

HSU'S WORK IN MULTIVARIATE ANALYSIS

BY T. W. ANDERSON

Stanford University

From 1938 to 1945 Hsu published papers in the forefront of the development of the mathematical theory of multivariate analysis. It can be presumed that he was influenced by his proximity to R. A. Fisher, who was also at University College, London. After 1945 he lectured on multivariate analysis at Columbia University and the University of North Carolina, where he trained students who pursued research in this area. As a highly-trained mathematician Hsu promoted the use of matrix theory in statistical theory as well as proving new theorems about matrices.

A crucial element of multivariate theory is the distribution of the sample covariance matrix \mathbf{S} . When the p -component vectors are independently distributed each according to $N(\mathbf{0}, \Sigma)$, $(N - 1)\mathbf{S} = \mathbf{A} = \sum_{\alpha=1}^N (\mathbf{X}_\alpha - \bar{\mathbf{X}})(\mathbf{X}_\alpha - \bar{\mathbf{X}})'$ has the so-called Wishart distribution with density (for \mathbf{A} positive definite) of

$$(1) \quad w(\mathbf{A}|\Sigma, n) = K(\Sigma, n) |\mathbf{A}|^{\frac{1}{2}(n-p-1)} e^{-\frac{1}{2} \text{tr } \Sigma^{-1}\mathbf{A}},$$

where $K(\Sigma, n)$ is a constant depending on Σ , its order p , and the "degrees of freedom" $n = N - 1$. For $p = 2$ the density of a_{11} , a_{22} , and $r_{12} = a_{12}a_{11}^{-\frac{1}{2}}a_{22}^{-\frac{1}{2}}$ was obtained by Fisher (1915) in his famous paper, which marked the beginning of rigorously derived exact small-sample distribution theory. Wishart's paper of 1928 derived the density (1) by use of a geometric argument, which was, roughly speaking, a generalization of that of Fisher. Since Wishart's publication many alternative proofs have been given. That of Hsu [5], based on algebra and analysis, is particularly elegant. To derive the density for p and n Hsu assumed it for $p - 1$ and $n - 1$. In addition to the matrix with this density there is needed a $(p - 1)$ -component normal vector and an n -component normal vector. With a little algebraic manipulation only the analytic derivation of the χ_n^2 -distribution for the norm squared of the n -component vector is needed to complete the proof.

Mahalanobis, Bose and Roy (1937) approached the distribution of \mathbf{A} by writing $\mathbf{A} = \mathbf{T}\mathbf{T}'$, where \mathbf{T} is lower triangular ($t_{ij} = 0, i < j$), and derived the distribution of the $p(p + 1)/2$ elements of T , called "rectangular coordinates", from the distribution of $\mathbf{X}_1, \dots, \mathbf{X}_N$. The (nonzero) elements of \mathbf{T} were shown to be independent when $\Sigma = \mathbf{I}$. Each off-diagonal element of \mathbf{T} has the standard univariate normal distribution, and the i th (nonnegative) diagonal element has a χ -distribution with $n - i + 1$ degrees of freedom. Hsu's derivation of the distribution of rectangular coordinates in [8] is algebraic and analytic instead of the more geometric method of Mahalanobis, Bose and Roy. It is in the same spirit as the derivation of the Wishart distribution in [5].

Received June 1978.

AMS 1970 subject classifications. Primary 01A70; secondary 62H10.

Key words and phrases. Obituary, multivariate analysis.

The next most basic distributions to be derived in multivariate analysis were those of the roots of certain determinantal equations. That this was the "natural" development is signified by the fact that such distributions were discovered independently and almost simultaneously by Hsu [6], Fisher (1939), Girshick (1939), Mood (1951) and Roy (1939). (For Girshick and Mood the research was intended for doctoral dissertations, but having been "beat out" by Fisher and Hsu they turned to other topics. Bose (1977) has some interesting remarks on the independence of the work of Fisher and Roy.) For **A** positive semidefinite and **B** positive definite let $\theta_1 \geq \theta_2 \geq \dots \geq \theta_p \geq 0$ be the roots of

$$(2) \quad |\mathbf{A} - \theta(\mathbf{A} + \mathbf{B})| = 0.$$

If **A** and **B** are independently distributed according to $W(\Sigma, m)$ and $W(\Sigma, n)$, respectively, with $m \geq p$ and $n \geq p$, the density of $\theta_1, \dots, \theta_p$ is a constant times

$$(3) \quad \prod_{i=1}^p \theta_i^{\frac{1}{2}(m-p-1)} \prod_{i=1}^p (1 - \theta_i)^{\frac{1}{2}(n-p-1)} \prod_{i=1}^p \prod_{j=i+1}^p (\theta_i - \theta_j).$$

If the roots of $|\mathbf{A} - \phi\mathbf{B}| = 0$ are $\phi_1 \geq \dots \geq \phi_p \geq 0$, then $\phi_i = \theta_i(1 - \theta_i)^{-1}$, $i = 1, \dots, p$. Hsu made the transformation $\mathbf{A} = \mathbf{W}\mathbf{D}_\phi\mathbf{W}'$, $\mathbf{B} = \mathbf{W}\mathbf{W}'$, where \mathbf{D}_ϕ is a diagonal matrix with ϕ_1, \dots, ϕ_p as diagonal elements. The transformation is one-to-one if $w_{ij} > 0$ (with probability 1). The derivation involves (i) substituting the above expressions into the density of **A** and **B**, namely, $w(\mathbf{A}|\Sigma, m)w(\mathbf{B}|\Sigma, n)$, (ii) multiplying by the Jacobian of the transformation, (iii) integrating out the elements of **W** to obtain the marginal density of ϕ_1, \dots, ϕ_p , and (iv) converting to $\theta_1, \dots, \theta_p$. The difficult part of this procedure is the calculation of the Jacobian. In fact, in this paper Hsu states the Jacobian for arbitrary p but proves it only for $p = 3$. (The 12×12 matrix of partial derivatives is written explicitly and evaluated.)

A tractable method of evaluating the Jacobian was subsequently developed by Hsu and was presented in his lectures at the University of North Carolina. Unfortunately Hsu never wrote up his lecture notes, but several students recorded them. Deemer and Olkin (1951) elaborated the methods and results and made them available (with the approval of Hsu, then in China). The case of $m < p$ was also treated in [6]. The roots of $|\mathbf{A} - \lambda\Sigma| = 0$ have the density which is a constant times

$$(4) \quad \prod_{i=1}^p \lambda_i^{\frac{1}{2}(m-p-1)} e^{\frac{1}{2}\sum_{i=1}^p \lambda_i^{-1}} \prod_{i=1}^p \prod_{j=i+1}^p (\lambda_i - \lambda_j)$$

for $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$. The exposition in Chapter 13 of Anderson (1958) is based on the papers of Hsu and Deemer and Olkin. The constants in the densities are not given here in order to contain the exposition, but they are, in fact, important (in obtaining moments of likelihood ratio test criteria, for instance) and require careful treatment.

In obtaining the distribution of the roots it is crucial that the Wishart distributions be central. Suppose $\widehat{\mathbf{E}} \mathbf{X}_{\alpha t} = \boldsymbol{\mu}_t$, $\alpha = 1, \dots, N_t$, $t = 1, \dots, k$. Let $\mathbf{A} = \sum_{t=1}^k N_t (\bar{\mathbf{X}}_t - \bar{\mathbf{X}})(\bar{\mathbf{X}}_t - \bar{\mathbf{X}})'$ and $\mathbf{B} = \sum_{t=1}^k \sum_{\alpha=1}^{N_t} (\mathbf{X}_{\alpha t} - \bar{\mathbf{X}}_t)(\mathbf{X}_{\alpha t} - \bar{\mathbf{X}}_t)'$. Then **A** and **B**

are independent, \mathbf{B} has a central Wishart distribution and \mathbf{A} has the noncentral Wishart distribution depending on $\Psi = N^{-1} \sum_{\alpha=1}^k N_t (\mu_t - \bar{\mu})(\mu_t - \bar{\mu})'$ as well as on Σ and $k - 1$. [The noncentral Wishart distributions for ranks of Ψ being one or two were obtained independently by Anderson and Girshick (1944)—again, intended as dissertation topics]. In [10] Hsu treated the roots of $|\mathbf{A} - \phi \mathbf{B}| = 0$ when \mathbf{B} has a central Wishart distribution and \mathbf{A} has a noncentral Wishart distribution. The distribution of these random roots depends on μ_1, \dots, μ_k , and Σ only through the roots of

$$(5) \quad |\Psi - \lambda \Sigma| = 0.$$

The exact distribution of ϕ_1, \dots, ϕ_p when $\Psi \neq \mathbf{0}$ is very complicated. Hsu treated the asymptotic problem as $N_t \rightarrow \infty$ such that N_t/N_s , $t, s = 1, \dots, k$, and Ψ remain fixed. Then the elements of $(1/N^{\frac{1}{2}})(\mathbf{A} - \Sigma - \Psi)$ and $(1/N^{\frac{1}{2}})(\mathbf{B} - \Sigma)$ have a joint limiting normal distribution. If the roots of (5) are different and positive, then the properly normalized elements of \mathbf{D}_ϕ or \mathbf{D}_θ and \mathbf{W} , have a joint asymptotic normal distribution.

Hsu treated the more general situation where the roots of (5) can have arbitrary multiplicity and some may be 0. Let the different nonzero roots of (5) be $\lambda_1 > \dots > \lambda_\nu > 0$ with multiplicities m_1, \dots, m_ν , respectively; then the root 0 has multiplicity $p - \sum_{j=1}^\nu m_j$. Let

$$(6) \quad \zeta_i = N^{\frac{1}{2}}(\phi_i - \lambda_h) / (2\lambda_h^2 + 4\lambda_h)^{\frac{1}{2}},$$

$$i = m_1 + \dots + m_{h-1} + 1, \dots, m_1 + \dots + m_h, \quad h = 1, \dots, \nu,$$

$$(7) \quad \zeta_i = N\phi_i, \quad i = m_1 + \dots + m_\nu + 1, \dots, p.$$

Then the $\nu + 1$ sets of normalized roots are independent in the limiting distribution. The density for a set corresponding to $\lambda_h > 0$ is a constant times

$$(8) \quad e^{-\frac{1}{2} \sum_{i=1}^m x_i^2} \prod_{i=1}^m \prod_{j=i+1}^m (x_i - x_j)$$

for $x_1 > \dots > x_m > 0$ ($m = m_h$). The density for the set corresponding to the zero root is a constant times (4) where p is replaced by the number of zero roots of (5) and n is replaced by the sum of $k - 1$ and the number of zero roots of (5) (if $k - 1 \geq p$). As we shall see, the provision for zero roots is essential to treat problems of rank; the inclusion of multiple positive roots is for mathematical generality. The treatment of such generality required great ingenuity and mathematical technique. The corresponding treatment of \mathbf{D}_ϕ (or \mathbf{D}_θ) and simultaneously \mathbf{W} was done by Anderson (1951b).

If $\mathbf{A} = \sum_{\alpha=1}^N (\mathbf{X}_\alpha - \bar{\mathbf{X}})(\mathbf{X}_\alpha - \bar{\mathbf{X}})'$ and Σ are partitioned into p_1 and p_2 rows and columns,

$$(9) \quad \mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

the sample and population canonical correlations are the roots of

$$(10) \quad \begin{vmatrix} -\lambda \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & -\lambda \mathbf{A}_{22} \end{vmatrix} = 0, \quad \begin{vmatrix} -\lambda \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & -\lambda \boldsymbol{\Sigma}_{22} \end{vmatrix} = 0,$$

respectively. In [11] Hsu finds the asymptotic distribution of the normalized sample roots for arbitrary multiplicities of the canonical correlations including 0. This paper parallels [10].

In [9] Hsu considers Wilks' likelihood ratio criterion, equivalent to $W = |\mathbf{B}|/|\mathbf{A} + \mathbf{B}|$, and Lawley's trace criterion $V = \text{tr } \mathbf{A}\mathbf{B}^{-1}$ for testing the null hypothesis that $\mu_1 = \dots = \mu_k$. He showed that $NV + N \log W \rightarrow 0$ in probability if $N\Psi$ as defined above has a limit. In effect this implies that the tests behave similarly when the alternatives approach the null hypothesis at the rate of the reciprocal of the square root of the sample sizes. An implication of the reduction to canonical form [13] is that this conclusion holds for the general linear hypothesis. The problem considered in [14] can be stated in the above terms. That μ_1, \dots, μ_k lie on an l -dimension hyperplane in the p -space is equivalent to Ψ being of rank l or that the number of 0 roots of (5) is $p - l$. To test this hypothesis Fisher (1938) proposed the test criterion $\sum_{j=l+1}^p \phi_j$ (Anderson (1951a) showed that the likelihood ratio test is based on $\prod_{j=l+1}^p (1 + \phi_j)$.) Hsu deduced from [11] that N times Fisher's criterion has a limiting χ^2 -distribution with $(p - l)(k - 1 - l)$ degrees of freedom.

Asymptotic distributions of ratios. In [18] and [22] Hsu applied Cramér's (1937) theory of asymptotic expansions and Berry's (1941) theorem to the distributions of ratios. Let $\bar{Y}_m = \sum_{i=1}^m Y_i/m$ and $\bar{X}_n = \sum_{i=1}^n X_i/n$, where the Y_i 's and X_i 's are independent, Y_1, \dots, Y_m are independently, identically distributed random variables with finite positive absolute moment of some specified order and X_1, \dots, X_n are independently, identically distributed positive random variables with finite positive moment of the same order. The probability distribution of \bar{Y}/\bar{X} can be written

$$(11) \quad \Pr\{\bar{Y}/\bar{X} \leq z\} = \Pr\{-z\bar{X} + \bar{Y} \leq 0\}.$$

Each of the distribution functions of \bar{X} and \bar{Y} suitably normalized can be expanded with the Berry bound on the error. Then by means of the convolution of these two distributions Hsu found an expansion of (11) with a bound on the error.

A serial correlation which is used to test independence of X_1, \dots, X_N normally distributed with the same means and variances can be written as $T = Q/S$, where $Q = \sum_{i,j=1}^N a_{ij}(X_i - \bar{X})(X_j - \bar{X})$ and $S = \sum_{i=1}^N (X_i - \bar{X})^2$. By an orthogonal transformation of X_1, \dots, X_N to Y_1, \dots, Y_N with a change of scale, one can write $Q = \sum_{i=1}^{N-1} \lambda_i Y_i^2$ and $S = \sum_{i=1}^{N-1} Y_i^2$ where under the hypothesis of independence each Y_i has the distribution $N(0, 1)$, $i = 1, \dots, N - 1$. Then

$$(12) \quad \Pr\{T \leq z\} = \Pr\{\sum_{i=1}^{N-1} (\lambda_i - z) Y_i^2 \leq 0\}.$$

If a_{ij} depends on N , then $\lambda_1, \dots, \lambda_{N-1}$ depend on N . Under certain conditions on

these sequences of sets of roots the distribution (12) has an asymptotic expansion as $N \rightarrow \infty$.

In [24], which is a further development of the earlier [15], Hsu considers the limiting distribution of a function $f(\bar{\mathbf{u}}_1, \dots, \bar{\mathbf{u}}_k)$, where $\bar{\mathbf{u}}_1, \dots, \bar{\mathbf{u}}_k$ are independent vector sample means, as the respective sample sizes tend to infinity. On the basis of the central limit theorem applied to the means, and the Taylor expansion of $f(\cdot)$, Hsu obtains in the limit a normal distribution or the distribution of a weighted sum of squares of normal variables (if the variances of the linear terms tend to zero). Hsu applies his general results to obtain the asymptotic distributions of many test criteria, particularly in multivariate analysis.

REFERENCES

- ANDERSON, T. W. (1951a). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *Ann. Math. Statist.* **22** 327–351.
- ANDERSON, T. W. (1951b). The asymptotic distribution of certain characteristic roots and vectors. *Proc. Second Berkeley Symp. Math. Statist. Probability*. Univ. California Press, Berkeley and Los Angeles, 103–130.
- ANDERSON, T. W. (1958). *An Introduction to Multivariate Statistical Analysis*. John Wiley and Sons, New York.
- ANDERSON, T. W. AND GIRSHICK, M. A. (1944). Some extensions of the Wishart distribution. *Ann. Math. Statist.* **15** 345–357.
- BERRY, A. C. (1941). The accuracy of the Gaussian approximation to the sum of independent variates. *Trans. Amer. Math. Soc.* **49** 122–136.
- BOSE, R. C. (1977). Early history of multivariate statistical analysis. *J. Mult. Anal.* **4** 3–22.
- CRAMÉR, H. (1937). *Random Variables and Probability Distributions*. Chap. 7. Cambridge Univ. Press.
- DEEMER, WALTER L. and OLKIN, INGRAM (1951). The Jacobians of certain matrix transformations useful in multivariate analysis. Based on lectures of P. L. Hsu at the University of North Carolina, 1947. *Biometrika* **38** 345–367.
- FISHER, R. A. (1915). Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. *Biometrika* **10** 507–521.
- FISHER, R. A. (1938). The statistical utilization of multiple measurements. *Ann. Eugen.* **8** 376–386.
- FISHER, R. A. (1939). The sampling distribution of some statistics obtained from non-linear equations. *Ann. Eugen.* **9** 238–249.
- GIRSHICK, M. A. (1939). On the sampling theory of roots of determinantal equations. *Ann. Math. Statist.* **10** 203–224.
- LAWLEY, D. N. (1938). A generalization of Fisher's z -test. *Biometrika* **30** 180–187.
- MAHALANOBIS, P. C., BOSE, R. C. and ROY, S. N. (1937). Normalisation of statistical variates and the use of rectangular co-ordinates in the theory of sampling distributions. *Sankhyā* **3** 1–40.
- MOOD, A. M. (1951). On the distribution of the characteristic roots of normal second-moment matrices. *Ann. Math. Statist.* **22** 266–273.
- ROY, S. N. (1939). p -statistics or some generalisations in analysis of variance appropriate to multivariate problems. *Sankyā* **4** 381–396.
- WISHART, JOHN (1928). The generalized product moment distribution in samples from a normal multivariate population. *Biometrika* **20** 32–52.

DEPARTMENT OF STATISTICS
STANFORD UNIVERSITY
STANFORD, CALIFORNIA 94305