

UNIVERSITY OF NORTH CAROLINA
Department of Statistics
Chapel Hill, N. C.

Mathematical Sciences Directorate
Air Force Office of Scientific Research
Washington 25, D. C.

AFOSR Report No.

ON A NEW DERIVATION OF A WELL KNOWN DISTRIBUTION

by

S. N. Roy

April, 1960

Contract No. AF 49(638)-213

It is well known that so far the joint distribution of the latent roots associated with normal multivariate analysis of variance has been considerably more difficult to derive if the effective number of variates is greater than the number of components of the linear hypothesis than if it is the other way around. This report offers both on the null and the non null hypothesis a simple method of derivation of the distribution for the former case by throwing it back on the distribution for the latter case, and in this tie-up a pivotal role is played by the distribution of the latent roots connected with the testing of the hypothesis of independence between two sets of variates.

Qualified requestors may obtain copies of this report from the ASTIA Document Service Center, Arlington Hall Station, Arlington 12, Virginia. Department of Defense contractors must be established for ASTIA services, or have their "need to know" certified by the cognizant military agency of their project or contract.

Institute of Statistics
Mimeograph Series No. 254

ON A NEW DERIVATION OF A WELL KNOWN DISTRIBUTION¹

by

S. N. Roy
Department of Statistics
University of North Carolina

Summary and Introduction.

The joint distribution of the characteristic roots of the quotient of two matrices (in the sense of one matrix times the inverse of the other) was obtained under the null hypothesis and, up to a certain point, under the non null hypothesis, by a number of workers in the late thirties and during the forties [1,3], as part of an attempt to solve certain inference problems in multivariate normal distributions, including, in particular, those of testing (i) the hypothesis of equality of two dispersion matrices, (ii) the hypothesis of independence between two sets of variates and (iii) multivariate linear hypothesis. It is well known that the three sets of characteristic roots connected with the three problems mentioned just now have, under the respective null hypotheses for (i), (ii), (iii), the same form of joint distribution, but have different forms of joint distribution under the respective non null hypotheses [1,3]. Under the case (iii), there are two subcases, namely (a) $u \leq s$ and (b) $u > s$, where u denotes the "effective number of variates" and s the "components of a linear hypothesis" (for example, the number of treatment contrasts), both phrases [3,5] to be

1. This research was supported by the United States Air Force through the Air Force Office of Scientific Research of the Air Research and Development Command, under Contract No. AF 49(638)-213. Reproduction in whole or in part is permitted for any purpose of the United States Government.

explained in the next section. While the two sets of characteristic roots for the two subcases (a) and (b) under (iii) have the same form of joint distribution, it is also well known that by most methods of derivation, the distribution for the subcase (a) is considerably easier to derive than for the subcase (b). For example two alternative methods of deriving the distribution for the subcase (b) are given in [3], but even the shorter of the two is lengthier and more involved than that for (a). Given the distribution for the case (a), this paper offers a method of throwing back the distribution for the case (b) on that for the case (a) under both the null and the non null hypothesis by using certain arguments that are reminiscent and, in a sense, a generalization of the simple reasoning by which Fisher [2] threw back on the F-distribution, the distribution of what was essentially Hotelling's T^2 , at a time when he was not aware of Hotelling's prior derivation of the distribution by a more formal and elaborate method..

2. Preliminaries.

The model for multivariate linear hypothesis [3,5]. Let $X = \begin{bmatrix} x_1 & x_2 & \dots & x_p \end{bmatrix}'_p$ be n independent $N \left[\begin{matrix} \xi \\ \Sigma \end{matrix} \right]_{px1}$, where Σ is an unknown symmetric p.d. matrix of parameters called the population dispersion matrix, and the expectations are given by

$$(2.1) \quad \mathcal{E}(X') = n \begin{bmatrix} A_1 & : & A_2 \\ r^1 & & m^2_r \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} \begin{matrix} r \\ m-r \end{matrix},$$

where $A = \begin{bmatrix} A_1 & : & A_2 \\ r^1 & & m^2_r \end{bmatrix}_{nxm}$ is the structure matrix given by the design of the experiment and what is usually called the model by experimental

statistically and $\xi = \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix}_{m \times p}$,

A_1 is a basis (not necessarily unique) of the matrix A . It is also assumed that $\text{rank}(A) = r \leq m < n$, and $r < n-p$.

The multivariate linear hypothesis $[3,5]$. Under this model the hypothesis \mathcal{H}_0 to be tested and the alternative \mathcal{H} against which it is supposed to be tested are supposed to be given by

$$(2.2) \quad \mathcal{H}_0: s \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} \begin{bmatrix} \xi_1 \text{ (rxp)} \\ \xi_2 \text{ (m-rxp)} \end{bmatrix} \begin{matrix} W \\ pxu \end{matrix} = 0$$

against

$$\mathcal{H} : \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} \begin{bmatrix} \xi_1 \\ \xi_2 \end{bmatrix} W \neq 0 = \eta \text{ (say),} \\ \text{sxu}$$

where \mathcal{H}_0 is assumed to be a "testable" $[3,5]$ hypotheses for which a necessary and sufficient condition is that

$$(2.3) \quad \text{rank} [A] = \text{rank} \begin{bmatrix} A \\ C \end{bmatrix}, \text{ where } C = \begin{bmatrix} C_1 \\ C_2 \end{bmatrix} \text{ s..} \\ \text{sxm} \quad \text{r} \quad \text{m-r}$$

It is assumed that $\text{rank} [C] = s \leq r$, and $\text{rank} [W] = u \leq p$. The given matrices C and W together are called the hypothesis matrices; u is called the "effective number of variates" and s the "number of components of the linear hypothesis." In many problems $u=p$, in which case, W will disappear if we post multiply both sides of (2.2) by W^{-1} [notice that in this case, if $u=p$, W will be non-singular]. A partition of A into A_1 and A_2 induces a partition of ξ into ξ_1 and ξ_2 and that in turn induces a partition of C into C_1 and C_2 . No matter whether we use the λ criterion, or the largest root criterion or the sum of the roots criterion the test will come out in terms of two matrices which play a pivotal role and are given by

$$(2.4) \quad \begin{matrix} sS^* \\ uxu \end{matrix} = W' X Q_{\mathcal{H}_0} X' W; \quad \begin{matrix} (n-r) S \\ uxu \end{matrix} = W' X Q_{\text{error}} X' W,$$

where Q_{H_0} and Q_{error} are $n \times n$ matrices (called respectively the matrices due to the hypothesis and the error) and are given [3,5] in terms of A_1 and C_1 , and shown to be invariant under the choice of a basis A_1 for A and a consequent choice of C_1 from C . The matrices S^* and S themselves might be called respectively the ^{sample} dispersion matrices due to the hypothesis and to the error. S^* and S are symmetric matrices, S being, almost everywhere, p.d. and S^* being, almost everywhere, at least p.s.d. of rank $t = \min(u, s)$. Each of the three tests mentioned above comes out in terms of the characteristic roots of $[S^* S^{-1}]$, except that for the λ criterion the primitive form is one in terms of the roots of $[S (\frac{sS^* + n-rS}{s+n-r})^{-1}]$, which, of course, can be expressed in terms of the roots of $[S^* S^{-1}]$. For the two other tests, though not for the λ criterion, the joint distribution of the roots becomes an indispensable first step toward the test construction. Almost everywhere, the number of positive roots of $[S^* S^{-1}]$ is equal to $t = \min(u, s)$, the other $u-t$ roots being zero, so that, if $u \leq s$, all the roots are positive. The two subcases (a) $u \leq s$ and (b) $u > s$, are thus seen to arise in a natural manner. Under the non null hypothesis ~~the~~ the sampling distribution of these roots (and hence the power of anyone of the three tests mentioned) involves as parameters, aside from the degrees of freedom u, s and $n-r$, a set of "noncentrality parameters" $\zeta_1 \zeta_2 \dots \zeta_t$, given by the positive roots of

$$(2.5) \quad \eta' [C_1 (A_1' A_1)^{-1} C_1']^{-1} \eta, \quad ,$$

where $\eta_{s \times u}$ is a set of "deviation parameters." These roots again are invariant under a choice of A_1 (and of C_1). For the distribution of the

roots of $[S^* S^{-1}]$ (to be called c_1, c_2, \dots, c_t) we have, as a starting point, the distribution of the canonical matrices $Z^* [= (Z_{ij}^*)]$ and $Z_{u \times n-r} [= (Z_{ij})]$, given by

$$(2.6) \quad \text{const exp } [-\frac{1}{2} \left\{ \text{tr } Z^* Z^{*'} + \text{tr } Z Z' + \sum_{i=1}^t \left(\gamma_i - \frac{2}{\sum_{i=1}^t \gamma_i} \right) Z_{ii}^* \right\}] dZ^* dZ,$$

where the roots of $[S^* S^{-1}]$ are the same as of $\frac{n-r}{s} [(Z^* Z^{*'}) (Z Z')^{-1}]$

In this paper, assuming that the distribution of the u (positive roots) is known for the case $u \leq s$, we shall show how the distribution of the s (positive roots) for the case $u > s$, can be thrown back on the former case. This is done by tying up the distribution for both cases with the distribution of roots for the problem (ii) of independence between two sets of variates.

3. Independence between two sets of variates.

For independence between a p -set and a q -set (with a $p+q$ multivariate normal distribution), given a sample of size $n^* > p+q$, we are concerned with the joint distribution of a set of p roots ($c_1^*, c_2^*, \dots, c_p^*$) which can be expressed [3,4] as the characteristic roots of $[UU'(VV')^{-1}]$ where $U [= (u_{ij})]$, $V [= (v_{ij})]$ and T (a triangular matrix with positive diagonal elements and zero upper off diagonal elements) are canonical matrices having the distribution

$$(3.1) \quad \text{const exp } [-\frac{1}{2} \left(\sum_{i=1}^p \sum_{j=1}^q (u_{ij} - \gamma_{ij} t_{ij})^2 + \sum_{i=1}^q \sum_{j=1}^i t_{ij}^2 + \sum_{i=1}^p \sum_{j=q+1}^{n^*} v_{ij}^2 \right)] \\ \times dU dV \prod_{i=1}^q t_{ii}^{n^*-1} dT,$$

where $\gamma_{ij} = \rho_i / \sqrt{1 - \rho_i^2} = \gamma_i$ (say), for $j = 1, 2, \dots, i$ and $i = 1, 2, \dots, p$; and $= 0$, otherwise. The ρ_i^2 's are the characteristic roots of $\begin{bmatrix} \Sigma_{11}^{-1} & \Sigma_{12} & \Sigma_{22}^{-1} & \Sigma'_{12} \end{bmatrix}$, where the symmetric p. d. population dispersion matrix for the $(p + q)$ set of variates is given by

$$(3.2) \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma'_{12} & \Sigma_{22} \end{bmatrix} \begin{matrix} p \\ q \end{matrix}$$

$\begin{matrix} p & q \end{matrix}$

It may be noted that if

$$(3.3) \quad S = \begin{bmatrix} S_{11} & S_{12} \\ S'_{12} & S_{22} \end{bmatrix} \begin{matrix} p \\ q \end{matrix}$$

$\begin{matrix} p & q \end{matrix}$

is the sample dispersion matrix, then denoting by e_i 's the squares of the sample canonical correlation coefficients, given by the characteristic roots of $\begin{bmatrix} S_{11}^{-1} & S_{12} & S_{22}^{-1} & S'_{12} \end{bmatrix}$, we have that

$$(3.4) \quad c_i^* = e_i / (1 - e_i) \quad (i = 1, 2, \dots, p).$$

We observe further that if $p \leq q$, then, almost everywhere, all p of e_i 's and hence of c_i^* 's are positive, and for the joint distribution of these roots we can start from (3.1) as the canonical form. On the other hand, if $p > q$, then while the form (3.1) would be still permissible, another canonical form would be more convenient. To obtain this second canonical form we first observe that even if $p > q$, the positive roots of

$\begin{bmatrix} S_{11}^{-1} & S_{12} & S_{22}^{-1} & S'_{12} \end{bmatrix}$ are the same as the roots (all positive, almost everywhere) of $\begin{bmatrix} S^{-1} & S'_{12} & S_{11}^{-1} & S_{12} \end{bmatrix}$. To obtain a canonical form for the distribution problem of the roots as expressed in terms of the

latter matrix, we reverse the roles of the p-set and the q-set and end up with the joint distribution of canonical matrices. $U_1 \begin{bmatrix} (u_{lij}) \end{bmatrix}$, $q \times p$

$V_1 \begin{bmatrix} (v_{lij}) \end{bmatrix}$ and T_1 , given by $p \times p$

$$(3.5) \text{ const exp } \left[-\frac{1}{2} \left\{ \sum_{i=1}^q \sum_{j=1}^p (u_{lij} - \gamma_{lij} t_{lij})^2 + \sum_{i=1}^p \sum_{j=1}^1 t_{lij}^2 + \sum_{i=1}^q \sum_{j=p+1}^{n^*} v_{lij}^2 \right\} \right] dU_1 dV_1 \prod_{i=1}^p t_{lii}^{n^*-i} dT_1,$$

where $\gamma_{ij} = \rho_i / \sqrt{1 - \rho_i^2} = \gamma_i$ (say), for $j = 1, 2, \dots, i$ and $i = 1, 2, \dots, q$; and $= 0$ otherwise. Notice that ρ_i^2 are the characteristic roots of $\begin{bmatrix} \Sigma_{22}^{-1} & \Sigma_{1\gamma}' & \Sigma_{11}^{-1} & \Sigma_{12} \end{bmatrix}$, and the positive roots of this matrix are the same as the positive roots of $\begin{bmatrix} \Sigma_{11}^{-1} & \Sigma_{12} & \Sigma_{22}^{-1} & \Sigma_{12}' \end{bmatrix}$ which would justify using the same symbol ρ_i for both cases. Thus, if in (3.1) $p > q$, then for the distribution of the positive characteristic roots of $\begin{bmatrix} UU' & (VV')^{-1} \end{bmatrix}$ of (3.1), we use the canonical form (3.5), after having observed that the positive roots of $\begin{bmatrix} UU' & (VV')^{-1} \end{bmatrix}$ are the same as those of $\begin{bmatrix} U_1 U_1' & (V_1 V_1')^{-1} \end{bmatrix}$ and with the same multiplicity. Almost everywhere, all roots of the latter matrix are positive. This means that if, starting from (3.1), we obtain, as we can, the distribution of the roots of $\begin{bmatrix} U U' & (V V')^{-1} \end{bmatrix}$ (all positive, almost everywhere) when $p \leq q$, then all we have to do to obtain the distribution of the positive roots of this matrix when $p > q$, is to take the first distribution and replace p by q , q by p and $n^* - q$ by $n^* - p$.

4. Tie-up between the problems of sections 3 and 4.

Comparing (2.6) with (3.1) we observe that both on the null and the non-null hypothesis the distribution of the roots of $\begin{bmatrix} Z^* Z^{*'} & (Z Z')^{-1} \end{bmatrix}$

is the same as of the conditional distribution, given T , of the roots of $\int \bar{U} U' (V V')^{-1} \int$, if we put $p = u$, $q = s$, $n^* - q = n - r$, $t_{ii} = 1$, $t_{ij} (i \neq j) = 0$ and $\gamma_{ii} = \zeta_i$. This is otherwise obvious, once we obtain from (3.1) the unconditional distribution of T . This unconditional distribution is obtained by observing that the elements of U and V vary from $-\infty$ to $+\infty$, and, integrating these out, we obtain, for T , the distribution

$$(4.1) \quad \text{const exp} \int^{-\frac{1}{2}} \sum_{i=1}^q \sum_{j=1}^i t_{ij}^2 \int \prod_{i=1}^q t_{ii}^{n^* - i} dT.$$

Hence, the distribution of the positive roots of $\int Z^* Z^* (Z Z')^{-1} \int$ will be the same as the conditional distribution, given T , of the positive roots of $\int \bar{U} U' (V V')^{-1} \int$, after proper identification between the two sets of parameters and proper specification of the elements of T , held fixed. The conditional distribution of the positive roots of $\int \bar{U} U' (V V')^{-1} \int$ again will be the same as the conditional distribution of the roots (all positive, almost everywhere) of $\int \bar{U}_1 U'_1 (V_1 V'_1)^{-1} \int$. Hence it is obvious that if, starting from (2.6), we obtain, as we can, the distribution of the roots of $\int \bar{U} U' (V V')^{-1} \int$, when $u \leq s$, then, to obtain the distribution of the positive roots of the same matrix, when $u > s$, all we have to do is to take the first distribution and replace u by s , s by u and $n - r$ by $n - r - u + s$; this replacement follows if we recall the nature of the tie-up between the problems of section 2 and section 3. We have, for the first case ($p \leq q$, $u \leq s$), $p = u$, $q = s$, $n^* - q = n - r$, i.e., $n^* = n - r + q = n - r + s$. This n^* is to stay the same for both cases of section 3. Also for the second case ($p > q$; $u > s$), again, $p = u$, $q = s$; if now for this second case we put $n^* - p = n - r - u + s$, then it turns out that $n^* = n - r - u + s + p = n - r - u + s + u = n - r + s$, which is a consistent result in that n^* should be

the same for both cases of section 3, and $n-r+s$ should be the same for both cases of section 2. Thus, if we relate the tie-up between the distribution problems of the two cases of section 3 to the tie-up between the distributions of the two cases for section 2 we observe that

$$(4.2) \quad [p \rightarrow q; q \rightarrow p; n^* \rightarrow n^*] \longleftrightarrow [u \rightarrow s, s \rightarrow u, n-r \rightarrow n-r-u+s].$$

Special Cases. Consider, as a special case $s = 1$ in section 2. We have, for the single positive root, Hotellings T^2 and, by the tie-up between the problems of sections 2 and 3, the distribution of Hotelling's T^2 is the same as the conditional distribution of the F-transform of the square of the multiple correlation of a u -set and a 1 -set, that is, of a 1 -set and a u -set, which is the usual F-distribution. This was essentially the argument used by Fisher. Notice that here the chain is $(u > s) \rightarrow (p > q) \rightarrow (p \leq 1)$, and we stop at the third stage and utilize the well-known F-distribution. This works when $s = 1$. However, when both u and s are greater than 1 the chain that we use is $(u > s) \rightarrow (p > q) \rightarrow (p \leq q) \rightarrow (u \leq s)$.

We might conclude by giving another example [1,3]. In section 2, under \mathcal{H}_0 the distribution of the roots for the case $u \leq s$, is given by

$$(4.3) \quad \text{const} \prod_{i=1}^u \frac{c_i^{\frac{s-u-1}{2}} dc_i}{(1 + \frac{s}{n-r} c_i)^{\frac{n-r+s}{2}}} \prod_{i>j} (c_i - c_j), \quad (0 \leq c_1 \leq \dots \leq c_u < \infty).$$

Hence under \mathcal{H}_0 the distribution, for the case $u > s$, will be given by

$$(4.4) \quad \text{const} \prod_{i=1}^s \frac{c_i^{\frac{u-s-1}{2}} dc_i}{\left(1 + \frac{u}{n-r-u+s} c_i\right)^{\frac{n-r+s}{2}}} \prod_{i>j} (c_i - c_j), \quad (0 \leq c_1 \leq \dots \leq c_s < \infty).$$

Both results are, of course, well-known.

5. Concluding remarks.

It will be seen that in throwing back the second case of section 2 on the first case the argument that is used is basically very simple. Most of the space in this paper has been taken up just to explain the background of the pivotal distributions (2.6), (3.1) and (3.5). If these pivotal distributions and their background were better known and could be taken for granted, then section 4, or rather the part of it that precedes the two illustrations, is all that is needed for the derivation of the distribution for the case $u > s$.

Bibliography

- [1] Anderson, T.W., An introduction to multivariate statistical analysis, John Wiley and Sons, New York, (1958).
- [2] Fisher, R. A., "The use of multiple measurements in taxonomic problems," Annals of Eugenics, Vol.7 (1936), pp. 179-188.
- [3] Roy, S. N., Some aspects of multivariate analysis, John Wiley and Sons, (1958).
- [4] Roy, S. N. and Mikhail, W. F., "On the monotonic character of the power functions of two multivariate tests," North Carolina Institute of Statistics mimeograph series no. 238 (1959).
- [5] Roy, S. N. and Gnanadesikan, R., "Contributions to ANOVA in one or more dimensions, Parts I and II", Annals of Mathematical Statistics, Vol. 30 (1959) pp. 304-340.