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 \((\Sigma_1, m_1)\) and Wishart \((\Sigma_2, n)\) respectively. Let

\[ L_r = E^{-\frac{1}{2}} A_r(E^{-\frac{1}{2}}) \', \]

\( E \)' being a lower triangular matrix such that \( E'(E')' = E = \sum_{r=1}^{q} A_r + B. \) 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)),

\[
\int_{E > 0} \exp \left( -\frac{1}{2} \text{tr } RE \right) |E|^{p - \frac{1}{2}(p + 1)} C_k(SE) dE = \Gamma_p(\alpha) |R|^{\alpha - \frac{1}{2}(p + 1)} C_k(R^{-1} S).
\]

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

\[
\int_{0 < Z < I} \left| Z \right|^{\alpha - \frac{1}{2}(p + 1)} |I - Z|^{b - \frac{1}{2}(p + 1)} C_k(RZ) dZ = \left( \Gamma_p(\alpha) \Gamma_p(b) / \Gamma_p(\alpha + b) \right) \left( (\alpha)_k / (\alpha + b)_k \right) C_k(R).
\]

**Lemma 3.** If \( R \) is a positive definite symmetric matrix of order \( p, \) then

\[
\prod_{r=1}^{q} \Gamma_p(\alpha) |L_r|^{\alpha - \frac{1}{2}(p + 1)} |I - \sum_{r=1}^{q} L_r|^{b - \frac{1}{2}(p + 1)} C_k(R) \prod_{r=1}^{q} dL_r
\]

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

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

\[
\int_{Z=\sum_{r=1}^{q} L_r} \phi(L_1, \cdots, L_q) \prod_{r=1}^{q} dL_r
\]

\( = \left( \prod_{r=1}^{q} \Gamma_p(\alpha_r) / \Gamma_p(\alpha_r + b) \right) \left( (\alpha)_k / (\alpha + b)_k \right) \left( (\alpha - \frac{1}{2}(p + 1)) / (\alpha + b)_k \right) \).

Hence (2.3) can be written as

\[
\int_{0 < Z < I} \left( \sum_{r=1}^{q} L_r \right) \phi(L_1, \cdots, L_q) \prod_{r=1}^{q} dL_r C_k(RZ) dZ
\]

\( = \left( \prod_{r=1}^{q} \Gamma_p(\alpha_r - \frac{1}{2}m_r) / \Gamma_p(\alpha_r + b) \right) \int_{0 < Z < I} \left| Z \right|^{\alpha - \frac{1}{2}(p + 1)} |I - Z|^{b - \frac{1}{2}(p + 1)} C_k(RZ) dZ. \)

Received February 28, 1969; revised November 21, 1969.
Using (2.2), the lemma is proved.

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

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

$$
(1/\Gamma_p(\frac{1}{2}m)) \prod_{r=1}^q \Gamma_p(\frac{1}{2}m_r) \left| \Sigma_1^{-1} \Sigma_2 \right| \left| \Sigma_2 \right|^{-\frac{1}{2}(m+n)} \cdot \prod_{r=1}^q |L_r|^{\frac{1}{2}(m_r+n_r-1)} \cdot \left| I - \sum_{r=1}^q L_r \right|^{\frac{1}{2}(n-n_r-1)} \\
\cdot \int_{E>0} \exp \left( -\frac{1}{2} \text{tr} \Sigma_2^{-1} E \right) |E|^{\frac{1}{2}(m+n-1)} \\
\cdot \left( \frac{1}{2} \right)^{(q-1)E} \int_{E>0} \exp \left( -\frac{1}{2} \text{tr} \Sigma_2^{-1} E \right) |E|^{\frac{1}{2}(m+n-1)} dE,
$$

(3.1)

Proof. The joint density of $L_r (r=1, \ldots, q)$ and $E$ is easily obtained from the joint density of $A_r (r=1, \ldots, q)$ and $B$. The Jacobian being $J(A_r, \ldots, A_q, B \rightarrow L_1, \ldots, L_q, E) = |E|^{\frac{1}{2}(q+1)}$. 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.

$$
E(\prod_{r=1}^q |L_r|) = (\Gamma_p(\frac{1}{2}(m+n)) \prod_{r=1}^q \Gamma_p(\frac{1}{2}m_r)) \left( \Gamma_p(\frac{1}{2}m + h) \prod_{r=1}^q \Gamma_p(\frac{1}{2}m_r) \right)
$$

(3.2)

$$
E \left| I - \sum_{r=1}^q L_r \right| = (\Gamma_p(\frac{1}{2}(m+n)) \Gamma_p(\frac{1}{2}m + h) \prod_{r=1}^q \Gamma_p(\frac{1}{2}m_r)) \left( \sum_{r=1}^q h_r \right)
$$

(3.3)

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

$$
g(t) = c(x^t \prod_{r=1}^q x_r x_r)^{-2ipt}
$$

(4.1)

where $c$ is a constant such that $g(0) = 1$, $\rho$ 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

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

(4.2)

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

$$
f = \sum_{r=1}^q ((q-1)(1-2v_j) + \gamma_j),
$$

(4.3)

$$
\omega = (\frac{1}{2}) \sum_{r=1}^q ((\sum_{j=1}^q (1/x_r) - 1/x) \cdot (v_j^2 - v_j + \frac{1}{2} - \gamma_j^2 + 2v_j \gamma_j - \gamma_j - (1-\rho)f).
$$

(4.4)
The following theorems can now be proved:

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

\[
P(Q - \rho \sum_{r=1}^{q} m_r \log |L_r| \leq z) = P(\chi_f^2 \leq z) + O(m^{-2})
\]

where

\[
Q = \rho \sum_{r=1}^{q} m_r \log(\frac{1}{2m_r}) - m \log(\frac{1}{2m}),
\]

\[
\rho = 1 + \frac{1}{2}(n - p - 1)/m + 2d/mp - (2p^2 + 3p - 1)/12n,
\]

\[
d = |\Sigma_1^{-1}\Sigma_2|^{1/2} \sum_k \sum_k k(\frac{1}{2}m)_k C_k (I - \Sigma_1^{-1}\Sigma_2)/k!
\]

and \( \chi_f^2 \) is a chi-square variate with

\[
f = \frac{1}{2}p(2n + (q - 1)(p + 1))
\]

degrees of freedom.

**Proof.** Let

\[
W = ((\frac{1}{2}m)^{\sum_{r=1}^{q} (\frac{1}{2}m_r)^{\sum_{r=1}^{q} m_r} m/r}) \prod_{r=1}^{q} |L_r|^{\frac{1}{2}m_r}
\]

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

\[
\phi(t) = ((\frac{1}{2}m)^{\sum_{r=1}^{q} (\frac{1}{2}m_r)^{\sum_{r=1}^{q} m_r}} m/r) E \prod_{r=1}^{q} |L_r|^{-it \rho m_r}
\]

\[
= |\Sigma_1^{-1}\Sigma_2|^{1/2} \sum_k \sum_k g_1(t) g_2(t) C_k (I - \Sigma_1^{-1}\Sigma_2)/k!
\]

where

\[
g_1(t) = ((\frac{1}{2}m)^{\sum_{r=1}^{q} (\frac{1}{2}m_r)^{\sum_{r=1}^{q} m_r} m/r})^{-2it \rho} \Gamma_{\rho}(\frac{1}{2}m)
\]

\[
\cdot \prod_{r=1}^{q} \Gamma_{\rho}(\frac{1}{2}m_r(1 - 2it \rho))/(\Gamma_{\rho}(\frac{1}{2}m(1 - 2it \rho))) \prod_{r=1}^{q} \Gamma_{\rho}(\frac{1}{2}m_r)
\]

and

\[
g_2(t) = \Gamma_{\rho}(\frac{1}{2}m(1 - 2it \rho), K) \Gamma_{\rho}(\frac{1}{2}(m + n), K)/(\Gamma_{\rho}(\frac{1}{2}m, K) \Gamma_{\rho}(\frac{1}{2}m(1 - 2it \rho) + \frac{1}{2}n, K)).
\]

Use was made of the expression \( (a)_K = \Gamma_{\rho}(a, K)/\Gamma_{\rho}(a) \) where

\[
\Gamma_{\rho}(a, K) = \prod_{j=1}^{K - 1} \Gamma(a + j(1 - j) + K_j),
\]

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

\[
\Gamma_{\rho}(a) = \Gamma_{\rho}(a, 0).
\]

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 \( x_r = \frac{1}{2}m_r, \gamma_j = \frac{1}{4}(1 - j) \) 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

\[
\omega_1 = (-1/2p)((1 - \rho)f_1 - (p \sum_{r=1}^{q} (1/m_r) - p/m)(2p^2 + 3p - 1)/12).
\]
Comparing $g_2(t)$ with (4.1) it follows that $q = 1$, $x = \frac{1}{4}m$, $v_j = \frac{1}{4}(1-j) + K_j$ and $\gamma_j = \frac{1}{4}n$. Therefore $g_2(t)$ can be written as

$$g_2(t) = (1-2it)^{-\frac{1}{4}f}(1+\omega_2((1-2it)^{-1} - 1)) + O(m^{-2})$$

with $f_2 = m$ and

$$\omega_2 = \frac{(-1/2\rho)((1-\rho)f_2 + n(m_1 m - 1/m)(2p^2 + 3p - 1))/12.$$ 

Hence

$$\Phi(t) = (1-2it)^{-\frac{1}{4}f}(1-\omega_k((1-2it)^{-1} - 1)) + O(m^{-2})$$

where $f = f_1 + f_2 = \frac{1}{4}p(2n + (q-1)(p+1))$ and

$$\omega_k = \frac{(-1/2\rho)((1-\rho)f + n(m_1 m - 1/m)(2p^2 + 3p - 1))/12.$$ 

Substituting (4.11) in (4.10),

$$\phi(t) = (1-2it)^{-\frac{1}{4}f} + ((1-2it)^{-\frac{1}{4}(f+2)} - (1-2it)^{-\frac{1}{4}f})$$

$$\cdot |\Sigma_1^{-1}\Sigma_2|^{1/2} \sum_k \omega_k C_k(I-\Sigma_1^{-1}\Sigma_2)/k! + O(m^{-2}).$$

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

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

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

$$\rho M = M - \rho \sum_{r=1}^q m_r \log |L_r|,$$

$Q$ being given in (4.6), becomes

$$\phi(t) = (1-2it)^{-\frac{1}{4}f} + O(m^{-2}).$$

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

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

$$P(-np \log |I - \sum_{r=1}^q L_r| \leq z)$$

$$= |\Sigma_1^{-1}\Sigma_2|^{1/2} \sum_k \omega_k C_k(I-\Sigma_1^{-1}\Sigma_2)/k!$$

$$+ |\Sigma_1^{-1}\Sigma_2|^{1/2} \sum_k \omega_k C_k(I-\Sigma_1^{-1}\Sigma_2)/k! + O(n^{-2}),$$

where

$$\rho = 1 + \frac{1}{2}(m-p-1)/n,$$
\[
(4.19) \quad \omega_k = (-1/\rho)(k(1-\rho + m/n) + (\sum_j K_j^2 - \sum_j K_{jj})/n)
\]

and \( \chi^2_{f_k} \) a chi-square variate with

\[
(4.20) \quad f_k = mp + 2k
\]

degrees of freedom.

**Proof.** Let \( U = |I - \sum_{j=1}^q L_j|^4 \) and \( N = -2 \log U \). From (3.3), the characteristic function of \( \rho N \) can be written as

\[
(4.21) \quad \phi(t) = |\Sigma_1^{-1}\Sigma_2|^{1/2} \sum_k \left( \frac{j}{m} \right) g_3(t) C_k(I - \Sigma_1^{-1}\Sigma_2)/k!
\]

where

\[
g_3(t) = \Gamma_{\rho}(\frac{1}{2}(m + n), K)\Gamma_{\rho}(\frac{1}{2}(1 - 2it\rho))/\Gamma_{\rho}(\frac{1}{2}(1 - 2it\rho) + \frac{j}{m}, K).
\]

\( g_3(t) \) 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),

\[
(4.22) \quad \phi(t) = |\Sigma_1^{-1}\Sigma_2|^{1/2} \sum_k \left( \frac{j}{m} \right) g_3(t) (1 - 2it)^{-k/2} C_k(I - \Sigma_1^{-1}\Sigma_2)/k! + O(n^{-2})
\]

with \( f_k \) given in (4.20) and

\[
(4.23) \quad \omega_k = (-1/2\rho)((1-\rho)f_k + \frac{1}{2}m(m-p-1) + 2mk/n + 2(\sum_j K_j^2 - \sum_j K_{jj})/n).
\]

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

**References**


