ON THE DISTRIBUTION OF THE LARGEST LATENT ROOT OF THE COVARIANCE MATRIX

By T. Sugiyama[1]

Columbia University

1. Summary. The distribution of the largest latent root of the covariance matrix calculated from a sample from the normal multivariate population with population covariance matrix $\sigma^2 I$ are presented by author [10] in 1966. The purpose of this paper is to find the distribution of the largest latent root for arbitrary Σ.

2. Preliminary result. Let Λ be a diagonal matrix with diagonal elements $1 > \lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$, then the following result on a beta-function integral is shown in [10]

$$(2.1)\quad \int_{1>\lambda_1>\lambda_2>\cdots>\lambda_p>0} |\Lambda|^{t-(p+1)/2} |I - \Lambda|^{u-(p+1)/2} C_\kappa(\Lambda) \prod_{i<j} (\lambda_i - \lambda_j) \prod_{i=1}^p d\lambda_i$$
$$= (\Gamma_p(p/2)/\pi^{p^2/2})(\Gamma_p(t, \kappa)\Gamma_p(u)/\Gamma_p(t + u, \kappa))C_\kappa(I_p)$$

where I_p is the identity matrix $(p \times p)$, $C_\kappa(\Lambda)$ is zonal polynomials defined for each partition $\kappa(k_1, \cdots, k_p)$, $k_1 \geqq \cdots \geqq k_p \geqq 0$ of k into not more than p parts as certain symmetric polynomials of $\lambda_1, \lambda_2, \cdots, \lambda_p$, and

$$\Gamma_p(t, \kappa) = \pi^{p(p-1)/4} \prod_{i=1}^p \Gamma(t + k_i - (\mathrm{i} - 1)/2),$$
$$\Gamma_p(t) = \pi^{p(p-1)/4} \prod_{i=1}^p \Gamma(t - (i - 1)/2).$$

A detailed discussion of zonal polynomials may be found in A. T. James (6), and (7).

In (2.1), let $u = (p + 1)/2$, then the left hand side of (2.1) is written as follows

$$(2.2)\quad \int_{1>\lambda_1>\lambda_2>\cdots>\lambda_p>0} |\Lambda|^{t-(p+1)/2} C_\kappa(\Lambda) \prod_{i<j} (\lambda_i - \lambda_j) \prod_{i=1}^p d\lambda_i .$$

Now, let $l_i = \lambda_i/\lambda_1$, $i = 2, \cdots, p$, then we can write (2.2) as follows

$$(2.3)\quad \int_0^1 \lambda_1^{pt+k-1}\, d\lambda_1$$
$$\cdot \int_{1>l_2>\cdots>l_p>0} |\Lambda_l|^{t-(p+1)/2} C_\kappa({}^1\Lambda_l) \prod_{i=2}^p (1 - l_i) \prod_{i<j} (l_i - l_j) \prod_{i=2}^p dl_i$$

where

$$(2.4)\quad \Lambda_l = \begin{pmatrix} l_2 & & 0 \\ & \ddots & \\ 0 & & l_p \end{pmatrix}, \qquad {}^1\Lambda_l = \begin{pmatrix} 1 & 0 \\ 0 & \Lambda_l \end{pmatrix}.$$

Since $\int_0^1 \lambda_1^{pt+k-1}\, d\lambda_1 = 1/(pt + k)$, we have the following lemma.

Received 21 November 1966; revised 14 February 1967.

[1] On leave from Aoyama Gakuin University, Japan.

LEMMA. *Let Λ_l be a diagonal matrix with diagonal elements $1 > l_2 > \cdots > l_p > 0$ and let κ be a partition of k. Then*

$$\int_{1>l_2>\cdots>l_p>0} |\Lambda_l|^{t-(p+1)/2} C_\kappa({}^1\Lambda_l) \prod_{i=2}^p (1 - l_i) \prod_{i<j} (l_i - l_j) \prod_{i=2}^p dl_i$$
$$= (pt + k)(\Gamma_p(p/2)/\pi^{p^2/2})$$
$$\cdot(\Gamma_p(t, \kappa)\Gamma_p((p + 1)/2)/\Gamma_p(t + (p + 1)/2, \kappa))C_\kappa(I_p).$$

3. The distribution of the largest latent root. Suppose the sample consists of N observations from a normal p-variate population with covariance matrix Σ. After the usual orthogonal transformation to eliminate the sample means, we have a $(p \times n)$ matrix, $n = N - 1$,

$$X = \begin{pmatrix} x_{11} & \cdots & x_{1n} \\ \cdot & & \\ x_{p1} & \cdots & x_{pn} \end{pmatrix}.$$

Then, the joint distribution of the latent roots $\lambda_1, \lambda_2, \cdots, \lambda_p$ of the symmetric matrix XX' is written as follows:

$$(3.1)\quad \text{Const} \sum_{k=0}^\infty \sum_\kappa (C_\kappa(-\tfrac{1}{2}\Sigma^{-1})/k!\, C_\kappa(I_p)) \cdot|\Lambda|^{(n-p-1)/2} C_\kappa(\Lambda)\prod_{i<j} (\lambda_i - \lambda_j) \prod_{i=1}^p d\lambda_i$$

where

$$\Lambda = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_p \end{pmatrix}$$

with $\infty > \lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$, and

$$\text{Const} = \pi^{p^2/2} |\Sigma|^{-n/2}/2^{np/2}\Gamma_p(p/2)\Gamma_p(n/2).$$

The derivation of the joint distribution of the latent roots $\lambda_1, \lambda_2, \cdots, \lambda_p$ has been described in detail in A. T. James [6] and Y. Tumura [11].

Now, let

$$(3.2)\qquad l_i = \lambda_i/\lambda_1, \qquad i = 2, \cdots, p.$$

Integrating with respect to $l_2, \cdots, l_p$, from (3.1) we have the following formula:

$$(3.3)\quad \text{Const} \sum_{k=0}^\infty \sum_\kappa (C_\kappa(-\tfrac{1}{2}\Sigma^{-1})/k!C_\kappa(I_p))\lambda_1^{pn/2+k-1}\int_{1>l_2>\cdots>l_p>0}|\Lambda_l|^{n/2-(p+1)/2} C_\kappa({}^1\Lambda_l)\prod_{i=2}^p(1-l_i)\prod_{i<j}(l_i-l_j)\prod_{i=2}^p dl_i$$

where Λ_l and ${}^1\Lambda_l$ are the same matrices as (2.4). Therefore, using the lemma, we have the probability element of the largest latent root λ_1

$$(3.4)\quad (|\Sigma|^{-n/2}\Gamma_p((p + 1)/2)/2^{np/2}\Gamma_p(n/2)) \cdot\sum_{k=0}^\infty \sum_\kappa ((pn/2 + k)\Gamma_p(n/2, \kappa)/\Gamma_p((n + p + 1)/2, \kappa)k!)C_\kappa(-\tfrac{1}{2} \Sigma^{-1}) \cdot\lambda_1^{(pn+2k)/2-1}\, d\lambda_1 .$$

As usual, if a is such that the gamma functions are defined, then $(a)_\kappa = \Gamma_p(a, \kappa)/\Gamma_p(a)$. Therefore, (3.4) may be rewritten as

$$(3.5)\qquad \begin{aligned}&(|\Sigma|^{-n/2}\Gamma_p((p+1)/2)/2^{np/2}\Gamma_p((n+p+1)/2))\\ &\cdot\textstyle\sum_{k=0}^{\infty}\sum_\kappa ((pn/2+k)(n/2)_\kappa/((n+p+1)/2)_\kappa k!)C_\kappa(-\frac{1}{2}\Sigma^{-1})\\ &\cdot\lambda_1^{(pn+2k)/2-1}\,d\lambda_1 .\end{aligned}$$

Thus we have the following theorem:

THEOREM. *Let U has the Wishart distribution $W(p, n, \Sigma)$. Then the distribution of the largest latent root of the positive definite symmetric matrix U is given by* (3.5).

Let $\Sigma = I$ in (3.1), and using

$$(3.6)\qquad \exp(\operatorname{tr} Z) = \textstyle\sum_{k=0}^{\infty}\sum_\kappa C_\kappa(Z)/k!,$$

we rewrite (3.1) as follows:

$$(3.7)\qquad \text{Const}\,|\Lambda|^{(n-p-1)/2}\exp(-\tfrac{1}{2}\operatorname{tr}\Lambda)\textstyle\prod_{i<j}(\lambda_i-\lambda_j)\prod_{i=1}^{p}d\lambda_i .$$

Using (3.6) again, (3.7) can be rewritten as follows:

$$(3.8)\qquad \begin{aligned}&\text{Const}\cdot\lambda_1^{(n-p-1)/2}\exp(-\tfrac{1}{2}\lambda_1)\\ &\cdot\textstyle\sum_{k=0}^{\infty}\sum_\kappa |\Lambda_1|^{(n-p-1)/2}C_\kappa(-\frac{1}{2}\Lambda_1)\prod_{i<j}(\lambda_i-\lambda_j)\prod_{i=1}^{p}d\lambda_i\end{aligned}$$

where

$$\Lambda_1 = \begin{pmatrix}\lambda_2 & & 0\\ & \ddots & \\ 0 & & \lambda_p\end{pmatrix}$$

with $\lambda_1 > \lambda_2 > \cdots > \lambda_p > 0$.

Let $l_i = \lambda_i/\lambda_1$, $i = 2, \cdots, p$, in (3.8). Then, using (2.1), from (3.8) we have the probability element of λ_1 in the null case

$$(3.9)\qquad \begin{aligned}&(\pi^{p/2}\Gamma_{p-1}((p+2)/2)/2^{np/2}\Gamma(p/2)\Gamma_p(n/2))\\ &\cdot\textstyle\sum_{k=0}^{\infty}\sum_\kappa (\Gamma_{p-1}((n-1)/2,\kappa)/\Gamma_{p-1}((n+p+1)/2,\kappa)k!)C_\kappa(-\frac{1}{2}I_{p-1})\\ &\cdot\lambda_1^{(pn+2k)/2-1}\exp(-\lambda_1/2)\,d\lambda_1 .\end{aligned}$$

This formula has been obtained by the author [10]. And also, by letting $\Sigma = I$ in (3.4), we have the probability element of λ_1 in the null case

$$(3.10)\qquad \begin{aligned}&(\Gamma_p((p+1)/2)/2^{np/2}\Gamma_p(n/2))\\ &\cdot\textstyle\sum_{k=0}^{\infty}\sum_\kappa ((pn/2+k)\Gamma_p(n/2,\kappa)/\Gamma_p((n+p+1)/2,\kappa)k!)C_\kappa(-\frac{1}{2}I_p)\\ &\cdot\lambda_1^{(pn+2k)/2-1}\,d\lambda_1 .\end{aligned}$$

Therefore, the two expressions of (3.9) and (3.10) are equal. But, direct verification of this is left.

Acknowledgment. The author wishes to express very sincere thanks to Professor Theodore W. Anderson of Columbia University for his generous discussion and guidance during the preparation of the paper. Finally, I also wish to express my very sincere thanks to the referee for pointing some modifications in the earlier version of the paper.

REFERENCES

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

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

[3] CONSTANTINE, A. G. (1963). Some non-central distribution problems in multivariate analysis. *Ann. Math. Statist.* **34** 1270–1285.

[4] Herz, C. S. (1955). Bessel functions of matrix argument. *Ann. Math.* **61** 474–523.

[5] JAMES, A. T. (1954). Normal multivariate analysis and the orthogonal group. *Ann. Math. Statist.* **25** 40–75.

[6] JAMES, A. T. (1960). The distribution of the latent roots of the covariance matrix. *Ann. Math. Statist.* **31** 151–158.

[7] JAMES, A. T. (1961). Zonal polynomials of the real positive definite symmetric matrices. *Ann. Math.* **74** 456–469.

[8] JAMES, A. T. (1964). Distribution of matrix variates and latent roots derived from normal sample. *Ann. Math. Statist.* **35** 475–501

[9] ROY, S. N. (1957). *Some Aspects of Multivariate Analysis*. Wiley, New York.

[10] SUGIYAMA, T. (1966). On the distribution of the largest latent root and corresponding latent vector for principal component analysis. *Ann. Math. Statist.* **37** 995–1001.

[11] TUMURA, Y. (1965). The distribution of latent roots and vectors. *Tokyo Rika Univ. Math.* **1** 1–16.