The Annals of Probability
1978, Vol. 6, No. 1, 1-18

THE STRONG LIMITS OF RANDOM MATRIX SPECTRA FOR SAMPLE MATRICES OF INDEPENDENT ELEMENTS[1]

BY KENNETH W. WACHTER

Harvard University

This paper proves almost-sure convergence of the empirical measure of the normalized singular values of increasing rectangular submatrices of an infinite random matrix of independent elements. The limit is the limit as both dimensions grow large in some ratio. The matrix elements are required to have uniformly bounded central $2+\delta$th moments, and the same means and variances within a row. The first section (relaxing the restriction on variances) proves any limit-in-distribution to be a constant measure rather than a random measure, establishes the existence of subsequences convergent in probability, and gives a criterion for almost-sure convergence. The second section proves the almost-sure limit to exist whenever the distribution of the row variances converges. It identifies the limit as a nonrandom probability measure which may be evaluated as a function of the limiting distribution of row variances and the dimension ratio. These asymptotic formulae underlie recently developed methods of probability plotting for principal components and have applications to multiple discriminant ratios and other linear multivariate statistics.

0. Introduction and notation. The limiting distributions of the singular values of rectangular random matrices have as interesting applications to multivariate statistics as the well-known limiting distributions of eigenvalues of square symmetric random matrices have to nuclear physics. The latter work, primarily connected with Wigner's "Semicircle Law," described in books by Mehta (1967), Porter (1965) and Bharucha-Reid (1972) pages 86–88, has been reviewed for statisticians in the *Sixth Berkeley Symposium* by Olson and Uppuluri (1973). Recent papers of Ludwig Arnold (1973) and (1976) and Rudolf Wegmann (1976a) and (1976b) treat with rigour many aspects of square symmetric and Hermitian random matrix eigenvalues.

The singular values of rectangular multivariate sample matrices of concern to statisticians and the "random spectra" of these matrices, the random empirical measures of their normalized singular values, do not fall under the theorems motivated by the physics applications. It is true that the squares of the singular values of mean-centered sample matrices are eigenvalues of square symmetric sample covariance matrices. But elements of the square matrices are not independent when elements of the rectangular matrices are, and so the natural assumptions under the square and rectangular formulations differ. Previous study

Received August 9, 1976; revised March 17, 1977.

[1] Portions of this work were supported by The Master and Fellows of St. Catherine's College, Oxford, and by the National Science Foundation under grant MPS75-09835.

AMS 1970 *subject classifications*. Primary 60F15; Secondary 62H25.

Key words and phrases. Random matrix spectra, singular values, limit distributions, principal components.

of rectangular samples with independent elements has been confined to very special cases. For situations where variances of all elements are equal and all moments exist, limits of expected random spectra due to Stein have been reported by Olson and Uppuluri (1973) and proofs of convergence of random spectra have been announced by Mallows and Wachter (1970). This paper tackles convergence of random spectra for rectangular matrices of independent elements in the more general case where means and variances may vary from row to row and only a bound on the moments of order $2 + \delta$ for some $\delta > 0$ is imposed.

Section 1 of this paper proves existence and stochastic degeneracy of subsequence limits of random spectra. Section 2 marshals this degeneracy in proofs of formulae for limiting forms and almost-sure convergence to them. The degeneracy of asymptotic random spectra is known to be a ticklish matter and demands great care with the details of proofs. The complicated arguments which follow, however, are rewarded by results believed to be the strongest known of their kind. They have already served as the basis for methods of probability plotting for principal components in Wachter (1975) and (1976a) and for discriminant ratios and canonical correlations in Wachter (1976c) and (1976d).

The combinatorial labour of the proofs is expended showing that the assumptions of independence turn any limit-in-distribution of the random spectrum into a constant rather than a random probability measure. This degeneracy is equivalent by Corollary 3.2 of Wachter (1974b) to the asymptotic independence of the singular values. A simple zero-one law proved in Wachter (1974a) Theorem 6.1 forces degeneracy of any limit-in-probability of random spectra for matrices with merely independent columns, but problems of measurability make that law insufficient for present purposes.

The asymptotic limit dealt with here is the limit when both p and m go to infinity and their ratio approaches a finite, nonzero constant α, where p and m are the row and column dimensions of the sample matrix, or the "dimension" and "degrees of freedom," or, following Dempster (1969) pages 6–8, the number of variables and the number of individuals. This asymptotic situation must be sharply distinguished from the much simpler limit where p remains fixed and m goes to infinity alone, discussed for instance in Anderson (1963).

Most of the previous work dealing with our type of asymptotics has concentrated on cases of convergence to one special limit, Wigner's Semicircle Law. It tends to convey the misleading impression that this measure is a kind of canonical limit like the Gaussian distribution for partial sums. The great variety of limiting measures derived in this paper shows how special and restricted the semicircle cases are. Arnold (1973) and, as reported there, W. Schmale have likewise identified a broad family of limits for square symmetric matrices. This paper, though very different in substance, owes much to the spirit of that work.

Theorem 1.1 was originally proved by more complicated methods in the author's Ph. D. dissertation, Cambridge, 1974. Conclusions along the lines of Theorem 2.1 were first obtained under more stringent hypotheses jointly by

Colin L. Mallows of Bell Telephone Laboratories and the present author, then also at B.T.L. The key equivalence asserted in Lemma 2.4 is largely due to Dr. Mallows. The other lemmas and proof of the theorems from them are new.

Depending upon context, Z will denote either an infinite-dimensional matrix of real or complex random variables Z_{ij}, or else the rectangular p by m dimensional matrix which is the upper-left-hand corner of this infinite dimensional random matrix, with elements Z_{ij}, $i = 1, \cdots, p$; $j = 1, \cdots, m$. Z^* is the conjugate transpose of Z.

Given $p + m \leqq n$ we form the $n \times n$ square symmetric matrix

$$\begin{pmatrix} 0 & 0 & Z \\ 0 & 0 & 0 \\ Z^* & 0 & 0 \end{pmatrix}$$

whose set of n eigenvalues, all real, is symmetric about the origin. We define the singular values of Z for p, m and n to be these eigenvalues and the random spectrum of Z to be the empirical measure of the n singular values of $Z/n^{\frac{1}{2}}$. Authors who insist that singular values be positive differ from our convention. The distribution function corresponding to the random spectrum is a random step function with a step of size $1/n$ at each singular value of $Z/n^{\frac{1}{2}}$, counting multiplicities. There is a step of size at least $1 - \min(m, p)/n$ at zero.

A "path of dimensions" consists of two nondecreasing sequences of positive integers $p(n)$ and $m(n)$ such that $p(n) + m(n) \leqq n$ with $p(n+1) - p(n) \leqq 1$ and $m(n+1) - m(n) \leqq 1$ and, as $n \to \infty$, $p(n) \to \infty$ and $m(n) \to \infty$. Such a path picks out a sequence of submatrices of Z. We keep the normalization n separate from p and m with a view toward the applications to multiple discriminants in Wachter (1976b). It is necessary to restrict the jumps in $p(n)$ and $m(n)$ for Proposition 1.5 below.

E denotes expectation and $\mathbb{P}$ denotes probability.

The random spectrum is a random element in the space Prob $\mathbb{R}$ of probability measures on the real line with the topology of weak convergence and the Borel sigmafield. A sequence of measures converges in Prob $\mathbb{R}$ if the expectations they assign to bounded continuous functions converge. The probability distributions of random spectra for different n are elements of Prob (Prob $\mathbb{R}$). For definiteness and convenience we introduce a metric D for Prob $\mathbb{R}$ in the preamble to Proposition 1.4 below. In terms of this metric we say that two random matrices Y and Z "have the same limiting random spectra with probability one" if for every path of dimensions their random spectra R_n and S_n, whether or not they themselves converge, still satisfy $D(R_n, S_n) \to 0$ as $n \to \infty$ with probability one.

1. Degeneracy of limit spectra. Our object now is to prove a portmanteau theorem on degeneracy:

THEOREM 1.1. *Let Z be an infinite matrix of real or complex-valued random* *variables which satisfy for all* $i, j = 1, 2, \cdots$: Z_{ij} *independent*; $EZ_{ij} = c_i$ *regardless of* j; $E|Z_{ij} - c_i|^{2+\delta} < A^2$ *for some* $\delta > 0$ *and some* $A > 1$. *Let the nondecreasing sequences* $p(n)$ *and* $m(n)$ *with* $p(n) + m(n) \leqq n$ *form a "path of dimensions" and let* R_n *be the random spectrum of the* p *by* m *dimensional upper-left submatrix of* Z. *Then*

(1.1.1) *Any limit-in-distribution of a subsequence of* R_n *has a degenerate (non-random) distribution.*

(1.1.2) *Any fixed element in the space* Prob $\mathbb{R}$ *to which a subsequence of* R_n *converges in distribution is a probability measure determined by its moments.*

(1.1.3) *Some subsequence of any subsequence of* R_n *does converge in probability to such a limit* F *in* Prob $\mathbb{R}$.

(1.1.4) R_n *converges almost surely in* Prob $\mathbb{R}$ *as* n *goes to infinity through the full sequence or a subsequence* τ *if and only if the expectations of the positive real random variables* $H(n, k)$ *in Lemma* 1.2 *below converge for each* k *as* $n \to \infty$ *through* τ. *Then* $F = \lim R_n$ *if and only if* $\int x^k\, dF = \lim EH(n, k)$ *for all* k.

PROOF. We assume Lemmas 1.2 and 1.3 and Propositions 1.4 and 1.5 below and proceed with the proof of the theorem from them. We begin by truncating Z_{ij}. Put $T(j) = 3A((2 + \log j)^3/\log j)^{1/\delta}$ for $j > 2$ and $T(1) = T(2) = 3A$. $T(j)$ is nondecreasing and $\sum (1/n)^2 \sum_{j=1}^{n} 1/T^\delta(j)$ is finite. It is possible to prove Theorem 1.1 with the less delicate truncation $(i + j)^{\frac{1}{4}}$, but the property $T^{2k}(n)/n \to 0$ for all k as n goes to infinity greatly eases our task. Proposition 1.4 allows us to replace our original random matrix by a random matrix Z with independent elements with $|Z_{ij}| < T(j)$, with zero means, with new variances differing from the old by less than $T^{-\delta}(j)$ and with the same limiting random spectra with probability one. From here on we assume these conditions.

We prove Theorem 1.1 by examining the random moments $\int x^{2k}\, dR_n = (2/n) \operatorname{tr} (ZZ^*/n)^k$ of the random measure R_n. By our convention on the signs of singular values, all odd moments vanish identically, and we ignore them. The key to Theorem 1.1 is a bound of order $1/n$ on the variance of $\int x^{2k}\, dR_n$. Lemma 1.3 bounds this variance by $(1/n)(A^{4k} + T^{4k}(n)/n)(2k)^{2k+2}$.

Foreseeing the role of arbitrary subsequences in 1.1.1, we allow ourselves to start with any sequence τ of increasing integers. Define a subsequence $N_1, N_2, \cdots$ of τ by $N_{t+1} = \inf\{n \text{ in } \tau : n > (1 + 2/t)N_t\}$ with $N_0 = 1$. Note for future reference that $1 \leqq n/\sup\{N_t : N_t < n\} < 1 + 2/t \to 1$ as $n \to \infty$ through τ. Raabe's ratio test on page 355 of Hardy (1955) keeps $\sum_{t=1}^{\infty} 1/N_t$ finite. Inasmuch as $T^{4k}(n)/n \to 0$ for all k, $\sum_{t=1}^{\infty} \operatorname{Var} (\int x^{2k}\, dR_{N_t})$ is finite. Hence $\int x^{2k}\, dR_{N_t} - E \int x^{2k}\, dR_{N_t}$ converges almost surely to zero as $t \to \infty$ for every k.

We have already essentially disposed of the randomness of our random spectra, but we must take care to assemble facts about the moment problem relating $\int x^{2k}\, dR_n$ to R_n in the right order. First, any time that $E \int x^k\, dR_n$ approach limits for all k as $n \to \infty$ through τ, these limits must be the moments of a distribution determined by its moments. For, Lemma 1.2 obliges us with the inequality $|E \int x^{2k}\, dR_n| < (kAA)^k + k^k T^{2k}(n)/n$ so that

$$\limsup_{k\to\infty} (1/k)(\limsup_{n\to\infty} E \int x^{2k}\, dR_n)^{1/2k} = \limsup_{k\to\infty} (kAA)^{k/2k}/k$$

which equals zero. This familiar condition, as justified for instance on page 182 of Breiman (1968), guarantees uniqueness of any measure whose moments are limits of $E \int x^k\, dR_n$. Given uniqueness, existence follows, since such limits would equal the limits of moments, namely of the moments $\int x^k\, dR_{N_t}$ for any fixed sample point outside the null set where convergence fails.

We now know that if $E \int x^{2k}\, dR_n$ converges as $n \to \infty$ through τ, R_{N_t} converges almost surely to a constant measure F determined by its moments. We want to extend convergence from the sequence N_t back to the original subsequence τ. Since we have arranged $\sup \{N_t \colon N_t < n\}/n \to 1$, Proposition 1.5 tells us that R_n then converges to F almost surely through τ.

To complete the proof of 1.1.4, note that the random variables $H(n, 2k)$ whose properties we establish in Lemma 1.2 have expectations with the same limits as $E \int x^{2k}\, dR_n$ since $|E \int x^{2k}\, dR_n - EH(n, 2k)| < k^k T^{2k}(n)/n \to 0$ as $n \to \infty$.

The variables to which Lemma 1.2 is being applied are of course those based on the matrix Z after truncation. But we can prove that $EH(n, 2k)$ for the original Z have the same limits if any, since we have provided in Proposition 1.4 that the variance of the old Z_{ij} is greater but no more than $T^{-\delta}(j)$ greater than that of the new Z_{ij}. The difference of the EH's is less than

$$(2/n) \sum_{s=1}^{m} T^{-\delta}(s)((k-1)AA)^{k-1}k$$

which goes to zero as n goes to infinity since $T^{-\delta}(j)$ decreases and $\sum (1/n)^2 \sum_1^n T^{-\delta}(j)$ is finite.

To prove 1.1.3, suppose we are handed some subsequence of increasing integers. By Lemma 1.2, $EH(n, 2k)$ is confined to the compact interval $[-(kAA)^k, (kAA)^k]$. For each k and so, by diagonal selection, for all k simultaneously, we can find a subsequence τ for which $EH(n, 2k)$ converge. Then R_n converges almost surely and so in probability to a constant limit F as $n \to \infty$ through τ.

The obvious equality between any limit-in-distribution for a subsequence and the limit of a subsequence finally proves 1.1.1 and 1.1.2. □

LEMMA 1.2. *For each n and k let*

$$H(n, 2k) = (2/n^{k+1}) \sum |Z_{s_1 s_2}|^2 |Z_{s_3 s_4}|^2 \cdots |Z_{s_{2k-1} s_{2k}}|^2 .$$

The summation ranges over all 2k-tuples $s_1 \cdots s_{2k}$ *(writing* $s_0 = s_{2k}$*) satisfying, for* $i = 1, \cdots, k$, $s_{2i-1} \leqq p$ *and* $s_{2i} \leqq m$, *such that the cardinality of the set of distinct values of* s_{2i-1} *plus the cardinality of the set of distinct values of* s_{2i} *equals* $k+1$, *and such that there is a permutation* π *of* $1, \cdots, k$ *for which each pair* $\langle s_{2i-1}, s_{2i}\rangle$ *equals* $\langle s_{2\pi(i)-1}, s_{2\pi(i)-2}\rangle$.

Then $\int x^{2k}\, dR_n = H(n, 2k) + H' + H''$, *where*

$$EH(n, 2k) \leqq (kAA)^k$$
$$EH'' = 0$$
$$|H'| < (k^k/n) \sup\{|Z_{ij}|^{2k} : i \leqq p, j \leqq m\}$$

H' being the sum of less than $k^k n^k/2$ terms each bounded by $(2/n^{k+1}) \sup\{|Z_{ij}|^{2k}\}$.

PROOF. We shall elucidate the somewhat murky combinatorial characterization of H with a geometric construction below. The moment $\int x^{2k}\, dR_n = (2/n) \operatorname{tr}(ZZ^*/n)^k = (2/n^{k+1}) \sum X(s)$ where

$$X(s) = Z_{s_1 s_2} \bar{Z}_{s_3 s_2} Z_{s_3 s_4} \bar{Z}_{s_5 s_4} \cdots Z_{s_{2k-1} s_{2k}} \bar{Z}_{s_1 s_{2k}}$$

and the summation over s spreads over $2k$ separate indices $s_1 \cdots s_{2k}$. For odd i, s_i is a row index running from 1 to $p(n)$. For even i, s_i is a column index running from 1 to $m(n)$. We denote the set of all such indices by 'pm'^{k}. Given s in 'pm'^{k} we define positive integers b and w. The integer b equals the number of distinct integer values which occur among the k row subscripts $s_1, s_3 \cdots s_{2k-1}$. The integer w equals the number of distinct integer values which occur among the k column subscripts $s_2, s_4 \cdots s_{2k}$. Define $H' = (2/n^{k+1}) \sum X(s)$ with a summation over all s such that $b + w \leqq k$ and $H'' = \int x^{2k}\, dR_n - H - H'$.

We first show that $EX(s)$ vanishes if $b + w \geqq k + 2$, a step which is the crux of our proof and disposes of most of our terms. Our claim becomes perspicuous when for each $2k$-tuple s we construct a graph to represent the pattern of interlocking coincident indices. Writing a line of integers from 1 to p and below them a line of integers from 1 to m, we connect an integer r on the top line to an integer c on the bottom line if and only if there exists i among $1, 3, 5, \cdots, 2k - 1$ such that $s_i = r$ and $s_{i+1} = c$ or $s_{i-1} = c$, in other words if and only if the factor Z_{rc} or $\bar{Z}_{rc}$ occurs in the term $X(s)$.

Notice that the graph we have constructed must be a connected graph. For, by traversing the nodes labelled $s_1, s_2, s_3 \cdots s_{2k}, s_1$ in that order, alternating between the top and bottom lines of nodes, we eventually visit every integer that occurs as a subscript, and each jump from one integer to another passes over a branch or link which corresponds to a pair r, c and a factor Z_{rc} or $\bar{Z}_{rc}$.

The number of nodes touched by branches is $b + w$. There are $2k$ factors in all, but any factor which occurs unrepeated and without its conjugate makes $EX(s)$ vanish, since different elements Z_{ij} are assumed independent and $EZ_{ij} = E\bar{Z}_{ij} = 0$. For $EX(s)$ to be unequal to zero, an index pair r, c or a factor Z_{rc} must occur never or at least twice, so there must be at most k distinct factors affiliated to at most k branches on our constructed graph. Because the graph is connected, the maximum number of nodes it can have is $k + 1$, so either $b + w \leqq k + 1$ or else $EX(s) = 0$.

A connected graph with the maximum number of nodes on k branches is a tree. A tree has no loops, so that if $s_i = r$ and s_j is the first return to r after i, we must have $s_{i+1} = s_{j-1}$. There always is a first return, when we traverse $s_i, s_{i+1} \cdots s_{2k}, s_1 \cdots s_i$. Thus for those terms, for each factor Z_{rc} which occurs, the factor $\bar{Z}_{rc}$ follows it before any new occurrence of some $Z_{rc'}$. Thus those terms have the form $X(s) = |Z_{s_1 s_2}|^2 \cdots |Z_{s_{2k-1} s_{2k}}|^2$. Recapitulating, they are the terms with $b + w = k + 1$ and in general with $EX(s) \neq 0$, in other words, terms for which the factors are matched by their conjugates, so that there is a permutation π for which each pair $\langle s_{2i-1}, s_{2i} \rangle$ equals $\langle s_{2\pi(i)-1}, s_{2\pi(i)-2} \rangle$, and $b + w = k + 1$. They are exactly the terms in $H(n, 2k)$, and the terms with $EX(s) = 0$ and $b + w \geqq k + 1$ are exactly the terms in H''.

Crude upper bounds on numbers of terms are easy. We can choose b integers out of $1, 2, \cdots, p$ in $\binom{p}{b}$ ways and assign each of $s_1, s_3 \cdots s_{2k-1}$ one of them in b^k ways; similarly for m and w. Recalling Stirling's formula and $p + m \leqq n$, we find no more than $n^{k+1}k^k$ terms with $b + w = k + 1$. The power n^{k+1} cancels the divisor in $n^{-k-1} \sum X(s)$ to make $EH(n, 2k) \leqq (kAA)^k$.

Similar calculations disclose no more than $k^k/2$ terms with $b + w \leqq k$. Thanks to the truncation, each is a product of $2k$ factors each bounded by $\sup \{|Z_{ij}|\}$. Hence $|H'| < (k^k/n) \sup \{|Z_{ij}|\}$. □

LEMMA 1.3. *If Z is a random matrix with independent columns, the variance of $\int x^{2k}\, dR_n$ is less than*

$$(2k^2/n)E^+ \int x^{4k}\, dR_n + E^+ \int x^{2k}\, dR_n\, E^+\left(\int x^{2k}\, dR_n - H(n, 2k)\right) .$$

Here the operator E^+ applied to a sum $\sum X(s)$ of products of elements of Z is given by $E^+ \sum X(s) = \sum |EX(s)|$, and $H(n, 2k)$ is as in Lemma 1.2.

If all the elements of Z are independent with zero means and $|Z_{ij}| < T(j)$ for increasing $T(j)$, the variance is less than $(1/n)(A^{4k} + T^{4k}(n)/n)(2k)^{2k+2}$.

PROOF. The bound which requires independence only of matrix columns is of separate interest for the study of degeneracy of limiting random spectra: column independence alone rules out random spectra with tractable moments and non-degenerate limits.

We use the notation of Lemma 1.2 to write $\int x^{2k}\, dR_n = (2/n^{k+1}) \sum X(r)$ for r ranging over 'pm'^{k};

$$\mathrm{Var}\left(\sum X(r)\right) = \sum\nolimits_L EX(r)X(s) + \sum\nolimits_L (-1)EX(r)EX(s)$$

where $L = \{r, s\colon \text{there exist } a, b \text{ such that } r_{2a} = s_{2b}\}$. L includes all pairs r, s with some column index in common. The covariance between other $X(r)$ and $X(s)$ pairs vanishes, because products of elements from disjoint sets of columns of Z are independent by assumption.

It will be easy to prove both sums over L real, and if we knew that the second sum were also positive, we could discard it, but all we shall need is to discard from it some obvious positive terms. We can write the second sum as a sum over all r in 'pm'^{k} as follows:

$$\sum -EX(r)E \operatorname{tr}\left((ZZ^*)^k - (Z(I - D(r))Z^*)^k\right)$$

where $(D(r))_{ij} = 0$ unless there exists $a: i = j = r_{2a}$, when $(D(r))_{ij} = 1$. We may assure ourselves that $(Z(I - D(r))Z^*)^k$ contains all the terms in $(ZZ^*)^k$ except those terms in which an element of a column represented in $X(r)$ appears. The trace of this difference of self-adjoint matrices must be real, and the same value of the trace factor multiplies $X(r)$ as multiplies $X(r') = \bar{X}(r)$, where $r_i' = r_{2k-i}$; thus the whole sum is real. The difference $ZZ^* - Z(I - D(r))Z^* = ZD(r)Z^*$ is nonnegative definite. We can simultaneously diagonalize $(ZZ^*)^k$ and $(Z(I - D(r))Z^*)^k$ to prove the trace of their difference nonnegative, following Dempster (1969), Section 5.1, or its generalization to complex matrices. We infer that the coefficient of every $EX(r)$ is positive.

Since each $X(r)$ included in H is positive by Lemma 1.2 and has $b + w = k + 1$, we achieve

$$\textstyle\sum_L (-EX(r)EX(s)) \leqq \sum_{r:b+w\neq k+1} (-EX(r) \sum_{s:r,s \text{ in } L} EX(s))$$

which is smaller than $\sum |EX(r)| \sum |EX(s)|$ with the same ranges of summation, and so less than $E^+(\int x^{2k}\, dR_n - H)E^+ \int x^{2k}\, dR_n$.

Alternatively, under the stronger hypotheses of all Z_{ij} independent with zero means and $|Z_{ij}| < T(j)$ we can recall the arguments of Lemma 1.2 to count how few s can satisfy $\langle r, s\rangle$ in L. If r has w_r distinct column subscripts, there are fewer such s than

$$\textstyle\sum w_r \binom{m-1}{w-1} w^k \binom{p}{b} b^k$$

with the summation over w and $b: w + b = k + 1$. Now $|EX(r)EX(s)| < T^{4k}(n)$ and rapid calculations give $\sum_L -EX(r)EX(s) < k^{2k+1}T^{4k}(n)/2n^2$.

We turn now to the apparently more formidable sum $\sum_L EX(r)X(s)$ in which terms $X(r)$ and $X(s)$ with zero expectations themselves can join to yield nonzero contributions. We notice that our index set

$$L = \textstyle\bigcup_{a=1}^k \bigcup_{b=1}^k L_{ab}$$

where $L_{ab} = \{r, s: r_{2a} = s_{2b}\}$. The sets L_{ab} are not disjoint, but each term $EX(r)X(s)$ for r, s in L_{ab} corresponds to a different term in the expansion of the higher-order moment $\int x^{4k}\, dR_n$. In particular, define t in 'pm'^{2k} by

$$t_{4k} = t_0 = r_{2a} = s_{2b}\,; \qquad t_1 = s_{2b+1} \cdots t_{2k-1} = s_{2b-1}\,;$$
$$t_{2k} = s_{2b} = r_{2a}\,; \qquad t_{2k+1} = r_{2a+1} \cdots t_{4k-1} = r_{2a-1}\,.$$

We loop around the chain of indices s beginning and ending on the common column index s_{2b} which equals r_{2a} and brings us onto a similar loop around the chain of indices r back to our starting point, the common column index. We have $X(t) = X(r)X(s)$. The k^2 sets L_{ab} give us at most k^2 copies of any one term from $X(t) = (n^{2k+1}/2) \int x^{4k}\, dR_n$. Applying the operator E^+ in order that missing terms from the $4k$th moment only add to the sum, we may be gratified that the divisor n^{2k+1} on this moment compared to the divisor $(1/n^{2k+1})^2$ wins us the divisor of n which appears in the bound.

Under the stronger hypotheses that allow us to cite Lemma 1.2, $\sum_L EX(r)X(s) < (1/n)(2k)^{2k+2}(A^{4k} + T^{4k}/n)$. □

We come now to Propositions 1.4 and 1.5, on which our proof of Theorem 1.1 depends. Proposition 1.4 covers our truncation. Proposition 1.5 covers the generalization from subsequence limits to sequence limits. We sketch their proofs, omitting certain technical details found in Wachter (1974a), 5.7 and 5.9. The key idea is to define a metric D for the space Prob $\mathbb{R}$ specially suited to calculations with random spectra. We define the metric with reference to a sequence of test functions h_i by the sum over $i = 1, 2, \cdots$,

$$D(F, G) = \sum |\int h_i \, dF - \int h_i \, dG|/2^i \, .$$

For our test functions we take any countable collection uniformly dense in the set of (bounded Lipschitz) functions from $\mathbb{R}$ to $[0, 1]$ satisfying $|h(x) - h(y)| \leqq |x - y|$ for all x and y. That such a choice exists and does impose on Prob $\mathbb{R}$ the topology of weak convergence follows from Dudley (1966) or Wachter (1974a), 3.6 and 3.9.

PROPOSITION 1.4. *Let Y be an infinite random matrix with $EY_{ij} = c_i$ and $E|Y_{ij} - c_i|^{2+\delta} < A^2$. Then there exists a random matrix Z whose element Z_{ij} depends only on Y_{ij} and satisfies $EZ_{ij} = 0$, $E|Z_{ij}|^{2+\delta} < A^2$, $|Z_{ij}| < T(j)$ and $|E|Y_{ij} - c_i|^2 - E|Z_{ij}|^2| < T^{-\delta}(j)$. Here T is any nondecreasing positive function bounded below by $3A$ which keeps $\sum (1/n)^2 \sum_{j=1}^{n} T^{-\delta}(j)$ finite. Let σ_{ij} be positive constants and $p(n)$, $m(n)$ a path of dimensions determining two sequences of random spectra R_n and S_n for the matrices with elements $\sigma_{ij} Y_{ij}$ and $\sigma_{ij} Z_{ij}$. Then*

$$\mathbb{P}\{\lim_{n\to\infty} D(R_n, S_n) = 0\} = 1 \, .$$

Furthermore, for all $b > 0$ and $c > 0$ we can find N such that for all $n > N$ and all σ_{ij} such that $\sup \sigma_{ij} < A$,

$$\mathbb{P}\{D(R_n, S_n) < b\} > 1 - c \, .$$

SKETCH OF PROOF. The last conclusion amounts to the claim that R_n and S_n converge to each other uniformly in probability. This claim is in fact independent of the choice of metric, since Prob $\mathbb{R}$ inherits its topology from a linear topological space, namely, the dual space of the bounded continuous functions with the weakstar topology.

Notice that no assumption of independence of matrix elements is required.

We call a transformation from an infinite matrix Y to an infinite matrix Z "allowable" if for every path of dimensions the conclusions of the proposition hold for their random spectra. Allowability is a transitive relation.

We first claim that transforming Y_{ij} to $Y_{ij} - c_i$, which amounts to subtracting a matrix of rank one from Y, is allowable.

If U and V are two $p(n)$-by-$m(n)$ dimensional matrices and F and G are the empirical measures of the singular values of $U/n^{\frac{1}{2}}$ and $Z/n^{\frac{1}{2}}$ respectively, then the inequalities of Ky Fan (1951) allow us to write

$$\sup_x |F\{(-\infty, x)\} - G\{(-\infty, x)\}| < (1/n)\,\mathrm{rank}\,(U - V)\,.$$

The Lipschitz bound on the test functions defining our metric D then guarantees that, for any constant $a > 0$,

$$D(F, G) \leqq (1/n)(2a + 3)\,\mathrm{rank}\,(U - V) + 2F\{|x| > a\}\,.$$

We refer to this result as the "rank inequality."

Putting $U_{ij} = Y_{ij}$ and $V_{ij} = Y_{ij} - c_i$ and $a = n^{\frac{3}{4}}$ we have $D(R_n, S_n) < 3n^{-\frac{1}{4}} + 2n^{-\frac{3}{2}} \int x^2\, dR_n$. Since $\sum n^{-\frac{3}{2}} E \int x^2\, dR_n < A^2 \sup \sigma_{ij}^2 \sum n^{-\frac{3}{2}}$ which is finite, $D(R_n, S_n)$ converges almost surely to zero. Convergence in probability uniform over the choice of the σ_{ij} follows in the same fashion via Chebychev's inequality.

At each step, besides proving allowability, we must verify that our moment conditions on the matrix elements are preserved. The calculations are laborious but elementary and may be found in Wachter (1974a), pages 90–95.

For the allowability of the transformation from Y to Z which sets $Z_{ij} = Y_{ij}$ if $|Y_{ij}| < T(j)/2$ and $Z_{ij} = 0$ else, we need a generalization to singular values of rectangular matrices of the eigenvalue bound for diagonal perturbations of square matrices on page 96 of Arnold (1971). The perturbation estimate (21) of Wielandt (1955) applied to the absolute value function and the ordered singular values, along with the Lipschitz bound on our test function, shows that any two $p(n)$ by $m(n)$ matrices U and V with random spectra F and G satisfy

$$D^2(F, G) \leqq (2/n^2) \sum\sum |U_{ij} - V_{ij}|^2\,.$$

We refer to this result as the "square sum" inequality. Suppose W_n is the right-hand side of this inequality when $U_{ij} = Y_{ij}\sigma_{ij}$ and $V_{ij} = Z_{ij}\sigma_{ij}$. We claim that W_n converges almost surely to zero. W_n is almost surely a Cauchy sequence because the sum over n of $|W_n - W_{n-1}|$ is almost surely finite since the sum of $E|W_n - W_{n-1}|$ is finite. The last fact holds and the limit of W_n equals zero because the sum of EW_n/n is finite thanks to the bound $E|Y_{ij}|^{2+\delta} < A^2$ and the constraint on the double sum of $1/T^\delta(j)$. Since multiplication of Y_{ij} and Z_{ij} by σ_{ij} cannot increase the bound on W_n by more than a factor of $\sup \sigma_{ij}^2$ we not only have almost sure convergence of $D(R_n, S_n)$, but convergence in probability uniform over the choice of σ_{ij}. Thus our truncation is allowable.

The truncation may lead to nonzero means for matrix elements, but the square sum inequality plus Kronecker's lemma make removal of means an allowable transformation, and loosen our bound on Z_{ij} at worst from $T(j)/2$ to $T(j)$. □

PROPOSITION 1.5. *If* ζ *and* τ *are increasing sequences of integers such that* $\sup\{j$ *in* $\zeta : j < n\}/n \to 1$ *as* $n \to \infty$ *through* τ, *then the convergence of* R_n *in* Prob $\mathbb{R}$ *through* ζ *entails convergence in* Prob $\mathbb{R}$ *through* τ.

SKETCH OF PROOF. Consider $D(R_n, R_r)$ where $r = \sup\{j \text{ in } \zeta : j < n\}$. Define H to be the random spectrum of a $p(n)$ by $m(n)$ dimensional matrix which agrees when $i \leqq p(r)$ and $j \leqq m(r)$ with the matrix whose random spectrum is R_n and which has zeros elsewhere. The rank inequality adduced in the proof of 1.4 gives a bound on $D(R_n, H)$. The Lipschitz property of the test functions chosen for our metric D gives a bound on $D(H, R_r)$. Putting the bounds together yields $D(R_n, R_r) \leqq (8 + 3a)(n - r)/2n + 4R_r\{|x| > a\}$. Since by assumption R_r converges in Prob $\mathbb{R}$ as $r \to \infty$ through ζ, $\{R_r : r \text{ in } \zeta\}$ is uniformly tight. For any $\varepsilon > 0$ we can find a value $a > 0$ to make the last term less than ε for all r. Our hypothesis about r/n then makes $D(R_n, R_r) \to 0$ as $n \to \infty$ through τ. □

2. Almost sure convergence. Now that we know what kind of object the limiting random spectrum is—a fixed measure determined by its moments—we can proceed to find the limits and prove convergence to them. We do so by calculating limiting moments by enumeration of terms as functions of matrix element variances. We obtain a broad family of alternative limits to which, we prove, the random spectra of the matrices almost surely converge.

The chief additional assumptions over Theorem 1.1 in Theorem 2.1 are that elements Z_{ij} in the same row have the same variance and that the distribution of row variances approaches a limit as the number of rows goes to infinity. A transform of the limit F of R_n of the form $\int (z - x)^{-1}\, dF$ turns out to be a function of the same transform of the limiting variance distribution.

THEOREM 2.1. *Let Z be an infinite random matrix satisfying Z_{ij} independent; $EZ_{ij} = c_i$ regardless of j; $E|Z_{ij} - c_{ij}|^{2+\delta} < A^2$ for some $\delta > 0$ and some $A > 1$; and $E|Z_{ij} - c_i|^2 = \sigma_i^2$ regardless of j.*

Let R_n be the random spectrum at n defined by a path of dimensions $p(n)$, $m(n)$ such that $p(n)/n \to \beta$ and $m(n)/n \to \mu$ as n goes to infinity with $0 < \beta/\mu = \alpha < 1$.

Suppose there is a probability measure K such that for all k the averages $(1/p) \sum_1^p \sigma_i^{2k}$ converge to $\int t^{2k}\, dK(t)$. Then R_n converges almost surely in the weak topology on Prob $\mathbb{R}$ *to a fixed, nonrandom probability measure F on the interval $[-A(2\mu + 2\beta)^{\frac{1}{2}}, A(2\mu + 2\beta)^{\frac{1}{2}}]$ whose distribution function $F(t)$, continuous except at zero, is given off zero by*

$$F(t) = \lim_{y \to 0} (1/\pi i) \int_{-\infty}^{t} x(M(x - iy) - M(x + iy))\, dx$$

where $M(z)$ is the unique analytic function off the real axis smaller than $1/2A^2$ at $z = \infty$ and satisfying

$$z^2 = \frac{\mu}{M} + \beta \int \frac{dK(t)}{(1/t^2) - M}\,; \tag{2.1.1}$$

$$M(z) = \frac{2\mu - 1}{2z^2} - \tfrac{1}{2} \int \frac{dF(t)}{t^2 - z^2}\,. \tag{2.1.2}$$

PROOF. The almost sure convergence of R_n to a degenerate limit F is by 1.1.4 equivalent to the convergence for each k of $EH(n, 2k)$ as defined in 1.2. We postpone until Lemma 2.2 the proof that these numbers do converge and devote this proof to identifying the limit measure F.

We pinpoint F by an expression involving its moments. Lemma 2.2 implies that if $m/n \to \mu$ and $p/n \to \beta$ and for all k $(1/p) \sum_1^p \sigma_i^{2k} \to \int x^{2k}\, dK$, then

$$\int x^{2k}\, dF = \sum_a 2\phi(a)\mu^{k+1} \left(\frac{\beta}{\mu} \int x^2\, dK\right)^{a(1)} \left(\frac{\beta}{\mu} \int x^4\, dK\right)^{a(2)} \cdots \left(\frac{\beta}{\mu} \int x^{2k}\, dK\right)^{a(k)}$$

in which the sum is taken over k-tuples $a(1) \cdots a(k)$ of nonnegative integers such that $\sum ja(j) = k$ and in which the coefficient $\phi(a)$ is the number of trees of a special type identified in 2.3. We can enumerate these trees. Suppose M is a formal power series with arguments $y, b_1, b_2, \cdots$ such that the coefficient $\phi(a)$ of $y^{k+1} b_1^{a(1)} b_2^{a(2)} b_3^{a(3)} \cdots$ in $M - y$ is the number of alternating bichromatic rooted plane trees with black root and $a(j)$ black nodes of valency j and k branches for $k = \sum ja(j)$. It is proved in Mallows and Wachter (1972) that this enumerator satisfies the formal relation $M = y(1 + b_1 M + b_2 M^2 + b_3 M^3 + \cdots)$. This relation is the key result from which the theorem follows when we reinterpret M to be an analytic function rather than a formal power series.

The rest of our argument resembles a jigsaw puzzle, fitting together interlocking pieces. We discover some things about M only from properties of F, for instance that M is a transform of a positive measure. On the other hand, we discover some things about F only from properties of M, for instance that $\int (z - x)^{-1}\, dF$ has a Taylor expansion about infinity. It takes some prudence to avoid building a circular argument.

Regarded as a formal power series in y, M has coefficients which are polynomials in $b_1, b_2, \cdots$ determined uniquely by successive substitution for M in the right-hand side of the relation. Any analytic function of y which satisfies the relation in a neighborhood of zero for appropriate numerical choices of $b_1, b_2, \cdots$ must have the numerical values of these same coefficients in its Taylor expansion about zero. Given values of b_j smaller than αA^{2j} we find such an analytic function in the region $|y| < 1/2A^2(1 + \alpha)$. Writing $h(M) = (1 + b_1 M + b_2 M^2 + \cdots)$, we notice that $h(M)$ is an analytic function of M bounded by $1 + \alpha A^2|M|/(1 - A^2|M|)$ inside the circle $|M| = 1/(2A^2)$. As long as $|y| < 1/2A^2(1 + \alpha)$, the inequality $|yh(M)| < |M|$ holds along the contour $|M| = 1/2A^2$, so that within this contour Rouché's theorem (cf. Ahlfors (1966), page 152) provides for fixed y a unique root $M(y)$ for $M - yh(M) = 0$ and $M(y)$ is an analytic function. Inserting the coefficients from the formal power series M, we conclude that the Taylor expansion $M(y)$ about zero must equal its formal expansion

$$y + \sum_{k=1}^{\infty} y^{k+1} \sum_{a: \sum ja(j) = k} \phi(a) b_1^{a(1)} b_2^{a(2)} \cdots b_k^{a(k)}\ .$$

The values $b_j = (\beta/\mu) \int \sigma^{2j}\, dK(\sigma)$ satisfy the requirement $b_j \leqq \alpha A^{2j}$ so we may substitute them for the parameters in $M(y)$. Then

$$h(M) = 1 + (\beta M/\mu) \int \frac{dK(\sigma)}{(1/\sigma^2) - M}$$

for $|M| < 1/2A^2 < A^2$ and putting $y = \mu/z^2$, the equation defining M takes the form

$$z^2 = \frac{\mu}{M} + \beta \int \frac{dK(\sigma)}{(1/\sigma^2) - M} \tag{2.1.1}$$

where $M(\mu/z^2)$ is the unique and analytic solution inside $|z| > A(2\mu + 2\beta)^{\frac{1}{2}}$ such that $|M| < 1/2A^2$.

We now interpret the Taylor expansion of $M(\mu/z^2)$ about $z = \infty$ in terms of moments of F. For $k \geqq 1$, twice the coefficient of $1/z^{2k+1}$ in $zM(\mu/zz)$ equals $\int x^{2k}\, dF$ as evaluated in 2.3. For $k = 0$, twice the coefficient of $1/z$ in $zM(\mu/zz)$ is 2μ instead of $1 = \int dF$. Hence

$$2zM(\mu/zz) - (2\mu - 1)/z = \sum_0^\infty (1/z)^{2k+1} \int x^{2k}\, dF\,.$$

This Taylor expansion must converge on the region $|z| > A(2\mu + 2\beta)^{\frac{1}{2}}$ where M is analytic, and this convergence has consequences for F. It forces the limit as $k \to \infty$ of

$$(1/2AA(\mu + \beta))^k \int x^{2k}\, dF \geqq F\{|x| > A(2\mu + 2\beta)^{\frac{1}{2}}\}$$

to go to zero. Hence F is concentrated on the interval $[-A(2\mu + 2\beta)^{\frac{1}{2}}, A(2\mu + 2\beta)^{\frac{1}{2}}]$. It follows that

$$\int \frac{dF(x)}{z - x}$$

has a Taylor expansion about $z = \infty$, which, now that it is shown to exist, must be the expansion $\sum (1/z)^{2k+1} \int x^{2k}\, dF$, which by 1.1.2 is a series which determines F.

The values of the unique root M on $|z| > A(2\mu + 2\beta)^{\frac{1}{2}}$ smaller than $1/2A^2$ of equation (2.1.1) determine the unique F for which (2.1.2) holds:

$$M(\mu/zz) = \frac{2\mu - 1}{2zz} + \frac{1}{2z} \int \frac{dF(x)}{z - x} = \frac{2\mu - 1}{2zz} + \tfrac{1}{2} \int \frac{dF(x)}{z^2 - x^2}\,.$$

But this equation then defines values of M for other z, indeed values of M such that $\bar{M}(\mu/zz) = M(\mu/\bar{z}\bar{z})$ for all z off the real support of F. Are these values also solutions of (2.1.1)?

Define

$$W(M) = \mu/M + \beta \int \frac{dK(t)}{(1/t^2) - M}\,.$$

W is analytic off the real axis, hence over the range of $M(\mu/zz)$ for z off the real axis. Thus $W(M(\mu/zz))$ is analytic off the real axis. Since it agrees with z^2 for $|z|^2 > 2AA(\mu + \beta)$, it must agree with z^2 for all nonreal z. Hence the transform $M(\mu/zz)$ of F does satisfy (2.1.1) for all nonreal z. It is not the only solution of (2.1.1), but it is the direct analytic continuation of the solution less than $1/(2AA)$ at infinity.

(2.1.1) implies that F has no atoms except at zero. For, if F had an atom of size q at $\pm a$, then at $z^2 = a^2 + iy^2$, by (2.1.2), $-\operatorname{Im}(M(\mu/(a^2 + iy^2))) > q/y + (\mu - \frac{1}{2})y/(a^4 + y^4)$. On the other hand, by (2.1.1), $|z|^4 = a^4 + y^4 < \mu^2/(\operatorname{Im} M)^2$. Since y can be arbitrarily small, a can only be zero.

It remains only to remark that for any test function $g(t)$ continuous off an F-null set (and hence for the unit step function with step anywhere but zero)

$$\lim_{y\to 0} \int (1/2\pi i)\left(\int \frac{dF(t)}{x-iy-t} - \int \frac{dF(t)}{x+iy-t}\right) g(t)\,dt = \int g(t)\,dF(t)\,. \quad \square$$

We now proceed to the proofs of the lemmas on which Theorem 2.1 is based.

LEMMA 2.2. *Under the assumptions of Theorem* 2.1, *the expectation of the random variable* $H(n, 2k)$ *defined in Lemma* 1.2 *differs by less than* $(1/n)k^{2k+1}A^{2k}$ *from*

$$\sum_a 2(m/n)^{k+1}\phi(a)\left(\frac{p}{m}\cdot\frac{1}{p}\sum_1^p \sigma_i^2\right)^{a(1)} \cdots \left(\frac{p}{m}\cdot\frac{1}{p}\sum_1^p \sigma_i^{2k}\right)^{a(k)}$$

where the summation ranges over k*-tuples of nonnegative integers* $a(1)\cdots a(k)$ *such that* $\sum ja(j) = k$ *and* $\phi(a)$ *is the number of* $2k$*-tuples of positive integers* $s_1\cdots s_{2k}$ *satisfying* $s_{2i-1} \leqq b$, $s_{2i} \leqq k+1-b$, *where* $b = \sum a(j)$, *and* $a(j) = \text{card}\,\{t\colon j = \text{card}\,\{i\colon s_{2i-1} = t\}\}$ *and there exists a one-to-one map from the pairs* s_{2i-1}, s_{2i} *to the pairs* s_{2i-1}, s_{2i-2}, *where* $s_0 = s_{2k}$.

PROOF. The assumption that variances are equal within rows reduces each term

$$E|Z_{s_1s_2}|^2 E|Z_{s_3s_4}|^2 E|Z_{s_{2k-1}s_{2k}}|^2$$

in $H(n, 2k)$ to a product $(\sigma_{s_1}\cdots\sigma_{s_{2k-1}})^2$ indexed only by row subscripts. This expression depends only on which distinct integers, say $r(1)\cdots r(b)$, occur with what multiplicities $y(1)\cdots y(b)$ in turn among the s_{2i-1}, so we split up the sum in $EH(n, 2k)$ into

$$(2/n^{k+1})\sum_b \sum_y \sum_r \sum_c \sum_{s \text{ in } L} (\sigma_{s_1}\sigma_{s_2}\cdots\sigma_{s_{2k-1}})^2$$

where the summations range as follows:

(i) b over 1 to k;

(ii) y over all b-tuples of positive integers $y(1)\cdots y(b)$ satisfying $\sum_1^b y(t) = k$;

(iii) r over all b-tuples of distinct positive integers $r(1)\cdots r(b)$ all smaller than $p(n)$;

(iv) c over all w-tuples of distinct positive integers $c(1)\cdots c(w)$ all smaller than $m(n)$, with $w = k+1-b$;

(v) $L = \{2k$-tuples such that s_{2i-1} ranges over $r(1)\cdots r(b)$ and s_{2i} ranges over $c(1)\cdots c(k+1-b)$ and pairs $\langle s_{2i-1}, s_{2i}\rangle$ correspond one-to-one to $\langle s_{2i-1}, s_{2i-2}\rangle\}$.

The summand does not depend on L (or on c) and the cardinality of L is a purely combinatorial quantity fixed by b, k, and $y(1)\cdots y(b)$, always equal to the value it takes, for instance, when $r(i) = i$ and $c(i) = i$. We have

$$EH = (2/n^{k+1})\sum_b \sum_y \text{card}\,\{L\} \sum_c \sum_r \sigma_{r(1)}^{2y(1)} \cdots \sigma_{r(n)}^{2y(b)}\,.$$

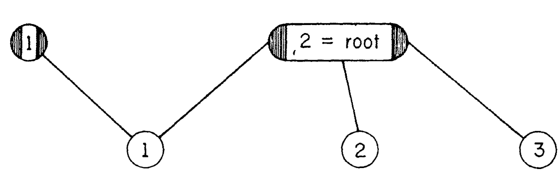
The sums over r and c differ from the sums ranging over all tuples of not-necessarily-distinct integers by $p^b m^w - p!\,m!/(p-b)!\,(m-w)!$ terms each bounded by A^{2k}. We have $\sum_y$ card $\{L\} \leqq b^k w^k \leqq ((k+1)/2)^{2k}$. Thus EH differs by less than $k^{2k+1}A^{2k}/n$ from

$$(2/n^{k+1}) \sum_b \sum_y \text{card } \{L\} m^w \sum_1^p \sigma_i^{2y(1)} \cdots \sum \sigma_i^{2y(b)} .$$

Plugging in $a(j)$ for the number of factors in which $y(i) = j$ produces the expression in the lemma. □

LEMMA 2.3. *The rooted plane bichromatic trees with alternating black and white nodes with $a(j)$ black nodes of valency j for $j = 1, 2, 3, \cdots$, one of whose black nodes is the root node, stand in one-to-one correspondence, when $k = \sum ja(j)$ and $b = \sum a(j)$ and $w = k + 1 - b$ with the $2k$-tuples of positive integers $s_1 \cdots s_{2k}$ specified in Lemma* 2.2.

PROOF. For each $2k$-tuple of subscripts we construct a graph by the same procedure as in Lemma 1.2. The graph is a tree because it is connected, has k branches, and $b + w = k + 1$ nodes. If we colour row subscript nodes labeled 1 to b along a top line black and we colour column subscript nodes labeled 1 to w along a bottom line white, our tree is clearly an alternating, bichromatic tree. Traversing the tree by the route $s_1, s_2, s_3 \cdots s_{2k}, s_1$, each of the $b + w$ nodes is visited. Each branch is crossed once in each direction (black to white and white to black) thanks to the one-to-one correspondence between s_{2i-1}, s_{2i} and s_{2i-1}, s_{2i-2}. In our traverse, we are beginning at a preferred black node, s_1, and we are establishing an orientation: at each node any branches leading to nodes farther from the root node are ordered in the order we traverse them. The preferred node makes the tree a rooted tree. The orientation makes the tree a plane tree. The tree could be drawn so as to inherit its orientation from the clockwise sense of the Euclidean plane. The valency of the tth black node, the number of branches emanating from it, equals card $\{i: s_i = t \text{ for } i = 1, 3, \cdots, 2k - 1\}$. The terminology accords with Chapter 6 of Riordan (1958) and with Mallows and Wachter (1972). In this way each of our $2k$-tuples corresponds to a tree of the specified type. For example, $s = \langle 2, 2, 2, 3, 2, 1, 1, 1 \rangle$ which indexes the term $Z_{22}\bar{Z}_{22}Z_{23}\bar{Z}_{23}Z_{21}\bar{Z}_{11}Z_{11}\bar{Z}_{21}$ in $H(n, 8)$ for $k = 4$, $b = 2$, $w = 3$, and $a = \langle 1, 0, 1, 0 \rangle$ corresponds to the tree



Conversely, from any such tree, we can recover a $2k$-tuple satisfying the given conditions, namely, the same $2k$-tuple that yields the tree. Our correspondence is a bijection. We recover the $2k$-tuple by traversing the tree and putting s_i equal to the label of the ith node traversed. We begin at the black root node, proceeding from each node along the branch with the highest precedence in the orientation at that node among all branches not previously traversed toward nodes farther from the root node, and, when such branches are exhausted, returning to the unique adjacent node nearer the root node. By this rule we never cross a branch in the same direction twice. After $2k$ steps we must have passed each of the k branches in both directions. No branch to a node nearer the root node from our position can remain to cross, so we must have returned to the root node s_1. As before, the valencies match with the parameters $a(j)$, and the correspondence is established. □

Theorem 2.1 in its present form covers the asymptotic distribution of sample standard deviations of principal components. Tables of quantiles for cases where K is concentrated on one or two points along with discussion of their use in probability plotting occur in Wachter (1975). More formal analysis with computer programs for these quantiles may be found in Wachter (1976a). The case of general finite discrete K is treated in Wachter (1976b). Applications to plotting of discriminant ratios in Wachter (1976c), however, call for stronger conclusions than those of Theorem 2.1 in the direction of uniform convergence over matrices with different sequences of row variances. Uniformity is needed when we let row variances themselves be random variables. We already have in hand the tools for a good uniform assertion:

THEOREM 2.4. *For any sequence σ of positive numbers σ_i in* $(0, A)$, *let $K_n\langle\sigma\rangle$ be the empirical measure of $\sigma_1 \cdots \sigma_{p(n)}$ and let $G_n\langle\sigma\rangle$ be the solution* (F) *to equations* (2.1.1) *and* (2.1.2) *when $K = K_n\langle\sigma\rangle$. Let $R_n\langle\sigma\rangle$ be the random spectrum of $Z\langle\sigma\rangle$ where $Z_{ij}\langle\sigma\rangle = \sigma_i Y_{ij}$ for a random matrix Y satisfying the conditions of Theorem* 2.1 *but with unit variances. Then $R_n\langle\sigma\rangle$ and $G_n\langle\sigma\rangle$ converge to each other in probability uniformly over σ.*

PROOF. Proposition 1.4 is formulated so that if our claim here of uniformity holds for Z truncated at $T(j)$ as in Proposition 1.4, it holds for the original Z. The existence of $G_n\langle\sigma\rangle$ is furnished us by Theorem 2.1, since the sequence $\xi_i = \sigma_{i \bmod p(n)}$ satisfies

$$|(1/l)\sum_1^l \xi_i^{2k} - \int x^{2k}\, dK_n\langle\sigma\rangle| < 2p(n)A^{2k}/l \to 0 \qquad \text{as} \quad l \to \infty\,.$$

For each n we know that $G_n\langle\sigma\rangle$ is a fixed measure concentrated on $[A(2\mu + 2\beta)^{\frac{1}{2}}, A(2\mu + 2\beta)^{\frac{1}{2}}]$ determined by its moments.

We begin by estimating the probability that the differences between the moments of R_n and G_n stay under control.

$$\begin{aligned} \mathbb{P}\{\exists\, k : |\textstyle\int x^{2k}\, dR_n\langle\sigma\rangle - \int x^{2k}\, dG_n\langle\sigma\rangle| > sc(k)\} \\ \leqq \textstyle\sum_{k=1} (1/sc(k))^2\{\mathrm{Var}\,(\int x^{2k}\, dR_n) + (E\int x^{2k}\, dR_n - EH(n, 2k))^2 \\ + (EH(n, 2k) - \textstyle\int x^{2k}\, dG_n)^2\}\,. \end{aligned}$$

The three terms on the right are bounded according to Lemmas 1.3, 1.2 and 2.2 by $(1/n)(A^{4k} + T^{4k}(n)/n)(2k)^{2k+2} + (k^k T^{2k}(n)/n)^2 + (A^{2k}k^{2k+1}/n)^2$ and $\sup\{T^{4k}(n)/n : n = 1, 2, \cdots\}$ is some finite function of k. Thus we may choose $c(k)$ to increase so rapidly with k that the infinite sum is less than $1/ns^2$. $\int x^{2k}\, dG\langle\sigma\rangle < (2AA(\mu + \beta))^k$, so the moment sequences of all measures whose odd moments vanish and whose $2k$th moments differ from $\int x^{2k}\, dG_n\langle\sigma\rangle$ by at most $c(k)$ inhabit a coordinatewise bounded (closed) subset of the space $\mathbb{R}^\infty$, that is to say, a subset compact in the product topology. The map which associates a determinate moment sequence with the measure it determines is continuous between the topology induced on these sequences by the product topology on all of $\mathbb{R}^\infty$ and the topology induced on Prob $\mathbb{R}$ by the weakstar topology on the whole dual space of the bounded continuous functions on $\mathbb{R}$, an induced topology which is, after all, just the familiar topology of weak convergence. Uniformity can be defined by reference to neighborhoods of zero in the larger linear spaces; on a compact subset the moments-to-measures map must be uniformly continuous.

The uniform continuity means that we can keep $R_n\langle\sigma\rangle - G_n\langle\sigma\rangle$ in any neighborhood U of zero by making the first k differences of moments small enough for some finite k. That we can certainly achieve by making each $2k$th difference smaller than $sc(k)$ for small enough s, and then the moments stay nicely in the compact set as well. We deduce that for any neighborhood U of zero,

$$\exists s \quad \forall \sigma \quad \mathbb{P}\{R_n\langle\sigma\rangle - G_n\langle\sigma\rangle \text{ in } U\} > 1 - 1/(ns^2)\,. \qquad \square$$

REFERENCES

[1] AHLFORS, LARS (1966). *Complex Analysis*, Second Edition. McGraw-Hill, New York.

[2] ANDERSON, T. W. (1963). Asymptotic theory for principal component analysis. *Ann. Math. Statist.* **34** 122–148.

[3] ARNOLD, LUDWIG (1971). On Wigner's Semicircle Law for the eigenvalues of random matrices. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **19** 191–198.

[4] ARNOLD, LUDWIG (1973). The possible asymptotic distributions of eigenvalues of random symmetric matrices. *Trans. Sixth Prague Conf. Information Theor. Statist. Decision Functions, Random Processes*. Academia, Prague 51–62.

[5] ARNOLD, LUDWIG (1976). Deterministic version of Wigner's Semicircle Law for the distribution of matrix eigenvalues. *Linear Algebra Appl.* **13** 185–199.

[6] BHARUCHA-REID, A. T. (1972). *Random Integral Equations*. Academic Press, New York, 86–88.

[7] BREIMAN, LEO (1968). *Probability*. Addison-Wesley, Reading, Mass.

[8] DEMPSTER, ARTHUR P. (1969). *Elements of Continuous Multivariate Analysis*. Addison-Wesley, Reading, Mass.

[9] DUDLEY, RICHARD M. (1966). Convergence of Baire measures. *Studia Math.* **27** 251–268.

[10] FAN, KY (1951). Maximum properties and inequalities for the eigenvalues of completely continuous operators. *Proc. Nat. Acad. Sci. U.S.A.* **37** 760–766.

[11] HARDY, G. H. (1955). *Pure Mathematics*, Tenth Edition. Cambridge Univ. Press.

[12] MALLOWS, COLIN and WACHTER, KENNETH (1970). The asymptotic configuration of Wishart eigenvalues (abstract). *Ann. Math. Statist.* **41** 1384.

[13] MALLOWS, COLIN and WACHTER, KENNETH (1972). Valency enumeration of rooted plane trees. *J. Austral. Math. Soc.* **13** 472–476.

[14] MEHTA, M.L. (1967). *Random Matrices and the Statistical Theory of Energy Levels*. Academic Press, New York.

[15] OLSON, W. H. and UPPULURI, V. R. RAO (1973). Asymptotic distribution of eigenvalues of random matrices. *Proc. Sixth Berkeley Symp. Math. Statist. Prob.* **3** 615–644. Univ. of California Press.

[16] PORTER, CHARLES E., ed. (1965). *Statistical Theories of Spectra: Fluctuations.* Academic Press, New York.

[17] RIORDAN, JOHN (1958). *An Introduction to Combinatorial Analysis.* Wiley, New York.

[18] WACHTER, KENNETH (1974a). *Foundations of the Asymptotic Theory of Random Matrix Spectra.* Ph. D. dissertation, Cambridge Univ.

[19] WACHTER, KENNETH (1974b). Exchangeability and asymptotic random matrix spectra. *Progress in Statistics.* Joe Gani, K. Sarkadi and I. Vince, eds. North Holland Press.

[20] WACHTER, KENNETH (1975). User's guide and tables to probability plotting for large scale principal component aualysis. Research report W-75-1, Dept. Statist. Harvard Univ.

[21] WACHTER, KENNETH (1976a). Probability plotting points for principal components. *Proc. Ninth Interface Symp. Computer Science and Statist.* (Hoaglin and Welsch, eds.). Prindle, Weber and Schmidt. Boston, Mass. 299–308.

[22] WACHTER, KENNETH (1976b). Asymptotic distributions of principal component variances: the k-atom case. Research report W-76-1, Dept. Statist. Harvard Univ.

[23] WACHTER, KENNETH (1976c). The limiting empirical measure of multiple discriminant ratios. Research report W-76-3, Dept. Statist. Harvard Univ.

[24] WACHTER, KENNETH (1976d). Probability plotting of multiple discriminant ratios. *Amer. Statist. Assoc. Proc. Social Statist. Section 1976, Part II.* Washington, D.C. 830–833.

[25] WEGMANN, RUDOLF (1976a). Bermerkungen zum halbkreissatz für die eigenwerte hermitescher matrizen. *Arch. Math.* **27** 535–538.

[26] WEGMANN, RUDOLF (1976b). The asymptotic eigenvalue distribution for a certain class of random matrices. *J. Math. Anal. Appl.* **56** 113–132.

[27] WIELANDT, HELMUT (1955). An extremum property of sums of eigenvalues. *Proc. Amer. Math. Soc.* **6** 106–110.

DEPARTMENT OF STATISTICS
HARVARD UNIVERSITY
1 OXFORD STREET
CAMBRIDGE, MASS. 02138