC II) *hold.*

(I) *There exist a neighbourhood $U \subset \mathbb{R}^{k(k+1)/2}$ of the point $v(\Sigma)$ and a continuously differentiable mapping $\tilde{P} = \tilde{P}\big(v(\Sigma^*)\big)$ of the argument $v(\Sigma^*) \in U$, taking values in the set of $k \times h$ matrices with unit-norm orthogonal columns such that $\tilde{P}\big(v(\Sigma)\big) = (p_1, \dots, p_h)$, where P is the matrix from the spectral decomposition of Σ, and for every symmetric $k \times k$ matrix Σ^* with $v(\Sigma^*) \in U$*

$$\tilde{P}\big(v(\Sigma^*)\big)' \Sigma^* \tilde{P}\big(v(\Sigma^*)\big) = \Lambda^* , \qquad \Lambda^* = \operatorname{diag}(\lambda_1^*, \dots, \lambda_h^*) ,$$

with $\lambda_1^* \geq \cdots \geq \lambda_k^*$ denoting the eigenvalues of Σ^*.

(II) *Suppose* $\{X_m\}_{m=1}^{\infty}$ *are i.i.d. random vectors with distribution* $\mathscr{L}(X_i) = \mathscr{L}(X)$. *Let*

$$(\hat{p}_1, \dots, \hat{p}_h) = \tilde{P}\big(v(S_n)\big) \tag{2.4}$$

denote the first h *eigenvectors of the sample covariance matrix chosen in the continuous way, described in* (I). *Then*

$$\sqrt{n}\big(\mathrm{vec}(\hat{p}_1, \dots, \hat{p}_h) - \mathrm{vec}(p_1, \dots, p_h)\big) \longrightarrow N_{hk}(0, M_h) \tag{2.5}$$

in distribution. The covariance matrix of this normal distribution

$$M_h = \Gamma\, W\, \Gamma' , \qquad \Gamma' = (\Gamma_1', \dots, \Gamma_h') \tag{2.6}$$

where $W = \mathrm{var}\{\mathrm{vec}(\,(X-\mu)(X-\mu)')\}$ *and* (cf. (2.1))

$$\begin{gathered} \Gamma_j = \Big(\mathrm{vec}(Pe_je_1'C_j)\,, \mathrm{vec}(Pe_je_2'C_j)\,, \dots, \mathrm{vec}(Pe_je_k'C_j)\Big)', \\ C_j = (\lambda_j I_k - \Sigma)^+ = P\ \mathrm{diag}\Big(\tfrac{1}{\lambda_j-\lambda_1}, \dots, \tfrac{1}{\lambda_j-\lambda_{j-1}}, 0, \tfrac{1}{\lambda_j-\lambda_{j+1}}, \dots, \tfrac{1}{\lambda_j-\lambda_k}\Big)P'\,. \end{gathered} \tag{2.7}$$

In the case $h < k-1$ the previous theorem allows the covariance matrix Σ to be singular. We remark that another form of the asymptotic covariance matrix is presented in [6, Theorem 3.1.9, p. 299] (this Theorem 3.1.9 is formulated for $h = k$ and therefore requires the non-singularity of Σ). An advantage of the formulas (2.6) and (2.7) is that they yield the explicit formula

$$M_{h,(i,j)} = \Gamma_i W \Gamma_j' \tag{2.8}$$

for asymptotic covariance of the eigenvectors $\hat{p}_i$, $\hat{p}_j$.

The previous theorem is based on the differentiability of $\tilde{P}$. Therefore one could ask whether the value of the eigenvectors computed in a particular case is compatible with such a condition. Since the eigenvectors corresponding to a simple eigenvalue form a one-dimensional linear subspace, to achieve this differentiability it is enough to find a rule which chooses the eigenvectors in a unique way continuous at $v(\Sigma)$. To ensure this, the quoted [6, Theorem 3.1.9] requires that the ith coordinate $p_i(i) > 0$ for $i = 1, \dots, k$. However, this assumption is unnecessarily restrictive, because the convergence (2.5) holds also for covariance matrices Σ with $p_i(i) = 0$, like $\Sigma = \mathrm{diag}(1, 2)$. Instead of this rule we therefore propose to choose the eigenvectors p_i in such a way that

$$p_{j_i i} > 0\,, \qquad j_i = \min\Big\{j:\ |p_{ji}| = \max\{|p_{1i}|, \dots, |p_{ki}|\Big\}\,, \tag{2.9}$$

because this choice makes the function $v(\Sigma) \longrightarrow p_i$ continuous provided that the eigenvalue λ_i is simple.

Although the eigenvalues are in (I) of the previous theorem chosen in a specific way, this does not hinder an application of this theorem in the further text, because the statistic proposed in Theorem 2.4 remains the same both for a given value of the sample eigenvector and for its value obtained by the multiplication with -1.

In deriving algebraic properties of the asymptotic covariance matrix (2.6) the following assertion will be useful.

Theorem 2.2. *Assume that the regularity condition* (RC II) *holds. If X possesses a density with respect to the Lebesgue measure, then both the covariance matrix Σ and the matrix*

$$V = V(X) = \operatorname{var}\Big\{v\Big((X-\mu)(X-\mu)'\Big)\Big\}, \qquad \mu = E(X), \tag{2.10}$$

are non-singular.

To avoid a confusion with the previous notation, in the next theorem the symbol $e_j^{(h)} = (0,\dots,1,\dots,0)'$ denotes the jth basis vector from $\mathbb{R}^h$. The symbol $\mathscr{M}(A)$ will denote the linear space generated by the columns of the matrix A and the symbol $\langle\, G\,\rangle$ the linear space generated by the set G.

Theorem 2.3. *Suppose that* (RC I), (RC II) *hold and both the covariance matrix Σ and the matrix* (2.10) *are non-singular. Then for the matrix M_h defined by means of* (2.6), (2.7) *the set equality*

$$\begin{aligned}\mathscr{M}(M_h) = \Big\langle \{\, e_j^{(h)} \otimes p_r - e_r^{(h)} \otimes p_j :\ 1 \le j < r \le h\} \\ \cup \{e_j^{(h)} \otimes p_r :\ 1 \le j \le h,\ \ h+1 \le r \le k\,\}\Big\rangle\end{aligned} \tag{2.11}$$

holds. All the vectors appearing on the right hand side of (2.11) *are linearly independent, if $1 \le h < k$ then $\operatorname{rank}(M_h) = hk - \frac{h(h+1)}{2}$ and if $h = k$, then $\operatorname{rank}(M_h) = \frac{(k-1)k}{2}$.*

If (RC I), (RC II) hold and X possesses a density with respect to the Lebesgue measure, then according to Theorem 2.2 the assumptions of Theorem 2.3 are fulfilled. Hence in this case the formulas for the rank and the column space of the asymptotic covariance matrix M_h given in Theorem 2.3 hold and the column space $\mathscr{M}(M_h)$ is not influenced by the matrix W appearing in (2.6).

The main topic of this paper is testing of the null hypothesis that the eigenvectors corresponding to the h largest eigenvalues of covariance matrices of k-dimensional populations are common to all underlying distributions. A test statistic for testing this hypothesis and its asymptotic distribution are presented in the next Theorem 2.4. To describe the setting of this problem, suppose throughout the rest of this paper that $q > 1$, $1 \le h < k$ are fixed integers and for $i = 1,\dots,q$ the random sample $X_1^{(i)},\dots,X_{n_i}^{(i)}$ of size n_i is drawn from a distribution with a $k \times k$ non-singular covariance matrix $\Sigma^{(i)}$, the random vector $X_1^{(i)}$ and its covariance matrix fulfil the assumptions of theorems 2.1, 2.3 and these random samples are independent. Thus for $i = 1, \ldots, q$ the covariance matrix $\Sigma^{(i)}$ has eigenvalues

$$\lambda_1^{(i)} \geq \cdots \geq \lambda_k^{(i)} > 0\,, \tag{2.12}$$

where the inequalities $\lambda_1^{(i)} > \cdots > \lambda_{h+1}^{(i)}$ hold and

$$\Sigma^{(i)} = P^{(i)} \operatorname{diag}(\lambda_1^{(i)}, \ldots, \lambda_k^{(i)}) P^{(i)'}\,, \qquad P^{(i)} = \left(p_1^{(i)}, \ldots, p_k^{(i)}\right), \;\; P^{(i)} P^{(i)'} = I_k\,.$$

Since the unit-length eigenvectors are uniquely defined up to the multiplication with -1, the null hypothesis of partial common principal components $p_j^{(1)} = \cdots = p_j^{(q)}$, $j = 1, \ldots, h$, is equivalent to

$$p_j^{(1)} p_j^{(1)'} = \cdots = p_j^{(q)} p_j^{(q)'}\,, \qquad j = 1, \ldots, h\,. \tag{2.13}$$

Further it is assumed for all i that the vector $X_1^{(i)}$ has all mixed moments of the fourth order finite and the covariance matrix (cf. (2.3))

$$V^{(i)} = V(X_1^{(i)}) = \operatorname{var}\Big\{ v\Big((X_1^{(i)} - \mu^{(i)})(X_1^{(i)} - \mu^{(i)})' \Big) \Big\} \tag{2.14}$$

is non-singular (here $\mu^{(i)} = E(X_1^{(i)})$). Suppose further that $S_{n_i}^{(i)}$ denotes the sample covariance matrix of the random sample from the ith population and

$$\begin{gathered} S_{n_i}^{(i)} = \hat{P}^{(i)} \operatorname{diag}(\hat{\lambda}_1^{(i)}, \ldots, \hat{\lambda}_k^{(i)}) \hat{P}^{(i)\prime}\,, \\ \hat{P}^{(i)} = \left(\hat{p}_1^{(i)}, \ldots, \hat{p}_k^{(i)}\right), \qquad \hat{P}^{(i)} \hat{P}^{(i)\prime} = I_k\,, \qquad \hat{\lambda}_1^{(i)} \geq \cdots \geq \hat{\lambda}_k^{(i)}\,. \end{gathered} \tag{2.15}$$

Estimate the covariance matrix (2.6) from the ith sample by the estimate

$$\hat{M}_h^{(i)} = \hat{\Gamma}^{(i)} \hat{W}^{(i)} \hat{\Gamma}^{(i)'}\,, \qquad \hat{\Gamma}^{(i)'} = (\hat{\Gamma}_1^{(i)'}, \ldots, \hat{\Gamma}_h^{(i)'})\,, \tag{2.16}$$

$$\begin{gathered} \hat{\Gamma}_j^{(i)} = \Big(\operatorname{vec}(\hat{P}^{(i)} e_j e_1' \hat{C}_j^{(i)})\,, \operatorname{vec}(\hat{P}^{(i)} e_j e_2' \hat{C}_j^{(i)})\,, \ldots, \operatorname{vec}(\hat{P}^{(i)} e_j e_k' \hat{C}_j^{(i)}) \Big)'\,, \\ \hat{C}_j^{(i)} = (\hat{\lambda}_j^{(i)} I_k - S_{n_i}^{(i)})^+\,, \end{gathered} \tag{2.17}$$

$$\begin{aligned} &\hat{W}^{(i)} = \tfrac{1}{n_i} \sum_{m=1}^{n_i} Y_{i,m} Y_{i,m}' - \overline{Y}_i \overline{Y}_i'\,, \quad Y_{i,m} = \operatorname{vec}\Big((X_m^{(i)} - \overline{X^{(i)}})(X_m^{(i)} - \overline{X^{(i)}})' \Big)\,, \\ &\overline{Y_i} = \tfrac{1}{n_i} \sum_{m=1}^{n_i} Y_{i,m}\,, \quad \overline{X^{(i)}} = \tfrac{1}{n_i} \sum_{m=1}^{n_i} X_m^{(i)}\,. \end{aligned}$$

Now define for $p_1, \ldots, p_h \in \mathbb{R}^k$ the block diagonal matrix by the formula

$$\overline{\partial}\big(p_1, \ldots, p_h\big) = \operatorname{diag}\left(D_k^+ \partial\,(p_1), \ldots, D_k^+ \partial\,(p_h) \right)\,, \qquad \partial\,(p) = (p \otimes I_k) + (I_k \otimes p)\,. \tag{2.18}$$

Here D_k is the duplication matrix, i.e., according to [8, (5), p.49], the equality

$$v(A) = D_k^+ \operatorname{vec}(A) \tag{2.19}$$

holds for every symmetric $k \times k$ matrix A. Obviously

$$D_k^+ = \Big(B_1', \dots, B_j', \dots, B_k'\Big)', \qquad B_j = \Big(e^{(k^2)}_{(j-1)k+j}, e^{(k^2)}_{(j-1)k+j+1}, \dots, e^{(k^2)}_{jk}\Big)' . \tag{2.20}$$

Use (2.18) and put $\overline{\partial}\Big(p_1, \dots, p_h\Big) = \operatorname{diag}\Big(D_k^+ \partial(p_1), \dots, D_k^+ \partial(p_h)\Big)$. To construct a test statistic for testing the hypothesis of partial common principal components, assume that

$$\hat\Psi_h^{(i)} = \overline{\partial}\Big(\hat p_1^{(i)}, \dots, \hat p_h^{(i)}\Big) \hat M_h^{(i)} \overline{\partial}\Big(\hat p_1^{(i)}, \dots, \hat p_h^{(i)}\Big)', \qquad \tilde\Psi = \sum_{i=1}^{q} \hat a_i \hat\Psi_h^{(i)+}, \tag{2.21}$$

where $\hat a_i = \frac{n_i}{n}$, $n = n_1 + \dots + n_q$. Employ the spectral decomposition

$$\tilde\Psi = \sum_{j=1}^{\vartheta} \tilde\lambda_j \tilde p_j \tilde p_j', \qquad \tilde\lambda_1 \ge \dots \ge \tilde\lambda_\vartheta, \quad \vartheta = \frac{k(k+1)}{2} h, \quad \tilde p_i' \tilde p_j = \delta_{i,j},$$

where $\delta_{i,j}$ stands for the Kronecker delta and put

$$\overline{\pi} = \sum_{j=1}^{r} \tilde p_j \tilde p_j', \quad \overline{\Psi} = \sum_{j=1}^{r} \tilde\lambda_j \tilde p_j \tilde p_j', \quad \overline{\Psi}_h^{(i)} = \overline{\pi}\, \hat\Psi_h^{(i)+} \overline{\pi}, \qquad r = hk - \frac{h(h+1)}{2}. \tag{2.22}$$

Theorem 2.4. *Suppose that* $\{p_j^{(i)}\}_{j=1}^k$ *are eigenvectors of* $\Sigma^{(i)}$, *i.e.* (cf. (2.12))

$$\Sigma^{(i)} p_j^{(i)} = \lambda_j^{(i)} p_j^{(i)}, \qquad j = 1, \dots, k. \tag{2.23}$$

In accordance with (2.15) *let*

$$\tau_h^{(i)} = \Big(v(p_1^{(i)} p_1^{(i)\prime})', \dots, v(p_h^{(i)} p_h^{(i)\prime})'\Big)', \quad \hat\tau_h^{(i)} = \Big(v(\hat p_1^{(i)} \hat p_1^{(i)\prime})', \dots, v(\hat p_h^{(i)} \hat p_h^{(i)\prime})'\Big)',$$
$$\overline{\tau} = \overline{\Psi}^+ \sum_{i=1}^{q} \hat a_i \overline{\Psi}_h^{(i)} \hat\tau_h^{(i)}.$$

Define the test statistic by the formula

$$Z_{n_1,\dots,n_q} = \sum_{i=1}^{q} n_i (\hat\tau_h^{(i)} - \overline{\tau})' \, \overline{\Psi}_h^{(i)} (\hat\tau_h^{(i)} - \overline{\tau}).$$

If $\tau_h^{(1)} = \tau_h^{(2)} = \dots = \tau_h^{(q)}$ *then the distribution of* $Z_{n_1,\dots,n_q}$ *is asymptotically chi-squared with* $(q-1)\Big(hk - \frac{h(h+1)}{2}\Big)$ *degrees of freedom as* $n_1 \to \infty, \dots, n_q \to \infty$.

Since the hypothesis of the partial principal components is equivalent to the equalities $\tau_h^{(1)} = \tau_h^{(2)} = \dots = \tau_h^{(q)}$, in accordance with the previous theorem this hypothesis is rejected whenever the statistic $Z_{n_1,\dots,n_q}$ exceeds the $1-\alpha$ quantile of the chi-square distribution with $(q-1)(hk - \frac{h(h+1)}{2})$ degrees of freedom.

3. Lemmas and proofs

Lemma 3.1. *Suppose that Σ is a symmetric $k \times k$ matrix and its spectral decomposition is described in* (2.1), (2.2). *Let j be a fixed integer from $\{1, \dots, k\}$ and λ_j be a simple eigenvalue of Σ. There exist a neighbourhood $U \subset \mathbb{R}^{k(k+1)/2}$ of the point $v(\Sigma)$, and a continuously differentiable mapping $\hat{p}_j = \hat{p}_j(v(\Sigma^*))$ of the argument $v(\Sigma^*) \in U$ taking values in $\mathbb{R}^k$, such that the following statements* (I)–(III) *are true.*

(I) *Let Σ^* be a symmetric $k \times k$ matrix. If $v(\Sigma^*) \in U$, then $\Sigma^* \hat{p}_j(v(\Sigma^*)) = \lambda_j^* \hat{p}_j(v(\Sigma^*))$, where $\lambda_1^* \geq \dots \geq \lambda_k^*$ denote the eigenvalues of Σ^*.*

(II) *$\hat{p}_j(v(\Sigma)) = p_j$ and $\|\hat{p}_j(v(\Sigma^*))\| = 1$ for all $v(\Sigma^*) \in U$.*

(III) *Let $\{S_n\}_{n=1}^{\infty}$ be symmetric random matrices and*

$$\sqrt{n}(S_n - \Sigma) = \mathcal{O}_P(1)\,. \tag{3.1}$$

If $v(S_n) \in U$, then for $t = 1, \dots, k$ the difference of coordinates

$$\hat{p}_{tj} - p_{tj} = \operatorname{vec}(p_j e_t' C_j)' \operatorname{vec}(S_n - \Sigma) + o_P(n^{-1/2})\,, \tag{3.2}$$

where the matrix $C_j = (\lambda_j I_k - \Sigma)^+$ and the abbreviated notation $\hat{p}_j = \hat{p}_j(S_n)$ is used.

Proof. According to [7, Theorem 7, p. 158] (cf. also [5, Formula (15.3), p. 565]) there exist a neighbourhood $U \subset \mathbb{R}^{k(k+1)/2}$ of the point $v(\Sigma)$ and a continuously differentiable $\mathbb{R}^k$ valued mapping $\hat{p}_j = \hat{p}_j(v(\Sigma^*))$ of the argument $v(\Sigma^*) \in U$ such that (I) and (II) of this lemma hold and

$$\frac{\partial \hat{p}_j\Big(v(\Sigma^*)\Big)}{\partial v(\Sigma^*)} = \Big(\lambda_j^* I_k - \Sigma^*\Big)^+ \frac{\partial \Sigma^*}{\partial v(\Sigma^*)} \hat{p}_j\Big(v(\Sigma^*)\Big)\,. \tag{3.3}$$

Hence if $v(S_n) \in U$ then by the Taylor expansion and (3.1) the difference of coordinates (the argument of the mapping is omitted)

$$\hat{p}_{tj} - p_{tj} = \sum_{r=1}^{k} \sum_{w=r}^{k} \frac{\partial p_{tj}}{\partial \Sigma_{rw}} \Big(S_{n,rw} - \Sigma_{rw}\Big) + o_P(n^{-1/2})\,. \tag{3.4}$$

Since the matrices S_n, Σ are symmetric and

$$\frac{\partial \Sigma}{\partial \Sigma_{rw}} = \begin{cases} e_r e_r' & r = w\,, \\ e_r e_w' + e_w e_r' & r < w\,, \end{cases}$$

from (3.3) and (3.4) after some computation one obtains (3.2). □

Proof of Theorem 2.1. Since all the h largest eigenvalues of Σ are simple, the assumptions of Lemma 3.1 hold for $j = 1, \dots, h$, from which (I) can be easily proved.

Suppose that the assumptions of (II) are fulfilled. Then (RC II) and the central limit theorem imply that $\sqrt{n}(\mathrm{vec}(S_n)-\mathrm{vec}(\Sigma))$ converges in distribution to $N_{k^2}(0,W)$, which together with delta method and (III) of the previous lemma yields (2.5) – (2.7). $\square$

P r o o f o f T h e o r e m 2.2. Suppose that $g \in \mathbb{R}^{k(k+1)/2}$ is a non-zero vector. Since $g'Vg = E\{h(X)\}$, $h(x) = \Big\{g'v\Big((x-\mu)(x-\mu)'\Big) - g'v(\Sigma)\Big\}^2$, it is sufficient to prove that the function $\xi(y) = g'v(yy') - g'v(\Sigma)$ equals zero on a set of Lebesgue measure zero. Since ξ is a polynomial in coordinates of y, acccording to [10, Lemma] this is true if there exists a vector $y \in \mathbb{R}^k$ such that $\xi(y) \neq 0$. To prove this assume that ξ is the zero polynomial, i.e.,

$$g'v(yy') = g'v(\Sigma) \tag{3.5}$$

for each $y \in \mathbb{R}^k$. Label by g_{jr} the coordinate in g on the position, corresponding in $v(\Sigma)$ to Σ_{jr}. Plug e_j and $2e_j$ into (3.5). Then $g'v(\Sigma) = g_{jj} = 4g_{jj}$, $g_{jj} = 0$ and $g'v(yy')$ is identically zero. Thus for $1 \le j < r \le k$ the equality $0 = g'v\{(e_j+e_r)(e_j+e_r)'\} = g_{jr}$ holds and g is a zero vector, which contradicts the assumption. $\square$

P r o o f o f T h e o r e m 2.3. According to [8, (3), p. 49] the equality $\mathrm{vec}(A) = D_k v(A)$, where D_k is the duplication matrix, holds for every symmetric $k \times k$ matrix. Therefore $\mathrm{vec}((X-\mu)(X-\mu)') = D_k v((X-\mu)(X-\mu)')$, which implies that the asymptotic covariance matrix in (2.5) is given by the formula

$$M_h = FVF', \qquad F = \Big(F_1' \ldots, F_h'\Big)', \qquad F_j = \Gamma_j D_k, \quad j = 1,\ldots,k, \tag{3.6}$$

where Γ_j is defined in (2.7) and V is the matrix (2.10). But according to the assumptions the matrix V is symmetric and positive definite, which together with (3.6) means, that $\mathscr{M}(M_h) = \mathscr{M}(FV^{1/2}) = \mathscr{M}(F)$. Further, let Z be a random vector which is $N_k(\mu,\Sigma)$ distributed. Since the matrix Σ is assumed to be non-singular, this normal distribution possesses a density with respect to the Lebesgue measure. Hence according to Theorem 2.2 the matrix $\tilde{V} = \mathrm{var}\Big\{v\Big((Z-\mu)(Z-\mu)'\Big)\Big\}$ is non-singular, and by the preceding argument for the asymptotic covariance matrix $\tilde{M}_h$ of the first h eigenvectors computed from the random sample from the normal $N_k(\mu,\Sigma)$ distribution the equality $\mathscr{M}(\tilde{M}_h) = \mathscr{M}(F)$ holds. Thus $\mathscr{M}(M_h) = \mathscr{M}(\tilde{M}_h)$ and we shall assume without the loss of generality that X is normally distributed. Under this assumption the matrix

M_h can be expressed in the block notation by the formula

$$M_h = (M_{h,st})\,, \quad s,t = 1,\dots,h\,, \qquad M_{h,st} = \begin{cases} \sum\limits_{\substack{w=1\\ w\neq s}}^{k} g_{sw} p_w p_w' & s = t\,, \\ -g_{st} p_t p_s' & s \neq t\,, \end{cases}$$

$$g_{sw} = \tfrac{\lambda_s \lambda_w}{(\lambda_s - \lambda_w)^2}\,. \tag{3.7}$$

We remark that this formula can be obtained either directly from (3.6), or from results of [1], in the case that all the eigenvalues of Σ are distinct, from [9, Theorem 8.3.3].

To make the notation more concise, put $b_j(x) = e_j^{(h)} \otimes x$, $b_{j,r}(x,y) = b_j(x) + b_r(y)$. Since the vectors $b_r(p_j)$, $1 \le r \le h$, $1 \le j \le k$ span $\mathbb{R}^{kh}$, for matrix (3.7) the equality

$$\mathscr{M}(M_h) = \big\langle \{M_h b_r(p_j):\ r = 1,\dots,h,\ j = 1,\dots,k\}\big\rangle \tag{3.8}$$

holds. Now denote by the notation $x \sim x^*$ the fact, that there exists a non-zero constant c such that $x = c\,x^*$. Then for $r = 1,\dots,h$

$$M_h b_r(p_j) = \begin{cases} b_{jr}(-g_{jr}p_r, g_{jr}p_j) \sim b_{jr}(p_r, -p_j) & j = 1,\dots,r-1\,, \\ 0 & j = r\,, \\ b_{rj}(g_{rj}p_j, -g_{rj}p_r) \sim b_{rj}(p_j, -p_r) & j = r+1,\dots,h\,, \\ b_r(g_{rj}p_j) \sim b_r(p_j) & j = h+1,\dots,k\,, \end{cases}$$

which together with (3.8) yields (2.11).

Since the proof of the formula for rank is similar in both cases, assume that $1 \le h < k$. If

$$\sum_{j=1}^{h}\Big(\sum_{r=j+1}^{h} \alpha_{jr} b_{jr}(p_r, -p_j) + \sum_{r=h+1}^{k} \beta_{jr} b_j(p_r)\Big) = 0\,,$$

then $\alpha_{12} = \dots = \alpha_{1h} = \beta_{1h+1} = \dots = \beta_{1k} = 0$ and

$$0 = \sum_{j=2}^{h}\Big(\sum_{r=j+1}^{h} \alpha_{jr} b_{jr}(p_r, -p_j) + \sum_{r=h+1}^{k} \beta_{jr} b_j(p_r)\Big)\,.$$

Thus in this fashion the nullity of all coefficients α_{jr}, β_{jr} can be obtained and the rest of the proof is obvious. □

Lemma 3.2. *Suppose that the assumptions of* Theorem 2.1 *are fulfilled and* (cf. (2.4))

$$\hat\tau = \Big(v(\hat p_1 \hat p_1')', \dots, v(\hat p_h \hat p_h')'\Big)', \qquad \tau = \Big(v(p_1 p_1')', \dots, v(p_h p_h')'\Big)'.$$

Then, as $n \to \infty$, the convergence in distribution (cf. (2.18)), (2.6)

$$\sqrt{n}(\hat{\tau} - \tau) \to N(0, \Psi_h)\,, \qquad \Psi_h = \overline{\partial}\big(p_1, \dots, p_h\big) M_h \, \overline{\partial}\big(p_1, \dots, p_h\big)' \tag{3.9}$$

holds. The column space and the rank of this matrix fulfil the relations

$$\operatorname{rank}(\Psi_{)} = \operatorname{rank}(M_h)\,, \qquad \mathscr{M}(\Psi_h) = \big\{\overline{\partial}\big(p_1, \dots, p_h\big)x : \ x \in \mathscr{M}(M_h)\,\big\}\,. \tag{3.10}$$

P r o o f. Since the equality $\left(\frac{\partial \operatorname{vec}(pp')}{\partial p(1)}, \dots, \frac{\partial \operatorname{vec}(pp')}{\partial p(k)}\right) = (I_k \otimes p) + (p \otimes I_k)$ holds, (3.9) can be easily proved by means of (2.19), delta method and Theorem 2.1. Further, since the matrix M_h is symmetric and positive semi-definite, the second equality in (3.10) obviously holds. Hence the first formula of (3.10) will be proved by showing that for any non-zero vector $p \in \mathbb{R}^k$ in the notation (2.18) the implication $D_k^+ \partial(p) y = 0 \Longrightarrow y = 0$ is true. However, this can be verified by means of the fact, that the equality $B_j \partial(p) y = 0$, where B_j is the matrix from (2.20), is equivalent to

$$2p_j y_j = 0, \quad p_{j+1} y_j + p_j y_{j+1} = 0, \ \dots \ p_k y_j + p_j y_k = 0\,.$$

$\square$

Lemma 3.3. *Let $M_h^{(i)}$, $\Gamma_j^{(i)}$, $W^{(i)}$ be the quantities defined in* (2.6), (2.7) *by means of $X = X_1^{(i)}$ (i.e., $M_h^{(i)}$ is the asymptotic covariance matrix of eigenvalues of $S_{n_i}^{(i)}$). Put* (cf. (2.18), (2.23))

$$\Psi_h^{(i)} = \overline{\partial}\Big(p_1^{(i)}, \dots, p_h^{(i)}\Big)\, M_h^{(i)} \, \overline{\partial}\Big(p_1^{(i)}, \dots, p_h^{(i)}\Big)'\,. \tag{3.11}$$

(I) *The convergence*

$$\hat{M}_h^{(i)} \to M_h^{(i)}\,, \qquad \hat{\Psi}_h^{(i)} \to \Psi_h^{(i)} \tag{3.12}$$

holds a.s..

(II) *Let*

$$\hat{a}_i \to a_i\,, \qquad i = 1, \dots, q\,, \tag{3.13}$$

and $\Psi = \sum\limits_{i=1}^{q} a_i \Psi_h^{(i)+}$. Suppose that for $j = 1, \dots h$ the vector $p_j^{(i)} = p_j$ does not depend on the population index i. Then (cf. (2.22))

$$\operatorname{rank}(\Psi_h^{(1)}) = \cdots = \operatorname{rank}(\Psi_h^{(q)}) = \operatorname{rank}(\Psi) = r \tag{3.14}$$

and a.s.

$$\overline{\Psi}_h^{(i)} \to \Psi_h^{(i)+}\,, \qquad \overline{\Psi}^+ \to \Psi^+\,. \tag{3.15}$$

P r o o f.

(I) Since $S_{n_i}^{(i)} \to \Sigma_i$ a.s., the convergence $\hat{p}_j^{(i)} \to p_j^{(i)}$ holds for $j = 1, \dots, h$ almost surely and the convergence $\hat{C}_j^{(i)} \to (\lambda_j^{(i)} I_k - \Sigma_i)^+$ can be proved similarly as in the proof of Lemma 3.1. Hence the first convergence in (3.12) can be easily established by means of the law of large numbers, the second one follows from the first one.

(II) Since $M_h^{(i)}$ is symmetric and positive semidefinite,

$$\mathscr{M}(\Psi_h^{(i)}) = \mathscr{M}\Big(\partial(p_1,\dots,p_h)M_h^{(i)}\Big) .$$

Hence by Theorem 2.3 the equalities $\mathscr{M}(\Psi_h^{(1)}) = \cdots = \mathscr{M}(\Psi_h^{(q)})$ hold, the dimension of this linear space equals r and since $\sum_{i=1}^{q} a_i\Psi_h^{(i)+} \gg a_i\Psi_h^{(i)+}$ in the sense of positive definitness, the validity of (3.14) is obvious. Further, similarly as in (3.6) put

$$\hat F_j^{(i)} = \hat\Gamma_j^{(i)} D_k ,$$

where D_k is the duplication matrix. Then the (j,r)th block of the matrix (2.16)

$$\hat\Gamma_j^{(i)}\hat W^{(i)}\hat\Gamma_r^{(i)\prime} = \hat F_j^{(i)}\hat V^{(i)}\hat F_r^{(i)\prime} ,$$

where with the notation $U_{i,m} = v\Big(\big(X_m^{(i)} - \overline{X}^{(i)}\big)\big(X_m^{(i)} - \overline{X}^{(i)}\big)'\Big)$

$$\hat V^{(i)} = \frac{1}{n_i}\sum_{m=1}^{n_i} U_{i,m}U_{i,m}' - v(S_{n_i}^{(i)})v(S_{n_i}^{(i)})'$$

is the sample counterpart of (2.14). Hence if the matrices $\hat V^{(i)}$, $S_{n_i}^{(i)}$ are nonsingular, then similarly as in the proof of Theorem 2.3

$$\begin{aligned} \mathscr{M}(\hat M_h^{(i)}) =& \big\langle\{\, b_{j,r}(\hat p_r^{(i)}, -\hat p_j^{(i)}) :\ 1 \le j < r \le h\} \\ & \cup \{b_j(\hat p_r^{(i)}) :\ 1 \le j \le h\,,\ h+1 \le r \le k\,\}\big\rangle , \\ & \operatorname{rank}(\hat M_h^{(i)}) = \operatorname{rank}(M_h^{(i)}) = r\,. \end{aligned} \tag{3.16}$$

But the matrix $\hat V^{(i)}$ is computed from the random sample of size n_i and therefore by the law of large numbers $\hat V^{(i)} \to V^{(i)}$. Since the limiting matrix is nonsingular, (3.16) holds a.s. for all u sufficiently large. But similarly as in the proof of Lemma 3.2

$$\operatorname{rank}(\hat\Psi_h^{(i)}) = \operatorname{rank}(\hat M_h^{(i)}) = r = \operatorname{rank}(\Psi_h^{(i)}) ,$$

and by (3.12) and (3.13) almost surely $\hat\Psi_h^{(i)+} \to \Psi_h^{(i)+}$, $\tilde\Psi \to \Psi$. Further, $\tilde\lambda_r > \tilde\lambda_{r+1}$ a.s. for all $n_1,\dots,n_q$ sufficiently large, which together with (3.14) means that $\overline{\pi} \to \pi$ a.s., where π is the projection on the linear space $\mathscr{M}(\Psi)$. Thus a.s.

$$\overline{\Psi}_h^{(i)} = \overline{\pi}\,\hat\Psi^{(i)+}\,\overline{\pi} \to \pi\Psi_h^{(i)+}\pi = \Psi_h^{(i)+} , \qquad \overline{\Psi} = \overline{\pi}\,\tilde\Psi\,\overline{\pi} = \sum_{i=1}^{q} \hat a_i\overline{\Psi}_h^{(i)} \to \Psi ,$$

and since by (2.22) $\operatorname{rank}(\,\overline{\Psi}\,) = r = \operatorname{rank}(\Psi)$ for all sample sizes $n_1,\dots,n_q$ sufficiently large a.s., also the second convergence in (3.15) holds. □

Proof of Theorem 2.4. Suppose without the loss of generality that also all assumptions of the previous lemma are fulfilled. Obviously $\sum_{i=1}^{q} \hat{a}_i \overline{\psi}_h^{(i)} = \overline{\pi}\,\tilde{\Psi}\,\overline{\pi} = \overline{\Psi}$ and $\mathscr{M}(\overline{\psi}_h^{(i)}) \subset \mathscr{M}(\overline{\psi})$. Therefore $\overline{\psi}_h^{(i)} = \overline{\psi}_h^{(i)}\,\overline{\psi}^{+}\,\overline{\psi}$ and for $\tau_h^{(1)} = \tau_h^{(2)} = \cdots = \tau_h^{(q)} = \tau$ the equality

$$\sum_{i=1}^{q} n_i(\hat{\tau}_i - \tau)'\,\overline{\psi}_h^{(i)} = n(\overline{\tau} - \tau)'\,\overline{\psi}$$

holds. Now if $\xi_i = \sqrt{n}_i(\hat{\tau}_h^{(i)} - \tau)$, $\xi = (\xi_1', \dots, \xi_q')'$, then by means of these results

$$\begin{aligned} Z_{n_1,\dots,n_q} &= \sum_{i=1}^{q} \xi_i' \,\overline{\psi}_h^{(i)} \xi_i - \; n(\overline{\tau} - \tau)'\,\overline{\Psi}(\overline{\tau} - \tau) \\ &= \sum_{i=1}^{q} \xi_i' \,\overline{\psi}_h^{(i)}\, \xi_i - \sum_{i=1}^{q}\sum_{j=1}^{q} \sqrt{\hat{a}_i \hat{a}_j}\, \xi_i' \,\overline{\psi}_h^{(i)}\,\overline{\Psi}^{+}\,\overline{\psi}_h^{(j)} \xi_j\,. \end{aligned} \tag{3.17}$$

The convergence in distribution (cf. (3.11))

$$\xi \to N(0, G)\,, \qquad G = \operatorname{diag}(\Psi_h^{(1)}, \dots, \Psi_h^{(q)}) \tag{3.18}$$

follows from Lemma 3.2. Further, combine (3.13), (3.17), (3.15) and (3.18) to establish that $Z_{n_1,\dots,n_q} = \xi' A \xi + o_P(1)$, $A = H - C$, where the block diagonal matrix $H = \operatorname{diag}(\Psi_h^{(1)+}, \dots, \Psi_h^{(q)+})$ and the (i,j)th block of the matrix C is $C_{ij} = \sqrt{a_i a_j}\Psi_h^{(i)+}\Psi^{+}\Psi_h^{(j)+}$. Therefore, by means of (3.14), the trace

$$\operatorname{tr}(HG) = \sum_{i=1}^{q} tr(\Psi_h^{(i)+}\Psi_h^{(i)}) = qr\,, \quad \operatorname{tr}(CG) = \sum_{i=1}^{q} tr(C_{ii}\Psi_h^{(i)}) = \operatorname{tr}(\Psi^{+}\Psi) = r\,. \tag{3.19}$$

Since (3.18), (3.19) hold and the matrix AG is idempotent, the rest of the proof can be carried out by means of [11, Theorem 9.2.1]. □

REFERENCES

[1] ANDERSON, T. W.: *Asymptotic theory for principal component analysis*, Ann. Math. Statist. **34** (1963), 122–148.

[2] FLURY, B. N.: *Asymptotic theory for common principal component analysis*, Ann. Statist. **14** (1986), 418–430.

[3] FLURY, B. N.: *Two generalizations of the common principal component model*, Biometrika **74** (1987), 59–69.

[4] FLURY, B. N.: *Common principal components in k groups*, J. Amer. Statist. Assoc. **79** (1984), 892–898.

[5] HARVILLE, D. A.: *Matrix Algebra from a Statistical Perspective*, New York, Springer-Verlag, 1997.

[6] KOLLO, T.—VON ROSEN, D.: *Advanced Multivariate Statistics with Matrices*, Dordrecht, Springer, 2005.
[7] MAGNUS, J. R.: *On differentiating eigenvalues and eigenvectors*, Econometric Theory **1** (1985), 179–191.
[8] MAGNUS, J. R.—NEUDECKER, H.: *Matrix Differential Calculus with Applications in Statistics and Econometrics*, New York, Wiley & Sons, 1991.
[9] MARDIA, K. V.—KENT, J. T.—BIBBY, J. M.: *Multivariate Analysis*, London, Academic Press, 1979.
[10] OKAMOTO, M.: *Distinctness of the eigenvalues of a quadratic form in a multivariate sample*, Ann. Statist. **1** (1973), 763–765.
[11] RAO, C. R.—MITRA, S. K.: *Generalized Inverse of Matrices and its Applications*, New York, Wiley & Sons, 1971.
[12] RUBLÍK, F.: *Tests of some hypotheses on characteristic roots of covariance matrices not requiring normality assumptions*, Kybernetika **37** (2001), 61–78.
[13] SCHOTT, J. R.: *Some tests for common principal component subspaces in several groups*, Biometrika **78** (1991), 771–777.
[14] SCHOTT, J. R.: *Asymptotics of eigenprojections of correlation matrices with some applications in principal component analysis*, Biometrika **84** (1997), 327–337.
[15] SCHOTT, J. R.: *Partial common principal subspaces*, Biometrika **86** (1999), 899–908.
[16] TYLER, D. A.: *Asymptotic inference for eigenvectors*, Ann. Statist. **9** (1981), 725–736.
[17] TYLER, D. A.: *A class of asymptotic tests for principal component vectors*, Ann. Statist. **11** (1983), 1243–1250.

Received 26. 11. 2008

Institute of Measurement Science
Dúbravská cesta 9
SK–84104 Bratislava
SLOVAKIA
E-mail: umerrubl@savba.sk