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 $\xi$, 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 $\Sigma$. There is an orthogonal matrix

$$L = \begin{bmatrix} l_{ij} \end{bmatrix} = \begin{bmatrix} l_{(1)} & l_{(2)} & \cdots & l_{(p)} \end{bmatrix}'$$

such that

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

where $\sigma_i^2$ $(i = 1, \ldots, p)$ are the latent roots of $\Sigma$ 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

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

then

$$y_i = l_{(i)}'x$$

is called the $i$-th principal component. The transformation

$$y' = [y_1 \ldots y_p] = x' L'$$

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

Let

$$X = \begin{bmatrix} x_{ir} \end{bmatrix}_{n \times p} \quad i=1, \ldots, p \quad r=1, \ldots, n$$

be a sample of size $n$ from the distribution of $x$. The maximum likelihood estimate of $\Sigma$ is

$$\frac{1}{n} A,$$

where

$$A = [a_{ij}] = X'X$$

be a sample of size $n$ from the distribution of $x$. The maximum likelihood estimate of $\Sigma$ is

$$\frac{1}{n} A,$$

where

$$A = [a_{ij}] = X'X$$

be a sample of size $n$ from the distribution of $x$. The maximum likelihood estimate of $\Sigma$ is

$$\frac{1}{n} A,$$

where

$$A = [a_{ij}] = X'X$$
is the matrix of the sums of squares and products of the sample observations. There exists an orthogonal matrix $G$,

$$G = [g_{ij}] = [g(1) \ | \ g(2) \ | \ \ldots \ | \ g(p)]'$$

(7)

such that

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

(8)

where

$$\theta_1^2 > \theta_2^2 > \ldots \ldots > \theta_p^2$$

(9)

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

$$Z_i = g'() \times$$

(10)

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

$$Z' = [z_1, z_2, \ldots, z_p] = x'G'$$

(11)

is obviously

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

(12)

**A SINGLE NON-ISOTROPIC PRINCIPAL COMPONENT**

If all the roots $\sigma_i^2$ are equal, the variation of the $x$'s is isotropic. However, if all the roots except $\sigma_1^2$, 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 $\sigma_1^2$, to be unity. In the case of such a single non-isotropic principal component, $\Sigma$ is completely determined by $\sigma_1^2$ 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. 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 $\Sigma$ 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 $\theta_i^2$ and the residual roots $\phi_i^2$ of $A$ is obtained here,
The hypothetical function \( \xi = h'x \) can be expressed in terms of the sample principal components \( Z \) by using (11). Thus
\[
\xi = h'x = h'G'z = w'z = w_1z_1 + \ldots + w_pz_p
\]  
where,
\[
w = Gh
\]  
and we assume that for normalization
\[
w'w = 1
\]  
The conditional covariance in the sample between \( z_i \) and \( z_j \) when \( \xi \) is fixed is (from 12)
\[
\text{Cov.}(z_i, Z_j|\omega'z) = \frac{\text{cov.}(Z_i, Z_j) - \frac{\text{cov.}(Z_i, \omega'z) \text{cov.}(Z_j, \omega'z)}{V(\omega'z)}}{V(\omega'z)}
\]  
where \( \delta_{ij} \) is the Kronecker delta and \( i, j \) run from 1 to \( p \). The conditional covariance matrix of the \( z \)'s when \( \xi \) is fixed is, therefore,
\[
\frac{1}{n} \left[ \theta_i^2 \delta_{ij} - \frac{1}{\lambda^2} \omega_i \omega_j \theta_i^2 \theta_j^2 \right]
\]  
where
\[
\frac{1}{n} \lambda^2 = \frac{1}{n} \sum_{i=1}^{p} \omega_i^2 \theta_i^2 = \text{the sample variance of } \xi.
\]  
The latent roots of the above ‘conditional’ covariance matrix of the \( z \)'s when \( \xi \) is eliminated are called the residual roots of the \( z \)'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
\[
w_1z_1 + \ldots + w_pz_p = 0
\]  
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 \( \phi^2 \):
\[
\left| \begin{array}{cc}
\theta_i^2 & \delta_{ij} - \frac{1}{\lambda^2} \omega_i \omega_j \theta_i^2 \theta_j^2
\end{array} \right| = 0
\]  
This equation simplifies to
\[
1 - \sum_{i=1}^{p} \frac{\omega_i^2 \theta_i^4}{\lambda^2 (\theta_i^2 - \phi^2)} = 0
\]
and can also be written as
\[
\sum_{i=1}^{p} \frac{\omega_i^2 \theta_i^2}{\theta^2 - \phi^2} = 0 \quad \text{as} \quad \sum_{i=1}^{p} \omega_i^2 \theta_i^2 = \lambda^2 \tag{21}
\]

Let \( \phi^2 \) \((k = 1, \ldots, p-1)\) be the roots of this \((p-1)\)th degree equation in \( \phi^2 \). Collecting the coefficients of \((\phi^2)^{-1}, (\phi^2)^{-2}\) and the constant term, it can be easily shown that
\[
\sum_{k=1}^{p-1} \frac{\phi_k^2}{\theta^2 - \phi^2} = \frac{1}{\lambda^2} \sum_{i=1}^{p} \theta_i^2 \tag{22}
\]

and
\[
\sum_{i=1}^{p} \phi_k^2 = \sum_{i=1}^{p} \theta_i^2 - \frac{1}{\lambda^2} \sum_{i=1}^{p} \omega_i^2 \theta_i^2 \tag{23}
\]

From (21) Williams has proved that
\[
\omega_i^2 = \frac{\lambda}{\theta_i^2} \sum_{j=1}^{p-1} (\phi_j^2 - \theta_i^2) \quad (i = 1, \ldots, p) \tag{24}
\]

**Distribution of the Residual Roots**

Since \( \phi_k^2 \) \((k = 1, \ldots, p-1)\) 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 \( \theta_i^2 \), with \( n \) replaced by \( n-1 \) and \( p \) by \( p-1 \). The more important distribution is, however, of the original roots \( \theta_i^2 \) and the residual roots \( \phi_k^2 \) when \( \lambda^2 \) is held fixed. Fortunately it so happens that this latter distribution does not involve the nuisance parameter \( \sigma_i^2 \) and is, therefore, useful for deriving exact tests. This is so because \( \lambda^2 \) is a sufficient statistic for \( \sigma_i^2 \).

We obtain this distribution under the null hypothesis:

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

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. \((\sigma_i^2, 1, \ldots, 1)\) and hence the distribution of \( B \) is the Wishart distribution

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

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

\[
y = LGz = Wz \tag{28}
\]
where
\[ L G' = W = [W_{ij}] \] (29)

\( W \) is orthogonal because \( L \) & \( G \) are so, \( i.e. \)
\[ W W' = I_p, \] (30)

where \( I_p \) denotes the identity matrix of order \( p \). From (28), the true non-isotropic principal component is
\[ y_1 = w_{11} z_1 + \cdots + w_{1p} z_p \] (31)

The assigned function is \( \xi = w'z' \). Hence if \( H \) of (25) is true, the two vectors \( u' \) and \( [w_{11}, \ldots, w_{1p}] \) are the same. Also from (26) and (8) with (29) we have
\[ B = LAL' = LG' \text{ diag. } [\theta_1^2, \ldots, \theta_p^2] \text{ } GL' = W \Theta W' \] (32)

where
\[ \Theta = \text{ diag. } (\theta_1^2, \ldots, \theta_p^2) \] (33)

In the distribution of \( B \), transform from \( B \) to \( \Theta \) 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 \( W_{ij} \) (\( j > i; i, j = 1, \ldots, p \)). We shall denote by \( dW \), the product of the differential of these \( p(p-1)/2 \) elements. The transformation from \( B \) to \( \Theta \) and \( W \) is unique only if we further impose the condition \( w_{i,j} > 0 \) for all \( i \). The Jacobian of this transformation \( \frac{p(p-1)}{2} \) is the absolute value of
\[ \prod_{i < j} (\theta_i^2 - \theta_j^2) \left| \begin{array}{c}
\sum_{q=1}^{p-1} W_{i,q} \end{array} \right| \] (34)

where \( W_q \) is the matrix of the first \( q \) rows and \( q \) columns of \( W \). The joint distribution of \( \Theta \) and \( W \), therefore, comes out to be
\[ \text{const. } \exp \left[ -\frac{1}{2} \lambda^2 \left( \frac{1}{\sigma^2} - 1 \right) \right] \prod_{q=1}^{p-1} |W_q| d\Theta dW. \] (35)

where \( \sigma \) or \( \sigma^2 \)
\[ p(\Theta) = \prod_{i=1}^{p} \left\{ (\theta_i^2)^{(n-p-1)/2} \exp \left( -\frac{1}{2} \theta_i^2 \right) \right\} \prod_{i < j} (\theta_i^2 - \theta_j^2) \] (36)

We shall now integrate out all the elements of \( W \), except those in its first row, viz, \( w_{ii} \) (\( i = 2, \ldots, p \)). For this we need
\[ I = \int \frac{1}{|W_q|} dw_q \] (37)

where
\[ W'_q = [w_{q,q+1}, \ldots, w_{q,p}] \] (38)

and the range of integration is such that \( w w' = I_p \).

Let
\[ \Delta = \text{ the matrix of the first } q-1 \text{ rows and } p \text{ columns of } W, \] (39)
\[ C = \text{ The matrix of the first } q \text{ rows and } p \text{ columns of } W \]
\[ = \left[ \begin{array}{c|c}
| & \Delta \\
| & w'_q \\
\end{array} \right] \] (40)
Hence

\[ C C' = I_q = W_q W_q' + \left[ \begin{array}{c} \Delta \\ \Delta' \\ \hline w_q' \\ \Delta' \\ w_q' w_q \end{array} \right] \]

or

\[ W_q W_q^{-1} = \left[ \begin{array}{c} I_{q-1} - \Delta \\ -w_q' \Delta' \\ \hline 1 - w_q' w_q \end{array} \right] \]

(41)

Taking determinants,

\[ |W_q| = (1 - w_q' D w_q)^{1/2} \quad |I_q - \Delta \Delta'|^{1/2} \]

where

\[ D = I_{p-q} + \Delta' (I_{q-1} - \Delta \Delta')^{-1} \Delta = (I_{p-q} - \Delta' \Delta)^{-1} \]

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

\[ m_q' = w_q' T \]

The Jacobian of the transformation is

\[ |T|^{-1} = |D|^{-1/2} = |I_{p-q} - \Delta' \Delta|^{1/2} = |I_{q-1} - \Delta \Delta'|^{1/2} \]

(45)

\[ = |w_q| (1 - w_q' D w_q)^{-1} = |w_q| (1 - m_q' m_q)^{-1/2} \]

Hence

\[ I = \int \left( 1 - m_q' m_q \right)^{-1/2} \, dm_q = \frac{\Gamma\left(\frac{p-q+1}{2}\right)}{\Gamma\left(\frac{p-q+1}{2}\right)^2} \]

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

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

(46)

where

\[ W_{11} = + (1 - w_{12}^2 - \ldots - w_{1p}^2)^{1/2} \]

(47)

as \( W \) is orthogonal.

We are deriving the distribution of \( \theta_i^2 \) and \( \phi_k^2 \) under the null hypothesis \( H \) and so \( w_i = \theta_i \) for all \( i \), where the \( w_i \)'s are the coefficients in the assigned function \( \xi \). We now transform from the \( w_i (i = 2, \ldots, p) \) to \( \phi_k^2 (k = 1, \ldots, p-1) \) by (24). We find on using (22)

\[ \frac{\partial W_i^2}{\partial \theta_i^2} = \frac{\omega_i^2 \theta_i^2}{\phi_k^2} \quad \frac{\partial \phi_k^2}{\partial \theta_i^2} = \phi_k^2 \cdot (\phi_k^2 - \theta_i^2) \]

or

\[ \frac{\partial W_i^2}{\partial \theta_i^2} = \frac{\omega_i^2 \theta_i^2}{2\phi_k^2 \cdot (\phi_k^2 - \theta_i^2)} \]

for all \( k \) and \( i \).
The Jacobian of transformation from \( w_i \) \( (i=2, \ldots, p) \) to \( \phi_k \) \( (k=1, \ldots, p-1) \) comes out to be (after a little algebra) the absolute value of

\[
\text{Const.} \prod_i \left( w_i \left/ \theta_i^2 \right. \right) \prod_{i \neq j} \left( \theta_i^2 - \theta_j^2 \right) \prod_{k \neq k} \left( \phi_k^2 - \phi_k^2 \right) \prod_k \left( \phi_k^2 - \theta_i^2 \right) \\
W_1 \prod_k \phi_k^2 \prod_{k \neq k} \left( \phi_k^2 - \theta_k^2 \right) \prod_{i, i \neq 1} \left( \theta_1^2 - \theta_i^2 \right) 
\]

(48)

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

\[
\text{Const.} \ p \left( \Theta \right) \exp \left[ -\frac{1}{2} \lambda^2 \left( \frac{1}{\sigma_1^2} - 1 \right) \right] \prod_i \left( w_i \theta_i^2 \right) \prod_{i \neq j} \left( \theta_i^2 - \theta_j^2 \right) \prod_{h \neq k} \left( \phi_h^2 - \phi_k^2 \right) \\
\lambda^2 \phi_1 \theta_1^2 \prod_k \phi_k^2 \prod_{k \neq k} \phi_k^2 - \theta_i^2 \\
\prod_{i \neq 1} \phi_i^2 - \theta_i^2 
\]

(49)

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

\[
\text{Const.} \ p \left( \Theta \right) \exp \left[ -\frac{1}{2} \lambda^2 \left( \frac{1}{\sigma_1^2} - 1 \right) \right] \prod_{i \neq j} \left( \theta_i^2 - \theta_j^2 \right) \prod_{h \neq k} \phi_h^2 - \phi_k^2 \\
\lambda^2 \prod_i \theta_i^2 \prod_k \phi_k^2 - \theta_i^2 
\]

(50)

Since \( \lambda^2 \) is the s.s. of the sample observations on \( y_i \), which is \( N \ (0, \sigma_1) \), \( \lambda^2/\sigma_1^2 \) has a \( \chi^2 \) distribution with \( n \) degrees of freedom. Hence the conditional distribution of \( \Theta \) and \( \phi \) when \( \lambda^2 \) is fixed is obtained by dividing (50) by the distribution of \( \lambda^2 \). On using (22), it comes out as

\[
\text{Const.} \ \prod_k \left( \phi_k^2 \right)^{\frac{n-p-2}{2}} \exp \left( -\frac{1}{2} \sum_i \theta_i^2 \right) \prod_{i \neq j} \left( \theta_i^2 - \theta_j^2 \right) \prod_{h \neq k} \phi_h^2 - \phi_k^2 \\
\exp \left( -\frac{1}{2} \lambda^2 \right) \prod_{k, i} \phi_k^2 - \theta_i^2 \ d\lambda^2 
\]

(51)

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

Thus, this conditional distribution does not involve the nuisance parameter \( \sigma_1^2 \) as \( \lambda^2 \) 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 \( x_1, \ldots, x_p \) is \( \sum \theta_i^2 \), the sum of the original roots. When \( \xi \) is eliminated, the residual roots are \( \phi^2_k \) \((k = 1, \ldots, p-1) \) and the residual variation is thus \( \sum \phi^2_k \). The s.s. of \( \xi \) itself is \( \lambda^2 \). Hence, if \( H \) is true, we expect

\[
U = \sum_{i=1}^{p} \theta_i^2 - \sum_{i=1}^{p-1} \phi^2_i - \lambda^2
\]

(52)

to be insignificant. Since \( U \) is a function of \( \theta_i^2, \phi^2_k \) and \( \lambda^2 \), its distribution does not involve \( \sigma_1^2 \). In fact, the author has shown elsewhere that \( U \) is a \( \chi^2 \) with \((p-1)\) d.f. This, therefore, is an exact test for the goodness of fit of the assigned function \( \xi \). It should be noted that the hypothesis \( H \), of (25), comprises of two parts.

\( H_1 \): All the roots of \( \Sigma \) except \( \sigma_1^2 (> 1) \) are unity and \( H_2 \): the principal component corresponding to \( \sigma_1^2 \) is

\[
\xi = \lambda' x = w' z
\]

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

\[
\nu = \sum_{i=1}^{p} \theta_i^2 - \lambda^2
\]

(53)

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

In \( H_1 \), the common value of all the roots of \( \Sigma \) excluding \( \sigma_1^2 \), is assumed to be unity. However, if this is not so, it is \( \sigma^2 \) and if \( \sigma^2 \) is unknown, we can use \( \frac{U/(p-1)}{(\nu - U)/(n-1)(p-1)} \) as an \( F \)-ratio with \((p-1) \) and \((n-1)(p-1) \) d.f. because, in that case \( U/\sigma^2 \) 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

8. —— ibid., 4 (1951), 1.