

NON-CENTRAL DISTRIBUTIONS OF THE LARGEST LATENT ROOTS OF THREE MATRICES IN MULTIVARIATE ANALYSIS*

K. C. S. PILLAI AND T. SUGIYAMA

(Received Jan. 30, 1969; revised June 2, 1969)

1. Introduction and summary

The cdf of the largest latent root of the generalized B statistic in multivariate analysis in the central case is given by Pillai [8], [9], [11], and also useful formulae [10] approximating at the upper end the cdf of the largest latent root. Further, the above cdf has been obtained by Pillai as a series of incomplete beta functions [8], [12] and also independently by Sugiyama and Fukutomi [13]. Recently, Sugiyama [15] has obtained the cdf of the same, as power series. In the non-central MANOVA case, Hayakawa [5] and Khatri and Pillai [7] have obtained the density in a beta function series form. The purpose of this paper is to find simpler power series expressions than these obtained by the above authors for the non-central density function and the cdf of the largest latent root in the MANOVA situation, both in the generalized beta case and by usual transformation in the generalized F case. We will also obtain similar formulae for the non-central density functions of the largest roots in the case of canonical correlation and test of equality of two covariance matrices.

2. Preliminary results

In this section we state the following two lemmas which will be used in the sequel.

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

$$\int_{1 > l_2 > \dots > l_p > 0} |D_i|^{t-(p+1)/2} C_i(D_i) \prod_{i=2}^p (1-l_i) \prod_{i < j} (l_i - l_j) \prod_{i=2}^p dl_i \\ = (pt+k) \left(\frac{\Gamma_p(p/2)}{\pi^{p^2/2}} \right) \left(\frac{\Gamma_p(t, \kappa) \Gamma_p((p+1)/2)}{\Gamma_p(t+(m+1)/2, \kappa)} \right) C_r(I_p),$$

* This research was supported by the National Science Foundation, Grant No. GP-7663.

where the zonal polynomial $C_i(\mathbf{T})$, and $\Gamma_p(\alpha, \kappa)$ are defined in [1], [2], [3], [4].

LEMMA 2. Let $S(p \times p)$ be a symmetric matrix, and $C_i(S)$ and $C_s(S)$ be zonal polynomials of degree k and s respectively corresponding to the partition $\kappa = (k_1 \geq k_2 \geq \dots \geq k_p \geq 0)$ and $\sigma = (s_1 \geq s_2 \geq \dots \geq s_p \geq 0)$. Then

$$C_i(S)C_s(S) = \sum_j g_{i,s}^j C_j(S),$$

where $\delta = (\delta_1 \geq \delta_2 \geq \dots \geq \delta_p \geq 0)$, $\sum_{i=1}^p \delta_i = k + s$ and $g_{i,s}^j$ are constants.

Lemma 1 has been discussed by Sugiyama [14] and [15]. Tables of the coefficients $g_{i,s}^j$ of Lemma 2 are given by Hayakawa [5] and Khatri and Pillai [7] for various values of k and s .

3. Non-central distribution of the largest latent root in the MANOVA case

Let X be a $p \times n_1$ matrix variate ($p \leq n_1$) and Y a $p \times n_2$ matrix variate ($p \leq n_2$) and the columns be all independently normally distributed with covariance matrix Σ , $E(X) = M$ and $E(Y) = 0$. Then it is well known that $XX' = U_1$ is non-central Wishart with n_1 degrees of freedom and $YY' = U_2$ is central Wishart with n_2 degrees of freedom and the covariance matrix Σ , respectively. The generalized non-central statistics are defined as the latent roots of

$$L = (U_1 + U_2)^{-1/2} U_1 (U_1 + U_2)^{-1/2}.$$

Let $1 > l_1 > \dots > l_p > 0$ be the ordered latent roots of the matrix L , namely the roots of the following determinantal equation

$$|U_1 - l(U_1 + U_2)| = 0,$$

then the joint density function of l_1, \dots, l_p is given by Constantine [1] as

$$(1) \quad C(p, n_1, n_2) \exp(\text{tr} - \Omega) |L|^{(n_1 - p - 1)/2} |I_p - L|^{(n_2 - p - 1)/2} \\ \cdot \prod_{i < j} (l_i - l_j) \sum_{k=0}^{\infty} \sum_{\epsilon} \frac{((n_1 + n_2)/2)_{\epsilon}}{(n_1/2)_{\epsilon}} \frac{C_{\epsilon}(\Omega) C_{\epsilon}(L)}{C_{\epsilon}(I_p) k!},$$

where Ω is the non-centrality matrix, $1/2 \cdot M' \Sigma^{-1} M$, determinants $|L|$ and $|I_p - L|$ expressed as products of the latent roots of their matrices, and $C(p, n_1, n_2) = \pi^{p^2/2} \Gamma_p((n_1 + n_2)/2) / \Gamma_p(p/2) \Gamma_p(n_1/2) \Gamma_p(n_2/2)$. In this section, we obtain first the density and c.d.f. of l_1 .

First we use lemma 2 and write

$$|I_p - L|^{(n_2 - p - 1)/2} C_{\epsilon}(L) = \sum_{s=0}^{\infty} \sum_{\sigma} ((p + 1 - n_2)/2)_{\sigma} C_{\sigma}(L) C_{\epsilon}(L) / s!$$

$$= \sum_{s=0}^{\infty} \sum_{\sigma} \sum_{\delta} ((p+1-n_2)/2)_{\sigma} g_{\sigma, \delta}^s C_{\delta}(\mathbf{L})/s! ,$$

and from (1), we get

$$(2) \quad C(p, n_1, n_2) \exp(\text{tr} - \mathbf{Q}) | \mathbf{L} |^{(n_1-p-1)/2} \\ \cdot \prod_{i < j} (l_i - l_j) \sum_{k=0}^{\infty} \sum_{\kappa} \frac{((n_1+n_2)/2)_{\kappa}}{(n_1/2)_{\kappa}} \frac{C_{\kappa}(\mathbf{Q})}{C_{\kappa}(\mathbf{I}_p)k!} \\ \cdot \sum_{s=0}^{\infty} \sum_{\sigma, \delta} g_{\sigma, \delta}^s ((p+1-n_2)/2)_{\sigma} C_{\delta}(\mathbf{L})/s! .$$

Now consider the integral

$$(3) \quad \int_{l_1 > l_2 > \dots > l_p > 0} | \mathbf{L} |^{(n_1-p-1)/2} C_{\delta}(\mathbf{L}) \prod_{i < j} (l_i - l_j) \prod_{i=2}^p dl_i .$$

In (3) transform $q_i = l_i/l_1, i = 2, \dots, p$ and integrate with respect to q_2, \dots, q_p , we get

$$(4) \quad l_1^{pn_1/2+k+s-1} \int_{1 > q_2 > \dots > q_p > 0} | \mathbf{D}_q |^{n_1/2-(p+1)/2} C_{\delta}(\mathbf{D}_q) \prod_{i=2}^p (1-q_i) \cdot \prod_{i < j} (q_i - q_j) \prod_{i=2}^p dq_i \\ = l_1^{pn_1/2+k+s-1} (pn_1/2+k+s) \left(\frac{\Gamma_p(p/2)}{\pi^{p^2/2}} \right) \\ \cdot \frac{\Gamma_p(n_1/2, \delta) \Gamma_p((p+1)/2)}{\Gamma_p((n_1+p+1)/2, \delta)} C_{\delta}(\mathbf{I}_p) , \quad (\text{by Lemma 1}).$$

Hence from (2) and (4), using the result

$$(5) \quad \Gamma_p(\alpha, \delta) = \Gamma_p(\alpha)(\alpha)_{\delta} ,$$

we obtain the density of the largest latent root in the following form

$$(6) \quad C_1(p, n_1, n_2) \sum_{k=0}^{\infty} \sum_{\kappa} \frac{((n_1+n_2)/2)_{\kappa}}{(n_1/2)_{\kappa}} \frac{C_{\kappa}(\mathbf{Q})}{C_{\kappa}(\mathbf{I}_p)k!} \sum_{s=0}^{\infty} ((pn_1/2+k+s)/s!) \\ \cdot \sum_{\sigma, \delta} g_{\sigma, \delta}^s \frac{((p+1-n_2)/2)_{\sigma} (n_1/2)_{\delta}}{((n_1+p+1)/2)_{\delta}} C_{\delta}(\mathbf{I}_p) \cdot l_1^{pn_1/2+k+s-1}$$

where $1 > l_1 > 0$, and

$$C_1(p, n_1, n_2) = \frac{\Gamma_p((p+1)/2) \Gamma_p((n_1+n_2)/2)}{\Gamma_p(n_2/2) \Gamma_p((n_1+p+1)/2)} e^{\text{tr} - \mathbf{Q}} .$$

Further, the c.d.f. of the largest latent root is given by

$$(7) \quad P(l_1 < x) = C_1(p, n_1, n_2) \sum_{k=0}^{\infty} \sum_{\kappa} \frac{((n_1+n_2)/2)_{\kappa}}{(n_1/2)_{\kappa}} \frac{C_{\kappa}(\mathbf{Q})}{C_{\kappa}(\mathbf{I}_p)k!} \\ \cdot \sum_{s=0}^{\infty} \sum_{\sigma, \delta} g_{\sigma, \delta}^s \frac{((p+1-n_2)/2)_{\sigma} (n_1/2)_{\delta}}{((n_1+p+1)/2)_{\delta} s!} C_{\delta}(\mathbf{I}_p) x^{pn_1/2+k+s} .$$

Let $\Omega=0$ in (6). Then, since $g_{0,\sigma}^{\delta}=1$ and $\delta=\sigma$, we obtain the following formula given by Sugiyama [15]:

$$(8) \quad C_1(p, n_1, n_2)_2 F_1((-n_2+p+1)/2, n_1/2; (n_1+p+1)/2; x I_p) x^{pn_1/2}.$$

Since the roots l_1, \dots, l_p of the generalized beta case are related to the roots f_1, \dots, f_p of the generalized F case in the following manner:

$$l_1 = \frac{f_1}{1+f_1}, \dots, l_p = \frac{f_p}{1+f_p},$$

we obtain from (9), the c.d.f. of the largest latent root in the non-central generalized F case in the form

$$(9) \quad P(f_1 < y) = C_1(p, n_1, n_2) \sum_{k=0}^{\infty} \sum_x \frac{((n_1+n_2)/2)_x}{(n_1/2)_x} \frac{C_x(\Omega)}{C_x(I_p)k!} \\ \cdot \sum_{s=0}^{\infty} \sum_{\sigma, \delta} g_{\sigma, \delta}^s \frac{((p+1-n_2)/2)_s (n_1/2)_s}{s!((n_1+p+1)/2)_s} C_s(I_p) (y/(1+y))^{pn_1/2+k+s}.$$

THEOREM 1. *Let U_1 be a matrix having non-central Wishart distribution with n_1 degrees of freedom and matrix of non-centrality parameter Ω , and U_2 be a matrix having the Wishart distribution with n_2 degrees of freedom. Then the pdf and the cdf of the largest latent root l_1 of the equation $|U_1 - (U_1 + U_2)l| = 0$ are given by (6) and (7) respectively. And the cdf of the largest latent root f_1 of the equation $|U_1 - U_2 f| = 0$ is given by (9).*

4. Distribution of the largest latent root in the canonical correlation case

Let the columns of $\begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ be n independent $(p+p)$ -dimensional variates ($p \leq q$) with zero means and covariance matrix

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{pmatrix}.$$

Let R be the diagonal matrix with diagonal elements r_1, r_2, \dots, r_p , where $r_1^2, r_2^2, \dots, r_p^2$ are the latent roots of the equation $|X_1 X_2' (X_2 X_2')^{-1} X_2 X_1' - r^2 X_1 X_1'| = 0$ and also P be the diagonal matrix with diagonal elements $\rho_1, \rho_2, \dots, \rho_p$, where $\rho_1^2, \rho_2^2, \dots, \rho_p^2$ are the latent roots of the equation $|\Sigma_{12} \Sigma_{22}^{-1} \Sigma'_{12} - \rho^2 \Sigma_{11}| = 0$. Then, the distribution of $r_1^2, r_2^2, \dots, r_p^2$, is given by Constantine [1] in the following form:

$$(10) \quad C(n, p, q) |I_p - P^2|^{n/2} |R^2|^{(q-p-1)/2} |I_p - R^2|^{(n-p-q-1)/2}$$

$$\cdot \prod_{i < j} (r_i^2 - r_j^2) \sum_{k=0}^{\infty} \sum_{\kappa} \frac{(n/2)_{\kappa} (n/2)_{\kappa}}{(q/2)_{\kappa}} \frac{C_{\kappa}(\mathbf{R}^2) C_{\kappa}(\mathbf{P}^2)}{C_{\kappa}(\mathbf{I}_p) k!},$$

where

$$(11) \quad C(n, p, q) = \frac{\Gamma_p(n/2) \pi^{p^2/2}}{\Gamma_p(q/2) \Gamma_p((n-q)/2) \Gamma_p(p/2)}.$$

By the same method as before, namely using lemmas 1 and 2, we obtain the density of r_1^2 in the following form:

$$(12) \quad C_2(n, p, q) | \mathbf{I}_p - \mathbf{P}^2 |^{n/2} \sum_{k=0}^{\infty} \sum_{\kappa} \frac{(n/2)_{\kappa} (n/2)_{\kappa}}{(q/2)_{\kappa}} \frac{C_{\kappa}(\mathbf{P}^2)}{C_{\kappa}(\mathbf{I}_p) k!} \\ \cdot \sum_{s=0}^{\infty} ((pq/2 + k + s)/s!) \sum_{\alpha, \beta} g_{\alpha, \beta}^s ((p + q + 1 - n)/2)_{\alpha} \\ \cdot \frac{(q/2)_{\beta}}{((q + p + 1)/2)_{\beta}} C_{\beta}(\mathbf{I}_p) (r_1^2)^{pq/2 + k + s - 1},$$

where

$$C_2(n, p, q) = \frac{\Gamma_p((p+1)/2) \Gamma_p(n/2)}{\Gamma_p((n-q)/2) \Gamma_p((q+p+1)/2)}.$$

Integrating (12) from 0 to x with respect to r_1^2 , we have the following cdf of the largest latent root in the canonical correlation case

$$(13) \quad P(r_1^2 < x) = C_2(n, p, q) | \mathbf{I}_p - \mathbf{P}^2 |^{n/2} \sum_{k=0}^{\infty} \sum_{\kappa} \frac{(n/2)_{\kappa} (n/2)_{\kappa}}{(q/2)_{\kappa}} \frac{C_{\kappa}(\mathbf{P}^2)}{C_{\kappa}(\mathbf{I}_p) k!} \\ \cdot \sum_{s=0}^{\infty} g_{\alpha, \beta}^s ((p + q + 1 - n)/2)_{\alpha} \frac{(q/2)_{\beta}}{((q + p + 1)/2)_{\beta}} \\ \cdot \frac{C_{\beta}(\mathbf{I}_p)}{s!} \cdot x^{pq/2 + k + s}.$$

THEOREM 2. Let $\begin{pmatrix} \mathbf{X}_1 \\ \mathbf{X}_2 \end{pmatrix}$ be n independent normal $(p+q)$ -dimensional variates ($p \leq q$) with zero means and covariance matrix, Σ . Then the pdf and the cdf of the largest latent root r_1^2 of the equation $|\mathbf{X}_1 \mathbf{X}_1' - (\mathbf{X}_2 \mathbf{X}_2')^{-1} \mathbf{X}_2 \mathbf{X}_1' - r^2 \mathbf{X}_1 \mathbf{X}_1'| = 0$ is given by (12) and (13) respectively.

5. Non-central distribution of the largest latent root for test of equality of two covariance matrices

Let S_1 and S_2 be independently distributed as Wishart $W(n_1, p, \Sigma_1)$ and $W(n_2, p, \Sigma_2)$, respectively. Let the latent roots of $S_1 S_2^{-1}$ and $\Sigma_1 \Sigma_2^{-1}$ be g_1, \dots, g_p and $\delta'_1, \dots, \delta'_p$, respectively such that $\infty > g_1 > \dots > g_p > 0$ and $\infty > \delta'_1 \geq \dots \geq \delta'_p > 0$.

Let

$$\omega_i = \lambda g_i / (1 + \lambda g_i), \quad i = 1, \dots, p,$$

where λ is a given positive constant in the test of the null-hypothesis H that $\lambda \mathbf{A}' = \mathbf{I}$ and $\mathbf{A}' = \text{diag.}(\delta'_1, \dots, \delta'_p)$. Then the joint distribution of ω_i 's is given by Khatri [6] in the following form:

$$C(p, n_1, n_2) |\lambda \mathbf{A}'|^{-n_1/2} |\mathbf{W}|^{(n_1-p-1)/2} |\mathbf{I}_p - \mathbf{W}|^{(n_2-p-1)/2} \prod_{i < j} (\omega_i - \omega_j) \\ \cdot \sum_{k=0}^{\infty} \sum_x ((n_1 + n_2)/2)_x \frac{C_x(\mathbf{I}_p - (\lambda \mathbf{A}')^{-1}) C_x(\mathbf{W})}{C_x(\mathbf{I}_p) k!}$$

where $\mathbf{W} = \text{diag.}(\omega_1, \dots, \omega_p)$. Then, by the same method as before, we can obtain the density function of the largest latent root ω_1 in the following form:

$$(14) \quad C_3(p, n_1, n_2) |\lambda \mathbf{A}'|^{-n_1/2} \sum_{k=0}^{\infty} \sum_x ((n_1 + n_2)/2)_x \frac{C_x(\mathbf{I}_p - (\lambda \mathbf{A}')^{-1})}{C_x(\mathbf{I}_p) k!} \\ \cdot \sum_{s=0}^{\infty} \sum_{\sigma, \delta} (pn_1/2 + k + s) g_{x, \sigma}^s \frac{((p+1-n_2)/2)_\sigma (n_1/2)_\delta}{s! ((n_1 + p + 1)/2)_\delta} C_\delta(\mathbf{I}_p) \omega_1^{pn_1/2 + k + s - 1}$$

where $1 > \omega_1 > 0$, and $C_3(p, n_1, n_2) = \frac{\Gamma_p((p+1)/2) \Gamma_p((n_1+n_2)/2)}{\Gamma_p(n_2/2) \Gamma_p((n_1+p+1)/2)}$. Let $\lambda \mathbf{A}' = \mathbf{I}$, namely the central case. Then, since $g_{0, \sigma}^s = 1$ and $\sigma = \delta$, the cdf of ω_1 is

$$P(\omega_1 < x) = C_3(p, n_1, n_2) {}_2F_1((p+1-n_2)/2, n_1/2; (n_1+p+1)/2; \omega_1 \mathbf{I}_p) \omega_1^{pn_1/2}.$$

This is the same formula as given by Sugiyama [14]. We note that if $(p+1-n_2)/2$ is an integer, the summation of s will be terminated in finite number of terms. Further, $g_{x, \sigma}^s$'s are constants which do not exceed unity [7]. Again when $n_2 = p+1$ we get

$$P(\omega_1 < x) = |\lambda \mathbf{A}'|^{-n_1/2} {}_1F_0(n_1/2; x(\mathbf{I}_p - (\lambda \mathbf{A}')^{-1})) x^{pn_1/2}.$$

Let $x=1$, $a=n_1/2$, and $\mathbf{\Omega} = \mathbf{I}_p - (\lambda \mathbf{A}')^{-1}$. Then we have ${}_1F_0(a; \mathbf{\Omega}) = |\mathbf{I}_p - \mathbf{\Omega}|^{-a}$.

THEOREM 3. *Let \mathbf{S}_1 and \mathbf{S}_2 be matrices having Wishart distributions $W(n_1, p, \mathbf{\Sigma}_1)$ and $W(n_2, p, \mathbf{\Sigma}_2)$, respectively. Then the pdf of $\omega_1 = \lambda g_1 / (1 + \lambda g_1)$, where g_1 is the largest latent root of the equation*

$$|\mathbf{S}_1 - g \mathbf{S}_2| = 0,$$

is given by (14).

It may be pointed out that Khatri [6] has given the density of g_1 but (14) does not follow from his result by transformation.

PURDUE UNIVERSITY
AOYAMA GAKUIN UNIVERSITY*

* Now at The University of Adelaide

REFERENCES

- [1] A. G. Constantine, "Some non-central distribution problem in multivariate analysis," *Ann. Math. Statist.*, 34 (1963), 1270-1285.
- [2] A. T. James, "The distribution of the latent roots of the covariance matrix," *Ann. Math. Statist.*, 31 (1960), 151-158.
- [3] A. T. James, "Zonal polynomials of the real positive definite symmetric matrices," *Ann. Math.*, 74 (1961), 456-469.
- [4] A. T. James, "Distribution of matrix variates and latent roots derived from normal samples," *Ann. Math Statist.*, 35 (1964), 475-501.
- [5] T. Hayakawa, "On the distribution of the maximum latent root of a positive definite symmetric random matrix," *Ann. Inst. Statist. Math.*, 19 (1967), 1-17.
- [6] C. G. Khatri, "Some distribution problems connected with the characteristic root of $S_1 S_2^{-1}$," *Ann. Math. Statist.*, 38 (1967), 944-948.
- [7] C. G. Khatri and K. C. S. Pillai, "On the non-central distributions of two test criteria in multivariate analysis of variance," *Ann. Math. Statist.*, 39 (1968), 215-226.
- [8] K. C. S. Pillai, "On the distribution of the largest or the smallest root of a matrix in multivariate analysis," *Biometrika*, 43 (1956), 122-127.
- [9] K. C. S. Pillai, *Statistical Tables for Tests of Multivariate Hypotheses*, The Statistical Center, University of the Philippines, 1960.
- [10] K. C. S. Pillai, "On the distribution of the largest characteristic root of a matrix in multivariate analysis," *Biometrika*, 52 (1965), 405-411.
- [11] K. C. S. Pillai, "Upper percentage points of the largest root of a matrix in multivariate analysis," *Biometrika*, 54 (1967), 189-194.
- [12] K. C. S. Pillai, "On the distribution of the largest root of a matrix in multivariate analysis," *Ann. Math. Statist.*, 38 (1967), 616-617.
- [13] T. Sugiyama and K. Fukutomi, "On the distribution of the extreme characteristic roots of the matrices in multivariate analysis," Reports of statistical application on research, Union of Japanese Scientists and Engineers, 13, 1966.
- [14] T. Sugiyama, "On the distribution of the largest latent root of the covariance matrix," *Ann. Math. Statist.*, 38 (1967), 1148-1151.
- [15] T. Sugiyama, "Distribution of the largest latent root and the smallest latent root of the generalized B statistic and F statistic in multivariate analysis," *Ann. Math. Statist.*, 38 (1967), 1152-1159.