Patterned Random Matrices and Method of Moments

Arup Bose*, Rajat Subhra Hazra, and Koushik Saha

In memory of Ashok Prasad Maitra

Abstract

We present a unified approach to limiting spectral distribution (LSD) of patterned matrices via the moment method. We demonstrate relatively short proofs for the LSD of common matrices and provide insight into the nature of different LSD and their interrelations. The method is flexible enough to be applicable to matrices with appropriate dependent entries, banded matrices, and matrices of the form $A_p = \frac{1}{n}XX'$ where X is a $p \times n$ matrix with real entries and $p \to \infty$ with $n = n(p) \to \infty$ and $p/n \to y$ with $0 \leq y < \infty$.

This approach raises interesting questions about the class of patterns for which LSD exists and the nature of the possible limits. In many cases the LSD are not known in any explicit forms and so deriving probabilistic properties of the limit are also interesting issues.

Mathematics Subject Classification (2000). Primary 60B20; Secondary 60B10.

Keywords. Moment method, large dimensional random matrix, eigenvalues, empirical and limiting spectral distributions, Wigner, Toeplitz, Hankel, circulant, reverse circulant, symmetric circulant, sample covariance and XX' matrices, band matrix, balanced matrix, linear dependence.

1. Introduction

Consider a sequence of patterned matrices with random entries. Examples include Wigner, sample variance covariance, Toeplitz and Hankel matrices. Finding the asymptotic properties of the spectrum as the dimension increases has been a major focus of research. We concentrate on such real symmetric matrices and provide an overview of a unified moment approach in deriving their limiting spectral distribution (LSD). After developing a unified framework, we

*Research supported by J.C. Bose National Fellowship, DST, Govt. of India.
Stat. Math. Unit, Indian Statistical Institute, 203 B.T Road, Kolkata 700108.
E-mails: bosearu@gmail.com, rajat_r@isical.ac.in, koushik_r@isical.ac.in.

present selective sketches of proofs for a few of these matrices. We also discuss extensions to situations where the entries come from a dependent sequence or the matrix is of the form XX', thus generalizing the sample variance covariance matrix. Finally we discuss in brief a few other matrices as well as methods for deriving the LSD.

2. Moment Method

Suppose $\{Y_n\}$ is a sequence of real valued random variables. Suppose that there exists some (nonrandom) sequence $\{\beta_h\}$ such that $\mathrm{E}[Y_n^h] \to \beta_h$ for every positive integer h where $\{\beta_h\}$ satisfies *Carleman's condition:*

$$\sum_{h=1}^{\infty} \beta_{2h}^{-1/2h} = \infty.$$

Then there exists a distribution function F, such that for all h,

$$\beta_h = \int x^h dF(x) \text{ and } Y_n \text{ converges to } F \text{ in distribution.}$$

As an illustration, suppose $\{x_i\}$ are i.i.d. random variables with mean zero and variance one and all moments finite. Let $Y_n = n^{-1/2}(x_1 + x_2 + \ldots + x_n)$. By using binomial expansion and taking term by term expectation and then using elementary order calculations, $\mathrm{E}[Y_n^{2k+1}] \to 0$ and $\mathrm{E}[Y_n^{2k}] \to \frac{2k!}{2^k k!}$. Using Stirling's approximation, it can be easily checked that $\{\beta_{2k} = \frac{2k!}{2^k k!}\}$ satisfies Carleman's condition. Since β_{2k} are the $2k$-th moments of standard Normal distribution, $Y_n \xrightarrow{\mathcal{D}} N(0,1)$.

This idea has traditionally been used for establishing the LSD for example of the Wigner and the sample variance covariance matrices. There the trace formula replaces the binomial expansion. However, the calculation/estimation of the leading term and the bounding of the lower order terms lead to combinatorial issues which usually have been addressed on a case by case basis.

3. Limiting Spectral Distribution and Moments

For any random $n \times n$ matrix A_n, if $\lambda_1, \lambda_2, \ldots, \lambda_n$ are all its eigenvalues, then its *empirical spectral distribution function (ESD)* is given by

$$F^{A_n}(x,\ y) = n^{-1} \sum_{i=1}^{n} I\{\mathrm{Re}\lambda_i \le x,\quad \mathrm{Im}\lambda_i \le y\}.$$

The *expected spectral distribution function* of A_n is defined as $\mathrm{E}[F^{A_n}(\cdot)]$. The limiting spectral distribution (LSD) of a sequence $\{A_n\}$ as $n \to \infty$, is the weak limit of the sequence $\{F^{A_n}\}$, if it exists, either almost surely (a.s.) or in probability. *We shall deal with only real symmetric matrices and hence all eigenvalues are real.* The h-th moment of the ESD of A_n has the following nice form:

$$h\text{-th moment of the ESD of } \ A_n = \frac{1}{n}\sum_{i=1}^{n}\lambda_i^h = \frac{1}{n}\mathrm{tr}(A_n^h) = \beta_h(A_n) \text{ (say)}.$$

The following easy Lemma links convergence of moments and LSD. Consider the following conditions:

(M1) For every $h \geq 1$, $\mathrm{E}[\beta_h(A_n)] \to \beta_h$ and $\{\beta_h\}$ satisfies Carleman's condition.

(M2) $\mathrm{Var}[\beta_h(A_n)] \to 0$ for every $h \geq 1$.

(M4) $\sum_{n=1}^{\infty} \mathrm{E}[\beta_h(A_n) - \mathrm{E}(\beta_h(A_n))]^4 < \infty$ for every $h \geq 1$.

Lemma 1. *If (M1) and (M2) hold, then $\{F^{A_n}\}$ converges in probability to F determined by $\{\beta_h\}$. If further (M4) holds, then this convergence is a.s.*

4. A Unified Approach

The sequence of variables which is used to construct the matrix will be called the **input sequence**. It shall be of the form $\{x_i; i \geq 0\}$ or $\{x_{ij};\ i,j \geq 1\}$.

4.1. Link function.

Let $\mathbb{Z}$ be the set of all integers and let $\mathbb{Z}_+$ denote the set of all nonnegative integers. Let $L_n : \{1, 2, \dots n\}^2 \to \mathbb{Z}^d,\ n \geq 1$ be a sequence of functions such that $L_{n+1}(i,j) = L_n(i,j)$ whenever $1 \leq i,j \leq n$. We shall write $L_n = L$ and call it the **link** function and by abuse of notation we write $\mathbb{Z}_+^2$ as the common domain of $\{L_n\}$. The matrices we consider will be of the form $((x_{L(i,j)}))$. Here are some well known matrices and their link functions:

(i) Wigner matrix $W_n^{(s)}$. $L : \mathbb{Z}_+^2 \to \mathbb{Z}^2$ where $L(i,j) = (\min(i,j), \max(i,j))$.

$$W_n^{(s)} = \begin{bmatrix} x_{11} & x_{12} & x_{13} & \dots & x_{1(n-1)} & x_{1n} \\ x_{12} & x_{22} & x_{23} & \dots & x_{2(n-1)} & x_{2n} \\ & & & \vdots & & \\ x_{1n} & x_{2n} & x_{3n} & \dots & x_{(n-1)n} & x_{nn} \end{bmatrix}.$$

(ii) Symmetric Toeplitz matrix $T_n^{(s)}$. $L : \mathbb{Z}_+^2 \to \mathbb{Z}_+$ where $L(i,j) = |i-j|$.

$$T_n^{(s)} = \begin{bmatrix} x_0 & x_1 & x_2 & \dots & x_{n-2} & x_{n-1} \\ x_1 & x_0 & x_1 & \dots & x_{n-3} & x_{n-2} \\ x_2 & x_1 & x_0 & \dots & x_{n-4} & x_{n-3} \\ & & & \vdots & & \\ x_{n-1} & x_{n-2} & x_{n-3} & \dots & x_1 & x_0 \end{bmatrix}.$$

(iii) Symmetric Hankel matrix $H_n^{(s)}$. $L : \mathbb{Z}_+^2 \to \mathbb{Z}_+$ where $L(i,j) = i+j$.

$$H_n^{(s)} = \begin{bmatrix} x_2 & x_3 & x_4 & \dots & x_n & x_{n+1} \\ x_3 & x_4 & x_5 & \dots & x_{n+1} & x_{n+2} \\ x_4 & x_5 & x_6 & \dots & x_{n+2} & x_{n+3} \\ & & & \vdots & & \\ x_{n+1} & x_{n+2} & x_{n+3} & \dots & x_{2n-1} & x_{2n} \end{bmatrix}.$$

(iv) Reverse Circulant $R_n^{(s)}$. $L : \mathbb{Z}_+^2 \to \mathbb{Z}_+$ where $L(i,j) = (i+j) \mod n$.

$$R_n^{(s)} = \begin{bmatrix} x_2 & x_3 & x_4 & \dots & x_0 & x_1 \\ x_3 & x_4 & x_5 & \dots & x_1 & x_2 \\ x_4 & x_5 & x_6 & \dots & x_2 & x_3 \\ & & & \vdots & & \\ x_1 & x_2 & x_3 & \dots & x_{n-1} & x_0 \end{bmatrix}.$$

(v) Symmetric circulant $C_n^{(s)}$. $L : \mathbb{Z}_+^2 \to \mathbb{Z}$ where $L(i,j) = n/2 - |n/2 - |i-j||$.

$$C_n^{(s)} = \begin{bmatrix} x_0 & x_1 & x_2 & \dots & x_2 & x_1 \\ x_1 & x_0 & x_1 & \dots & x_3 & x_2 \\ x_2 & x_1 & x_0 & \dots & x_2 & x_3 \\ & & & \vdots & & \\ x_1 & x_2 & x_3 & \dots & x_1 & x_0 \end{bmatrix}.$$

(vi) Doubly symmetric Hankel matrix DH_n. The symmetric circulant is also a "doubly symmetric" Toeplitz matrix. The doubly symmetric Hankel matrix DH_n with link function $L(i,j) = n/2 - |n/2 - (i+j) \bmod n|,\ 1 \le i,j \le n$ is

$$DH_n = \begin{bmatrix} x_0 & x_1 & x_2 & \dots & x_3 & x_2 & x_1 \\ x_1 & x_2 & x_3 & \dots & x_2 & x_1 & x_0 \\ x_2 & x_3 & x_4 & \dots & x_1 & x_0 & x_1 \\ & & & \vdots & & & \\ x_2 & x_1 & x_0 & \dots & x_5 & x_4 & x_3 \\ x_1 & x_0 & x_1 & \dots & x_4 & x_3 & x_2 \end{bmatrix}.$$

(vii) Palindromic matrices PT_n and PH_n. For these symmetric matrices, the first row is a palindrome. PT_n is given below and PH_n is defined similarly.

$$PT_n = \begin{bmatrix} x_0 & x_1 & x_2 & \dots & x_2 & x_1 & x_0 \\ x_1 & x_0 & x_1 & \dots & x_3 & x_2 & x_1 \\ x_2 & x_1 & x_0 & \dots & x_4 & x_3 & x_2 \\ & & & \vdots & & & \\ x_1 & x_2 & x_3 & \dots & x_1 & x_0 & x_1 \\ x_0 & x_1 & x_2 & \dots & x_2 & x_1 & x_0 \end{bmatrix}.$$

(viii) Sample variance covariance matrix: Often called the S matrix, is defined as

$$A_p(W) = n^{-1} W_p W_p' \quad \text{where} \quad W_p = ((x_{ij}))_{1 \le i \le p, 1 \le j \le n}. \tag{1}$$

It is convenient in this case to think of the link function as a *pair*, given by:

$$L1, L2 : \mathbb{Z}_+^2 \times \mathbb{Z}_+^2 \to \mathbb{Z}_+^2, \; L1(i,j) = (i,j), \; L2(i,j) = (j,i).$$

(ix) Taking a cue from (viii), one may consider XX' where X is a suitable nonsymmetric matrix.

All the link functions above possess Property B given below with $f(x) = x$. Unless otherwise specified, we shall always assume that $f(x) = x$. The general form for f is needed to deal with matrices with dependent entries.

Property B: Let $f : \mathbb{Z}_+^d \to \mathbb{Z}$. Then (L, f) is said to satisfy Property B if

$$\Delta(L, f) = \sup_n \sup_{t \in \mathbb{Z}_+^d} \sup_{1 \le k \le n} \#\{l : 1 \le l \le n, \; f(|L(k,l) - t|) = 0\} < \infty. \tag{2}$$

4.2. Scaling. Assume that $\{x_i\}$ have mean zero and variance 1. Let F_n denote the ESD of $T_n^{(s)}$ and let X_n be the corresponding random variable. Then

$$\mathrm{E}(X_n) = \frac{1}{n} \sum_{i=1}^{n} \lambda_{i,n} = \frac{1}{n} \operatorname{Tr}(T_n^{(s)}) = x_0 \quad \text{and} \quad \mathrm{E}[\mathrm{E}(X_n)] = 0,$$

$$\begin{aligned} \mathrm{E}(X_n^2) &= \frac{1}{n} \sum_{i=1}^{n} \lambda_{i,n}^2 = \frac{1}{n} \operatorname{Tr}\left(T_n^{(s)2}\right) \\ &= \frac{1}{n}[n x_0^2 + 2(n-1)x_1^2 + \ldots + 2x_{n-1}^2]] \quad \text{and} \quad \mathrm{E}[\mathrm{E}(X_n^2)] = n. \end{aligned}$$

Hence, it is appropriate to consider $n^{-1/2} T_n^{(s)}$. The same holds for all matrices except XX', for which the issue is more complicated.

4.3. Trace formula and circuits. Let $A_n = ((a_{L(i,j)}))$. Then the h-th moment of $F^{n^{-1/2}A_n}$ is given by

$$\frac{1}{n}\operatorname{Tr}\left(\frac{A_n}{\sqrt{n}}\right)^h = \frac{1}{n^{1+h/2}} \sum_{1\le i_1, i_2, \ldots, i_h \le n} a_{L(i_1,i_2)} a_{L(i_2,i_3)} \cdots a_{L(i_{h-1},i_h)} a_{L(i_h,i_1)}. \tag{3}$$

Circuit: $\pi : \{0,1,2,\cdots,h\} \to \{1,2,\cdots,n\}$ with $\pi(0) = \pi(h)$ is called a **circuit** of **length** $l(\pi) := h$. The dependence of a circuit on h and n will be suppressed. Clearly, (M1), (M2) and (M4) may be written in terms of circuits. For example,

(M1) $\mathrm{E}[\beta_h(n^{-1/2}A_n)] = \mathrm{E}[\frac{1}{n}\operatorname{Tr}\left(\frac{A_n}{\sqrt{n}}\right)^h] = \frac{1}{n^{1+h/2}}\sum_{\pi:\ \pi \text{ circuit}} \mathrm{E}\,\mathrm{X}_\pi \to \beta_h$
where

$$\mathrm{X}_\pi = x_{L(\pi(0),\pi(1))} x_{L(\pi(1),\pi(2))} \cdots x_{L(\pi(h-2),\pi(h-1))} x_{L(\pi(h-1),\pi(h))}.$$

Matched Circuits: For any π, any $L(\pi(i-1),\pi(i))$ is an **L-value**. If an L-value is repeated exactly e times, we say that it has an **edge of order** e ($1 \le e \le h$). If π has all $e \ge 2$, then it is called **L-matched** (in short *matched*). For any nonmatched π, $\mathrm{E}[\mathrm{X}_\pi] = 0$ and hence only matched π are relevant. If π has only order 2 edges, then it is called **pair matched**.

To verify (M2), we need multiple circuits: k circuits $\pi_1, \pi_2, \cdots, \pi_k$ are **jointly matched** if each L-value occurs at least twice across all circuits. They are **cross matched** if each circuit has at least one L-value which occurs in at least one of the other circuits.

To deal with dependent inputs, we need the following: a π is **(L,f)-matched** if for each i, there is at least one $j \neq i$ such that $f\Big(\big|L(\pi(i-1),\pi(i)) - L(\pi(j-1),\pi(j))\big|\Big) = 0$. The earlier L matching is a special case with $f(x) = x$. The concepts of jointly matching and cross matching can be similarly extended.

Equivalence of circuits: The following defines an equivalence relation between the circuits: π_1 and π_2 are **equivalent** if and only if

$$\big\{L(\pi_1(i-1),\pi_1(i)) = L(\pi_1(j-1),\pi_1(j)) \Leftrightarrow L(\pi_2(i-1),\pi_2(i)) = L(\pi_2(j-1),\pi(j))\big\}.$$

4.4. Words. Any equivalence class induces a partition of $\{1,2,\cdots,h\}$. To any partition we associate a **word** w of length $l(w) = h$ of letters where the first occurrence of each letter is in alphabetical order. For example if $h = 5$ then the partition $\{\{1,3,5\},\{2,4\}\}$ is represented by the word *ababa*.

The class $\Pi(w)$**:** Let $w[i]$ denote the i-th entry of w. The equivalence class corresponding to w will be denoted by

$$\Pi(w) = \{\pi : w[i] = w[j] \Leftrightarrow L(\pi(i-1), \pi(i)) = L(\pi(j-1), \pi(j))\}.$$

The number of partition blocks corresponding to w will be denoted by $|w|$. If $\pi \in \Pi(w)$, then clearly, $\#\{L(\pi(i-1), \pi(i)) : 1 \le i \le h\} = |w|$.

The above notions carry over to words. For instance *ababa* is matched, and *abcadbaa* is nonmatched with edges of order 1, 2 and 4 and the corresponding partition is $\{\{1, 4, 7, 8\}, \{2, 6\}, \{3\}, \{5\}\}$.

For technical reasons it becomes easier to deal with a class bigger than Π. Let

$$\Pi^*(w) = \{\pi : w[i] = w[j] \Rightarrow L(\pi(i-1), \pi(i)) = L(\pi(j-1), \pi(j))\}.$$

4.5. Reduction to bounded case.

We first show that in general, it is enough to work with input sequences which are uniformly bounded. The proof of the following lemma is available in Bose and Sen (2008)[21].

Assumption I $\{x_i, x_{ij}\}$ are independent and uniformly bounded with mean zero and variance 1.

Assumption II $\{x_i, x_{ij}\}$ are i.i.d. with mean zero and variance 1.

Let $\{A_n\}$ be a sequence of $n \times n$ random matrices with link function L_n. Let

$$k_n = \#\{L_n(i,j) : 1 \le i, j \le n\}, \quad \alpha_n = \max_k \#\{(i,j) : L_n(i,j) = k,\ 1 \le i,\ j \le n\}.$$

Lemma 2. *Suppose* $k_n \to \infty$ *and* $k_n \alpha_n = O(n^2)$*. If* $\{F^{n^{-1/2} A_n}\}$ *converges to a nonrandom* F *a.s. when the input sequence satisfies Assumption I, then the same limit holds if it satisfies Assumption II.*

4.6. Only pair matched words contribute.

From the discussion in Section 4.3 it is enough to consider matched circuits. The next lemma shows that we can further restrict attention to pair matched words. Its proof is easy and is available in Bose and Sen (2008)[21].

Let $N_{h,3+}$ be the number of (L, f) matched circuits on $\{1, 2, \ldots, n\}$ of length h with at least one edge of order ≥ 3.

Lemma 3. *(a) If* (L, f) *satisfies Property B, then there is a constant* C *such that*

$$N_{h,3+} \le C n^{\lfloor (h+1)/2 \rfloor} \quad \text{and as } n \to \infty, \quad n^{-(1+h/2)} N_{h,3+} \to 0. \tag{4}$$

(b) Suppose $\{A_n\}$ is a sequence of $n \times n$ random matrices with input sequence $\{x_i\}$ satisfying Assumption I and (L, f) with $f(x) = x$ satisfying Property B. Then

$$\text{if } h \text{ is odd,} \quad \lim_n \mathrm{E}[\beta_h(n^{-1/2}A_n)] = \lim_n \mathrm{E}\left[\frac{1}{n}\operatorname{Tr}\left(\frac{A_n}{\sqrt{n}}\right)^h\right] = 0 \tag{5}$$

$$\text{and if } h = 2k, \text{ then} \quad \sum_{\substack{w \text{ has only} \\ \text{order 2 edges}}} \lim_n \frac{1}{n^{1+k}} |\Pi^*(w) - \Pi(w)| = 0 \tag{6}$$

and provided the limit in the right side below exists,

$$\lim_n \mathrm{E}[\beta_{2k}(n^{-1/2}A_n)] = \sum_{\substack{w \text{ has only} \\ \text{order 2 edges}}} \lim_n \frac{1}{n^{1+k}} |\Pi(w)|. \tag{7}$$

Define, for each fixed matched word w of length $2k$ with $|w| = k$,

$$p(w) = \lim_n \frac{1}{n^{1+k}} |\Pi(w)| = \lim_n \frac{1}{n^{1+k}} |\Pi^*(w)| \tag{8}$$

whenever the limit exists. This limit will be positive and finite only if the number of elements in the set is of exact order n^{k+1}. In that case, Lemma 3 implies that the limiting $(2k)$-th moment is

$$\beta_{2k} = \sum_{w:|w|=k,\ l(w)=2k} p(w).$$

The next Lemma verifies (M4). Its proof is easy and is available in Bose and Sen (2008)[21]. Let $Q_{h,4}$ be the number of quadruples of circuits $(\pi_1, \pi_2, \pi_3, \pi_4)$ of length h which are jointly matched and cross matched with respect to (L, f).

Lemma 4. *(a) If (L, f) obeys Property B, $Q_{h,4} \leq K n^{2h+2}$ for some constant K.*

(b) If $\{A_n\}$ is a sequence of $n \times n$ random matrices with the input sequence $\{x_i\}$ satisfying Assumption I and (L, f) with $f(x) = x$ satisfying Property B, then the following holds. As a consequence, (M4) holds.

$$\mathrm{E}\left[\frac{1}{n}\operatorname{Tr}\left(\frac{A_n}{\sqrt{n}}\right)^h - \mathrm{E}\,\frac{1}{n}\operatorname{Tr}\left(\frac{A_n}{\sqrt{n}}\right)^h\right]^4 = \mathrm{O}(n^{-2}). \tag{9}$$

4.7. Vertex, generating vertex and Carleman's condition.

Any $\pi(i)$ is a **vertex**. It is **generating** if either $i = 0$ or $w[i]$ is the *first* occurrence of a letter. For example, if $w = abbcab$ then $\pi(0), \pi(1), \pi(2), \pi(4)$ are generating. By Property B a circuit is completely determined, *up to a finitely many choices*, by its generating vertices. The number of generating vertices is $|w| + 1$ and hence $|\Pi^*(w)| = \mathrm{O}(n^{|w|+1})$. The following result is due to Bose and Sen (2008)[21].

Theorem 4.1. *Let* $\{A_n = ((x_{L(i,j)}))_{i,j=1}^n\}$ *with the input sequence satisfying Assumption I and* (L, f) *satisfying Property B with* $f(x) = x$*. Then* $\{F^{n^{-1/2}A_n}\}$ *is tight a.s. Any subsequential limit* G *satisfies, for all nonnegative integers* k*, (i)* $\beta_{2k+1}(G) = 0$ *and (ii)* $\beta_{2k}(G) \leq \frac{(2k)!\Delta(L,f)^k}{k!2^k}$*. Hence* G *is sub Gaussian. The (nonrandom) LSD exists for* $\{n^{-1/2}A_n\}$ *iff for every* h*, a.s.,*

$$\lim \beta_h(n^{-1/2}A_n) = \beta_h \text{ (say).} \tag{10}$$

In particular, $\{\beta_h\}$ *automatically satisfies Carleman's condition.*

5. The LSD for Some Specific Matrices

To derive any LSD, it is enough to obtain (10) or (8). It turns out that for $C_n^{(s)}$, PT_n, PH_n and DH_n, $p(w) = 1$ for all w. For other matrices only certain words contribute in the limit. Properties of $p(w)$ for different matrices is given in Tables 1 and 2. Two special type of words which arise are the following:

Symmetric and Catalan words: A pair matched word is *symmetric* if each letter occurs once each in an odd and an even position. A pair matched word is *Catalan* if, sequentially deleting all double letters leads to the empty word. For example, *abccbdda* is a Catalan word whereas *abab* is not. The following result gives the count of these words. The proof of the first part of (a) is also available in Chapter 2 of Anderson, Guionnet and Zeitouni (2009)[3] and Bose and Sen (2008)[21].

Lemma 5. *(a) The number of Catalan words and symmetric words of length* $2k$ *are respectively,* $\frac{(2k)!}{(k+1)!k!}$ *and* $k!$*.*

(b) Let $M_{t,k}$ = $\#\{$ *Catalan words of length* $2k$ *with* $(t+1)$ *even generating vertices and* $(k-t)$ *odd generating vertices*$\}$*. Then*

$$M_{t,k} = \binom{k-1}{t}^2 - \binom{k-1}{t+1}\binom{k-1}{t-1} = \frac{1}{t+1}\binom{k}{t}\binom{k-1}{t}.$$

Proof. (a) For any Catalan word, mark the first and second occurrences of a letter by $+1$ and -1 respectively. For example, *abba* and *abccbdda* are represented respectively by $(1, 1, -1, -1)$ and $(1, 1, 1, -1, -1, 1, -1, -1)$. This provides a bijection between the Catalan words of length $2k$ and sequences $\{u_l\}_{1 \leq l \leq 2k}$ satisfying: each $u_l = \pm 1$, $S_l = \sum_{j=1}^{l} u_j \geq 0 \ \forall \ l \geq 1$ and $S_{2k} = 0$. By reflection principle, the total number of such paths is easily seen to be $\frac{(2k)!}{(k+1)!k!}$. We omit the details. The proof of the second part is trivial.

(b) We know from part (a) that,

$$\#\{\text{Catalan words of length } 2k\} = \#\left\{\{u_l\}_{1 \leq l \leq 2k} : \ u_l = \pm 1, \ S_l \geq 0, \ S_{2k} = 0\right\}.$$

Note that the conditions $\{S_l \geq 0$ and $S_{2k} = 0\}$ imply $u_1 = 1$ and $u_{2k} = -1$. Define

$$N_{e,1} := \#\{l : u_l = 1,\ l \text{ even}\} \text{ and } N_{o,-1} := \#\{l : u_l = -1,\ l \text{ odd}\}.$$

Clearly, $N_{e,1} \leq k-1$ and $N_{o,-1} \leq k-1$. Define

$$\begin{aligned}
C_0 &= \{\{u_l\}:\ \ S_{2k-1} = 1,\ N_{e,1} = t,\ N_{o,-1} = t\}\,, \\
C_1 &= \{\{u_l\}:\ S_l < 0 \text{ for some } l,\ S_{2k-1} = 1,\ N_{e,1} = t,\ N_{o,-1} = t\}\,, \\
C_2 &= \{\{u_l\}:\ S_{2k-1} = -3,\ N_{e,1} = t-1,\ N_{o,-1} = t+1\}\,.
\end{aligned}$$

Then it is easy to see that

$$\begin{aligned}
\#C_0 &= \binom{k-1}{t}^2 \quad \text{and} \quad \#C_2 = \binom{k-1}{t-1}\binom{k-1}{t+1}, \\
M_{t,k} &= \#\{\{u_l\}:\ S_l \geq 0\ \forall l \text{ and } S_{2k-1} = 1,\ \ N_{e,1} = t,\ N_{o,-1} = t\} \\
&= \#C_0 - \#C_1.
\end{aligned}$$

Now we will show $\#C_1 = \#C_2$. Note that for $\{u_l\} \in C_1$ there exist l, such that $S_{l-1} = -1$. Similarly for $\{u_l\} \in C_2$ there exist l, such that $S_{l-1} = -1$. Let

$$\begin{aligned}
l_1 &= \sup\{l :\ S_{l-1} = -1,\ \{u_l\} \in C_1\}, \\
l_2 &= \sup\{l :\ S_{l-1} = -1,\ \{u_l\} \in C_2\}.
\end{aligned}$$

Then

$$u_{l_1} = u_{l_1+1} = 1 \quad \text{and} \quad u_{l_2} = u_{l_2+1} = -1.$$

Now define a map $f : C_1 \to C_2$ as follows: $f(\{u_l\}) = \{u_l'\}$ where

$$u_l' = u_l\ \forall\ l \neq l_1, l_1+1 \text{ and } u_{l_1}' = -u_{l_1},\ u_{l_1+1}' = -u_{l_1+1}.$$

Similarly define $g : C_2 \to C_1$ as $g(\{u_l\}) = \{u_l'\}$ where

$$u_l' = u_l\ \forall\ l \neq l_2, l_2+1 \text{ and } u_{l_2}' = -u_{l_2},\ u_{l_2+1}' = -u_{l_2+1}.$$

It is easy see that f and g are injective. Hence $\#C_1 = \#C_2$. Therefore

$$M_{t,k} = \#C_0 - \#C_1 = \#C_0 - \#C_2 = \binom{k-1}{t}^2 - \binom{k-1}{t+1}\binom{k-1}{t-1}.$$

□

We now provide brief sketches of the steps verifying the existence of the limit (8) for different matrices.

5.1. Wigner matrix: the semicircular law. The **semi-circular law** $\mathcal{L}_W$ arises as the LSD of $n^{-1/2}W_n^{(s)}$. It has the density function

$$p_W(s) = \begin{cases} \frac{1}{2\pi}\sqrt{4-s^2} & \text{if } |s| \leq 2, \\ \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

All its odd moments are zero. The even moments are given by

$$\beta_{2k}(\mathcal{L}_W) = \int_{-2}^{2} s^{2k} p_W(s) ds = \frac{(2k)!}{k!(k+1)!}. \tag{12}$$

Wigner (1958)[47] assumed the entries $\{x_i\}$ to be i.i.d. real Gaussian and established the convergence of $\mathrm{E}[F^{n^{-1/2}W_n^{(s)}}(\cdot)]$ to the semi-circular law (11). Subsequent improvements and extensions can be found in Grenander (1963, pages 179 and 209)[26], Arnold (1967)[2] and Bai (1999)[6].

Theorem 5.1. *Let $W_n^{(s)}$ be the $n \times n$ Wigner matrix with the entries $\{x_{ij} : 1 \leq i \leq j,\ j \geq 1\}$ satisfying Assumption I or II. Then with probability one, $\{F^{n^{-1/2}W_n}\}$ converges weakly to the semicircular law $\mathcal{L}_W$ given in (11).*

Proof. By Lemma 3, Lemma 5 and Theorem 4.1, it is enough to show that for pair matched word w,

$$\frac{1}{n^{1+k}}|\Pi^*(w)| \to 1 \text{ or } 0 \text{ according as } w \text{ is or is not a Catalan word.} \tag{13}$$

Note that if $\pi \in \Pi^*(w), w[i] = w[j] \Rightarrow L(\pi(i-1), \pi(i)) = L(\pi(j-1), \pi(j))$. Then

$$(\pi(i-1),\ \pi(i)) = \begin{cases} (\pi(j-1),\ \pi(j)) & \text{(constraint } C1\text{) or} \\ (\pi(j),\ \pi(j-1)) & \text{(constraint } C2\text{).} \end{cases}$$

For any matched word w, there are k such constraints. Since each constraint is either $C1$ or $C2$, there are at most 2^k choices in all. Let λ be a typical choice of k constraints and $\Pi^*_\lambda(w)$ be the subset of $\Pi^*(w)$ corresponding to λ and so,

$$\Pi^*(w) = \bigcup_\lambda \Pi^*_\lambda(w), \quad \text{a disjoint union.} \tag{14}$$

Fix w and λ. For $\pi \in \Pi^*_\lambda(w)$, consider the graph with vertices $\pi(0), \pi(1), \cdots, \pi(2k)$ Vertices within the following pairs are connected with a single edge:

(i) the pairs $(\pi(i-1), \pi(j-1))$ and $(\pi(i), \pi(j))$ if $w[i] = w[j]$ yields constraint C1.

(ii) the pairs $(\pi(i'-1), \pi(j'))$ and $(\pi(i'), \pi(j'-1))$ if $w[i'] = w[j']$ yields constraint C2.

(iii) the pair $(\pi(0), \pi(2k))$, ensuring that π is indeed a circuit.

So, the graph has a total of $(2k+1)$ edges. These may include loops and double edges. By abuse of notation, $\pi(i)$ thus denotes both, a vertex and its numerical value. One shows by an easy argument that the graph has $(k+1)$ connected components if and only if w is Catalan and all constraints are $C2$. See Bose and Sen (2008)[21] for details. Denote by λ_0 the case when all constraints are $C2$. Note that

$$|\Pi^*_{\lambda_0}(w)| = n^{k+1}. \tag{15}$$

If w is Catalan and not all constraints are $C2$, or, w is not Catalan and λ is arbitrary, then the corresponding graph has at most k connected components and hence $|\bigcup_{\lambda\neq\lambda_0} \Pi^*_\lambda(w)| \leq 2^k n^k$ implying

$$\frac{1}{n^{k+1}} |\bigcup \Pi^*_{\lambda\neq\lambda_0}(w)| \to 0. \tag{16}$$

Now (13) follows by combining (14), (15) and (16), and the proof is complete. □

Remark 1. *Robustness of the semicircle law:*

(i) For the Wigner matrix, $\Delta(L, f) = 1$ with $f(x) = x$ and $\alpha_n = 2$. The following can be proved using the approach described here: If A_n is symmetric where L satisfies $\Delta(L, f) = 1$ with $f(x) = x$ and $\alpha_n = O(1)$ and the input sequence satisfies Assumption I or II, then $F^{n^{-1/2}A_n}$ converges a.s. to the semicircle law. This and other related results on the Wigner matrix may be found in Bannerjee (2010) [4].

(ii) Consider Wigner matrices with the input random variables having possibly different variances which repeat periodically modulo some integer m_n. Then the LSD turns out to be a scaled semicircular distribution. The details are available in Sen (2010) [43].

(iii) Anderson and Zeitouni (2006)[1] considers an $n \times n$ symmetric random matrix with on-or-above-diagonal terms of the form $\frac{1}{\sqrt{n}} f(\frac{i}{n}, \frac{j}{n})\xi_{ij}$ where ξ_{ij} are zero mean unit variance i.i.d. random variables with all moments bounded and f is a continuous function on $[0,1]^2$ such that $\int_0^1 f^2(x,y)dy = 1$. They show that the empirical distribution of eigenvalues converges weakly to the semi-circular law.

5.2. Toeplitz and Hankel matrices.

5.2.1. Standard Toeplitz and Hankel. Their LSD were established by Bryc, Dembo and Jiang (2006)[22] and Hammond and Miller (2005)[28].

Theorem 5.2. *If $\{x_i\}$ satisfies Assumption I or II then a.s., $\{F^{n^{-1/2}T_n^{(s)}}\}$ and $\{F^{n^{-1/2}H_n^{(s)}}\}$ converge to symmetric distributions, $\mathcal{L}_T$ and $\mathcal{L}_H$ respectively.*

$\mathcal{L}_T, \mathcal{L}_H$ have unbounded support. Their moments may be expressed as volumes of certain subsets of hypercubes. Obtaining further properties of the LSD is an open problem.

Proof of Theorem 5.2. We sketch the main steps in the proof for the Toeplitz matrix. Since the L-function satisfies Property B with $f(x) = x$, it is enough to obtain $\lim\limits_{n\to\infty} \frac{1}{n^{1+k}}|\Pi^*(w)|$. From Bryc, Dembo and Jiang (2006)[22], this limit is equal to $\lim_{n\to\infty} \frac{1}{n^{1+k}}|\Pi^{**}(w)|$, where

$$\Pi^{**}(w) = \big\{\pi : w[i] = w[j] \Rightarrow \pi(i-1) - \pi(i) + \pi(j-1) - \pi(j) = 0\big\}.$$

Let $v_i = \pi(i)/n$ and $U_n = \{0, 1/n, 2/n, \ldots, (n-1)/n\}$. The number of elements in $\Pi^{**}(w)$ then equals

$$\#\Big\{(v_0, v_1, \cdots, v_{2k}) : v_0 = v_{2k}, v_i \in U_n \text{ and } \ v_{i-1} - v_i + v_{j-1} - v_j = 0 \text{ if } w[i] = w[j]\Big\}.$$

Let $S = \{0\} \cup \{\min(i,j) : w[i] = w[j], i \neq j\}$ be the set of all indices corresponding to the generating vertices of word w and clearly, $|S| = k+1$. If $\{v_i\}$ satisfy k equations then each v_i is a unique linear combination of $\{v_j\}$ where $j \in S$ and $j \leq i$. Denoting $\{v_i : i \in S\}$ by v_S, we write $v_i = L_i^T(v_S)\ \forall\ i = 0, 1, \cdots, 2k$. Note that these linear functions $\{L_i^T\}$ also depend on the word w. Clearly, $L_i^T(v_S) = v_i$ if $i \in S$ and also summing the k equations would imply $L_{2k}^T(v_S) = v_0$. So

$$|\Pi^{**}(w)| = \#\big\{v_S : L_i^T(v_S) \in U_n \ \text{ for all } \ i = 0, 1, \cdots, n\big\}. \tag{17}$$

Since $\frac{1}{n^{1+k}}|\Pi^{**}(w)|$ is nothing but the $(k+1)$ dimensional Riemann sum for the function $\mathrm{I}(0 \leq L_i^T(v_S) \leq 1,\ \forall\ i \notin S \cup \{2k\})$ over $[0,1]^{k+1}$,

$$\lim_{n\to\infty} \frac{1}{n^{1+k}}|\Pi^{**}(w)| = \underbrace{\int_0^1 \cdots \int_0^1}_{k+1} \mathrm{I}(0 \leq L_i^T(v_S) \leq 1,\ \forall\, i \notin S \cup \{2k\}) dv_S := p_T(w) \tag{18}$$

and $\beta_{2k}(\mathcal{L}_T) = \sum_{\substack{w \text{ matched,} \\ l(w)=2k, |w|=k}} p_T(w)$.

Similarly $\beta_{2k}(\mathcal{L}_H) = \sum_{\substack{w \text{ matched,} \\ l(w)=2k, |w|=k}} p_H(w)$, where $p_H(w)$ is given by

$$\underbrace{\int_0^1 \cdots \int_0^1}_{k+1} \mathrm{I}(0 \leq L_i^H(v_S) \leq 1,\ \forall\ i \notin S \cup \{2k\})\, \mathrm{I}(v_0 = L_{2k}^H(v_S)) dv_S \tag{19}$$

□

5.2.2. Balanced Toeplitz and Hankel matrices. Excluding the diagonal, each variable in the Wigner matrix appears equal number of times (twice). The Toeplitz and Hankel matrices do not have this property and it seems natural to consider the following balanced versions, first considered by Sen (2006)[42]. For proof of the next theorem, see Basak and Bose (2009)[12]. Let

$$BH_n = \begin{bmatrix} \frac{x_1}{\sqrt{1}} & \frac{x_2}{\sqrt{2}} & \frac{x_3}{\sqrt{3}} & \cdots & \frac{x_{n-1}}{\sqrt{n-1}} & \frac{x_n}{\sqrt{n}} \\ \frac{x_2}{\sqrt{2}} & \frac{x_3}{\sqrt{3}} & \frac{x_4}{\sqrt{4}} & \cdots & \frac{x_n}{\sqrt{n}} & \frac{x_{n+1}}{\sqrt{n-1}} \\ \frac{x_3}{\sqrt{3}} & \frac{x_4}{\sqrt{4}} & \frac{x_5}{\sqrt{5}} & \cdots & \frac{x_{n+1}}{\sqrt{n-1}} & \frac{x_{n+2}}{\sqrt{n-2}} \\ & & & \vdots & & \\ \frac{x_n}{\sqrt{n}} & \frac{x_{n+1}}{\sqrt{n-1}} & \frac{x_{n+2}}{\sqrt{n-2}} & \cdots & \frac{x_{2n-2}}{\sqrt{2}} & \frac{x_{2n-1}}{\sqrt{1}} \end{bmatrix}.$$

$$BT_n = \begin{bmatrix} \frac{x_0}{\sqrt{n}} & \frac{x_1}{\sqrt{n-1}} & \frac{x_2}{\sqrt{n-2}} & \cdots & \frac{x_{n-2}}{\sqrt{2}} & \frac{x_{n-1}}{\sqrt{1}} \\ \frac{x_1}{\sqrt{n-1}} & \frac{x_0}{\sqrt{n}} & \frac{x_1}{\sqrt{n-1}} & \cdots & \frac{x_{n-3}}{\sqrt{3}} & \frac{x_{n-2}}{\sqrt{2}} \\ \frac{x_2}{\sqrt{n-2}} & \frac{x_1}{\sqrt{n-1}} & \frac{x_0}{\sqrt{n}} & \cdots & \frac{x_{n-4}}{\sqrt{4}} & \frac{x_{n-3}}{\sqrt{3}} \\ & & & \vdots & & \\ \frac{x_{n-1}}{\sqrt{1}} & \frac{x_{n-2}}{\sqrt{2}} & \frac{x_{n-3}}{\sqrt{3}} & \cdots & \frac{x_1}{\sqrt{n-1}} & \frac{x_0}{\sqrt{n}} \end{bmatrix}.$$

Theorem 5.3. *If $\{x_i\}$ satisfies Assumption I or II then $\{F^{BT_n}\}$, $\{F^{BH_n}\}$ converge a.s. to symmetric distributions having unbounded support and finite moments.*

5.3. The reverse circulant and the palindromic matrices.

Bose and Mitra (2002)[18] studied the LSD of $n^{-1/2}R_n^{(s)}$ under finiteness of the third moment. Massey, Miller and Sinsheimer (2007)[34] established the Gaussian limit for $F^{n^{-1/2}PT_n}$ and $F^{n^{-1/2}PH_n}$. The following result may be proved using arguments similar but simpler than those given earlier for the Wigner and Toeplitz matrices. See Bose and Sen (2008)[21] for details. Let $\mathcal{L}_R$ be the distribution with density and moments

$$f_R(x) = |x|\exp(-x^2),\ -\infty < x < \infty,\ \beta_{2k+1}(\mathcal{L}_R) = 0 \text{ and } \beta_{2k}(\mathcal{L}_R) = k!\ k \geq 0.$$

Theorem 5.4. *If $\{x_i\}$ satisfies Assumption I or II, then a.s., $\{F^{n^{-1/2}R_n^{(s)}}\}$ converges to $\mathcal{L}_R$ and $\{F^{n^{-1/2}A_n}\}$ for $A_n = PT_n, C_n^{(s)}, PH_n$ and DH_n, converge to the standard Gaussian distribution.*

Proof. First consider $R_n^{(s)}$. It is enough to show that

(i) If w is pair matched and not symmetric then $\lim_{n\to\infty} \frac{1}{n^{k+1}}|\Pi^*(w)| = 0$.

(ii) If w is symmetric then for every choice of the generating vertices there is exactly one choice for the nongenerating vertices.

Proof (i) Since w is pair matched, let $\{(i_s, j_s), 1 \leq s \leq k\}$ be such that $w[i_l] = w[j_l]$ and $j_s, 1 \leq s \leq k$, is in ascending order and $j_k = 2k$. We use the notation from the proof of Theorem 5.2. So, $|\Pi^*(w)| = \sum_{\nu=(\nu_1,\nu_2,\ldots,\nu_k)\in\{-1,0,1\}^k}$ $\#\Big\{(v_0, v_1, \cdots, v_{2k}) : v_0 = v_{2k}, v_i \in U_n, \text{ and } v_{i_s-1} + v_{i_s} - v_{j_s-1} - v_{j_s} = \nu_s\Big\}$. Observe that $v_i = L_i^H(v_S) + a_i^{(\nu)}, i \notin S$ for some integer $a_i^{(\nu)}$. As in the Hankel case, we easily reach the following equality (compare with (19)),

$\lim_{n\to\infty} \frac{1}{n^{k+1}} |\Pi^*(w)| =$

$$\sum_\nu \underbrace{\int_0^1 \cdots \int_0^1}_{k+1} \mathrm{I}(0 \leq L_i^H(v_S) + a_i^{(\nu)} \leq 1, \ \forall\, i \notin S \cup \{2k\})\, \mathrm{I}(v_0 = L_{2k}^H(v_S) + a_{2k}^{(\nu)})\, dv_S.$$

For the integral to be non zero, we must have $v_0 = L_{2k}^H(v_S) + a_{2k}^{(\nu)}$.

Let $t_i = v_{i-1} + v_i$. From the definition of $\Pi^*(w)$, $v_{2k} = v_{2k} + \sum_{s=1}^k \alpha_s (t_{i_s} - t_{j_s} - \nu_s)$ for some $\{\alpha_i\}$. We choose them as follows: Let $\alpha_k = 1$. Having fixed α_k, α_{k-1}, $\ldots$, α_{s+1}, we choose α_s as follows: (a) if $j_s + 1 \in \{i_m, j_m\}$ for some $m > s$, then set $\alpha_s = \pm\alpha_m$ according as $j_s + 1$ equals i_m or j_m, (b) if there is no such m, choose α_s arbitrarily. By this choice of $\{\alpha_s\}$, $v_{2k} = v_{2k} + \sum_{s=1}^k \alpha_s (t_{i_s} - t_{j_s} - \nu_s) = L_{2k}^H(v_S) + a_{2k}^{(\nu)}$. Hence $v_{2k} + \sum_{s=1}^k \alpha_s (t_{i_s} - t_{j_s}) + a_{2k}^{(\nu)} - v_0 = 0$ and thus coefficient of each v_i in the left side has to be zero including the constant term. This implies that w is symmetric, proving (i).

(ii) First fix the generating vertices. Then we determine the nongenerating vertices from left to right. Consider $L(\pi(i-1), \pi(i)) = L(\pi(j-1), \pi(j))$ where $i < j$ and $\pi(i-1), \pi(i)$ and $\pi(j-1)$ have been determined. We rewrite it as

$$\pi(j) \bmod n = Z \quad \text{where} \quad Z = (L(\pi(i-1), \pi(i)) - \pi(j-1)) \bmod n \in \{0, 1, \ldots, n\}.$$

This determines $\pi(j)$ uniquely, since $1 \leq \pi(j) \leq n$. Continuing, we obtain the whole circuit uniquely and the result is proved for $\{n^{-1/2} R_n^{(s)}\}$.

For other matrices, similar arguments show that (ii) holds for *all* pair-matched words. We omit the details. This completes the proof. □

Remark 2. *In a recent paper, Jackson, Miller and Pham (2010) [29] studied the situation when there is more than one palindrome in the first row of a symmetric Toeplitz matrix and used method of moments to show that under certain moment assumptions, the limiting spectral distribution exists and has an unbounded support.*

5.4. XX' matrices.

Table 1. Words and moments for symmetric X.

MATRIX	w Cat.	w sym. not Cat.	Other w	β_{2k} or LSD
$C_n^{(s)}$	1	1	1	$\frac{(2k)!}{2^k k!}$, $N(0,1)$
PT_n	1	1	1	ditto
PH_n	1	1	1	ditto
DH_n	1	1	1	ditto
$R_n^{(s)}$	1	1	0	$k!$, $\mathcal{L}_R$
$T_n^{(s)}$	1	$0 < p_T(w) < 1$	$0 < p_T(w) < 1$	$\frac{(2k)!}{k!(k+1)!} \leq \beta_{2k} \leq \frac{(2k)!}{k!2^k}$
$H_n^{(s)}$	1	$0 < p_H(w) = p_T(w) < 1$	0	$\frac{(2k)!}{k!(k+1)!} \leq \beta_{2k} \leq k!$
$W_n^{(s)}$	1	0	0	$\frac{(2k)!}{k!(k+1)!}$, $\mathcal{L}_W$

5.4.1. Sample covariance matrix. For historical information on the LSD of $S = A_p(W)$, see Bai and Yin (1988)[9], Marčenko, and Pastur (1967)[33], Grenander and Silverstein (1977)[27], Wachter (1978)[46], Jonsson (1982)[30], Yin and Krishnaiah (1985)[49], Yin (1986)[48] and Bai and Zhou (2008)[7].

We first describe the Marčenko-Pastur law denoted by $\mathcal{L}_{MPy}$: It has a positive mass $1 - \frac{1}{y}$ at the origin if $y > 1$. Elsewhere it has a density:

$$p_{MPy}(x) = \begin{cases} \frac{1}{2\pi xy}\sqrt{(b-x)(x-a)} & \text{if } a \leq x \leq b, \\ 0 & \text{otherwise} \end{cases} \tag{20}$$

where $a = a(y) = (1-\sqrt{y})^2$ and $b = b(y) = (1+\sqrt{y})^2$.

Moments of $\mathcal{L}_{MPy}$ are (see Bai (1999)[6] or Bai and Silverstein (2006)[8]):

$$\beta_k(\mathcal{L}_{MPy}) = \sum_{t=0}^{k-1} \frac{1}{t+1}\binom{k}{t}\binom{k-1}{t} y^t, \ k \geq 1.$$

Theorem 5.5. *(a)Suppose $\{x_{ij}\}$ satisfy Assumption I or II and $p \to \infty$. If $p/n \to y \in (0,\infty)$, then $\{F^{A_p(W)}\}$ converges to $\mathcal{L}_{MPy}$ a.s..*

(b) Suppose $\{x_{ij}\}$ satisfy Assumption I or they are i.i.d. with mean 0, variance 1 and bounded forth moment and $p \to \infty$. If $p/n \to 0$, then $\{F^{\sqrt{\frac{n}{p}}(A_p(W)-I_p)}\}$ converges to $\mathcal{L}_W$ a.s. where I_p is the identity matrix of order p.

Proof. (a) We apply mutas mutandis the proof given for the Wigner matrix.

$$\beta_k(S) = p^{-1}n^{-k}\sum_{\pi} x_{L1(\pi(0),\pi(1))}x_{L2(\pi(1),\pi(2))}x_{L1(\pi(2),\pi(3))}\cdots x_{L2(\pi(2k-1),\pi(2k))}.$$

A circuit π now has the *non uniform* range $1 \le \pi(2m) \le p,\ 1 \le \pi(2m+1) \le n$. It is said to be matched if it is matched within the same $Li,\ i = 1,\ 2$ or across. For any w, let $\widetilde{\Pi}(w)$ be the possibly larger class of circuits with the range $1 \le \pi(i) \le \max(p, n),\ 1 \le i \le 2k$. Likewise define $\widetilde{\Pi}^*(w)$.

Lemma 3 remains valid in the present case. See Bose and Sen (2008)[21]. Hence only the pair matched circuits potentially contribute and we need to calculate

$$\lim_n \sum_w \sum_{\pi \in \Pi(w)} \frac{1}{n^k p} \mathrm{E}[x_{L1(\pi(0),\pi(1))} \cdots x_{L2(\pi(2k-1),\pi(2k))}] = \lim_n \sum_{\substack{w:\ \text{matched},\\ |w|=k}} \frac{|\Pi^*(w)|}{n^k p}.$$

We need *exactly* $(k+1)$ generating vertices (hence k nongenerating vertices) for a contribution. There is an obvious similarity between the matching restrictions here and the ones we encountered for the Wigner link function. Note that $L1(\pi(i-1), \pi(i)) = L2(\pi(j-1), \pi(j))$ for $i \ne j$ implies a $C2$ constraint. On the other hand, $Lt(\pi(i-1), \pi(i)) = Lt(\pi(j-1), \pi(j)),\ t = 1$ or 2, yields a $C1$ constraint. However, unlike the Wigner matrix, $w[i] = w[j]$ implies exactly one of the constraints is satisfied ($C1$ if i and j have same parity and $C2$ otherwise). Hence there is a *unique* $\bar{\lambda}$ (depending on w) such that $\Pi^*(w) = \Pi^*_{\bar{\lambda}}(w)$.

As before, let λ_0 be the index when all constraints in $\Pi^*_{\lambda_0}(w)$ are $C2$. Let $\widetilde{\Pi}^*_{\bar{\lambda}}(w)$ denote the class $\Pi^*_{\lambda}(w)$ but where $1 \le \pi(i) \le \max(p, n),\ i = 0, 1, \ldots 2k$.

If w is not Catalan then it is easy to see that $\bar{\lambda} \ne \lambda_0$. Hence it follows that

$$n^{-k} p^{-1} |\Pi^*_{\bar{\lambda}}(w)| \le C[\max(p, n)]^{-(k+1)} \left| \bigcup_{\lambda \ne \lambda_0} \widetilde{\Pi}^*_{\lambda}(w) \right| \to 0.$$

For any $0 \le t \le k-1$, if w is Catalan with $(t+1)$ generating even vertices (with range p) and $(k-t)$ generating odd vertices (with range n) then

$$\lim_{n\to\infty} n^{-k} p^{-1} |\Pi^*_{\lambda_0}(w)| = \lim_{n\to\infty} n^{-k} p^{-1} (p^{t+1} n^{k-t}) = y^t.$$

Hence $\lim \mathrm{E}[\beta_k(S)] = \sum_{t=0}^{k-1} M_{t,k} y^t$ and the result follows from Lemma 5 (b).

(b) We assume that the input sequence satisfy Assumption I. The proof when they are i.i.d. with bounded forth moment is available in Bai (1999)[6]. We now sketch briefly how the Wigner link function and hence the semicircle law appears. The details can be found in Bose and Sen (2008)[21] Theorem 5. Note that $\mathrm{E}\big[\beta_k(\sqrt{\frac{n}{p}}(A_p(W) - I_p)\big] = \frac{1}{n^{k/2} p^{1+k/2}} \sum_\pi \mathrm{E}[X_\pi]$, where X_π is equal to

$$\big(x_{\pi(0),\pi(1)} x_{\pi(2),\pi(1)} - \delta_{\pi(0),\pi(2)}\big) \cdots \big(x_{\pi(2k-2),\pi(2k-1)} x_{\pi(2k),\pi(2k-1)} - \delta_{\pi(2k-2),\pi(2k)}\big),$$

with $\delta_{ij} = \mathrm{I}\{i = j\}$. Now $\mathrm{E}[X_\pi] = 0$ if $(\pi(2i), \pi(2i-1))$ or $(\pi(2i), \pi(2i+1))$ occurs only once in the product and if for some j, the value of $\pi(2j+1)$ does not occur at least twice among $\pi(1), \pi(3), \ldots, \pi(2k-1)$. Define a graph $G = (V, E)$ with $V = V_1 \cup V_2$ and $V_1 = \{\pi(2j), 0 \leq j \leq k\}$ and $V_2 = \{\pi(2j-1), 1 \leq j \leq k\}$. Let $E = \{(\pi(2l), \pi(2l+1)), (\pi(2l+2), \pi(2l+1)) : 0 \leq l \leq k-1\}$ (multiple edges count as one). Fix a matching (out of finitely many choices) among the even vertices and one among the odd vertices, such that $\mathrm{E}[X_\pi] \neq 0$. There are at most $p^{|V_1|} n^{|V_2|}$ corresponding circuits. So the contribution of that term is

$$O\left(\frac{p^{|V_1|} n^{|V_2|}}{p^{k/2+1} n^{k/2}}\right) = O\left(\left(\frac{p}{n}\right)^{k/2-|V_2|}\right). \tag{21}$$

If k is odd then $|V_2| < k/2$ and since $p/n \to 0$, (21) immediately implies that $\mathrm{E}[\beta_k(A_p)] \to 0$.

If $k = 2m$, we look for π which produce nontrivial contribution. From (21), we must have $|V| = 2m+1, |E| = 2m, |V_2| = m$ and $|V_1| = m+1$. Observe that:

(i) $|V_2| = m$ implies a pair partition of odd vertices. Denote it by a word w of length k. So, $\pi(2i-1) = \pi(2j-1)$ iff $w[i] = w[j]$.

(ii) Each pair in E must occur exactly twice.

(iii) If $(\pi(2l), \pi(2l+1)) = (\pi(2l+2), \pi(2l+1))$ or equivalently $\pi(2l) = \pi(2l+2)$, then $\mathrm{E}[X_\pi] = 0$. So, consecutive even vertices cannot be equal.

(iv) Note that (ii) and (iii) together imply that $\mathrm{E}[X_\pi] = 1$. Suppose

$$w[i] = w[j] \text{ i.e. } \pi(2i-1) = \pi(2j-1) \tag{22}$$

and they are different from the rest of the odd vertices. If we fixed w, then independent of w, there are exactly $N_1(n) = n(n-1)\ldots(n-m+1)$ choices of odd vertices satisfying the pairing imposed by w.

Consider the four pairs of vertices from E, $(\pi(2i-2), \pi(2i-1))$, $(\pi(2i), \pi(2i-1))$, $(\pi(2j-2), \pi(2j-1))$ and $(\pi(2j), \pi(2j-1))$.

By (22) and (ii), they have to be matched in pairs among themselves. Also, (iii) rules out the possibility that the first pair is matched with the second and the third is matched with the fourth. So the other two combinations are the only possibilities. It is easy to verify that this is the same as saying that

$$L(\pi(2i-2), \pi(2i)) = L(\pi(2j-2), \pi(2j)) \tag{23}$$

where L is the Wigner link function. Let $\pi^*(i) = \pi(2i)$. Equation (23) implies, π^* is a matched circuit of length k. Let $\Pi^*(w)$= all circuits π^* satisfying Wigner link function. Then $\lim \frac{1}{p^{m+1}} |\Pi^*(w)| = 1$ or 0 according as w is or is not Catalan. Hence, the following equalities hold and (M1) is established.

$$\begin{aligned}
\lim_{n,p} \mathrm{E}[\beta_k(A_p)] &= \lim_{n,p} \frac{1}{p^{m+1} n^m} \sum_{w:\text{matched},\ |w|=m} N_1(n) |\Pi^*(w)| \\
&= \lim_{p} \frac{1}{p^{m+1}} \sum_{w:\text{matched},\ |w|=m} |\Pi^*(w)| = \frac{(2m)!}{(m+1)!m!}.
\end{aligned}$$

□

Remark 3. *Simulated eigenvalue distribution of the sample autocovariance matrix and a close cousin of it were given in Sen (2006)[42]. The former is defined as* $\Gamma_n = n^{-1}((\sum_{t=1}^{n-|i-j|} x_t x_{t+|i-j|}))_{i,j=1,\ldots,n}$. *This is also a Toeplitz matrix but with a dependent input sequence. Assuming that* $\{x_t\}$ *satisfies Assumption II, Basak (2009)[10] showed that the LSD exists, and Basak, Bose and Sen (2010)[14] showed that the LSD exists when* $x_t = \sum_{j=0}^{d} a_j \epsilon_{t-j}$ *with* $\{\epsilon_t\}$ *satisfying Assumption II. They also showed that the modified matrix* $\bar{\Gamma}_n = n^{-1}((\sum_{t=1}^{n} x_t x_{t+|i-j|}))_{i,j=1,\ldots,n}$ *which is not nonnegative definite also has an LSD.*

5.4.2. XX' matrices with Toeplitz, Hankel and reverse circulant structures. Let $L_p : \{1, 2, \ldots, p\} \times \{1, 2, \ldots, n = n(p)\} \to \mathbb{Z}$ be a sequence of link functions. Define the following generalization of the S matrix:

$$A_p(X) = (1/n)XX', \quad \text{where} \quad X = ((x_{L_p(i,j)}))_{1 \le i \le p,\ 1 \le j \le n}.$$

In particular, consider the following choices for X:

(Asymmetric) Toeplitz $T = ((x_{i-j}))_{p \times n}$.

(Asymmetric) Hankel H with (i, j)th entry x_{i+j} if $i > j$ and $x_{-(i+j)}$ if $i \le j$.

(Asymmetric) reverse circulant R with $L(i, j) = (i + j) \bmod n$ for $i \le j$ and $-[(i + j) \bmod n]$ for $i > j$.

(Asymmetric) circulant C where $L(i, j) = (n - i + j) \bmod n$.

Also let $H_p^{(s)}$ and $R_p^{(s)}$ be the $p \times n$ rectangular versions of $H_n^{(s)}$ and $R_n^{(s)}$.

Assumption III. $\{x_i\}$ *are independent with mean zero and variance 1. Further,* $\lambda \ge 1$ *is such that* $p = O(n^{1/\lambda})$ *and* $\sup_i \mathrm{E}(|x_i|^{4(1+1/\lambda)+\delta}) < \infty$ *for some* $\delta > 0$.

For a proof of the following theorem see Bose, Gangopadhyay and Sen (2009)[20].

Theorem 5.6. *(a) If Assumption I or II holds and* $\frac{p}{n} \to y \in (0, \infty)$, *then* $\{F^{A_p(X)}\}$, *where* X *is* T, H, R *or* C, *converge in distribution a.s. to nonrandom distributions which do not depend on the distribution of* $\{x_i\}$.

(b) If Assumption III holds, $p \to \infty$ *and* $p/n \to 0$, *then* $F^{\sqrt{\frac{n}{p}}(A_p(X)-I_p)} \to \mathcal{L}_T$ *a.s. for* X *equal to* T, H, R, C, $H_p^{(s)}$ *or* $R_p^{(s)}$.

Table 2. Words and moments for XX' matrices.

MATRIX	w Cat.	Other w	β_k and LSD
$p/n \to 0$ $\sqrt{\frac{n}{p}}(S - I_p)$ $(S = n^{-1}W_pW_p')$ $\sqrt{\frac{n}{p}}(A_p(X) - I_p)$ $(X = T,\ H,\ R,\ C)$	1 (Cat. in p)	0	$\frac{(2k)!}{k!(k+1)!}$, $\mathcal{L}_W$ $\mathcal{L}_T$
$p/n \to y \neq 0, \infty$ $S = n^{-1}W_pW_p'$ $A_p(X)$ $(X = T, H, R_p, C_p)$	1	0	$\sum_{t=0}^{k-1} \frac{1}{t+1}\binom{k}{t}\binom{k-1}{t} y^t$, $\mathcal{L}_{MPy}$ different, but universal

5.5. Band matrices. If the top right corner and the bottom left corner elements of a matrix are zeroes, we call it a band matrix. The amount of banding may change with the dimension of the matrix and may alter the LSD drastically. See for example Popescu (2009)[41]. In this section we discuss the Toeplitz, Hankel and circulant band matrices. Similar band matrices have been considered by Kargin (2009)[31] and Liu and Wang (2009)[32]. Proofs of the next two theorems are available in Basak and Bose (2009)[11]. Let $\{m_n\}$ be a sequence of integers. For simplicity we write m for m_n. Consider the following assumptions.

Assumption I* $\{x_i\}$ are independent with mean 0 and variance 1 and satisfy
(i) $\sup \mathrm{E}[|x_i|^{2+\delta}] < \infty$ for some $\delta > 0$,
(ii) For all large t, $\lim n^{-2} \sum_{i=0}^{n} \mathrm{E}[|x_i|^4 I(|x_i| > t)] = 0$.

Assumption IV $\{m_n\} \to \infty$ and $\lim\limits_{n\to\infty} m_n/n = \alpha < \infty$.

Assumption V $\sum\limits_{n=1}^{\infty} m_n^{-2} < \infty$. (Holds trivially when $\alpha \neq 0$).

(i) **Type I banding**. For any A_n, the Type I band matrix A_n^b is the matrix A_n with input $\{x_i I(i \leq m) + 0I(i > m)\}$.

Let $N(0, 2)$ denote a normal random variable with mean zero and variance 2. Let $\Rightarrow$ denote weak convergence of probability measures.

Theorem 5.7. *Suppose Assumption IV holds and one of the following holds: (A) Assumption I, (B) Assumption II or (C) Assumption I*(i), (ii). Then in probability,*

(a) If $m_n \leq n/2$ *then* $F^{m_n^{-1/2}A} \Rightarrow N(0,2)$ *for* $A = C_n^{(s)b}, DH_n^b, PT_n^b$ *and* PH_n^b.

(b) If $m_n \leq n$ *then* $F^{m_n^{-1/2}R_n^{(s)b}} \Rightarrow \mathcal{L}_R$.

(c) If $m_n \leq 2n$ *then* $F^{m_n^{-1/2}H_n^{(s)b}} \Rightarrow H_\alpha^b$ *which is symmetric and* H_0^b *is the degenerate distribution at zero.*

(d) If $m_n \leq n$ *then* $F^{m_n^{-1/2}T_n^{(s)b}} \Rightarrow T_\alpha^b$ *which is symmetric and* ${T_0}^b = N(0,2)$.

If Assumption V holds, the convergence are a.s. in cases (A) and (B).

(ii) **Type II banding**. The Type II band version H_n^B of $H_n^{(s)}$ is defined with the input sequence $\{x_i I(|i-n| \leq m) + 0I(|i-n| > m)\}$. The Type II band versions R_n^B of $R_n^{(s)}$ and T_n^B of $T_n^{(s)}$ are defined with the input sequence $\{x_i I(\{i \leq m\} \cup \{i \geq n-m\}) + 0I(m < i < n-m)\}$. Type II banding does not yield any nontrivial situations for symmetric, doubly symmetric and palindromic matrices.

Theorem 5.8. *Suppose Assumption IV holds and any one of the following holds: (A) Assumption I, (B) Assumption II or (C) Assumption I*(i), (ii). Then in probability,*

(a) If $m_n \leq n/2$ *then* $F^{(2m_n)^{-1/2}R_n^B} \Rightarrow \mathcal{L}_R$.

(b) If $m_n \leq n$ *then* $F^{(2m_n)^{-1/2}H_n^B} \Rightarrow H_\alpha^B$ *which is symmetric and* $H_0^B = \mathcal{L}_R$.

(c) If $m_n \leq n/2$ *then* $F^{m_n^{-1/2}T_n^B} \Rightarrow T_\alpha^B$ *which is symmetric and* $T_0^B = N(0,2)$.

If Assumption V holds, the convergence are a.s. in cases (A) and (B).

5.6. Matrices with dependent entries.

Let $x_t = \epsilon_t \epsilon_{t+1} \cdots \epsilon_{t+d-1}$ where $\{\epsilon_i\}$ are i.i.d. To deal with this kind of dependence between $\{x_i\}$, we extend the concept of matching.

Matched circuits: Let L be a link function. Let $M^\pi = ((m_{i,j}))$ be the $d \times h$ matrix where $m_{i,j} = L(\pi(j-1), \pi(j)) + i - 1$. We say that π is d**-matched** if every element of M^π appears at least twice. This notion is extended to d-joint matching and d-cross matching in the obvious way. Note the following facts:

1. No two entries belonging to the same column of M^π can be equal.

2. If some entry in the j_1-th column of M^π is equal to some entry in its j_2-th column then $|L(\pi(j_1-1), \pi(j_1)) - L(\pi(j_2-1), \pi(j_2))| \leq d-1$.

Let $N^M_{h,3^+}$= Number of d-matched circuits of length h with at least one entry of M^π repeated at least thrice, and let $Q^M_{h,4}$= Number of circuits $(\pi_1, \pi_2, \pi_3, \pi_4)$ of length h which are jointly d-matched and jointly d-cross matched. The following lemma was proved in Bose and Sen (2008)[21].

Lemma 5.9. *Suppose (L, f) with $f(x) = x$ satisfies Property B.*
(a) There are constants $C_{h,d}$ and $K_{h,d}$ such that

$$N^M_{h,3^+} \leq C_{h,d} n^{\lfloor (h+1)/2 \rfloor} \text{ and } Q^M_{h,4} \leq K_{h,d} n^{2h+2}. \tag{24}$$

(b) Suppose $x_t = \epsilon_t \epsilon_{t+1} \cdots \epsilon_{t+d-1}$ where $\{\epsilon_i\}$ satisfies Assumption I. Let $A_{n,d} = ((x_{L(i,j)}))_{n\times n}$ where (L, f) satisfies Property B with $f(x) = x$. Then for every h

$$\mathrm{E}\left[\frac{1}{n}\operatorname{Tr}\left(\frac{A_{n,d}}{\sqrt{n}}\right)^h - \mathrm{E}\,\frac{1}{n}\operatorname{Tr}\left(\frac{A_{n,d}}{\sqrt{n}}\right)^h\right]^4 = \mathrm{O}(n^{-2}). \tag{25}$$

As a consequence (M4) holds too.

Lemma 5.10. *Each d-matched circuit π with only pair matchings is also pair-matched w.r.t. L and vice versa. Hence if $l(\pi) = h$ is odd, then no d-matched circuit π can be pair-matched.*

Detailed proof of the following theorem is given in Bose and Sen (2008)[21].

Theorem 5.11. *Let $x_t = \epsilon_t \epsilon_{t+1} \cdots \epsilon_{t+d-1}$ where $\{\epsilon_i\}$ satisfies Assumption I. Let $A_{n,d} = ((x_{L(i,j)}))_{n\times n}$ where (L, f) satisfies Property B with $f(x) = x$. If LSD of $\{F^{n^{-1/2}A_{n,d}}\}$ exists a.s. for $d = 1$, then the same LSD holds a.s. for $d \geq 2$.*

Sketch of proof of Theorem 5.11. Let $F_{n,d}$ denote the ESD of $n^{-1/2}A_{n,d}$. Lemma 5.9 and 5.10 imply that for every h, d,

$$\beta_h(F_{n,d}) - \mathrm{E}[\beta_h(F_{n,d})] \to 0 \text{ almost surely.}$$

On the other hand

$$\begin{aligned}\mathrm{E}[\beta_h(F_{n,d})] = \frac{1}{n}\mathrm{E}\big[\operatorname{Tr}(n^{-1/2}A_{n,d})^h\big] &= \frac{1}{n^{1+h/2}}\sum_\pi \mathrm{E}[X_\pi] \\ &= \frac{1}{n^{1+h/2}}\sum_{\pi\ d\text{-matched}} \mathrm{E}[X_\pi]\end{aligned}$$

where $X_\pi = \prod_{i=1}^h \epsilon_{L(\pi(i-1),\pi(i))}\epsilon_{L(\pi(i-1),\pi(i))+1}\cdots\epsilon_{L(\pi(i-1),\pi(i))+d-1}$.

Lemma 5.9(a) and Lemma 5.10 imply that if h is odd then $\lim \mathrm{E}[\beta_h(F_{n,d})] = 0$, and hence for every d, $\lim \beta_h(F_{n,d}) = 0$ a.s..

Now suppose $h = 2k$ is even. Let $\Pi(w)$ be as defined in Section 4 for ordinary L-matching. From Theorem 4.1, a.s.,

$$\lim n^{-(k+1)} \sum_{w:|w|=k} \Pi(w) = \lim \beta_{2k}(n^{-1/2}A_{n,1}) = \lim \mathrm{E}[\beta_{2k}(n^{-1/2}A_{n,1})].$$

On the other hand, Lemma 5.9 and Lemma 5.10 imply that for all d

$$\lim n^{-(k+1)} \sum_{w:|w|=k} \Pi(w) = \lim \mathrm{E}[\beta_{2k}(F_{n,d})] = \lim \beta_{2k}(F_{n,d}), \quad \text{almost surely.}$$

□

Here is another result in a dependent situation. For proof see Bose and Sen (2008)[21].

Theorem 5.12. *Let* $x_t = \sum_{j=0}^{\infty} a_j \epsilon_{t-j}$ *with* $\{a_j\}$ *satisfying* $\sum_j |a_j| < \infty$ *and* $\{\epsilon_i\}$ *satisfying Assumption I and* $\sum_j j a_j^2 < \infty$*. Then with the input sequence* $\{x_t\}$, $\{F^{n^{-1/2}T_n^{(s)}}\}$ *and* $\{F^{n^{-1/2}H_n^{(s)}}\}$ *converge weakly to nonrandom symmetric probability measures* T_a *and* H_a *respectively. These LSD do not depend on the distribution of* ϵ_1*. The* $(2k)$*-th moment of* T_a *and* H_a *are given by*

$$\beta_{2k}(T_a) = A_{2k}\beta_{2k}(\mathcal{L}_T) \text{ and } \beta_{2k}(H_a) = A_{2k}\beta_{2k}(\mathcal{L}_H)$$

where $\mathcal{L}_T$ *and* $\mathcal{L}_H$ *are as in Theorem 5.2 and* $A_{2k} = \sum_{d=0}^{\infty} \Big(\sum_{\substack{m_1,\ldots,m_k \geq 0:\\ \sum_1^k m_i = d}} \prod_{j=1}^{k} a_{m_j} \Big)^2$.

6. Moment Method Applied to Other Matrices

6.1. Mod $\lfloor np \rfloor$ link functions. Recall that the Hankel and reverse circulant link functions are respectively, $L(i,j) = i + j$ and $L(i,j) = i + j \bmod n$. Define a class of link functions $\{L_p : p \in (0,2]\}$, where $L_p(i,j) = i + j \bmod \lfloor np \rfloor$. Then the previous two link functions are special cases. Some results on the LSD with i.i.d. inputs and link function L_p have been established by Basak and Bose (2010)[13]. In particular, when $p = \frac{1}{m}$ for some integer m, the LSD is $(1-p)\delta_0 + p\sqrt{m}R$, where δ_0 is the degenerate distribution at 0 and R has distribution $\mathcal{L}_R$. Similar extensions were also obtained for mod $\lfloor np \rfloor$ versions of Toeplitz and symmetric circulant link functions.

6.2. Tridiagonal matrices. Let A_n be the tridiagonal random matrix

$$A_n = \begin{bmatrix} d_n & b_{n-1} & 0 & 0 & \dots & 0 & 0 \\ b_{n-1} & d_{n-1} & b_{n-2} & 0 & \dots & 0 & 0 \\ 0 & b_{n-2} & d_{n-2} & b_{n-3} & \dots & 0 & 0 \\ & & & \vdots & & & \\ 0 & 0 & 0 & \dots & b_2 & d_2 & b_1 \\ 0 & 0 & 0 & \dots & 0 & b_1 & d_1 \end{bmatrix}.$$

Popescu (2009)[41] used the trace formula and the moment method to obtain many interesting limit distributions. It is also assumed that, the sequence $\{d_j, b_j\}$ is independent, moments of d_n are bounded uniformly in n and $\mathrm{E}[(b_n/n^\alpha)^k] \to m_k$ for every k, as $n \to \infty$. Let $X_n = A_n/n^\alpha$ and tr denotes the normalized trace i.e., $tr(I) = 1$. He showed that $\mathrm{E}[tr(X_n^k)] \to L_k$ for some $\{L_k\}$ which can be expressed in terms of $\{m_i\}$. Moreover, if $X_{i,n}$ are several independent matrices, then the joint moments $\mathrm{E}[tr(X_{i_1,n}^{k_1} \cdots X_{i_s,n}^{k_s})]$ converge and the limits can be expressed in terms of the corresponding $\{m_r\}$. For example, if $\alpha = 1/2$ and $m_k = 1$ for all k, then the $L_k, k \geq 1$ are moments of the semicircle law. For other values of α, $\{L_k\}$ determines probability distributions whose densities are available in explicit forms.

6.3. Block Matrices. Oraby (2007)[37] discussed the a.s. limiting spectral distributions of some block random matrices. Under the strong assumption that the ESD of the blocks themselves converge a.s. to some limiting spectral distribution, an easy consequence from the theory of polynomials is the a.s. limiting behavior of the spectrum of the block matrix. The proof of the main theorem involves the method of moments.

Let B_k be a block matrix with Hermitian structure of order k (fixed) with blocks formed by independent Wigner matrices of size n. Oraby (2007)[38] showed that its LSD exists and depends only on the structure of the block matrix. When the block structure is circulant, the LSD is a weighted sum of two semicircle laws. In particular, the LSD of a Wigner matrix with k-weakly dependent entries need not be the semicircle law. Bannerjee (2010)[4] considered the case where B_k is symmetric and derived an explicit formula for the moments in terms of the link function L of B_k. In particular, only Catalan words contribute and the support of the LSD lies within $[-2\sqrt{\Delta(L,f)},\ 2\sqrt{\Delta(L,f)}]$ with $f(x) = x$.

7. Some Other Methods and Matrices

7.1. Normal approximation and the *k* circulant matrix. For the circulant matrix, apart from conjugacy, the eigenvalues are asymptotically normal and asymptotically independent. LSD proofs can be developed by appropriate usage of normal approximation methods. See for example, Bose and Mitra (2002)[18] (reverse circulant and symmetric circulant) and Meckes (2009)[35]. Recently Bose, Mitra and Sen (2008)[19] and Bose, Hazra and Saha (2009)[17] used normal approximation to establish LSD for some specific type of k-circulant matrices with independent and dependent inputs respectively.

7.2. Stieltjes transform and the Wigner and sample covariance matrices.

Stieltjes transform plays an important role in the study of spectral distribution. For any probability distribution G on the real line, its Stieltjes transform s_G is defined on $\{z: \ u+iv, v \neq 0\}$ as

$$s_G(z) = \int_{-\infty}^{\infty} \frac{1}{x-z} G(dx).$$

If A has real eigenvalues $\lambda_i, \ 1 \leq i \leq n$, then the Stieltjes transform of the ESD of A is

$$s_A(z) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i - z} = \frac{1}{n} \mathrm{Tr}[(A - zI)^{-1}].$$

Let $\{A_n\}$ be a sequence of random matrices with real eigenvalues and let the corresponding sequence of Stieltjes transform be $\{m_n\}$. If $m_n \to m$ in some suitable manner, where m is a Stieltjes transform, then the LSD of the sequence $\{A_n\}$ is the unique probability on the real line whose Stieltjes transform is the function m. It is shown that $\{m_n\}$ satisfies some (approximate) recursion equation. Solving the limiting form of this equation identifies the Stieltjes transform of the LSD. See Bai (1999)[6] for this approach in deriving the LSD for the Wigner and the S matrices and in studying the rate of convergence. Incidentally, no Stieltjes transform based proof is available for the Toeplitz and the Hankel matrices.

7.2.1. Wigner matrix with heavy tailed input. Consider the Wigner matrix $W_n^{(s)}$ with i.i.d. entries belonging to domain of attraction of an α-stable law with $\alpha \in (0, 2)$. Ben Arous and Guionnet (2006)[16] prove that with an appropriate slowly varying function $l(\cdot)$, $\{\mathrm{E}[F^{(l(n)n^{1/\alpha})^{-1} W_n^{(s)}}]\}$ converges to some law μ_α which depends only on α. This law is symmetric, heavy-tailed, and is absolutely continuous with respect to the Lebesgue measure, except possibly on a compact set of capacity zero. Some similar results for the S matrix and band matrices can be found in Belinschi, Dembo and Guionnet (2009)[15].

7.2.2. I.I.D. Matrix and the Circular Law. Let A_n be the $n \times n$ random matrix with mean 0 and variance 1 i.i.d. complex entries. Then $\{F^{n^{-1/2} A_n}\}$ converges a.s., to the uniform distribution on the unit disk (called the circular law). This was first established for Gaussian entries by Mehta (1967)[36]. Girko (1984)[23] suggested a method of proof for the general case. Bai (1997)[5] considered smooth densities and bounded sixth moment of the entries and showed the result to be true. Gotze and Tikhomirov (2007)[24] showed the result for subgaussian entries and the moment conditions were further relaxed by Pan and Zhou (2010)[39], Gotze and Tikhomirov (2007)[25] and Tao and Vu (2008)[44]. The result in its final form was derived by Tao, Vu and Krishnapur (2009)[45]. The moment method fails for this matrix as all the moments of the circular distribution are zero and they do not determine the distribution uniquely. The Stieltjes transform method was used to show that the ESD converges. The laws of the singular value distribution of $n^{-1/2}A_n - zI$ for complex z also played a crucial role in determining the convergence of the ESD.

8. Discussion

(i) We have seen that under the boundedness property of the link function, convergence of the moments is a necessary and sufficient condition for the LSD to exist. Moreover, subsequential limits exist. It is not known if suitable restrictions on the link function guarantees the existence of limits of moments.

(ii) Similarly, given a specific subclass of words, can an appropriate (bounded) link function be devised for which the LSD contribution comes only from these words?

(iii) Under what conditions on the link does the LSD have bounded or unbounded support? Bannerjee (2010)[4] has shown that if Property B is satisfied with $f(x) = x$ and $\alpha_n = o(n)$, then only the Catalan words contribute to the moments and the support of the LSD is a subset of $[-2\sqrt{\Delta(L,f)},\ 2\sqrt{\Delta(L,f)}]$.

(iv) We have used the moment method only for real symmetric matrices. Using the moment method for nonsymmetric matrices or for matrices with complex entries does not appear to be convenient. However, more thought on this is needed.

(v) The d-matching helped us to address linear dependence. One can also think of extending the results to input sequences which admit other types of dependence, for example for martingale differences.

(vi) Recall that for the S matrix, there is a positive mass equal to $1 - y^{-1}$ when $p/n \to y > 1$. It is evident from simulations that a similar phenomenon exists for general XX' matrices. See Bose, Sen and Gangopadhyay (2009)[20]. However, detailed information on the quantum of mass at zero and the gap between 0 and the next point in the support of the LSD is not known.

Acknowledgement

We are grateful to Sayan Bannerjee, Anirban Basak and Sanchayan Sen for helpful discussions.

References

[1] Anderson, Greg W. and Zeitouni, Ofer (2006). A CLT for a band matrix model. *Probability Theory and Related Fields*, 134, no. 2, 283–338.

[2] Arnold, L. (1967). On the asymptotic distribution of the eigenvalues of random matrices. *J. Math. Anal. Appl.*, 20, 262–268.

[3] Anderson, G. W.; Guionnet, A. and Zeitouni, O. (2009). *An Introduction to Random Matrices.* Cambridge University Press.

[4] Bannerjee, Sayan (2010). Large dimensional random matrices. *M. Stat. Mid Year Project Report*, Februaty 2010. Indian Statistical Institute, Kolkata.

[5] Bai, Z. D. (1997). Circular law. *Ann. Probab*, 25, 494–529.

[6] Bai, Z. D. (1999). Methodologies in spectral analysis of large dimensional random matrices, a review. *Statistica Sinica*, 9, 611–677 (with discussions).

[7] Bai, Z. D. and Zhou, Wang (2008). Large sample covariance matrices without independent structure in columns. *Statist. Sinica*, 18, 425–442.

[8] Bai, Z. D. and Silverstein, J. (2006). *Spectral Analysis of Large Dimensional Random Matrices.* Science Press, Beijing.

[9] Bai, Z. D. and Yin, Y. Q. (1988). Convergence to the semicircle law. *Ann. Probab.*, 16, no. 2, 863–875.

[10] Basak, Anirban (2009). Large dimensional random matrices. *M. Stat. Project Report*, June 2009. Indian Statistical Institute, Kolkata.

[11] Basak, Anirban and Bose, Arup (2009). Limiting spectral distribution of some band matrices. *Technical Report R16/2009, Stat-Math Unit, Indian Statistical Institute, Kolkata.* To appear in *Periodica Hungarica.*

[12] Basak, Anirban and Bose, Arup (2010). Balanced random Toeplitz and Hankel Matrices. *Technical Report R01/2010, Stat-Math Unit, Indian Statistical Institute, Kolkata.* To appear in *Elec. Comm. Probab.*.

[13] Basak, Anirban and Bose, Arup (2010). Limiting spectral distribution of a class of patterned matrices. *In preparation.*

[14] Basak, Anirban; Bose, Arup and Sen, Sanchayan (2010). Limiting spectral distribution of sample autocovariance matrices. *In preparation.*

[15] Belinschi, Serban; Dembo, Amir and Guionnet, Alice (2009). Spectral measure of heavy tailed band and covariance random matrices. *Comm. Math. Phys.*, 289, no. 3, 1023–1055.

[16] Ben Arous, Gérard and Guionnet, Alice (2008). The spectrum of heavy tailed random matrices. *Comm. Math. Phys.*, 278, no. 3, 715–751.

[17] Bose, Arup; Hazra, Rajat Subhra and Saha, Koushik (2009). Limiting spectral distribution of circulant type matrices with dependent inputs. *Electron. J. Probab.*, 14, no. 86, 2463–2491.

[18] Bose, Arup and Mitra, Joydip (2002). Limiting spectral distribution of a special circulant. *Stat. Probab. Letters*, 60, 1, 111–120.

[19] Bose, Arup; Mitra, Joydip and Sen, Arnab (2008). Large dimensional random k-circulants. *Technical Report No.R10/2008, Stat-Math Unit, Indian Statistical Institute, Kolkata.*

[20] Bose, Arup; Gangopadhyay, Sreela and Sen, Arnab (2009). Limiting spectral distribution of XX' matrices. To appear in *Ann. Inst. Henri Poincare Probab. Stat.*

[21] Bose, Arup and Sen, Arnab (2008). Another look at the moment method for large dimensional random matrices. *Elec. J. Probab.*, 13, 588–628.

[22] Bryc, W.; Dembo, A. and Jiang, T. (2006). Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Ann. Probab.*, 34, no. 1, 1–38.

[23] V. L. Girko. (1984). Circular law. *Theory Probab. Appl.*, 4, 694–706.

[24] Gotze, F. and Tikhomirov, A.N. (2007). On the circular law. *arXiv:math/0702386v1 [math.PR].*

[25] Gotze, F. and Tikhomirov, A.N. (2007). The circular law for random matrices. *arXiv:0709.3995v3 [math.PR].*

[26] Grenander, U. (1963). *Probabilities on Algebraic Structures.* John Wiley & Sons, Inc., New York-London; Almqvist & Wiksell, Stockholm-Göteborg-Uppsala.

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

[28] Hammond, C. and Miller, S. J. (2005). Distribution of eigenvalues for the ensemble of real symmetric Toeplitz matrices. *J. Theoret. Probab.* 18, no. 3, 537–566.

[29] Jackson, S., Miller, S. J. and Pham, T. (2010) Distribution of eigenvalues of highly palindromic Toeplitz matrices. *arXiv:1003.2010[math.PR]*

[30] Jonsson, D. (1982). Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.* 12, no. 1, 1–38.

[31] Kargin, Vladislav (2009). Spectrum of random Toeplitz matrices with band structures. *Elect. Comm. in Probab.* 14 (2009), 412–421.

[32] Liu, Dang-Zheng and Wang, Zheng-Dong (2009). Limit Distributions for Random Hankel, Toeplitz Matrices and Independent Products. *arXiv:0904.2958v2 [math.PR].*

[33] Marčenko, V. A. and Pastur, L. A. (1967). Distribution of eigenvalues in certain sets of random matrices, (Russian) *Mat. Sb. (N.S.)* 72 (114), 507–536.

[34] Massey, A.; Miller, S. J. and and Sinsheimer, J. (2007). Distribution of eigenvalues of real symmetric palindromic Toeplitz matrices and circulant matrices. *J. Theoret. Probab.* 20, 3, 637–662.

[35] Meckes, Mark W. (2009). Some results on random circulant matrices. *arXiv:0902.2472v1 [math.PR].*

[36] Mehta, M. L. (1967). *Random matrices and the statistical theory of energy levels.* Academic Press.

[37] Oraby, Tamer (2007). The limiting spectra of Girko's block-matrix *J. Theoret. Probab.* 4, 959–970.

[38] Oraby, Tamer (2007). The spectral laws of Hermitian block-matrices with large random blocks. *Electron. Comm. Probab.* 12, 465–476.

[39] Pan, G. and Zhou, W. (2010). Circular law, extreme singular values and potential theory. *J. Multivariate Anal.* 101, no. 3, 645–656.

[40] Pastur, L. (1972). The spectrum of random matrices. (Russian) *Teoret. Mat. Fiz.* 10 no. 1, 102–112.

[41] Popescu, Ionel (2009). General tridiagonal random matrix models, limiting distributions and fluctuations. *Probab. Theory Relat. Fields*, 144, 179–220.

[42] Sen, Arnab (2006). Large dimensional random matrices. *M. Stat. Project Report*, May 2006. Indian Statistical Institute, Kolkata.

[43] Sen, Sanchayan (2010). Large dimensional random matrices. *M. Stat. Mid Year Project Report*, Februaty 2010. Indian Statistical Institute, Kolkata.

[44] Tao, T. and Vu, V. (2008). Random Matrices: The circular Law, *Communications in Contemporary Mathematics*, 10, 261–307.

[45] Tao, T.; Vu, V. and Krishnapur M. (Appendix) (2009). Random matrices: universality of the ESDs and the circular law. To appear in *Ann. Probab.*.

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

[47] Wigner, E. P. (1958). On the distribution of the roots of certain symmetric matrices. *Ann. of Math.*, (2), 67, 325–327.

[48] Yin, Y. Q. (1986). Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.*, 20, no. 1, 50–68.

[49] Yin, Y. Q. and Krishnaiah, P. R. (1985). Limit theorem for the eigenvalues of the sample covariance matrix when the underlying distribution is isotropic. *Theory Probab. Appl.*, 30, no. 4, 810–816.