DISTRIBUTION OF THE RESIDUAL ROOTS IN PRINCIPAL COMPONENTS ANALYSIS

A. M. Kshirsagar

Defence Science Laboratory, Delhi

The distribution of latent roots of the covariance matrix of normal variables, when a hypothetical linear function of the variables is eliminated, is derived in this paper. The relation between the original roots and the residual roots—after elimination of ξ, is also derived by an analytical method. An exact test for the goodness of fit of a single non-isotropic hypothetical principal components, using the residual roots, is then obtained.

Let $x' = [x_1, \ldots, x_p]$ be a (row) vector having a p-variate normal distribution with zero means and variance-covariance matrix Σ. There is an orthogonal matrix

$$L = [l_{ij}] = [l_{(1)} \mid l_{(2)} \mid \ldots\ldots \mid l_{(p)}]' \tag{1}$$

such that

$$\Sigma = L' \; diag. \, (\sigma_1^2, \; \sigma_2^2, \; \ldots, \; \sigma_p^2) \; L \tag{2}$$

where σ^2_i $(i = 1, \ldots, p)$ are the latent roots of Σ and $l_{(i)}$ are the corresponding (column) latent vectors (diag. stands for a diagonal matrix, the elements of which are written in the adjoining bracket). If the roots are arranged in descending order of magnitude as

$$\sigma_1^2 \geqslant \sigma_2^2 \geqslant \ldots \geqslant \sigma_p^2$$

then

$$y_i = l'_{(i)} x \tag{3}$$

is called the i-th principal component. The transformation

$$y' = [y_1 \ldots y_p] = x' L' \tag{4}$$

to the principal components shows that the y_i are normal independent variables with zero means and variances σ_i^2.

Let

$$\underset{n \times p}{X} = [x_{ir}] \qquad \begin{array}{l} i=1,\ldots\ldots, p \\ r=1,\ldots\ldots, n \end{array} \tag{5}$$

be a sample of size n from the distribution of x. The maximum likelihood estimate of Σ is $\frac{1}{n} A$, where

$$A = [a_{ij}] = X'X \tag{6}$$

is the matrix of the sums of squares and products of the sample observations. There exists an orthogonal matrix G,

$$G = [g_{ij}] = [g_{(1)} \mid g_{(2)} \mid \ldots\ldots \mid g_{(p)}]' \tag{7}$$

such that

$$A = G' diag. (\theta_1^2 , \ldots , \theta_p^2) G \tag{8}$$

where

$$\theta_1^2 > \theta_2^2 > \ldots\ldots\ldots\ldots > \theta_p^2 \tag{9}$$

are the latent roots of A and $g_{(i)}$ are the corresponding latent vectors. Then $\frac{1}{n} \theta_i^2$ are the sample roots and

$$Z_i = g'_{(i)} x \tag{10}$$

are the sample principal components. They are the maximum likelihood estimates of the corresponding population parameters. The sample variance-covariance matrix of

$$Z' = [z_1, z_2, \ldots\ldots\ldots\ldots z_p] = x' G' \tag{11}$$

is obviously

$$\frac{1}{n} \text{diag.} (\theta_1^2, \ldots, \theta_p^2) \tag{12}$$

A SINGLE NON-ISOTROPIC PRINCIPAL COMPONENT

If all the roots σ_i^2 are equal, the variation of the x's is isotropic. However, if all the roots except σ_1^2, the largest root, are equal, the variation is not isotropic. It is so because of y_1, the first principal component. Hence y_1 is called the single non-isotropic principal component. There is no loss of generality in assuming the common value of all the roots, except σ_1^2, to be unity. In the case of such a single non-isotropic principal component, Σ is completely determined by σ_1^2 and $l_{(1)}$ the direction vector of y_1. Such a situation arises in factor analysis if there is only one (common) factor besides the specific factor in a factor-structure[1]. If this is the case, the problem of testing the goodness of fit of a single non-isotropic hypothetical principal component arises. Thus, if $h'x$ is a hypothetical function, we desire to test the hypothesis that $h'x$ is the same as the true non-isotropic component $l'_{(1)}x$. Since in the population, when $l'_{(1)}x$ is eliminated, the remaining roots $\sigma_2^2, \ldots, \sigma_p^2$ of Σ are equal, one feels that the criterion of the hypothesis can be based on the 'residual' sample roots of A when the hypothetical function $h'x$ is eliminated. The relationship of the original roots θ_i^2 and the residual roots ϕ_k^2 of A is obtained here,

RESIDUAL ROOTS

The hypothetical function $\xi = h'x$ can be expressed in terms of the sample principal components Z by using (11). Thus

$$\xi = h'x = h'G'z = w'z = w_1 z_1 + \ldots + w_p z_p \tag{13}$$

where,

$$w = Gh \tag{14}$$

and we assume that for normalization

$$w'w = 1 \tag{15}$$

The conditional covariance in the sample between z_i and z_j when ξ is fixed is (from 12)

$$\text{Cov.}\,(z_i\, Z_j \,|\omega' z) = \text{cov.}\,(Z_i, Z_j) - \frac{\text{cov.}\,(Z_i, \omega' z)\,\text{cov.}\,(Z_j, \omega' z)}{V\,(\omega' z)}$$

$$= \frac{1}{n}\,\delta_{ij}\,\theta_i^2 - \frac{\omega_i\,\theta_i^2 .\ \omega_j\,\theta_j^2}{n \sum\limits_1^p \omega_i^2\ \theta_i^2} \tag{16}$$

where δ_{ij} is the Kronecker delta and i, j run from 1 to p. The conditional covariance matrix of the z's when ξ is fixed is, therefore,

$$\frac{1}{n}\left[\theta_i^2\,\delta_{ij} - \frac{1}{\lambda^2}\,\omega_i\,\omega_j\,\theta_i^2\,\theta_j^2\right] \tag{17}$$

where

$$\frac{1}{n}\lambda^2 = \frac{1}{n}\sum_1^p \omega_i^2\,\theta_i^2 = \text{the sample variance of } \xi. \tag{18}$$

The latent roots of the above 'conditional' covariance matrix of the z's when ξ is eliminated are called the residual roots of the x's. The idea of residual roots is originally due to Williams[2]; he derived them by considering the intersection of the ellipsoid

$$\frac{z_1^2}{\theta_1^2} + \ldots + \frac{z_p^2}{\theta_p^2} = 1$$

and the hyperplane

$$w_1\,z_1 + \ldots + w_p\,z_p = 0$$

However, this geometrical derivation can be replaced by the above analytical method. Thus the residual roots are $\frac{1}{n}$ times the roots of the determinantal equation in ϕ^2:

$$\left|\,\theta_i^2\,\delta_{ij} - \frac{1}{\lambda^2}\,\omega_i\,\omega_j\,\theta_i^2\,\theta_j^2 - \phi^2\,\delta_{ij}\right| = 0 \tag{19}$$

This equation simplifies to

$$1 - \sum_1^p \frac{\omega_i^2\,\theta_i^4}{\lambda^2\,(\theta_i^2 - \phi^2)} = 0 \tag{20}$$

and can also be written as

$$\sum_{1}^{p} \frac{\omega_i^2 \ \theta_i^2}{\theta^2 - \phi^2} = 0 \qquad \text{as } \Sigma \ \omega_i^2 \ \theta_i^2 = \lambda^2 \tag{21}$$

Let ϕ_k^2 $(k = 1, \ldots, p-1)$ be the roots of this $(p-1)$ th degree equation in ϕ^2. Collecting the coefficients of $(\phi^2)^{p-2}$, $(\phi^2)^{p-2}$ and the constant term, it can be easily shown that

$$\prod_{k=1}^{p-1} \phi_k^2 = \frac{1}{\lambda^2} \prod_{i=1}^{p} \theta_i^2 \tag{22}$$

and

$$\sum_{1}^{p-1} \phi_k^2 = \sum_{1}^{p} \theta_i^2 - \frac{1}{\lambda^2} \sum_{1}^{p} \omega_i^2 \ \theta_i^4 \tag{23}$$

From (21) Williams[2] has proved that

$$\omega_i^2 = \frac{\frac{\lambda}{\theta_i^2} \prod_{j=1}^{p-1} (\phi_j^2 - \theta_i^2)}{\prod_{j=1, j\neq i}^{p} (\theta_j^2 - \theta_i^2)}, \ (i = 1, \ldots, p) \tag{24}$$

DISTRIBUTION OF THE RESIDUAL ROOTS

Since ϕ_k^2 $(k = 1, \ldots, p-1)$ are the residual roots, *i.e.*, derived from a conditional covariance matrix, it is obvious that their distribution is the same as those of the original roots θ_i^2, with n replaced by $n-1$ and p by $p-1$. The more important distribution is, however, of the original roots θ_i^2 and the residual roots ϕ_k^2 when λ^2 is held fixed. Fortunately it so happens that this latter distribution does not involve the nuisance parameter σ_1^2 and is, therefore, useful for deriving exact tests. This is so because λ^2 is a sufficient statistic for σ_1^2. We obtain this distribution under the null hypothesis:

H : Σ has one root $\sigma_1^2 > 1$; the remaining roots are all unity and the principal component corresponding to this root is the assigned function $\xi = h'x = w' z$. (25)

Since A is the matrix of the sums of squares and products (S.S. & S.P.) of the sample observations on x, it follows from (4) that the S.S. & S.P. matrix of the true principal components y is

$$B = LAL' \tag{26}$$

The variance-covariance matrix of y is diag. $(\sigma_1^2, 1, \ldots, 1)$ and hence the distribution of B is the Wishart distribution

$$\text{const.} \left| B \right|^{\frac{1}{2}(n-p-1)} \exp\left[-\tfrac{1}{2} \left(\frac{1}{\sigma_1^2} b_{11} + b_{22} + \ldots + b_{pp} \right) \right] dB, \tag{27}$$

where dB stands for the product of the differentials of the $p(p+1)/2$ distinct elements of B. From (4) and (11)

$$y = L G' z = W z \tag{28}$$

where

$$L\,G' = W = [W_{ij}] \tag{29}$$

W is orthogonal because L & G are so, *i.e.*

$$W\,W' = I_p\,, \tag{30}$$

where I_p denotes the identity matrix of order p. From (28), the true non-isotropic principal component is

$$y_1 = w_{11}\,z_1 + \ldots + w_{1p}\,z_p \tag{31}$$

The assigned function is $\xi = w'z'$. Hence if H of (25) is true the two vectors w' and $[w_{11}, \ldots, w_{1p}]$ are the same. Also from (26) and (8) with (29) we have

$$\mathrm{B} = \mathrm{LAL}' = LG'\,diag.\,[\theta_1^2, \ \ldots, \ \theta_p^2\,)\,GL'. = W\,\Theta\,W' \tag{32}$$

where

$$\Theta = diag.\,(\theta^2, \ldots, \theta_p^2\,) \tag{33}$$

In the distribution of B, transform from B to Θ and W by (32) and (30). Since W is orthogonal, there are $p\,(p+1)/2$ constraints on the elements of W and only $p(p-1)/2$ elements are functionally independent. They can be taken to be W_{ij} ($j > i$; $i, j = 1, \ldots\ldots\ldots, p$). We shall denote by dW, the product of the differential of these $p(p-1)/2$ elements. The transformation from B to Θ and W is unique only if we further impose the condition $w_{1j} > o$ for all i. The Jacobian of this transformation [3,4] is the absolute value of

$$\prod_{i<j} (\theta_i^2 - \theta_j^2) \div \prod_{q=1}^{p-1} \left| w_q \right| \tag{34}$$

where Wq is the matrix of the first q rows and q columns of W. The joint distribution of Θ and W, therefore, comes out to be

$$\text{const.}\ p\,(\Theta).\ \sigma_1^{-n} \exp\left[-\frac{1}{2}\lambda^2\left(\frac{1}{\sigma_1^2} - 1\right)\right] \prod_{q=1}^{p-1} \left| W_q \right|^{-1} d\Theta\,dW. \tag{35}$$

where [5 or 6]

$$p\,(\Theta) = \prod_{i=1}^{p} \left\{ (\theta_i^2)^{(n-p-1)/2} \exp\left(-\frac{1}{2}\theta_i^2\right) \right\} \pi_{i<j}\ (\theta_i^2 - \theta_j^2) \tag{36}$$

We shall now integrate out all the elements of W, except those in its first row, viz, w_{li} ($i = 2, \ldots\ldots\ldots\ldots\ldots\ldots, p$). For this we need

$$I = \int \frac{1}{\left| w_q \right|}\,dw_q \tag{37}$$

where

$$W'_q = [w_{q,q+1} \ \ldots\ldots\ldots\ \ w_{q,p}\,] \tag{38}$$

and the range of integration is such that $ww' = I_p$.

Let

$\triangle$ = the matrix of the first q–1 rows and p columns of W, and (39)
C = The matrix of the first q rows and p columns of W

$$= \left[\ w_q \left| \frac{\Delta}{w'_q} \right. \right]_1^{q-1} \tag{40}$$

Hence

$$C\,C' = I_q = W_q\,W'_q + \left[\frac{\Delta\ \Delta'}{w'_q\ \Delta'}\Bigg|\frac{\Delta\ w_q}{w'_q\,w_q}\right]$$

or

$$W_q\,W_{q-1} = \left[\frac{I_{q-1}-\Delta\ \Delta'}{-\ w'_q\ \Delta'}\Bigg|\frac{-\ \Delta\ w_q}{1-w'_q\ w_q}\right] \quad (41)$$

Taking determinants,

$$|\,W_q\,| = (\,1-w'_q\ D\ w_q\,)^{\frac{1}{2}}\ \ |\,I_{q-1}-\Delta\ \Delta'\,|^{\frac{1}{2}} \quad (42)$$

where

$$D = I_{p-q} + \Delta'\ (\,I_{q-1}-\Delta\ \Delta'\,)^{-1}\ \Delta\ = (I_{p-q}-\Delta'\ \Delta)^{-1} \quad (43)$$

Let D $= T\ T'$ where T is a lower triangular matrix. Transform from w_q to $m_q = [m_{q,q+1}\ldots\ldots m_{q,\ p}]'$ by

$$m'_q = w'_q\ T \quad (44)$$

The Jacobian of the transformation is

$$|T|^{-1} = |D|^{-\frac{1}{2}} = |\,I_{p-q}-\Delta'\ \Delta\,|^{\frac{1}{2}} = |\,I_{q-1}-\Delta\ \Delta'\,|^{\frac{1}{2}} \quad (45)$$

$$= |w_q\,|\ (1-w'_q\ D\ w_q\,)^{-\frac{1}{2}} = |w_q\,|\ (1-m_q'\ m_q\,)^{-\frac{1}{2}}$$

Hence

$$I = \int_{m'_q\ m_q\ <\ 1} (1-m'_q\ m_q\,)^{-\frac{1}{2}}\ \mathrm{d}m_q = \frac{\Pi^{(p-q+1)/2}}{\Gamma^{(p-q+1)/2}}$$

Proceeding in this manner for all q from p-1 to 2, for integrating out elements of w, except w_{1i} $(i = 2,\ldots\ldots, p)$, we obtain the joint distribution of Θ and w_{1i} $(i = 2,\ldots\ldots, p)$ as

$$\text{const. p}\ \ (\Theta)\sigma_1^{-n}\ \exp\left[-\frac{1}{2}\lambda^2\left(\frac{1}{\sigma_1^2}-1\right)\right]\frac{2}{w_{11}}\ d\Theta\ \prod_{i=2}^{p} d\,w_{1i} \quad (46)$$

where

$$W_{11} = +\ (\,1-w_{12}^2-\ldots\ldots-w_{1p}^2\,)^{\frac{1}{2}} \quad (47)$$

as W is orthogonal.

We are deriving the distribution of θ_i^2 and ϕ_k^2 under the null hypothesis H and so $w_{1i} = w_i$ for all i, where the w's are the coefficients in the assigned function ξ. We now transform from the w_i $(i$=2, ..., $p)$ to ϕ_k^2 $(k$=1, ..., p-1) by (24). We find on using (22)

$$\frac{\partial\ W_i^2}{\partial\ \phi_k^2} = \frac{\omega_i^2\ \theta_i^2}{\phi_k^2\ .\ (\phi_k^2-\theta_i^2)}$$

or

$$\frac{\partial W_i}{\partial\phi_k^2} = \frac{\omega_i\ \ \theta_i^2}{2\phi_k^2\ .\ (\phi_k^2-\theta_i^2)} \quad \text{for all } k \text{ and } i.$$

The Jacobian of transformation from w_i $(i=2, \ldots\ldots, p)$ to $\phi_k{}^2$ $(k=1, \ldots\ldots\ldots\ldots, p-1)$ comes out to be (after a little algebra) the absolute value of

$$\frac{\text{Const.} \prod_i (w_i\ \theta_i^2) \prod_{i\neq j} (\theta_i^2 - \theta_j^2) \prod_{h\neq k} (\phi_h^2 - \phi_k^2) \prod_k (\phi_k^2 - \theta_1^2)}{W_1\, \theta_2^1 \prod_k \phi_k^2 \prod_k \prod_i (\phi_k^2 - \theta_i^2) \prod_{i, i\neq 1} (\theta_1^2 - \theta_i^2)} \tag{48}$$

where $h, k = 1, \ldots\ldots\ldots\ldots\ldots, p-1$ and $i, j = 1, \ldots\ldots\ldots\ldots\ldots, p$. The joint distribution of Θ and ϕ_k^2 $(k=1, \ldots\ldots\ldots\ldots\ldots, p-1)$ therefore comes out to be

$$\frac{\text{Const.}\ p(\Theta) \exp\left[-\frac{1}{2}\lambda^2\left(\frac{1}{\sigma_1^2} - 1\right)\right] \prod_i (w_i\ \theta_i^2) \prod_{i\neq j} |\theta_i^2 - \theta_j^2| \prod_{h\neq k} (\phi_h^2 - \phi_k^2)}{\sigma_1^n\, w_1^2\, \theta_1^2 \prod_k \phi_k^2 \prod_k \prod_i |\phi_k^2 - \theta_i^2|} \cdot \frac{\prod_k |\phi_k^2 - \theta_1^2|\ d\Theta\ d\phi}{\prod_{i\neq 1} |\theta_1^2 - \theta_i^2|} \tag{49}$$

where $d\phi = \prod_k d\phi_k^2$. Substitute for all w_i from (24) in terms of θ_i^2 and ϕ_k^2 and use (22). After a little simplification, (49) reduces to

$$\frac{\text{Const.}\ p(\Theta) \exp\left[-\tfrac{1}{2}\lambda^2\left(\frac{1}{\sigma_1^2} - 1\right)\right] (\lambda^2)^{p/2} \prod_{h\neq k} |\phi_h^2 - \phi_k^2|\ d\Theta\ d\boldsymbol{\phi}}{\sigma_1^n \prod_i \theta_i \prod_{k,i} |\phi_k^2 - \theta_i^2|^{\frac{1}{2}}} \tag{50}$$

Since λ^2 is the s.s. of the sample observations on y_1, which is $N\ (0, \sigma_1)$, λ^2/σ_1^2 has a χ^2 distribution with n degrees of freedom. Hence the conditional distribution of Θ and ϕ when λ^2 is fixed is obtained by dividing (50) by the distribution of λ^2. On using (22), it comes out as

$$\frac{\text{Const.} \prod_k (\phi_k^2)^{\frac{n-p-2}{2}} \exp\left(-\tfrac{1}{2}\sum_i \theta_i^2\right) \prod_{i\neq j} |\theta_i^2 - \theta_j^2| \prod_{h\neq k} |\phi_h^2 - \phi_k^2|\ d\Theta d\phi}{\exp\left(-\tfrac{1}{2}\lambda^2\right) \prod_{k, i} |\phi_k^2 - \theta_i^2|^{\frac{1}{2}}\ d\lambda^2} \tag{51}$$

(One of the θ_i^2 and ϕ_k^2 must be replaced by its expression in terms of λ^2, using (22) but this is not explicitly carried out to preserve symmetry.)

Thus, this conditional distribution does not involve the nuisance parameter σ_1^2 as λ^2 is a sufficient statistic. This, therefore, forms the basis of an exact test for the goodness of fit of the assigned function.

A TEST FOR H

A measure of the total variation of the characters $x_1 \ldots\ldots\ldots\ldots, x_p$ is $\Sigma\, \theta_i^2$, the sum of the original roots. When ξ is eliminated, the residual roots are ϕ_k^2 $(k = 1, \ldots\ldots\ldots, p-1)$ and the residual variation is thus $\Sigma\, \phi_k^2$. The s.s. of ξ itself is λ^2. Hence, if H is true, we expect

$$U = \sum_1^p \theta_i^2 - \sum_1^{p-1} \phi_k^2 - \lambda^2 \qquad (52)$$

to be insignificant. Since U is a function of θ_i^2, ϕ_k^2 and λ^2, its distribution does not involve σ_1^2. In fact, the author has shown elsewhere that U is a χ^2 with $(p-1)$ $d.f.$ This, therefore, is an exact test for the goodness of fit of the assigned function ξ. It should be noted that the hypothesis H, of (25), comprises of two parts.

H_1 : All the roots of Σ except σ_1^2 (> 1) are unity and H_2 : the principal component corresponding to σ_1^2 is

$$\xi = h'\, x = w'\, z$$

A test for H_1 is given by Bartlett[7,8] or Lawley[9]. The more important part is, therefore, of testing H_2, which deals with the direction of ξ. An overall test of H, as the author[1] has shown is provided by

$$\nu = \sum_{i=1}^{p} \theta_i^2 - \lambda^2 \qquad (53)$$

which is a χ^2 with $n(p-1)$ $d.f.$ For H_2 alone, however, we should use U. It was shown by the author that U and $\nu - U$ are independently distributed.

In H_1, the common value of all the roots of Σ excluding σ_1^2, is assumed to be unity. However, if this is not so, it is σ^2 and if σ^2 is unknown, we can use $\dfrac{U/(p-1)}{(\nu-U)/(n-1)(p-1)}$ as an F-ratio with $(p-1)$ and $(n-1)$ $(p-1)$ $d.f.$ because, in that case U/σ^2 and $\nu - U/\sigma^2$ are independent chi-squared variables. This, therefore, provides an exact test for H_2.

A numerical example and the use of U for obtaining confidence intervals has been given in the earlier paper.

ACKNOWLEDGEMENTS

I am very much grateful to Dr. P.V. Krishna Iyer for encouragement and valuable advice. I am also indebted to Dr. C. G. Khatr i for suggesting the method of evaluation of I of (37)and the Jacobian.

REFERENCES

1. KSHIRSAGAR, A. M., *Biometrika,* **48** (1961), 397.
2. WILLIAMS, E. J., *ibid,* **43** (1952), 17.
3. OLKIN, "On distribution problems in Multivariate analysis" (Institute of Statistics Mimiograph Series No. 43) 1951, p. 1
4. MAULDON, J. G., *J. Roy. Statist. Soc.,* **B17** (1955), 79.
5. FISHER, R. A., *Ann. Engen.,* **9** (1939), 238.
6. HSU, P. L., *Ibid.,* **9** (1939), 250.
7. BARTLETT, M. S., *Brit. J. Psychol. (Stat. Sec),* **3** (1950), 77.
8. —— *ibid,* **4** (1951), 1.
9. LAWLEY, D. N., *Biometrika,* **43** (1956), 128.