The Annals of Mathematical Statistics
1968, Vol. 39, No. 1, 227–232

ON THE DISTRIBUTION OF A MULTIPLE CORRELATION MATRIX: NON-CENTRAL MULTIVARIATE BETA DISTRIBUTIONS

By M. S. Srivastava

University of Toronto

Summary. Of several possible versions of multiple correlation matrix between two sets of variables **x** and **y** (see, e.g., Khatri, 1964), we derive using the techniques of A. T. James (zonal polynomials), the non-null distribution of one version when (i) one of the two sets of variables is fixed, i.e., multivariate analysis of variance and covariance case (MANOVA), and when (ii) both sets of variables are random variables, i.e., canonical correlations case. These distributions are non-central multivariate β-distributions in much the same way as the two cases of multiple correlation commonly known as the multiple correlation of the second and the first kind respectively.

1. Introduction. There are several ways of defining multiple correlation matrix between the two sets of variables $\mathbf{x}' = (x_1, x_2, \cdots, x_p)$ and $\mathbf{y}' = (y_1, \cdots, y_q)$ such that for $q = 1$ it is the square of the multiple correlation of y on the **x**-set, and for $q \geqq 1$, the roots of the matrix are the cannonical correlations between the **y**-set and the **x**-set. In the case of *MANOVA* (multivariate analysis of variance and covariance), Khatri (1964), gave several versions of it as follows:

Let $X = (x_{ir}):p \times n$ and $Y = (y_{jq}):q \times n$ be n independent observations on the two vectors **x** and **y**. Considering the variables of **x** as fixed, we have the usual multivariate linear regression analysis of **y** on **x**-set in the following table.

Source	df	Matrix of order $q \times q$ of s.s. and s.p.
Linear regression coefficients of **y**-set on **x**-set	p	$YX'(XX')^{-1}XY' = B$
Residual	$n - p$	$YY' - YX'(XX')^{-1}XY' = A - B$
Total due to **y**-set	n	$YY' = A$.

Let $A = TT'$ be any nonsingular factorization (including triangular and positive definite). Then the matrix L defined by

$$B = TLT'$$

is a version of multiple correlation matrix. Khatri (1964) considered yet another version of multiple correlation matrix (see Equation 3.4 of this paper) and

Received 5 May 1967; revised 24 July 1967.

derived its non-null distribution in the linear case. The object of this paper is to derive the non-null distribution when the matrix of non-centrality parameters is of full rank and also for the case when **x** is not fixed. These distributions are non-central multivariate β-distributions. The method of attack is that of A. T. James (1960, 1961a) and rests heavily on some integrals evaluated by him and Constantine (1963) in terms of zonal polynomials. The zonal polynomials of any symmetric and positive definite $m \times m$ matrix S are certain homogeneous symmetric polynomials in the characteristic roots of S.

2. Some results on integration. We shall write $0(m)$ for orthogonal group of $m \times m$ orthogonal matrices H, $d(H)$ for the invariant Haar measure on the orthogonal group normalized to make the volume of the group manifold unity and $Z_p(s)$ for the zonal polynomial (in the characteristic roots of $m \times m$ p.d. and symmetric matrix S) of degree m corresponding to the partition $p = (f_1, f_2, \cdots, f_m)$ of the integer f into not more than m parts, where $f_1 \geqq f_2 \geqq \cdots \geqq f_m \geqq 0, f_1 + \cdots + f_m = f$. $R(t)$ means the real part of t. We shall denote

$$(2.1) \quad \Gamma_m(t, p) = \pi^{\frac{1}{4}m(m-1)} \prod_{i=1}^{m} \Gamma(t + f_i - \tfrac{1}{2}(i-1)), \qquad R(t) > \tfrac{1}{2}(m-1).$$

In the present investigation, we need the following results on integration:

LEMMA 2.1.

$$\int_{0(m)} Z_p(AHBH')\, d(H) = Z_p(A)Z_p(B)/Z_p(I_m),$$

where A and B are $m \times m$ symmetric matrices, and $Z_p(I_m)$, the value of $Z_p(A)$ at $A = I$, is a polynomial of degree f in m.

For proof, refer to James (1960).

LEMMA 2.2. *Let W be a $k \times m$ matrix, $k \leqq m$, then*

$$\int_{0(m)} (\operatorname{tr} WH)^{2f}\, dH = \sum_{p \varepsilon P(f,k)} \chi_p(1) Z_p(WW')/Z_p(I_m),$$

where $P(f, k)$ is the set of partitions $p = (f_1, f_2, \cdots, f_k)$ of the positive integer f into not more than k parts, and $\chi_p(1)$ is the dimension of the representation $[2f_1, 2f_2, \cdots, 2f_k]$ of the symmetric group.

For proof, refer to James (1961b).

LEMMA 2.3. *Let R be a complex symmetric matrix whose real part is positive definite, and let T be an arbitrary complex symmetric matrix. Then*

$$\int_{S>0} (\operatorname{etr} - RS)(\det S)^{t-\frac{1}{2}(m+1)} Z_p(ST)\, dS = \Gamma_m(t, p)(\det R)^{-t} Z_p(TR^{-1}),$$

the integration being over the space of positive definite $m \times m$ matrices, and valid for all complex numbers t satisfying $R(t) > \frac{1}{2}(m-1)$. The constant $\Gamma_m(t, p)$ has been defined in (2.1).

For proof, refer to Constantine (1963).

3. Distribution of a multiple correlation matrix in Manova.

Let each column vector of the $p \times n$ matrix X be independently normally distributed with nonsingular $p \times p$ covariance matrix Σ, and let

$$(3.1) \qquad E(X \mid Y) = \beta Y,$$

where β is a $p \times q$, $q \leqq p$, regression matrix and Y is a $q \times n$ matrix of known constants, i.e., Y is fixed. Hence the pdf of X given Y fixed is given by

$$(3.2) \qquad (2\pi)^{-\frac{1}{2}pn}|\Sigma|^{-\frac{1}{2}n}[\text{etr} - \tfrac{1}{2}\Sigma^{-1}(X - \beta Y)(X - \beta Y)'].$$

It is known (see, e.g., Roy, 1958) that Y can be written as

$$Y = TM_1,$$

where T is a $q \times q$ triangular matrix with positive diagonal elements (hence unique) and M_1 is a $q \times n$ semi-orthogonal matrix; $M_1M_1' = I_q$, I_q is the $q \times q$ identity matrix. Completing the matrix M_1 by a $n - q \times n$ matrix M_2 such that $M' = (M_1', M_2')$ is an orthogonal matrix and making the transformation $XM' = (U_1, U_2)$, $U_1:p \times q$ and $U_2:p \times (n - q)$, we find that the joint pdf of U_1 and U_2 is given by

$$(3.3) \quad (2\pi)^{-\frac{1}{2}pn}|\Sigma|^{-n/2}[\text{etr} - \tfrac{1}{2}\Sigma^{-1}\{(U_1 - \beta T)(U_1 - \beta T)' + U_2U_2'\}].$$

A version of multiple correlation matrix is

$$(3.4) \qquad R = U_1'(U_1U_1' + U_2U_2')^{-1}U_1.$$

Note that R is invariant under the transformations $U_1 \to \Sigma^{-\frac{1}{2}}U_1$ and $U_2 \to \Sigma^{-\frac{1}{2}}U_2$, where $\Sigma^{-\frac{1}{2}}$ is any nonsingular factorization of Σ. Hence instead of (3.3), we may take the joint pdf of U_1 and U_2

$$(3.5) \quad (2\pi)^{-\frac{1}{2}pN}[\text{etr} - \tfrac{1}{2}\{(U_1 - \Sigma^{-\frac{1}{2}}\beta T)(U_1 - \Sigma^{-\frac{1}{2}}\beta T)' + U_2U_2'\}].$$

In (3.5), we first use Hsu's theorem (Anderson, 1958, Lemma 13.3.1, page 319) for deriving the distribution of $S_2 = U_2U_2'$ and then use the transformation

$$(3.6) \qquad S = S_2 + U_1U_1' \quad \text{and} \quad W = S^{-\frac{1}{2}}U_1.$$

The Jacobian of the transformation is $J(S_2, U_1 \to S, W) = J(S_2 \to S)J(U_1 \to W) = |S|^{\frac{1}{2}q}$. Hence, the joint pdf of S and W is

$$(3.7) \quad c\cdot(\text{etr} - \tfrac{1}{2}\mu\mu')|S|^{\frac{1}{2}(n-p-1)}|I_p - WW'|^{\frac{1}{2}(n-q-p-1)}(\text{etr} - \tfrac{1}{2}S)(\text{etr } S^{\frac{1}{2}}W\mu'),$$

where

$$(3.8) \qquad \begin{aligned} \mu &= \Sigma^{-\frac{1}{2}}\beta T, \\ c &= (2\pi)^{-\frac{1}{2}pn}\cdot\{\pi^{\frac{1}{2}p[n-q-\frac{1}{2}(p-1)]}/\textstyle\prod_{i=1}^{p}\Gamma[\tfrac{1}{2}(n - q + 1 - i)]\}. \end{aligned}$$

Since R is invariant under the transformation $W \to HW$, where H is a $p \times p$ orthogonal matrix, we find that the joint pdf of R and S is

$$(3.9) \quad c\cdot(\text{etr} - \tfrac{1}{2}\mu\mu')|S|^{\frac{1}{2}(n-p-1)}|R|^{\frac{1}{2}(p-q-1)}|I_q - R|^{\frac{1}{2}(n-p-q-1)} \cdot(\text{etr} - \tfrac{1}{2}S)\left(\textstyle\int_{0(p)} \text{etr } (W\mu'S^{\frac{1}{2}}H)\, dH\right),$$

where dH stands for the invariant measure on the group $0(p)$ of $p \times p$ orthogonal matrices H normalized so that $\int dH = 1$. Using Lemma (2.1), the joint pdf of R and S is given by

$$(3.10)\quad c\cdot(\text{etr} - \tfrac{1}{2}\mu\mu')|S|^{\frac{1}{2}(n-p-1)}(\text{etr} - \tfrac{1}{2}S)|R|^{\frac{1}{2}(p-q-1)}|I_q - R|^{\frac{1}{2}(n-p-q-1)} \cdot \textstyle\sum_{f=0}^{\infty}(1/(2f)\,!)\sum_{r\epsilon P(f,q)}\chi_r(1)Z_r(\mu' S\mu R)/Z_r(I_p),$$

since the matrix $W\mu' S^{\frac{1}{2}}$ is a $p \times p$ matrix of rank q.

Since $n > p + q - 1$, we find from Lemma 2.3, that the pdf of R is

$$(3.11)\quad c\cdot(2)^{\frac{1}{2}np}(\text{etr} - \tfrac{1}{2}\beta'\Sigma^{-1}\beta TT')|R|^{\frac{1}{2}(p-q-1)}|I_q - R|^{\frac{1}{2}(n-p-q-1)}, \cdot \textstyle\sum_{f=0}^{\infty}(1/(2f)!)\sum_{r\epsilon P(f,q)}[\chi_r(1)Z_r(2RT'\beta'\Sigma^{-1}\beta T)/Z_r(I_p)]\Gamma_q((n-p)/2, r),$$

$(n - p) > (q - 1)$, where c is defined in (3.8). This may be called the noncentral multivariate β-distribution.

4. Distribution of a multiple correlation matrix of the first kind (canonical correlations case). In the preceding section we derived the distribution of the correlation matrix R defined by (3.4) under the assumption that Y is fixed. In this section we drop this restriction and assume that $(\mathbf{x}, \mathbf{y})$ are jointly normally distributed with covariance matrix

$$(4.1)\quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{12}' & \Sigma_{22} \end{pmatrix}.$$

Without any loss of generality, we may assume that the means are zero. The sample may then be represented by the $(p + q) \times n$ partitioned matrix $\binom{X}{Y}$, where X and Y are $p \times n$ and $q \times n$ matrices, respectively. The sample covariance matrix is then

$$(4.2)\quad \begin{pmatrix} XX' & XY' \\ YX' & YY' \end{pmatrix}.$$

The joint density of X and Y can be written as the product of the conditional density of X given Y, and the marginal density of Y. The conditional density of X given Y is normal with mean matrix $\Sigma_{12}\Sigma_{22}^{-1}Y = \Delta Y$, and covariance matrix $\Sigma_{11.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{12}'$.

Proceeding exactly as in Section 3, and making an orthogonal transformation $XM' = (U_1, U_2)$, we find that the conditional (given Y) joint pdf of U_1 and U_2 is given by

$$(4.3)\quad (2\pi)^{-\frac{1}{2}pn}|\Sigma_{11\cdot 2}|^{-\frac{1}{2}}\,\text{etr} - \tfrac{1}{2}\Sigma_{11.2}^{-1}[(U_1 - \Delta T)(U_1 - \Delta T)' + U_2U_2'],$$

where $Y = TM_1$, $M' = (M_1', M_2')$, T is a $q \times q$ triangular matrix and M_1 is a semi-orthogonal matrix. We wish to find that distribution of the multiple correlation matrix R defined by

$$R = U_1'(U_1U_1' + U_2U_2')^{-1}U_1,$$

where $q \leqq p$ and Y is not fixed. The case where $q > p$ can be considered similarly. It is clear that R is invariant under the group of transformations G of $p \times p$

non-singular matrices q operating as $U_1 \to qU_1$ and $U_2 \to gU_2$. Hence, we may take the conditional density of U_1 and U_2 as

$$(4.4) \quad (2\pi)^{-pn/2} \text{ etr } -\tfrac{1}{2}[(U_1 - \Sigma_{11.2}^{-\frac{1}{2}}\Delta T)(U_1 - \Sigma_{11.2}^{-\frac{1}{2}}\Delta T)' + U_2U_2').$$

Making first the transformation $S_2 = U_2U_2'$ and then

$$(4.5) \qquad S = S_2 + U_1U_1', \qquad W = S^{-\frac{1}{2}}U_1,$$

we find that the conditional joint pdf of S and W is

$$(4.6) \quad c\cdot|S|^{\frac{1}{2}(n-p-1)}(\text{etr} - \tfrac{1}{2}S)|(I_p - WW')|^{\frac{1}{2}(n-q-p-1)} \cdot(\text{etr} - \tfrac{1}{2}\Delta'\Sigma_{11.2}^{-1}\Delta TT')(\text{etr } S^{\frac{1}{2}}WT'\Delta'\Sigma_{11.2}^{-\frac{1}{2}})$$

where c has been defined in (3.8). Following Section 3, we find that the conditional density of R given Y is

$$(4.7) \quad c\cdot 2^{\frac{1}{2}np}|R|^{\frac{1}{2}(p-q-1)}|(I_q - R)|^{\frac{1}{2}(n-p-q-1)}(\text{etr} - \tfrac{1}{2}\Delta'\Sigma_{11.2}^{-1}\Delta TT') \cdot\sum_{f=0}^{\infty}(1/(2f)\,!)\sum_{r\varepsilon P(f,q)}\chi_r(1)Z_r(2RT'\Delta'\Sigma_{11.2}^{-1}\Delta T) \cdot\Gamma_q((n-p)/2, r)/Z_r(I_p)$$

Since, R is invariant under the triangular group of transformations operating on T, we can write the joint density of R and T as

$$(4.8) \quad kc2^{\frac{1}{2}np}|R|^{\frac{1}{2}(p-q-1)}|I_q - R|^{\frac{1}{2}(n-p-q-1)}(\text{etr} - \tfrac{1}{2}[I + \Omega)TT')(2^p\textstyle\prod_{i=1}^{q}t_{ii}^{n-1}) \cdot\sum_{f=0}^{\infty}(1/(2f)\,!)\sum_{r\varepsilon P(t,q)}\chi_r(1)Z_r(2RT'\Omega T)\Gamma_p(n/2, r)/Z_r(I_p),$$

where

$$(4.9) \qquad k^{-1} = 2^{\frac{1}{2}nq}\pi^{q(q-1)/4}\prod_{i=1}^{q}\Gamma[\tfrac{1}{2}(n+1-i)],$$

and

$$(4.10) \qquad \Omega = \tilde{\Sigma}_{22}^{-\frac{1}{2}}\,\Sigma_{12}'\,\Sigma_{11.2}^{-1}\,\Sigma_{12}\,\tilde{\Sigma}_{22}^{-\frac{1}{2}\prime},$$

$\tilde{\Sigma}_{22}^{-\frac{1}{2}}$ is the triangular factorization $\Sigma_{22} = \tilde{\Sigma}_{22}^{-\frac{1}{2}}\,\tilde{\Sigma}_{22}^{-\frac{1}{2}\prime}$.

Letting $B = TT'$, we find that the joint pdf of R and B is

$$(4.11) \quad k^*|R|^{\frac{1}{2}(p-q-1)}|I_q - R|^{\frac{1}{2}(n-p-q-1)}(\text{etr} - \tfrac{1}{2}[I_q + \Omega]B) \cdot|B|^{(n-q-1)/2}\sum_{f=0}^{\infty}(1/(2f)\,!)\sum_{r\varepsilon P(f,q)}\chi_r(1)\Gamma_q((n-p)/2, r)/Z_r(I_p) \cdot\int_{0(q)}Z_r(2RHB^{\frac{1}{2}}\Omega B^{\frac{1}{2}}H)\,dH,$$

where k^* is given by (4.14), and where dH and $0(\mathrm{q})$ have been defined before; H is a $q \times q$ orthogonal matrix. Using Lemma (2.1), we find that the joint pdf of R and B is

$$(4.12) \quad k^*|R|^{\frac{1}{2}(p-q-1)}|I_q - R|^{\frac{1}{2}(n-p-q-1)}(\text{etr} - \tfrac{1}{2}[I + \Omega]B)|B|^{(n-q-1)/2}\sum_{f=0}^{\infty}1/(2f)\,! \cdot\sum_{r\varepsilon P(f,q)}[\chi_r(1)\Gamma_q((n-p)/2, r)/Z_r(I_p)Z_r(I_q)]Z_r(R)Z_r(2\Omega B).$$

Hence from Lemma 2.3, the pdf of R is given by

$$k^*|R|^{\frac{1}{2}(p-q-1)}|I_q - R|^{\frac{1}{2}(n-p-q-1)}[\det(\tfrac{1}{2}(I+\Omega))]^{-(n-p)/2}$$

$$\cdot \textstyle\sum_{f=0}^{\infty}(1/(2f)\,!)\sum_{r\varepsilon P(f,q)}[\chi_r(1)\Gamma_q((n-p)/2,\, r)\Gamma_q(n/2,\, r)Z_r(R)/ \tag{4.13}$$

$$Z_r(I_p)Z_r(I_q)]Z_r(4\Omega(I_q+\Omega)^{-1}))$$

where

$$k^{*-1} = 2^{\frac{1}{2}nq}\pi^{\frac{1}{2}p[q+\frac{1}{2}(p-1)]}(\textstyle\prod_{i=1}^{p}\Gamma[\tfrac{1}{2}(n-q+1-i)])\pi^{q(q-1)/4} \tag{4.14}$$

$$(\textstyle\prod_{i=1}^{q}\Gamma[\tfrac{1}{2}(n+1-i)]).$$

REFERENCES

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

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

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

JAMES, A. T. (1961a). The distribution of non-central means with known covariance. *Ann. Math. Statist.* **32** 874–882.

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

KHATRI, C. G. (1964). Distribution of the "Generalized" multiple correlation matrix in the dual case. *Ann. Math. Statist.* **35** 1801–1806.

ROY, S. N. (1958). *Some Aspects of Multivariate Analysis.* Wiley, New York.