The Annals of Mathematical Statistics
1970, Vol. 41, No. 3, 1091–1095

DISTRIBUTIONS CONNECTED WITH A MULTIVARIATE BETA STATISTIC

BY D. J. DE WAAL

University of Cape Town

1. Introduction. Let $A_r(p \times p)(r = 1, \cdots, q)$ and $B(p \times p)$ be independently distributed according to Wishart (Σ_1, m_r) and Wishart (Σ_2, n) respectively. Let

$$L_r = E^{-\frac{1}{2}} A_r (E^{-\frac{1}{2}})', \tag{1.1}$$

$E^{\frac{1}{2}}$ being a lower triangular matrix such that $E^{\frac{1}{2}}(E^{\frac{1}{2}})' = E = \sum_{r=1}^{q} A_r + B$. The purpose of this paper is to derive the joint density of $L_1, L_2, \cdots, L_q$, an asymptotic distribution for $\prod_{r=1}^{q} |L_r|$ and an asymptotic distribution for $|I - \sum_{r=1}^{q} L_r|$.

2. Preliminary results.

LEMMA 1. *If R and S are two positive definite symmetric matrices of order p, then* (*Constantine* (1963)),

$$\int_{E>0} \exp(-\tfrac{1}{2} \operatorname{tr} RE) |E|^{\alpha - \frac{1}{2}(p+1)} C_K(SE)\, dE = \Gamma_p(\alpha) |R|^{-\alpha} (\alpha)_K\, C_K(R^{-1}S). \tag{2.1}$$

LEMMA 2. *If R is a positive definite symmetric matrix of order p, then* (*Constantine* (1963))

$$\int_{0<Z<I} |Z|^{\alpha - \frac{1}{2}(p+1)} |I - Z|^{b - \frac{1}{2}(p+1)} C_K(RZ)\, dZ = (\Gamma_p(\alpha)\Gamma_p(b)/\Gamma_p(\alpha+b))((\alpha)_K/(\alpha+b)_K\, C_K(R). \tag{2.2}$$

LEMMA 3. *If R is a positive definite symmetric matrix of order p, then*

$$\int \cdots \int_{0<L_r<I,\, 0<\sum_{r=1}^{q} L_r < I} \prod_{r=1}^{q} |L_r|^{\alpha_r - \frac{1}{2}(p+1)} |I - \textstyle\sum_{r=1}^{q} L_r|^{b - \frac{1}{2}(p+1)} \cdot C_K(R \textstyle\sum_{r=1}^{q} L_r) \prod_{r=1}^{q} dL_r = (\Gamma_p(\alpha) \prod_{r=1}^{q} \Gamma_p(\alpha_r)/\Gamma_p(\alpha+b))((\alpha)_K/(\alpha+b)_K) C_K(R) \tag{2.3}$$

where $\alpha = \sum_{r=1}^{q} \alpha_r$.

PROOF. Let $\phi(L_1, \cdots, L_q) = \prod_{r=1}^{q} |L_r|^{a_r - \frac{1}{2}(p+1)} |I - \sum_{r=1}^{q} L_r|^{b - \frac{1}{2}(p+1)}$. It follows from Tan (1960) that the integral of ϕ w.r.t. $L_1, \cdots, L_q$ over the space $Z = \sum_{r=1}^{q} L_r$ is

$$\int_{Z=\sum_{r=1}^{q} L_r} \phi(L_1, \cdots, L_q) \prod_{r=1}^{q} dL_r = (\prod_{r=1}^{q} \Gamma_p(a_r)/\Gamma_p(a)) |Z|^{a - \frac{1}{2}(p+1)} |I - Z|^{b - \frac{1}{2}(p+1)}. \tag{2.4}$$

Hence (2.3) can be written as

$$\int_{0<Z<I} (\int_{\sum_{r=1}^{q} L_r = Z} \phi(L_1, \cdots, L_q) \prod_{r=1}^{q} dL_r) C_K(RZ)\, dZ = (\prod_{r=1}^{q} \Gamma_p(\tfrac{1}{2}m_r)/\Gamma_p(\tfrac{1}{2}m)) \int_{0<Z<I} |Z|^{\frac{1}{2}(m-p-1)} |I - Z|^{\frac{1}{2}(n-p-1)} C_K(RZ)\, dZ.$$

Received Febuary 28, 1969; revised November 21, 1969.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Mathematical Statistics.*
www.jstor.org

Using (2.2), the lemma is proved.

3. The joint density of $L_1, \cdots, L_q$.

THEOREM 1. *If L_r is defined as in* (1.1), *then the joint density of $L_1, \cdots, L_q$ is given by*

$$(3.1)\quad \begin{aligned}&(1/\Gamma_p(\tfrac{1}{2}n)\textstyle\prod_{r=1}^q \Gamma_p(\tfrac{1}{2}m_r))|\Sigma_1^{-1}\Sigma_2|^{\frac{1}{2}m}|2\Sigma_2|^{-\frac{1}{2}(m+n)}\\ &\cdot\textstyle\prod_{r=1}^q |L_r|^{\frac{1}{2}(m_r-p-1)}|I-\sum_{r=1}^q L_r|^{\frac{1}{2}(n-p-1)}\\ &\cdot\textstyle\int_{E>0}\exp(-\frac{1}{2}\operatorname{tr}\Sigma_2^{-1}E)|E|^{\frac{1}{2}(m+n-p-1)}\\ &\cdot{}_0F_0(\tfrac{1}{2}(I-\Sigma_1^{-1}\Sigma_2)\Sigma_2^{-1}E^{\frac{1}{2}}\textstyle\sum_r^q L_r(E^{\frac{1}{2}})'\,dE,\end{aligned}$$

$$L_r > 0,\ \textstyle\sum_{r=1}^q L_r < I,\ m = \sum_{r=1}^q m_r.$$

PROOF. The joint density of $L_r(r = 1, \cdots, q)$ and E is easily obtained from the joint density of $A_r(r = 1, \cdots, q)$ and B. The Jacobian being $J(A_1, \cdots, A_q, B \to L_1, \cdots, L_q, E) = |E|^{\frac{1}{2}q(p+1)}$. Integration with respect to E gives the desired result.

From (3.1) we can obtain the moments of $\prod_{r=1}^q |L_r|$ and $|I-\sum_{r=1}^q h_r|$ using (2.3) and (2.1) successively, i.e.

$$(3{\cdot}2)\quad \begin{aligned}&E(\textstyle\prod_{r=1}^q |L_r|^{h_r}) = (\Gamma_p(\frac{1}{2}(m+n)\prod_{r=1}^q \Gamma_p(\frac{1}{2}m_r+h)/(\Gamma_p(\frac{1}{2}(m+n)+h)\prod_{r=1}^q \Gamma_p(\frac{1}{2}m_r)))\\ &\cdot|\Sigma_1^{-1}\Sigma_2|^{\frac{1}{2}m}{}_2F_1(\tfrac{1}{2}m+h, \tfrac{1}{2}(m+n); \tfrac{1}{2}(m+n)+h; I-\Sigma_1^{-1}\Sigma_2), \qquad h = \textstyle\sum_{r=1}^q h_r,\end{aligned}$$

and

$$(3.3)\quad \begin{aligned}&E|I-\textstyle\sum_{r=1}^q L_r|^h = (\Gamma_p(\frac{1}{2}m+n))\Gamma_p(\frac{1}{2}m+h)/(\Gamma_p(\frac{1}{2}n)\Gamma_p(\frac{1}{2}(m+n)+h)))\\ &\cdot|\Sigma_1^{-1}\Sigma_2|^{\frac{1}{2}m}{}_2F_1(\tfrac{1}{2}m, \tfrac{1}{2}(m+n); \tfrac{1}{2}(m+n)+h; I-\Sigma_1^{-1}\Sigma_2).\end{aligned}$$

4. Asymptotic distributions. Asymptotic distributions for $\prod_{r=1}^q|L_r|$ and $|I-\sum_{r=1}^q L_r|$ will now be derived using the moments (3.2) and (3.3) respectively. The following approximation procedure will be used (Box (1949), Anderson (1958)) (see also de Waal (1968)): Consider a function g of the form

$$(4.1)\quad \begin{aligned}&g(t) = c(x^x/\textstyle\prod_{r=1}^q x_r^{x_r})^{-2it\rho p}\\ &\cdot\textstyle\prod_{j=1}^p(\prod_{r=1}^q \Gamma(\rho x_r(1-2it)+\beta_r+v_j)/\Gamma(\rho x(1-2it)+\beta+v_j+\gamma_j))\end{aligned}$$

where c is a constant such that $g(0) = 1$, ρ a chosen constant $\sum_{r=1}^q x_r = x$, $\beta_r = (1-\rho)x_r$, $\beta = (1-\rho)x$ and $i = (-1)^{\frac{1}{2}}$. Ignoring terms of order x^{-2} and higher, this function can be written as

$$(4.2)\quad g(t) = (1-2it)^{-\frac{1}{2}f}(1+\omega((1-2it)^{-1}-1))+O(x^{-2}),$$

assuming $O(x^{-2})$ small, where

$$(4.3)\quad f = \textstyle\sum_{j=1}^p((q-1)(1-2v_j)+\gamma_j),$$

$$(4.4)\quad \begin{aligned}&\omega = (\tfrac{1}{2}\rho)(\textstyle\sum_{j=1}^p((\sum_{r=1}^q(1/x_r)-1/x)\\ &\cdot(v_j^2-v_j+\tfrac{1}{6})-(\gamma_j^2+2v_j\gamma_j-\gamma_j)/x)-(1-\rho)f).\end{aligned}$$

The following theorems can now be proved:

THEOREM 2. *If* $L_1, \cdots, L_q$ *are distributed according to* (3.1) *then*

$$P(Q-\rho\textstyle\sum_{r=1}^{q} m_r \log|L_r| \leqq z) = P(\chi_f^2 \leqq z)+O(m^{-2}) \tag{4.5}$$

where

$$Q = \rho p(\textstyle\sum_{r=1}^{q} m_r \log(\frac{1}{2}m_r) - m\log(\frac{1}{2}m)), \tag{4.6}$$

$$\rho = 1+\tfrac{1}{2}(n-p-1)/m+2d/mp-(2p^2+3p-1)/12n, \tag{4.7}$$

$$d = |\Sigma_1^{-1}\Sigma_2|^{\frac{1}{2}m}\textstyle\sum_k\sum_K k(\frac{1}{2}m)_K C_K(I-\Sigma_1^{-1}\Sigma_2)/k! \tag{4.8}$$

and χ_f^2 *is a chi-square variate with*

$$f = \tfrac{1}{2}p(2n+(q-1)(p+1)) \tag{4.9}$$

degrees of freedom.

PROOF. Let

$$W = ((\tfrac{1}{2}m)^{\frac{1}{2}mp}/\textstyle\prod_{r=1}^{q}(\frac{1}{2}m_r)^{\frac{1}{2}m_r p})\prod_{r=1}^{q}|L_r|^{\frac{1}{2}m_r}$$

and $M = -2\log W$. From (3.2), the characteristic function of ρM can be written as

$$\begin{aligned}\phi(t) &= ((\tfrac{1}{2}m)^{\frac{1}{2}mp}/\textstyle\prod_{r=1}^{q}(\frac{1}{2}m_r)^{\frac{1}{2}m_r p})E\prod_{r=1}^{q}|L_r|^{-it\rho m_r} \\ &= |\Sigma_1^{-1}\Sigma_2|^{\frac{1}{2}m}\textstyle\sum_k\sum_K(\frac{1}{2}m)_K\, g_1(t)g_2(t)C_K(I-\Sigma_1^{-1}\Sigma_2)/k!\end{aligned} \tag{4.10}$$

where

$$\begin{aligned}g_1(t) = {}&((\tfrac{1}{2}m)^{\frac{1}{2}mp}/\textstyle\prod_{r=1}^{q}(\frac{1}{2}m_r)^{\frac{1}{2}pm_r})^{-2it\rho}\Gamma_p(\frac{1}{2}m) \\ &\cdot\textstyle\prod_{r=1}^{q}\Gamma_p(\frac{1}{2}m_r(1-2it\rho))/(\Gamma_p(\frac{1}{2}m(1-2it\rho))\prod_{r=1}^{q}\Gamma_p(\frac{1}{2}m_r))\end{aligned}$$

and

$$g_2(t) = \Gamma_p(\tfrac{1}{2}m(1-2it\rho), K)\Gamma_p(\tfrac{1}{2}(m+n), K)/(\Gamma_p(\tfrac{1}{2}m, K)\Gamma_p(\tfrac{1}{2}m(1-2it\rho)+\tfrac{1}{2}n, K)).$$

Use was made of the expression $(a)_K = \Gamma_p(a, K)/\Gamma_p(a)$ where

$$\Gamma_p(a, K) = \textstyle\prod^{\frac{1}{4}p(p-1)}\prod_{j=1}^{p}\Gamma(a+\frac{1}{2}(1-j)+K_j),$$

$K = (K = (K_1, \cdots, K_p)$ being a partition of k, and

$$\Gamma_p(a) = \Gamma_p(a, 0).$$

Using these expressions for the gamma functions, $g_1(t)$ and $g_2(t)$ can be written in the form (4.1).

Comparing $g_1(t)$ with (4.1), it follows that $x_r = \frac{1}{2}m_r$, $v_j = \frac{1}{2}(1-j)$ and $\gamma_j = 0$. Hence, from (4.2), $g_1(t)$ can be written as

$$g_1(t) = (1-2it)^{-\frac{1}{2}f_1}(1+\omega_1((1-2it)^{-1}-1))+O(m^{-2})$$

with $f_1 = \frac{1}{2}p(q-1)(p+1)$ and

$$\omega_1 = (-1/2\rho)((1-\rho)f_1-(p\textstyle\sum_{r=1}^{q}(1/m_r)-p/m)(2p^2+3p-1)/12).$$

Comparing $g_2(t)$ with (4.1) it follows that $q = 1$, $x = \frac{1}{2}m$, $v_j = \frac{1}{2}(1-j)+K_j$ and $\gamma_j = \frac{1}{2}n$. Therefore $g_2(t)$ can be written as

$$g_2(t) = (1-2it)^{-\frac{1}{2}f_2}(1+\omega_2((1-2it)^{-1}-1))+O(m^{-2})$$

with $f_2 = mp$ and

$$\omega_2 = (-1/2\rho)((1-\rho)f_2+n(\tfrac{1}{2}np-\tfrac{1}{2}p^2+2k-\tfrac{1}{2}p)/m).$$

Hence

$$(4.11)\qquad g_1(t)g_2(t) = (1-2it)^{-\frac{1}{2}f}(1-\omega_k((1-2it)^{-1}-1))+O(m^{-2})$$

where $f = f_1+f_2 = \frac{1}{2}p(2n+(q-1)(p+1))$ and

$$\begin{aligned}\omega_k &= \omega_1+\omega_2\\ (4.12)\qquad &= (-1/2\rho)((1-\rho)f+n(\tfrac{1}{2}np-\tfrac{1}{2}p^2+2k-\tfrac{1}{2}p)/m\\ &\quad -p(\textstyle\sum_{r=1}^{q}(1/m_r)-1/m)(2p^2+3p-1)/12).\end{aligned}$$

Substituting (4.11) in (4.10),

$$\begin{aligned}(4.13)\quad \phi(t) &= (1.2it)^{-\frac{1}{2}f}+((1-2it)^{-\frac{1}{2}(f+2)}-(1-2it)^{-\frac{1}{2}f})\\ &\quad\cdot|\Sigma_1{}^{-1}\Sigma_2|^{\frac{1}{2}m}\textstyle\sum_k\sum_K(\frac{1}{2}m)_K\,\omega_k\,C_K(1-\Sigma_1{}^{-1}\Sigma_2)/k!+O(m^{-2}).\end{aligned}$$

Choosing ρ such that $\omega_d = 0$, we have ρ as given in (4.7). Hence ω_k can therefore be written as

$$(4.14)\qquad \omega_k = n(d-k)m\rho.$$

Substituting (4.14) in (4.13) we note that the second term vanishes if d is chosen equal to (4.8). Hence the characteristic function of

$$(4.15)\qquad \rho M = Q-\rho\textstyle\sum_{r=1}^{q} m_r\log|L_r|,$$

Q being given in (4.6), becomes

$$(4.16)\qquad \phi(t) = (1-2it)^{-\frac{1}{2}f}+O(m^{-2}).$$

This is the characteristic function of a chi-square distribution with f degrees of freedom and hence the theorem follows.

THEOREM 3. *If $L_1, \cdots, L_q$ are distributed according to* (3.1) *then*

$$\begin{aligned}&P(-n\rho\log|I-\textstyle\sum_{r=1}^{q}L_r|\leqq z)\\ (4.17)\qquad &= |\Sigma_1{}^{-1}\Sigma_2|^{\frac{1}{2}m}\textstyle\sum_k\sum_K(\frac{1}{2}m)_K P(\chi^2_{f_k}\leqq z)C_K(I-\Sigma_1{}^{-1}\Sigma_2)/k!\\ &\quad+|\Sigma_1{}^{-1}\Sigma_2|^{\frac{1}{2}m}\textstyle\sum_k\sum_K(\frac{1}{2}m)_K(P(\chi^2_{f_k+2}\leqq z)-P(\chi^2_{f_k}\leqq z))\\ &\quad\cdot\omega_k\,C_K(I-\Sigma_1{}^{-1}\Sigma_2)/k!+O(n^{-2}),\end{aligned}$$

where

$$(4.18)\qquad \rho = 1+\tfrac{1}{2}(m-p-1)/n,$$

$$(4.19) \qquad \omega_k = (-1/\rho)(k(1-\rho+m/n)+(\sum_j K_j{}^2-\sum_j K_j j)/n)$$

and $\chi^2_{f_k}$ *a chi-square variate with*

$$(4.20) \qquad f_k = mp+2k$$

degrees of freedom.

PROOF. Let $U = |I-\sum_{r=1}^q L_r|^{\frac{1}{2}n}$ and $N = -2\log U$. From (3.3), the characteristic function of ρN can be written as

$$(4.21) \qquad \phi(t) = |\Sigma_1{}^{-1}\Sigma_2|^{\frac{1}{2}m}\sum_k\sum_K(\tfrac{1}{2}m)_K\, g_3(t)\, C_K(I-\Sigma_1{}^{-1}\Sigma_2)/k!$$

where

$$g_3(t) = \Gamma_p(\tfrac{1}{2}(m+n), K)\Gamma_p(\tfrac{1}{2}n(1-2it\rho))/(\Gamma_p(\tfrac{1}{2}n)\Gamma_p(\tfrac{1}{2}n(1-2it\rho)+\tfrac{1}{2}m, K)).$$

$g_3(t)$ can be written in the form (4.1) with $q = 1$, $x = \frac{1}{2}n$, $v_j = \frac{1}{2}(1-j)$ and $\gamma_j = \frac{1}{2}m+K_j$. Hence, using (4.2),

$$(4.22) \qquad \phi(t) = |\Sigma_1{}^{-1}\Sigma_j|^{\frac{1}{2}m}\sum_k\sum_K(\tfrac{1}{2}m)_K(1-2it)^{-\frac{1}{2}f_k} \cdot(1+\omega_k((1-2it)^{-1}-1))C_K(I-\Sigma_1{}^{-1}\Sigma_2)/k!+O(n^{-2})$$

with f_k given in (4.20) and

$$(4.23) \quad \omega_k = (-1/2\rho)((1-\rho)f_k+\tfrac{1}{2}m(m-p-1)+2mk/n+2(\sum_j K_j{}^2-\sum_j K_j j)/n).$$

Let $\omega_0 = 0$ and solve for ρ. Hence ρ becomes as given in (4.18) and ω_k can be written as (4.19). Taking the inverse of (4.22) (Anderson (1958), page 206) the theorem is proved.

REFERENCES

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

BOX, G.E.P. (1949). A general distribution theory for a class of likelihood criteria. *Biometrika* **36** 317–346.

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

DE WAAL, D. J. (1968). An asymptotic distribution for the determinant of a non-central B statistic in multivariate analysis. *S. Afr. Statist. J.* **2** 77–84.

TAN, W. Y. (1969). Note on the multivariate and the generalized multivariate Beta distributions. *J. Amer. Statist. Assoc.* **64** 230–241.