ON THE DISTRIBUTION OF THE LATENT VECTORS FOR PRINCIPAL COMPONENT ANALYSIS

By T. Sugiyama

Tokyo College of Science

1. Summary. The distribution of the latent vectors of a sample covariance matrix was found by T. W. Anderson [1] (1951) when the population covariance matrix is a scalar matrix, $\Sigma = \sigma^2 I$. The asymptotic distribution for arbitrary Σ, also, was obtained by T. W. Anderson [3] in 1963. The elements of each latent vector are the coefficients of a principal component (with sum of squares of coefficients being unity). The object of the paper is to obtain the exact distribution of the latent vectors when the observations are obtained from bi-variate normal distribution.

2. The distribution of the latent vectors when the observations are from a bi-variate normal distribution. Let $X_1, \cdots, X_n$ be a sample from 2-dimensional distribution $N(\mu, \Sigma)$. Then the elements of $U = \sum_{\alpha=1}^{N} (X_\alpha - \bar{X})(X_\alpha - \bar{X})'$ have the pdf

$$g(\{u_{ij}\}) = \text{Const. } |U|^{(n-3)/2} \exp\left(-\tfrac{1}{2} \operatorname{tr} \Sigma^{-1} U\right) \tag{1}$$

for U positive definite and 0 otherwise, where $n = N - 1$, $\Sigma = \|\sigma_{ij}\|$ is positive definite, $\Sigma^{-1} = \|\sigma^{ij}\|$, $U = \|u_{ij}\|$ and

$$\text{Const} = |\Sigma^{-1}|^{n/2}/2^n \pi^{\frac{1}{2}} \Gamma(n/2) \Gamma((n-1)/2). \tag{2}$$

Let the characteristic roots of U be $l_1 > l_2 > 0$. Then U can be written as follows

$$\begin{pmatrix} u_{11} & u_{12} \\ u_{12} & u_{22} \end{pmatrix} = \begin{pmatrix} \cos\varphi & -\sin\varphi \\ \sin\varphi & \cos\varphi \end{pmatrix} \begin{pmatrix} l_1 & 0 \\ 0 & l_2 \end{pmatrix} \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} \tag{3}$$

where $0 \leqq \varphi < \pi$. Both the $(\cos\varphi \ \sin\varphi)'$ and $(-\sin\varphi \ \cos\varphi)'$ may be called the latent vectors of the principal component. Then the Jacobian of the transformation

$$|\partial(u_{11}, u_{12}, u_{22})/\partial(l_1, l_2, \varphi)| = l_1 - l_2. \tag{4}$$

From (1), (3) and (4) we have

$$h(l_1, l_2, \varphi) = \text{Const. } (l_1 l_2)^{(n-3)/2} \exp\left[-\tfrac{1}{2}(a_1 l_1 + a_2 l_2)\right](l_1 - l_2), \tag{5}$$

where

$$a_1 = \sigma^{11}\cos^2\varphi + 2\sigma^{12}\sin\varphi\cos\varphi + \sigma^{22}\sin^2\varphi,$$

$$a_2 = \sigma^{11}\sin^2\varphi - 2\sigma^{12}\sin\varphi\cos\varphi + \sigma^{22}\cos^2\varphi.$$

By integrating (5) with respect to l_2 over the range 0 to l_1, and using

Received 10 March 1965; revised 12 July 1965.

$$\int_0^p t^{z-1}e^{-t}\,dt = e^{-p}\sum_{i=0}^{\infty}\{p^{z+i}\Gamma(z)/\Gamma(z+1+i)\} \tag{6}$$

we have

$$\begin{aligned} h_1(l_1,\varphi) = \text{Const}\sum_{i=0}^{\infty}\{&\Gamma(n-1)/2)/\Gamma((n+1)/2+i) \\ &-\Gamma((n+1)/2)/\Gamma((n+3)/2+i)\}(a_2/2)^i \\ &\cdot l_1^{n-1+i}\exp(-\tfrac{1}{2}(a_1+a_2)l_1). \end{aligned} \tag{7}$$

By integrating (7) with respect to l_1 over the range 0 to ∞, and using the Gauss' hypergeometric series

$$\begin{aligned} F(\alpha,\beta,\gamma;z) \\ = \sum_{i=0}^{\infty}[\Gamma(\alpha+i)/\Gamma(\alpha)][\Gamma(\beta+i)/\Gamma(\beta)][\Gamma(\gamma)/\Gamma(\gamma+i)][z^i/i!] \end{aligned} \tag{8}$$

we have

$$\begin{aligned} h_2(\varphi) = \text{Const}\cdot(2/(a_1+a_2))^n\{&(\Gamma((n-1)/2)\Gamma(n)/\Gamma[(n+1)/2]) \\ &\cdot F(1,n,(n+1)/2;a_2/(a_1+a_2)) \\ -(\Gamma((n+1)/2)\Gamma(n)/\Gamma((n+3)/2))&F(1,n,(n+3)/2;a_2/(a_1+a_2))\} \end{aligned} \tag{9}$$

Since

$$\begin{aligned} (n+1)F(1,n,(n+1)/2;z) - (n-1)F(1,n,(n+3)/2;z) \\ = 2F(2,n,(n+3)/2;z) \end{aligned}$$

if we let the characteristic roots of Σ be $\lambda_1 \geqq \lambda_2 > 0$, from (2) and (9) we obtain

$$(1/\pi(n+1))(\lambda_1\lambda_2/\bar{\lambda}^2)^{n/2}F(2,n,(n+3)/2;x) \tag{10}$$

where

$$x = (\sigma_{11}\cos^2\varphi + 2\sigma_{12}\sin\varphi\cos\varphi + \sigma_{22}\sin^2\varphi)/(\sigma_{11}+\sigma_{22}),$$

$$\bar{\lambda} = (\lambda_1+\lambda_2)/2.$$

Thus we obtain the following:

THEOREM. *Let U have the Wishart distribution $W(2, n, \Sigma)$, and let $l_1 > l_2 > 0$ be the characteristic roots and $(\cos\varphi\ \sin\varphi)'$, $(-\sin\varphi\ \cos\varphi)'$ the corresponding vectors of U, then φ has the* pdf (10).

Acknowledgment. The author wishes to express sincere thanks to Dr. Y. Tsumura for his instructive suggestions throughout this study.

REFERENCES

[1] ANDERSON, T. W. (1951). The asymptotic distribution of certain characteristic roots and vectors. *Proc. Second Berkeley Symp. Math. Statist. Prob.* 103–130. Univ. of California Press.

[2] ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

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