**Undergraduate Research Opportunity Programme in Science**

# The Moment Method in Determining Limiting Densities of Singular Values of Powers of Random Matrices

NG SAY-YAO, MARK

Department of Mathematics

National University of Singapore

2015/2016

# The Moment Method in Determining Limiting Densities of Singular Values of Powers of Random Matrices

Ng M.[1]

*Department of Mathematics, National University of Singapore*
*10 Lower Kent Ridge Road, Singapore 119076*

## Abstract

In this paper, we establish three significant results in the theory of random matrices. They are Wigner's semicircle law [1, 3, 7, 5, 15], the Marchenko-Pastur law [11, 16], and finally the asymptotic distribution of singular values of powers of random matrices [2]. These results correspond to the limiting densities of eigenvalues of three different random matrix ensembles. The key technique here is the moment method – in particular, we focus on proving convergence of expected moments in order to conclude convergence in distribution [5]. More importantly, we learn about the importance of combinatorics and graphs in establishing the convergence of expected moments. We attempt to explain thoroughly what is observed for each random matrix ensemble, and provide detailed descriptions of the combinatorial problems we need to solve.

## 1 Introduction

Random matrix theory is the study of matrices with random variable entries [15]. There are various random matrix models in which this study has made much progress over the years. One example of such is a random $n \times n$ Hermitian matrix. As its name would suggest, this model consists of a random matrix which is necessarily Hermitian and therefore always has $n$ eigenvalues (up to multiplicity) $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ located along the real line. Naturally, one could ask for various statistics of these eigenvalues – such as a histogram, the largest/smallest eigenvalue, gaps between some consecutive eigenvalues etc.

This brings about the phenomenon of universality [14] where for large classes of random matrix models, the limiting distribution of such statistics converges to a universal distribution (usually after appropriate normalization) as the dimension $n \to \infty$. In other words, they end up looking the 'same' in many ways. Furthermore the so-called universal limit usually has a rich algebraic structure. This is analogous to a simpler phenomenon for *scalar* random variables, known as the Central Limit Theorem (CLT). In the case of the CLT, the standard normal (or Gaussian) distribution is the universal limiting distribution – indeed rich in algebraic structure. The development of mathematics with respect to the CLT has brought about deeper understanding of such a universality for example, through further generalisations of the CLT.

On the other hand, not the same can be said in terms of universality for random matrices. Most

[1] Student, National University of Singapore

progress in this area involves only certain classes of random matrix models, also known as ensembles. Examples of such include the Wigner and Wishart ensembles. This paper focuses on using a technique known as the moment method in proving the limiting density of eigenvalues for noticeable ensembles of random matrices. At this point it is useful to note that the eigenvalues themselves are a random vector, depending on the entries of the random matrix. The underlying principle of the moment method observed in this paper is to prove that the expected moments of the eigenvalue distribution convergences to the moments of the limiting distribution as the dimension of the matrix gets larger. Under appropriate conditions [6], this is sufficient to prove convergence in distribution (also known as *weak* convergence) of the spectral measure of eigenvalues[2]. We will therefore present moment method-style proofs used to establish the limiting densities for three types of random matrix ensembles. The first and second are commonly known as Wigner's Semicircle Law and the Marchenko-Pastur Law respectively. The third is from our Reference [2] which concerns the limiting density of singular values of powers of random matrices.

The moment method has its merits, especially in determining results on a global [3] or as some call it, macroscopic [14] scale. More importantly, we learn to appreciate the role of enumerative combinatorics as well as graphs in such a scenario. Random matrix theory itself draws from a wide variety of concepts – containing at the very least probability theory, linear algebra, and mathematical analysis. The relation between a the sum of eigenvalues of a matrix and its trace brings a good insight as to how we should first approach the problem. Furthermore certain terms contained in the trace of a matrix have interesting combinatorial and graph-like properties. We find that the Catalan numbers and it's generalisation – the Fuss Catalan numbers [2, 9] plays an important role in finding solutions to the combinatorial problem at hand.

In the case of Wigner's Semicircle Law, we will demonstrate how to use convergence in expected moments to explicitly prove convergence in distribution. This requires us to first establish some control over the variance [3] of the eigenvalue distribution. Then we use some relatively simpler arguments involving the Weierstrass Approximation Theorem, Markov Inequality, and obtaining some upper bound on the moments to show the desired result. For the other two limiting densities, we stick to the proving the convergence of expected moments. This is because our main focus is to understand the combinatorial problems observed when applying the moment method.

[2]To be more precise, the *random* probability measure of eigenvalues converges weakly to some distribution, in probability.

# 2 Preliminaries

We begin by developing essential concepts and results applied in the moment method-style proofs for establishing limiting densities of eigenvalues, or more generally – the singular values of random matrices. These concepts are central to the idea of the proofs established later on.

## 2.1 Linear Algebra

Although we generally assume the well-known results linear algebra throughout, we state and prove those which are emphasized greatly later on.

**Proposition 2.1.1.** Let $A$ be a $n \times n$ complex-valued matrix. Given any $k \in \mathbb{N}$,

$$\text{tr}\, A^k = \sum_{i=1}^{n} \lambda_i^k. \tag{1}$$

*Proof.* Write $A = P^{-1}JP$ where $J$ is the Jordan Canonical form of the matrix $A$. Note that $J$ is necessarily upper triangular, with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_n$ occurring along its diagonal. Therefore by property of the trace

$$\begin{aligned}\text{tr}\, A^k &= \text{tr}\,(P^{-1}JP)^k \\ &= \text{tr}\, P^{-1}J^kP \\ &= \text{tr}\, J^kP^{-1}P \\ &= \text{tr}\, J^k \\ &= \sum_{i=1}^{n} \lambda_i^k.\end{aligned} \tag{2}$$

□

**Proposition 2.1.2.** Let $A$ be a $n \times n$ complex-valued matrix. Then given $k \in \mathbb{N}$ we have

$$(A^k)_{ij} = \sum_{1 \le \ell_1, \dots, \ell_{k-1} \le n} A_{i\ell_1} A_{\ell_1\ell_2} \cdots A_{\ell_{k-1}j}. \tag{3}$$

*Proof.* When $k = 1, 2$ is it clear. We prove this by induction. If it were true for some $k \in \mathbb{N}$ then by definition of matrix multiplication

$$\begin{aligned}(A^{k+1})_{ij} &= \sum_{\ell=1}^{n} (A^k)_{i\ell} A_{\ell j} \\ &= \sum_{\ell=1}^{n} \left( \sum_{1 \le \ell_1, \dots, \ell_{k-1} \le n} A_{i\ell_1} A_{\ell_1\ell_2} \cdots A_{\ell_{k-1}\ell} \right) A_{\ell j} \\ &= \sum_{1 \le \ell_1, \dots, \ell_{k-1}, \ell_k \le n} A_{i\ell_1} A_{\ell_1\ell_2} \cdots A_{\ell_{k-1}\ell_k} A_{\ell_k j},\end{aligned} \tag{4}$$

as was to be shown. □

## 2.2 Catalan Numbers

**Definition 2.2.1.** The Catalan numbers are members of a sequence $(C_k)_{k=1}^{\infty}$ given by the recursive relation defined by $C_0 = 1$ and

$$C_{k+1} = \sum_{i=0}^{k} C_i C_{k-i}, \tag{5}$$

for all $k \geq 0$.

Catalan numbers often appear in problems related to enumerative combinatorics. The applications are vast, in particular Catalan numbers play a fundamental role in this paper. We establish important properties related to Catalan numbers in this section [9].

**Fact 2.2.2.** The $k$-th Catalan number $C_k$ has the following closed form expression:

$$C_k = \frac{1}{k+1}\binom{2k}{k} = \frac{(2k)!}{(k+1)!k!}. \tag{6}$$

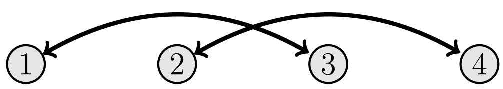
**Definition 2.2.3.** The set of $n$ letters is defined to be the set $S_n = \{1, 2, \ldots, n\}$, where $n \in \mathbb{N}$

*Remark* 2.2.4. As a convention, we also define $S_0 = \emptyset$.

**Definition 2.2.5.** A partition on the set $S_n$ into cells is said to be noncrossing if for any quadruple of elements in $S_n$ satisfying $a_1 < b_1 < a_2 < b_2$ we must not have $a_1, a_2 \in A$ and $b_1, b_2 \in B$ where $A \neq B$ are distinct cells.

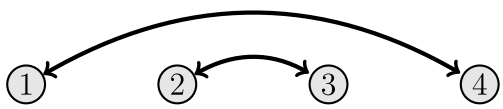
*Example* 2.2.6. For $S_4$ all partitions are noncrossing except the partition $\{\{1,3\},\{2,4\}\}$. This is because $1 < 2 < 3 < 4$ but $2, 4 \in \{2,4\}$ and $1, 3 \in \{1,3\}$. There are a total of $14$ distinct noncrossing partitions on $S_4$. This idea is presented in Figure 1.

Figure 1: Some Partitions on $S_4$




(i) Crossing Partition, $\{\{1,3\},\{2,4\}\}$


(ii) Noncrossing Partition, $\{\{1,2\},\{3,4\}\}$




(iii) Noncrossing Partition, $\{\{1,4\},\{2,3\}\}$

*Remark* 2.2.7. The noncrossing condition can be applied to any set with a linear order.

**Lemma 2.2.8.** The number of noncrossing partitions on $S_n$ is equal to $C_n$.

*Proof.* For $n \geq 0$, let $P_n$ be the number of noncrossing partitions on $S_n$. Now consider $S_{n+1}$. For each $k \in \{1, \ldots, n+1\}$ consider a noncrossing partition in which $k$ and $n+1$ belong to the same cell. Consider the following sets:

$$\begin{aligned} A &= \{m \in S_{n+1} : k \leq m < n+1\} \\ B &= \{m \in S_{n+1} : 0 < m < k\} \end{aligned} \tag{7}$$

It is clear that any other cell in the partition cannot contain an element from $A$ as well as an element from $B$. Therefore the number of noncrossing partitions given that $n+1$ and $k$ (possibly equal numbers) are in the same cell is $P_{k-1}P_{n+1-k}$. Consequently we have

$$\begin{aligned} P_{n+1} &= \sum_{k=1}^{n+1} P_{k-1}P_{n+1-k} \\ &= \sum_{i=0}^{n} P_i P_{n-i}. \end{aligned} \tag{8}$$

It is clear that $P_0 = P_1 = 1$, since there is only one possible noncrossing partition on the empty set as well as a singleton. Hence $P_n = C_n$ for all $n \geq 0$. $\square$

**Definition 2.2.9.** A Dyck path of order $n$ is defined to be a path along the lattice $\mathbb{Z}^2$ starting from $(0,0)$ and ending at $(2n, 0)$ in a total of $2n$ steps. Each step must be taken in either $\vec{u} = (1,1)$ or $\vec{d} = (1,-1)$ with the condition that the position $(x, y)$ at any point must satisfy $y \geq 0$.

**Lemma 2.2.10.** The number of Dyck paths of order $n$ is equal to $C_n$.

*Proof.* Let $D_n$ denote the number of Dyck paths of order $n$. It is clear that $D_0 = D_1 = 1$. We want to show that for all $n \in \mathbb{N}$ the following recurrence relation holds:

$$D_{n+1} = \sum_{i=0}^{n} D_i D_{n-i}. \tag{9}$$

Now, for $n+1 > 1$ we may write a Dyck path $p = up_idp_j$ where $u$ is the compulsory first movement $\vec{u} = (1,1)$ and $d$ is a movement $\vec{d} = (1,-1)$. Note that $p_i, p_j$ are Dyck paths of order strictly less than $n+1$. Apply the induction hypothesis here. Each variation of the position of $d$ creates a unique Dyck path $p$. From this we conclude that the recurrence relation holds. Therefore $D_n = C_n$ for all $n \in \mathbb{N}$. $\square$

*Remark* 2.2.11. A Dyck path of order $n$ can be equivalently defined as a staircase walk on the lattice $\mathbb{Z}^2$ from $(0,0)$ to $(n,n)$ that lies below (but may touch) the line $y = x$. We present[3] this idea informally via Figures 2 and 3.

[3]Figures 2 and 3 are drawn using Python 2.7.

Figure 2: Dyck Paths of Order $3$

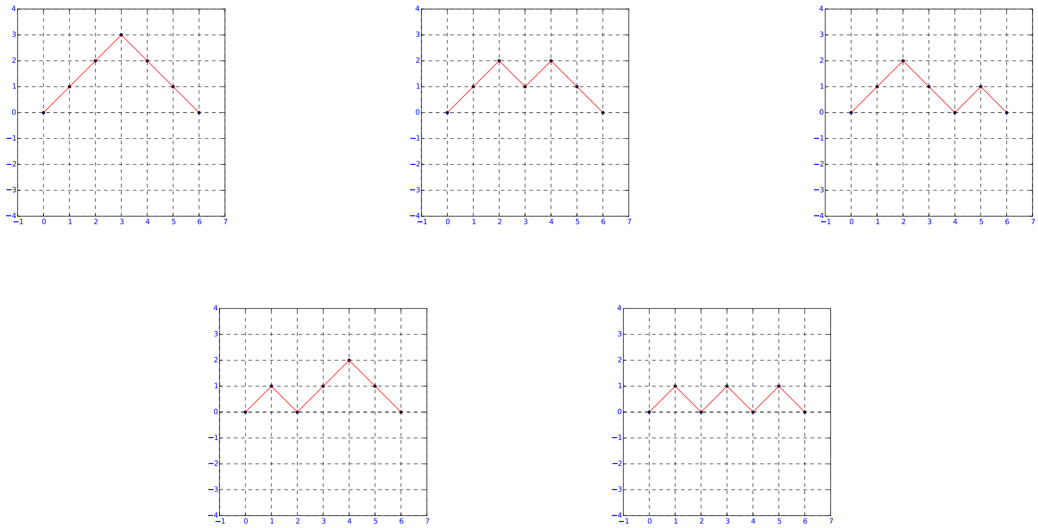
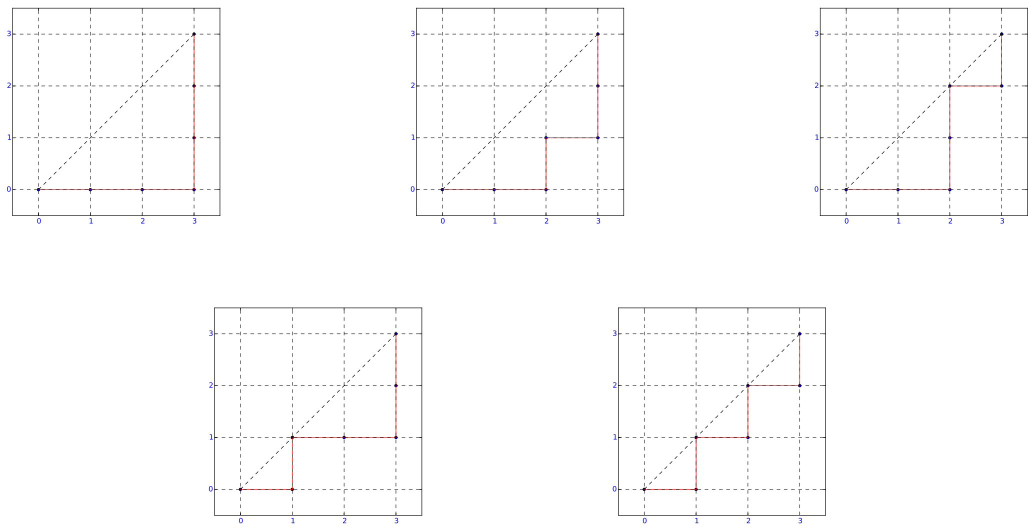


Figure 3: Equivalent Staircase Walks from $(0, 0)$ to $(3, 3)$



## 2.3 Fuss Catalan Numbers

**Definition 2.3.1.** A single-parameter Fuss Catalan number is defined to be a sequence $(C_{m,k})_{k=0}^{\infty}$ for every $m \in \mathbb{N}$ satisfying the recurrence relation given by $C_{m,0} = C_{m,1} = 1$ for all $m \in \mathbb{Z}_{\geq 0}$ and

$$C_{m,k} = \sum_{\Omega_k} \left( \prod_{i=0}^{m} C_{m,k_i} \right), \tag{10}$$

where the underlying domain of the sum is defined to be

$$\Omega_{m,k} = \left\{ (k_0, k_1, \dots, k_m) \in \mathbb{Z}_{\geq 0}^{m+1} : \sum_{i=0}^{m} k_i = k-1 \right\}. \tag{11}$$

**Fact 2.3.2.** For each $m, k \in \mathbb{Z}_{\geq 0}$ and the number $C_{m,k}$ has the following closed form expression:

$$C_{m,k} = \frac{1}{mk+1}\binom{mk+k}{k}. \tag{12}$$

*Remark* 2.3.3. The Fuss Catalan numbers are a generalisation of the Catalan numbers. In particular, when $m = 1$ and $k \in \mathbb{Z}_{\geq 0}$

$$C_{1,k} = C_k. \tag{13}$$

*Remark* 2.3.4. Although this is not much of a concern, it is good to know that for a given $m, k \in \mathbb{Z}_{\geq 0}$ with $k > 1$ the cardinality of $\Omega_{m,k}$ is the number of ways of writing $k-1$ as a sum of $m+1$ nonnegative integers. Any way of writing a positive integer like this is called a $m+1$-weak composition of the integer $k-1$.

$$|\Omega_{m,k}| = \binom{(k-1)+(m+1)-1}{(m+1)-1} = \binom{k+m-1}{m}. \tag{14}$$

## 2.4 Graphs

**Definition 2.4.1.** A graph $\mathcal{G}$ is an ordered pair $(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ are sets. Elements of $\mathcal{V}$ are known as vertices, and elements of $\mathcal{E}$ consist of ordered, or unordered pairs of vertices $u, v \in \mathcal{V}$. $\mathcal{G}$ is said to be directed if $\mathcal{E}$ is a set of ordered pairs $(u, v)$ for some $u, v \in \mathcal{V}$. $\mathcal{G}$ is said to be undirected if $\mathcal{E}$ is a set of unordered pairs $\{u, v\}$ for some $u, v \in \mathcal{V}$.

**Definition 2.4.2.** A walk on a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a finite sequence $w = (v_1, v_2, \dots, v_k)$ of vertices $v_1, v_2, \dots, v_k \in \mathcal{V}$ and for each $1 \leq i \leq k-1$ if the $\mathcal{G}$ is directed then $(v_i, v_{i+1}) \in \mathcal{E}$; otherwise if $\mathcal{G}$ is undirected then $\{v_i, v_{i+1}\} \in \mathcal{E}$.

*Remark* 2.4.3. Another form of notation we will use for describing the walk $w$ above is

$$v_1 \to v_2 \to \cdots \to v_k. \tag{15}$$

**Definition 2.4.4.** A graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is said to be connected if for any two of vertices $v_1, v_2 \in \mathcal{V}$ there exists a walk $w$ on $\mathcal{G}$ such that

$$w = (v_1, \dots, v_2). \tag{16}$$

**Definition 2.4.5.** A simple cycle in a graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is defined to be a walk $w$ of the form

$$w = (u, v_1, \dots, v_k, u), \tag{17}$$

for some $u \in \mathcal{V}$ such that no edges and vertices (other than $u$ itself) are repeated.

**Definition 2.4.6.** A tree is an undirected, connected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ such that $|\mathcal{V}| = |\mathcal{E}| - 1$.

*Remark* 2.4.7. The notion of a tree is important in this paper. In particular, we assume without proof that a tree is equivalently defined as an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ which is connected and has no cycles.

## 2.5 Probability Theory

For the sake of clarity, we state some important definitions and results in probability theory [12].

**Definition 2.5.1.** A sequence of random variables $\{X_n\}_{n=1}^{\infty}$ with corresponding cumulative distributions $\{F_n\}_{n=1}^{\infty}$ is said to converge weakly to a random variable $X$ with cumulative distribution $F$ if

$$\lim_{n\to\infty} F_n(x) = F(x), \tag{18}$$

for every $x \in \mathbb{R}$ in which $F$ is continuous. The shorthand notation for this convergence is

$$X_n \xrightarrow{d} X. \tag{19}$$

*Remark* 2.5.2. Weak convergence is equivalently defined to be the condition that for any bounded continuous function $f : \mathbb{R} \to \mathbb{R}$ we have

$$\lim_{n\to\infty} \int_{\mathbb{R}} f(x)\, \mathrm{d}F_n = \int_{\mathbb{R}} f(x)\, \mathrm{d}F. \tag{20}$$

The bounded continuous function $f$ is referred to as a test function.

**Definition 2.5.3.** A sequence of random variables $\{X_n\}_{n=1}^{\infty}$ is said to converge in probability to a random variable $X$ if for any given $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbf{P}\left(|X_n - X| \geq \varepsilon\right) = 0. \tag{21}$$

The shorthand notation for this convergence is

$$X_n \xrightarrow{p} X. \tag{22}$$

**Definition 2.5.4.** A sequence of random variables $\{X_n\}_{n}^{\infty}$ is said to said converge almost surely to a random variable $X$ if

$$\mathbf{P}\left(\lim_{n\to\infty} X_n = X\right) = 1. \tag{23}$$

The shorthand notation for this convergence is

$$X_n \xrightarrow{a.s.} X. \tag{24}$$

*Remark* 2.5.5. It can be shown that the following relation holds:

Almost Sure Convergence $\implies$ Convergence in Probability $\implies$ Weak Convergence.

**Definition 2.5.6.** For $k \in \mathbb{N}$, the $k$-th moment of a random variable $X$ (or distribution $F$) is defined to be

$$\mathbf{E}X^k = \int_{\Omega} x^k\, \mathrm{d}F. \tag{25}$$

The following are well-known results which will be used establishing proofs for convergence stated in the later part of this paper.

**Fact 2.5.7** (Markov Inequality)**.** Let $X$ be a nonnegative random variable. For any $\varepsilon > 0$,

$$\mathbf{P}(X \geq \varepsilon) \leq \frac{\mathbf{E}X}{\varepsilon}. \tag{26}$$

**Fact 2.5.8** (Chebyshev's Inequality)**.** Let $X$ be a random variable with $\mathbf{E}X < \infty$. For any $\varepsilon > 0$,

$$\mathbf{P}\left(|X - \mathbf{E}X| \geq \varepsilon\right) \leq \frac{\mathbf{Var}X}{\varepsilon^2}. \tag{27}$$

## 2.6 Weierstrass Approximation Theorem

We state a a well-known theorem in analysis and approximation theory.

**Fact 2.6.1** (Weierstrass Approximation Theorem)**.** Let $f : [a, b] \to \mathbb{R}$ be a continuous function. Then given $\delta > 0$ there exists a polynomial $p_\delta$ such that

$$\sup_{x\in[a,b]} \{|f - p_\delta|\} < \delta. \tag{28}$$

This theorem will be utilized in the later parts of this paper. In particular, it is used to explicitly prove weak convergence for Wigner's semicircle law. Before doing so however, consider the case where the interval is $[0, 1]$. We provide an interesting probabilistic proof for this case, as stated in the following proposition. The proof is adapted from Reference [10]

**Proposition 2.6.2.** Let $f : [0, 1] \to \mathbb{R}$ be a continuous function. Then given $\delta > 0$ there exists a polynomial $B_\delta$ such that

$$\sup_{x\in[0,1]} \{|f - B_\delta|\} < \delta. \tag{29}$$

*Proof.* We construct a Bernstein polynomial of $f$

$$B_n = \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} x^k (1-x)^{n-k} f(\frac{k}{n}). \tag{30}$$

Let $\alpha > 0$. Then

$$\begin{aligned} |B_n(x) - f(x)| &= \left| \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} x^k (1-x)^{n-k} (f(\frac{k}{n}) - f(x)) \right| \\ &\leq \sum_{k:|\frac{k}{n}-x|<\alpha} \frac{n!}{k!(n-k)!} x^k (1-x)^{n-k} \left| f(\frac{k}{n}) - f(x) \right| \\ &+ \sum_{k:|\frac{k}{n}-x|\geq\alpha} \frac{n!}{k!(n-k)!} x^k (1-x)^{n-k} \left| f(\frac{k}{n}) - f(x) \right| \\ &= I_1 + I_2. \end{aligned} \tag{31}$$

Now, $f$ is necessarily uniformly continuous. Take $\alpha > 0$ such that $|f(k/n) - f(x)| < \delta/2$. Since

$$0 \leq \sum_{k:|\frac{k}{n}-x|<\alpha} \frac{n!}{k!(n-k)!} x^k (1-x)^{n-k} \leq 1, \tag{32}$$

we have

$$I_1 < \frac{\delta}{2}. \tag{33}$$

Next, $f$ is necessarily bounded on $[0, 1]$. This means there exists $M \in \mathbb{R}$ such that $|f(x)| \leq M$ over $[0, 1]$. Therefore

$$I_2 \leq 2M \sum_{k:|\frac{k}{n}-x|\geq\alpha} \frac{n!}{k!(n-k)!} x^k (1-x)^{n-k}. \tag{34}$$

Let $X \sim \text{Binomial}(n, x)$. Observe that

$$\begin{aligned}\sum_{k:|\frac{k}{n}-x|\geq\alpha} \frac{n!}{k!(n-k)!}x^k(1-x)^{n-k} &= \mathbf{P}\left(|\frac{1}{n}X - x| \geq \alpha\right) \\ &= \mathbf{P}\left(|X - \mathbf{E}X| \geq n\alpha\right).\end{aligned} \tag{35}$$

By Chebyshev's Inequality and sufficiently large $n$ we obtain

$$\begin{aligned}\mathbf{P}\left(|X - \mathbf{E}X| \geq n\alpha\right) &\leq \frac{nx(1-x)}{n^2\alpha^2} \\ &= \frac{x(1-x)}{n\alpha^2} \\ &< \frac{\delta}{4M}.\end{aligned} \tag{36}$$

This implies that

$$I_2 < \frac{\delta}{2}. \tag{37}$$

Therefore

$$|B_n(x) - f(x)| < \frac{\delta}{2} + \frac{\delta}{2} = \delta. \tag{38}$$

$\square$

# 3 Wigner's Semicircle Law

In this section we will prove Wigner's semicircle law using the moment method. We gain some insight on the combinatorial and graph properties associated with the trace of some random matrices, as well as their moments. For the following two sections, a similar train of thought applies.

## 3.1 Wigner Random Matrices

**Definition 3.1.1.** A Wigner random matrix is a random Hermitian matrix $H_n = (H_{ij})$ such that all of its upper triangular entries are independent and

1. The diagonal entries satistfy $\mathbf{E}H_{ii} = 0$ and $\mathbf{E}H_{ii}^2 = 1$.
2. For $i < j$, $\mathbf{E}H_{ij} = 0$ and $\mathbf{E}|H_{ij}|^2 = 1$.
3. All entries have bounded moments, which means for each $k \in \mathbb{N}$ there exists $M_k \in \mathbb{R}$ such that $\mathbf{E}|H_{ij}|^k \leq M_k$.

By the Hermitian property, $H_n$ has necessarily $n$ real eigenvalues which we may assume to follow the notation $\lambda_1 \leq \cdots \leq \lambda_n$. This motivates us to define the following.

**Definition 3.1.2.** The Empirical Spectral Measure (ESM) of eigenvalues of a Wigner matrix $H_n$ is defined to be

$$\mu_n(x) = \frac{1}{n} \sum_{i=1}^{n} \delta_{\frac{\lambda_i}{\sqrt{n}}}(-\infty, x] \tag{39}$$

Here $\delta_{\frac{\lambda}{\sqrt{n}}}$ denotes the Dirac measure on $\mathbb{R}$. In particular we have

$$\delta_{\frac{\lambda_i}{\sqrt{n}}}(-\infty, x] = \begin{cases} 1 & \text{if } \frac{\lambda_i}{\sqrt{n}} \in (-\infty, x] \\ 0 & \text{if } \frac{\lambda_i}{\sqrt{n}} \notin (-\infty, x] \end{cases} \tag{40}$$

*Remark* 3.1.3. The ESM is essentially the cumulative distribution of eigenvalues corresponding to the Wigner matrix $H$ along the real lines.

*Remark* 3.1.4. Note that the ESM, $\mu_n$, is a random variable in it's own right. More precisely, the ESM is a random probability measure on the real line [14]. This is because it depends on the eigenvalues arising from the random matrix $H_n$.

**Definition 3.1.5.** The semicircle distribution over the interval $[-2, 2]$ is defined by the probability density function $\sigma$ as follows

$$\sigma(x) = \frac{1}{2\pi}\sqrt{4 - x^2} \cdot \chi_{[-2,2]} \tag{41}$$

## 3.2 Convergence of Expected Moments

Our first aim is to show that the expected moments of $\mu_n$ converges to the moments of the semicircle distribution $\sigma$. In order to do so, we establish an critical identity between the moments of the semicircle distribution and the Catalan numbers.

**Lemma 3.2.1.** The moments of the semicircle distribution is given by the following identity

$$\mathbf{E}X^k = \int_{-2}^{2} x^k \sigma(x)\,\mathrm{d}x = \begin{cases} C_{k/2} & \text{if } k \equiv 0 \mod 2 \\ 0 & \text{if } k \equiv 1 \mod 2 \end{cases} \tag{42}$$

*Proof.* Let $k \geq 0$ be given. We split the proof into two cases.

Case #1: $k \equiv 1 \mod 2$. Then $x^k\sigma(x)$ is an odd function. By symmetry we have

$$\mathbf{E}X^k = \int_{-2}^{2} x^k \sigma(x)\,\mathrm{d}x = 0. \tag{43}$$

Case #2: $k \equiv 0 \mod 2$. Then we write $2k$ for the sake of simplicity and partially evaluate the integral as follows

$$\begin{aligned}
\mathbf{E}X^{2k} &= \int_{-2}^{2} x^{2k}\sigma(x)\,\mathrm{d}x \\
&= \int_{-2}^{2} \frac{1}{2\pi} x^{2k}\sqrt{4-x^2}\,\mathrm{d}x \\
&= \frac{2^{2k+1}}{\pi}\int_{-\pi/2}^{\pi/2} \sin^{2k}\theta\cos^2\theta\,\mathrm{d}\theta \\
&= \frac{2^{2k+1}}{\pi}\left(\int_{-\pi/2}^{\pi/2} \sin^{2k}\theta \mathrm{d}\,\theta - \int_{-\pi/2}^{\pi/2} \sin^{2k+2}\theta\,\mathrm{d}\theta\right) \\
&= \frac{2^{2k+1}}{\pi}\left(\int_{-\pi/2}^{\pi/2} \sin^{2k}\theta\,\mathrm{d}\theta - (2k+1)\int_{-\pi/2}^{\pi/2} \sin^{2k}\theta\cos^2\theta\,\mathrm{d}\theta\right) \\
&= \frac{2^{2k+1}}{\pi}\int_{-\pi/2}^{\pi/2} \sin^{2k}\,\mathrm{d}\theta - (2k+1)\mathbf{E}X^{2k}.
\end{aligned} \tag{44}$$

This defines a recrurrence relation between $\mathbf{E}X^{2k}$ and $\mathbf{E}X^{2k-2}$ since the above equality implies

$$\begin{aligned}
\mathbf{E}X^{2k} &= \frac{2^{2k+1}}{(2k+2)\pi}\int_{-\pi/2}^{\pi/2} \sin^{2k}\theta\,\mathrm{d}\theta \\
&= \frac{2^{2k+1}}{(2k+2)\pi}\left((2k-1)\int_{-\pi/2}^{\pi/2} \sin^{2k-2}\theta\cos^2\theta\,\mathrm{d}\theta\right) \\
&= \frac{2^2(2k-1)}{(2k+2)}\mathbf{E}X^{2k-2}.
\end{aligned} \tag{45}$$

The Catalan numbers also satisfy a similar recurrence relation since

$$\begin{aligned} C_k &= \frac{(2k)!}{(k+1)!(k)!} \\ &= \frac{(2k)(2k-1)(2k-2)!}{(k+1)(k)(k)!(k-1)!} \\ &= \frac{(2k)(2k-1)}{(k+1)(k)} C_{k-1} \\ &= \frac{2^2(2k-1)}{(2k+2)} C_{k-1}. \end{aligned} \tag{46}$$

We check that for $k = 0$, $\mathbf{E}X^0 = 1 = C_0$. By induction $\mathbf{E}X^{2k} = C_k$ for all integers $k \geq 0$. $\square$

The following lemma concerns the convergence of *expected* moments of the ESM. This is the core of the moment method, which we will observe in future sections as well. We will also see how Lemma 3.2.1 helps us establish a proof for such convergence.

**Lemma 3.2.2.** Fix any $k \in \mathbb{N}$ and let $H_n$ be a Wigner random matrix for $n \in \mathbb{N}$. Then

$$\lim_{n\to\infty} \mathbf{E} \int_{\mathbb{R}} x^k \, \mathrm{d}\mu_n = \int_{\mathbb{R}} x^k \sigma(x) \, \mathrm{d}x. \tag{47}$$

*Proof.* By linearity of expectation, we express the sequence on the left-hand side as follows

$$\begin{aligned} \mathbf{E} \int_{\mathbb{R}} x^k \, \mathrm{d}\mu_n &= \mathbf{E} \, \frac{1}{n} \sum_{i=1}^{n} \left( \frac{\lambda_i}{\sqrt{n}} \right)^k \\ &= \mathbf{E} \, \frac{1}{n^{k/2+1}} \sum_{i=1}^{n} \lambda_i^k \\ &= \mathbf{E} \, \frac{1}{n^{k/2+1}} \mathrm{tr} H_n^k \\ &= \frac{1}{n^{k/2+1}} \mathbf{E} \, \mathrm{tr} H_n^k \\ &= \frac{1}{n^{k/2+1}} \sum_{1 \leq \ell_0, \ell_1, \ldots, \ell_{k-1} \leq n} \mathbf{E} H_{\ell_0\ell_1} H_{\ell_1\ell_2} \cdots H_{\ell_{k-1}\ell_0}. \end{aligned} \tag{48}$$

We focus on the terms of the summation above. Our primary concern is to identify the terms which are nonzero. Since $\mathbf{E}H_{ij} = 0$ for all $1 \leq i, j \leq n$ we deduce that each independent factor must occur at least twice, otherwise $\mathbf{E}H_{\ell_0\ell_1} H_{\ell_1\ell_2} \cdots H_{\ell_{k-1}\ell_0} = 0$. After removing the zero terms from the sum, there are at most $k/2$ independent $H_{ij}$'s in each term.

Define a sequence of digits $\rho = \ell_0\ell_1 \cdots \ell_{k-1}\ell_0$ for each term $\mathbf{E}H_{\ell_0\ell_1} H_{\ell_1\ell_2} \cdots H_{\ell_{k-1}\ell_0}$. Note that each $\ell_i \in S_n = \{1, \ldots, n\}$. Furthermore for each $\rho$, define its weight to be $\omega(\rho) = |\{\ell_0, \ell_1, \ldots, \ell_{k-1}\}|$. In other words, $\omega(\rho)$ is the number of distinct digits in the sequence $\rho$. If the term is nonzero, then $\omega(\rho) \leq k/2 + 1$ since $\omega(\rho) > k/2 + 1$ implies that there are more than $k/2$ independent $H_{ij}$'s in the term.

Define a relation $\sim$ between any two $\rho = \ell_0\ell_1\cdots\ell_{k-1}$ and $\rho' = \ell_0'\ell_1'\cdots\ell_{k-1}'$ via $\rho \sim \rho'$ if and only if there exists a bijection $f : S_n \to S_n$ such that $f(\ell_i) = \ell_i'$ for all $0 \le i \le k-1$. It is clear that $\sim$ is an equivalence relation, and $\rho \sim \rho'$ implies that $\omega(\rho) = \omega(\rho')$ and $\mathbf{E}H_{\ell_0\ell_1}H_{\ell_1\ell_2}\cdots H_{\ell_{k-1}\ell_0} = \mathbf{E}H_{\ell_0'\ell_1'}H_{\ell_1'\ell_2'}\cdots H_{\ell_{k-1}\ell_0}$.

Let $\mathcal{P}_{\omega_0}$ be an equivalence class induced by $\sim$ such that for all $\rho \in \mathcal{P}_{\omega_0}$, $\omega(\rho) = \omega_0 < k/2+1$. We claim that the contribution of the corresponding terms $\mathbf{E}H_{\ell_0\ell_1}H_{\ell_1\ell_2}\cdots H_{\ell_{k-1}\ell_0}$ goes to zero as $n \to \infty$. Since there are $\omega_0$ distinct ordered digits in $\rho \in \mathcal{P}_{\omega_0}$ due to bijection, and these digits are taken from $S_n$ we conclude that

$$\begin{aligned} |\mathcal{P}_{\omega_0}| &= n(n-1)(n-2)\cdots(n-\omega_0+1) \\ &= O(n^{\omega_0}). \end{aligned} \tag{49}$$

By the assumption that all moments of $H_{ij}$ are bounded we have

$$\begin{aligned} \frac{1}{n^{k/2+1}}\sum_{\rho\in\mathcal{P}_{\omega_0}} \mathbf{E}H_{\ell_0\ell_1}H_{\ell_1\ell_2}\cdots H_{\ell_{k-1}\ell_0} &= \frac{1}{n^{k/2+1}}\sum_{\rho\in\mathcal{P}_{\omega_0}} O(1) \\ &\le O\left(\frac{n^{\omega_0}}{n^{k/2+1}}\right) \to 0. \end{aligned} \tag{50}$$

Therefore, whenever $\omega(\rho) < k/2+1$ the corresponding term and it's equivalent terms go to zero as $n \to \infty$. This proves the case where $k \equiv 1 \mod 2$, since it is impossible for $\omega(\rho) = k/2+1$ if $k$ is odd.

We are left to prove the case where $k \equiv 0 \mod 2$. By our previous work, we only need to consider terms $\mathbf{E}H_{\ell_0\ell_1}H_{\ell_1\ell_2}\cdots H_{\ell_{k-1}\ell_0}$ fulfilling the following conditions

1. Each independent factor $H_{ij}$ occurs at least twice.
2. The corresponding sequence $\rho$ has the property that $\omega(\rho) = k/2+1$.

We claim that each independent factor $H_{ij}$ along the term $\mathbf{E}H_{\ell_0\ell_1}H_{\ell_1\ell_2}\cdots H_{\ell_{k-1}\ell_0}$ occurs exactly twice. To see why this is true, construct a undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ where the vertices are given by the set $\mathcal{V} = \{\ell_0, \ell_1, \ldots, \ell_{k-1}\}$ and undirected edges by $\mathcal{E} = \{\{\ell_0,\ell_1\}, \{\ell_1,\ell_2\}, \ldots, \{\ell_{k-1},\ell_0\}\}$. Note that we use unordered pair notation $\{u, v\}$ to represent an undirected edge between vertices $u$ and $v$. Since the graph $\mathcal{G}$ is connected we have the following relation

$$\omega(\rho) = k/2+1 = |\mathcal{V}| \le |\mathcal{E}|+1. \tag{51}$$

Note that each unordered pair, or simply undirected edge $\{i, j\}$, represents a distinct independent factor $H_{ij}$ in the term $\mathbf{E}H_{\ell_0\ell_1}H_{\ell_1\ell_2}\cdots H_{\ell_{k-1}\ell_0}$. If there exists an independent factor $H_{ij}$ occurring more than twice, then $|\mathcal{E}| < k/2$, which violates the above inequality. Hence each independent factor $H_{ij}$ occurs exactly twice.

Note that this necessarily implies that $\mathcal{G}$ is a tree, since $|\mathcal{E}| = k/2$ and

$$k/2+1 = |\mathcal{V}| = |\mathcal{E}|+1. \tag{52}$$

The tree property leads us to the conclusion that each independent pair consists of one $H_{ij}$ and one $H_{ji}$. This is justified as follows. For the term $\mathbf{E}H_{\ell_0\ell_1}H_{\ell_1\ell_2}\cdots H_{\ell_{k-1}\ell_0}$ we may define a walk such that the order of vertices visited is

$$\ell_0 \to \ell_1 \to \ell_2 \to \cdots \to \ell_{k-1} \to \ell_{k-1}. \tag{53}$$

By the fact that every edge occurs exactly twice, we find that it is impossible to have $H_{ij}$ followed by another copy $H_{ij}$. If this were true, then the order of vertices visted due to the walk is

$$\ell_0 \to \cdots \to i \to j \to \cdots \to i \to j \to \cdots \to \ell_{k-1} \to \ell_0. \tag{54}$$

which clearly indicates that there exists a simple cycle in $\mathcal{G}$. This is impossible as $\mathcal{G}$ is a tree. Hence every factor $H_{ij}$ of the term $\mathbf{E}H_{\ell_0\ell_1}H_{\ell_1\ell_2}\ldots H_{\ell_{k-1}\ell_0}$ must occurs exactly twice; once as $H_{ij}$ and another time as $H_{ji} = \bar{H}_{ij}$

Let $\mathcal{P}_{\omega_1}$ be an equivalence class induced by $\sim$ such that for all $\rho \in \mathcal{P}_{\omega_1}, \omega(\rho) = \omega_1 = k/2+1$. We want to count how many equivalence classes there are of this form. To do this first define the path $\Gamma(\rho) = \gamma_0\gamma_1\gamma_2\gamma_3\cdots\gamma_k$ to be a sequence of digits where each $\gamma_i$ defined based on the following rule

1. $\gamma_0 = 0$.
2. For $1 < i \leq k$, if $\ell_j \neq \ell_i$ for all $0 \leq j < i$ then $\gamma_i = \gamma_{i-1} + 1$. Otherwise there exists $0 \leq j < i$ such that $\ell_j = \ell_i$, in this case $\gamma_i = \gamma_i - 1$.

Such a path $\Gamma(\rho)$ consists of only nonnegative digits. We prove this by noting that the path $\Gamma(\rho)$ is a record of 'new' and 'old' vertices visited along the walk defined on $\mathcal{G}$. If there exists a negative digit at say position $1 \leq m \leq k$, we may write $m = m_+ + m_-$ where $m_+$ (respectively $m_-$) is the number of digits in $\ell_0\ell_1\ldots\ell_m$ which correspond to an increase (respectively decrease) along $\gamma_1\gamma_2\ldots\gamma_m$. It follows that $m_+ < m_-$. This means that we must have recorded more old vertices than new vertices visited. Given that we can only visit each edge once in each opposite direction and $\mathcal{G}$ is a tree, this is impossible.

Finally, we will always have $\gamma_k = 0$. This is true since every edge is visited twice; once in each opposite direction. In particular, each time we visit a new edge, the digit increases. On the other hand, each time we revisit an edge, the digit decreases.

We conclude that $\Gamma(\rho)$ is a Dyck path. Furthermore $\rho \sim \rho'$ if and only if $\Gamma(\rho) = \Gamma(\rho')$. In other words, each equivalence class $\mathcal{P}_{\omega_1}$ corresponds to a unique Dyck Path of order $k/2$. By Lemma 2.2.10, the total number of such paths is $C_{k/2}$.

The cardinality of each $\mathcal{P}_{\omega_1}$ is given by

$$\begin{aligned} |\mathcal{P}_{\omega_1}| &= n(n-1)\cdots(n-\omega_1+1) \\ &= n^{\omega_1}(1+o(1)) \\ &= O(n^{k/2+1}). \end{aligned} \tag{55}$$

Since $H_{ij}$ occurs twice in conjugate pairs and $\mathbf{E}H_{ij}H_{ji} = 1$, we have $\mathbf{E}H_{\ell_0\ell_1}H_{\ell_1\ell_2}\cdots H_{\ell_{k-1}\ell_0} = 1$.

Therefore it follows that

$$
\begin{aligned}
\frac{1}{n^{k/2+1}}\sum_{\mathcal{P}_{\omega_1}}\sum_{\rho\in\mathcal{P}_{\omega_\infty}}\mathbf{E}H_{\ell_0\ell_1}H_{\ell_1\ell_2}\cdots H_{\ell_{k-1}\ell_0} &= \frac{1}{n^{k/2+1}}\sum_{\mathcal{P}_{\omega_1}}\sum_{p\in\mathcal{P}_{\omega_1}}1 \\
&= \frac{1}{n^{k/2+1}}\sum_{\mathcal{P}_{\omega_1}}O(n^{k/2+1}) \\
&= \sum_{\mathcal{P}_{\omega_1}}(0+o(1)) \\
&= C_{k/2}\cdot O(1)\to C_{k/2}.
\end{aligned}
\tag{56}
$$

This completes the proof for the convergence of expected moments. □

## 3.3 Proof of the Semicircle Law

We require one more lemma in before establishing Wigner's semicircle law. Note that in the previous subsection, Lemma 3.2.2 concerns a convergence which is in fact deterministic. The following lemma however, concerns one which is probabilistic.

**Lemma 3.3.1.** Fix any $k\in\mathbb{N}$ and let $H_n$ be a Wigner random matrix. Then

$$
\int_{\mathbb{R}} x^k\,\mathrm{d}\mu_n - \mathbf{E}\int_{\mathbb{R}} x^k\,\mathrm{d}\mu_n \xrightarrow{p} 0. \tag{57}
$$

*Proof.* Note that we are essentially trying to prove that

$$
\lim_{n\to\infty}\mathbf{P}\left(\left|\int_{\mathbb{R}} x^k\,\mathrm{d}\mu_n - \mathbf{E}\int_{\mathbb{R}} x^k\,\mathrm{d}\mu_n\right|\geq\varepsilon\right)=0. \tag{58}
$$

Let $\varepsilon>0$ be given. Apply Chebyshev's Inequality and to obtain the following expression

$$
\begin{aligned}
\mathbf{P}\left(\left|\int_{\mathbb{R}} x^k\,\mathrm{d}\mu_n - \mathbf{E}\int_{\mathbb{R}} x^k\,\mathrm{d}\mu_n\right|\geq\varepsilon\right) &\leq \frac{1}{\varepsilon^2}\mathbf{Var}\int_{\mathbb{R}} x^k\,\mathrm{d}\mu_n \\
&= \frac{1}{\varepsilon^2}\left(\mathbf{E}\left(\int_{\mathbb{R}} x^k\,\mathrm{d}\mu_n\right)^2 - \left(\mathbf{E}\int_{\mathbb{R}} x^k\,\mathrm{d}\mu_n\right)^2\right) \\
&= \frac{1}{\varepsilon^2}\left(\mathbf{E}\left(\frac{1}{n^{k/2+1}}\mathrm{tr}\,H_n^k\right)^2 - \left(\mathbf{E}\frac{1}{n^{k/2+1}}\mathrm{tr}\,H_n^k\right)^2\right) \\
&= \frac{1}{\varepsilon^2 n^{k+2}}\left(\mathbf{E}(\mathrm{tr}\,H_n^k)^2) - \left(\mathbf{E}\,\mathrm{tr}H_n^k\right)^2\right) \\
&= \frac{1}{\varepsilon^2 n^{k+2}}\sum_{\mathcal{H},\mathcal{H}'}\mathbf{E}\mathcal{H}\mathcal{H}' - \mathbf{E}\mathcal{H}\mathbf{E}\mathcal{H}'.
\end{aligned}
\tag{59}
$$

where $\mathcal{H}=H_{\ell_0\ell_1}H_{\ell_1\ell_0}\cdots H_{\ell_{k-1}\ell_0}$ and $\mathcal{H}'=H_{\ell_0'\ell_1'}H_{\ell_1'\ell_2'}\cdots H_{\ell_{k-1}\ell_0}$ are all possible terms consisting of factors $H_{ij}$ which taken to be entries off the matrix $H_n$ and $1\leq\ell_i,\ell_i'\leq n$. From this step we can already conclude two properties concerning the summation above. In particular if a term $\mathbf{E}\mathcal{H}\mathcal{H}'-\mathbf{E}\mathcal{H}\mathbf{E}\mathcal{H}'$ were to be nonzero, then

1. The terms $\mathcal{H}$, $\mathcal{H}'$ and $\mathcal{H}\mathcal{H}'$ must have the property that each independent factor $H_{ij}$ occurs at least twice.

2. Every $\mathcal{H}\mathcal{H}'$ must have an independnt term $H_{ij}$ in common, otherwise by independence we will have $\mathbf{E}\mathcal{H}\mathcal{H}' = \mathbf{E}\mathcal{H}\mathbf{E}\mathcal{H}'$.

After eliminating the zero terms identified above, consider the remaining terms. We identify each remaining term with an ordered pair $(\mathcal{H}, \mathcal{H}')$. Define an equivalence relation $\simeq$ on the set of remaining terms via $(\mathcal{H}_1, \mathcal{H}_1') \simeq (\mathcal{H}_2, \mathcal{H}_2')$ if and only if $\rho_1 \sim \rho_2$ and $\rho_1' \sim \rho_2'$. Here $\sim$ refers to the equivalence relation defined in the proof for convergence of expectec moments. Also, $\rho_1, \rho_1', \rho_2, \rho_2'$ are the sequence of digits derived from $\mathcal{H}_1, \mathcal{H}_1', \mathcal{H}_2, \mathcal{H}_2'$ respectively. For example, if $\mathcal{H} = H_{\ell_0\ell_1} H_{\ell_1\ell_2} \cdots H_{\ell_{k-1}\ell_0}$ then $\rho = \ell_0\ell_1 \cdots \ell_{k-1}\ell_0$. Similarly, we define $\omega(\rho, \rho')$ to be the total number of distinct digits in the concatenated sequence of digits $\rho\rho'$.

Just as we have done previously, consider the equivalence classes induced by $\simeq$. Let $\mathcal{P}_{\omega_0}$ be an equivalence class with representative $\rho\rho'$ such that $\omega(\rho, \rho') = \omega_0 < k + 2$. The cardinality of $\mathcal{P}_{\omega_0}$ is

$$\begin{aligned} |\mathcal{P}_{\omega_0}| &= n(n-1)\cdots(n-\omega_0+1) \\ &\leq n^{\omega_0} \\ &< n^{k+2}. \end{aligned} \tag{60}$$

By our assumption of bounded moments we obtain

$$\begin{aligned} \frac{1}{n^{k+2}} \sum_{\rho \in \mathcal{P}_{\omega_0}} (\mathbf{E}\mathcal{H}\mathcal{H}' - \mathbf{E}\mathcal{H}\mathbf{E}\mathcal{H}') \frac{1}{n^{k+2}} \sum_{\rho \in \mathcal{P}_{\omega_0}} O(1) & \\ &= \frac{1}{n^{k+2}} O(n^{\omega_0}) \\ &= O\left(\frac{n^{\omega}}{n^{k+2}}\right) \to 0. \end{aligned} \tag{61}$$

Therefore we are left to consider the equivalence classes $\mathcal{P}_{\omega_1}$ with representative $\rho\rho'$ such that $\omega(\rho, \rho') = \omega_1 \geq k+2$. But there are a total of $2k$ terms in the product $\mathcal{H}\mathcal{H}'$, implying that there are at most $k$ independent factors $H_{ij}$ as we have argued above. Therefore the terms $\mathbf{E}\mathcal{H}\mathcal{H}' - \mathbf{E}\mathcal{H}\mathbf{E}\mathcal{H}'$ corresponding to $\mathcal{P}_{\omega_1}$ are in fact zero terms.

We conclude that

$$\int_{\mathbb{R}} x^k \,\mathrm{d}\mu_n - \mathbf{E} \int_{\mathbb{R}} x^k \,\mathrm{d}\mu_n \xrightarrow{p} 0 \tag{62}$$

□

We are now ready to explicitly prove Wigner's semicircle law. This law roughly states that the distribution of eigenvalues of a Wigner matrix converges to the semicircle distribution[4]. The proof is surprisingly simple, given what we have established so far.

[4] We would like to emphasize once more that that $\mu_n$ is a random measure. Hence, the formal statement of Wigner's theorem might appear odd at first sight – as it single-handedly involves two probabilistic modes of convergence.

**Theorem 3.3.2** (Wigner)**.** The Empirical Spectral Measure (ESM) of eigenvalues of a Wigner random matrix converges weakly to the semicircle distribution, in probability. That is, $\mu_n$ converges weakly to $\sigma$, in probability.

*Proof.* We are essentially trying to prove that for any given $\varepsilon > 0$ and any bounded, continuous test function $f : \mathbb{R} \to \mathbb{R}$

$$\lim_{n\to\infty} \mathbf{P}\left( \left| \int_{\mathbb{R}} f(x)\,\mathrm{d}\mu_n - \int_{\mathbb{R}} f(x)\sigma(x)\mathrm{d}x \right| > \varepsilon \right) = 0. \tag{63}$$

By the Weierstrass Approximation Theorem, given a compact interval $I$ and $\delta > 0$ we can find a polynomial $p_\delta$ such that

$$\sup_{x\in I}\{|f - p_\delta|\} < \delta. \tag{64}$$

By the triangle inequality

$$\begin{aligned}
\left| \int_{\mathbb{R}} f(x)\,\mathrm{d}\mu_n + \int_{\mathbb{R}} f(x)\sigma(x)\,\mathrm{d}x \right| &\leq \left| \int_{\mathbb{R}} f(x) - p_\delta(x)\,\mathrm{d}\mu_n - \int_{\mathbb{R}} (f(x) - p_\delta(x))\sigma(x)\,\mathrm{d}x \right| \\
&\quad + \left| \int_{\mathbb{R}} p_\delta(x)\mathrm{d}\mu_n - \int_{\mathbb{R}} p_\delta(x)\sigma(x)\,\mathrm{d}x \right| \\
&\leq \left| \int_I f(x) - p_\delta(x)\,\mathrm{d}\mu_u \right| + \left| \int_I (f(x) - p_\delta(x))\sigma(x)\,\mathrm{d}x \right| \\
&\quad + \left| \int_{I^c} f(x) - p_\delta(x)\,\mathrm{d}\mu_n \right| + \left| \int_{I^c} (f(x) - p_\delta(x))\sigma\,\mathrm{d}x \right| \\
&\quad + \left| \int_{\mathbb{R}} p_\delta(x)\,\mathrm{d}\mu_n - \int_{\mathbb{R}} p_\delta\sigma(x)\,\mathrm{d}x \right| \\
&< 2\delta + \left| \int_{I^c} f(x) - p_\delta(x)\,\mathrm{d}\mu_n \right| + \left| \int_{I^c} (f(x) - p_\delta)\sigma(x)\mathrm{d}x \right| \\
&\quad + \left| \int_{\mathbb{R}} p_\delta(x)\,\mathrm{d}\mu_n - \int_{\mathbb{R}} p_\delta\sigma(x)\,\mathrm{d}x \right|.
\end{aligned} \tag{65}$$

For a given $\varepsilon > 0$, let $\delta = \varepsilon/4$. Then

$$\begin{aligned}
\mathbf{P}\left(\left|\int_{\mathbb{R}} f(x)\,\mathrm{d}\mu_n - \int_{\mathbb{R}} f(x)\sigma(x)\,\mathrm{d}x\right| \geq \varepsilon\right) \leq\, & \mathbf{P}\left(\left|\int_{I^c} f(x) - p_\delta(x)\,\mathrm{d}\mu_n\right| \geq \varepsilon/2\right) \\
& + \mathbf{P}\left(\left|\int_{I^c} (f(x) - p_\delta(x))\sigma(x)\,\mathrm{d}x\right| \geq \varepsilon/2\right) \\
& + \mathbf{P}\left(\left|\int_{\mathbb{R}} p_\delta(x)\mathrm{d}\mu_n - \int_{\mathbb{R}} p_\delta(x)\sigma(x)\,\mathrm{d}x\right| \geq \varepsilon/2\right) \\
\leq\, & \mathbf{P}\left(\left|\int_{I^c} f(x) - p_\delta(x)\,\mathrm{d}\mu_n\right| \geq \varepsilon/2\right) \\
& + \mathbf{P}\left(\left|\int_{I^c} (f(x) - p_\delta(x))\sigma(x)\,\mathrm{d}x\right| \geq \varepsilon/2\right) \\
& + \mathbf{P}\left(\left|\int_{\mathbb{R}} p_\delta(x)\mathrm{d}\mu_n - \mathbf{E}\int_{\mathbb{R}} p_\delta(x)\,\mathrm{d}\mu_n\right| \geq \varepsilon/4\right) \\
& + \mathbf{P}\left(\left|\mathbf{E}\int_{\mathbb{R}} p_\delta(x)\,\mathrm{d}\mu_n - \int_{\mathbb{R}} p_\delta(x)\sigma(x)\,\mathrm{d}x\right| \geq \varepsilon/4\right).
\end{aligned} \tag{66}$$

In the last expression, we may apply convergence in probability and convergence of expected moments to conclude that the third and fourth term goes to zero as $n \to \infty$. Choose $I = [-m, m]$ for any $m > 4$. Then the second term is identically zero due to the support of $\sigma(x)$. For the first term, we show that it also convergences to zero by the following construction.

Observe that $f$ is bounded. This means there exists $M \in \mathbb{R}$ such that $|f(x)| \leq M$ for all $x \in \mathbb{R}$. Define a polynoomial $q_M(x) = M - p_\delta(x)$ and apply the Markov Inequality to get

$$\begin{aligned}
\mathbf{P}\left(\left|\int_{I^2} f(x) - p_\delta(x)\,\mathrm{d}\mu_n\right| > \varepsilon/2\right) =\, & \mathbf{P}\left(\left|\int_{|x|>m} f(x) - p_\delta(x)\,\mathrm{d}\mu_n\right| > \varepsilon/2\right) \\
\leq\, & \mathbf{P}\left(\left|\int_{|x|>m} q_M(x)\,\mathrm{d}\mu_n\right| > \varepsilon/2\right) \\
\leq\, & \mathbf{P}\left(\int_{|x|>m} |q_M(x)|\,\mathrm{d}\mu_n > \varepsilon/2\right) \\
\leq\, & \frac{1}{\varepsilon/2}\mathbf{E}\int_{|x|>|m} |q_M(x)|\,\mathrm{d}\mu_n.
\end{aligned} \tag{67}$$

Since $q_M$ is a polynomial, it suffices to show that for each $k \in \mathbb{N}$

$$\lim_{n\to\infty} \mathbf{E}\int_{|x|>m} |x|^k\,\mathrm{d}\mu_n = 0. \tag{68}$$

Notice that for each $k \in \mathbb{N}$

$$
\begin{aligned}
\mathbf{E}\int_{|x|>m} |x|^k \,\mathrm{d}\mu_n &\leq \mathbf{E}\int_{|x|>m} \frac{|x|^{2k}}{m^k}\mathrm{d}\mu_n \\
\implies \limsup_{n\to\infty}\mathbf{E}\int_{|x|>m} |x|^k \,\mathrm{d}x &\leq \limsup_{n\to\infty}\mathbf{E}\int_{|x|>m} \frac{|x|^{2k}}{m^k}\,\mathrm{d}\mu_n \\
&\leq \frac{1}{m^k}\limsup_{n\to\infty}\mathbf{E}\int_{\mathbb{R}} x^{2k}\,\mathrm{d}\mu_n \\
&= \frac{1}{m^k}\int_{\mathbb{R}} x^{2k}\sigma(x)\,\mathrm{d}x \\
&= \frac{C_k}{m^k} \\
&\leq \frac{4^k}{m^k} \to 0 \text{ as } k\to\infty.
\end{aligned} \tag{69}
$$

Note that the last equality is argued by Stirling's approximation for the factorial. Therefore the term convergences to zero, since

$$
0 \leq \lim_{n\to\infty}\mathbf{E}\int_{|x|>m} |x|^k \,\mathrm{d}\mu_n \leq \limsup_{n\to\infty}\mathbf{E}\int_{|x|>m} |x|^k \,\mathrm{d}\mu_n = 0. \tag{70}
$$

We conclude that for any bounded, continuous text function $f:\mathbb{R}\to\mathbb{R}$ and $\varepsilon>0$

$$
\lim_{n\to\infty}\mathbf{P}\left(\left|\int_{\mathbb{R}} f(x)\,\mathrm{d}\mu_n - \int_{\mathbb{R}} f(x)\sigma(x)\mathrm{d}x\right| > \varepsilon\right) = 0. \tag{71}
$$

□

We end this section with a some comments. The core of the moment method [5] is to show convergence of expected moments. This is exactly the statement in Lemma 3.2.2. The basic idea is to exploit combinatorial and graph-theoretical properties of the trace. Note that the proof also relies on a relation between Catalan numbers and the moments of the semicircle distribution stated by Lemma 3.2.1.

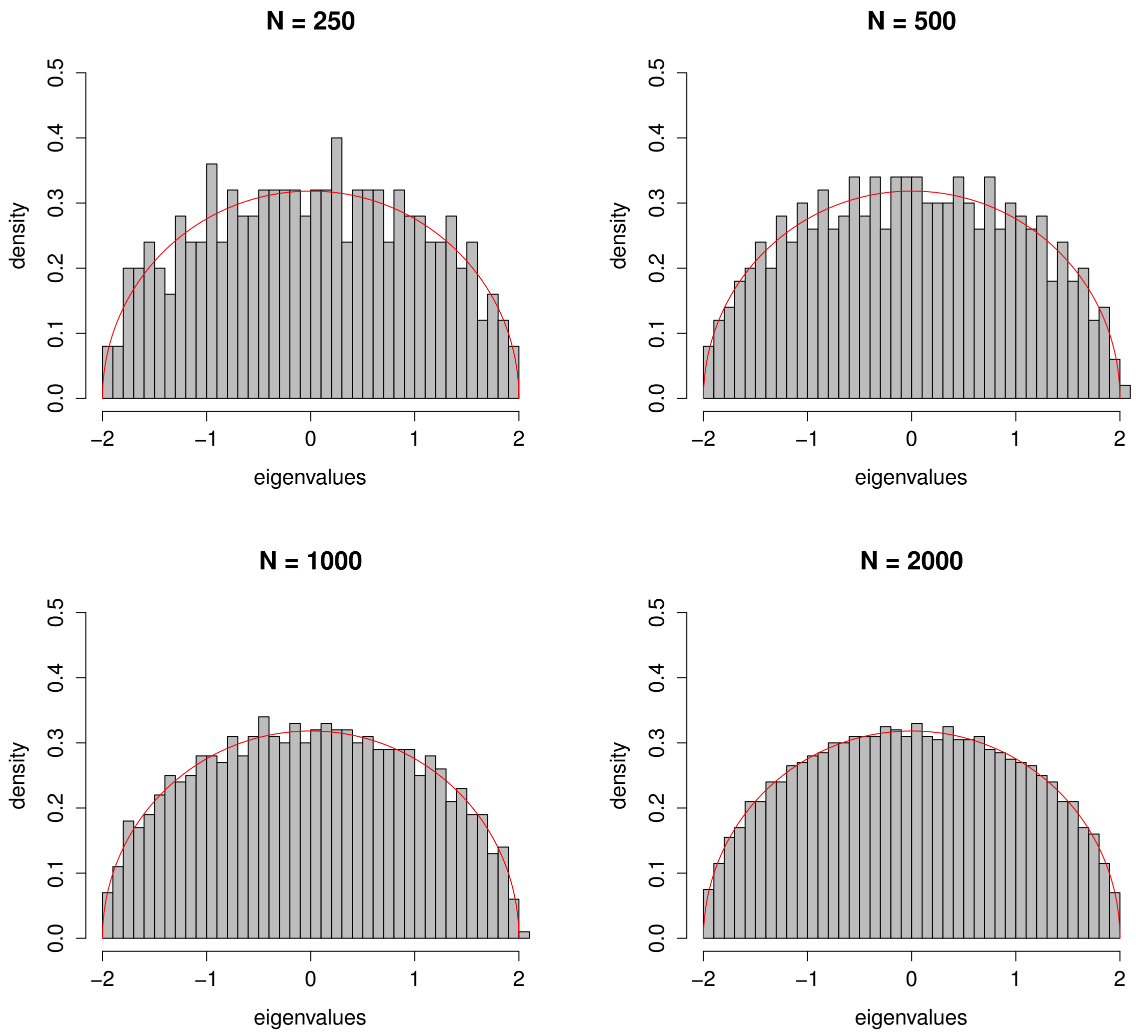
To prove the final result, we also need to control the variance [3] of the ESM. This is the underlying motivation for Lemma 3.3.1, which states that the ESM converges in probability to its expected moments. This statement follows almost immediately by Chebyshev's Inequality and the assumption of bounded moments.

Theorem 3.3.2 states that the ESM converges weakly to the semicircle distribution, in probability. The proof for this uses the result of Lemma 3.2.2 and 3.3.1, the Weierstrass Approximation Theorem and obtaining appropriate bounds for the Catalan numbers via Stirlings approximation for the factorial.

Given certain assumptions, namely that of bounded moments as well as compact support of the limiting density function, one might guess that it is sufficient to *just* prove convergence of expected moments for similar cases. This is true, although false in general – when we drop these assumptions [6]. Therefore, in the later sections we will focus purely on establishing the required convergence of expected moments.

Figure 4 demonstrates sample eigenvalue distributions of a few symmetric $n \times n$ random matrices $H$ with iid entries $H_{ij} \sim \mathrm{N}(0,1)$ for $1 \le i \le j \le n$ and $n = 250, 500, 1000, 2000$. These are a special case of Wigner matrices, simply referred to as Symmetric Wigner Ensembles [15]. For more details on generating random matrices, refer to Appendix A.

Figure 4: Eigenvalues of a Wigner $N \times N$ matrix



# 4 Marchenko-Pastur Law

The Marchenko-Pastur Law is named after Ukrainian mathematicians Vladimir Marchenko and Leonid Pastur [11]. In this section we focus on proving the required convergence of expected moments. Not surprisingly, the proof is ultimately reduced to a combinatorial problem – similar to what we have seen in the semicircle law. We provide some elaboration on the convergence of $k$-th moments when $k = 2, 3$, so as to formulate some idea of the combinatorial problem we need to solve. The proof for the general case $k \in \mathbb{N}$ is adapted from [16], with arguments slightly modifed[5].

## 4.1 Wishart Random Matrices

**Definition 4.1.1.** A Wishart random matrix is defined to be a $p_n \times p_n$ matrix $XX^*$ resulting from the post-multiplication of a $p_n \times n$ random complex-valued matrix $X$ with its adjoint $X^*$. The random matrix $X$ here satisfies the following conditions:

1. $\mathbf{E}X_{ij} = 0$ and $\mathbf{E}|X_{ij}|^2 = 1$ for all $1 \leq i \leq p_n$ and $1 \leq j \leq n$.
2. All entries have bounded moments, which means for each $k \in \mathbb{N}$ there exists $M_k \in \mathbb{R}$ such that $\mathbf{E}|X_{ij}|^k \leq M_k$.
3. The integer sequence $p_n$ has the property that

$$\lim_{n\to\infty} \frac{p_n}{n} = y \in (0, 1]. \tag{72}$$

Similar to the Wigner case, the matrix $XX^*$ has $p_n$ real eigenvalues which we may assume to follow the notation $\lambda_1 \leq \cdots \leq \lambda_{p_n}$. Due to the differing structure of $XX^*$, we require some modification with respect to the scaling factor used in the ESM.

**Definition 4.1.2.** The Empirical Spectral Measure (ESM) of eigenvalues of a Wishart matrix $XX^*$ is defined to be

$$\nu_{p_n}(x) = \frac{1}{p_n} \sum_{i=1}^{p_n} \delta_{\frac{\lambda_i}{n}}(-\infty, x]. \tag{73}$$

where $\delta_{\frac{\lambda_i}{n}}$ refers to the Dirac measure on $\mathbb{R}$ as defined previously.

**Definition 4.1.3.** The Marchenko-Pastur distribution is defined by the probability density function

$$\phi(x) = \frac{1}{2\pi xy} \sqrt{(b-x)(x-a)} \cdot \chi_{[a,b]}, \tag{74}$$

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

[5]For example, Valkó uses constructs a new sequence by 'reflection' in order to count the total number of terms.

## 4.2 Convergence of Expected Moments

**Lemma 4.2.1.** The moments of a random variable $X$ following the Marchenko-Pastur distribution satisfies the following identity

$$\mathbf{E}X^k = \int_{\mathbb{R}} x^k \phi(x)\,\mathrm{d}x = \sum_{r=0}^{k-1} \frac{y^r}{r+1}\binom{k}{r}\binom{k-1}{r}. \tag{75}$$

*Proof.* We evaluate the integral by substitution $z = \frac{x-1-y}{\sqrt{y}}$

$$\begin{aligned}
\mathbf{E}X^k &= \int_{\mathbb{R}} x^k \phi(x)\,\mathrm{d}x \\
&= \int_a^b \frac{x^k}{2\pi x y}\sqrt{(b-x)(x-a)}\,\mathrm{d}x \\
&= \int_a^b \frac{x^k}{2\pi x y}\sqrt{(2\sqrt{y}-(x-1-y))\left((x-1-y)+2\sqrt{y}\right)}\,\mathrm{d}x \\
&= \int_a^b \frac{x^{k-1}}{2\pi\sqrt{y}}\sqrt{\left(2-\frac{x-1-y}{\sqrt{y}}\right)\left(\frac{x-1-y}{\sqrt{y}}+2\right)}\,\mathrm{d}x \\
&= \int_{-2}^{2} \frac{(\sqrt{y}z+1+y)^{k-1}}{2\pi}\sqrt{(2-z)\,(z+2)}\,\mathrm{d}z \\
&= \int_{-2}^{2} \frac{(\sqrt{y}z+1+y)^{k-1}}{2\pi}\sqrt{4-z^2}\,\mathrm{d}z.
\end{aligned} \tag{76}$$

By Lemma 3.2.1, for each $r \in \mathbb{Z}_{\geq 0}$ we have

$$\int_{-2}^{2} \frac{z^{2r}}{2\pi}\sqrt{4-z^2}\,\mathrm{d}z = C_r = \frac{1}{r+1}\binom{2r}{r}. \tag{77}$$

Applying Binomial Expansion gives

$$\int_{-2}^{2} \frac{(\sqrt{y}z+1+y)^{k-1}}{2\pi}\sqrt{4-z^2}\,\mathrm{d}z = \int_{-2}^{2} \frac{\sum_{r=0}^{k-1}\binom{k-1}{r}(\sqrt{y}z)^r(1+y)^{k-1-r}}{2\pi}\sqrt{4-z^2}\,\mathrm{d}z. \tag{78}$$

Note that the terms corresponding to odd values of $r$ in the sum above are zero. Therefore the expression above can be further simplified to

$$\int_{-2}^{2} \frac{\sum_{r=0}^{\lfloor (k-1)/2 \rfloor}\binom{k-1}{2r}(\sqrt{y}z)^{2r}(1+y)^{k-1-2r}}{2\pi}\sqrt{4-z^2}\,\mathrm{d}z. \tag{79}$$

Then by Lemma 3.2.1 we obtain

$$\sum_{r=0}^{\lfloor \frac{k-1}{2} \rfloor} \frac{y^r}{r+1}\binom{2r}{r}\binom{k-1}{2r}(1+y)^{k-1-2r}. \tag{80}$$

Applying Binomial Expansion once more gives

$$\begin{aligned}
&\sum_{r=0}^{\lfloor\frac{k-1}{2}\rfloor}\sum_{s=0}^{k-1-2r}\frac{y^r}{r+1}\binom{2r}{r}\binom{k-1}{2r}\binom{k-1-2r}{s}y^s \\
&=\sum_{r=0}^{\lfloor\frac{k-1}{2}\rfloor}\sum_{s=0}^{k-1-2r}\frac{y^{r+s}}{r+1}\frac{(2r)!(k-1)!(k-1-2r)!}{r!r!(2r)!(k-1-2r)!s!(k-1-2r-s)!} \\
&=\sum_{r=0}^{\lfloor\frac{k-1}{2}\rfloor}\sum_{s=0}^{k-1-2r}y^{r+s}\frac{(k-1)!}{(r+1)!r!s!(k-1-2r-s)!}.
\end{aligned} \tag{81}$$

Let $t = r + s$ then we can write the above expression as

$$\sum_{t=0}^{k-1}\sum_{r=0}^{\min(t,k-1-t)}y^t\frac{(k-1)!}{(r+1)!r!(t-r)!(k-1-r-t)!}. \tag{82}$$

Further simplification and Vandermonde's identity gives us the required result

$$\begin{aligned}
&\sum_{t=0}^{k-1}\sum_{r=0}^{\min(t,k-1-t)}y^t\frac{(k-t)(k-t-1)\cdots(k-r-t)t(t-1)\cdots(t-r+1)}{k(r+1)!r!}\binom{k}{t} \\
&=\sum_{t=0}^{k-1}\sum_{r=0}^{\min(t,k-1-t)}y^t\frac{(k-t)!t!}{k(r+1)!r!(k-1-r-t)!(t-r)!}\binom{k}{t} \\
&=\sum_{t=0}^{k-1}\sum_{r=0}^{\min(t,k-1-t)}\frac{y^t}{k}\binom{t}{r}\binom{k-t}{r+1}\binom{k}{t} \\
&=\sum_{t=0}^{k-1}\frac{y^t}{k}\binom{k}{t}\sum_{r=0}^{\min(t,k-1-t)}\binom{t}{r}\binom{k-t}{r+1} \\
&=\sum_{t=0}^{k-1}\frac{y^t}{k}\binom{k}{t}\binom{k}{t+1} \\
&=\sum_{t=0}^{k-1}\frac{y^t}{t+1}\binom{k-1}{t}\binom{k}{t}.
\end{aligned} \tag{83}$$

□

Similar to the case of the Semicircle law, the result of Lemma 4.2.1 will be useful in proving the required convergence of expected moments.

We take some time to first provide separate proofs for the convergence of the $k$-th moments, where $k = 2, 3$. These proofs give us some insight about the underlying combinatorial problem associated with the general case covering all $k$-th moments.

**Proposition 4.2.2.** Let $XX^*$ be a Wishart random matrix. Then

$$\lim_{n\to\infty} \mathbf{E} \int_{\mathbb{R}} x^2 \, \mathrm{d}\nu_{p_n} = \int_{\mathbb{R}} x^2 \phi(x) \, \mathrm{d}x. \tag{84}$$

*Proof.* The sequence on the left-hand side of the equation can be evalulated as follows

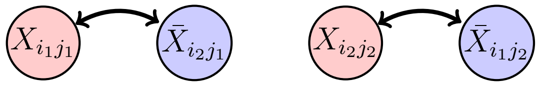
$$\begin{aligned}
\mathbf{E} \int_{\mathbb{R}} x^2 \, \mathrm{d}\nu_{p_n} &= \mathbf{E} \frac{1}{p_n} \sum_{i=1}^{p_n} \left( \frac{\lambda_i}{n} \right)^2 \\
&= \frac{1}{p_n n^2} \mathbf{E} \sum_{i=1}^{p_n} \lambda_i^2 \\
&= \frac{1}{p_n n^2} \mathbf{E} \operatorname{tr}(XX^*)^2 \\
&= \frac{1}{p_n n^2} \sum_{\substack{1 \le i_1, i_2 \le p_n \\ 1 \le j_1, j_2 \le n}} \mathbf{E} X_{i_1 j_1} \bar{X}_{i_2 j_1} X_{i_2 j_2} \bar{X}_{i_1 j_2}.
\end{aligned} \tag{85}$$

Now consider the terms of the sum, $\mathbf{E} X_{i_1 j_1} \bar{X}_{i_2 j_1} X_{i_2 j_2} \bar{X}_{i_1 j_2}$. We are concerned with the number of independent factors $X_{ij}$ in each term. Let $1 \le D \le 4$ denote this number. It is clear that if $D > 2$ then there must exist an independent factor $X_{ij}$ occurring only once. In this case we immediately see that $\mathbf{E} X_{i_1 j_1} \bar{X}_{i_2 j_1} X_{i_2 j_2} \bar{X}_{i_1 j_2} = 0$. On the other hand, since $1 \le i_1, i_2 \le p_n$ and $1 \le j_1 j_2 \le n$, there are $p_n n$ possible terms with $D = 1$. For this case, let $\mathcal{P}_1$ denote the set of all terms with $D = 1$ and apply the assumption of bounded moments to observe that

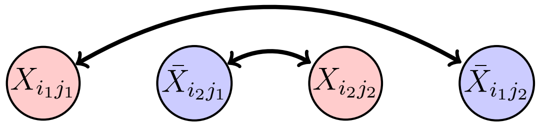
$$\begin{aligned}
\frac{1}{p_n n^2} \sum_{\mathcal{P}_1} \mathbf{E} X_{i_1 j_1} \bar{X}_{i_2 j_1} X_{i_2 j_2} \bar{X}_{i_1 j_2} &= \frac{1}{p_n n^2} \sum_{\mathcal{P}_1} O(1) \\
&= \frac{1}{p_n n^2} O(p_n n) \\
&= O\left( \frac{p_n n}{p_n n^2} \right) \to 0.
\end{aligned} \tag{86}$$

Therefore we only need to consider terms with $D = 2$. From this point onwards we assume that $D = 2$. This means that each independent factor $X_{ij}$ occurs exactly twice; as a pair. The 3 possibles types of pairings are shown in Figure 5.

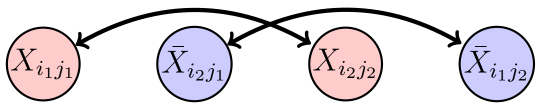
Figure 5: Possible Pairing Types when $k = 2$.




(i) Pairing Type 2-I




(ii) Pairing Type 2-II




(iii) Pairing Type 2-III

We first show that the Pairing of Type 2-III does not apply. Suppose such a pairing exists. This implies that $i_1 = i_2$ and $j_1 = j_2$, and hence $X_{i_1j_1} = X_{i_2j_1} = X_{i_2j_2} = X_{i_1j_2}$. This means that the number of independent terms is $D = 1$, a contradiction to our assumption.

On the other hand, Pairing of Type 2-I only implies that $i_1 = i_2$ which does not necessarily contradict our assumption that $D = 2$. We just need $j_1 \neq j_2$ in this case.

Similary, Pairing of Type 2-II only implies that $j_1 = j_2$, which also does not necessarily contradict our assumption that $D = 2$. We just need $i_1 \neq i_2$ in this case.

To finish this proof, we just need to count the number of terms $\mathbf{E}X_{i_1j_1}\bar{X}_{i_2j_1}X_{i_2j_2}\bar{X}_{i_1j_2}$ with $D = 2$ and either Pairing of Type 2-I or Pairing of Type 2-II. For Pairing of Type 2-I, we have $1 \leq i_1 = i_2 \leq p_n$, $1 \leq j_1, j_2 \leq n$ and $j_1 \neq j_2$. The number of such terms is $p_n n^2$. For Pairing of Type 2-II, we have $1 \leq i_1, i_2 \leq p_n$, $i_1 \neq i_2$ and $1 \leq j_1 = j_2 \leq n$. The number of such terms is $p_n^2 n = p_n n^2 y_n^1$, where $y_n = p_n/n \to y$.

Note that since $D = 2$ and the pairings are as described, we have $\mathbf{E}X_{i_1j_1}\bar{X}_{i_2j_1}X_{i_2j_2}\bar{X}_{i_1j_2} = 1$.

Therefore

$$
\begin{aligned}
\frac{1}{p_n n^2} \sum_{\substack{1\le i_1,i_2\le p_n \\ 1\le j_1,j_2\le n}} \mathbf{E} X_{i_1j_1}\bar{X}_{i_2j_1}X_{i_2j_2}\bar{X}_{i_1j_2} &= \frac{1}{p_n n^2}\left(p_n n^2 + p_n n^2 y_n^1\right)\cdot 1 \\
&= \left(\frac{p_n n^2}{p_n n^2} + \frac{p_n n^2 y_n^1}{p_n n^2}\right) \\
&\to 1 + y^1 \qquad (87) \\
&= \frac{y^0}{0+1}\binom{2}{0}\binom{1}{0} + \frac{y^1}{1+1}\binom{2}{1}\binom{1}{1} \\
&= \sum_{r=0}^{1} \frac{y^r}{r+1}\binom{k}{r}\binom{k-1}{r}.
\end{aligned}
$$

□

We now state and prove the result for the case $k = 3$. The proof follows the same idea as the previous case.

**Proposition 4.2.3.** Let $XX^*$ be a Wishart random matrix. Then

$$
\lim_{n\to\infty} \mathbf{E}\int_{\mathbb{R}} x^3 \,\mathrm{d}\nu_{p_n} = \int_{\mathbb{R}} x^3 \phi(x)\,\mathrm{d}x \qquad (88)
$$

*Proof.* We obtain

$$
\begin{aligned}
\mathbf{E}\int_{\mathbb{R}} x^3 \,\mathrm{d}\nu_{p_n} &= \mathbf{E}\frac{1}{p_n}\sum_{i=1}^{p_n}\left(\frac{\lambda_i}{n}\right)^3 \\
&= \frac{1}{p_n n^3}\mathbf{E}\sum_{i=1}^{p_n}\lambda_i^3 \\
&= \frac{1}{p_n n^3}\mathbf{E}\mathrm{tr}(XX^*)^3 \\
&= \frac{1}{p_n n^3}\sum_{\substack{1\le i_1,i_2,i_3\le p_n \\ 1\le j_1,j_2,j_3\le n}} \mathbf{E}X_{i_1j_1}\bar{X}_{i_2j_1}X_{i_2j_2}\bar{X}_{i_3j_2}X_{i_2j_3}\bar{X}_{i_1j_2}.
\end{aligned} \qquad (89)
$$

Considering the terms of the sum, $\mathbf{E}X_{i_1j_1}\bar{X}_{i_2j_1}X_{i_2j_2}\bar{X}_{i_1j_2}$, we let $1 \le D \le 6$ be the number of independent factors $X_{ij}$.

If $D > 3$ then there must exist an independent factor $X_{ij}$ occurring exactly once, in which we have $\mathbf{E}X_{i_1j_1}\bar{X}_{i_2j_1}X_{i_2j_2}\bar{X}_{i_1j_2} = 0$.

If $D = 1$ then there are $p_n n$ of such terms. If $D = 2$, then # (of distinct $i$'s) + # (of distinct $j$'s) cannot be more than $D + 1 = 3$. Hence the number of such terms is either $O(p_n n^2)$ or $O(p_n^2 n)$. For $D = 1, 2$ it is clear that the contribution of these terms tends to zero as $n \to \infty$.

We conclude that we only need to consider terms with $D = 3$. This means that every independent factor $X_{ij}$ in $\mathbf{E}X_{i_1j_1}\bar{X}_{i_2j_1}X_{i_2j_2}\bar{X}_{i_3j_2}X_{i_3j_3}\bar{X}_{i_1j_3}$ occurs exactly twice; as a pair. There are a total of 15 distinct pairings in this case. They are shown in Figure 6.

Figure 6: Possible Pairing Types when $k = 3$

(i) Pairing Type 3-I; (ii) Pairing Type 3-II; (iii) Pairing Type 3-III; (iv) Pairing Type 3-IV; (v) Pairing Type 3-V; (vi) Pairing Type 3-VI; (vii) Pairing Type 3-VII; (viii) Pairing Type 3-VIII; (ix) Pairing Type 3-IX; (x) Pairing Type 3-X; (xi) Pairing Type 3-XI; (xii) Pairing Type 3-XII; (xiii) Pairing Type 3-XIII; (xiv) Pairing Type 3-XIV; (xv) Pairing Type 3-XV

Direct verification tells us that only Pairing Types 3-I, 3-II, 3-III, 3-IV and 3-V apply when $D = 3$. Again, we just need to count the number of such terms $\mathbf{E}X_{i_1j_1}\bar{X}_{i_2j_1}X_{i_2j_2}\bar{X}_{i_3j_2}X_{i_3j_3}\bar{X}_{i_1j_3}$. Let $y_n = p_n/n$. For Pairing Type 3-I, there are $p_nn^3y_n^0$ terms. For Pairing Type 3-II and III, there are $p_nn^3y_n^1$ terms. For Pairing Type 3-IV, there are $p_nn^3y_n^2$ terms. Finally for Pairing Type 3-V, there are $p_nn^3y_n^1$ terms.

Observe that since $D = 3$ and we must obey such pairings, $\mathbf{E}X_{i_1j_1}\bar{X}_{i_2j_1}X_{i_2j_2}\bar{X}_{i_3j_2}X_{i_3j_3}\bar{X}_{i_1j_3} = 1$

for these given terms. Therefore

$$\frac{1}{p_n n^3} \sum_{\substack{1 \le i_1, i_2, i_3 \le p_n \\ 1 \le j_1, j_2, j_3 \le n}} \mathbf{E} X_{i_1 j_1} \bar{X}_{i_2 j_1} \cdots X_{i_2 j_3} \bar{X}_{i_1 j_2} = \frac{1}{p_n n^2} (p_n n^3 y_n^0 + 2 p_n n^3 y_n^1 + p_n n^3 y_n^2 + p_n n^3 y_n^1) \cdot 1$$

$$= \frac{p_n n^3 y_n^0}{p_n n^3} + \frac{2 p_n n^3 y_n^1}{p_n n^3} + \frac{p_n n^3 y_n^2}{p_n n^3} + \frac{p_n n^3 y_n^1}{p_n n^3}$$

$$1 + 3y + y^2 \to$$

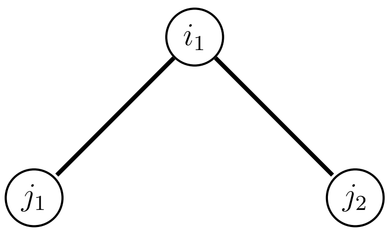
$$= \sum_{r=0}^{2} \frac{y^r}{r+1} \binom{k}{r} \binom{k-1}{r}. \tag{90}$$

□

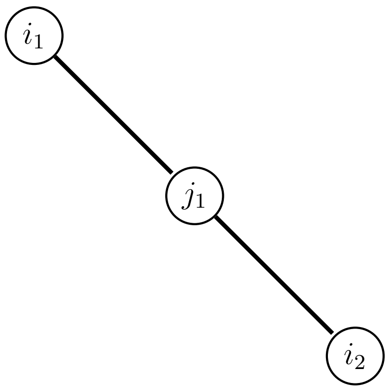
*Remark* 4.2.4. One important property we have noticed for $k = 2, 3$ is that the significant terms are exactly those whose with $k$ independent factors $X_{ij}$ each occurring as noncrossing conjugate pairs. In terms of the Figure 5 and 6 presented above; note that every red node must pair with a blue node (conjugate pairs) in order to be considered a valid pairing. This property is true for the general case $k \in \mathbb{N}$ as well, and will be established later on.

*Remark* 4.2.5. The proof of the case $k = 2, 3$ allows us to avoid entirely a graph-theoretic argument used in the proof of general case $k \in \mathbb{N}$. We show *some* of the corresponding walk along the graph (tree) $\mathcal{G}$ in Figure 7 and 8. Note that we always start and end at the same vertex $i_1$.

Figure 7: Possible Walk Types when $k = 2$.

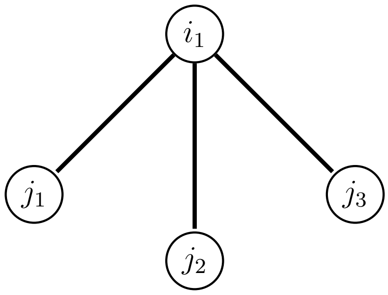



(i) Walk Type 2-I: $(i_1,\, j_1,\, i_1,\, j_2, i_1)$. Note that $i_1 = i_2$.

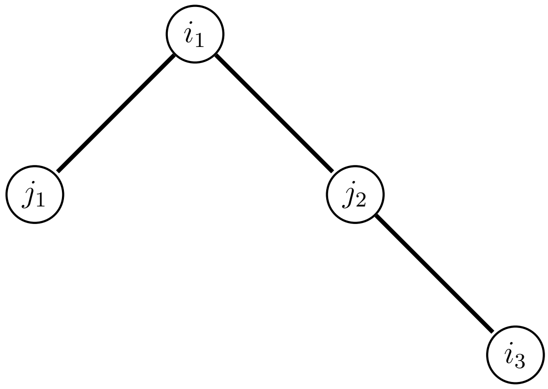



(ii) Walk Type 2-II: $(i_1,\, j_1,\, i_2,\, j_1,\, i_1)$. Note that $j_1 = j_2$.

Figure 8: Possible Walk Types when $k = 3$.




(i) Walk Type 3-I: $(i_1,\, j_1,\, i_1,\, j_2,\, i_1,\, j_3,\, i_1)$. Note that $i_1 = i_2 = i_3$.




(ii) Walk Type 3-II: $(i_1,\, j_1,\, i_1,\, j_2,\, i_3,\, j_2,\, i_1)$. Note that $i_1 = i_2$ and $j_2 = j_3$.

We now proceed with a proof for the general case.

**Lemma 4.2.6.** Fix any $k \in \mathbb{N}$ and let $XX^*$ be a Wishart matrix as defined previously. Then

$$\lim_{n\to\infty} \mathbf{E} \int_{\mathbb{R}} x^k \,\mathrm{d}\nu_{p_n} = \int_{\mathbb{R}} x^k \phi(x)\,\mathrm{d}x \tag{91}$$

*Proof.* First express the sequence on the left-hand side as follows

$$\begin{aligned}
\mathbf{E} \int_{\mathbb{R}} x^k \,\mathrm{d}\nu_{p_n} &= \mathbf{E} \frac{1}{p_n} \sum_{i=1}^{p_n} \left(\frac{\lambda_i}{n}\right)^k \\
&= \frac{1}{p_n n^k} \mathbf{E} \sum_{i=1}^{p_n} \lambda_i^k \\
&= \frac{1}{p_n n^k} \mathbf{E}\, \mathrm{tr}(XX^*)^k \\
&= \frac{1}{p_n n^k} \sum_{\substack{1 \le i_1,\dots,i_k \le p_n \\ 1 \le j_1,\dots,j_k \le n}} \mathbf{E} X_{i_1 j_1} \bar{X}_{i_2 j_1} X_{i_2 j_2} \bar{X}_{i_3 j_2} \cdots X_{i_k j_k} \bar{X}_{i_1 j_k}
\end{aligned} \tag{92}$$

For each term $\mathbf{E} X_{i_1 j_1} \bar{X}_{i_2 j_1} \cdots X_{i_k j_k} \bar{X}_{i_1 j_k}$ in the sum, we construct a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. The set of vertices in $\mathcal{G}$ is given by $\mathcal{V} = \mathcal{I} \sqcup \mathcal{J}$, where $\mathcal{I} = \{i_1, \dots, i_k\}$ and $\mathcal{J} = \{j_1, \dots, j_k\}$.

Note that any two $i \in \mathcal{I}$ and $j \in \mathcal{J}$ are always considered to be different vertices, although it is certainly possible to have $i = i'$ for some $i, i' \in \mathcal{I}$ or $j = j'$ for some $j, j' \in \mathcal{J}$. Hence $|\mathcal{V}| \leq 2k$. The set of edges in $\mathcal{G}$ is given by $\mathcal{E} = \{e(X_{i_1j_1}), e(\bar{X}_{i_2j_1}), \ldots, e(X_{i_kj_k}), e(\bar{X}_{i_1j_k})\}$. Similarly $|\mathcal{E}| \leq 2k$.

We take some time to elaborate on the construction of the set of edges $\mathcal{E}$. Here $e(\cdot)$ is a function which defines a directed edge based on a given factor $X_{ij}$. We use ordered pair notation to represent directed edges. For example, $(u, v)$ represents a directed edge from vertex $u$ to vertex $v$. We define $e(\cdot)$ as follows

$$\vec{e}(*) = \begin{cases} (i, j) & \text{if } * = X_{ij}, \text{ and} \\ (j, i) & \text{if } * = \bar{X}_{ij}. \end{cases} \tag{93}$$

Given $\mathcal{G}$ and its corresponding term $\mathbf{E}X_{i_1j_1}\bar{X}_{i_2j_1} \cdots X_{i_kj_k}\bar{X}_{i_1j_k}$, we can define a walk along the graph itself. This walk is done by visiting the directed edges in the following order

$$\vec{e}(X_{i_1j_1}), \vec{e}(\bar{X}_{i_2j_1}), \ldots, \vec{e}(X_{i_kj_k}), \vec{e}(\bar{X}_{i_1j_k}). \tag{94}$$

The order of vertices visited is therefore

$$i_1 \to j_1 \to i_2 \to j_1 \to \cdots \to i_k \to j_k \to i_1. \tag{95}$$

Much can be said about the graph $\mathcal{G}$, in particular $\mathcal{G}$ is connected since there exists a path which reaches all vertices as shown above. Furthermore $\mathcal{G}$ is clearly a bipartite graph, as observed by the disjoint vertex sets $\mathcal{I}$ and $\mathcal{J}$.

We return to to the sum of terms $\mathbf{E}X_{i_1j_1}\bar{X}_{i_2j_1} \cdots X_{i_kj_k}\bar{X}_{i_1j_k}$ and eliminate the zero terms. To do this, let $1 \leq D \leq 2k$ be the number of independent factors in the term $\mathbf{E}X_{i_1j_1}\bar{X}_{i_2j_1} \cdots X_{i_kj_k}\bar{X}_{i_1j_k}$. We claim that we only need to examine terms with $D = k + 1$. We justify this by ruling out the following cases.

Case #1: $D > k$. Since each term has a total of $2k$ factors, this implies that there must be exist some factor $X_{ij}$ occurring exactly once. As $\mathbf{E}X_{ij} = 0$ for all $1 \leq i \leq p_n$ and $1 \leq j \leq n$, we conclude that for terms with $D > k$ independent factors, $\mathbf{E}X_{i_1j_1}\bar{X}_{i_2j_1} \cdots X_{i_kj_k}\bar{X}_{i_0j_k} = 0$.

Case #2: $D < k$. Then consider $|\mathcal{V}| = |\mathcal{I}| + |\mathcal{J}|$. Since the graph is connected, it is easy to see that $|\mathcal{V}| \leq D + 1$. Let $|\mathcal{I}| = a$, $|\mathcal{J}| = b$ and let $\mathcal{P}_D$ denote the set of all terms with $D$ independent factors. We count the number of such terms. As $1 \leq i \leq p_n$, $1 \leq j \leq n$, and $a+b \leq D+1 < k+1$ we have

$$|\mathcal{P}_D| = O(p_n^a n^b). \tag{96}$$

By assumption of bounded moments we have

$$\begin{aligned} \frac{1}{p_n n^k} \sum_{\mathcal{P}_D} \mathbf{E}X_{i_1j_1}\bar{X}_{i_2j_1} \cdots X_{i_kj_k}\bar{X}_{i_1j_k} &= \frac{1}{p_n n^k} \sum_{\mathcal{P}_D} O(1) \\ &= \frac{1}{p_n n^k} O(p_n^a n^b) \\ &= O\left(\frac{p_n^a n^b}{p_n n^k}\right) \to 0. \end{aligned} \tag{97}$$

We conclude that $D = k$ and $|\mathcal{V}| = k + 1$. Note that $D = k$ means that if we disregard the direction, there are exactly $k$ distinct edges. This means every independent factor $X_{ij}$ occurs exactly twice. Therefore if we derive an undirected graph $\mathcal{G}'$ with the same vertices as $\mathcal{G}$ but using the $k$ distinct undirected edges as previously described instead, then $\mathcal{G}'$ is a tree.

The fact that $\mathcal{G}'$ is a tree tells us more about $\mathcal{G}$. More specifically it allows us to find out more about the corresponding term $\mathbf{E}X_{i_1j_1}\bar{X}_{i_2j_1}\cdots X_{i_kj_k}\bar{X}_{i_1j_k}$. This is done by considering the walk defined on $\mathcal{G}$.

Firstly, we claim that for every directed edge $(i, j)$ in $\mathcal{G}$ there exists another directed edge of the form $(j, i)$. Note that this means both directed edges $(i, j)$ and $(j, i)$ occur exactly once in the set $\mathcal{E}$. In other words, if $X_{ij}$ is a factor occurring in the term $\mathbf{E}X_{i_1j_1}\bar{X}_{i_2j_1}\cdots X_{i_kj_k}\bar{X}_{i_1j_k}$ then $\bar{X}_{ij}$ must occur exactly once as well. We justify this as follows. Suppose without loss of generality that $X_{ij}$ occurs twice. By considering the walk defined on $\mathcal{G}$ we have the following subpath

$$i \to j \to \cdots \to i \to j. \tag{98}$$

This is a contradiction, as $\mathcal{G}'$ is a tree and hence has no simple cycles. On the other hand, assuming $X_{ij}$ occurs twice implies that we may find a simple cycle from the subpath above. Note that there is no edge $(j, i)$ under such assumptions.

Secondly, since it is now clear that each $X_{ij}$'s repeated paired with $\bar{X}_{ij}$, we claim that this pairing is noncrossing with respect to the order of appearance along the term $\mathbf{E}X_{i_1j_1}\bar{X}_{i_2j_1}\cdots X_{i_kj_k}\bar{X}_{i_1j_k}$. Again, suppose not. Then we can find some factors appearing in the following order

$$X_{ij}\ldots X_{i'j'}\ldots \bar{X}_{ij}\ldots \bar{X}_{i'j'}. \tag{99}$$

By considering the walk defined on $\mathcal{G}$ we have the following subpath

$$i \to j \to \ldots i' \to j' \to \cdots \to j \to i \to \cdots \to j' \to i'. \tag{100}$$

Again, this contradicts the property that $\mathcal{G}'$ is a tree as we have found at least one simple cycle.

Hence for the significant terms $\mathbf{E}X_{i_1j_1}\bar{X}_{i_2j_1}\cdots X_{i_kj_k}\bar{X}_{i_1j_k}$ of the sum, every independent factor $X_{ij}$ occurs twice, in the form of noncrossing conjugate pairs.

Now, for each of these terms $\mathbf{E}X_{i_1j_1}\bar{X}_{i_2j_1}\cdots X_{i_kj_k}\bar{X}_{i_1j_k}$ we may define a sequence of digits $\rho = i_1j_1i_2j_2\cdots i_kj_ki_1$. Note that this sequence corresponds to the order of vertices visited due to the walk on $\mathcal{G}$. Using the same equivalence relation $\sim$ in the proof for Wigner's Semicircle Law, we partition the set of all such sequences $\rho$ into equivalence classes $\tilde{\rho}$. Furthermore, for each equivalence class $\tilde{\rho}$ we have

$$\begin{aligned} |\tilde{\rho}| &= O(p_n^{r+1}n^{k-r}) \\ &= O(p_n n^k y_n^r), \end{aligned} \tag{101}$$

where $|\mathcal{I}| = r + 1$, $|\mathcal{J}| = k - r$ correspond to the number of unique $i$'s and $j$'s in the equivalence class, and $y_n = \frac{p_n}{n}$. Hence we need to show that for each $0 \le r \le k - 1$ the number of such equivalence classes is

$$\frac{1}{r+1}\binom{k}{r}\binom{k-1}{r}. \tag{102}$$

To do so, for each sequence $\rho$ (and corresponding walk on a tree $\mathcal{G}$) define a sequence $\xi = \xi_1\xi_2\cdots\xi_{2k}$ based on the following rule

1. $\xi_1 = 0$.
2. For $1 < m \leq 2k$, we consider the $m$-th edge visited during the walk.
   (a) (*First Arrival*). If the edge brings us to an $i$ vertex which we have not visited yet, then let $\xi_m = \xi_{m-1} + 1$
   (b) (*Last Departure*). If the edge brings us away from an $i$ vertex which we will never visit again, then let $\xi_m = \xi_{m-1} - 1$.
   (c) If neither condition (a) nor (b) is satisfied, let $\xi_m = \xi_{m-1}$.

We note a few important properties of the sequence $\xi$. Firstly, $\xi_m$ increases by one compared to $\xi_{m-1}$ only if $m$ is even. Secondly, $\xi_m$ decreases by one compared to $\xi_{m-1}$ only if $m$ is odd. This is because the edges visited during even steps correspond to some $X_{ij}$, whereas the edges visited during odd steps correspond to some $\bar{X}_{ij}$.

Another important property is $\xi_1 = \xi_{2k} = 0$ and $\xi_m \geq 0$ for all $1 \leq m \leq 2k$. This is due to the fact that every edge is visited once in each opposite direction and that $\mathcal{G}$ is a tree.

Let $\Xi$ be the set of all such sequences $\xi$. Let $S$ be the set of all sequences $s = s_1s_2\cdots s_{2k}$ defined by $s_1 = 0$, choosing $r$ of the $k$ possible elements from the set $\{s_2, s_4, \ldots, s_2\}$ to be $+1$ more than their predecessor and $r$ out of the $k-1$ possible elements from the set $\{s_3, s_5, \ldots, s_{2k-1}\}$ to be $-1$ less than their predecessor. Then $s_{2k} = 0$, and $\Xi \subset S$. Further more is is clear that

$$|S| = \binom{k}{r}\binom{k-1}{r}. \tag{103}$$

Note that the elements which belong to $S$ but not $\Xi$ are therefore those which at least one negative digit. Define a truncation $\tau(\xi)$ to be the a subsequence of $\xi \in \Xi$ such that only terms which do not differ from their predecessor are removed. More generally we may evaluate $\tau(s)$ for any $s \in S$. We note that the resultant length of $\tau(s)$ must be exactly $2r$. Also, $\tau(\xi)$ is necesarily a Dyck path, although $\tau(s)$ might not. We define an equivalence relation $\approx$ on the set $S$ via $s \approx s'$ if and only if $\tau(s) = \tau(s')$. Note that $\approx$ partitions the set $S$ into equal cells.

Now, any given equivalence class represented by some $\tau(s)$ corresponds to a unique 2-dimensional lattice walk from the points $(0,0)$ to $(r,r)$ in $\mathbb{Z}^2$. On the other hand, any given equivalence class represented by some $\tau(\xi)$ corresonds to a unique Dyck path of order $r$ (also equivalent to a unique 2-dimensional lattic walk from the points $(0,0)$ to $(r,r)$ in $\mathbb{Z}^2$). Using Lemma 2.2.10, we calculate the following

$$\frac{\#\text{of Dyck Paths}}{\#\text{of Lattice Walks}} = \frac{\frac{1}{r+1}\binom{2r}{r}}{\binom{2r}{r}} = \frac{1}{r+1}. \tag{104}$$

which implies that the number of equivalence classes in $\Xi$ is

$$\frac{1}{r+1}\binom{k}{r}\binom{k-1}{r}. \tag{105}$$

This completes the proof since

$$
\begin{aligned}
\frac{1}{p_n n^k} \sum_{\substack{1 \le i_1, \ldots, i_k \le p_n \\ 1 \le j_1, \ldots, j_k \le n}} \mathbf{E} X_{i_1 j_1} \bar{X}_{i_2 j_1} \cdots X_{i_k j_k} \bar{X}_{i_1 j_k} &= \frac{1}{p_n n^k} \sum_{\mathcal{P}_D,\, D=k} \mathbf{E} X_{i_1 j_1} \bar{X}_{i_2 j_1} \cdots X_{i_k j_k} \bar{X}_{i_1 j_k} \\
&= \frac{1}{p_n n^k} \sum_{r=0}^{k-1} \frac{1}{1+r} \binom{k}{r} \binom{k-1}{r} O(p_n n^k y_n^r) \\
&= \sum_{r=0}^{k-1} \frac{y^r}{1+r} \binom{k}{r} \binom{k-1}{r} (1 + o(1)) \\
&\to \sum_{r=0}^{k-1} \frac{y^r}{1+r} \binom{k}{r} \binom{k-1}{r}.
\end{aligned} \tag{106}
$$

□

*Remark* 4.2.7. One might notice that the moment method proof above suffered a great deal of complication. Some of this is attributed to the presence of the arbitrary constant $y \in (0, 1]$, or equivalently the sequence $p_n$. For the last limiting density in the Section 5, we have $p_n = n$ and hence $y = 1$. This allows us to fully exploit some combinatorial properties of the random matrix of interest, and hopefully for some – make the counting simpler.

We end this section in a similar fashion as the semicircle law. Figure 9 below shows sample eigenvalues of a $P \times P$ Wishart random matrix $XX^*$. The underlying $P \times N$ matrix $X$ is real-valued with iid entries $X_{ij} \sim N(0, 1)$ for all $1 \le i \le P$, $1 \le j \le N$ and $P = \lfloor \frac{N}{2} \rfloor$. It is clear that in this case we have

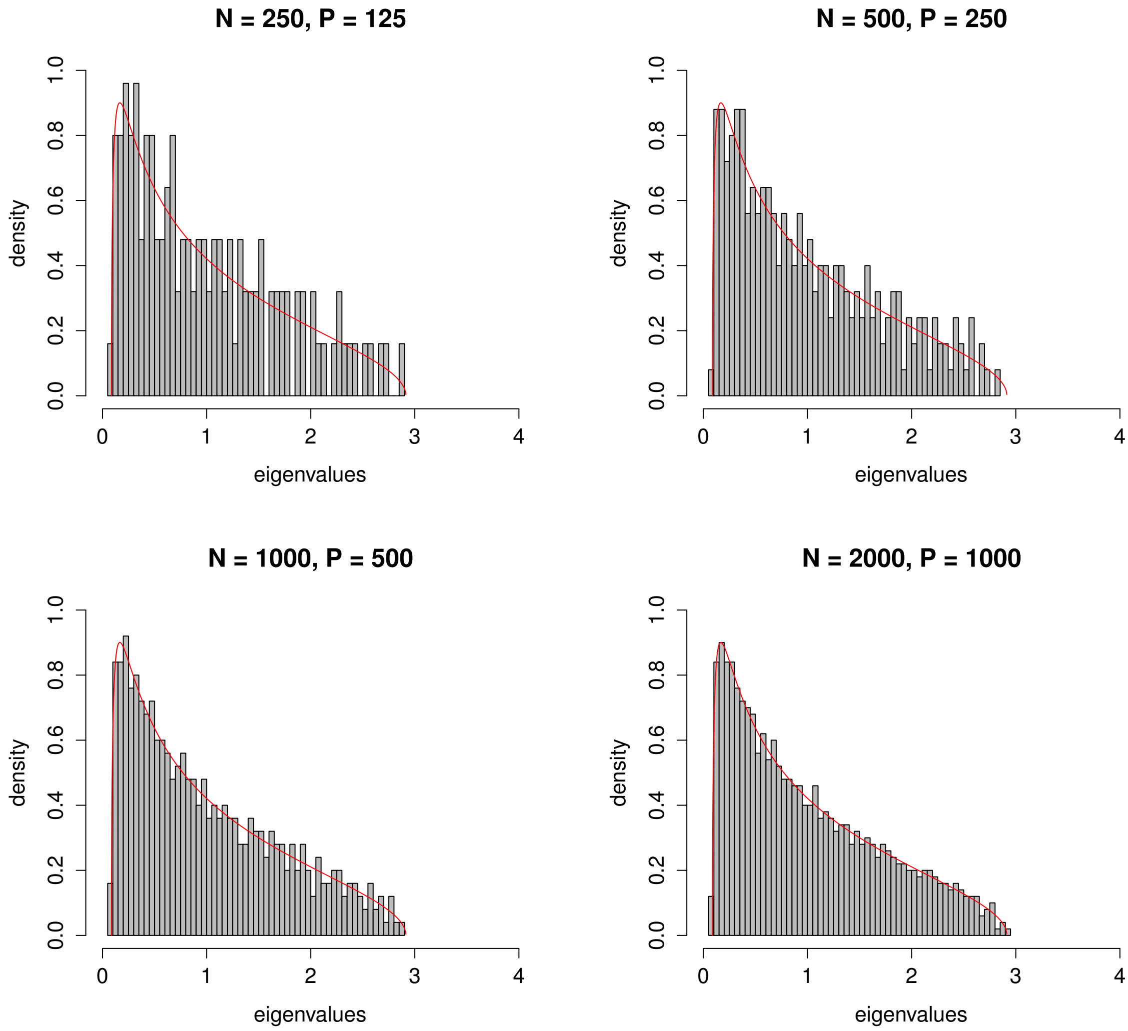
$$
\lim_{N \to \infty} \frac{P}{N} = y = \frac{1}{2}, \tag{107}
$$

and the limiting density is

$$
\phi(x) = \frac{1}{\pi x} \sqrt{(b-x)(x-a)} \chi_{[a,b]}, \tag{108}
$$

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

Figure 9: Eigenvalues of a Wishart $P \times P$ matrix.



# 5 Singular Values of Matrix Powers

## 5.1 Random Matrices of Interest

We shift our attention to a particular generalisation of the Marchenko-Pastur Law. The idea of such a random matrix ensemble is taken from [2]. Define the random matrices of interest for this section as follows

**Definition 5.1.1.** The random matrix of interest is a $n \times n$ matrix of the form $X^{(m)}X^{*(m)}$ where $m \in \mathbb{N}$ and $X$ is a random $n \times n$ matrix satisfying the following conditions

1. $\mathbf{E}X_{ij} = 0$ and $\mathbf{E}|X_{ij}|^2 = 1$ for all $1 \leq i, j \leq n$.
2. All entries have bounded moments, which means for each $k \in \mathbb{N}$ there exists $M_k \in \mathbb{R}$ such that $\mathbf{E}|X_{ij}|^k \leq M_k$

*Remark* 5.1.2. To avoid ambiguity, we define

$$X^{(m)}X^{*(m)} = \underbrace{XX\cdots X}_{m \text{ times}} \cdot \underbrace{X^*X^*\cdots X^*}_{m \text{ times}}. \tag{109}$$

*Remark* 5.1.3. The order of matrix multiplication matters. If we consider $(XX^*)^m$ instead, this would result in a different matrix and hence different limiting distribution.

*Remark* 5.1.4. The matrix $X$ is the same as that stated in Section 4. In particular we simply set $p_n = n$.

Similar to before, $X^{(m)}X^{*(m)}$ has $n$ real eigenvalues which we may assume to follow the notation $\lambda_1 \leq \cdots \leq \lambda_n$. To obtain meaningful statistics as before, we apply appropriate modification to the scaling factor in the ESM.

**Definition 5.1.5.** The Empirical Spectral Measure (ESM) of eigenvalues of the random matrix of interest $X^{(m)}X^{*(m)}$ is defined to be

$$\eta_n^{(m)}(x) = \frac{1}{n}\sum_{i=1}^{p_n} \delta_{\frac{\lambda_i}{n^m}}(-\infty, x]. \tag{110}$$

where $\delta_{\frac{\lambda_i}{n^m}}$ refers to the Dirac measure on $\mathbb{R}$ as defined previously.

**Definition 5.1.6.** Given $m \in \mathbb{N}$, the limiting density has the property that its moments are the Fuss Catalan numbers with parameter $m$. Formally, the random variable of such a density satifies

$$\mathbf{E}X^k = \frac{1}{mk+1}\binom{mk+k}{k} = C_{m,k}. \tag{111}$$

## 5.2 Convergence of Expected Moments

We present a proof for the convergence of expected moments. The proof presented has significant differences from our main Reference [2]. In particular, we devise a slightly different[6] combinatorial problem to solve.

**Lemma 5.2.1.** Fix any $k, m \in \mathbb{N}$ and let $X^{(m)}X^{*(m)}$ be the random matrix of interest as defined previously. Then

$$\lim_{n\to\infty} \mathbf{E} \int_{\mathbb{R}} x^k \mathrm{d}\eta_n^{(m)} = \frac{1}{mk+1}\binom{mk+k}{k}. \tag{112}$$

*Proof.* We have

$$\begin{aligned}
\mathbf{E} \int_{\mathbb{R}} x^k \mathrm{d}\eta_n^{(m)} &= \mathbf{E}\frac{1}{n}\sum_{i=1}^{n}\left(\frac{\lambda_i}{n^m}\right)^k \\
&= \frac{1}{n^{mk+1}}\mathbf{E}\sum_{i=1}^{n}\lambda_i^k \\
&= \frac{1}{n^{mk+1}}\mathbf{E}\,\mathrm{tr}\left(X^{(m)}X^{*(m)}\right)^k \\
&= \frac{1}{n^{mk+1}}\sum\left(\mathbf{E}\prod_{j=0}^{2mk-1} X^{\epsilon(j)}_{\ell_j\ell_{j+1}}\right).
\end{aligned} \tag{113}$$

where $1 \leq \ell_j, \ell_{j+1} \leq n$, $\ell_0 = \ell_{2mk}$ and $\epsilon(j)$ denotes a 'spin' variable

$$X^{\epsilon(j)}_{\ell_j\ell_{j+1}} = \begin{cases} X_{\ell_j\ell_{j+1}} & \text{if } j \equiv 0, 1, \ldots, m-1 \mod 2m, \\ \bar{X}_{\ell_{j+1}\ell_j} = X_{\ell_j\ell_{j+1}} & \text{if } j \equiv m, m+1, \ldots, 2m-1 \mod 2m. \end{cases} \tag{114}$$

In other words, $\epsilon(j)$ is a way of shortening notation for

$$\mathbf{E}\,\underbrace{\underbrace{X_{|j=1}\cdots X_{|j=m-1}}_{m \text{ terms}}\underbrace{\bar{X}_{|j=m}\cdots\bar{X}_{|j=2m-1}}_{m \text{ terms}}\cdots\cdots\underbrace{X_{|j=2mk-2m}\cdots X_{|j=2mk-m-1}}_{m \text{ terms}}\underbrace{\bar{X}_{|j=2mk-m}\cdots\bar{X}_{|j=2mk-1}}_{m \text{ terms}}}_{m\times 2k \text{ terms}}. \tag{115}$$

At this point we construct a directed graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ for each term in the sum such that

$$\mathcal{V} = \{\ell_0, \ell_1, \ldots, \ell_{2mk-1}, \ell_{2mk} = \ell_0\}, \tag{116}$$

$$\mathcal{E} = \left\{\vec{e}(X^{\epsilon(0)}_{\ell_0\ell_1}), \vec{e}(X^{\epsilon(1)}_{\ell_1\ell_2}), \ldots, \vec{e}(X^{\epsilon(2mk-2)}_{\ell_{2mk-2}\ell_{2mk-1}}), \vec{e}(X^{\epsilon(2mk-1)}_{\ell_{2mk-1}\ell_{2mk}})\right\}. \tag{117}$$

Here $\vec{e}(\cdot)$ refers to the edge constructing function as defined previously.

We also construct an undirected graph $\mathcal{G}' = (\mathcal{V}, \mathcal{E}')$, where $\mathcal{E}'$ consists of all undirected edges $\{u, v\}$ for which there exists a corresponding directed edge $(u, v) \in \mathcal{E}$.

We claim that we only need to consider the terms with corresponding graphs $\mathcal{G}' = (\mathcal{V}, \mathcal{E}')$ such that the following conditions hold:

[6] This allows us to entirely avoid the argument using $(m, p)$-regular graphs presented in [2].

Condition 1: $|\mathcal{V}| = mk + 1$.

Condition 2: $|\mathcal{E}'| = mk$.

At this point it might be useful for readers to see the proofs for Lemma 3.2.2 and Lemma 4.2.6 for analogous but simpler arguments. We note that $\mathcal{G}$ is connected. This is because the edges of $\mathcal{G}$ induce a walk $w$, which visits all vertices $v \in \mathcal{V}$, while starting and ending at same vertex $\ell_0$

$$w = (\ell_0, \dots, \ell_{m-1}, \ell_m, \dots, \ell_{2m-1}, \dots\dots, \ell_{2mk-2m}, \dots, \ell_{2mk-m-1}, \ell_{2mk-m}, \dots, \ell_{2mk-1} = \ell_0). \tag{118}$$

This means $\mathcal{G}'$ is connected as well. Hence if $|\mathcal{V}| > mk + 1$, then $|\mathcal{E}'| \geq |\mathcal{V}| - 1 > mk$. So the number of independent factors $X_{ij}$ of the term is strictly greater than $mk$. This implies that there exists an independent factor $X_{ij}$ of the term occurring exactly once. Since $\mathbf{E}X_{ij} = 0$ this implies

$$\mathbf{E} \prod_{j=0}^{2mk-1} X_{\ell_j \ell_{j+1}}^{\epsilon(j)} = 0. \tag{119}$$

If $|\mathcal{V}| = \omega_0 < mk + 1$, we count the number of such possible terms. Let $\mathcal{P}_{\omega_0}$ be the set of such terms, then

$$\begin{aligned} |\mathcal{P}_{\omega_0}| &= n(n-1)\cdots(n-\omega_0+1) \\ &= O(n^{\omega_0}). \end{aligned} \tag{120}$$

By assumption of bounded moments, the contribution of such terms to the sum as $n \to \infty$ goes to zero, since

$$\begin{aligned} \frac{1}{n^{mk+1}} \sum_{\mathcal{P}_{\omega_0}} \left( \mathbf{E} \prod_{j=0}^{2mk-1} X_{\ell_j \ell_{j+1}}^{\epsilon(j)} \right) &= \frac{1}{n^{mk+1}} \sum_{\mathcal{P}_{\omega_0}} O(1) \\ &= \frac{1}{n^{mk+1}} O(n^{\omega_0}) \\ &= O\left( \frac{n^{\omega_0}}{n^{mk+1}} \right) \to 0. \end{aligned} \tag{121}$$

Hence, Condition 1 holds. Since $|\mathcal{V}| = mk + 1$ and $\mathcal{G}'$ is connected, $|\mathcal{E}'| \geq |\mathcal{V}| - 1 = mk$. We have shown that $|\mathcal{E}'| > mk$ is impossible, hence Condition 2 holds as well. We conclude that $\mathcal{G}' = (\mathcal{V}, \mathcal{E}')$ is a tree.

By the tree property, we further conclude that:

1. Every independent factor $X_{ij}$ occurs twice, in conjugate pairs.
2. The pairing is noncrossing.

There are $2mk$ factors in total, but $mk$ of them are independent, therefore each independent factor occurs twice. To see why conjugate pairing is necessary, suppose the contrary for some independent factor $X_{ij}$. Then by observing the walk $w$ we would have found a path

$$i \to j \to \cdots \to i \to j. \tag{122}$$

which contradicts the tree property of $\mathcal{G}'$, since we are able to find a simple cycle off such a path. This shows that the first conclusion is true. To see why the second is true, we again suppose the contrary. Then we would have some order of appearance $X_{ij}, X_{i'j'}, \bar{X}_{ij}, \bar{X}_{i'j'}$ which implies the existence of a path

$$i \to j \to \cdots \to i' \to j' \to \cdots \to j \to i \to \cdots \to j' \to i'. \tag{123}$$

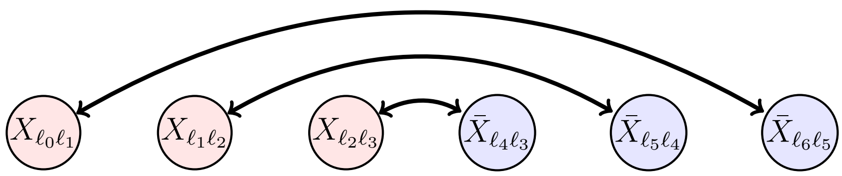
Again, this contradicts the tree property of $\mathcal{G}'$. It is important to note that a simple cycle can be found precisely because every edge occurs only twice.

Therefore we have reduced the original problem to one which is purely combinatorial. The question is, how many noncrossing conjugate pairings can be made given the format

$$\underbrace{\underbrace{X \cdots X}_{m \text{ terms}} \underbrace{\bar{X} \cdots \bar{X}}_{m \text{ terms}} \cdots\cdots \underbrace{X \cdots X}_{m \text{ terms}} \underbrace{\bar{X} \cdots \bar{X}}_{m \text{ terms}}}_{m \times 2k \text{ terms}}. \tag{124}$$

For example, when $m = 3$ and $k = 1$ there is only one way of doing do.

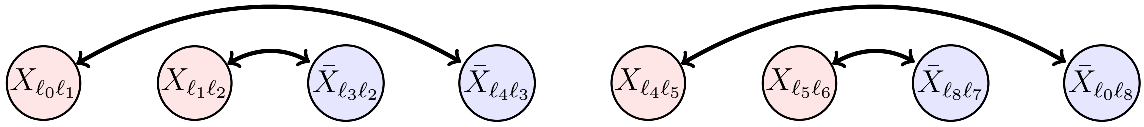
Figure 10: Possible Pairing Type when $m = 3$, $k = 1$.



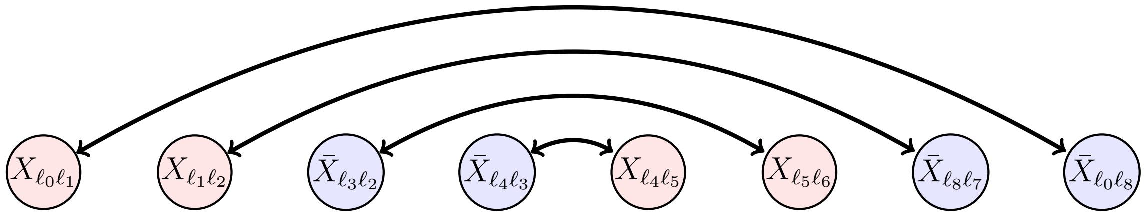

(i) Pairing Type $(m_3k_1)$-I

On the other hand, when $m = 2$ and $k = 2$ there are a total of 3 ways of doing so.

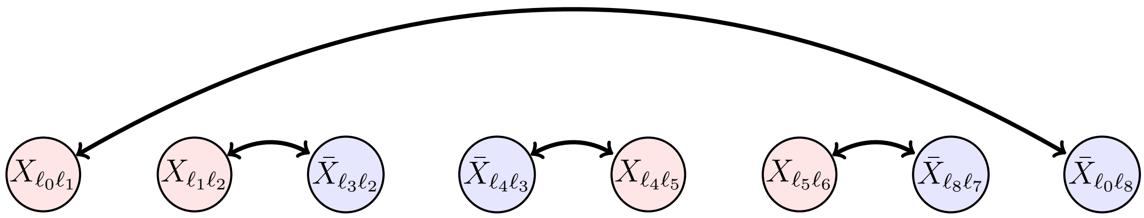
Figure 11: Possible Pairing Types when $m = 2$, $k = 2$.




(i) Pairing Type $(m_2k_2)$-I




(ii) Pairing Type $(m_2k_2)$-II




(iii) Pairing Type $(m_2k_2)$-III

Given $m, k \in \mathbb{N}$, denote the number of pairings by $P_{m,k}$. We claim that this is equal to the single-parameter Fuss-Catalan number $C_{m,k}$. To do so, it suffices to show that $P_{m,k}$ satisfies the same recurrence relation of the single-parameter Fuss-Catalan numbers that is

$$C_{m,k} = \sum_{\Omega_{m,k}} \left( \prod_{i=0}^{m} C_{m,k_i} \right), \tag{125}$$

where the underlying domain of the sum is defined to be

$$\Omega_{m,k} = \left\{ (k_0, k_1, \ldots, k_m) \in \mathbb{Z}_{\geq 0}^{m+1} : \sum_{i=0}^{m} k_i = k - 1 \right\}. \tag{126}$$

For $m, k \in \mathbb{N}$ we start by establish some properties with respect to the pairing.

Define the index of a factor $X_{\ell_j\ell_{j+1}}$, to be the number $0 \leq j \leq 2mk - 1$. This allows us to number the factors $X_{\ell_j\ell_{j+1}}$ from $0, 1, 2, \ldots, 2mk - 1$. We say an index $j$ is of positive spin if $j \equiv 0, 1, \ldots, m-1 \mod 2m$. We say an index $j$ is of negative spin if $j \equiv m, m+1, \ldots, 2m-1 \mod 2m$. Note that this naming is indeed related to the the 'spin' variable $\epsilon(j)$ defined earlier. Furthermore we are interested in pairing two indices of opposite spin in a noncrossing manner.

The first property is that, for a $j_1$ index and a $j_2$ index to pair we must have

$$j_2 \equiv 2m - j_1 \mod 2m. \tag{127}$$

To see why this is true, suppose $j_1 < j_2$ are the indices two factors in a noncrossing conjugate pairing. Consider the set of indices $j$ between $j_1$ and $j_2$.

$$S_{j_1,j_2} = \{j \in \mathbb{Z} : j_1 < j < j_2\}. \tag{128}$$

If $S_{j_1,j_2} = \emptyset$ then $j_1$ and $j_2$ are adjacent factors of opposite spin, hence $j_2 \equiv 2m - j_1 \mod 2m$.

If $S_{j_1,j_2} \neq \emptyset$ then our first observation is $|S_{j_1,j_2}|$ must be an even number, otherwise we will be forced to make a crossing-type pairing. Furthermore the number of positive spin indices $j \equiv 0, 1, \ldots, m-1 \mod 2m$ and negative spin indices $j' \equiv m, m+1, \ldots, 2m-1 \mod 2m$ must be equal and add up to a multiple of $2m$. This is enough to show that $j_2 \equiv 2m - j_1 \mod 2m$.

The second property is, if we take the leftmost group of positive spin indices, in particular those indexed $0, 1, \ldots, m-1$ and pair them first with negative spin indices $j_0^-, j_1^-, \ldots, j_{m-1}^-$, then we must have $j_0^- < j_1^- < \cdots < j_{m-1}^-$. This is clear as we would be forced to make a crossing otherwise.

Given $m, k \in \mathbb{N}$ and the two properties above we can find the recurrence relation on $P_{m,k}$. Apply the second property and for each possible pairing between the leftmost $m$ indices $0, 1, m-1$ and $j_0^-, j_1^-, \ldots, j_{m-1}^-$ as described above, consider the sets $S_{j_L,j_0^-}, S_{j_0^-,j_1^-}, \ldots, S_{j_{m-2}^-,j_{m-1}^-}, S_{j_{m-1}^-,j_R}$ where we denote $j_L = m-1$ and $j_R = 2mk-1$. By the first property we deduce that the cardinality of each set is some multiple of $2m$. This implies that

$$k_0 = \frac{|S_{j_L,j_0^-}|}{2m}, k_1 = \frac{|S_{j_0^-,j_1^-}|}{2m}, \ldots, k_{m-1} = \frac{|S_{j_{m-2}^-,j_{m-1}^-}|}{2m}, k_m = \frac{|S_{j_{m-1}^-,j_R}|}{2m} \in \mathbb{Z}_{\geq 0}, \tag{129}$$

and

$$k_0 + k_1 + \ldots k_{m-1} + k_m = k - 1. \tag{130}$$

Note that the possible pairings within each of the sets $S_{j_L,j_0^-}, S_{j_0^-,j_1^-}, \ldots, S_{j_{m-2}^-,j_{m-1}^-}, S_{j_{m-1}^-,j_R}$ are therefore $P_{m,k_0}, P_{m,k_1}, \ldots, P_{m,k_{m-1}}, P_{m,k_m}$ respectively. Therefore

$$P_{m,k} = \sum_{\Omega_k} \left( \prod_{i=0}^{m} P_{m,k_i} \right). \tag{131}$$

We further note $P_{m,1}$ for all $m \in \mathbb{N}$, just as the base cases for the single-parameter Fuss-Catalan numbers. Inductively, we obtain $P_{m,k} = C_{m,k}$ for all $m, k \in \mathbb{N}$. This is what was to be shown, since a closed-form expression for the Fuss-Catalans is given by

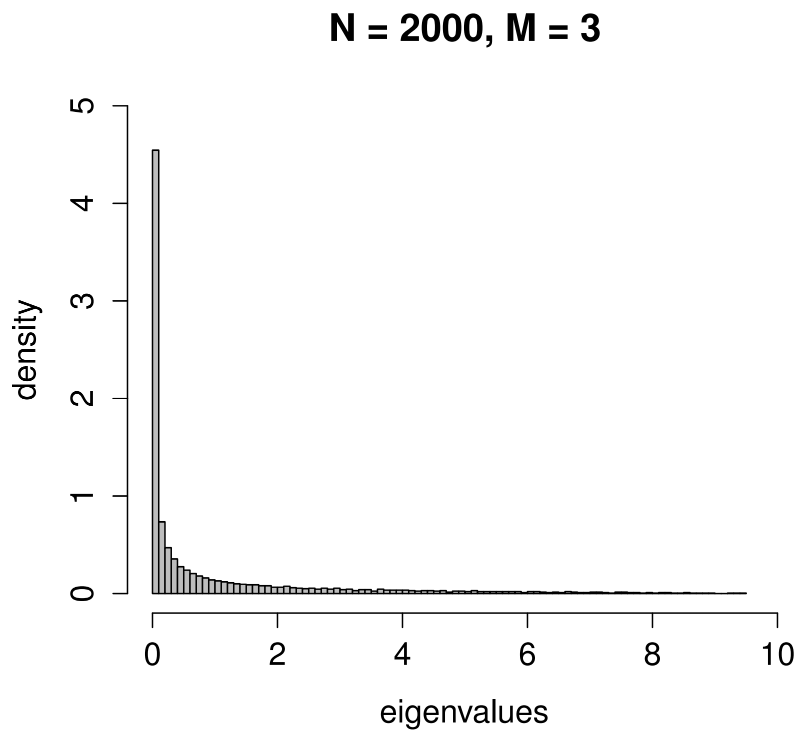
$$C_{m,k} = \frac{1}{mk+1} \binom{mk+k}{k}. \tag{132}$$

$\square$

*Remark* 5.2.2. We have not described the limiting density of such eigenvalues, as that is not the main purpose of the paper. It can be shown that the limiting density, based on these exact moments exists [8] and is unique [4].

In order to provide a rough idea of what this density look like, a sample histogram of eigenvalues of the random matrix $X^{(m)}X^{*(m)}$, when the underlying $X$ is a $2000 \times 2000$ random matrix with iid entries $X_{ij} \sim \mathrm{N}(0,1)$ and $m = 3$ is shown in Figure 12.

Figure 12: Eigenvalues of the matrix $X^{(3)}X^{*(3)}$



# 6 Conclusion

We have established convergence of expected moments of eigenvalue distributions (ESM) for three different random matrix ensembles. In these cases, we note that this is sufficient to conclude the limiting distribution of the ESM from the convergence of expected moments. We also observe several interesting combinatorial and graph arguments when we proceed with the moment method.

The three proofs for convergence of expected moments also share similar ideas. It is almost customary to reduce the initial expression to in terms of the trace of the underlying matrix. The next steps usually involve getting rid of the insignificant terms and then counting the significant terms in the trace expression. To do so, we required the construction of a graph and/or a sequence of digits.

# 7 Acknowledgement

I would like to thank my supervisor, Assistant Professor Wang Dong for his patience and helpfulness throughout the semester. The experience gained throughout this process was invaluable, and I have learnt a lot more than I have expected to. I am especially grateful for all the academic advice and time dedicated during our almost-weekly meetings.

# References

[1] Greg W. Anderson, Alice Guionnet, Ofer Zeitouni, *"An Introduction to Random Matrices"*. Cambridge University Press.

[2] N. Alexeev, F. Götze, A. Tikhomirov. *"Asymptotic Distribution of Singular Values of Powers of Random Matrices"*. Lithuanian Mathematical Journal, Vol. 50, No. 2, 2010, pp. 121-132.

[3] Jinho Baik, *"Topics in Analysis: Random Matrices, Preliminary Lecture Notes"*. Department of Mathematics, University of Michigan, 2009.

[4] T. Carleman, *Les fonctions quasi-analytiques*. Paris, 1926.

[5] Nicholas Cook, *"Two proofs of Wigner's semicircular law"*. University of California, Los Angeles Mathematics Department (2012).

[6] Rick Durrett, *"Probability: Theory and Examples"*. 4th Edition, Cambridge University Press, URL: https://services.math.duke.edu/~rtd/PTE/PTE4_1.pdf (2009).

[7] Adina Roxana Feier, Michael P. Brenner, *"Methods of Proof in Random Matrix Theory"*. Harvard University Department of Mathematics.

[8] Ira M. Gessel, Guoce Xin, *The generating function of ternary trees and continued fractions.* Electron. J. Combin. 13 (2006), no. 1.

[9] Ronald L. Graham, Donald E. Knuth, Oren Patashnik, *"Concrete Mathematics"*. Addison-Wesley Publishing Company, 1994.

[10] Leonid B. Koralov, Yakoc G. Sinai, *"Theory of Probability and Random Processes"*. Second Edition, Springer Berlin Heidelberg, 2007.

[11] Xiaochuan Lu, Hitoshi Murayama, *"Universal Asymptotic Eigenvalue Distribution of Large N Random Matrices"*. UCB-PTH-14/36, IPMU14-0317 (2015).

[12] Sheldon Ross, *A First Course in Probability*. Pearson Education Limited, 2013.

[13] Terence Tao, *"Discrete random matries and universality"*. University of California, Los Angeles. Mahler Lecture Series, 2009.

[14] Terence Tao, *Minerva Lectures 2013 - Terence Tao Talk 3: Universality for Wigner random matrices*. URL: https://www.youtube.com/watch?v=tihxQGGrMcc, 2013.

[15] Terence Tao, *"Topics in random matrix theory"*. Department of Mathematics, UCLA, Los Angeles, CA. 2009.

[16] Benedek Valkó, *"Lectures 6-7: The Marchenko-Pastur Law"*. Department of Mathematics, University of Wisconsin-Madison, 2009.

# A Generating Random Matrices

## A.1 Wigner Random Matrices

```
# generate wigner matrix
N <- 1000
H <- matrix(rnorm(N*N, mean = 0, sd = 1), N, N)
H[lower.tri(H)] <- t(H)[lower.tri(H)]
eigenvalues <- eigen(H)$values /  sqrt(N)
hist(eigenvalues, freq = FALSE, breaks = 50,
     xlab="eigenvalues", ylab = "density", col = "gray",
     xlim = c(-2,2), ylim = c(0,0.5))
# plot semicircle distribution
semicircle <- function(x) {
        sqrt(4- x^2) / (2*pi)
}
xpt <- seq(from = -2, to = 2, by = 0.001)
lines(xpt, semicircle(xpt), col = 'red')
```

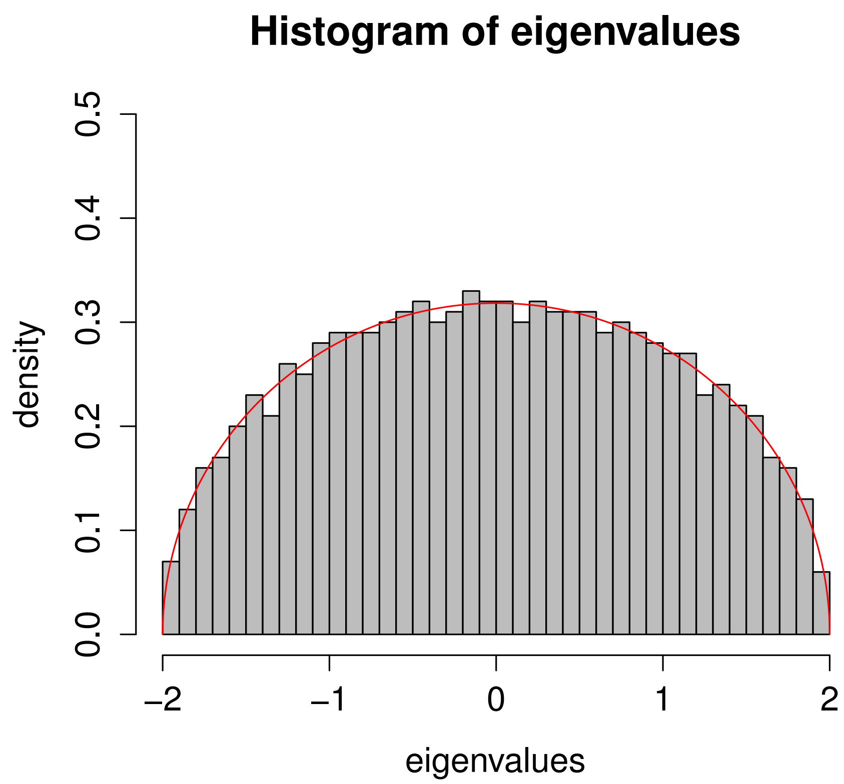



Figure 13: R Code for Simulating Wigner Random Matrices (above), Histogram of Eigenvalues with superimposed Semicircle Density (below). Note that $n = 1000$.

## A.2 Wishart Random Matrices

```
# generate wishart matrix
N <- 1000
P <- floor(N/2)
X <- matrix(rnorm(P*N, mean = 0, sd = 1), P, N)
eigenvalues <- eigen(X %*% t(X))$values / N
hist(eigenvalues, freq = FALSE, breaks = 50,
     xlab="eigenvalues", ylab = "density", col = "gray",
     xlim = c(0,4), ylim = c(0,1))
# plot marchenko pastur distribution
a <- (1 - sqrt(0.5))^2
b <- (1 + sqrt(0.5))^2
marchenko_pastur <- function(x) {
        sqrt((x-a)*(b-x)) / (2*pi*x*0.5)
}
xpt <- seq(from = a, to = b, by = 0.001)
lines(xpt, marchenko_pastur(xpt), col = 'red')
```

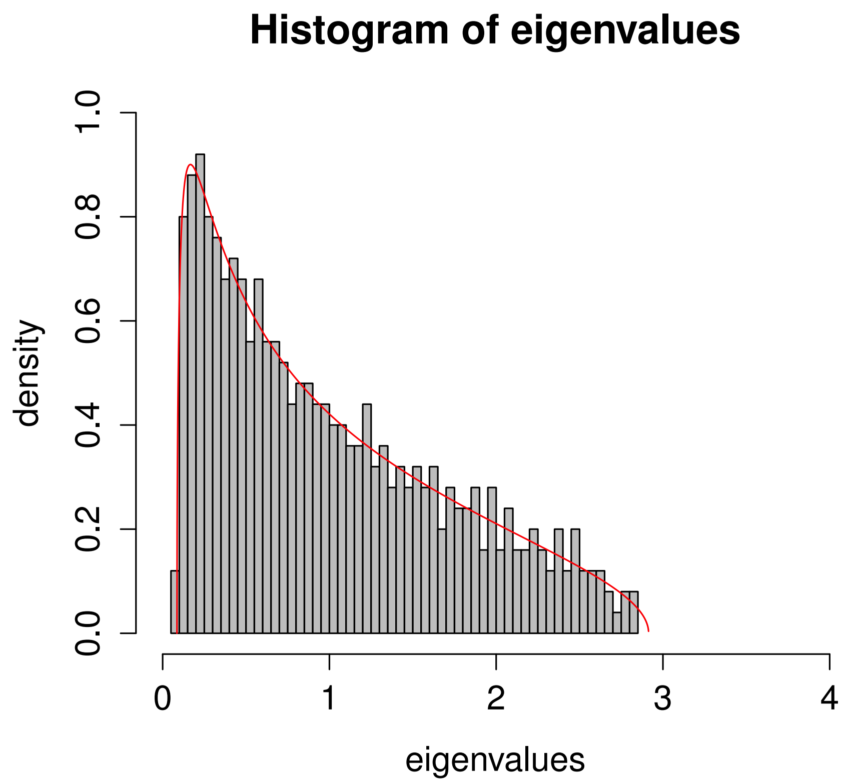



Figure 14: R Code for Simulating Wishart Random Matrices (above), Histogram of Eigenvalues with superimposed Marchenko-Pastur Density (below). Note that $n = 1000$ and $p_n = \lfloor \frac{n}{2} \rfloor = 500$.

## A.3 Singular Values of Powers of Random Matrices

```
# generate third-type matrix
library("expm") # for matrix exponentiation
M <- 3
N <- 2000
X <- matrix(rnorm(N*N, mean = 0, sd = 1), N, N)
eigenvalues <- eigen((X %^% M) %*% (t(X) %^% M))$values / (N^M)
hist(eigenvalues, freq = FALSE, breaks = 100,
     xlab="eigenvalues", ylab = "density", col = "gray",
     xlim = c(0,10), ylim = c(0,5))
```

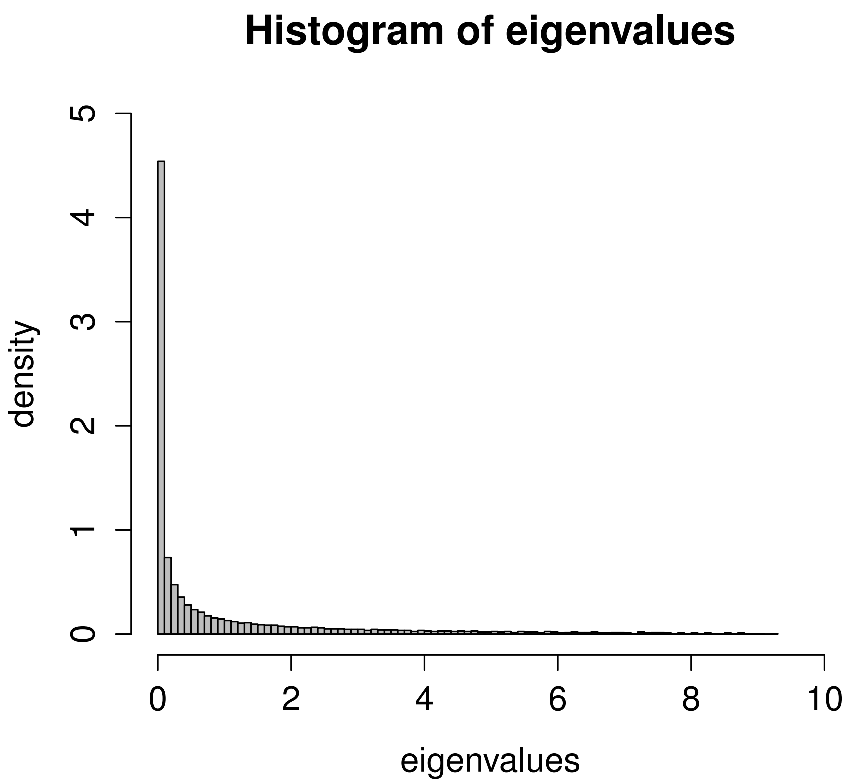



Figure 15: R Code for Simulating the third type of Random Matrices $X^{(3)}X^{*(3)}$ (above), Histogram of Eigenvalues with (below). Note that the dimension is $n = 2000$ and power is $m = 3$.

# B Some Fuss-Catalan Numbers

Table 1 provides reference for some values of Catalan numbers.

Table 1: List of Catalan Numbers, $C_k$ for $0 \leq k \leq 10$

| $k$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $C_k$ | 1 | 1 | 2 | 5 | 14 | 42 | 132 | 429 | 1430 | 4862 | 16796 |

Table 2 provides reference for some values of Fuss-Catalan numbers. They are, in fact a generalisation of the Catalan numbers $C_k$. This is identified by $C_{1,k} = C_k$ for all $k \in \mathbb{N}$.

In some texts these referred to as single-parameter Fuss-Catalan numbers, so as to avoid confusion with the two-parameter Fuss-Catalan numbers. The 'single-parameter' refers to the integer $m$.

Table 2: List of Fuss Catalan Numbers, $C_{m,k}$ for $0 \leq m \leq 5$ and $0 \leq k \leq 10$

| $k$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $C_{0,k}$ | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 1 |
| $C_{1,k}$ | 1 | 1 | 2 | 5 | 14 | 42 | 132 | 429 | 1430 | 4862 | 16796 |
| $C_{2,k}$ | 1 | 1 | 3 | 12 | 55 | 273 | 1428 | 7752 | 43263 | 246675 | 1430715 |
| $C_{3,k}$ | 1 | 1 | 4 | 22 | 140 | 969 | 7084 | 53820 | 420732 | 3362260 | 27343888 |
| $C_{4,k}$ | 1 | 1 | 5 | 35 | 285 | 2530 | 23751 | 231880 | 2330445 | 23950355 | 250543370 |
| $C_{5,k}$ | 1 | 1 | 6 | 51 | 506 | 5481 | 62832 | 749398 | 9203634 | 115607310 | 1478314266 |

The equality $C_{1,k} = C_k$ is demonstrated for small values of $0 \leq k \leq 10$ in Tables 1 and 2.