

ON THE DISTRIBUTION OF THE LARGEST ROOT OF A MATRIX IN MULTIVARIATE ANALYSIS¹

BY K. C. SREEDHARAN PILLAI

Purdue University

1. Introduction and summary. Distribution problems in multivariate analysis are often related to the joint distribution of the characteristic roots of a matrix derived from sample observations. This well-known Fisher-Girshick-Hsu-Mood-Roy distribution (under certain null hypotheses) of s non-null characteristic roots can be expressed in the form

$$(1.1) \quad f(\theta_1, \dots, \theta_s) = C(s, m, n) \prod_{i=1}^s \theta_i^m (1 - \theta_i)^n \prod_{i > j} (\theta_i - \theta_j),$$

$$0 < \theta_1 \leq \dots \leq \theta_s < 1,$$

where

$$(1.2) \quad C(s, m, n) = \pi^{\frac{1}{2}s^2} \Gamma_s(m + n + s + 1) /$$

$$\{\Gamma_s(\frac{1}{2}(2m + s + 1)) \Gamma_s(\frac{1}{2}(2n + s + 1)) \Gamma_s(\frac{1}{2}s)\},$$

$\Gamma_s(\cdot)$ is the multivariate gamma function defined in [2], and m and n are defined differently for various situations described in [4], [6]. Pillai [3], [5] has given the density function of the larger of two roots, i.e. when $s = 2$, as a hypergeometric function. In this paper, the result is extended to the general case giving the density function of the largest of s roots as a generalized hypergeometric function [1], [2]. The density function of the largest root of a sample covariance matrix derived by Sugiyama [7] can be obtained from the one derived here by considering the transformation $\frac{1}{2}\lambda_s = n\theta_s$ and making n tend to infinity.

2. The distribution of the largest root. Let us recall first the definition of the hypergeometric function of matrix argument [2]. If \mathbf{S} and \mathbf{T} are $(p \times p)$ symmetric matrices, then

$$(2.1) \quad {}_2F_1(a_1, a_2; b; \mathbf{S}) = [\Gamma_p(b) / \Gamma_p(a_1) \Gamma_p(b - a_1)]$$

$$\cdot \int_0^1 |\mathbf{I} - \mathbf{ST}|^{-a_2} |\mathbf{T}|^{a_1 - \frac{1}{2}(p+1)} |\mathbf{I} - \mathbf{T}|^{b - a_1 - \frac{1}{2}(p+1)} (d\mathbf{T}).$$

Now make the following transformation [3], [5], [7] in (1.1):

$$(2.2) \quad \theta_s = \theta_s, \quad l_i = \theta_i / \theta_s, \quad i = 1, \dots, s - 1.$$

We get

$$(2.3) \quad f_1(l_1, l_2, \dots, l_{s-1}, \theta_s) = C(s, m, n) \theta_s^{ms + (s-1)(1 + \frac{1}{2}s)}$$

$$\cdot (1 - \theta_s)^n \prod_{i=1}^{s-1} \{l_i^m (1 - l_i \theta_s)^n (1 - l_i)\} \prod_{i > j=1}^{s-1} (l_i - l_j).$$

Received 7 September 1966.

¹ This research was supported by the National Science Foundation, Grant No. GP-4600.

Now for the integration of (2.3) with respect to the l 's in the range $0 < l_1 \leq \dots \leq l_{s-1} < 1$, note that in the multivariate beta function form the $(s-1)$ -fold integral will reduce to (2.1) with $a_1 = m + \frac{1}{2}s$, $a_2 = -n$, $b = m + s + 1$, $\mathbf{S} = \theta_s \mathbf{I}_{s-1}$, $p = s - 1$, and the limits $\mathbf{0}$ and \mathbf{I}_{s-1} except that the result thus obtained should be multiplied by $\pi^{-\frac{1}{2}(s-1)^2} \Gamma_{s-1}((s-1)/2)$ since the integrand of (2.1) is equivalent to that of (2.3) only after an orthogonal transformation to diagonalize the matrix \mathbf{T} and integrating out the elements of this orthogonal matrix. Thus we get

$$(2.4) \quad f_2(\theta_s) = \{C(s, m, n) \Gamma_{s-1}(m + \frac{1}{2}s) \Gamma_{s-1}(1 + \frac{1}{2}s) \cdot \Gamma_{s-1}((s-1)/2) / \pi^{\frac{1}{2}(s-1)^2} \Gamma_{s-1}(m + s + 1)\} \cdot \theta_s^{m_s + (s-1)(1+\frac{1}{2}s)} (1 - \theta_s)^n {}_2F_1(m + \frac{1}{2}s, -n; m + s + 1; \theta_s \mathbf{I}_{s-1}).$$

When $s = 2$, (2.4) reduces to the result given by Pillai [3], [5]. Further, the density function of the smallest root, θ_1 , can be obtained from (2.4) by changing $1 - \theta_s$ to θ_1 and m to n [3], [5]. In addition, the density function of the largest root of a sample covariance matrix [7] can be obtained from (2.4) by the method given in the last section.

REFERENCES

- [1] CONSTANTINE, A. G. (1963). Some non-central distribution problems in multivariate analysis. *Ann. Math. Statist.* **34** 1270-1285.
- [2] JAMES, A. T. (1964). Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Statist.* **35** 475-501.
- [3] PILLAI, K. C. S. (1954). On some distribution problems in multivariate analysis. Mimeograph Series No. 88, Institute of Statistics, University of North Carolina.
- [4] PILLAI, K. C. S. (1955). Some new test criteria in multivariate analysis. *Ann. Math. Statist.* **26** 117-121.
- [5] PILLAI, K. C. S. (1956). On the distribution of the largest or the smallest root of a matrix in multivariate analysis. *Biometrika* **43** 122-127.
- [6] PILLAI, K. C. S. (1960). *Statistical Tables for Tests of Multivariate Hypotheses*. The Statistical Center, University of the Philippines, Manila.
- [7] SUGIYAMA, T. (1966). On the distribution of the largest latent root and the corresponding latent vector for principal component analysis. *Ann. Math. Statist.* **37** 995-1001.