The Annals of Statistics
1980, Vol. 8, No. 5, 937–957

THE LIMITING EMPIRICAL MEASURE OF MULTIPLE DISCRIMINANT RATIOS[1]

BY KENNETH W. WACHTER[2]

University of California at Berkeley

Consider the positive roots of the determinental equation $\det|YJY^* - x^2YY^*| = 0$ for a $p(n)$ by n sample matrix of independent unit Gaussians Y with transpose Y^* and a projection matrix J of rank $m(n)$. We prove that the empirical measure of these roots converges in probability to a nonrandom limit F as $p(n)$, $m(n)$, and n go to infinity with $p(n)/n \to \beta$ and $m(n)/n \to \mu$ in $(0, 1)$. Along with possible atoms at zero and one, F has a density proportional to $((x - A)(x + A)(B - x)(B + x))^{\frac{1}{2}}/[x(1 - x)(1 + x)]$ between $A = |(\mu - \mu\beta)^{\frac{1}{2}} - (\beta - \mu\beta)^{\frac{1}{2}}|$ and $B = |(\mu - \mu\beta)^{\frac{1}{2}} + (\beta - \mu\beta)^{\frac{1}{2}}|$. On the basis of this result, tables of quantiles are given for probability plotting of multiple discriminant ratios, canonical correlations, and eigenvalues arising in MANOVA under the usual null hypotheses when the dimension and degree of freedom parameters are large.

This paper determines the asymptotic behavior of the empirical measure of the roots of a determinental equation which sprang into prominence in 1939 when R. A. Fisher, S. N. Roy, P. L. Hsu, and M. A. Girshick simultaneously published papers deriving the joint probability distribution of the roots under a null hypothesis. The roots go under many names, since they figure equally in discriminant analysis, canonical correlation analysis, and in invariant tests of linear hypotheses in the multivariate analysis of variance. We single out discriminant ratios in our title merely for concreteness. The roots can be represented, as we shall see in Lemma 1.1 below, as the singular values of a rectangular submatrix of a Haar-distributed orthogonal matrix, or as the square roots of eigenvalues of a symmetric random matrix often called the MANOVA or multivariate beta matrix. We define them as functions of a p by n dimensional matrix Y of independent unit Gaussians with $p \leqslant n$. Letting Y^* denote the transpose of Y and J denote an n by n dimensional projection matrix with rank m, our roots are the p largest nonnegative solutions $x = L_i$ for $i = 1$ to p to $\det|YJY^* - x^2YY^*| = 0$.

In Theorem 3.1 we prove that the empirical measure of these roots converges in distribution as p, m, and n go to infinity together in a suitable fashion and the limit is a fixed measure with a density proportional, for suitable A and B, to

$$((x + A)(x - A)(B + x)(B - x))^{\frac{1}{2}}/[x(1 - x)(1 + x)].$$

Received September 1978; April 1979.

[1]Research supported in early stages by National Science Foundation Grant SOC-75-15702.

[2]Presently Miller Fellow, University of California

AMS 1970 *subject classifications.* Primary 62H30, 60F05.

Key words and phrases. Discriminant analysis, canonical correlations, MANOVA, asymptotic distributions, random matrix spectra, empirical measure, probability plotting, Wilks' likelihood ratio.

Depending on the dimension limits, there may also be atoms at zero and unity. This result gives a standard against which to compare the multiple discriminant ratios, canonical correlations, or MANOVA eigenvalues from a data sample, where the numbers of observations and variables are both large.

The technique of quantile-quantile plotting using this limit has already been presented in Wachter (1976d), and in Section 5 we present tables of quantiles to facilitate such applications. This probability plotting technique parallels the methods for principal components described in Wachter (1975) and (1976a). Unlike the case of principal components, however, limiting forms under alternative hypotheses to the null hypothesis do not appear accessible, so the methods here are more limited in scope.

The empirical measure of the collection $L_1 \cdot \cdot \cdot L_p$, the pure-point probability measure placing mass $1/p$ at each of these p numbers, represents neatly the sample configuration of the roots. Its distribution function, usually called an "empirical cumulative distribution function" or "e.c.d.f." is a step function with p steps of size $1/p$ at the roots. The expectation of the empirical measure is the marginal distribution of any one of the unordered L_i or of a randomly selected L_i out of the ordered set. It does not of course give as much information for formal hypothesis testing as exact distributions of various of the ordered roots, but in the eminently multivariate case when p, m, and n are all large, formal tests may have low power and more exploratory inference is appropriate. In contrast to methods using approximations to joint densities, probability plotting using the quantiles in Section 5 is simple and direct.

The different roles of our roots are discussed in practically every text on multivariate analysis. We describe in Section 4 the reduction of null hypotheses for multiple discriminant analysis, canonical correlations, and multivariate analysis of variance to our standard form.

The literature on the L_i is extensive, but it has concentrated on functions of them other than their empirical measure, and it has not by and large led to results handy for data analysis. Three recent survey articles summarize the work to date. Expressions for marginal distribution functions of single roots in terms of multiple integrals and for marginal densities in terms of zonal polynomials are treated on pages 165 to 177 of Pillai (1976). Page 145 of Krishnaiah (1978) is a good guide to tables of percentage points computed for low dimensional cases, mostly for p below 8. We compare our results against some of these tables in Section 4. Asymptotic results for fixed p and large n or n and m are helpfully collected on pages 18 to 20 and 24 to 30 of Muirhead (1978).

Calculation of the limiting form in Theorem 3.1 from expectations of moments is joint work of Dr. Colin L. Mallows of Bell Telephone Laboratories and the present author dating from 1969. Proof of convergence and particularly of stochastic degeneracy of the limit is trickier than the simple limiting form would suggest, and it has had to await recent refinements in convergence theorems for random spectra

of matrices with independent elements proved in Wachter (1978). An earlier proof is found in Wachter (1976c).

Section 1 proves an identity between singular values of matrices of different dimensions, a key lemma in the proof of Theorem 3.1. Section 2 sets out the notation for asymptotic assertions and states the results on random spectra on which we depend. The relationship to standard null hypotheses and test statistics becomes our theme in Section 4, and Section 5 presents tables of quantiles for probability plotting based on the limiting empirical measure we have found.

1. A singular value identity. Our first quarry is a representation of our roots as scale factors in an identity between singular values of matrices of differing dimensions.

In order to cope simultaneously with submatrices of differing dimensions, we frame our definition of singular values so as to include extra zeros up to a specified total number of singular values n. We start with a "dimension triple" p, m, and n along with a p by m dimensional (possibly complex) matrix X. Choose n' bigger than $p + m$ and than n and form the n' by n' symmetric matrix with X in its upper-right corner, the (conjugate) transpose X^* in its lower-left, and zeros elsewhere.

$$\begin{bmatrix} 0 & 0 & X \\ 0 & 0 & 0 \\ X^* & 0 & 0 \end{bmatrix}$$

The n' eigenvalues of this matrix are all real and at least $n' - 2\min(m,p)$ of them are zero. Deleting $n' - n$ of the zeros, we call the remaining n eigenvalues of the symmetric matrix the "singular values" of the original matrix X. The phrase "positive singular values" refers to the $\min(p,m)$ largest of these eigenvalues which, together with their negatives, account for all nonzero eigenvalues if X has full rank. Had we chosen instead the popular convention that singular values all be positive, nonvanishing odd moments would plague our convergence theorems for random spectra. The "random spectrum" of X for the triple p, m, and n is the empirical measure of the n singular values of $X/n^{\frac{1}{2}}$.

Our representation lemma is an exercise with Haar distributions on rotation groups, for whose properties Dempster (1969), pages 292-294 and 304, or Loomis (1953), Chapter 6, are ready references.

LEMMA 1.1. *Let L be a p by p diagonal matrix whose diagonal elements are the positive singular values of a p by m submatrix of an n by n random rotation with Haar measure for $p \leqslant m \leqslant n$. Let J be an n by n projection of rank m. Then there exist identically distributed (though not independent) p by n matrices Y and Z such that YJ and Z^*L have the same singular values, Z is independent of L, and the diagonal elements of L are the positive solutions to* $\det|YJY^* - x^2YY^*| = 0$. *We may arrange for Y and Z each to have independent unit normals as elements.*

PROOF. To mediate between dimensions, we define a p by n matrix P by the condition $P_{ij} = 1$ if $i = j$ and $P_{ij} = 0$ else. All matrices in this proof except for P and the sample matrices X, Y, and Z are square. The dimensions are as follows:

$$p \text{ by } p : D, F, G, H, H_0, K, K_0, L$$
$$p \text{ by } n : P, X, Y, Z$$
$$n \text{ by } n : J, Q, R, R_0, T, U, V.$$

Any p by n matrix X has a singular value decomposition $X = H_0 DPR_0$ described among other places on pages 388 to 394 of Golub (1969) and pages 111 to 115 of Chambers (1977). Here H_0 is a p by p rotation whose columns are eigenvectors of XX^*, D is a p by p diagonal matrix of positive singular values of X, and R_0 is an n by n rotation whose columns are eigenvectors of X^*X augmented to a complete orthonormal set.

If we form $Y = GXV$ from X with Haar-distributed random rotations G and V, we find $Y = (GH_0)DP(R_0V)$. Our Haar rotation V has the same distribution as $R_0V = R$ even conditional on G, H_0, D and R_0, purely because R_0 is a rotation independent of V. Likewise G and $H = GH_0$ have the same Haar distribution conditional on H_0, D, and R. Therefore $Y = HDPR$ where H and R are Haar rotations and H, R, and D are independent. Our construction of H and R might seem roundabout, but it neatly avoids technical snags posed by nonuniqueness of singular value decompositions. We now have any random X yielding a matrix $Y = GXV = HDPR$ whose distribution can boast two-sided rotational invariance.

Our key equations are $Y = HDPR$ and a similar decomposition $PRJ = KLPQ$ involving the rotation R. We form this decomposition of PRJ with the aid of another p by p independent Haar rotation F, which we augment to an n by n rotation T by adjoining the $n - p$ dimensional identity and suitable zeros, so that $FPT^* = P$. The Haar distribution of R makes T^*R be Haar independent of T. We take a singular value decomposition $PT^*RJ = K_0LPQ$. Since F is Haar, so is $K = FK_0$, independent of K_0, L, and Q. Thus $PRJ = FPT^*RJ = KLPQ$, where K is Haar independent of L and Q. In our proof we need this equality between PRJ and $KLPQ$ themselves and not merely between their distributions.

The diagonal elements of L are its p positive singular values when $p \leqslant m$, our only reason for requiring $p \leqslant m$. They are then the positive singular values of $PRJ = KLPQ$ since pre-and postmultiplications by rotations do not alter singular values. Premultiplication by P picks out p rows of R. The projection J may not be of the form to pick out m columns, but we can substitute a singular value decomposition of J for J in PRJ. We then find R postmultiplied by a rotation, which preserves its Haar distribution, by a diagonal matrix of zeros and ones which does pick out m columns, and by a rotation which does not change the singular values. Thus the diagonal elements of L are the positive singular values of a p by m submatrix of a Haar rotation, as the condition of the lemma requires.

We show that the diagonal elements of L are the positive solutions to $\det|YJY^* - x^2YY^*| = 0$ by writing down

$$YY^* = HDPRR^*P^*D^*H^* = HDD^*H^* = HDKK^*D^*H^*$$

$$YJY^* = YJJ^*Y^* = HDKLPQQ^*P^*L^*K^*D^*H^* = HDKL^2K^*D^*H^*.$$

The determinant equals $\det|HDK|\det|L^2 - x^2I|\det|K^*D^*H^*|$ and $\det|HDK| \neq 0$ when rank D, which equals rank Y, is equal to p, as occurs with probability one when $p \leqslant n$.

We are now ready to construct Z from YJ. The matrix YJ equals $HDPRJ = HDKLPQ$, which has the same singular values as $DKLP$, since prerotations and postrotations leave singular values unaltered. D, K, and L are all p by p matrices, and postmultiplication of DKL by P merely adds $n - p$ columns of zeros, just as premultiplication of DKL by P^* would add $n - p$ rows of zeros. Therefore $DKLP$ has the same singular values as P^*DKL or as $(U^*P^*DK)L$, where U is an independent n by n Haar-distributed random rotation. But $Z^* = U^*P^*DK$ then has the same distribution as $Y^* = R^*P^*DH^*$, since U and K are n by n and p by p Haar rotations independent of $D = D^*$ and of each other just like R and H^*. We have Z independent of L because U, D, and K are. Thus YJ and Z^*L share the same singular values and Y and Z share the same distribution.

We can make Y and Z each have independent unit normals as elements by starting with X of this form and appealing to invariance of this distribution under rotations on both sides. □

The construction of Y makes its distribution invariant under rotations on both sides but not necessarily Gaussian. We need to make Y Gaussian in Theorem 3.1 not for the YJ and Z^*L relationship but for the convergence theorems for random spectra which require both Y and Z to have independent elements.

The arguments in Lemma 1.1 and throughout this paper remain valid when X, Y, and Z have complex elements, so long as we reinterpret "rotation" as "unitary matrix" rather than "orthogonal matrix" and * as "adjoint" or "conjugate transpose" rather than "transpose" (our reason for using * instead of T). The elements of L are always real. Whenever we begin the proof with a matrix X of identically distributed elements whose real and imaginary parts are general (not necessarily spherical) bivariate normals, we obtain Y_{ij} whose real and imaginary parts are spherical normal. Our proof was specially designed for this bonus. Since we can replace Y by X in $\det|YJY^* - x^2YY^*|$ if we also replace J by the projection VJV^*, we shall be able to replace "a matrix Y of independent unit Gaussians" by "a matrix Y of independent exchangeable complex Gaussians" in the statement of Theorem 3.1. This generalization is noteworthy. Submatrices of Haar orthogonal and Haar unitary matrices have different joint distributions of singular values, so it is surprising to find the asymptotic empirical measures turning out the same.

Applications as described in Section 4 often invoke sample matrices with more than n columns, whose equivalence to the n-column setup we now prove. Our

notation employs double primes to avoid confusion between single primes and a popular symbol for transpose.

LEMMA 1.2. *For any $n'' \geqslant n$ we can start with a p by n'' matrix Y'' of independent unit Gaussians and two n'' by n'' projections N'' and J'' of ranks n and m and obtain J, L, Y, and Z satisfying Lemma* 1.1 *with the diagonal elements of L also being the positive solutions to*

$$\det|Y''(N''J''N'')Y''^* - x^2Y''N''Y''^*| = 0.$$

PROOF. Postmultiplying $Y''N''$ and $J''N''$ by the inverse of the final rotation in the singular value decomposition of N'' leaves matrices whose n'' columns include $n'' - n$ columns of zeros. Deleting these zero columns gives a p by n matrix to call Y and an n'' by n matrix which we premultiply by its transpose to obtain J. Putting $X = G^*YV^*$ in the proof of Lemma 1.1 and citing the invariance of the distributions of matrices of independent unit normals under rotations on both sides, we are left only with elementary details to verify. □

Our representation via the n by n Haar rotation R reveals intriguing symmetry relationships. If we split R into the four rectangular submatrices with sides of dimensions p and $n - p$ by m and $n - m$, the identity $RR^* = R^*R = I$ means that we obtain the squares of the ordered positive singular values of one rectangle by subtracting from unity the squares of the reverse-ordered positive singular values of either adjacent rectangle. Thus our limiting measure to be derived in Section 3 must be invariant under each triple of interchanges $p \leftrightarrow n - p$, $m \leftrightarrow m$, $x \leftrightarrow (1 - x^2)^{\frac{1}{2}}$; $p \leftrightarrow p$, $m \leftrightarrow n - m$, $x \leftrightarrow (1 - x^2)^{\frac{1}{2}}$; and $p \leftrightarrow n - p$, $m \leftrightarrow n - m$, $x \leftrightarrow x$.

One way of phrasing these symmetries is to take along with the empirical measure of positive singular values for one rectangle the measures for the other three rectangles. In the measures for the adjacent rectangles replace x by $(1 - x^2)^{\frac{1}{2}}$ and leave the measure for the opposite rectangle alone. Then our original measure equals a superposition of these four measures, with arbitrary weights summing to one.

Now it turns out, remarkably enough, that we obtain the limit measure of Theorem 3.1 from a set of parallel operations. First replace each rectangle with dimensions r by c from our Haar matrix with a rectangular matrix of independent unit Gaussians of dimensions $r - rc/n$ by $c - rc/n$. Form the limiting random spectra of these four rectangular matrices. Choosing one rectangle, transform the measures for adjacent rectangles by $x \leftrightarrow (1 - x^2)^{\frac{1}{2}}$, and leave the opposite one alone. Then it turns out that the superposition of the absolutely continuous positive parts of these four measures with equal weights scaled to give total mass one produces our limiting empirical measure in Theorem 3.1 in the atomless case. The mass of the term from the r by c rectangle in the superposition is $(\min(r, c) - rc/n)/2p$. We make no use of this characterization in our proof, but it suggests the possibility of a different derivation from convergence theorems for random spectra using suitably clever conditioning and superposition.

2. Paths to the asymptotic limit. In this section we embed the p by m and p by n matrices that figure in the representation in Section 1 into sequences of matrices whose dimensions go to infinity together, and we state the results about asymptotic limits which will enter our proof. We begin with a collection of definitions which hold throughout Sections 2 and 3.

DEFINITIONS 2.1. Y and Z are independent infinite matrices of independent random variables with unit variances, the same means within rows, and a uniform bound on some central moment above the second.

$p(n)$ and $m(n)$ are nondecreasing sequences of positive integers for $n \geqslant 2$ such that $p(n)/n \to \beta > 0$ and $m(n)/n \to \mu > 0$ as n goes to infinity.

σ is any infinite sequence of elements in (0, 1].

$Z\langle\sigma\rangle$ is the matrix with elements $\sigma_i Z_{ij}$.

$K_n\langle\sigma\rangle$ is the empirical measure of $\sigma_1 \cdot \cdots \cdot \sigma_{p(n)}$.

$R_n\langle\sigma\rangle$ is the empirical measure of the singular values of the $p(n)$ by n upper-left submatrix of $Z\langle\sigma\rangle/n^{\frac{1}{2}}$ for the dimension triple p, n, n.

Δ is the measure that puts unit mass at unity.

$M(F, \mu, w)$ is an analytic function of the complex argument w off the real axis for each probability measure F and positive real μ given by

$$M(F, \mu, w) = (2\mu - 1)/(2w) - (1/2)\int_{-\infty}^{\infty} \frac{dF(t)}{t^2 - w}.$$

$W(K, \mu, \beta, z)$ is an analytic function of the complex argument z off the real axis for each probability measure K and positive reals μ and β given by

$$W(K, \mu, \beta, z) = (\mu/z) + \beta\int_0^{\infty} \frac{dK(t)}{(1/t)^2 - z}.$$

$\mathbb{R}$ denotes the real line.

$\mathbb{P}$ denotes probability.

Prob S for a topological space S is the space of Borel probability measures on S with the topology of weak convergence, under which a sequence of measures converges if and only if the expectations they assign to bounded continuous functions converge.

D is any metric for Prob $\mathbb{R}$ which imposes the same sense of uniform convergence as we obtain from neighborhoods of zero in the linear space (containing Prob $\mathbb{R}$) of all bounded linear functionals on the bounded continuous functions on $\mathbb{R}$ with the weak-star topology. For instance, D may be the metric defined on page 59 of Wachter (1974) or page 9 of Wachter (1978).

The sequences $p(n)$, $m(n)$ form a "path of dimensions" if $p(n) + m(n) \leqslant n$, $p(n+1) - p(n) \leqslant 1$ and $m(n+1) - m(n) \leqslant 1$ for $n \geqslant 2$.

Our theorem in Section 3 rests on the following two propositions, which are proved as Theorem 2.1 and Theorem 2.4 of Wachter (1978):

PROPOSITION 2.2. *If $p(n)$, $m(n)$ form a path of dimensions, if $\beta \leqslant \mu$, and if the empirical distribution $K_n\langle\sigma\rangle$ of the row scale factors converges in* Prob $\mathbb{R}$ *to a*

measure K, the random spectrum $R_n\langle\sigma\rangle$ *converges almost surely in* Prob $\mathbb{R}$ *to a fixed, nonrandom measure F determined by the condition for nonreal w that* $W(K, \mu, \beta, M(F,\mu, w)) = w$.

PROPOSITION 2.3. *If* $p(n)$, $m(n)$ *form a path of dimensions and* $G_n\langle\sigma\rangle$ *is the measure determined by the condition for nonreal w that*

$$W(K_n\langle\sigma\rangle, m(n)/n, p(n)/n, M(G_n\langle\sigma\rangle, m(n)/n, w)) = w,$$

then $D(R_n\langle\sigma\rangle, G_n\langle\sigma\rangle)$ *converges in probability to zero uniformly over* σ.

The restrictions on the sequences $p(n)$, $m(n)$ needlessly complicate manipulations with differing dimension triples, so we prove a technical lemma to relax them.

LEMMA 2.4. *The conclusions of Propositions* 2.2 *and* 2.3 *still hold whether or not* $p(n)$, $m(n)$ *are paths of dimensions, so long as they are nondecreasing sequences of positive integers with* $p(n) \leqslant m(n) \leqslant n$ *and* $p(n)/n \to \beta > 0$ *and* $m(n)/n \to \mu > 0$ *as n goes to infinity.*

PROOF. On a graph of the lattice of pairs of positive integers mark the pairs $p(n)$, $m(n)$ and connect them with a broken line. Among all step functions with steps only at integers and only of integral height starting at (0, 1), choose the highest which lies wholly below this broken line. Let $p(n')$, $m(n')$ be an enumeration of the pairs of integers that this step function traverses in order. Then $p(n')$, $m(n')$ forms a path of dimensions and $p(n') + m(n') = n'$. There is a finite nonzero limit to n'/n, namely $\mu + \beta$, and $p(n) \leqslant m(n)$ at the junctures of the broken line implies $p(n') \leqslant m(n')$ along the step-function in between. Furthermore, convergence of $p(n)/m(n)$ to β/μ implies convergence of $p(n')/m(n')$ to the same limit, since the step function falls never more than two units below or to the right of the broken line.

Replacing n by n' under our conventions rescales the singular values by the square root of $g = n/n'$ and increases or decreases the number of singular values fixed at zero by $n' - n = (1 - g)n'$. Now augmenting the atom and rescaling in this fashion converts any empirical measure F into

$$F^{\#g}(t) = gF\left(t(g)^{\frac{1}{2}}\right) + (1 - g)\Delta(t + 1).$$

Then R_n converges to F whenever n/n' converges to g and $R_{n'}^{\#n/n'}$ converges to $F^{\#g}$. We must now check that $M(F^{\#g}, \mu g, wg) = M(F, \mu, w)$ so that $W(K, \mu g, \beta g, M(F^{\#g}, \mu g, wg)) = wg$ is equivalent to $W(K, \mu, \beta, M(F, \mu, w)) = w$. □

3. Convergence of the empirical measure.

THEOREM 3.1. *The empirical measure of the* $p(n)$ *positive solutions x to the determinental equation* $\det|YJY^* - x^2YY^*| = 0$ *for a* $p(n)$ *by n matrix Y of independent unit Gaussians and an n by n projection matrix J of rank* $m(n)$ *converges in*

probability in Prob $\mathbb{R}$ *as n goes to infinity and* $p(n)/n \to \beta$ *and* $m(n)/n \to \mu$ *for* β *and* μ *in* $(0, 1)$. *The limit in distribution is a fixed, nonrandom probability measure with density between* $A = |(\mu - \mu\beta)^{\frac{1}{2}} - (\beta - \mu\beta)^{\frac{1}{2}}|$ *and* $B = |(\mu - \mu\beta)^{\frac{1}{2}} + (\beta - \mu\beta)^{\frac{1}{2}}|$ *given by* $((x - A)(x + A)(B - x)(B + x))^{\frac{1}{2}}/[\pi\beta x(1 - x)(1 + x)]$ *and atoms of size* $\max(0, 1 - \mu/\beta)$ *at zero and size* $\max(0, 1 - (1 - \mu)/\beta)$ *at unity.*

PROOF. By virtue of Lemma 1.1 we can begin with a sequence of independent matrices the nth of which is a Haar-distributed n by n orthogonal matrix. For each n we take a $p(n)$ by $m(n)$ rectangular submatrix of the nth matrix, and let $L_1(n) \cdots L_p(n)$ be its randomly shuffled $p(n)$ largest singular values. We augment each $L(n)$ to an infinite sequence by adjoining, say, values $\mu, \mu, \cdots$ on the right. $L_1(n) \cdots L_p(n)$ are the same as the diagonal elements of L in Lemma 1.1 which proves them equal to the roots of the determinental equation for suitable Gaussian Y when $p \leqslant m$. This claim is also true when $m \leqslant p$, for Lemma 1.1 then holds for the largest m roots and singular values other than one, while the remaining $p - m$ are ones or zeros.

Under Definitions 2.1, $K_n\langle L(n)\rangle$ is the object whose convergence we are trying to prove. We are going to work our way back to this convergence by a circuitous route, and the reasons for our detours are clearer if we see first what would prevent us from working forward if we knew that $K_n\langle L(n)\rangle$ did converge in some sense to a limit K. Then we should like to apply Proposition 2.2 to conclude that $R_n\langle L(n)\rangle$ converged to the limit F determined by $W(K, 1, \beta, M(F, 1, w)) = w$. But we cannot apply Proposition 2.2 because $L(n)$ is random rather than fixed and because $L(n + 1)$ is not an extension of $L(n)$ but a whole new sequence for each n. We are not riding the same sequence down to the finish line; rather, as it were, we are leaping from horse to horse in mid-gallop, changing sequences with every change of n, a circus stunt for which Proposition 2.2 was not designed.

We can at least use the condition $W(K_n\langle L(n)\rangle, 1, \beta, M(G_n\langle L(n)\rangle, 1, w)) = w$ to generate a sequence of measures $G_n\langle L(n)\rangle$ which with luck ought to converge to the solution F to $W(K, 1, \beta, M(F, 1, w)) = w$. We should then like to apply Proposition 2.3 to conclude that $R_n\langle L(n)\rangle$ also converged to F, although we cannot apply Proposition 2.3 without further argument, because $L(n)$ is random rather than fixed. We do, however, at least know from Lemma 1.1 that $R_n\langle L(n)\rangle$ has the same distribution as the random spectrum S_n of the matrix Y for the dimension triple $p(n), m(n), n$, a rather different triple from $p(n), n, n$ with which R_n itself is constructed. Furthermore, we can apply Proposition 2.2 to prove that S_n converges to a limit F determined by $W(\Delta, \mu, \beta, M(F, \mu, w)) = w$, and so we finally arrive at a limit for $R_n\langle L(n)\rangle$ courtesy of S_n.

We are using here the equivalence of the distributions of singular values of the scaled p by n matrix $Z\langle L(n)\rangle$ with the unscaled p by m matrix Y. We can describe our problem as a hunt for the distribution of random scale factors which makes a $p(n)$ by n scaled case look like a $p(n)$ by $m(n)$ unscaled case in the limit. The main fact in our favor as we now try to reverse our path and go back from S_n to $R_n\langle L(n)\rangle$ to $G_n\langle L(n)\rangle$ to $K_n\langle L(n)\rangle$ is the stochastic degeneracy of the limit F.

We begin, then, with the matrix Y and its random spectrum S_n for the dimension triple $p(n)$, $m(n)$, n. Lemma 2.1 tells us that, for all n, S_n has the same distribution as $R_n\langle L(n)\rangle$, the random spectrum of $Z\langle L(n)\rangle$ for the dimension triple $p(n)$, n, n. (If $m < p, p - m$ elements of $L(n)$ are always zero, as are $p - m$ of the p largest singular values that enter S_n.) We write Law S_n = Law $R_n\langle L(n)\rangle$. Almost sure convergence entails convergence in distribution, so Proposition 2.2 makes Law S_n converge in Prob(Prob $\mathbb{R}$) to unit mass at the measure F determined by the condition for nonreal w that $W(\Delta, \mu, \beta, M(F, \mu, w)) = w$. This is the defining condition for F to which we later refer. Strictly speaking, Proposition 2.2 requires $\beta \leqslant \mu$, but in case $\mu < \beta$ we apply it to Y^* with the triple $m(n)$, $p(n)$, n. The simple form of Δ enables elementary algebra to verify that $W(\Delta, \beta, \mu, M(F, \beta, w)) - w = 0$ implies $W(\Delta, \mu, \beta, M(F, \mu, w)) - w = 0$.

We now know that Law $R_n\langle L(n)\rangle$ also converges to unit mass at F. Since the limit is fixed rather than random, convergence in distribution to F entails convergence in probability. In terms of our metric D for Prob $\mathbb{R}$ we are justified in writing

$$\forall \varepsilon, \delta\ \exists N\ \forall n > N\ \mathbb{P}\{D(F, R_n\langle L(n)\rangle) < \delta\} > 1 - \varepsilon.$$

We resort to epsilons and deltas because we must be careful with the uniformity condition in Proposition 2.3 which allows us to pass from fixed sequences σ to random ones $L(n)$. Let $G_n\langle\sigma\rangle$ be the measure determined (as Proposition 2.2 assures us) by the condition for nonreal w that

$$W(K_n\langle\sigma\rangle, 1, \beta, M(G_n\langle\sigma\rangle, 1, w)) = w.$$

The inclusion of the support of $K_n\langle\sigma\rangle$ in $[0, 1]$ restricts the support of $G_n\langle\sigma\rangle$ to at most $[-2, 2]$. We now have by Proposition 2.3

$$\forall \varepsilon, \delta \exists N \forall n > N \forall \sigma\ \mathbb{P}\{D(R_n\langle\sigma\rangle, G_n\langle\sigma\rangle) < \delta\} > 1 - \varepsilon.$$

The same N works for all σ and gives us control over $R_n\langle L(n)\rangle$ conditional on $L(n) = \sigma$. The sequence σ is a fixed sequence. We are conditioning on a different event for every n. But, fortunately for us, we are able to operate with the values of n separately one at a time. Integrating the conditional probability over the distribution of $L(n)$, we have

$$\forall \varepsilon, \delta \exists N \forall n > N\ \mathbb{P}\{D(R_n\langle L(n)\rangle, G_n\langle L(n)\rangle) < \delta\} > 1 - \varepsilon.$$

The triangle inequality now squeezes $D(G_n\langle L(n)\rangle, F)$ below 2δ with probability at least $1 - 2\varepsilon$ for all n above some N, so $G_n\langle L(n)\rangle$ also converges in distribution to F.

Rudimentary weak-convergence theory carries us over the next hurdle in our proof. Writing simply K_n, G_n, and R_n for $K_n\langle L(n)\rangle$, $G_n\langle L(n)\rangle$ and $R_n\langle L(n)\rangle$ from here on, we note that K_n is the empirical measure of a collection of roots in $[0, 1]$. The set of all probability measures concentrated on $[0, 1]$ is a (closed, uniformly tight) compact subset of Prob $\mathbb{R}$. The set of all probability measures in Prob(Prob $\mathbb{R}$) concentrated on this compact subset of Prob $\mathbb{R}$ includes Law K_n for all the random measures K_n and is itself a compact subset of Prob(Prob $\mathbb{R}$).

Thus every subsequence of Law K_n has a subsubsequence convergent to a limit Law K. We want to find Law K and show it to be the same for all subsubsequences. By 4.4 on page 27 of Billingsley (1968) convergence in distribution of G_n to F and of K_n to K through our subsubsequence along with degeneracy of F is equivalent to joint convergence in distribution of K_n, G_n to K, F through the subsubsequence. In order to infer that $M(G_n, 1, W(K_n, 1, \beta, z))$, which equals z for all z in the range of M converges in distribution to $M(F, 1, W(K, 1, \beta, z))$, through the subsubsequence we must prove the last expression to be a jointly continuous function of K and F.

The functions $1/(z - (1/t^2))$ and $1/(w - t^2)$ are bounded continuous functions of t for each nonreal z and w, so the maps which integrate them by $dK(t)$ and $dF(t)$ are continuous maps between Prob $\mathbb{R}$ and $\mathbb{R}$. Thus $W(K, \mu, \beta, z)$ and $M(F, \mu, w)$ are continuous in K and F for fixed nonreal w and z, and simple differencing shows M to be jointly continuous in F and w for nonreal values of w like the values $W(K, 1, \beta, z)$ for z in the range of M. Thus we may assert the desired joint continuity and define K by the condition $M(F, 1, W(K, 1, \beta, z)) = z$ for z in the range of M. The next paragraphs show this condition to determine K uniquely, so that all the subsubsequence limits will coincide.

We possess all the ingredients for the calculation of $W(K, 1, \beta, z)$. In these calculations we always take z to be in the range of M and w to be nonreal. The defining condition for F provides

$$w = W(\Delta, \mu, \beta, M(F, \mu, w)).$$

The formula for M in Definitions 2.1 arranges

$$M(F, \mu, w) = M(F, 1, w) - (1 - \mu)/w$$

so that

$$w = W(\Delta, \mu, \beta, [M(F, 1, w) - (1 - \mu)/w]).$$

Substituting $W(K, 1, \beta, z)$ for w we obtain the horrendous but crucial formula

$$W(K, 1, \beta, z) = W(\Delta, \mu, \beta, [M(F, 1, W(K, 1, \beta, z)) - (1 - \mu)/W(K, 1, \beta, z)]).$$

Now our weak convergence argument defining K supplies

$$M(F, 1, W(K, 1, \beta, z)) = z$$

so the argument in brackets is $y = z - (1 - \mu)/W(K, 1, \beta, z)$. The formula for W in Definitions 2.1 gives

$$W(\Delta, \mu, \beta, y) = \mu/y + \beta/(1 - y).$$

Therefore $W(K, 1, \beta, z)$ satisfies a quadratic equation with coefficients which are quadratics in z. We find

$$W(K, 1, \beta, z) =$$

$$= (1/2z^2)\left(\frac{1 - \beta + 1 - \mu}{1/z} + \frac{1 - \beta - \mu}{1 - 1/z} \pm \frac{((A^2 - 1/z)(B^2 - 1/z))^{\frac{1}{2}}}{(1/z)(1 - 1/z)} \right)$$

with

$$A = |(\mu - \mu\beta)^{\frac{1}{2}} - (\beta - \mu\beta)^{\frac{1}{2}}|$$
$$B = |(\mu - \mu\beta)^{\frac{1}{2}} + (\beta - \mu\beta)^{\frac{1}{2}}|.$$

The proper branch of the square root is the one which makes it equal $\mu - \beta$ at $z = \infty$ and equal $\beta - (1 - \mu)$ at $z = 1$. This claim follows from equating the expression

$$2wM(F, 1, w) = 2 - \int (t^2/(t^2 - w))\, dF(t)$$

from Definitions 2.1 with the expression just derived for $2zW(K, 1, \beta, z)$ at $z = M(F, 1, w)$. The measure F, being the almost sure limit of S, has an atom of mass $1 - \min(2\mu, 2\beta)$ at the origin, so in the expression involving F as w goes to zero along the positive imaginary axis, $M(F, 1, w) = z$ goes to infinity and $2wz$ goes to $2 - \min(2\beta, 2\mu)$. That value has to equal $2 - \beta - \mu \pm (A^2B^2)^{\frac{1}{2}}$. Similarly, in the expression involving F we cannot have w go to infinity when z goes to 1, so $1 - \beta - \mu$ has to equal $-((A^2 - 1)(B^2 - 1))^{\frac{1}{2}}$. We notice parenthetically that $2zW(K, 1, \beta, z)$ goes to 2 as z goes to zero and that $W(K, 1, \beta, z)$, which equals $\overline{W}(K, 1, \beta, \bar{z})$ has positive imaginary part when z has negative imaginary part, and vice versa.

The formula for W in Definitions 2.1 gives

$$2z^2W(K, 1, \beta, z) - (2 - 2\beta)z = 2\beta \int_0^\infty \frac{dK(t)}{(1/z) - t^2}.$$

The expression we have derived for $W(K, 1, \beta, z)$ now supplies in the limit where $1/z = x^2 + iy$ and y decreases to zero a finite limit for

$$\int \frac{dK(t)}{x^2 + iy - t^2} = \frac{\beta - \mu}{2\beta(x^2 + iy)} + \frac{1 - \beta - \mu}{2\beta(1 - x^2 - iy)} + \frac{((A^2 - x^2 - iy)(B^2 - x^2 - iy))^{\frac{1}{2}}}{2\beta(x^2 + iy)(1 - x^2 - iy)}$$

for all finite x, except perhaps zero and unity. This limit has a nonzero imaginary part only for $A^2 < x^2 < B^2$ given there by $((x - A)(x + A)(B - x)(B + x))^{\frac{1}{2}}/(2\beta x(1 - x))$. If $\mu < \beta$ there is a pole at zero with residue $1 - \mu/\beta$ and if $1 - \mu < \beta$ there is a pole at unity with residue $1 - (1 - \mu)/\beta$. We can recover K from the imaginary part of these limits of its transform via a rarely quoted version of Fatou's Theorem for the half-plane (cf. Hoffman (1962), pages 34 and 123) which is not hard to verify directly: For any test function $g(t)$ continuous off a K-nullset, $\int g(t)\, dK(t)$ equals

$$\operatorname{limit}_{y \to 0} (1/2\pi i) \int_0^\infty \left[\int (1/(x^2 - iy - t^2))\, dK(t) - \int (1/(x^2 + iy - t^2))\, dK(t) \right] g(x)(2x)\, dx.$$

It follows that K itself has a density on $[A, B]$ given by

$$\frac{((B^2 - x^2)(x^2 - A^2))^{\frac{1}{2}}}{\pi\beta x(1 - x^2)}$$

along with atoms of size $\max(0, 1 - \mu/\beta)$ at zero and $\max(0, 1 - (1 - \mu)/\beta)$ at unity.

Since every subsequence of K_n has a subsubsequence converging in distribution to the same nonrandom limit K, K_n itself converges in probability to K. □

4. Discriminant ratios, canonical correlations, and MANOVA. We can relate our roots to familiar quantities in linear multivariate analysis concisely through the definition of $L_1 \cdots L_p$ established in Lemma 1.2 as solutions to $\det|YNJNY^* - x^2YNY^*| = 0$ where we now omit the double primes of 1.2 except on n'', letting Y be a p by n'' matrix of independent unit Gaussians, while J and N are n'' by n'' projection matrices of ranks m and n.

In multiple discriminant analysis we start with a sample Y of $n'' = n + 1$ p-vectors divided into m subgroups. The matrix N is the projection onto the complement of the space spanned by any sample grand mean vector and J is the projection onto the space spanned by subgroup sample means. The usual null hypothesis asserts equality of the subgroup means. Lachenbruch (1975), pages 63 to 72, is a standard reference.

For canonical analysis, we start with two samples Y and X of dimensions p by n'' and m by n'' respectively, where $n'' = n + 1$. Let N again denote projection onto the complement of any sample grand mean and for any matrix C let $\text{hat}(C)$ denote projection onto the space spanned by columns of C. Excepting singular cases, $\text{hat}(C) = C(C^*C)^{-1}C^*$, the "hat matrix" of Tukey in Hoaglin and Welsch (1978), page 17. Put $J = \text{hat}(NX^*)$ and observe $NJN = J$. The canonical correlations between Y and X are then the positive solutions to $\det|YJY^* - x^2YNY^*| = 0$ and the usual null hypothesis takes Y and X independent. Anderson (1958), pages 323-324, is a standard reference.

For the MANOVA model expectation $(Y^*) = XS$, where Y is a p by n'' sample, X an n'' by r design matrix of rank r, and S an r by p parameter matrix, consider a null hypothesis of the form $CS = 0$ for an s by r matrix C of rank s. Invariant tests are functions of our roots if we put $N = I - \text{hat}(X) + \text{hat}(X(X^*X)^{-1}C^*)$ and $J = NJN = I - \text{hat}(X)$. Here $n = n'' - r + s$ and $m = n'' - r$. Detailed references are Roy, Gnanadesikan, and Srivastava (1971), pages 32 to 36, and Timm (1975), pages 189 and 371.

The five well-known test statistics called Hotelling's Trace, Pillai's Trace, the Largest and the Smallest Root criteria, and Wilks' Likelihood Ratio are all functions of the empirical measure whose limit we have found. Without wanting to recommend hypothesis testing compared, for instance, to the plotting methods of Section 5, we may still set out consequences of Theorem 3.1 for these statistics.

COROLLARY 4.1 (to 3.1). *Under the null hypotheses covered by Theorem* 3.1 *for* $0 < \beta < \mu < \beta + \mu < 1$ *as* $p/n \to \beta$, $m/n \to \mu$ *and* n *goes to infinity, the following test statistics converge in probability*:

a) *Hotelling's Trace*	$(1/p)\Sigma - 1/(1 - 1/L_i^2)$	*to* $\mu/(1 - \mu - \beta)$
b) *Pillai's Trace*	$(1/p)\Sigma L_i^2$	*to* μ
c) *Largest Root*	maximum L_i^2	*to* $\mu + \beta - 2\mu\beta + 2(\mu\beta(1 - \mu)(1 - \beta))^{\frac{1}{2}}$
d) *Smallest Root*	minimum L_i^2	*to* $\mu + \beta - 2\mu\beta - 2(\mu\beta(1 - \mu)(1 - \beta))^{\frac{1}{2}}$
e) *Wilks' Likelihood Ratio*	$(1/p)\Sigma \log L_i$	*to* $\int \log(x)\, dK$

PROOF. Consider the five functions $\int(x^2/(1 - x^2))\, dF$, $\int x^2 dF$, $F^{-1}(1 - 1/2n)$, $F^{-1}(1/2n)$, and $\int \log(x)\, dF$, functions of n and of a probability measure F. These equal the five test statistics respectively when F is the empirical measure K_n. For $0 < \beta < \mu < \beta + \mu < 1$ there is no sequence of measures converging to the limiting empirical measure K for which the values of the five functions do not converge to their values at $F = K$. This assertion is the condition for applying Theorem 5.5 on page 34 of Billingsley (1968), which establishes convergence in probability. The integrals may be evaluated with help from the transform $\int(-1/(t^2 - 1/z))\, dK$ given at the end of the proof of Theorem 3.1. □

The limiting distributions demanded for significance tests are not the constants of 4.1 but rather the differences between the statistics and these constants, all multiplied by the square root of p, more delicate limits about which we have proved nothing. Yet by and large these centering constants themselves (like the limiting distributions) appear to be known already only when p/n goes to zero. A suggestion by Colin Mallows to treat p singular values like alternate order statistics from an independent sample of size $2p - 1$ from the limiting empirical distribution has met with success in studies by Lurie (1978) of principal component variances, and does generate guesses at significance levels. It is interesting to compare the centering constants and Mallows-style guesses against tabulated levels for the test statistics to address two questions. First, how easily does Theorem 3.1 extend the information in current tables? Second, how close to its asymptotic limit is the empirical measure for the rather small p, m, and n which current tables treat? Space constraints preclude a systematic study here, but three examples are instructive.

The rescaled version $((n + m - p - 1)/(mp - np))\Sigma \log L_i$ of Wilks' Likelihood Ratio is often compared against a chi-square variable divided by its mean with degrees of freedom equal to $p(n - m)$. For large p, m, and n Corollary 4.1 shows this scaling to be wrong unless p/n goes to zero. The factor $((1 + \mu - \beta)/(1 - \mu))\int \log(x)\, dF$ which connects the means of the scaled Wilks' statistic and the scaled chi-square does not equal unity unless $\beta = 0$. Exhibit 4.2 compares this factor against the factor connecting the ninetieth percentiles of the two statistics tabulated by Schatzoff (1966). His parameters p, q, M, and n equal our p, $n - m$,

EXHIBIT 4.2.
Comparisons on adjustment factors for Wilks' likelihood ratio

p	m	n	β	μ	Schatzoff 90% factor	$\frac{1+\mu-\beta}{1-\mu}\int \log(x)\, dK$
8	9	17	.471	.529	1.270	1.211
8	10	18	.444	.556	1.185	1.150
8	11	19	.421	.579	1.138	1.115
8	12	20	.400	.600	1.108	1.091
8	17	25	.320	.680	1.046	1.040
8	27	35	.229	.771	1.017	1.015
8	67	75	.107	.893	1.003	1.002
7	8	18	.389	.444	1.285	1.223
7	9	19	.368	.474	1.197	1.160
7	10	20	.350	.500	1.148	1.123
7	11	21	.333	.524	1.117	1.099
7	16	26	.269	.615	1.051	1.045
7	26	36	.194	.722	1.019	1.017
7	66	76	.092	.868	1.003	1.003

$m + 1 - p$, and m respectively. We choose p and $n - m$ values among the highest Schatzoff lists. The factors agree within 4% for β/μ below .80, suggesting 4.1 as a reasonable surrogate for Schatzoff's tables beyond the p, m, and n he treats.

Our second comparison ventures still further toward low dimensions, considering percentiles of the smallest and largest root given by Chang (1974) for $p = 5$. Chang's parameters p, m, and n equal our p, $(m - p - 1)/2$, and $(n - m - p - 1)/2$ respectively. The 90th percentile of the square of the smallest root from Chang, labeled 1:90%, is juxtaposed against an asymptotic percentile labeled K:22.6%. This is the squared 22.6th percentile from the limiting empirical measure, the 22.6th percentile of any distribution being the 95th percentile for the smallest of $2p - 1 = 9$ independent observations from it. Similarly, Chang's 95th percentile for the squared smallest root (1:95%) is juxtaposed against the asymptotic percentile K:28.3%. The differences between Chang's values and the asymptotic percentiles are substantially smaller than the differences between 90th and 95th percentiles, except for overshoot at 95% in the first and third rows. The last three columns of Exhibit 4.3 compare the squared upper bound B^2 for the support of the limiting empirical measure against the 95% and 99% quantiles for the squared largest (fifth) root, labeled 5:95% and 5:99%. Here the upper tail of the largest root distribution creeps beyond the asymptotic upper bound, but the size of the excess considering the low dimension is hardly discouraging.

EXHIBIT 4.3.
Comparisons on smallest and largest roots

p	m	n	β	μ	1:90%	K:22.6%	1:95%	K:28.3%	B^2	5:95%	5:99%
5	6	16	.313	.375	.088	.106	.113	.146	.902	.929	.961
5	36	46	.109	.783	.664	.664	.689	.697	.978	.985	.992
5	6	32	.156	.188	.035	.043	.045	.060	.569	.607	.676
5	36	62	.081	.581	.451	.454	.473	.482	.836	.857	.885

Among the tabulations in Krishnaiah and Schuurmann (1974) are various upper percentiles for all six roots of the complex MANOVA matrix when $p = 6$. The applicability of Theorem 3.1 to the complex case which we have discussed in Section 2 after Lemma 2.1 makes these tables relevant, though the medians of the six roots would make for better comparisons than the upper percentiles. Exhibit 4.4 gives the 95th percentiles for each root for four cases, followed by the p-value from 3.1 for each case, that is, by the probability that the limiting empirical measure assigns to values less than the tabled percentile point. The bottom line gives the same probability for the 95th percentiles of alternate order statistics from a sample of size $2p - 1 = 11$. Considering the variety in the percentile values, the close match down each column in the p-values argues that important structure for dimensions as low as 6 is already being captured by the asymptotic form. The alternate order statistic guesses, however, fit badly, at least this far out into the tails.

EXHIBIT 4.4.
Comparisons on upper percentiles of ordered roots in complex case

			95% quantiles of ordered squared roots					
p	m	n	#1	#2	#3	#4	#5	#6
6	6	17	.0444	.1646	.3356	.5324	.7281	.8972
6	21	32	.4389	.5796	.7025	.808	.8958	.9624
6	6	32	.0190	.0730	.1571	.2694	.4095	.5877
6	21	47	.2630	.3633	.4627	.563	.6669	.7799
β		μ	p-values from limiting empirical measure					
.353		.353	.183	.359	.529	.695	.855	.993
.188		.656	.194	.366	.533	.696	.856	.993
.188		.188	.183	.360	.530	.696	.857	.994
.128		.447	.195	.367	.535	.699	.859	.995
alternate order stats.			.238	.470	.650	.800	.921	.995

5. Quantiles of the limit measure for probability plotting. Probability plotting of multiple discriminant ratios, canonical correlations, and MANOVA eigenvalues offers a substitute for formal hypothesis testing when the range of interesting alternative hypotheses is hard to restrict and the kind of divergence from a null hypothesis, not merely the fact of divergence, is at issue. Methods based on our limiting empirical measure are described in Wachter (1976d). They apply when the dimension p and the two degree of freedom parameters m and $n - m$ are all comparably large, so that the asymptotic empirical measure is a plausible approximation to the expected empirical measure for the finite parameter values. The comparisons in Section 4 suggest that the methods begin to be useful when p, m, and n are about 10, but detailed simulations would be desirable to indicate more clearly the rapidity of convergence.

The definitions of the parameters p, m, and n and of the roots L_i appropriate to null hypotheses in each area of application are found at the start of Section 4. The

singular value decompositions in the proof of Lemma 1.1 are generally efficient ways of calculating the roots (cf. Chambers (1977), Chapter 5). If roots are output from packaged programs, it is essential to put them in a form equivalent to that in Section 4 or transform the empirical measure accordingly. Many programs reverse m and $n - m$ or output L_i^2, not the L_i with which we are dealing. The choice is purely conventional, our unsquared roots having the virtue of being on the familiar scale of correlations. The roots given in Wachter (1976d) are squares of the roots given here. Exhibit 5.1 shows quantiles of the limiting empirical measure for cases of β and μ satisfying $\beta \leqslant \mu < \mu + \beta \leqslant 1$ and so having no atoms. If we interchange β and μ, putting $\beta' = \mu$, $\mu' = \beta$ so that $\mu' < \beta'$, we create an atom of size $(\beta' - \mu')/\beta'$ at zero, but the quantiles of the distribution conditional on values greater than zero, which are the quantiles of real interest, remain unchanged. The further symmetries mentioned after Lemma 1.2, namely under interchanges

$$\beta \leftrightarrow 1 - \beta, x \leftrightarrow (1 - x^2)^{\frac{1}{2}}; \mu \leftrightarrow 1 - \mu, x \leftrightarrow (1 - x^2)^{\frac{1}{2}};$$

$$\beta \leftrightarrow 1 - \beta, \mu \leftrightarrow 1 - \mu$$

EXHIBIT 5.1
Quantiles of the limiting empirical measure of 3.1

F	μ = .50 β = .10	.50 .20	.50 .30	.50 .40	.50 .50
.00	.447	.316	.204	.101	.000
.04	.492	.380	.279	.179	.063
.08	.519	.420	.329	.235	.125
.12	.542	.455	.374	.288	.187
.16	.563	.487	.415	.339	.249
.20	.583	.517	.455	.388	.309
.24	.602	.546	.493	.435	.368
.28	.619	.573	.529	.481	.426
.32	.636	.599	.564	.526	.482
.36	.653	.625	.598	.569	.536
.40	.669	.649	.631	.611	.588
.44	.684	.673	.662	.651	.637
.48	.699	.696	.692	.689	.685
.52	.714	.718	.721	.725	.729
.56	.729	.739	.749	.759	.771
.60	.743	.760	.776	.792	.809
.64	.757	.781	.801	.822	.844
.68	.771	.800	.826	.850	.876
.72	.785	.819	.848	.876	.905
.76	.798	.838	.870	.900	.930
.80	.812	.856	.890	.922	.951
.84	.826	.873	.910	.941	.969
.88	.840	.890	.927	.957	.982
.92	.854	.907	.944	.972	.992
.96	.870	.925	.960	.984	.998
1.00	.894	.949	.979	.995	1.000

F	$\mu = .60$ $\beta = .10$	.60 .20	.60 .30	.60 .40
.00	.535	.410	.302	.200
.04	.579	.476	.383	.291
.08	.605	.516	.434	.352
.12	.628	.550	.479	.407
.16	.648	.582	.521	.458
.20	.666	.610	.559	.506
.24	.648	.638	.596	.552
.28	.700	.663	.630	.596
.32	.716	.688	.663	.638
.36	.731	.711	.694	.678
.40	.745	.733	.724	.715
.44	.759	.755	.752	.751
.48	.773	.775	.779	.784
.52	.786	.795	.804	.815
.56	.799	.813	.828	.844
.60	.811	.831	.850	.871
.64	.823	.848	.871	.895
.68	.835	.865	.891	.917
.72	.847	.880	.909	.936
.76	.858	.895	.925	.953
.80	.870	.909	.940	.967
.84	.881	.923	.954	.979
.88	.892	.935	.966	.988
.92	.904	.948	.976	.995
.96	.917	.960	.985	.999
1.00	.935	.976	.994	1.000

make our tables apply to all parameter combinations, so long as the atoms at zero when $\mu < \beta$ and at one when $1 < \mu + \beta$ are borne in mind.

In order to compare a set of roots for multivariate samples against their approximate expected empirical measure, we choose the column in Exhibit 5.1 with β and μ nearest to p/n and m/n. We then plot the ith smallest of the p observed roots $L_{(i)}$ on the y-axis against the value X_i on the x-axis which is the table entry for which $F(X_i)$ is closest to $i/(p + 1)$ or we interpolate for greater accuracy. Points falling away from the 45 degree line through the origin indicate divergence from the null hypothesis. An example of the application of these methods to actual data is given in Wachter (1976d), pages 832-833. That paper also discusses cases in which traditional practices searching for sets of large roots or gaps in the set of roots as clues to divergence from the null hypothesis can prove thoroughly misleading, as contrasted with the probability plotting methods advocated here.

The computation of quantiles in Exhibit 5.1 capitalizes on the super position representation of K described at the end of Section 1 to utilize general programs for quantiles of limiting random spectra already developed for principal components in Wachter (1976a). The quantiles have been checked for accuracy of one unit in the third decimal place by relying on the following result:

COROLLARY 5.2 (to 3.1). *The limiting measure of Theorem* 3.1 *when* $0 < \beta \leqslant \mu < \mu + \beta \leqslant 1$ *has a distribution function given by*

$$K(x) = (1 - \mu)K_0(\mu - \mu\beta, \beta - \mu\beta, x) + \mu\left(1 - K_0\left(1 - \beta - \mu + \mu\beta, \mu\beta, (1 - x^2)^{\frac{1}{2}}\right)\right.$$

where

$$K_0(a, b, x) = (1/\pi b)\left\{x(a - x^2U^2(a, b, x))^{\frac{1}{2}} - (a) \arccos\left(xU/(a)^{\frac{1}{2}}\right)\right.$$
$$\left. + (b) \arccos\left(x(U - 1)/(b)^{\frac{1}{2}}\right)\right\}$$

and $U(a, b, x) = (1 + (a - b)/x^2)/2$ *with range of the* arc cosine *between zero and* π.

The mean, variance, and third central moment of x^2 *under this distribution are* μ, $\beta\mu(1 - \mu)$, *and* $\mu\beta^2(1 - \mu)(1 - 2\mu)$ *respectively.*

PROOF. By differentiation and contour integration. □

F	$\mu = .70$ $\beta = .10$	.70 .20	.70 .30
.00	.621	.503	.400
.04	.663	.569	.484
.08	.688	.609	.537
.12	.709	.642	.581
.16	.727	.672	.622
.20	.745	.699	.659
.24	.760	.725	.693
.28	.775	.748	.725
.32	.789	.770	.755
.36	.803	.791	.784
.40	.815	.811	.810
.44	.828	.830	.834
.48	.839	.847	.857
.52	.851	.864	.878
.56	.861	.879	.898
.60	.872	.894	.915
.64	.882	.907	.931
.68	.892	.920	.946
.72	.901	.932	.958
.76	.910	.943	.969
.80	.919	.953	.979
.84	.928	.962	.986
.88	.937	.971	.992
.92	.945	.978	.996
.96	.954	.985	.999
1.00	.967	.993	1.000

F	$\mu = .80$ $\beta = .10$	.80 .20
.00	.707	.600
.04	.747	.665
.08	.770	.703
.12	.789	.735
.16	.806	.763
.20	.821	.788
.24	.835	.811
.28	.848	.832
.32	.860	.851
.36	.872	.869
.40	.882	.885
.44	.892	.901
.48	.902	.915
.52	.911	.927
.56	.919	.939
.60	.927	.950
.64	.935	.959
.68	.942	.968
.72	.949	.975
.76	.955	.982
.80	.962	.987
.84	.967	.992
.88	.973	.995
.92	.978	.998
.96	.983	.999
1.00	.990	1.000

F	$\mu = .90$ $\beta = .10$
.00	.800
.04	.836
.08	.857
.12	.873
.16	.888
.20	.900
.24	.912
.28	.922
.32	.931
.36	.940
.40	.948
.44	.955
.48	.961
.52	.967
.56	.973
.60	.978
.64	.982
.68	.986
.72	.989
.76	.992
.80	.995
.84	.997
.88	.998
.92	.999
.96	.999
1.00	1.000

Unfortunately, few authors have been accustomed to quote the values of the full set of roots when they perform discriminant or canonical correlation analysis or MANOVA tests. This lack precludes us from giving a good selection of illustrative examples in the present paper. It is to be hoped that quoting all the roots will become standard practice, so that good examples can be collected and agreement with null hypotheses reviewed on the basis of the asymptotic results.

The scope of our probability plotting methods would be enhanced if we possessed asymptotic forms for the empirical measure of our roots under alternatives to the null hypothesis, like those for principal components in Wachter (1975), (1976a) and (1976b) studied in Lurie (1978). But the obstacles to proving anything once we surrender the two-sided rotational invariance of the distribution of the matrix Y in 3.1 are formidable. Even the strengthening of Theorem 3.1 from convergence in probability to almost sure convergence in the null case appears nontrivial. Nor is the problem of limiting forms particularly amenable to simulation, because so many cases of the parameter values demand investigation. It may therefore be that further progress will have to await new breakthroughs in the theory of random matrix spectra.

REFERENCES

ANDERSON, THEODORE W. (1958). *An Introduction to Multivariate Statistical Analysis*. Wiley, New York.

BILLINGSLEY, PATRICK (1968). *Convergence of Probability Measures*. Wiley, New York.

CHAMBERS, JOHN M. (1977). *Computational Methods for Data Analysis*. Wiley, New York.