THE DISTRIBUTION OF THE LATENT ROOTS OF THE COVARIANCE MATRIX

By Alan T. James[1]

Yale University

1. Summary. The distribution of the latent roots of the covariance matrix calculated from a sample from a normal multivariate population, was found by Fisher [3], Hsu [6] and Roy [10] for the special, but important case when the population covariance matrix is a scalar matrix, $\Sigma = \sigma^2 I$. By use of the representation theory of the linear group, we are able to obtain the general distribution for arbitrary Σ.

2. An integral expression for the distribution. Suppose the sample consists of N observations from a normal k-variate population with covariance matrix Σ. After the usual orthogonal transformation to eliminate the sample means, we have a $k \times n$ matrix X, $n = N - 1$,

$$X = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ x_{k1} & \cdots & x_{kn} \end{bmatrix}$$

distributed as

$$(1) \qquad (2\pi)^{-\frac{1}{2}nk} \operatorname{etr}\left(-\tfrac{1}{2}\Sigma^{-1}XX'\right) |\Sigma|^{-(n/2)} \prod_{i,j} dx_{ij}$$

where the symbol etr stands for the exponential of the trace of a square matrix. If $A = (a_{ij})$, then $\operatorname{etr}(A) = \exp(a_{11} + a_{22} + \cdots + a_{nn})$. Our object is to find the distribution of the latent roots $t_1, t_2, \cdots, t_k$ of the matrix XX', $(t_1 \geqq t_2 \geqq \cdots \geqq t_k)$.

By expressing X as a function of the t_i and other variables and integrating with respect to the latter, Fisher, Hsu and Roy showed that

$$(2) \qquad \int_{\text{other variables}} \prod_{i,j} dx_{ij} = c \left(\prod t_i\right)^{\frac{1}{2}(n-k-1)} \prod_{i<j} (t_i - t_j) \prod dt_i$$

where

$$(3) \qquad c = 2^{-2k} \prod_{i=1}^{k} A(n - i + 1) \prod_{i=1}^{k} A(i), \qquad A(i) = \frac{2\pi^{\frac{1}{2}i}}{\Gamma(\frac{1}{2}i)}$$

Received May 11, 1959.

[1] This work was initiated in the Division of Mathematical Statistics of the Commonwealth Scientific and Industrial Research Organization, Australia, and completed at Yale University.

In the special case when $\Sigma = \sigma^2 I_k$, the density in (1) is a function of the t_i alone, and thus they obtained the distribution as

$$(4) \quad \frac{c}{(2\pi)^{\frac{1}{2}nk}\sigma^{nk}} \exp\left(-\frac{t_1 + t_2 + \cdots + t_k}{2\sigma^2}\right) \left(\prod t_i\right)^{\frac{1}{2}(n-k-1)} \prod_{i<j} (t_i - t_j) \prod dt_i$$

For general Σ, the density

$$(5) \quad \operatorname{etr}\left(-\tfrac{1}{2}\Sigma^{-1}XX'\right)$$

is no longer a function of the t_i. This is equivalent to saying the (5) is not invariant under congruence transformation by the orthogonal group $O(k)$ of $k \times k$ orthogonal matrices, H,

$$XX' \to HXX'H' \qquad H \varepsilon O(k).$$

However, by an argument similar to the one given for the derivation of the noncentral Wishart distribution in James [7], one sees that the distribution of the t_i is not altered if the density function in the initial distribution (1) is symmetrized, by which we mean that the function (5) occurring in (1) is replaced by its average

$$(6) \quad \int_{O(k)} \operatorname{etr}\left(-\tfrac{1}{2}\Sigma^{-1}H'XX'H\right) d(H)$$

with respect to the invariant measure, $d(H)$, on the orthogonal group, $O(k)$. The invariant measure is normalized to make the total measure of $O(k)$ unity. The symmetrized function (6) is now a function of the t_i, and, putting $XX' = A$, we have the

THEOREM 1. *The general distribution of the latent roots $t_1, \cdots, t_k$ of the matrix A of sums of squares and products about the means, of a sample of $N = n + 1$ observations from a k-variate normal population with covariance matrix Σ is*

$$(7) \quad (2\pi)^{-\frac{1}{2}nk} |\Sigma|^{-(n/2)} c \int_{O(k)} \operatorname{etr}\left(-\tfrac{1}{2}\Sigma^{-1}HAH'\right) d(H) \cdot \prod_{i=1}^{k} t_i^{\frac{1}{2}(n-k-1)} \prod_{i<j}^{k} (t_i - t_j) \prod dt_i$$

where the constant c is given by (3) *and $d(H)$ is the invariant measure on the orthogonal group $O(k)$. The integral is a symmetric function of the latent roots of Σ and the latent roots $t_1, \cdots, t_k$ of A.*

Formulae for the distribution, similar to (7), are well known. The real problem, to which we now turn, is to evaluate the integral which is of the form

$$(8) \quad \int_{O(k)} \operatorname{etr}\left(BHAH'\right) d(H)$$

where A and B are $k \times k$ symmetric matrices. Our results, as summarized in Theorem 2, are an expansion for the integral in a series of zonal polynomials which are given, up to fourth order, in the appendix.

One possible method of finding the integral would be to expand the integrand in a power series and calculate the integrals of the resulting monomials in the elements of H by using the generating function given by James [8]. This method was used to check our results up to third order but beyond this it became too cumbersome.

The use of the theory of spherical and zonal functions is far more powerful and much more enlightening.

3. Spherical and zonal functions. The initial distribution (1) of a normal multivariate sample is highly symmetrical and this provides the clue to the evaluation of the integral (8). The distribution (1) is clearly invariant under the group $O(n)$ of orthogonal matrices H of order n acting upon X from the right

$$X \to XH \qquad H \, \varepsilon \, O(n). \tag{9}$$

It is also invariant under the linear group $G(k)$ of all real $k \times k$ nonsingular matrices L acting upon X from the left and upon Σ, simultaneously, by congruence transformation

$$X \to LX \tag{10}$$

$$\Sigma \to L\Sigma L' \qquad L \, \varepsilon \, G(k) \tag{11}$$

Transformations (10) and (11) imply that $A = XX'$ is transformed by congruence transformation cogrediently, and the information matrix Σ^{-1}, contragrediently,

$$A \to LAL' \tag{12}$$

$$\Sigma^{-1} \to L^{-1\prime}\Sigma^{-1}L^{-1} \tag{13}$$

The function tr $(\Sigma^{-1}XX')$ upon which the probability in (1) depends and the volume element

$$| \Sigma |^{-(n/2)} \prod_{i,j} dx_{ij}$$

are invariants under the transformations. $\mathrm{tr}(\Sigma^{-1}A)$ is the sum of the latent roots λ_i of the determinantal equation

$$| \lambda\Sigma - A | = 0, \tag{14}$$

which, apart from their order, are a complete set of invariants of the pair of positive definite matrices Σ and A under their simultaneous transformation in (11) and (12). This is proved by the fact that one can choose an $L \, \varepsilon \, G(k)$ such that

$$\begin{aligned} L\Sigma L' &= I_k \\ LAL' &= \Lambda = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_k \end{bmatrix}. \end{aligned} \tag{15}$$

Hence $\lambda_1, \cdots, \lambda_k$ are a complete set of invariants.

One might think that, having obtained a complete set of invariants, one had exhausted all the information supplied by the symmetry. However, this is not so. Within the spaces of functions defined on the matrices there is a deeper and much more extensive structure associated with the symmetry, which is revealed by the theory of group representations. This leads to a generalized Fourier or harmonic analysis of the (scalar valued) functions on the positive definite symmetric matrices. They are seen to be a system of generalized spherical functions, under congruence transformation by the linear group, in the sense of Berezin and Gelfand [1] and Godement [5]. Functions of the latent roots of the determinantal equation (14), being functions of two matrices invariant under their simultaneous transformation, play a very special role; they are the *zonal* functions.

We shall restrict our study to complex valued polynomial functions of the elements of a positive definite real symmetric $k \times k$ matrix A. When we come to the evaluation of the integral (8), we can expand the exponential in the integrand in a power series

$$\sum_{f=0}^{\infty} \frac{1}{f!} (\mathrm{tr}(BHAH'))^f \tag{16}$$

whose terms will be polynomials in the elements of A and B and our results will be applicable to these.

Corresponding to a congruence transformation

$$A \to LAL' \qquad L \varepsilon G(k) \tag{17}$$

of the space of positive definite real symmetric matrices A, one can define an induced linear transformation of the polynomials $\varphi(A)$ in the elements of A with real or complex coefficients,

$$\varphi(A) \to (L\varphi)(A) = \varphi(L^{-1}AL^{-1'}) \qquad L \varepsilon G(k). \tag{18}$$

(18) is a representation of $G(k)$ in the vector space of all polynomials $\varphi(A)$.

The first problem is to decompose the space of polynomials into its irreducible invariant subspaces and find out which irreducible representations of $G(k)$ are present. Since the transformation (18) maps any monomial into a homogeneous polynomial of the same degree, the vector space V_f of homogeneous polynomials of degree f is an invariant subspace of the space of all polynomials. Let us concentrate upon this.

From the results of Littlewood [9] and Foulkes, [4] it can be shown that the space V_f decomposes into the direct sum of m irreducible invariant subspaces $V_{f,p}$ where m is the number of elements in the set $P(f, k)$ of partitions $p = (f_1, f_2, \cdots)$, of f into not more than k parts. $f = f_1 + f_2 + \cdots$

$$V_f = \bigoplus_{p \varepsilon P(f,k)} V_{f,p}\,. \tag{19}$$

In each of the $V_{f,p}$ a separate irreducible representation, namely $\{2f_1, 2f_2, \cdots\}$, of $G(k)$ acts. The symbol $\{2f_1, 2f_2, \cdots\}$ denotes the irreducible representation of $G(k)$ corresponding to the Young symmetry diagram whose rows are of length $2f_1, 2f_2, \cdots$ respectively. See Boerner [2].

Now consider the vector space $V_{f,p}$ for any p, under the transformations (18) but with the matrices L restricted to be orthogonal,

$$\varphi(A) \to \varphi(H^{-1}AH^{-1'}) \qquad H \in O(k) \tag{20}$$

Since $O(k)$ is only a subgroup of $G(k)$, $V_{f,p}$ will not be irreducible under it in general, but will decompose into a direct sum of irreducible invariant subspaces $V_{f,p,i}$

$$V_{f,p} = V_{f,p,1} \oplus V_{f,p,2} \oplus \cdots \tag{21}$$

of which there will be a unique one, say $V_{f,p,1}$, which is one-dimensional and is generated by a polynomial $Z_p(A)$ which is invariant under orthogonal transformations (20). $Z_p(A)$ is called a *zonal* polynomial. Being invariant under (20), it must be a symmetric function of the latent roots $t_1, \cdots, t_k$ of A. Zonal polynomials of low order are given in the appendix.

4. Evaluation of the integral.

THEOREM 2. *If A and B are symmetric $k \times k$ matrices, $O(k)$ the group of orthogonal $k \times k$ matrices and $d(H)$ its invariant measure, then*

$$\int_{O(k)} \operatorname{etr}(BHAH')\, d(H) = \sum_{f=0}^{\infty} \frac{1}{f!1.3.5.\cdots.(2f-1)} \cdot \sum_{p \in P(f,k)} \frac{c(p)}{Z_p(I)} Z_p(B) Z_p(A) \tag{22}$$

where $P(f, k)$ is the set of partitions $p = (f_1, f_2, \cdots)$ of f into not more than k parts and $Z_p(A)$ is the zonal polynomial corresponding to the partition p and hence to the representation $\{2f_1, 2f_2, \cdots\}$ of the linear group $G(k)$.

$Z_p(A)$ is a symmetric polynomial in the latent roots $t_1, \cdots, t_k$ of A, which can thus be written as a polynomial in the sums of powers $s_j = \sum_{\nu=1}^{k} t_\nu^j = \operatorname{tr}(A^j)$, of these roots. $c(p)$ is a constant and $Z_p(I)$, the value of $Z_p(A)$ at $A = I$, is a polynomial of degree f in k.

PROOF: Consider the function $(\operatorname{tr}(BA))^f$. Since it is a homogeneous polynomial of degree f in both B and A it belongs to the direct product $V_f \times V_f$ of V_f with itself. Assume a basis has been chosen in each $V_{f,p,i}$ for all p and i. Then any element of $V_f \times V_f$ is a unique linear combination of terms each of which is the product of a basis function of B with a basis function of A. In particular

$$(\operatorname{tr}(BA))^f = \sum_{p \in P(f,k)} \sum_{q \in P(f,k)} c_{pq} Z_p(B) Z_q(A) + \text{other terms} \tag{23}$$

where each of the other terms has at least one of its two factors in a $V_{f,p,i}$ with $i > 1$, i.e. not a zonal function.

Let A and B be transformed by an arbitrary $L \in G(k)$, A cogrediently and B contragrediently.

$$\begin{aligned} A &\to LAL' \\ B &\to L^{-1'}BL^{-1} \end{aligned} \qquad L \in G(k) \tag{24}$$

Then $\text{tr}(BA) \to \text{tr}(L^{-1\prime}BL^{-1}LAL') = \text{tr}(L^{-1\prime}BAL') = \text{tr}(BA)$, i.e. $\text{tr}(BA)$, and hence $(\text{tr}(BA))^f$, are invariant.

In a direct product of an irreducible covariant space $V_{f,p}$ with an irreducible contravariant space $V_{f,q}$, there will be an invariant if and only if the representation in $V_{f,q}$ is the contragredient representation corresponding to the cogredient representation in $V_{f,p}$, i.e. if and only if $p = q$. Hence the expansion (23) of $(\text{tr}(BA))^f$ contains only those terms whose factors belong to subspaces $V_{f,p}$ with the same index p. Thus $c_{pq} = 0$ if $p \neq q$, and both factors of the "other terms" must belong to subspaces with the same p, i.e. we cannot have terms belonging to $V_{f,p,i} \times V_{f,q,j}$ with $p \neq q$.

As $(\text{tr}(BA))^f$ is invariant under orthogonal transformations (20), we must likewise have $i = j$ for any nonzero term. In summary, all terms belong to subspaces of the form $V_{f,p,i} \times V_{f,p,i}$, and if one factor of a term is a zonal function, so is the other.

Averaging over $O(k)$ annihilates all irreducible invariant subspaces other than those in which the identity representation acts. These it leaves unaltered. Thus the vectors in the $V_{f,p,i}$ for $i > 1$ are all mapped on zero but the zonal polynomials in the $V_{f,p,1}$ remain unchanged. Therefore all the "other terms" in the expansion (23) disappear under the averaging process and we have

$$\int_{O(k)} (\text{tr}(BHAH'))^f \, d(H) = \sum_{p \in P(f,k)} c_{pp} Z_p(B) Z_p(A). \tag{25}$$

When calculating zonal functions, one can find coefficients $c(p)$ such that

$$(\text{tr}A)^f = \frac{1}{1.3.5.\cdots.(2f-1)} \sum_{p \in P(f,k)} c(p) Z_p(A). \tag{26}$$

Substituting $B = I_k$ in (25) we have $(\text{tr}\, A)^f = \sum c_{pp} Z_p(I) Z_p(A)$. Therefore

$$c_{pp} = \frac{c(p)}{1.3.5.\cdots.(2f-1)Z_p(I)}$$

and

$$\int_{O(k)} (\text{tr}(BHAH'))^f \, d(H) = \frac{1}{1.3.\cdots.(2f-1)} \sum_{p \in P(f,k)} \frac{c(p)}{Z_p(I)} Z_p(B) Z_p(A)$$

from which the theorem follows.

5. The distribution of the roots. If $l_1, \cdots, l_k$ are the latent roots of the covariance matrix $n^{-1}A$, then $t_i = nl_i$.

The convergence of the series can probably be improved by writing

$$\text{etr}\,(-\tfrac{1}{2}\Sigma^{-1}HAH') = \text{etr}\left(-\frac{1}{2\sigma^2} A\right) \text{etr}\,(BHAH') \tag{27}$$

where $B = (1/2\sigma^2)I - \frac{1}{2}\Sigma^{-1}$, and the constant σ^2 is chosen to give optimum convergence.

THEOREM 3. *The distribution of the latent roots $l_1, \cdots, l_k$ of the covariance matrix calculated from a sample of $N = n + 1$ observations from a normal k-variate population with covariance matrix Σ is*

$$\left(\frac{n}{2\pi}\right)^{\frac{1}{2}nk} c \left(\prod_{i=1}^{k} \sigma_i^2\right)^{-\frac{1}{2}n} \exp\left(-\frac{n}{2\sigma^2}\sum_{i=1}^{k} l_i\right)\left(\prod_{i=1}^{k} l_i\right)^{\frac{1}{2}(n-k-1)} \prod_{i<j}(l_i - l_j)$$

$$(28) \qquad \cdot \sum_{f=0}^{\infty} \frac{n^f}{f!1.3.5.\cdots.(2f-1)} \sum_{p \epsilon P(f,k)} \frac{c(p)}{Z_p(I)}$$

$$\cdot Z_p(\beta_1, \cdots, \beta_k) Z_p(l_1, \cdots, l_k)\, dl_1 \cdots dl_k$$

where $Z_p(l_1, \cdots, l_k)$ are zonal polynomials,

$$\beta_i = \frac{1}{2}\left(\frac{1}{\sigma^2} - \frac{1}{\sigma_i^2}\right) \qquad i = l, \cdots, k,$$

σ^2 is an arbitrary constant chosen to optimize the convergence of the series, σ_i^2, $i = 1, \cdots, k$ are the latent roots of the population covariance matrix and the constant c is given in (3).

The zonal polynomials $Z_p(l_1, \cdots, l_k)$ are listed up to $f = 4$ in the appendix as functions of $s_j = \sum_{\nu=1}^{k} l_\nu^j$, together with the constants $c(p)$ and $Z_p(I)$. The symbol $P(f, k)$ is explained in theorem 2.

Roy [11] discusses significance tests and confidence intervals based on the roots distribution.

APPENDIX

Zonal polynomials of the representation of the linear group in the space of polynomials in the elements of a positive definite real symmetric matrix. (s_i is the sum of the ith powers of the latent roots of the matrix.)

Degree f	Partition p	Zonal polynomial Zp	Constants	
			$c(p)$	$Z_p(I)$
1	(1)	s_1	1	k
2	(2)	$s_1^2 + 2s_2$	1	$k(k+2)$
	(1^2)	$s_1^2 - s_2$	2	$k(k-1)$
3	(3)	$s_1^3 + 6s_1s_2 + 8s_3$	1	$k(k+2)(k+4)$
	(21)	$s_1^3 + s_1s_2 - 2s_3$	9	$k(k+2)(k-1)$
	(1^3)	$s_1^3 - 3s_1s_2 + 2s_3$	5	$k(k-1)(k-2)$
4	(4)	$s_1^4 + 12s_1^2s_2 + 12s_2^2 + 32s_1s_3 + 48s_4$	1	$k(k+2)(k+4)(k+6)$
	(31)	$s_1^4 + 5s_1^2s_2 - 2s_2^2 + 4s_1s_3 - 8s_4$	20	$k(k+2)(k+4)(k-1)$
	(2^2)	$s_1^4 + 2s_1^2s_2 + 7s_2^2 - 8s_1s_3 - 2s_4$	14	$k(k+2)(k-1)(k+1)$
	(21^2)	$s_1^4 - s_1^2s_2 - 2s_2^2 - 2s_1s_3 + 4s_4$	56	$k(k+2)(k-1)(k-2)$
	(1^4)	$s_1^4 - 6s_1^2s_2 + 3s_2^2 + 8s_1s_3 - 6s_4$	14	$k(k-1)(k-2)(k-3)$

REFERENCES

[1] F. A. BEREZIN AND I. M. GELFAND, "Some remarks on the theory of spherical functions on symmetric Riemannian manifolds," *Trudy Moskov. Mat. Obšč.*, Vol. 5 (1956), pp. 311–351. Amer. Math. Soc. translation to appear shortly.

[2] H. BOERNER, *Darstellungen von Gruppen*, Springer (1955), Berlin.

[3] R. A. FISHER, "The sampling distribution of some statistics obtained from nonlinear equations," *Ann. Eugenics*, Vol. 9 (1939), pp. 238–249.

[4] O. H. FOULKES, "Plethysms of S-functions," *Philos. Trans. Roy. Soc. London. Ser. A*, Vol. 246 (1954), pp. 555–591.

[5] R. GODEMENT, "A theory of spherical functions I," *Trans. Amer. Math. Soc.*, Vol. 73 (1952), pp. 496–556.

[6] P. L. HSU, "On the distribution of the roots of certain determinantal equations," *Ann. Eugenics*, Vol. 9 (1939), pp. 250–258.

[7] A. T. JAMES, "The noncentral Wishart distribution," *Proc. Roy. Soc. London. Ser. A*, Vol. 229 (1955), pp. 364–366.

[8] A. T. JAMES, "A generating function for averages over the orthogonal group," *Proc. Roy. Soc. London. Ser. A*, vol. 229 (1955), pp. 367–375.

[9] DUDLEY E. LITTLEWOOD, *The theory of group characters and matrix representations of groups*, second edition, Oxford, 1950.

[10] S. N. ROY, "p-statistics or some generalizations in the analysis of variance appropriate to multivariate problems," *Sankhya*, Vol. 4 (1939), pp. 381–396.

[11] S. N. ROY, *Some Aspects of Multivariate Analysis*, John Wiley and Sons, New York, 1958.