The Annals of Probability
1988, Vol. 16, No. 2, 863–875

CONVERGENCE TO THE SEMICIRCLE LAW

BY Z. D. BAI AND Y. Q. YIN[1]

University of Pittsburgh and University of Arizona

This article proves that the spectral distribution of the random matrix $(1/2\sqrt{np})\,(X_pX_p')$, where $X_p = [X_{ij}]_{p\times n}$ and $[X_{ij}\colon\ i, j = 1, 2, \dots]$ has iid entries with $EX_{11}^4 < \infty$, $\mathrm{Var}(X_{11}) = 1$, tends to the semicircle law as $p \to \infty$, $p/n \to 0$, a.s.

1. Introduction. If A is a $p \times p$ matrix with real eigenvalues $\lambda_1 \le \cdots \le \lambda_p$, the distribution function

$$F^A(x) = \frac{1}{p}\#\{i\colon \lambda_i \le x\}$$

will be called the spectral distribution of A. Here $\#\{\cdots\}$ denotes the cardinality of the set $\{\cdots\}$.

Wigner (1958) proved that if $A_n = [X_{ij}]$ is an $n \times n$ symmetric matrix such that the entries X_{ij}, $1 \le i \le j \le n$, are independent random variables, and X_{ij}, $1 \le i < j \le n$, are distributed as $N(0, \sigma^2)$ but X_{ii}, $1 \le i \le n$, are distributed as $N(0, 2\sigma^2)$, then as $n \to +\infty$,

$$EF^{(1/2\sqrt{n\sigma^2})A_n}(x) \to w(x).$$

Here, $w(x)$ is the so-called semicircle law, i.e.,

$$w'(x) = \begin{cases} \dfrac{2}{\pi}\sqrt{1 - x^2}, & |x| < 1, \\ 0, & |x| > 1. \end{cases}$$

Many workers have been engaged in improving the above result. They either relax the requirements on A_n or strengthen the sense of convergence of the spectral distributions.

On the other hand, many papers are devoted to studying sample-covariance-type random matrices, i.e., matrices of the form $A_p = (1/n)X_pX_p'$, where the columns of X_p are iid random p vectors [cf. Grenander and Silverstein (1977), Wachter (1978), Jonsson (1982), Yin (1984), Yin and Krishnaiah (1985), among others]. They proved the convergence or computed the limits of F^{A_p} as $p \to \infty$ and $p/n \to y$, y a constant. Most of them consider the case $0 < y < \infty$. No limits are semicircle laws.

Received April 1986; revised September 1986.
[1] Partially supported by NSF Grant DMS-87-03090.
AMS 1980 *subject classifications.* Primary 60F99; secondary 62E20.
Key words and phrases. Random matrix, spectral distribution, semicircle law.

In this article we consider a "sample-covariance"-type matrix

$$A = \frac{1}{2\sqrt{np}}(XX' - nI),$$

where X is $p \times n$ with iid entries, and prove that the spectral distribution of A tends to the semicircle law as $p \to \infty$, $n \to \infty$ and $p/n \to 0$. More precisely, we prove the following theorem.

THEOREM. *Let* $\mathbf{X} = \{X_{ij}\colon i, j = 1, 2, \dots\}$ *be an infinite matrix with iid entries. Denote by* X_p *the submatrix* $[X_{ij}\colon i = 1, \dots, p;\ j = 1, \dots, n]$ *of* $\mathbf{X}$. *Here* $n = n(p) \to \infty$ *and* $p/n \to 0$ *as* $p \to \infty$.

If $E|X_{11}|^4 < +\infty$, $\operatorname{Var} X_{11} = 1$ *and* $A_p = (1/2\sqrt{np})(X_pX_p' - nI_p)$, *then as* $p \to \infty$,

$$F^{A_p}(x) \to w(x) \quad a.s.,$$

for any x. *Here* I_p *is the* $p \times p$ *identity matrix, and* $w(\cdot)$ *is the semicircle law.*

2. Proof of the theorem. First, we state some lemmas; their proofs are elementary and are omitted.

LEMMA 2.1. *If* $EX^2 < \infty$, *then for any* $\varepsilon > 0$, $P(|X| > \varepsilon n^{1/4}) = o(1/n)$.

LEMMA 2.2. *If* $E|X|^4 < \infty$, $\exists\ \varepsilon_p > 0$ *such that*

(1) $\varepsilon_p \downarrow 0$, *more slowly than any preassigned speed, and*
(2) $P(|X| \ge \varepsilon_p n^{1/4}) \le \varepsilon_p/n$.

LEMMA 2.3. *Let* $Y_1, Y_2, \dots$ *be iid*, $P(Y_1 = 1) = q = 1 - P(Y_1 = 0)$. *Then*

$$P(Y_1 + \cdots + Y_n - nq \ge n\varepsilon) \le e^{-nh(\varepsilon - qh)},$$

for all $\varepsilon > 0$, $n = 1, 2, \dots$, *and* $0 \le h \le 1/2$.

LEMMA 2.4. *Let* $F(x), G(x)$ *be two empirical distributions of two samples of size* n. *Then*

$$\int |F(x) - G(x)|\, dx = \frac{1}{n}\sum |\lambda_i - \mu_i|,$$

where $\lambda_1 \le \cdots \le \lambda_n$ *and* $\mu_1 \le \cdots \le \mu_n$ *are the two sets of order statistics.*

LEMMA 2.5. *Let* $\{(a_i, b_i)\colon i = 1, 2, \dots\}$ *be the set of all intervals with rational endpoints and with lengths less than* 1. *Let*

$$f_i(x) = \int_{-\infty}^{x} 1_{(a_i, b_i)}(t)\, dt$$

and

$$D(F, G) = \sum_i \left| \int f_i(x)\, d(F(x) - G(x)) \right| \frac{1}{2^i},$$

for any distributions F and G. Then $D(F_n, F) \to 0$ *implies* $F_n \to F$ *weakly.*

LEMMA 2.6. *If* A, B *are two* $p \times p$ *symmetric matrices with eigenvalues* $\{\lambda_1 \le \cdots \le \lambda_p\}$ *and* $\{\mu_1 \le \cdots \le \mu_p\}$, *respectively, then*

$$\sum (\lambda_i - \mu_i)^2 \le \operatorname{tr}(A - B)^2.$$

Now we prove the theorem.

At first we state a proposition. Its proof will be given in Section 3.

PROPOSITION. *For each* p *let* $Y_p = [X_{ijp}]$ *be a* $p \times n$ *random matrix with iid entries,* $n = n(p) \to \infty$, $p/n \to 0$ (*as* $p \to \infty$), *such that*

(1) $EX_{11p} = 0$, $EX_{11p}^2 = 1 + \xi_p$, $\xi_p \to 0$ *as* $p \to \infty$; *and*
(2) $|X_{11p}| \le \varepsilon_p n^{1/4}$, *where* $\varepsilon_p \downarrow 0$ *but* $\varepsilon_p p^{1/4} \uparrow +\infty$, *as* $p \to \infty$.

Let $B_p = [Z_{ij}]$ *be the* $p \times p$ *random matrix defined by*

$$Z_{ii} = 0,$$

$$Z_{ij} = \frac{1}{2\sqrt{np}} \sum_{l=1}^{n} X_{ilp} X_{jlp}, \quad \text{if } i \ne j.$$

Then, the spectral distribution $F^{B_p}(x)$ *of* B_p *tends to the semicircle law* $w(x)$, *as* $p \to \infty$ *for each* x, *with probability* 1.

Now we show that the proposition implies the theorem.

Suppose the proposition has been proved. Choose $\varepsilon_p \downarrow 0$ such that $\varepsilon_p p^{1/4} \uparrow \infty$ and $P(|X_{11}| \ge \varepsilon_p n^{1/4}) \le \varepsilon_p / n$. Define

$$\tilde{X}_p = \left[\tilde{X}_{ij}: i = 1, \dots, p;\ j = 1, \dots, n\right],$$

where

$$\tilde{X}_{ij} = X_{ij} I\left(|X_{ij}| < \varepsilon_p n^{1/4}\right).$$

$I(\cdot)$ denotes the indicator function. Note that $\tilde{X}_{ij}$ depend on p though this is not explicitly indicated.

Let

$$\tilde{A}_p = \frac{1}{2\sqrt{np}}\left(\tilde{X}_p \tilde{X}_p' - n I_p\right).$$

First, we prove that $\sup_x |F^{A_p}(x) - F^{\tilde{A}_p}(x)| \to 0$ as $p \to \infty$ a.s.

Let $\eta_{ij} = 1 - I(|X_{ij}| < \varepsilon_p n^{1/4})$, then by the Fan inequality [Fan (1951)],

$$\sup_x |F^{A_p}(x) - F^{\tilde{A}_p}(x)| = \sup_x |F^{X_p X_p'}(x) - F^{\tilde{X}_p \tilde{X}_p'}(x)| \le \frac{1}{p} \operatorname{rank}\left(X_p - \tilde{X}_p\right)$$

$$\le \frac{1}{p} \sum_{i=1}^{p} \sum_{j=1}^{n} \eta_{ij}.$$

By Lemma 2.3, since $P(\eta_{ij} = 1) = P(|X_{ij}| \geq \varepsilon_p n^{1/4}) = q_p$ (say), if $\delta > 0$,

$$P\Big(\sup_x |F^{A_p}(x) - F^{\tilde{A}_p}(x)| \geq \delta\Big) \leq P\bigg(\frac{1}{p}\sum_i\sum_j \eta_{ij} \geq \delta\bigg)$$
$$= P\bigg(\sum_i\sum_j \eta_{ij} - pnq_p \geq pn\Big(\frac{\delta}{n} - q_p\Big)\bigg)$$
$$\leq \exp\bigg(-nph\Big(\frac{\delta}{n} - q_p - q_p h\Big)\bigg)$$
$$\leq \exp\bigg(-nph\Big(\frac{\delta}{n} - (1+h)\frac{\varepsilon_p}{n}\Big)\bigg)$$
$$\leq \exp\Big(-p\frac{\delta h}{2}\Big),$$

for $\varepsilon_p < \delta/3$; here h can be chosen to be $1/2$. Thus, by the Borel–Cantelli lemma,

$$\sup_x |F^{A_p}(x) - F^{\tilde{A}_p}(x)| \to 0 \quad \text{a.s.}$$

By the Fan inequality, this is also true if we replace $\tilde{A}_p$ by $\hat{A}_p$, where $\hat{A}_p = (1/2\sqrt{np})(Y_pY_p' - nI)$, and $Y_p = [\tilde{X}_j - E\tilde{X}_{ij}\colon\ i = 1, \dots, p;\ j = 1, \dots, n]$. But by the proposition, we have

$$\lim F^{B_p}(x) = w(x), \quad \text{for any } x \text{ with probability } 1.$$

Here B_p is defined in the same way as in the statement of the proposition starting from Y_p. Thus, in order to prove the existence of $\lim F^{A_p}$, it is sufficient to show that

$$D(F^{\hat{A}_p}, F^{B_p}) = \sum_{i=1}^{\infty} \left|\int f_i(x)\, d\big(F^{\hat{A}_p}(x) - F^{B_p}(x)\big)\right| \frac{1}{2^i} \to 0,$$

where $\{f_i\}$ was defined in Lemma 2.5.

By integration by parts and Lemmas 2.4 and 2.6, we have

$$D^2\big(F^{\hat{A}_p}, F^{B_p}\big) \leq \left(\frac{1}{p}\sum_1^p |\lambda_i - \mu_i|\right)^2 \leq \frac{1}{p}\sum_1^p (\lambda_i - \mu_i)^2 \leq \frac{1}{p}\operatorname{tr}\big(\hat{A}_p - B_p\big)^2$$
$$= \frac{1}{4np^2}\sum_{i=1}^p\left(\sum_{l=1}^n (\hat{X}_{il}^2 - 1)\right)^2$$
$$\leq \frac{1}{2np^2}\sum_{i=1}^p\left(\sum_{l=1}^n (\hat{X}_{il}^2 - E\hat{X}_{il}^2)\right)^2 + \frac{n}{2p}(1 - E\hat{X}_{11}^2)^2.$$

Here $\hat{X}_{ij} = \tilde{X}_{ij} - E\tilde{X}_{ij}$, and $\{\lambda_i\}, \{\mu_i\}$ are eigenvalues of $\hat{A}_p$ and B_p, respectively.

For the second term on the right-hand side in the last inequality, we have for p sufficiently large

$$\frac{n}{2p}\left(1 - E\hat{X}_{11}^2\right) \le \frac{n}{2p}\frac{4}{n\varepsilon_p^4}E^2X_{11}^4 \to 0.$$

For the first term on the right-hand side of that inequality, we have

$$\frac{1}{2np^2}\sum_{i=1}^{p}\left(\sum_{l=1}^{n}\left(\hat{X}_{il}^2 - E\hat{X}_{il}^2\right)\right)^2$$
$$= \frac{1}{2np^2}\sum_{i=1}^{p}\sum_{l=1}^{n}\left(\hat{X}_{il}^2 - E\hat{X}_{il}^2\right)^2 + \frac{1}{2np^2}\sum_{i=1}^{p}\sum_{l_1\neq l_2}\left(\hat{X}_{il_1}^2 - E\hat{X}_{il_1}^2\right)\left(\hat{X}_{il_2}^2 - E\hat{X}_{il_2}^2\right)$$
$$= S_{1p} + S_{2p}.$$

But

$$\sum_{p=1}^{\infty} ES_{2p}^2 = \sum_{p=1}^{\infty}\frac{1}{4n^2p^4}\sum_{i=1}^{p}2n(n-1)E^2\left(\hat{X}_{11}^2 - E\hat{X}_{11}^2\right)^2$$
$$\le \sum_{p=1}^{\infty}\frac{1}{2p^3}E^2X_{11}^4 < \infty,$$

so $S_{2p} \to 0$ a.s.

For S_{1p} we have

$$S_{1p} = \frac{1}{2np^2}\sum_{i=1}^{p}\sum_{l=1}^{n}\left(\hat{X}_{il}^2 - E\hat{X}_{il}^2\right)^2$$
$$\le \frac{1}{2np^2}\sum_{i=1}^{p}\sum_{l=1}^{n}\left(\hat{X}_{il}^4 + E^2\hat{X}_{il}^2\right)$$
$$\le \frac{1}{2np^2}\sum_{i=1}^{p}\sum_{l=1}^{n}\left(\hat{X}_{il}^4 - E\hat{X}_{il}^4\right) + \frac{K}{p}EX_{11}^4$$
$$= \Delta_p + \frac{K}{p}EX_{11}^4,$$

for some $K > 0$, and

$$\sum_{p=1}^{\infty} E\Delta_p^2 = \sum_{p=1}^{\infty}\frac{1}{4n^2p^4}\sum_{i=1}^{p}\sum_{l=1}^{n}E\left(\hat{X}_{il}^4 - E\hat{X}_{il}^4\right)^2$$
$$\le \sum_{p=1}^{\infty}\frac{K^2}{4n^2p^4}\left(n^{1/4}\varepsilon_p\right)^4EX_{11}^4 < \infty.$$

Thus $S_{1p} \to 0$ a.s. So the proposition implies the theorem.

3. Proof of the proposition. We know that in order to prove our main theorem, it is enough to prove the proposition.

Let $Y_p = [X_{ij}\colon i = 1, \dots, p,\ j = 1, \dots, n]$, where $X_{ij} = X_{ij}(p)$ are iid random variables such that

$$(1) \qquad EX_{11} = 0, \qquad EX_{11}^2 = 1 + \xi_p, \qquad \xi_p \to 0 \quad \text{as } p \to \infty,$$

$$(2) \qquad |X_n| \le \varepsilon_p n^{1/4},$$

where $\varepsilon_p \downarrow 0$, $\varepsilon_p p^{1/4} \uparrow \infty$.

Let $B_p = [Z_{ij}]$ be the random matrix defined by

$$(3) \qquad Z_{ij} = \begin{cases} 0, & \text{if } i = j, \\ \dfrac{1}{2\sqrt{pn}} \displaystyle\sum_{l=1}^{n} X_{il} X_{jl}, & \text{if } i \ne j. \end{cases}$$

The kth moment of the spectral distribution of B_p will be denoted by $M_k = M_k(p)$. Then

$$(4) \qquad M_k = \frac{1}{p} \operatorname{tr} B_p^k = \frac{1}{p\left(2\sqrt{pn}\right)^k} \sum{}' X_{i_1 j_1} X_{i_2 j_1} X_{i_2 j_2} X_{i_3 j_2} \cdots X_{i_k j_k} X_{i_1 j_k};$$

here Σ' stands for the summation for $i_1, \dots, i_k$ running over $1, \dots, p$ and $j_1, \dots, j_k$ running over $1, \dots, n$, subject to the conditions that $i_1 \ne i_2$, $i_2 \ne i_3, \dots, i_k \ne i_1$.

We will prove

(I) $EM_k \to \int x^k\, dw(x)$ as $p \to \infty$, where $w(\cdot)$ is the semicircle law, and
(II) $\sum_{p=1}^{\infty} \operatorname{var}(M_k) < \infty$.

The conclusion of the proposition is a consequence of (I), (II) and that $w(\cdot)$ is uniquely determined by its moments.

Proof of (I). Definitions:

$$(5) \qquad \psi(e_1, \dots, e_m) := \text{number of distinct entities among } e_1, \dots, e_m,$$

$$(6) \qquad \begin{gathered} i := (i_1, \dots, i_k), \qquad j := (j_1, \dots, j_k), \\ 1 \le i_a \le p, \quad 1 \le j_b \le n, \quad a, b = 1, \dots, k, \end{gathered}$$

$$(7) \qquad r = \psi(i), \qquad c = \psi(j),$$

(8) $\Gamma(i, j)$ denotes the multigraph defined as follows.

Let I-line, J-line be two parallel lines, plot $i_1, \dots, i_k$ on the I-line, plot $j_1, \dots, i_j$ on the J-line; these are vertices. It has $2k$ distinct edges joining $i_1, j_1; i_2, j_1; i_2, j_2; i_3, j_2; \dots; i_k, j_k; i_1, j_k$.

If we merge two edges of $\Gamma(i, j)$ together when they have the same end sets, we get a new graph, denoted by $\overline{\Gamma}(i, j)$.

Let ν_m denote the number of edges in $\overline{\Gamma}(i, j)$, each of which is obtained by merging m edges of $\Gamma(i, j)$. Evidently,

$$\nu_1 + 2\nu_2 + \cdots + 2k\nu_{2k} = 2k. \tag{9}$$

Now define

$$\begin{aligned} A(r, c) := \{(i, j)\colon \psi(i) = r, \psi(j) = c, \\ i_1 \neq i_2, i_2 \neq i_3, \ldots, i_k \neq i_1; \nu_1 = 0\}. \end{aligned} \tag{10}$$

By the above definitions and (4) we can write

$$\begin{aligned} EM_k &= \frac{1}{p(2\sqrt{np})^k} \sum_{r, c=1}^{k} \sum_{A(r, c)} E\left(X_{i_1 j_1} X_{i_2 j_1} X_{i_2 j_2} X_{i_3 j_2} \cdots X_{i_k j_k} X_{i_1 j_k}\right) \\ &= \sum_{r, c=1}^{k} S(r, c), \end{aligned} \tag{11}$$

where

$$S(r, c) := \frac{1}{p(2\sqrt{np})^k} \sum_{A(r, c)} E\left(X_{i_1 j_1} X_{i_2 j_1} X_{i_2 j_2} X_{i_3 j_2} \cdots X_{i_k j_k} X_{i_1 j_k}\right). \tag{12}$$

Now we assert that

$$S(r, c) \to 0, \quad \text{as } p \to \infty, \text{except } r = k/2 + 1, c = k/2. \tag{13}$$

By (1), (2) and (9) we see that

$$\begin{aligned} &\left|E\left(X_{i_1 j_1} X_{i_2 j_1} X_{i_2 j_2} X_{i_3 j_2} \cdots X_{i_k j_k} X_{i_1 j_k}\right)\right| \\ &\quad \leq |EX_{11}|^{\nu_1} |EX_{11}^2|^{\nu_2} \cdots |EX_{11}^{2k}|^{\nu_{2k}} \\ &\quad \leq \left(1 + |\xi_p|\right)^{\nu_2 + \cdots + \nu_{2k}} \left(\varepsilon_p n^{1/4}\right)^{(3-2)\nu_3 + (4-2)\nu_4 + \cdots + (2k-2)\nu_{2k}} \\ &\quad \leq \left(1 + |\xi_p|\right)^k \left(\varepsilon_p n^{1/4}\right)^{2k - 2(\nu_2 + \cdots + \nu_{2k})}, \end{aligned}$$

when $\varepsilon_p n^{1/4} \geq 1$. But $\overline{\Gamma}(i, j)$ is a connected graph with $r + c$ vertices and $\nu_1 + \cdots + \nu_{2k}$ edges, so

$$r + c \leq \nu_1 + \cdots + \nu_{2k} + 1.$$

Since $\varepsilon_p n^{1/4} \uparrow +\infty$,

$$\left|E\left(X_{i_1 j_1} X_{i_2 j_1} \ldots X_{i_k j_k} X_{i_1 j_k}\right)\right| \leq \left(\varepsilon_p n^{1/4}\right)^{2k - 2(r + c - 1)} \left(1 + |\xi_p|\right)^k. \tag{14}$$

Therefore, by (12), (14) and $|A(r, c)| \leq \binom{p}{r} r^k \binom{n}{c} c^k \leq p^r n^c r^k c^k$,

$$\begin{aligned} |S(r, c)| &\leq \frac{1}{p(2\sqrt{np})^k} p^r n^c \left(\varepsilon_p n^{1/4}\right)^{2k - 2(r + c - 1)} \left(1 + |\xi_p|\right)^k r^k c^k \\ &= \varepsilon_p^{2(k - c - r + 1)} p^{r - k/2 - 1} n^{c/2 - r/2 + 1/2} 2^{-k} \left(1 + |\xi_p|\right)^k r^k c^k. \end{aligned} \tag{15}$$

We still need another inequality for $S(r, c)$. Suppose $(i, j) \in A(r, c)$, and let $l_1, \dots, l_c$ be the different values of $j_1, \dots, j_k$. Then

$$E = E\Big(X_{i_1j_1}X_{i_2j_1}X_{i_2j_2}X_{i_3j_2}\cdots X_{i_kj_k}X_{i_1j_k}\Big)$$
$$= \prod_{b=1}^{c} E\prod_{j_a=l_b}\Big(X_{i_aj_a}X_{i_{a+1}j_a}\Big)$$
$$= \prod_{b=1}^{c}\big(EX_{11}^{n_{b1}}EX_{11}^{n_{b2}}\cdots EX_{11}^{n_{bs}}\big),$$

where $n_{b1}, \dots, n_{bs}$ are all ≥ 2 (otherwise, E is 0 and the following inequalities are trivial) and $s \geq 2$ depends on b. So

$$(16)\qquad \begin{aligned}|E| &\leq \prod_{b=1}^{c}\Big(\varepsilon_p n^{1/4}\Big)^{\Sigma_a n_{ba}-2s}\big(1+|\xi_p|\big)^2\\ &\leq \Big(\varepsilon_p n^{1/4}\Big)^{\Sigma_b\Sigma_a n_{ba}-4c}\big(1+|\xi_p|\big)^k\\ &\leq \Big(\varepsilon_p n^{1/4}\Big)^{2k-4c}\big(1+|\xi_p|\big)^k\end{aligned}$$

and

$$(17)\qquad \begin{aligned}|S(r,c)| &\leq \frac{1}{p(2\sqrt{np}\,)^k}p^r n^c\Big(\varepsilon_p n^{1/4}\Big)^{2k-4c}\big(1+|\xi_p|\big)^k\\ &= \varepsilon_p^{2k-4c}p^{r-1-k/2}2^{-k}\big(1+|\xi_p|\big)^k r^k c^k.\end{aligned}$$

Now suppose $r \neq k/2 + 1$ or $c \neq k/2$. Consider the following cases.

Case 1. $r > k/2 + 1$. Since $c + r \leq k + 1$,

$$r - \frac{k}{2} - 1 + \frac{c}{2} - \frac{r}{2} + \frac{1}{2} = \frac{1}{2}(r + c - 1 - k) \leq 0.$$

By (15) and $p/n \to 0$, we get $|S(r, c)| \to 0$.

Case 2. $c > k/2$. Any j-vertex of $\overline{\Gamma}(i, j)$ cannot have degree 1 (since $i_1 \neq i_2, \dots, i_k \neq i_1$). Since every j-vertex has degree > 1, then there would be at least $4c$ edges in $\Gamma(i, j)$. $4c > 2k$, in our case. This is impossible.

Case 3. $r \leq k/2 + 1$, $c \leq k/2$, but not $r = k/2 + 1$ and $c = k/2$. $S(r, c) \to 0$ by (17).

In the following we compute $\lim_{p\to\infty} S(r, c)$ for $r = k/2 + 1$, $c = k/2$. In this case k must be an even number $k = 2m$, say, and then $r = m + 1$, $c = m$.

For $(i, j) \in A(r, c) = A(m + 1, m)$, since $\nu_1 = 0$, and

$$k + 1 = r + c \leq \nu_2 + \cdots + \nu_{2k} + 1 \leq \tfrac{1}{2}(2\nu_2 + \cdots + 2k\nu_{2k}) + 1 = k + 1,$$

we can conclude that

$$\nu_3 = \cdots = \nu_{2k} = 0, \qquad \nu_2 + 1 = r + c = k + 1.$$

So, each edge of $\bar{\Gamma}(i, j)$ is obtained by merging two edges of $\Gamma(i, j)$, and $\bar{\Gamma}(i, j)$ is a tree, since the number of edges equals the number of vertices $(r + c)$ minus one. Counting these facts,

$$S(m+1, m) = \frac{1}{p(2\sqrt{np})^k} \sum_{A(m+1,m)} (1+\xi_p)^k$$
$$= \frac{1}{p(2\sqrt{np})^k} |A(m+1, m)| (1+\xi_p)^k,$$

where $|A(m+1, m)|$ is the cardinal number of the set $A(m+1, m)$.

Two sequences $(a_1, \dots, a_s), (b_1, \dots, b_s)$ will be said to be equivalent iff $a_u = a_v \Leftrightarrow b_u = b_v$, $u, v = 1, \dots, s$, and we write $(a_1, \dots, a_s) \sim (b_1, \dots, b_s)$. But two sequences (i, j) and (i', j') are said to be equivalent iff $i \sim i'$ and $j \sim j'$. Suppose in $A(m+1, m)$ there are N_m equivalence classes with respect to the equivalence relation just defined. Then we have

$$S(m+1, m) = \frac{1}{p(2\sqrt{np})^{2m}} (p)_{m+1} (n)_m N_m (1 + |\xi_p|)^k = 2^{-2m} N_m (1 + o(1));$$

here $(a)_b = a(a-1) \cdots (a-b+1)$.

Next, we compute the number N_m.

To any $j = (j_1, \dots, j_k)$, we define a sequence $h(j) = (h_1, \dots, h_k)$ of ± 1 as follows:

$$h_u = \begin{cases} 1, & \text{if } j_u \text{ is new, i.e., } j_u \notin \{j_1, \dots, j_{u-1}\}, \\ -1, & \text{if } j_u \text{ is old, i.e., } j_u \in \{j_1, \dots, j_{u-1}\}, \end{cases} \quad \text{for } u = 1, 2, \dots, k.$$

Let Q be the set of all those sequences $(q_1, \dots, q_k)$ such that

$$(1)\ \text{each } q_i = \pm 1, \qquad (2)\ \sum_{i=1}^{k} q_i = 0, \qquad (3)\ \sum_{i=1}^{s} q_i \geq 0, \quad \text{for } s = 1, \dots, k.$$

Lemma. $N_m = |Q| = \binom{2m}{m}/(m+1)$.

Proof of the lemma. For any (i, j), let $H(i, j) = h(j)$. If $(i, j), (i', j')$ are in $A(m+1, m)$, and $i \sim i'$, $j \sim j'$, then evidently $H(i, j) = H(i', j')$ (in fact, $j \sim j'$ is enough). If $(i, j) \in A(m+1, m)$, then since $\psi(j) = m = k/2$, $\sum_{u=1}^{k} h_u(j) = 0$, and $\sum_{u=1}^{s} h_u(j) \geq 0$ for any $s = 1, \dots, k$.

Thus H can be regarded as a function defined on the quotient set $A(m+1, m)/\text{equivalent}$, with values in Q.

It is sufficient to show that H is a bijection.

First, we show that H is injective, i.e. for two vectors $(i, j), (i', j')$ of $A(m+1, m)$, if $H(i, j) = H(i', j')$, then (i, j) and (i', j') are equivalent. We prove this by induction. We suppose $H(i, j) = h(j) = h(j') = H(i', j')$.

Evidently, (i_1, i_2, j_1) and (i'_1, i'_2, j'_1) are equivalent since $i_1 \neq i_2$, $i'_1 \neq i'_2$.

Now suppose we have established that $(i_1, \dots, i_l) \sim (i'_1, \dots, i'_l)$ and $(j_1, \dots, j_{l-1}) \sim (j'_1, \dots, j'_{l-1})$. By $H(i, j) = H(i', j')$, j_l, j'_l are both new or old. Suppose j_l and j_l' are both new, then $(j_1, \dots, j_l) \sim (j_1', \dots, j_l')$. In this case i_{l+1} and i_{l+1}' must also both be new. For, suppose $i_{l+1} = i_g$, for some $g < l$, then we would get a cycle contained in the graph with edges $i_g j_g, i_{g+1} j_g, \dots, i_{l-1} j_{l-1}, i_l j_{l-1}, i_l j_l, i_g j_l$, since $i_g j_l, i_l j_l$ are new, it cannot be reduced to a tree by the merging process. This is impossible since $\overline{\Gamma}(i, j)$ is a tree. The same is true for i_{l+1}'. Thus $(i_1, \dots, i_{l+1}) \sim (i_1', \dots, i_{l+1}')$.

Suppose j_l, j_l' are old. First, we note that each j-vertex of the graph $\overline{\Gamma}(i, j)$ cannot have degree 1, since $i_1 \neq i_2, i_2 \neq i_3, \dots, i_k \neq i_1$. But there are $m = k/2$ j-vertices, and $k = 2m$ edges, and the total sum of degrees over all j-vertices is the number of edges $(= k)$. Therefore every j-vertex of $\overline{\Gamma}(i, j)$ for $(i, j) \in A(m+1, m)$ has degree 2.

So, all j_u's are classified into pairs, two j_u's are equal iff they are in a pair.

Suppose g is the largest integer $g < l$ such that j_g has not been paired among $j_1, \dots, j_{l-1}$, thus $j_l \neq j_{l-1}, j_{l-2}, \dots, j_{g+1}$. We assert that $j_l = j_g$. Otherwise if $j_l = j_f$ for some new j_f with $f < g$, we would get that $j_l i_l j_{l-1} \cdots i_{g+1} j_g i_g \cdots i_{f+1} j_f$ contains a cycle, since $i_{g+1} j_g$, $j_g i_g$ are single. For j', since $(j_1, \dots, j_{l-1}) \sim (j_1', \dots, j_{l-1}')$, $j_{l-1}', \dots, j_{g+1}'$ are also paired and j_g' is single. So $j_l' = j_g'$. Therefore $(j_1, \dots, j_l) \sim (j(_1', \dots, j_l')$. Thus $i_l = i_g$ or $i_l = i_{g+1}$. If $i_l = i_g$, then $i_{l+1} = i_{g+1}$, and if $i_l = i_{g+1}$, then $i_{l+1} = i_g$. The same for i'. Therefore $(i_1, \dots, i_{l+1}) \sim (i_1', \dots, i_{l+1}')$.

Thus we have proved that H is injective.

Now we show that H is surjective. Given any $(h_1, \dots, h_k) \in Q$, we will construct a $(i, j) \in A(m+1, m)$ such that $H(i, j) = (h_1, \dots, h_k)$.

Let $i_1 = 1$, $j_1 = 1$, $i_2 = 2$.

Suppose $i_1, \dots, i_l; j_1, \dots, j_{l-1}$ have been defined. If $h_l = 1$, let $j_l = \max_{1 \le f < l} j_f + 1$, $i_{l+1} = \max_{1 \le f \le l} i_f + 1$. Suppose $h_l = -1$. Since $\sum_{u=1}^{l} h_u \ge 0$, there is a largest $g \le l$ such that $\sum_{u=g}^{l} h_u = 0$. Let $j_l = j_g$, $i_{l+1} = i_g$.

Evidently, $\psi(j) = m$, $\psi(i) = m + 1$. We note a simple property that if $h_a + \cdots + h_b = 0$ and the partial sums are ≥ 0, then $i_a = i_{b+1}$. If $b - a = 1$, this is evident from definition. If $h_a + \cdots + h_b = h_a + \cdots + h_c + h_{c+1} + \cdots + h_b$, and $h_a + \cdots + h_c$, $h_{c+1} + \cdots + h_b$ have the same properties as $h_a + \cdots + h_b$, then, by the induction hypothesis, $i_{b+1} = i_{c+1}$, $i_{c+1} = i_a$, so $i_{b+1} = i_a$. If $h_a + \cdots + h_b$ cannot be split, by definition, $i_{b+1} = i_a$.

Now we show that $i_a \neq i_{a+1}$, suppose $i_1 \neq i_2, \dots, i_{a-1} \neq i_a$. If $h_a = 1$, $i_a \neq i_{a+1}$ is true, by definition. If $h_a = -1$, and $g < a$ is maximum such that $h_g + \cdots + h_a = 0$, then $i_{a+1} = i_g$. We must have $h_g = 1$, $h_{g+1} + \cdots + h_{a-1} = 0$ and partial sums ≥ 0. So $i_{g+1} = i_a$, but $i_g \neq i_{g+1}$ so $i_{a+1} \neq i_a$.

$\nu_1 = 0$ is evident.

So H is surjective. And the lemma is proved. [$|Q| = \binom{2m}{m}/(m+1)$ is evident.] □

By using the lemma,

$$EM_k \to \frac{1}{2^{2m}} \frac{1}{m+1} \binom{2m}{m} = \int x^k \, dw(x), \qquad k = 2m. \qquad \square$$

PROOF OF (II). Recalling that M_k is the kth moment of the spectral distribution of B_p, we have

$$M_k = \frac{1}{p(2\sqrt{np})^k} \sum{}' X_{i_1j_1} X_{i_2j_1} X_{i_2j_2} X_{i_3j_2} \cdots X_{i_kj_k} X_{i_1j_k},$$

where the summation is taken over all vectors $(i_1, \dots, i_k, j_1, \dots, j_k)$ subject to the conditions that $1 \le i_a \le p$, $1 \le j_a \le n$, $a = 1, 2, \dots, k$, and $i_1 \ne i_2$, $i_2 \ne i_3, \dots, i_k \ne i_1$.

Now we prove that $\sum_{p=1}^{\infty} \operatorname{Var} M_k < \infty$. We have

$$\begin{aligned}
\operatorname{var} M_k &= E(M_k - EM_k)^2 = EM_k^2 - E^2 M_k \\
&= \frac{1}{p^2(4np)^k} \Bigg(\sum_{i,j}{}' EX_{i_1j_1} X_{i_2j_1} \cdots X_{i_kj_k} X_{i_1j_k} X_{i_{k+1}j_{k+1}} X_{i_{k+2}j_{k+1}} \\
&\qquad \cdots X_{i_{2k}j_{2k}} X_{i_{k+1}j_{2k}} \\
&\qquad - \sum_{i,j}{}' EX_{i_1j_1} X_{i_2j_1} \cdots X_{i_kj_k} X_{i_1j_k} EX_{i_{k+1}j_{k+1}} X_{i_{k+2}j_{k+1}} \\
&\qquad \cdots X_{i_{2k}j_{2k}} X_{i_{k+1}j_{2k}} \Bigg).
\end{aligned}$$

Here $\sum'$ means the summation over all vectors $(i_1, \dots, i_{2k}, j_1, \dots, j_{2k})$ subject to the conditions that $1 \le i_a \le p$, $1 \le j_a \le n$, $a = 1, 2, \dots, 2k$, and $i_1 \ne i_2$, $i_2 \ne i_3, \dots, i_k \ne i_1$, $i_{k+1} \ne i_{k+2}$, $i_{k+2} \ne i_{k+3}, \dots, i_{2k} \ne i_{k+1}$. Let S be the set of all such vectors.

By independence, we can delete all terms in the above sums for which the graphs $\overline{\Gamma} = \overline{\Gamma}(i_1, \dots, i_k; j_1, \dots, j_k)$ and $\overline{\Gamma} = \overline{\Gamma}(i_{k+1}, \dots, i_{2k}; j_{k+1}, \dots, j_{2k})$ do not have common edges. Thus

$$\begin{aligned}
\operatorname{var} M_k = \frac{1}{4^k p^{k+2} n^k} \Bigg(&\sum{}'' EX_{i_1j_1} X_{i_2j_1} \cdots X_{i_kj_k} X_{i_1j_k} X_{i_{k+1}j_{k+1}} X_{i_{k+2}j_{k+1}} \\
&\cdots X_{i_{2k}j_{2k}} X_{i_{k+1}j_{2k}} \\
&- \sum{}'' EX_{i_1j_1} X_{i_2j_1} \cdots X_{i_kj_k} X_{i_1j_k} EX_{i_{k+1}j_{k+1}} X_{i_{k+2}j_{k+1}} \\
&\cdots X_{i_{2k}j_{2k}} X_{i_{k+1}j_{2k}} \Bigg).
\end{aligned}$$

Here the summation $\sum''$ is over S subject to the condition that $\overline{\Gamma}$ and $\overline{\Gamma}'$, defined above, have common edges.

Let the number of vertices of $\overline{\Gamma}, \overline{\Gamma}', \overline{\Gamma}^* = \overline{\Gamma} \cup \overline{\Gamma}'$ on the I line be r, r', r^*, respectively, and the number of vertices of $\overline{\Gamma}, \overline{\Gamma}', \overline{\Gamma}^*$ on the J line be c, c', c^*, respectively. If $\overline{\Gamma}$ and $\overline{\Gamma}'$ have common edges, we must have $c + c' > c^*$, $r + r' > r^*$.

Now we estimate

$$V_1 = \frac{1}{4^k p^{k+2} n^k} \sum{}'' EX_{i_1j_1} X_{i_2j_1} \cdots X_{i_kj_k} X_{i_1j_k} X_{i_{k+1}j_{k+1}} X_{i_{k+2}j_{k+1}} \cdots X_{i_{2k}j_{2k}} X_{i_{k+1}j_{2k}}$$

and

$$V_2 = \frac{1}{4^k p^{k+2} n^k} \sum{}'' E X_{i_1 j_1} X_{i_2 j_1} \cdots X_{i_k j_k} X_{i_1 j_k} E X_{i_{k+1} j_{k+1}} X_{i_{k+2} j_{k+1}} \cdots X_{i_{2k} j_{2k}} X_{i_{k+1} j_{2k}}.$$

First, we have inequalities similar to (14) and (16), namely,

$$\left| E\left(X_{i_1 j_1} X_{i_2 j_1} \cdots X_{i_k j_k} X_{i_1 j_k} X_{i_{k+1} j_{k+1}} X_{i_{k+2} j_{k+1}} \cdots X_{i_{2k} j_{2k}} X_{i_{k+1} j_{2k}} \right) \right| \leq \left(\varepsilon_p n^{1/4} \right)^{4k - 2(c^* + r^* - 1)} \left(1 + |\xi_p|\right)^{2k} \text{ or } \left(\varepsilon_p n^{1/4} \right)^{4k - 4c^*} \left(1 + |\xi_p|\right)^{2k}.$$

In the same way

$$\left| E\left(X_{i_1 j_1} X_{i_2 j_1} \cdots X_{i_k j_k} X_{j_1 j_k} \right) E\left(X_{i_{k+1} j_{k+1}} X_{i_{k+2} j_{k+1}} \cdots X_{i_{2k} j_{2k}} X_{i_{k+1} j_{2k}} \right) \right| \leq \left(\varepsilon_p n^{1/4} \right)^{4k - 2(c + r - 1) - 2(c' + r' - 1)} \left(1 + |\xi_p|\right)^{2k},$$

or

$$\left(\varepsilon_p n^{1/4} \right)^{4k - 4c - 4c'} \left(1 + |\xi_p|\right)^{2k}.$$

So, if we write

$$\begin{aligned} \operatorname{var} M_k = \frac{1}{4^k p^{k+2} n^k} \sum_{r^*, c^*} \sum_{\substack{\psi(i_1, \ldots, i_{2k}) = r^* \\ \psi(i_1, \ldots, j_{2k}) = c^*}} \Big[& E\left(X_{i_1 j_1} X_{i_2 j_1} \cdots X_{i_{k+1} j_{2k}} \right) \\ & - E\left(X_{i_1 j_1} \cdots X_{i_1 j_k} \right) E\left(X_{i_{k+1} j_{k+1}} X_{i_{k+2} j_{k+1}} \cdots X_{i_{k+1} j_{2k}} \right) \Big] \\ = \sum_{r^*, c^*} S^*(r^*, c^*), & \end{aligned}$$

we have

$$\text{(18)} \qquad \begin{aligned} |S^*(r^*, c^*)| &\leq \frac{C}{4^k p^{k+2} n^k} p^{r^*} n^{c^*} \left(\varepsilon_p n^{1/4} \right)^{4k - 2(c^* + r^* - 1)} \\ &\leq C p^{r^* - k - 2} n^{1/2(c^* - r^* + 1)}, \end{aligned}$$

by noting $c + c' > c^*$, $r + r' > r^*$, or

$$\text{(19)} \qquad |S^*(r^*, c^*)| \leq \frac{C}{4^k p^{k+2} n^k} p^{r^*} n^{c^*} \left(\varepsilon_p n^{1/4} \right)^{4k - 4c^*} \leq C p^{r^* - k - 2}.$$

Case 1. $r^* > c^* + 1$. Then $c - r^* + 1 < 0$, and

$$(r^* - k - 2) + \tfrac{1}{2}(c^* - r^* + 1) = \tfrac{1}{2} r^* + \tfrac{1}{2} c^* - k - \tfrac{3}{2} \leq -\tfrac{3}{2}.$$

So, by (18),

$$|S^*(r^*, c^*)| \leq C p^{-3/2}.$$

Case 2. $r^* \leq c^* + 1$. Then since $r^* + c^* \leq 2k$, so $r^* < k$, and $r^* + k - 2 \leq -2$. Using (19) we get

$$|S^*(r^*, c^*)| \leq C p^{-2} \leq C p^{-3/2}.$$

So

$$\operatorname{var} M_k \le C(k) p^{-3/2}$$

and

$$\sum_p \operatorname{var} M_k < \infty. \qquad \square$$

Acknowledgments. We are grateful to Professors C. M. Newman and J. Silverstein. They carefully read the earlier version of this paper and gave some suggestions. We also thank Susan Rollins for her patience in typing many versions of this paper.

REFERENCES

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

GRENANDER, U. and SILVERSTEIN, J. (1977). Spectral analysis of networks with random topologies. *SIAM J. Appl. Math.* **32** 499–519.

JONSSON, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.* **12** 1–38.

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

WIGNER, E. P. (1958). On the distribution of the roots of certain symmetric matrices. *Ann. of Math.* **67** 325–327.

YIN, Y. Q. (1986). Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.* **20** 50–68.

YIN, Y. Q. and KRISHNAIAH, P. R. (1985). Limit theorems for the eigenvalues of the sample covariance matrix when the underlying distribution is isotropic. *Theory Probab. Appl.* **30** 810–816.

CENTER FOR MULTIVARIATE ANALYSIS
UNIVERSITY OF PITTSBURGH
PITTSBURGH, PENNSYLVANIA 15260

DEPARTMENT OF MATHEMATICS
THE UNIVERSITY OF ARIZONA
TUCSON, ARIZONA 85721