Another look at the moment method and some new results

Arup Bose
Indian Statistical Institute, Kolkata,
abose@isical.ac.in

Arnab Sen
Indian Statistical Institute, Kolkata
• Some interesting matrices

• Empirical and Limiting Spectral Distributions

• The counting method

• The "volume" method of Bryc, Dembo and Jiang

• Some New Results
Some interesting random matrices

Sample variance covariance matrix Suppose \( \{x_{jk}, j, k = 1, 2, \ldots \} \) is a double array of i.i.d. complex random variables with mean zero and variance 1. Write \( x_k = (x_{1k}, \ldots, x_{pk})' \) and let \( X_n = [x_1, x_2, \ldots, x_n] \).

\[
S_n = n^{-1} X_n X_n^* 
\]

(in short an \( S \) matrix).

Wigner matrix Entries above the diagonal are independent complex random variables with zero mean and variance \( \sigma^2 \), and whose diagonal elements are i.i.d. real random variables. Scale by dividing by \( n^{1/2} \).
Let \( \{x_0, x_1, \ldots \} \) be a sequence of i.i.d. real random variables with mean zero and variance \( \sigma^2 \).

\[
\begin{pmatrix}
x_0 & x_1 & x_2 & \cdots & x_{n-2} & x_{n-1} \\
x_1 & x_0 & x_1 & \cdots & x_{n-3} & x_{n-2} \\
x_2 & x_1 & x_0 & \cdots & x_{n-4} & x_{n-3} \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
x_{n-1} & x_{n-2} & x_{n-3} & \cdots & x_1 & x_0
\end{pmatrix}
\]

The \((i, j)\)th entry is \( x_{|i-j|} \).

These give rise to the famous Toeplitz operators in mathematics.

The matrix of sample autocovariances is of this form, but there the entries are dependent.
Hankel Matrix

Let \( \{x_0, x_1, \ldots\} \) be a sequence of i.i.d. real random variables with mean zero and variance \( \sigma^2 \).

\[
n^{-1/2} H_n = \frac{1}{\sqrt{n}} \begin{bmatrix}
x_0 & x_1 & x_2 & \ldots & x_{n-2} & x_{n-1} \\
x_1 & x_2 & x_3 & \ldots & x_{n-1} & x_n \\
x_2 & x_3 & x_4 & \ldots & x_n & x_{n+1} \\
& & & \ddots & \vdots \\
x_{n-1} & x_n & x_{n+1} & \ldots & x_{2n-3} & x_{2n-2}
\end{bmatrix}.
\]

\((i, j)\)th entry is \( x_{i+j-2} \).

They are very closely related to the Toeplitz matrices.
Let \( \{x_i, \ i = 0, 1, \ldots\} \) be a sequence of independent and identically distributed (iid) random variables with mean zero and variance 1.

\[
X_n = \frac{1}{\sqrt{n}} \begin{bmatrix}
    x_0 & x_1 & x_2 & \cdots & x_{n-2} & x_{n-1} \\
    x_1 & x_2 & x_3 & \cdots & x_{n-1} & x_0 \\
    x_2 & x_3 & x_4 & \cdots & x_0 & x_1 \\
    & \ddots & \ddots & \ddots & \ddots & \ddots \\
    x_{n-1} & x_0 & x_1 & \cdots & x_{n-3} & x_{n-2}
\end{bmatrix}.
\]

\((i, j)\)th entry is \(x_{(i+j-2)} \mod n\).

Similar to Hankel, except mod n.
Symmetric circulant

The symmetric version of the usual circulant matrix may be defined as

\[
SC_n = \frac{1}{\sqrt{n}} \begin{bmatrix}
  x_0 & x_1 & x_2 & \ldots & x_2 & x_1 \\
  x_1 & x_0 & x_1 & \ldots & x_3 & x_2 \\
  x_2 & x_1 & x_0 & \ldots & x_2 & x_3 \\
  \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
  x_1 & x_2 & x_3 & \ldots & x_1 & x_0
\end{bmatrix}.
\]

So, the \((i, j)\)th element of the matrix is given by

\[
x(n/2 - |n/2 - |i - j||)。
\]

Even if \(n\) is odd, \((n/2 - |n/2 - |i - j||)\) is always an integer.
Empirical Spectral Distribution:

If $\lambda_1, \lambda_2, \ldots, \lambda_n$ are all the eigenvalues of $A_n$ then

$$F_n(x, y) = n^{-1} \sum_{i=1}^{n} I\{\Re \lambda_i \leq x, \ \Im \lambda_i \leq y\}.$$ 

Limiting Spectral Distribution (LSD):

The weak limit of Empirical Spectral Distribution (almost sure/in probability).

We shall deal only with symmetric matrices and hence we have only real eigenvalues.
The $r$-th moment of the ESD of an $n \times n$ matrix $A$, with characteristic roots $\lambda_1, \lambda_2, \ldots, \lambda_n$ has the following nice form:

$$\frac{1}{n} \sum_{i=1}^{n} \lambda_i^r = \frac{1}{n} \text{tr}(A^r) = \beta_r(A) \text{ (say)}. $$

Now, suppose $\{A_n\}$ is a sequence of random matrices such that

$$\beta_r(A_n) \longrightarrow \beta_r. $$

and $\{\beta_r\}$ (nonrandom), uniquely characterises a distribution. Then LSD of the sequence $\{A_n\}$ exists.

A sufficient condition for the uniqueness above is:

$$\sum_{r=1}^{\infty} \beta_2^{-1/2r} = \infty.$$
Combinatorial Problem

- We are tacitly assuming that the LSD has all moments finite.

- Involves computing the trace of $A_n^h$ or at least its leading term.

$$\text{tr}(A^r) = \sum_{1 \leq i_1, i_2, \ldots, i_r \leq n} a_{i_1} a_{i_2} i_3 \cdots a_{i_{r-1}} i_r a_{i_r} i_1.$$
Show that $E(\beta_r(A_n)) \longrightarrow \beta_r$ and $V(\beta_r(A_n)) \longrightarrow 0$.

- No guarantee that $E(\beta_r(A_n))$ are finite. The elements of $A_n$ are first appropriately truncated.....We shall assume that we are dealing with uniformly bounded random variables.

- If the entries are dependent, then there is no obvious way of carrying out this programme in general.

- Weak Convergence in probability. If $V(\beta_r(A_n)) = O(n^{-2})$, then the convergence is a.s.

In each case the proof involves complicated combinatorics. The Toeplitz and Hankel are particularly difficult–Bryc, Dembo and Jiang.
Limiting Spectral Distributions

**Wigner matrix: the Semi-Circular Law**

Wigner, Grenander, Bai and others....

**Theorem** The LSD of $n^{-1/2}W_n$ has the density function

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

Also known as *quarter circle law* (Girko and Repin (1995)).

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

$$\beta_{2k} = \int s^{2k} p_\sigma(s) ds = \frac{(2k)!\sigma^{2k}}{k!(k + 1)!}$$
Grenander, Silverstein, Wachter, Jonsson, Yin, Krishnaiah, Bai

Indexed by $\sigma^2$ and $0 < y < \infty$. It has a positive mass point mass $1 - \frac{1}{y}$ at 0 if $y > 1$. Elsewhere it has a density:

$$p_y(x) = \begin{cases} \frac{1}{2\pi xy\sigma^2} \sqrt{(b-x)(x-a)} & \text{if } a \leq x \leq b, \\ 0 & \text{otherwise.} \end{cases}$$

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

**Theorem** Suppose $\{x_{ij}\}$ are i.i.d. real variables with variance $\sigma^2$.

(i) If $p/n \to y \in (0, \infty)$ then the ESD of $S_n$ converges almost surely to the Marčenko-Pastur law with scale index $\sigma^2$.

(ii) If $p/n \to 0$ then the ESD of $W_n = \sqrt{n/p}(S_n - \sigma^2 I_p)$ converges almost surely to the semicircular law with scale index $\sigma^2$. 

S matrix: Marčenko-Pastur Law
Suppose $y = 1$.

If $W$ follows Wigner's law, then $S = W^2$ follows M-P law.

So the $k$th moment of M-P law is given by:

$$\beta_{2k} = \frac{(2k)!\sigma^{2k}}{(k + 1)!k!}$$

But why?
**Theorem** Bose and Mitra (2002): Let \( \{x_i\} \) be i.i.d. with mean zero and variance 1 and \( E|x_1|^3 < \infty \). Then the LSD (in probability) of \( X_n = n^{-1/2} A_n \) is given by:

\[
f(x) = |x| \exp(-x^2), \quad -\infty < x < \infty.
\]

True with only second moment assumption.

2\(k\)th moment is given by

\[
\beta_{2k} = k!
\]
Theorem  Bryc, Dembo and Jiang (2005): With probability one, the ESD of $\frac{1}{\sqrt{n}} T_n$ converges weakly as $n \to \infty$ to a non-random symmetric probability measure which does not depend on the distribution of the entries of $T_n$ and has unbounded support.

Not much is known about the limiting distribution.
Theorem Bryc, Dembo and Jiang (2003): Let the entries of the Hankel matrix $H_n$ be i.i.d. real-valued random variables with mean zero and variance one. With probability one, the ESD of $\frac{1}{\sqrt{n}}H_n$ converges weakly, as $n \to \infty$, to a non-random symmetric probability measure. The limiting distribution does not depend on the distribution of the entries of $H_n$, has unbounded support and is not unimodal.

Not much more is known about the limiting distribution.
Moment method: Another look

We shall assume that the underlying random variables are bounded and iid.

**Key Step 1:** \( \lim_{n \to \infty} \frac{1}{n} E \text{tr}(A_n^r) = \beta_r \)

Except for sample varcovar, all odd moments are zero.

**Key Step 2:** \( E \left[ \frac{1}{n} \text{tr}(A_n^r) - \frac{1}{n} E \text{tr}(A_n^r) \right]^4 = O \left( \frac{1}{n^2} \right) \).

Shall discuss only key step 1.

Recall:

\[
\text{tr}(A^r) = \sum_{1 \leq i_1, i_2, \ldots, i_r \leq n} a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_{r-1} i_r} a_{i_r i_1}.
\]

Idea borrowed from Bryc, Dembo and Jiang (Toeplitz Matrix, Hankel matrix and the Markov matrix)—relate counting to volume calculations.
Common structure: the $s$-function

Entries taken from one sequence:

**Toeplitz:** $X_{|i-j|}$. Define $s : \mathbb{N} \rightarrow \mathbb{N}$ by $s(i, j) = |i - j|$

**Hankel:** $X_{i+j-2}$ Define $s : \mathbb{N} \rightarrow \mathbb{N}$ by $s(i, j) = i + j - 2$

**Reverse circulant:** $X_{i+j-2 \mod n}$

Define $s : \mathbb{N} \rightarrow \mathbb{N}$ by $s(i, j) = i + j - 2 \mod n$

**Property (P):** Fix $p \in \mathbb{N}$. Then $|\{(k, l) : s(k, l) = p\}|$ is finite. Same if we fix one coordinate. This property is crucial in Key step 2.
Circuit: the $\pi$ function

CIRCUIT: $\pi : \{0, 1, 2, \cdots, r\} \rightarrow \{1, 2, \cdots, n\}$ such that $\pi(0) = \pi(2k)$. The length $L(\pi)$ of the circuit is defined to be $r$.

For the sample var-covar the range of $\pi(i)$ depends on whether $i$ is even or odd (because $p$ and $n$ need not be equal).

A circuit $\pi$ is said to be $s$-matched (or in short, matched) if given $i \exists j \neq i$ s.t.

$$s(\pi(i - 1), \pi(i)) = s(\pi(j - 1), \pi(j)).$$

Obviously, non $s$ matched circuits contribute zero amounts.

$\pi$ is said to have an Edge of order $h$ if some $s$ value is repeated $h$ times.

Later we show that circuits with only order 2 edges matter.
$s$ function over an array: Wigner

Entries taken from one array:

Wigner: $X_{\min(i,j),\max(i,j)}$

Define $s(i,j) = (\min(i,j), \max(i,j))$.

Property (P) holds (for a given pair of values).

$s$ matched circuit defined in the natural way.

Again, among $s$ matched circuits only the paired ones contribute.
Entries taken from two arrays:

Sample variance covariance matrix: Note that all moments are nonnegative. Here $\beta_k(S_n)$ equals:

$$p^{-1}n^{-k} \sum_{\pi} x_{s1}(\pi(0), \pi(1))x_{s2}(\pi(1), \pi(2)) x_{s1}(\pi(2), \pi(3))x_{s2}(\pi(3), \pi(4)) \cdots x_{s2}(\pi(2k-1), \pi(2k))$$

$s_1, s_2 : I^+ \times I^+ \rightarrow I^+ \times I^+$, $s_1(i, j) = (i, j)$, $s_2(i, j) = (j, i)$

Property (P) holds for both $s_1$ and $s_2$.

Now any circuit $\pi$ has the non-uniform range

$1 \leq \pi(2m) \leq p$, $1 \leq \pi(2m + 1) \leq n$

A circuit is said to be $(s_1, s_2)$ matched if it is matched within the same $s$ or across. Again, among matched circuits only the paired ones contribute.
Equivalence relation over $\pi$.

$\pi_1$ is equivalent to $\pi_2$ iff $\pi_1(i) = \pi_1(j) \iff \pi_2(i) = \pi_2(j)$.

Each equivalence class is uniquely indexed by a partition of $\{1, 2, \cdots, r\}$.

Consider any partition of $\{1, 2, \cdots, r\}$. Any such partition can be denoted by a word (of length $r$) of letters such that the first occurrence of each letter in the word is in alphabetical order.

For example let $r = 5$. Then the partition $\{\{1, 3, 5\}, \{2, 4\}\}$ can be represented by the word $ababa$.

The number of partition blocks corresponding to a word $w$ will be denoted by $|w|$. If $w = ababa$ then $|w| = 2$. 
A word $w$ has an edge of order $h$, if some letter of $w$ occurs exactly $h$ times.

$abccbaa$ has edges of order only 2 and 3.

A word is matched if all edges are of length two or more.

A word $w$ is non-matched if it has at least one edge of order 1.

Let $w[i]$ denote the $i$-th entry of word $w$. Define

$$
\Pi(w) = \{ \pi : \pi \text{ belongs to partition block indexed by } w \}
$$

i.e. $w[i] = w[j] \iff s(\pi(i - 1), s(\pi(i)) = s(\pi(j - 1), s(\pi(j)))$.
Association between circuits and words

A circuit $\pi$ is matched iff $\pi \in \Pi(w)$ where $w$ is matched.

A circuit $\pi$ has an edge of order $h$ iff $\pi \in \Pi(w)$, where $w$ has an edge of order $h$.

If $\pi \in \Pi(w)$, then number of distinct s-values for a circuit $\pi$ i.e. $|\{s(\pi(i-1), \pi(i)) : 1 \leq i \leq r\}| = |w|$. 
Let $\pi \in \Pi(w)$. For any circuit $\pi$, a vertex $\pi(i)$ is generating or independent if $w[i]$ corresponds to the first occurrence of a letter. We always take $\pi(0)$ as a generating vertex.

If $w = abbcab$ then $\pi(0), \pi(1), \pi(2), \pi(4)$ are generating vertices.

If we fix the generating vertices, other vertices get automatically fixed (upto a finitely many choices): thus a circuit is essentially determined by its generating vertices.

Let $\pi \in \Pi(w)$. Obviously, number of generating vertices in $\pi$ is $|w| + 1$. Hence

$$|\Pi(w)| \leq n|w| + 1$$
General reduction

Consider the sum \( \sum_{w} \sum_{\pi \in \Pi(w)} n^{-r} p^{-1} E(\cdots) \).

**Proposition** Let \( \lfloor x \rfloor \) denote the integral part of \( x \). Fix \( r \in \mathbb{N} \). Let \( N_r \) be the number of \( s \)-matched circuits of length \( r \) with at least one edge of order \( \geq 3 \). Then there is a constant \( C_r \) such that \( N_r \leq C_r n^{\lfloor (r+1)/2 \rfloor} \). In particular, \( n^{-(1+r/2)} N_r \to 0 \).

**Proof:** It is enough to show \( |\Pi(w)| \leq C_r n^{\lfloor (r+1)/2 \rfloor} \) for each such word because the number of such words is finite.

Fix such a word \( w \). Either \( r = 2k \) or \( r = 2k - 1 \). In either case \( |w| \leq k - 1 \). Hence, \( |\Pi(w)| \leq n^k = C_r n^{\lfloor (r+1)/2 \rfloor} \).

Thus only words with 2-edges contribute and each summand in the contributing sum is one and we are left with

\[
\sum_{w : w \text{ has only 2-edges}} [n^{-r} p^{-1}] |\Pi_w|.
\]
Number of such words is finite. BDJ showed that for each \( w \), for Toeplitz, Hankel and Markov matrices, the limit exists.

In general, for each word, we need to count the number of contributing circuits\( |\Pi(w)| \). Because of the Proposition, it is enough to count \( |\Pi^*(w)| \) where

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

Though the total number of word equals \( \frac{(2k)!}{2^k k!} \), only certain words contribute in the limit, depending on the type of matrix under consideration.

This count is obtained by fixing the independent vertices and then finding the number of choices of the dependent vertices. Because of the structure of the matrices, often the relations are "linear" or nearly so.
Symmetric word

Useful in Hankel, Reverse Circulant, and Symmetric Circulant matrices.

A word of length $2k$ is symmetric if each letter occurs exactly once in an odd and exactly once in an even position.

If $k = 3$ then $abccba, abcabc, abbacc$ are symmetric words.

$abcbca, ababcc$ are not symmetric words.

Symmetry automatically implies $\pi(0) = \pi(2k)$.

Number of symmetric words is $k!$. 
Details for Reverse circulant

Claim:

\[
\sum_{w: |w| = k} \frac{1}{n^{k+1}} |\Pi^*(w)| \to k!
\]

1. Only symmetric words contribute to the sum.

2. For any such word if we fix the independent vertices each of the remaining vertices has exactly one choice.

- Number of ways to choose the \((k + 1)\) independent vertices equals \(n^{k+1}\). So each symmetric word contributes 1.

- There are \(k!\) symmetric words.

Hence the limit is \(k!\).
Proof of 1

Let \( t_i = \pi(i - 1) - \pi(i) \).

1. \( \pi(0) - \pi(2k) = t_1 - t_2 + t_3 - t_4 + \ldots + t_{2k-1} - t_{2k} = 0 \). This must follow from the \( k \) independent restrictions on the word.

2. Consider a typical restriction, \( w[i] = w[j] \). This implies \( t_i - t_j = 0 \mod n \).

These imply that indices \( i \) and \( j \) occupy exactly one odd and one even position.

Thus the contributing words are symmetric.
Proof of 2

Suppose a typical restriction due to word looks like

\[ s_n(\pi(i - 1), \pi(i)) = s_n(\pi(j - 1), \pi(j)) \] where \( \pi(i - 1), \pi(i) \) and \( \pi(j - 1) \) are already fixed. The equation can be rewritten as follows

\[
\frac{1}{n} s_n(\pi(i - 1), \pi(i)) = \left[ \frac{1}{n}(\pi(j - 1) - 1) + \frac{1}{n}(\pi(j) - 1) \right] \mod 1.
\]

or, \[ s = x_{j-1} + x_j \mod 1 \]

where

\[ s, x_{j-1} \text{ and } x_j \in \{0, 1/n, 2/n, \ldots, 1 - 1/n\}. \]

So, \( x_j \) can be determined uniquely from the above equation. This in turn determines \( \pi(j) \) such that \( 1 \leq \pi(j) \leq n \). And we proceed inductively from left to right to get the whole circuit (i.e. \( \pi(0) = \pi(2k) \)) from only the independent vertices, uniquely.
Symmetric word in Hankel

For Hankel matrices, again the words producing nonzero contribution in the limit are only symmetric words by the same reasoning.

So, immediately we get

$$\limsup_{n} \beta_{2k}(n^{-1/2}H_n) \leq k!.$$ 

However, not every symmetric word contributes 1. So, it is not an equality.
Catalan words

Useful for Wigner and Sample variance covariance:

A word of length $2k$ is called a Catalan word if

(I) there is at least one pair of letters of the form $xx$.

(II) if we delete this pair, the remaining word is either empty or is of the type (I) and we can repeat to ultimately get an empty word.

Thus $abba$, $aabbcc$, $abccbdca$ are Catalan words.

$abab$, $abccab$, $abcddcab$ are not Catalan words.

Note that all Catalan word are also Symmetric words.
There is an obvious bijection between Catalan words of length \(2k\) and the set of sequences \((u_1, u_2, \cdots, u_{2k})\) such that

(I) \(u_i = +1\) or \(-1\)

(II) \(\sum_{i=1}^{2k} u_i = 0\) i.e. there are exactly \(k + 1\) and \(k - 1\) in the sequence

(III) \(\sum_{i=1}^{l} u_i \geq 0 \ \forall \ l \geq 1\)

\(i\)-th +1 in the sequence denotes the position of first occurrence of the \(i\)-th new letter (from left to right) in the Catalan word, and each \(-1\) denotes the position of the second occurrence of a letter in the Catalan word. For example the sequences

\((1, 1, -1, -1), (1, -1, 1, -1, 1, -1), (1, 1, 1, -1, -1, 1, -1, -1)\)

represent the words \(abba, aabbcc, abcdbda\)

The number of Catalan words of length \(2k\) is (use reflection principle) \(\frac{(2k)!}{(k+1)!k!}\).
\[
\lim_{n \to \infty} \sum_{w: |w|=k} n^{-k} p^{-1} |\Pi^*(w)|.
\]

Assume \( p/n \to y \leq 1 \).

Recall that

\[
\beta_k(S_n) = p^{-1} n^{-k} \sum x s_1(\pi(0),\pi(1)) x s_2(\pi(1),\pi(2)) x s_1(\pi(2),\pi(3)) x s_2(\pi(3),\pi(4)) \cdots x s_2(\pi(2k-1),\pi(2k))
\]

\[
= p^{-1} n^{-k} \sum x(\pi(0),\pi(1)) x(\pi(2),\pi(1)) x(\pi(2),\pi(3)) x(\pi(4),\pi(3)) \cdots x s_2(\pi(2k),\pi(2k-1))
\]

We need exactly \((k + 1)\) free vertices (hence \(k\) dependent vertices).
Only Catalan words contribute

Consider the set of vertices \( \{\pi(0), \pi(1), \ldots, \pi(2k - 1), \pi(2k)\} \).

Note that there are \((2k + 1)\) vertices, and \((k + 1)\) free vertices. Also \(\pi(0) = \pi(2k)\).

Hence by pigeon hole principle, \(\exists m \neq 0, 2k\) which is a singleton, That is, \(\pi(m) \neq \pi(j)\) for all \(j\).

This implies that

\[
(\pi(m - 1), \pi(m)) = (\pi(m + 1), \pi(m))
\]

OR

\[
(\pi(m), \pi(m - 1)) = (\pi(m), \pi(m + 1))
\]

This implies that there is a pair of letters of the form \(xx\).
Now consider the reduced set of \((2k - 1)\) vertices:

\[
\{\pi(0), \pi(1), \ldots, \pi(m - 1), \pi(m + 2), \ldots, \pi(2k)\}
\]

This generates a word of length \((2k - 2)\) obtained by deleting \(xx\).

The even vertices are still between 1 and \(p\) and the odd vertices are still between 1 and \(n\). Thus induction works....

Thus the word must be a Catalan word.

On the other hand, a Catalan word always satisfies relation \(\pi(0) = \pi(2k)\).

So we can concentrate our attention to Catalan words only.
We partition the number of Catalan words depending on how many free coordinate it generates with range $p$ and how many with range $n$.

Suppose it generate $(t + 1)$ free vertices with range $p$ or equivalently $(k - t)$ free vertices (or $t$ dependent vertices) with range $n$. $0 \leq t \leq k - 1$.

For a particular Catalan word with $(t + 1)$ free coordinate with range $p$,

$$\lim_{n \to \infty} n^{-k} p^{-1} |\Pi(w)| = \lim n^{-k} p^{-1} p^{t+1} n^{k-t} = y^t.$$
Contribution of Catalan words

Now note that a $+1$ in the even position produces a dependent coordinate with range $n$. So to get $(t + 1)$ free vertices with range $p$, we should look for the those sequences having exactly $t + 1$ in its even positions. Let $M_{t,k} \overset{\text{def}}{=} \text{Number of Catalan words with } (t + 1) \text{ free coordinate with range } p = \text{sequences of length } 2k \text{ satisfying (I)-(III) and also } t \text{ many } +1 \text{ in the even positions. Then}$

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

The proof can be done by a simple application of the reflection principle.

Thus

$$\lim_{n \to \infty} E/\beta_k(S_n) = \sum_{t=0}^{k-1} \frac{1}{t+1} \binom{k}{t} \binom{k-1}{t} y^t.$$
Proof based on reflection principle

Each sequence represents a simple random walk starting at origin \((0, 0)\) and ending at \((2k, 0)\) such that there are (i) exactly \(t\) upward moves in \(k\) of the even steps and (ii) no walk goes below the x-axis. This implies that \(u_1 = +1\) and \(u_{2k} = -1\). So consider the simple random walk starting at \((1,1)\) and ending at \((2k-1,1)\) satisfying condition (i) and (ii). First we count the number of paths without considering the constraint (ii). There are \(\binom{k-1}{t}\) many ways of choosing \(t + 1\)'s from \((k - 1)\) even steps and for each of them there are another \(\binom{k-1}{t}\) many ways of choosing \(t - 1\)'s from \((k - 1)\) odd steps (look at table 1). So, total number of such choices is \(\binom{k-1}{t}^2\).

<table>
<thead>
<tr>
<th></th>
<th>even</th>
<th>odd</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>+1</td>
<td>(t)</td>
<td>(k - 1 - t)</td>
<td>(k - 1)</td>
</tr>
<tr>
<td>-1</td>
<td>(k - 1 - t)</td>
<td>(t)</td>
<td>(k - 1)</td>
</tr>
<tr>
<td>total</td>
<td>(k - 1)</td>
<td>(k - 1)</td>
<td>(2(k - 1))</td>
</tr>
</tbody>
</table>
Number of paths which violate (ii)

Such a path touches \( y = -1 \) line at least once and hence has two consecutive upward movements. Consider the last time this occurs so that \( u_l = +1 \) and \( u_{l+1} = +1 \) in the sequence corresponding to above incidence. We consider a transformation

\[
(u_2, u_3, \cdots, u_l = +1, u_{l+1} = +1, \cdots, u_{2k-1}) \mapsto (u_2, u_3, \cdots, u_{l-1}, -1, -1, u_{l+2}, \cdots, u_{2k-1})
\]

(converting two suitably chosen +1s into two −1s leaving the rest of the sequence unaltered).

The resulting sequence is a path from \((1, 1)\) to \((2k − 1, −3)\). This is a bijection from the set all of paths from \((1, 1)\) to \((2k − 1, 1)\) satisfying (i) but violating (ii) to set of all paths from \((1, 1)\) to \((2k − 1, −3)\) having \((t − 1) + 1\) in the even steps (see table 2). Number of such paths is \((k−1−1)(k−1−1+t+1)\).

<table>
<thead>
<tr>
<th>table 2</th>
<th>even</th>
<th>odd</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>+1</td>
<td>(t−1)</td>
<td>(k−2−t)</td>
<td>(k−3)</td>
</tr>
<tr>
<td>−1</td>
<td>(k−t)</td>
<td>(t+1)</td>
<td>(k+1)</td>
</tr>
<tr>
<td>total</td>
<td>(k−1)</td>
<td>(k−1)</td>
<td>(2(k−1))</td>
</tr>
</tbody>
</table>
Catalan words and Wigner

By the SAME argument as in the $S$ matrix, only Catalan words matter and each contributes 1. So

$$\beta_{2k} = \frac{(2k)!}{k!(k+1)!}. $$

Odd moments are zero.

This also shows why the $k$th moment of $S$ matrix with $y = 1$ is the same as the $2k$th moment of Wigner.
Catalan words and Toeplitz

For Toeplitz matrices, every word of length $2k$ with exactly $k$ matches automatically implies $\pi(0) = \pi(2k)$.

- All words contribute

- Each Catalan word contributes 1.

So

$$\beta_{2k} \geq \frac{(2k)!}{k!(k+1)!}.$$

Odd moments are zero.
Catalan words and Hankel

Each Catalan word contributes 1.

So

\[ \beta_{2k} \geq \frac{(2k)!}{k!(k + 1)!}. \]

Odd moments are zero.
For Symmetric words, $p_T(w) = p_H(w)$. 

On the other hand, for words which are not symmetric, $p_H(w) = 0$ whereas $p_T(w) \geq 0$. 

Therefore,

$$\beta_{2k}(T) \geq \beta_{2k}(H).$$
In this case, ALL words contribute and contribute one each.

\[ \beta_{2k} = \frac{(2k)!}{2^k k!}. \]

Odd moments are zero.

So the LSD is the standard normal distribution.
The idea can be extended to some situations with dependent entries.

**Theorem:** Suppose \( X_t = \sum_{j=0}^{\infty} a_j \epsilon_{t-j} \) where \( \{\epsilon_t\} \) are i.i.d. with \( E(\epsilon_1) = 0 \), \( E(\epsilon_1^2) = 1 \), \( E(\epsilon_1^4) < \infty \).

Under certain nice conditions on \( a_j \)'s, with probability one, \( F_n(T_n/\sqrt{n}) \) and \( F_n(H_n/\sqrt{n}) \) converge weakly as \( n \to \infty \) to non-random symmetric probability measures \( \lambda_T \) and \( \lambda_H \) respectively. The LSDs \( \lambda_T \) and \( \lambda_H \) do not depend on the distribution of \( \epsilon_1 \).
It can be shown that in the Hankel case the moments of the LSD are given by

$$\frac{\beta_{2k}^*}{\beta_{2k}} = \sum_{m: m_1 + m_3 + \ldots + m_{2k-1} = m_2 + m_4 + \ldots + m_{2k}} a_{m_1} a_{m_2} \cdots a_{m_{2k}}$$

where $\beta_{2k}$ is the $2k$-th moment of the LSD of the Hankel random matrix with i.i.d. entries with variance 1.

**Corollary:** If $X_t = \epsilon_t + \theta \epsilon_{t-1}$ then

$$\beta_{2k}^* = \beta_{2k} \sum_{i=0}^{k} \binom{k}{i}^2 \theta^{2i}.$$
**Theorem** Suppose $X_t = \epsilon_t \epsilon_{t+1}$ where $\epsilon_t$ are i.i.d. with $E(\epsilon_1) = 0$ $Var(\epsilon_1) = 1$. Then the SAME limit as in the iid case holds for the Toeplitz matrix.

Figure 1: Kernel density estimates for the ESD for 5 simulated Toeplitz matrices of order 600 with $X_t = \epsilon_t \epsilon_{t+1}$, $\epsilon_t \sim N(0,1)$.
Toeplitz with varying scale

We can take a different scaling of the random Toeplitz matrix. For example, consider

\[
T_n^*(i,j) = \frac{x|i-j|}{\sqrt{n - |i-j|}}.
\]

Only partially completed proof.
Figure 2: Kernel density estimates for the ESD for 20 specially normalised simulated Toeplitz matrices of order 800 with $X_t \sim N(0,1)$. 
Autocovariance matrix

\[ \hat{\Gamma}_n = (\hat{\gamma}(i-j))_{i,j}, \quad \hat{\gamma}(k) = n^{-1} \sum_{i=1}^{n-|k|} x_i x_{i+|k|} \]

Figure 3: Empirical distribution of eigenvalues of 15 realizations of \( \hat{\Gamma}_n \) of order 400 with standardised Bernoulli(0.5) entries.
Modified autocovariance

\[ \hat{\Theta}_n = ((\hat{\theta}(i - j)))_{i, j}, \quad \hat{\theta}(k) = n^{-1} \sum_{i=1}^{n} x_i x_{i+k} \]

Figure 4: Empirical distribution of eigenvalues of 15 realizations of \( \Theta_n \) of order 400 with standardised Bernoulli(0.5) entries.