The Annals of Mathematical Statistics
1970, Vol. 41, No. 5, 1557–1562

ON THE NULL DISTRIBUTION OF THE SUM OF THE ROOTS OF A MULTIVARIATE BETA DISTRIBUTION[1]

By A. W. Davis

C.S.I.R.O., Adelaide and University of North Carolina at Chapel Hill

1. Introduction. The distribution of Pillai's V statistic [8] is shown to satisfy a homogeneous linear differential equation (d.e.) of Fuchsian type, which is related by a simple transformation to the author's d.e. for Hotelling's generalized T_0^2 [3]. This transformation implies certain relationships between the moments and asymptotic expansions of the two distributions. The adequacy of some approximations to V is checked by using the d.e. to tabulate some accurate percentage points.

2. Systems of differential equations. Let $\mathbf{S}_1$, $\mathbf{S}_2$ denote $m \times m$ matrices with independent null Wishart distributions on n_1, n_2 degrees of freedom respectively ($n_1, n_2 \geqq m$), estimating the same covariance matrix. The joint distribution of the latent roots $\theta_1, \cdots, \theta_m$ of $\mathbf{S}_1(\mathbf{S}_1+\mathbf{S}_2)^{-1}$ is well known to be

$$(2.1)\quad \phi_{n_1,n_2}(\theta_1, \cdots, \theta_m) = C(m; n_1, n_2)(\prod_{i=1}^m \theta_i)^{\frac{1}{2}(n_1-m-1)}(\prod_{i=1}^m (1-\theta_i))^{\frac{1}{2}(n_2-m-1)} \cdot \prod_{i<j}(\theta_i-\theta_j), \qquad (0 < \theta_m < \cdots < \theta_1 < 1),$$

where

$$(2.2)\qquad C(m; n_1, n_2) = \pi^{\frac{1}{2}m^2}\Gamma_m(\tfrac{1}{2}(n_1+n_2))/\Gamma_m(\tfrac{1}{2}m)\Gamma_m(\tfrac{1}{2}n_1)\Gamma_m(\tfrac{1}{2}n_2).$$

Pillai's V statistic is defined by

$$(2.3)\qquad V = \sum_{i=1}^m \theta_i$$

and Hotelling's generalized T_0^2 statistic by

$$(2.4)\qquad T = \sum_{i=1}^m \theta_i/(1-\theta_i) = T_0^2/n_2.$$

Following the method of [3], Section 2, we introduce the Laplace transforms (Lt's)

$$(2.5)\quad L_r(s) = \int_{R_m} \exp(-s\sum\theta_i)\, \phi_{n_1,n_2}(\theta_1, \cdots, \theta_m) \sum_{k_1<\cdots<k_r}[(1-\theta_{k_1})\cdots(1-\theta_{k_r})]^{-1} \cdot d\theta_1 \cdots d\theta_m, \qquad (r = 0, 1, \cdots, m),$$

where R_m is the region defined in (2.1), and the summation is extended over the $\binom{m}{r}$ selections of r distinct integers $k_1, \cdots, k_r$ from the set $1, 2, \cdots, m$. Thus, $L_0(s)$ is the Lt of $f_{n_1,n_2}(V)$, the density function of V. For $r \geqq 1$, the integrands in (2.5) are dominated by ϕ_{n_1,n_2-2}, and so the $L_r(s)$ exist only for $n_2 \geqq m+2$. This restriction will be preserved for the present. In general, we see that

$$(2.6)\qquad \int_{R_m} \exp(-s\sum\theta_i)\psi(\theta_1, \cdots, \theta_m)\, d\theta_1 \cdots d\theta_m$$

Received September 3, 1969; revised January 5, 1970.

[1] This research was partially supported by the National Institutes of Health Grant No. GM-12868.

is the ordinary Lt of

$$\Psi(V) = \int_{R_{m-1}(V)} \Psi(V-\theta_2-\cdots-\theta_m, \theta_2, \cdots, \theta_m)\, d\theta_2 \cdots d\theta_m, \tag{2.7}$$

where

$$\begin{aligned} R_{m-1}(V) &= R_{m-1} \cap \{\theta_2+\cdots+\theta_m > V-1\} \cap \{2\theta_2+\theta_3+\cdots+\theta_m < V\} \\ &= \{\max[\theta_{s+1}, V-(s-1)-\theta_{s+1}-\cdots-\theta_m] < \theta_s < s^{-1} \\ &\quad \cdot (V-\theta_{s+1}-\cdots-\theta_m);\ s = 2, \cdots, m\}, \qquad (\theta_{m+1} \equiv 0). \end{aligned} \tag{2.8}$$

Hence $L_r(s)$ is the Lt of $H_r(V)$, say, $(r = 0, 1, \cdots, m)$, which may be obtained in integral form from (2.5) and (2.7). Clearly, if $V = j, (j = 1, 2, \cdots, m-1)$, the left-hand sides of the inequalities in (2.8) reduce to θ_{s+1} for $s = j+1, \cdots, m$. The boundary of $R_{m-1}(V)$ therefore alters its character as V passes through the integer values $1, 2, \cdots, m-1$, corresponding to the passage of the hyperplane $\sum\theta_i = V$ through the vertices $(1, 0, \cdots, 0)$, $(1, 1, 0, \cdots, 0), \cdots, (1, 1, \cdots, 1, 0)$ of R_m. This results in $f_{n_1, n_2}(V)$ having a piecewise analytic nature which is reflected in the d.e.'s derived below.

A first-order system of d.e.'s relating the $L_r(s)$ may be obtained along the lines of [3], Section 2; in fact, the integrands at any stage of the argument may be derived formally from those given in this reference by making the transformation $w_i \to -\theta_i$, $n_2 \to m-n_1-n_2+1, s \to -s$.

This leads to the following system of d.e.'s:

$$\begin{aligned} -(m-r+1)sL_{r-1} + [s((d/ds)+r)+a_r+1]L_r - b_r L_{r+1} &= 0, \\ (r = 0, 1, \cdots, m-1;\ L_{-1} &\equiv 0), \\ ((d/ds)+m)L_m - L_{m-1} &= 0, \end{aligned} \tag{2.9}$$

where

$$a_r = \tfrac{1}{2}(m-r)(n_1+n_2-m+r-1)-1, \qquad b_r = \tfrac{1}{2}(r+1)(n_2-m+r-1). \tag{2.10}$$

Equation (2.9) may be obtained from [3] equations (2.19) and (2.20) by the transformations

$$s \to -s, \qquad n_2 \to m-n_1-n_2+1. \tag{2.11}$$

Inverting the Lt's, the following system of first order d.e.'s is found for the $H_r(V)$, $(n_2 \geqq m+2)$:

$$\begin{aligned} (m-r+1)\, dH_{r-1}/dV + [(V-r)\, d/dV - a_r]H_r + b_r H_{r+1} &= 0, \\ (r = 0, 1, \cdots, m-1;\ H_{-1} &\equiv 0), \\ H_{m-1} + (V-m)H_m &= 0. \end{aligned} \tag{2.12}$$

This system is related to [3] equations (2.21) and (2.22) by the transformations

$$T \to -V, \qquad n_2 \to m-n_1-n_2+1. \tag{2.13}$$

Since $b_r > 0$ for $n_2 \geqq m+2$, elimination of $H_1, \cdots, H_m$ from equation (2.12) will result in a linear homogeneous d.e. of order m for $H_0 = f$, having regular singularities at $V = 0, 1, \cdots, m$ and infinity.

3. Nature of the solution. The solution of (2.12) in the unit circle about $V = 0$ follows from [3] Section 3, using (2.13). Again the characteristic roots of the d.e. are $\frac{1}{2}mn_1 - 1$ and zero (with multiplicity m), the relevant solution following from the non-zero root. Recurrence relations for the coefficients in the power series for $f_{n_1, n_2}(V)$, $0 < V < 1$, are obtainable from [3] equation (3.11), and the multiplicative constant is the same as that for T, namely,

$$k(m; n_1, n_2) = \Gamma_m(\tfrac{1}{2}(n_1+n_2))/\Gamma(\tfrac{1}{2}mn_1)\Gamma_m(\tfrac{1}{2}n_2). \tag{3.1}$$

(Constantine [2]). This solution also serves to define the distribution in the interval $m-1 < V < m$, since from the definition of V,

$$f_{n_1, n_2}(V) = f_{n_2, n_1}(m-V), \qquad (0 < V < m). \tag{3.2}$$

Unfortunately, however, in the intervals between the singularities $1, 2, \cdots, m-1$, $f_{n_1, n_2}(V)$ will be specified by certain linear combinations of the full set of m linearly independent solutions. The calculation of the numerical coefficients in these linear combinations presents a formidable unsolved problem.

In the bivariate case $m = 2$, the differential equation for f_{n_1, n_2} is found to be

$$\begin{aligned} V(1-V)(2-V)f'' - [\tfrac{1}{2}(3n_1+3n_2-14)V^2 - 2(2n_1+n_2-7)V + 2(n_1-2)]f' \\ + \tfrac{1}{2}(n_1+n_2-4)[(n_1+n_2-4)V - 2(n_1-2)]f = 0, \end{aligned} \tag{3.3}$$

and the density function may be expressed in terms of the Gaussian hypergeometric function:

$$\begin{aligned} f_{n_1, n_2}(V) &= [2B(n_1, n_2-1)]^{-1}(\tfrac{1}{2}V)^{n_1-1}(1 = \tfrac{1}{2}V)^{n_2-3}\cdot \\ &\quad {}_2F_1(1, \tfrac{1}{2}(3-n_2); \tfrac{1}{2}(n_1+1); r^2), \qquad (0 < V < 1), \\ f_{n_1, n_2}(V) &= [2B(n_2, n_1-1)]^{-1}(\tfrac{1}{2}V)^{n_1-3}(1 = \tfrac{1}{2}V)^{n_2-1}\cdot \\ &\quad {}_2F_1(1, \tfrac{1}{2}(3-n_1); \tfrac{1}{2}(n_2+1); r^{-2}), \qquad (1 < V < 2), \end{aligned} \tag{3.4}$$

where $r = V/(2-V)$. These functions reduce to polynomials in V for odd $n_2 \geqq 3$ and odd $n_1 \geqq 3$, respectively.

So far, it has been assumed that $n_2 \geqq m+2$. In the cases $n_2 = m, m+1$ we note that f_{n_1, n_2} is a numerical multiple of the H_m function corresponding to f_{n_1, n_2+2}. Elimination of $H_0, \cdots, H_{m-1}$ from (2.12) with n_2 replaced by n_2+2 would show that f_{n_1, n_2} satisfies the general mth order d.e. in these cases. However, when $n_2 = m$, $m+1$, we have $b_1 = 0$, $b_0 = 0$ respectively, and the system (2.12), regarded as a d.e. for $H_0 = f_{n_1, n_2}$, degenerates into a second or first order d.e.:

$$\begin{aligned} V(1-V)H_0'' + [V(mn_1 - \tfrac{1}{2}m - \tfrac{1}{2}n_1 - 3) - (\tfrac{1}{2}mn_1 - 2)]H_0' \\ - (\tfrac{1}{2}mn_1 - \tfrac{1}{2}m - 1)(\tfrac{1}{2}mn_1 - \tfrac{1}{2}n_1 - 1)H_0 = 0, \qquad (n_2 = m), \end{aligned} \tag{3.5}$$

$$VH_0' - (\tfrac{1}{2}mn_1 - 1)H_0 = 0, \qquad (n_2 = m+1). \tag{3.6}$$

It may be shown that these d.e.'s validly specify the distribution in (0, 1), the solutions being

$$(3.7)\qquad f_{n_1,m}(V) = k(m; n_1, m)V^{\frac{1}{2}mn_1-1}\,{}_2F_1(\tfrac{1}{2}m, \tfrac{1}{2}n_1; \tfrac{1}{2}mn_1; V), \qquad (0 < V < 1),$$

$$f_{n_1,m+1}(V) = k(m; n_1, m+1)V^{\frac{1}{2}mn_1-1}, \qquad (0 < V < 1).$$

In virtue of (3.2), these results also define $f_{m,n}$ and $f_{m+1,n}$ in the interval $(m-1, m)$, V being replaced by $(m-V)$. It must be emphasized, however, that the degenerate d.e.'s (3.5)(3.6) do not hold throughout the entire range of V (with the exception of (3.5) when $m = 2$), although the general mth order d.e. does. The situation may be illustrated in the case $m = 3$, when

$$(3.8)\qquad f_{4,3}(V) = (6/7)(3-V)^{7/2}, \qquad (2 < V < 3),$$

$$f_{4,4}(V) = (3/8)(3-V)^5, \qquad (2 < V < 3).$$

These functions are not solutions of the d.e.'s (3.5), (3.6) respectively, but by taking each in turn as H_3 in (2.12) with $n_2 = 5, 6$, they may be shown to satisfy the general 3rd order d.e. for $m = 3$. That f_{n_1,n_2} may be cusped, with discontinuous first derivative, may be seen by taking $n_1 = n_2 = 3$ in (3.4).

4. Moments of V. From [3], Section 7, the system of d.e.'s (2.7) for the Lt. $L_0(s)$ of $f_{n_1,n_2}(V)$ has characteristic roots $-(a_r+1)$ at the regular singularity $s=0$. These are all negative with the exception of $-(a_m+1) = 0$, and the system has an analytic solution at the origin as we would expect, since V has a finite range, and all its moments exist.

By virtue of (2.11), a recurrence relation for $\mathscr{E}V^r$ may be obtained from equation (7.13) of [3] for $\mathscr{E}T^r$ by replacing n_2 by $m-n_1-n_2+1$ and multiplying by $(-1)^r$, $(r = 1, 2, \cdots)$. Pillai [9] has used the first four moments of V to fit a Pearson curve to the distribution. The following reduced form of Pearson's coefficient β_2 has been derived using the above recurrence relation:

$$(4.1)\qquad \beta_2 = 3(N-1)(N+2)A/mn_1n_2(N-m)(N-3)(N-2)(N+1)(N+4)(N+6),$$

where

$$N = n_1+n_2,$$

$$A = n_1n_2[(Nm-m^2)(N^3+5N^2+78N+72)-4N^2(5N+6)]$$
$$+4N^2[(m^2-Nm)(5N+6)+N(N^2+N+2)].$$

5. Itô-type expansions for large n_2. For completeness, we note that an Itô-type expansion [6] for the distribution of n_2V for large n_2 may be obtained from [3] Section 4. Noting that n_2V is asymptotically distributed as χ^2 on mn_1 degrees of freedom, a convenient approach is to expand the cumulant generating function of the statistic in a series of the type considered by Box [1]:

$$(5.1)\qquad \log L_0(s/n_2) \sim -\tfrac{1}{2}mn_1 \log(1+2s)+\textstyle\sum_{r=1}^{\infty} \omega_{r,V}[(1+2s)^{-r}-1].$$

Using the differential equations, the following set of recurrence relations may be obtained for the $\omega_{r,V}$:

$$(5.2)\qquad 2r\omega_{r,V} = 2(r-1)\omega_{r-1,V}+mn_1\delta_{1,r}-(1-(m+1)/n_2)\xi_{1,r}, \qquad (r = 1, 2, \cdots),$$

where the $\xi_{j,r}$ are defined by

$$\xi_{0,r} = \xi_{r,0} = \delta_{0,r},$$

$$(5.3)\quad \begin{aligned} j\xi_{j,r} = {} & \alpha_j\xi_{j-1,r-1}+(\beta_j+2(r-1))\xi_{j,r-1}/n_2 \\ & +[(j+1)/n_2-\gamma_j/n_2^2]\xi_{j+1,r-1}-[mn_1+2(r-2)]\xi_{j,r-2}/n_2 \\ & +2n_2^{-1}\textstyle\sum_{s=1}^{r-2}s\omega_{s,V}(\xi_{j,r-s-1}-\xi_{j,r-s-2}), \\ & \qquad\qquad (j = 1, \cdots, m;\ r = 1, 2, \cdots), \end{aligned}$$

$$\alpha_j = (m-j+1)(n_1-j+1), \qquad \beta_j = j(2m+n_1-2j+2), \qquad \gamma_j = (j+1)(m-j+1),$$

ξ's with negative subscripts being zero. Thus, in particular,

$$(5.4)\quad \begin{aligned} \omega_{1,V} &= mn_1(m+1)/2n_2, \\ \omega_{2,V} &= -\tfrac{1}{4}mn_1[(m+n_1+1)/n_2-(m+1)(2m+n_1+2)/n_2^2]. \end{aligned}$$

The first six ω's to order n_2^{-3} have been derived by Muirhead [7] using an independent approach, and the first eight to order n_2^{-4} by the present author. An analogue of Itô's expansion of T_0^2 percentiles in terms of $\chi^2_{mn_1}$ percentiles may be derived from a general Cornish–Fisher inversion of Box-type series given by the author [4]. To order n_2^{-2},

$$(5.5)\quad \begin{aligned} n_2V \sim {} & \chi^2+1/2n_2[\chi^2(m-n_1+1)-\chi^4(m+n_1+1)/(mn_1+2)] \\ & +1/24n_2^2\{\chi^2[7m^2-12m(n_1-1)+(7n_1^2-12n_1+1)] \\ & -\chi^4[11m^2+24m-13n_1^2+17]/(mn_1+2) \\ & +2\chi^6[2m^3n_1+m^2(2n_1+3n_1+10)+m(2n_1^3+3n_1^2+17n_1+18) \\ & +2(5n_1^2+9n_1+2)]/(mn_1+2)^2(mn_1+4) \\ & +6\chi^8(m-1)(m+2)(n_1-1)(n_1+2)/(mn_1+2)^2(mn_1+4)(mn_1+6)\} \\ & +O(n_2^{-3}), \end{aligned}$$

and the n_2^{-3} term has also been obtained.

An expansion of the type (5.1) also exists for T_0^2. In view of (2.11), the following relationship exists between the coefficients $\omega_{r,T}$ in this series and the $\omega_{r,V}$:

$$(5.6)\quad \begin{aligned} (-n_2)^r\omega_{r,T} = {} & mn_1(n_1-m-1)^r/2r \\ & +\textstyle\sum_{s=1}^{r}\binom{r-1}{s-1}(m-n_1-n_2+1)^s(n_1-m-1)^{r-s}\omega_{s,V}, \end{aligned}$$

where, in the $\omega_{s,V}$, n_2 is to be replaced by $m-n_1-n_2+1$. T and V may also be interchanged in this formula. The $\omega_{r,T}$ obtained from (5.6) and (5.2) check with those obtained to order n_2^{-3} by Muirhead (*loc. cit.*).

6. Examination of the approximations. In principle, the solution of (2.12) at the regular singularity $V = 0$ specifies the distribution of V in (0, 1) (or in $(m-1, m)$ when n_1 and n_2 are interchanged). For sufficiently large n_2 (or n_1, respectively), the upper 5% and 1% points of V lie in these intervals, and some investigation may be made of the accuracy of the available approximations. A corresponding study has been made for T in [5], where the d.e. was used to compute accurate percentage points by analytic continuation of the solution at $T = 0$. The same computer program, with the trivial modification (2.13), has been used to tabulate some percentiles of V in the range $m \leqq 5$, n_1 and $n_2 \leqq 200$. Except when n_1 and n_2 are both small integers, Pillai's Pearson curve approximation is accurate to four decimal places. The Itô-type approximation (5.5) is considerably improved by adding the n_2^{-3} term, and is a useful direct formula for large n_2 and small n_1, but its accuracy falls off rapidly as n_1 increases. In virtue of (3.2), a similar statement holds with n_1 and n_2 interchanged.

REFERENCES

[1] Box, G. E. P. (1958). A general distribution theory for a class of likelihood criteria. *Biometrika* **36** 317–46.

[2] Constantine, A. G. (1966). The distribution of Hotelling's generalized T_0^2. *Ann. Math. Statist.* **37** 215–225.

[3] Davis, A. W. (1968). A system of linear differential equations for the distribution of Hotelling's generalized T_0^2. *Ann. Math. Statist.* **39** 815–832.

[4] Davis, A. W. (1970). Cornish–Fisher-type inversions of Box's asymptotic series. To be submitted.

[5] Davis, A. W. (1970). Exact distributions of Hotelling's generalized T_0^2. *Biometrika* **57** 187–191.

[6] Itô, K. (1956). Asymptotic formulae for the distribution of Hotelling's generalized T_0^2 statistic. *Ann. Math. Statist.* **27** 1091–1105.

[7] Muirhead, R. J. (1970). Asymptotic distributions of some multivariate tests. *Ann. Math. Statist.* **41,** 1002–1010.

[8] Pillai, K. C. S. (1955). Some new test criteria in multivariate analysis. *Ann. Math. Statist.* **27** 117–121.

[9] Pillai, K. C. S. (1960). *Statistical Tables for Tests of Multivariate Hypotheses.* The Statistical Center Univ. of the Philippines