SIAM J. MATH. ANAL.
Vol. 12, No. 2, March 1981

0036-1410/81/1202-0012 $01.00/0

DESCRIBING THE BEHAVIOR OF EIGENVECTORS OF RANDOM MATRICES USING SEQUENCES OF MEASURES ON ORTHOGONAL GROUPS*

JACK W. SILVERSTEIN†

Abstract. A conjecture has previously been made on the chaotic behavior of the eigenvectors of a class of n-dimensional random matrices, where n is very large [J. Silverstein, SIAM J. Appl. Math., 37 (1979), pp. 235–245]. Evidence supporting the conjecture has been given in the form of two limit theorems, as $n \to \infty$, relating the random matrices to matrices formed from the Haar measure, h_n, on the orthogonal group $\mathcal{O}_n$.

The present paper considers a reformulation of the conjecture in terms of sequences of the form $\{\mu_n\}$, where for each n, μ_n is a Borel probability measure on $\mathcal{O}_n$. A characterization of μ_n being "close" to h_n for n large is developed. It is suggested that before a definition of what it means for $\{\mu_n\}$ to be *asymptotic Haar* is decided, properties $\{h_n\}$ possess should first be proposed as possible necessary conditions. The limit theorems are converted into properties on $\{\mu_n\}$. It is shown (Theorem 1) that one property is a consequence of the other. Another property is proposed resulting in the construction of measures on $D = D[0, 1]$ which converge weakly. It is shown (Theorem 2) that under this necessary condition for asymptotic Haar, not only is the conjecture in general not true, but that the behavior of the eigenvectors of large dimensional sample covariance matrices deviates significantly from being Haar distributed when the i.i.d. standardized components making up the matrix differ in the fourth moment from 3.

1. Toward a definition of asymptotic Haar. In [6], a class of large dimensional, symmetric, positive semidefinite random matrices resulted from a model for the generation of neural connections of a hypothetical organism at birth. Denote by W_n one of these random matrices which is $n \times n$, where n is very large. Briefly, W_n is of the form $(1/C_n) V_n V_n^T$, where $V_n = (v_{ij})$ is $n \times dn$ and d is fixed; the v_{ij}'s are independent; v_{ij} is 1 or -1 with equal probability, or zero; $P = (P_{ij})$ is $n \times dn$, where $P_{ij} = \text{Prob}\,(v_{ij}^2 = 1)$, is formed under rather general conditions, and, in particular, every row of P is a rotation of the first row; and C_n is the sum of the first row of P. It is shown in [6] that if $C_n \to \infty$ as $n \to \infty$, then the empirical distribution function $F_n(x)$ of the eigenvalues of W_n converges in probability as $n \to \infty$ for each x to a fixed continuous distribution function $F(x)$. This result complements those on large dimensional random matrices (see for example [2], [3], [5], [7], [9], [11], [12], [13]), in particular, results on sample covariance matrices and matrices associated with the statistical theory of spectra.

In [10], a question is raised as to the behavior of the eigenvectors of W_n. It has been conjectured that this behavior is completely chaotic, and an attempt at formalizing this conjecture has been the following: for each n let $\mathcal{O}_n$ denote the orthogonal group consisting of $n \times n$ orthogonal matrices, and let $O_n \in \mathcal{O}_n$ be distributed according to the normalized Haar measure, h_n, on $\mathcal{O}_n$. Let D_n be a nonrandom $n \times n$ diagonal matrix with diagonal elements arranged in nondecreasing order and such that the spectrum of D_n approaches F as $n \to \infty$. The conjecture is that, for n large, the distribution of $W'_n \equiv O_n D_n O_n^T$ is close (in some sense) to the distribution of W_n.

Evidence supporting the conjecture is provided in [10] in the form of results which demonstrate that W_n and W'_n have similar properties. Let $\{P_a(M^n)\}_{a=0}^{\infty}$ be the spectral family of $M^n = W_n$ or W'_n, let $\{x_n\}$, $x_n \in \mathbb{R}^n$, be any fixed sequence of unit vectors, and let M_1^n, M_2^n be two independent generations of M^n. Then it is proven in [10] that for $M^n = W_n$ or W'_n,

$$x_n^T P_a(M^n) x_n \xrightarrow{\text{i.p.}} F(a) \quad \text{as } n \to \infty \text{ for every } a \in [0, \infty), \tag{1.1}$$

* Received by the editors May 28, 1980, and in revised form September 3, 1980.
† Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27650.

and

$$(1.2)\quad \frac{1}{n}\operatorname{tr}\left[(P_{a_1}(M_1^n)-P_{a_2}(M_2^n))^2\right] \xrightarrow{\text{i.p.}} F(a_1)+F(a_2)-2F(a_1)F(a_2) \quad \text{as } n\to\infty,$$

for every $a_1, a_2 \in [0, \infty)$ where tr is the trace function. The belief has been that these two results are enough to prove the conjecture.

The validity of the conjecture will imply certain properties of the neural model. However, the same question can be asked of other classes of large dimensional random matrices, at least for those not constructed from Gaussian variables. It is known, for example, that the Wishart matrix $W(I, dn)$ behaves like W_n' except D_n is random (see [1, Chapt. 13]). We remark here that the results in [10] are true for sample covariance matrices in which the elements in the sample vectors are i.i.d., mean 0, having moments of all orders (the results in [10] rely totally on [6, Lemma 1], and the proof of this lemma can be slightly modified to include these cases). It is also believed that these results are valid for more general random matrices. Thus, statements concerning W_n are relevant for a large class of random matrices.

The present paper continues the investigation of the eigenvectors of W_n primarily by developing some ideas toward a well-defined statement of the conjecture. To begin with, it seems more fitting to shift the attention from W_n to the measure it induces on $\mathcal{O}_n$. Let $O_n \in \mathcal{O}_n$ be random, defined on the same probability space as W_n, and such that $O_n^T W_n O_n = \Lambda_n$, where Λ_n is diagonal with its diagonal elements arranged in nondecreasing order. We may as well assume that the distribution of O_n is the same as that of $O_n J$ for each diagonal J containing ± 1's along its diagonal. Also we may assume that, conditioned on any collection of subsets of eigenvalues of W_n being equal within each subset, the distribution of O_n is the same as that of $O_n K$ whenever $K \in \mathcal{O}_n$ transforms only among each subset of columns of O_n corresponding to a subset of equal eigenvalues, and leaves all other columns unchanged. Let ν_n be the Borel probability measure induced by O_n.

The conjecture can now be expressed in terms of ν_n and h_n being "close" for n large. We will use the expression *asymptotic Haar* to describe this, at present a vague property on sequences $\{\mu_n\}$ where, for each n, μ_n is a Borel probability measure on $\mathcal{O}_n$.

The most obvious and by far the strongest statement of asymptotic Haar is: for every $\varepsilon > 0$, we have for all n sufficiently large $|\mu_n(A) - h_n(A)| < \varepsilon$, for every $A \in \mathbb{B}_n \equiv$ the collection of Borel sets of $\mathcal{O}_n$ (the metric on $\mathcal{O}_n$ being induced from the operator norm). This definition is too restrictive if we do not want to exclude from being asymptotic Haar all sequences $\{\mu_n\}$ of atomic measures. If we let $S_{n,\delta}$ represent the collection of all open balls on $\mathcal{O}_n$ having Haar measure δ, then another definition which would allow certain sequences of atomic measures is: for every $\varepsilon > 0$, $1 \geqq \delta > 0$, we have for all n sufficiently large $|\mu_n(B) - h_n(B)| < \varepsilon$ for every $B \in S_{n,\delta}$. Several alternative definitions can certainly be proposed along the same lines.

It is the author's view that, instead of initially focusing on one definition of asymptotic Haar, attention should be drawn on intuitive and reasonable consequences of the definition. Various properties $\{h_n\}$ possess should be considered as necessary conditions for asymptotic Haar. Also, examples of sequences that should not be asymptotic Haar need to be found. For example, (1.1) and (1.2) can be restated in terms of the following properties.

We say that $\{u_n\}$ satisfies property I if for any sequence of unit vectors $\{x_n\}$, $x_n \in \mathbb{R}^n$, any number b such that $0 \leqq b \leqq 1$, and any sequence of integers $\{m_n\}$ satisfying

$0 \leqq m_n \leqq n$ and $m_n/n \to b$ as $n \to \infty$, we have,

$$x_n^T O_n D(n, m_n) O_n^T x_n \xrightarrow{\text{i.p}} b \quad \text{as } n \to \infty, \tag{1.3}$$

where O_n is μ_n-distributed, and where $D(n, m_n)$ is $n \times n$ and has 0 for all its entries except for 1's in the first m_n diagonal entries.

We say that μ_n satisfies property II if for any b_1, b_2 such that $0 \leqq b_1, b_2 \leqq 1$, and any two sequences of integers $\{m_n^1\}$, $\{m_n^2\}$ satisfying $0 \leqq m_n^i \leqq n$ and $m_n^i/n \to b_i$ as $n \to \infty$, $i = 1, 2$, we have,

$$\frac{1}{n} \operatorname{tr}\,[(O_n D(n, m_n^1) O_n^T - O_n' D(n, m_n^2) O_n'^T)^2] \xrightarrow{\text{i.p.}} b_1 + b_2 - 2b_1 b_2 \quad \text{as } n \to \infty, \tag{1.4}$$

where O_n and O_n' are independent and μ_n-distributed.

The sequence $\{h_n\}$ satisfies I. The easiest way of seeing this is to use the fact that $x_n^T O_n D(n, m_n) O_n^T x_n$ is beta-distributed with mean m_n/n which goes to b, and variance $2m_n(n - m_n)/n^3$ which goes to 0 (see [10, proof of Theorem 1]).

The sequence $\{\nu_n\}$ also satisfies I. The proof is elementary and technical and will be omitted.

Theorem 1 in the next section shows that II is a consequence of I, a somewhat surprising result. Thus, we have so far only one necessary condition for asymptotic Haar.

At this stage, we are in a position to consider whether I is enough to characterize asymptotic Haar. For each n let μ_n be absolutely continuous with respect to h_n, having density f_n. Let $\{x_n\}$, $\{m_n\}$ be as in I. Using the fact that $x_n^T O_n D(n, m_n) O_n^T x_n$ is beta-distributed when O_n is h_n-distributed, we get from the Cauchy-Schwarz inequality:

$$\begin{aligned} &\left(\int_{\mathcal{O}_n} \left| x_n^T O_n D(n, m_n) O_n^T x_n - \frac{m_n}{n} \right| f_n(O_n)\, dh_n(O_n)\right)^2 \\ &\quad \leqq \left[\int_{\mathcal{O}_n} \left(x_n^T O_n D(n, m_n) O_n^T x_n - \frac{m_n}{n}\right)^2 dh_n(O_n)\right]\left[\int_{\mathcal{O}_n} f_n^2(O_n)\, dh_n(O_n)\right] \\ &\quad = 2\,\frac{m_n}{n^2}\left(\frac{n - m_n}{n}\right) \int_{\mathcal{O}_n} f_n^2(O_n)\, dh_n(O_n). \end{aligned} \tag{1.5}$$

Thus, if $\int_{\mathcal{O}_n} f_n^2(O_n)\, dh_n(O_n) = o(n)$, then we get L^1-convergence in (1.3) so that $\{\mu_n\}$ satisfies I. This is true if $\{f_n\}$ is any uniformly bounded sequence of densities. For example, if $f_n = 2$ on a closed subset of $\mathcal{O}_n$ having Haar measure $\frac{1}{2}$, and 0 elsewhere, then $\{\mu_n\}$ satisfies I. Under the quite reasonable assumption that the above sequence should not be considered to be asymptotic Haar, then we must conclude that I is *not* enough to characterize asymptotic Haar.

Other properties of $\{h_n\}$ therefore, need to be considered.

The remainder of this paper is devoted to developing another property, and considering the consequences, if this property is to be a necessary condition for asymptotic Haar.

For $O_n \in \mathcal{O}_n$ Haar-distributed and any unit vector $x_n \in \mathbb{R}^n$, we have $O_n^T x_n$-distributed like $(\zeta_1, \zeta_2, \cdots, \zeta_n)/(\sum_{i=1}^n \zeta_i^2)^{1/2}$, where $\zeta_1, \zeta_2, \cdots, \zeta_n$ are i.i.d. $n(0, 1)$. Form

$$\begin{aligned} X_n(t) &= \frac{\sqrt{n}}{\sqrt{2}}\left(\frac{\sum_{i=1}^{[nt]} \zeta_i^2}{\sum_{i=1}^n \zeta_i^2} - \frac{[nt]}{n}\right) \\ &= \frac{n}{\sum_{i=1}^n \zeta_i^2} \frac{1}{\sqrt{2}} \frac{1}{\sqrt{n}}\left(\sum_{i=1}^{[nt]} (\zeta_i^2 - 1) - \frac{[nt]}{n} \sum_{i=1}^{h} (\zeta_i^2 - 1)\right), \end{aligned} \tag{1.6}$$

where $[s]$ is the greatest integer $\leqq s$. We have $X_n(t)$ a random element of $D=D[0,1]$ (the space of all r.c.l.l. functions on $[0,1]$) and from straightforward applications of Donsker's Theorem and the theory of weak convergence of measures [4], we have:

$$X_n \xrightarrow{\mathscr{D}} W^0, \tag{1.7}$$

where W^0 is a Brownian bridge. Hence, another necessary condition for asymptotic Haar:

We say that $\{\mu_n\}$ satisfies property III if, for every sequence $\{x_n\}x_n \in \mathbb{R}^n$ of unit vectors, if $(\zeta_1, \zeta_2, \cdots, \zeta_n) \equiv O_n^T x_n$ where O_n is μ_n-distributed, and if $X_n(t)$ is as in (1.6), then (1.7) holds.

This property seems to be a reasonable necessary condition for asymptotic Haar. It ensures that $O_n^T x_n$ be close to being uniformly distributed on the unit sphere in $\mathbb{R}^n$. In fact, $O_n^T x_n$ need only have a distribution resembling the distribution of $(Y_1, Y_2, \cdots, Y_n)/(\sum_{i=1}^n Y_i^2)^{1/2}$ where the Y_i's are i.i.d. with $\mathrm{E}(Y_1^2)=1$ and $\operatorname{var}(Y_1^2)=2$.

It would also seem reasonable that the behavior of the eigenvectors of large dimensional sample covariance matrices be a prime example for asymptotic Haar. But with the inclusion of III as a necessary condition, this is not the case. Let $\{u_{ij}\}$, $i, j = 1, 2, \cdots$, be i.i.d. random variables having mean 0, variance 1, and satisfying $\mathrm{E}(|u_{11}|^m) \leqq m^{\alpha m}$ for all integers $m>2$ and for some α. For each n let $U_n=(u_{ij})$, $i=1,2,\cdots,n$, $j=1,2,\cdots,s$, where $n/s \to y>0$ as $n\to\infty$, and let μ_n be the measure on $\mathcal{O}_n$ induced from $(1/s)U_nU_n^T$. Theorem 2 in the next section shows that if $\mathrm{E}(u_{ij}^4)\neq 3$, then $\{\mu_n\}$ does not satisfy III. The proof relies on standard tools used in the theory of weak convergence on metric spaces, along with a recent result on the almost sure convergence of the largest eigenvalue of sequences of sample covariance matrices [5], where the above growth condition on the moments of $|u_{11}|$ is assumed.

In the formation of W_n, letting $P_{ij}=p$ for all i, j where $p\neq \frac{1}{3}$, we are in the above case with $\mathrm{E}(u_{11}^4)=\mathrm{E}((v_{11}/\sqrt{p})^4)=1/p$. We must therefore conclude that with III as a necessary condition for asymptotic Haar, the original conjecture is, in general, false. It may be argued that III is too strong, and it may be possible to find interesting properties shared by $\{\nu_n\}$ and $\{h_n\}$. Moreover, $\{\nu_n\}$ may still satisfy III when $p=\frac{1}{3}$ or when the P_{ij}'s are not all the same. However, we feel that failure to satisfy III indicates significant departure from Haar measure.

The requirement that $\mathrm{E}(u_{11}^4)=3$ suggests that for sample covariance matrices, in order to satisfy III, the u_{ij}'s have to be near to being Gaussian distributed, as in the Wishart case. It appears worthwhile to determine what conditions on the u_{ij}'s are needed to ensure III.

In conclusion, it should be emphasized that one purpose of this paper is to begin an investigation on how to characterize the closeness of measures on $\mathcal{O}_n$ to Haar measure, where n is large. The considerations given are clearly the author's view on how to proceed in defining asymptotic Haar. We suggest continuing the characterization by finding other mappings of $\mathcal{O}_n$ onto a common metric space S, resulting in weak convergence of the measures on S induced by $\{h_n\}$. Intuitively, the mappings $F_n: \mathcal{O}_n \to S$ should all be similar, sort of invariant across dimensions. They should also illuminate the intrinsic uniformity of Haar measure.

We find it interesting that the W_n's do not in general fall into the present characterization of asymptotic Haar. Still, $\{\nu_n\}$ and $\{h_n\}$ are similar, and a first step toward determining just how similar they are would be to understand those sequences $\{\mu_n\}$ satisfying I.

The fact that sequences $\{\mu_n\}$ arising from sample covariance matrices do not in general satisfy III is of even greater interest, and this suggests a behavior of the eigenvectors of these matrices for large n which runs counter to our intuition. A description of this behavior is important to multivariate theory, and work in this area should be pursued.

2. The theorems.

THEOREM 1. I $\to$ II.

Proof. Assume $\{\mu_n\}$ satisfies I. Let $P(m_n, O_n) \equiv O_n D(n, m_n) O_n^T$. Convergence of $x_n^T P(m_n, O_n) x_n$ to b in probability is equivalent to

$$\text{(2.1)} \qquad \mathrm{E}(x_n^T P(m_n, O_n) x_n) = \int_{\mathcal{O}_n} x_n^T P(m_n, O_n) x_n \, d\mu_n(O_n) \to b \quad \text{as } n \to \infty$$

and

$$\text{(2.2)} \qquad \mathrm{E}((x_n^T P(m_n, O_n) x_n)^2) = \int_{\mathcal{O}_n} (x_n^T P(m_n, O_n) x_n)^2 \, d\mu_n(O_n) \to b^2 \quad \text{as } n \to \infty.$$

The expected values in (2.1) and (2.2) are polynomials in the components of x_n and are therefore continuous in x_n. Let $\{x'_n\}$ and $\{x''_n\}$ be the sequences such that $\mathrm{E}(x_n^T P(m_n, O_n) x_n)$ attains its maximum at x'_n and its minimum at x''_n. Since (2.1) holds for *all* sequences of unit vectors, it is certainly true for $\{x'_n\}$ and $\{x''_n\}$. Therefore,

$$\text{(2.3)} \quad \mathrm{E}(x_n^T P(m_n, O_n) x_n) = b + \alpha_n(x_n) \quad \text{where } |\alpha_n(x_n)| \leqq \alpha_n \text{ and } \alpha_n \to 0 \text{ as } n \to \infty.$$

Similarly,

$$\text{(2.4)} \qquad \mathrm{E}((x_n^T P(m_n, O_n) x_n)^2) = b^2 + \beta_n(x_n) \quad \text{where } |\beta_n(x_n)| \leqq \beta_n \text{ and } \beta_n \to 0 \text{ as } n \to \infty.$$

Also, for any two sequences $\{x_n\}$, $\{y_n\}$ we have

$$\text{(2.5)} \qquad (x_n^T P(m_n, O_n) x_n)(y_n^T P(m_n, O_n) y_n) \xrightarrow{\text{i.p.}} b^2 \quad \text{as } n \to \infty,$$

and as above we have

$$\text{(2.6)} \qquad \mathrm{E}((x_n^T P(m_n, O_n) x_n)(y_n^T P(m_n, O_n) y_n)) = b^2 + \gamma_n(x_n, y_n),$$

where $|\gamma_n(x_n, y_n)| \leqq \gamma_n$ and $\gamma_n \to 0$ as $n \to \infty$. Let $\{m_n^1\}$, $\{m_n^2\}$ be as in II. Since

$$\text{(2.7)} \qquad \begin{aligned} \frac{1}{n} \operatorname{tr} [(P(m_n^1, O_n) - P(m_n^2, O'_n))^2] = \frac{m_n^1}{n} + \frac{m_n^2}{n} - \frac{1}{n} \operatorname{tr} P(m_n^1, O_n) P(m_n^2, O'_n) \\ - \frac{1}{n} \operatorname{tr} P(m_n^2, O'_n) P(m_n^1, O_n), \end{aligned}$$

it is sufficient to prove

$$\text{(2.8)} \qquad \frac{1}{n} \operatorname{tr} P(m_n^1, O_n) P(m_n^2, O'_n) \xrightarrow{\text{i.p.}} b_1 b_2 \quad \text{as } n \to \infty.$$

We have

$$
\begin{aligned}
\frac{1}{n}\operatorname{tr} P(m_n^1, O_n)P(m_n^2, O_n') &= \frac{1}{n}\operatorname{tr} P(m_n^1, O_n)O_n'D(n, m_n^2)O_n'^T \\
&= \frac{1}{n}\operatorname{tr} O_n'^T P(m_n^1, O_n)O_n'D(n, m_n^2) \\
&= \frac{1}{n}\sum_{i=1}^{m_n^2} (O_n'^T P(m_n^1, O_n)O_n')_{ii} \\
&= \frac{1}{n}\sum_{i=1}^{m_n^2} o_{.i}'^T P(m_n^1, O_n)o_{.i}',
\end{aligned}
\tag{2.9}
$$

where $o_{.i}'$ is the ith column of O_n'. For fixed O_n' we have, from (2.3),

$$
\int_{\mathcal{O}_n} \frac{1}{n}\sum_{i=1}^{m_n^2} o_{.i}'^T P(m_n^1, O_n)o_{.i}'\, d\mu_n(O_n) = \frac{m_n^2}{n} b_1 + \frac{1}{n}\sum_{i=1}^{m_n^2} \alpha_n(o_{.i}')
\tag{2.10}
$$

and

$$
\frac{1}{n}\left|\sum_{i=1}^{m_n^2} \alpha_n(o_{.i}')\right| \leqq \alpha_n.
$$

Therefore,

$$
\mathrm{E}\left(\frac{1}{n}\operatorname{tr} P(m_n^1, O_n)P(m_n^2, O_n')\right) = \left(\frac{m_n^2}{n}\right) b_1 + \xi_n,
\tag{2.11}
$$

where $\xi_n \to 0$ as $n \to \infty$, and so

$$
\mathrm{E}\left(\frac{1}{n}\operatorname{tr} P(m_n^1, O_n)P(m_n^2, O_n')\right) \to b_1 b_2 \quad \text{as } n \to \infty.
\tag{2.12}
$$

We have

$$
\begin{aligned}
\left(\frac{1}{n}\operatorname{tr} P(m_n^1, O_n)P(m_n^2, O_n')\right)^2 &= \left(\frac{1}{n}\right)^2 \left(\sum_{i=1}^{m_n^2} o_{.i}'^T P(m_n^1, O_n)o_{.i}'\right)^2 \\
&= \frac{1}{n^2}\sum_{i=1}^{m_n^2} (o_{.i}'^T P(m_n^1, O_n)o_{.i}')^2 \\
&\quad + \frac{1}{n^2}\sum_{i_1 \neq i_2}^{m_n^2} (o_{.i_1}'^T P(m_n^1, O_n)o_{.i_1}')(o_{.i_2}' P(m_n^1, O_n)o_{.i_2}').
\end{aligned}
\tag{2.13}
$$

For fixed O_n' we have

$$
\begin{aligned}
&\int_{\mathcal{O}_n} \left(\frac{1}{n}\operatorname{tr} P(m_n^1, O_n)P(m_n^2, O_n')\right)^2 d\mu_n(O_n) \\
&\quad = \left(\frac{m_n^2}{n}\right)^2 b_1^2 + \frac{1}{n^2}\sum_{i=1}^{m_n^2} \beta_n(o_{.i}') + \frac{1}{n^2}\sum_{i_1 \neq i_2}^{m_n^2} \gamma_n(o_{.i_1}', o_{.i_2}'),
\end{aligned}
\tag{2.14}
$$

from (2.4) and (2.6). The absolute value of the sum of the last two terms is bounded by $\beta_n + \gamma_n$. Therefore,

$$
\mathrm{E}\left(\left(\frac{1}{n}\operatorname{tr} P(m_n^1, O_n)P(m_n^2, O_n')\right)^2\right) = \left(\frac{m_n^2}{n}\right)^2 b_1^2 + \eta_n,
\tag{2.15}
$$

where $\eta_n \to 0$ as $n \to \infty$, so that

$$\text{(2.16)} \qquad \mathrm{E}\left(\left(\frac{1}{n}\operatorname{tr} P(m_n^1, O_n)P(m_n^2, O_n')\right)^2\right) \to (b_1 b_2)^2 \quad \text{as } n \to \infty.$$

From (2.12) and (2.16) we get (2.8) and we are done.

THEOREM 2. Let $\{u_{ij}\} i, j = 1, 2, \cdots$, be i.i.d. random variables having mean 0, variance 1, and satisfying $\mathrm{E}(|u_{11}|^m) < m^{\alpha m}$ for all integers $m > 2$, and for some α. For each n let $U_n = (u_{ij})$, $i = 1, 2, \cdots, n$, $j = 1, 2, \cdots, s$, where $(n/s) \to y > 0$ as $n \to \infty$, and let μ_n be the measure on $\mathcal{O}_n$ induced from $M_n \equiv (1/s)U_n U_n^T$.

If $\{\mu_n\}$ satisfies III, then $\mathrm{E}(u_{11}^4) = 3$.

Proof. Let $F_n(a)$, $a \in [0, \infty)$ be the empirical distribution function of the eigenvalues of M_n. Let $F_y(a)$ be the limiting distribution function which is given in Theorem 2.1 of [7]. Since $F_y(a)$ is continuous for $a \in [0, \infty)$ we can conclude from Theorem 3.2 of [7] that

$$\text{(2.17)} \qquad \sup_{a \in [0,\infty)} |F_n(a) - F_y(a)| \xrightarrow{\text{a.s.}} 0 \quad \text{as } n \to \infty.$$

The functions $F_n(a)$ and $F_y(a)$ are elements of $D_0 = D_0[0, \infty) = \{x \in D[0, \infty): \lim_{t\to\infty} x(t) \text{ exists and is finite}\}$ [8]. From (2.17), it follows that

$$\text{(2.18)} \qquad F_n(a) \xrightarrow{\text{a.s.}} F_y(a) \quad \text{as } n \to \infty \text{ in } D_0.$$

Assume III and let $\{x_n\}$ be given. For our purpose $X_n(t)$ of (1.6) can be constructed directly from M_n. In fact, we have

$$\text{(2.19)} \qquad X_n(F_n(a)) = \frac{\sqrt{n}}{\sqrt{2}}(x_n^T P_a(M_n)x_n - F_n(a)),$$

where $\{P_a(M_a)\}$ is the spectral family of M_n. A simple extension of the material in [4, pp. 144–145] to nondecreasing functions in $D_0[0, \infty)$ and [4, Theorem 4.4] leads us to conclude that III and (2.18) imply

$$\text{(2.20)} \qquad X_n(F_n(a)) \xrightarrow{\mathcal{D}} W^0_{F_y(a)} \equiv W^y_a \quad \text{in } D_0.$$

For every positive integer r, we have

$$\text{(2.21)} \quad \frac{\sqrt{n}}{\sqrt{2}}\left(x_n^T M_n^r x_n - \frac{1}{n}\operatorname{tr} M_n^r\right) = \int_0^\infty a^r\, dX_n(F_n(a)) = -\int_0^\infty r a^{r-1} X_n(F_n(a))\, da,$$

where we have used the fact that with probability 1, $X_n(F_n(a))$ vanishes outside a bounded set.

For any $b > 0$, the mapping that takes $x \in D_0$ to $\int_0^b r a^{r-1} x(a)\, da$ is continuous. Therefore, from [4, Theorem 5.1],

$$\text{(2.22)} \qquad \int_0^b r a^{r-1} X_n(F_n(a))\, da \xrightarrow{\mathcal{D}} \int_0^b r a^{r-1} W_a^y\, da.$$

With the growth condition on $\mathrm{E}(|u_{11}|^m)$ we have from [5] that the maximum eigenvalue of M_n converges almost surely to $(1+\sqrt{y})^2$. Therefore, when $b > (1+\sqrt{y})^2$ we have

$$\text{(2.23)} \quad \int_0^\infty r a^{r-1} X_n(F_n(a))\, da - \int_0^b r a^{r-1} X_n(F_n(a))\, da \xrightarrow{\text{a.s.}} 0 \quad \text{as } n \to \infty.$$

Therefore,

$$(2.24)\qquad \frac{\sqrt{n}}{\sqrt{2}}\left(x_n^T M_n^r x_n - \frac{1}{n}\operatorname{tr} M_n^r\right) \xrightarrow{\mathscr{D}} -\int_0^b r a^{r-1} W_a^y \, da = -\int_0^{(1+\sqrt{y})^2} r a^{r-1} W_a^y \, da.$$

The limiting distribution is thus Gaussian, with mean and variance only depending on W_a^y.

Let $r=1$. We have

$$(2.25)\qquad \sqrt{n}\left(\frac{1}{n}\operatorname{tr} M_n - 1\right) = \frac{1}{\sqrt{ns}}\sum_{i,j}(u_{ij}^2 - 1),$$

which has mean 0 and variance $(1/s)(\mathrm{E}(u_{11}^4)-1)\to 0$ as $n\to\infty$. Therefore, we need only consider $(\sqrt{n}/\sqrt{2})(x_n^T M_n x_n - 1)$. Let $x_n = (1, 0, \cdots, 0)$. Then

$$(2.26)\qquad \frac{\sqrt{n}}{\sqrt{2}}(x_n^T M_n x_n - 1) = \frac{\sqrt{n}}{\sqrt{2}}\left(\frac{1}{s}\sum_j u_{1j}^2 - 1\right) = \frac{\sqrt{n}}{\sqrt{s}}\frac{1}{\sqrt{2}\sqrt{s}}\sum_j (u_{1j}^2 - 1),$$

which from the Central Limit Theorem converges in distribution to $n(0, (y/2)(\mathrm{E}(u_{11}^4) - 1))$. Therefore, III depends on the value of $\mathrm{E}(u_{11}^4)$ which must be 3, because in the Wishart case, u_{11} is $n(0, 1)$.

We remark that from preliminary work, it is believed that $\mathrm{E}(u_{11}^4) = 3$ is enough to ensure (2.24) for all $r \geqq 1$.

REFERENCES

[1] T. W. Anderson, *An Introduction to Multivariate Statistical Analysis*, John Wiley, New York, 1958.

[2] L. V. Arharov, *Limit theorems for the characteristic roots of a sample covariance matrix*, Soviet Math. Dokl., 12 (1971), pp. 1206–1209.

[3] L. Arnold, *On Wigner's semicircle law for the eigenvalues of random matrices*, Z. Wahrsch. Verw. Gebiete, 19 (1971), pp. 191–198.

[4] P. Billingsley, *Convergence of Probability Measures*, John Wiley, New York, 1968.

[5] S. Geman, *A limit theorem for the norm of random matrices*, Ann. Probab., 8 (1980), pp. 252–261.

[6] U. Grenander and J. W. Silverstein, *Spectral analysis of networks with random topologies*, SIAM J. Appl. Math., 32 (1977), pp. 499–519.

[7] D. Jonsson, *Some limit theorems for the eigenvalues of a sample covariance matrix*, Uppsala University, Department of Mathematics, Report No. 6, Uppsala, Sweden.

[8] T. Lindvall, *Weak convergence of probability measures and random functions in the function space $D[0,\infty)$*, J. Appl. Probab., 10 (1973), pp. 109–121.

[9] V. A. Marcenko and L. A. Pastur, *Distribution of eigenvalues for some sets of random matrices*, Math. USSR-Sb., 1 (1967), pp. 457–483.

[10] J. W. Silverstein, *On the randomness of eigenvectors generated from networks with random topologies*, SIAM J. Appl. Math., 37 (1979), pp. 235–245.

[11] K. W. Wachter, *The strong limits of random matrix spectra for sample matrices of independent elements*, Ann. Probab., 6 (1978), pp. 1–18.

[12] E. P. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Ann. of Math., 62 (1955), pp. 548–564.

[13] ———, *On the distribution of the roots of certain symmetric matrices*, Ann. of Math., 67 (1958), pp. 325–327.