The Eigenvalue Distribution of Random Matrices with Distortions

Kevin Hunt

April 2019

1. Introduction

Random matrix theory is often concerned with describing the limiting eigenvalue distribution of random matrix ensembles. For example, Wigner's semicircle law shows that the empirical distribution of eigenvalues for real symmetric matrices with independent, mean zero, and variance one random entries scaled by $\sqrt{N}$ converges weakly in probability to the semicircle law [1]. It has been shown that the eigenvalue distributions of Toeplitz and Hankel matrices with independent, mean zero, and variance one entries each converge almost surely to a unique distribution [3].

A natural next question is to study what happens to the limiting eigenvalue distribution when making slight changes to some of the entries of a matrix ensemble. The limiting distributions of matrix ensembles have been studied when the (i,j) and (j,i) entries are randomly multiplied by $\epsilon_{i,j} = \epsilon_{j,i} \in \{-1, 1\}$ where $\mathbb{P}(\epsilon_{i,j} = 1) = p$. For $p = 1/2$, the limiting eigenvalue distribution is the semicircle law when each random variable occurs $O(N)$ times in each row; this result holds for Toeplitz and circulant matrix ensembles [1]. The k-checkerboard matrix has been defined and studied, where each entry on the kth diagonal is equal to some constant w, and all other entries are independent, mean zero, variance one random variables. It has been shown that the eigenvalue distribution of this random matrix ensemble converges weakly in probability to the semicircle law. This can be thought of as changing the kth diagonals of a Wigner matrix to w.

In this thesis we study what happens when some entries of a random matrix ensemble are changed to zero. We consider adding these zero distortions randomly, and we consider when the distortions are added in a specific non-random way. In general, we show that under certain conditions the limiting distributions of the eigenvalues are not changed, either in a random or structured way, if "not too many" of the entries in the matrix are changed to zero.

1.1. Background

Here we introduce some definitions and notations that will be useful for our results.

Definition 1.1: Given a real symmetric NxN random matrix A_N, we can list its real eigenvalues as λ_i^N with $\lambda_1^N \leq \lambda_2^N \leq \cdots \leq \lambda_N^N$. The empirical distribution of the eigenvalues of A_N is then defined as the random probability measure on $\mathbb{R}$ given by

$$L_N := \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i^N}.$$

Definition 1.2: L_N converges weakly in probability to some distribution $\boldsymbol{\alpha}$ if for all functions $f \in C_b(\mathbb{R})$, the set of all bounded continuous functions on $\mathbb{R}$, and for all $\epsilon > 0$

$$\lim_{N\to\infty} \mathbb{P}(|\langle f, L_N\rangle - \langle f, \boldsymbol{\alpha}\rangle| > \epsilon) = 0.$$

Definition 1.3: Given two real NxN matrices A_N and B_N, the Hadamard product matrix $H_N = A_N \circ B_N$ is the NxN matrix whose $(i,j)^{th}$ entry $H_N(i,j) = A_N(i,j) \circ B_N(i,j)$.

Definition 1.4: Given a real symmetric NxN random matrix A_N and a symmetric NxN matrix 1_N whose entries are either 1 or 0, we can list the real eigenvalues of $1_N \circ A_N$ as ξ_i^N with $\xi_1^N \leq \xi_2^N \leq \cdots \leq \xi_N^N$. The empirical distribution of these eigenvalues is then defined as the random probability measure on $\mathbb{R}$ give by

$$D_N := \frac{1}{N}\sum_{i=1} \delta_{\xi_i^N} \quad .$$

Definition 1.5: Consider two sequences of i.i.d. random variables $\{Z_{i,j}\}_{1\leq i<j}$ and $\{\mathrm{Y_i}\}_{1\leq i}$ with mean zero and variance one that satisfy $\max\{EZ_{1,2}^k, EY_1^k\} < \infty$ for all $k \in \mathbb{N}$. Then a Wigner matrix is a real symmetric random matrix with entries

$$X_N(i,j) = X_N(j,i) = \begin{cases} \frac{Z_{i,j}}{\sqrt{N}} \ if\ i \neq j \\ \frac{Y_i}{\sqrt{N}} if\ i = j \end{cases} .$$

Definition 1.6: The semicircle law scaled by p is the distribution with density

$$\boldsymbol{\sigma_p}(x) := \frac{1}{2p\pi}\sqrt{4p - x^2} \cdot 1_{|x|<2\sqrt{p}} \quad .$$

Wigner's Semicircle Law: The empirical distribution of the eigenvalues of the Wigner matrix ensemble $\{X_N\}$ converges weakly in probability to $\boldsymbol{\sigma_1}$.

Definition 1.6: For an NxN non-random symmetric matrix 1_N whose entries are zero or one, we define

$$\theta(N) := \sum_{i,j=1}^{N} 1_N(i,j) \ , \qquad \text{and} \qquad \theta_i(N) := \sum_{j=1}^{N} 1_N(i,j) = \sum_{j=1}^{N} 1_N(j,i).$$

1.2 Results

The first result involves an ensemble of random matrices $\{A_N\}$ whose entries all have non-negative moments. That is, $EA_N(i,j)^k \geq 0$ for all $i,j,k,N \in \mathbb{N}$. The distorting matrix 1_N will have its entries be Bernoulli random variables.

Theorem 1.1: Suppose $\{A_N\}$ is any ensemble of symmetric random matrices whose entries have non-negative moments. Suppose $\{1_N\}$ is an ensemble of symmetric random matrices composed of Bernoulli random variables with parameter $p(N) \in (0,1]$ and $\lim_{N\to\infty} p(N) = 1$. If the empirical distribution of the eigenvalues of A_N converges weakly in probability to some distribution $\boldsymbol{\alpha}$ which has bounded support, then the empirical distribution of the eigenvalues of $1_N{}^\circ A_N$ converges to $\boldsymbol{\alpha}$ as well.

If $\lim_{N\to\infty} p(N) = p \in (0,1]$, then we can say something about the limiting distribution of the eigenvalues if we know the structure of the ensemble A_N. For example, suppose $\{A_N\}$ is an ensemble of Wigner matrices defined earlier, then $1_N \circ A_N$ has independent entries with mean zero and variance p. Then the matrix ensemble B_N with each entry multiplied by the constant $\frac{1}{\sqrt{p}}$

$$B_N = \frac{1}{\sqrt{p}} \cdot 1_N \circ A_N$$

will have mean zero, variance one, independent entries. Because B_N satisfies the conditions for Wigner's semicircle law, then the empirical distribution of the eigenvalues of B_N converges to $\boldsymbol{\sigma_1}$. This tells us that the eigenvalues of $1_N \circ A_N$ are just the eigenvalues of B_N multiplied by $\sqrt{p}$, and thus, the limiting distribution of the eigenvalues of $1_N \circ A_N$ is $\boldsymbol{\sigma_{\sqrt{p}}}$. Also, Wigner matrices can have negative moments beyond the second moment, so this result did not follow from Theorem 1.1 for $p = 1$.

We now consider distortions added in a non-random way, specifically for the Wigner random matrix ensemble.

Theorem 1.2: Suppose $\{X_N\}$ is an ensemble of Wigner random matrices, and $\{1_N\}$ are non-random matrices with entries equal to zero or one, and for all j, $\theta_j(1_N) = pN$ with $p \in (0,1]$. Then the empirical distribution of the eigenvalues of $1_N \circ X_N$ converges to $\boldsymbol{\sigma_{\sqrt{p}}}$.

Theorem 1.3: Suppose $\{X_N\}$ is an ensemble of Wigner random matrices, and $\{1_N\}$ are non-random symmetric matrices with entries equal to zero or one, and

$$\lim_{N\to\infty} \frac{\theta(1_N)}{N^2} = 0.$$

Then the empirical distribution of the eigenvalues of $1_N \circ X_N$ converges to $\boldsymbol{\sigma_1}$.

Theorems 1.2 and 1.3 only cover a limited number of possible 1_N matrices. We have yet to find a general formula for the limiting eigenvalue distribution of $1_N \circ X_N$, but we do have equations for the first eight moments, and a process for finding all higher moments. We will discuss these topics in section 3.

2. Method of Moments

All our proofs will be done by the Method of Moments as described in [1], which is described in Theorem 2.1. The proof requires two conditions, we will proof in each of the previously discussed cases.

Theorem 2.1: Let $\{A_N\}$ be an ensemble of NxN random matrices with empirical distribution of the eigenvalues L_N, and let $\boldsymbol{\alpha}$ be some distribution with bounded support $[-M, M]$. Suppose that

(i) For all $k \in \mathbb{N}$,

$$\lim_{N\to\infty} E\langle x^k, L_N\rangle = E\langle x^k, \boldsymbol{\alpha}\rangle.$$

(ii) For all $k \in \mathbb{N}$ and $\epsilon > 0$,

$$\lim_{N\to 0} \mathbb{P}\big(\big|\langle x^k, L_N\rangle - E\langle x^k, L_N\rangle\big| > \epsilon\big) = 0.$$

Then the empirical measure of the eigenvalues L_N converges weakly in probability, to the distribution $\boldsymbol{\alpha}$.

Assume that L_N and $\boldsymbol{\alpha}$ satisfy conditions (i) and (ii). We need to show that for all continuous bounded functions f and for all $\delta > 0$:

$$\lim_{N\to\infty} \mathbb{P}(|\langle f, L_N\rangle - \langle f, \boldsymbol{\alpha}\rangle| > \delta) = 0.$$

For such an f by the Weirstrass Approximation Theorem, we can find a polynomial $Q_\delta(x) \coloneqq \sum_{i=1}^{B} b_i x^i$ that satisfies:

$$\sup_{x:|x|\le R} |Q_\delta(x) - f(x)| \le \frac{\delta}{8}$$

where $R = (2M)^2$. Note,

$$\mathbb{P}(|\langle f, L_N\rangle - \langle f, \boldsymbol{\alpha}\rangle| > \delta) \le$$

$$\mathbb{P}\big(|\langle f, L_N\rangle - \langle f, \boldsymbol{\alpha}\rangle| \cdot 1_{|x|\le B} > \delta\big) + \mathbb{P}\big(|\langle f, L_N\rangle - \langle f, \boldsymbol{\alpha}\rangle| \cdot 1_{|x|> B} > \delta\big) \le$$

$$\mathbb{P}\left(|\langle f, L_N\rangle - E\langle f, L_N\rangle| \cdot 1_{|x|\le B} > \frac{\delta}{2}\right) + \mathbb{P}\left(|E\langle f, L_N\rangle - \langle f, \boldsymbol{\alpha}\rangle| \cdot 1_{|x|\le B} > \frac{\delta}{2}\right)$$
$$+ \mathbb{P}\left(|\langle f, L_N\rangle| \cdot 1_{|x|> B} > \frac{\delta}{2}\right) + \mathbb{P}\left(|\langle f, \boldsymbol{\alpha}\rangle| \cdot 1_{|x|> B} > \frac{\delta}{2}\right) =: P_1 + P_2 + P_3 + P_4.$$

First, $P_4 = 0$ because $\boldsymbol{\alpha}$ is only supported on $[-M, M]$, but $B = 2M$.

By definition of Q_δ,

$$P_1 = \mathbb{P}\left(|\langle f, L_N\rangle - E\langle f, L_N\rangle| \cdot 1_{|x|\le B} > \frac{\delta}{2}\right) \le \mathbb{P}\left(|\langle Q_\delta, L_N\rangle - E\langle Q_\delta, L_N\rangle| > \frac{\delta}{4}\right) \to 0$$

by Lemma 2.2. Similarly,

$$P_2 = \mathbb{P}\left(|E\langle Q_\delta, L_N\rangle - \langle f, \alpha\rangle| \cdot 1_{|x|\le B} > \frac{\delta}{2}\right) \le \mathbb{P}\left(|E\langle Q_\delta, L_N\rangle - \langle Q_\delta, \alpha\rangle| > \frac{\delta}{4}\right) \to 0$$

by Lemma 2.1.

Finally, by Chebyshev's inequality for all $\epsilon > 0$:

$$\mathbb{P}\big(\langle |x|^k 1_{|x|>B}, L_N\rangle > \epsilon\big) \le \frac{1}{\epsilon} E\langle |x|^k 1_{|x|>B}, L_N\rangle \le \frac{E\langle |x|^{2k}, L_N\rangle}{\epsilon B^k} \quad .$$

Then by Lemma 2.1,

$$\limsup_{N\to\infty} \mathbb{P}\big(\langle |x|^k 1_{|x|>B}, L_N\rangle > \epsilon\big) \le \frac{m_{2k}}{\epsilon B^k} \le \frac{M^{2k}}{\epsilon (2M)^{2k}} \quad .$$

The left-hand side of the above expression is increasing in k, but the right-hand side is decreasing in k, so

$$\limsup_{N\to\infty} \mathbb{P}\big(\langle |x|^k 1_{|x|>B}, L_N\rangle > \epsilon\big) = 0.$$

This implies that $P_3 \to 0$ as $N \to \infty$ which completes the proof of Theorem 2.3.

[There is an upward pointing arrow here in your previous comments, but I don't know what that meant.]

2.1. Proof of Theorem 1.1

Here we consider an ensemble of real symmetric random matrices A_N whose empirical distribution of eigenvalues L_N converges weakly in probability to some distribution $\boldsymbol{\alpha}$ with compact support. The matrices A_N also have entries with nonnegative moments, that is $EA_N^k(i,j) \ge 0\ \forall i, j, k, N \in \mathbb{N}$. These matrices are then distorted by multiplying $1_N \circ A_N$, where 1_N is a symmetric random matrix composed of Bernoulli random variables with parameter $p(N) \to 1$ as $N \to \infty$. The empirical distribution of the eigenvalues of $1_N \circ A_N$ is D_N defined earlier.

Proof of Condition (i): By the independence of entries in 1_N and A_N:

$$E\langle x^k, D_N\rangle = \frac{1}{N} tr(1_N \circ A_N)^k = \frac{1}{N}\sum_{i_1,i_2,\dots,i_k=1}^{N} E\, 1_N(i_1,i_2)A_N(i_1,i_2)\dots 1_N(i_k,i_1)A_N(i_k,i_1)$$

$$= \frac{1}{N}\sum_{i_1,i_2,\dots,i_k=1}^{N} E[1_N(i_1,i_2)\dots 1_N(i_k,i_1)]\cdot E[A_N(i_1,i_2)\dots A_N(i_k,i_1)] \quad .$$

Remember that every moment of the Bernoulli random variable is equal to its parameter $p(N) \le 1$. The $E[1_N(i_1,i_2)\dots 1_N(i_k,i_1)]$ term is the expectation of at most k independent Bernoulli random variables, which happens when none of the indices generate entries that are the same, and at least one Bernoulli random variable, which happens when the indices all generate the same entry. Therefore,

$$p(N)^k \le E[1_N(i_1,i_2)\dots 1_N(i_k,i_1)] \le p(N).$$

Since the moments of the entries in A_N are non-negative, we can create bounds on $E\langle x^k, D_N\rangle$ as follows:

$$p(N)^k \frac{1}{N}\sum_{i_1,i_2,\dots,i_k=1}^{N} E A_N(i_1,i_2)\dots A_N(i_k,i_1)$$

$$\le E\langle x^k, D_N\rangle \le$$

$$p(N)\frac{1}{N}\sum_{i_1,i_2,\dots,i_k=1}^{N} E A_N(i_1,i_2)\dots A_N(i_k,i_1).$$

The summations in the upper and lower bounds are the kth moments of the eigenvalue distribution of A_N, so

$$p(N)^k \cdot E\langle x^k, L_N\rangle \le E\langle x^k, D_N\rangle \le p(N)\cdot E\langle x^k, L_N\rangle.$$

By assumption, $p(N) \to 1$ and $E\langle x^k, L_N\rangle \to E\langle x^k, \alpha\rangle$. Then the above inequalities show that

$$\lim_{N\to\infty} E\langle x^k, D_N\rangle = E\langle x^k, \boldsymbol{\alpha}\rangle.$$

Proof of Condition (ii): By Markov's inequality,

$$\mathbb{P}\left(\left|\langle x^k, D_N\rangle - E\langle x^k, D_N\rangle\right| > \epsilon\right) \le \frac{1}{\epsilon} E\left|\langle x^k, D_N\rangle - E\langle x^k, D_N\rangle\right|$$

$$\leq \frac{1}{\epsilon}\left(E\left|\langle x^k, D_N\rangle - \langle x^k, L_N\rangle\right| + E\left|\langle x^k, L_N\rangle - E\langle x^k, \boldsymbol{\alpha}\rangle\right| + E\left|E\langle x^k, \boldsymbol{\alpha}\rangle - E\langle x^k, D_N\rangle\right|\right)$$

$$=: \frac{1}{\epsilon}(T_1 + T_2 + T_3).$$

First, T_2 goes to zero because L_N converges weakly in probability to $\boldsymbol{\alpha}$. Also, T_3 goes to zero because $E[\langle x^k, D_N\rangle]$ converges to $E[\langle x^k, \boldsymbol{\alpha}\rangle]$. Finally,

$$T_1 = \frac{1}{N} tr A_N^k - tr(1_N \circ A_N)^k$$

$$= \frac{1}{N} \sum_{i_1,\dots i_k=1}^{N} \left(1 - E1_N(i_1, i_2) \dots 1_N(i_k, i_1)\right) EA_N(i_1, i_2) \dots A_N(i_k, i_1).$$

The $1 - E1_N(i_1, i_2) \dots 1_N(i_k, i_1)$ term contains at most k independent Bernoulli random variables, and at least one, so

$$1 - p(N) \leq 1 - E1_N(i_1, i_2) \dots 1_N(i_k, i_1) \leq 1 - p(N)^k.$$

Therefore, $\left(1 - p(N)\right)E[\langle x^k, L_N\rangle] \leq T_1 \leq \left(1 - p(N)^k\right)E[\langle x^k, L_N\rangle]$. Taking the limit as $N \to \infty$ in the above expression shows $T_1 \to 0$. Therefore, $\lim_{N\to\infty} \mathbb{P}\left(\left|\langle x^k, D_N\rangle - E\langle x^k, D_N\rangle\right| > \epsilon\right) = 0$.

Now that we have shown conditions (i) and (ii) to be true under the conditions of Theorem 1.1, we know that the theorem holds.

2.2. Proof of Theorem 1.2

Now we consider the ensemble of Wigner random matrices X_N described in Definition 1.5. The eigenvalue distribution of X_N converges weakly in probability to the semicircle law $\boldsymbol{\sigma_1}$ described before. A full proof of this can be found in [1], which shows Wigner's original argument. We consider Wigner matrices X_N distorted by multiplying $1_N \circ X_N$, where 1_N is symmetric and composed of entries equal to zero or one. Also, $\theta_j(1_N) = pN$ for all j with $p \in (0,1)$. With a slight change to the argument found in [1], we show the eigenvalue distribution of $1_N \circ X_N$ converges to the semicircle law scaled by $\sqrt{p}$, denoted by $\boldsymbol{\sigma}_{\sqrt{\boldsymbol{p}}}$.

First, we find the moments of $\boldsymbol{\sigma}_{\sqrt{\boldsymbol{p}}}$. Let C_n denote the nth Catalan number for $n \in \mathbb{N}$, which is:

$$C_n = \frac{1}{n+1}\binom{2n}{n}.$$

If k is odd, then because $\boldsymbol{\sigma_p}$ is an even function and x^k is an odd function

$$E\,\langle x^k, \boldsymbol{\sigma}_{\sqrt{\boldsymbol{p}}}\rangle = \frac{1}{2\sqrt{p}\pi}\int x^k \boldsymbol{\sigma}_{\sqrt{\boldsymbol{p}}}(\boldsymbol{x}) 1_{|x|\le p^{\frac{1}{4}}}\, dx = 0.$$

For even moments,

$$E\,\langle x^{2k}, \boldsymbol{\sigma}_{\sqrt{\boldsymbol{p}}}\rangle = \frac{1}{2\sqrt{p}\pi}\int x^{2k} \boldsymbol{\sigma_p}(\boldsymbol{x}) 1_{|x|\le p^{\frac{1}{4}}}\, dx = \frac{2^{2k+1}p^k}{\pi}\int \sin^{2k}\theta\cos^2\theta\, d\theta$$

$$= \frac{2^{2k+1}p^k}{\pi}\int \sin^{2k}\theta\, d\theta - (2k+1)E\,\langle x^{2k}, \boldsymbol{\sigma}_{\sqrt{\boldsymbol{p}}}\rangle.$$

Therefore,

$$E\,\langle x^{2k}, \boldsymbol{\sigma}_{\sqrt{\boldsymbol{p}}}\rangle = \frac{2^{2k+1}p^k}{\pi(2k+2)}\int \sin^{2k}\theta\, d\theta = p^k\frac{4(2k-1)}{2k+2}m_{2k-2},$$

Using induction and starting with $E\,\langle x^0, \boldsymbol{\sigma}_{\sqrt{\boldsymbol{p}}}\rangle = 1$, we can see that $E\,\langle x^{2k}, \boldsymbol{\sigma}_{\sqrt{\boldsymbol{p}}}\rangle = p^k C_k$.

Proof of Condition (i): Let $\widetilde{\mathrm{X}}_N(i,j) = 1_N(i,j)\cdot X_N(i,j)$ and similarly for $\tilde{Y}_i$ and $\tilde{Z}_{i,j}$. Then

$$\mathrm{E}\langle x^k, D_N\rangle = \frac{1}{N}tr(1_N\circ X_N)^k = \frac{1}{\mathrm{N}}\sum_{i_1,\ldots,i_k=1} E\tilde{X}_N(i_1,i_2)\ldots\tilde{X}(i_k,i_1) = \frac{1}{\mathrm{N}}\sum_{\boldsymbol{i}} T_{\boldsymbol{i}}^N$$

where $\boldsymbol{i} = (i_1, i_2, \ldots i_k)$.

An N-word of length $m \ge 1$ is $w = a_1 a_2 \ldots a_m$ where $a_j \in \{1,\ldots,N\}$. The support of w denoted $\mathrm{supp}(w)$ is the unique elements appearing in w, and the weight of w denoted $\mathrm{wt}(w) = |\mathrm{supp}(w)|$. Each word w generates an undirected graph G_w with $m-1$ edges and $\mathrm{wt}(w)$ vertices. Denote the undirected edges of G_w by $E_w = \left\{\{i_s, i_{s+1}\} : s = 1,2,\ldots,m-1\right\}$, and the vertices of G_w by $V_w = \mathrm{supp}w$. The self-edges of w are $E_w^s = \{e \in E_w : e = \{b,b\}, b \in \mathrm{supp}(w)\}$, and the connecting edges of w are $E_w^c = E_w - E_w^s$. The number of times the graph G_w crosses edge e is denoted by N_w^e. A word w is closed if $a_1 = a_m$.

With these definitions, we see that each $\boldsymbol{i}$ is a closed connected N-word of length $k+1$ where $w_{\boldsymbol{i}} = i_1 i_2 \ldots i_k i_1$. Therefore,

$$T_{\boldsymbol{i}}^N = \frac{1}{N^{k/2}}\prod_{e\in E_{w_{\boldsymbol{i}}}^c} E\tilde{Z}_{1,2}^{N_{w_{\boldsymbol{i}}}^e}\prod_{e\in E_{w_{\boldsymbol{i}}}^s} E\tilde{Y}_{1,1}^{N_{w_{\boldsymbol{i}}}^e}.$$

If there is $e \in E_{w_{\boldsymbol{i}}}$ with $N_{w_{\boldsymbol{i}}}^e = 1$, then $T_{\boldsymbol{i}} = 0$ because $E\tilde{Z}_{1,2} = E\tilde{Y}_{1,1} = 0$. Thus, we only need to consider $\boldsymbol{i}$ where $N_{w_{\boldsymbol{i}}}^e \ge 2$ for all $e \in E_{w_{\boldsymbol{i}}}$. Because $w_{\boldsymbol{i}}$ has k edges and each vertex after the first accounts for two edges, $\mathrm{wt}(w_{\boldsymbol{i}}) \le k/2+1$.

If there is a bijection from w_i to w_j, these words are called equivalent and generate the same graph up to graph isomorphism, meaning that $T_i = T_j$. For an N-word w_i with $\mathrm{wt}(w_i) = u$, the number of words equivalent to w_i is given by $\psi_{N,u}$ where

$$N(pN-1)\dots(pN-u+1) \le \psi_{N,u} \le (pN)^{\frac{k}{2}} \tag{2.2.1}$$

Let $\mathbb{W}_{k,u}$ be the set of representatives for the equivalence classes of closed N-words w_i of length $k+1$ and weight u. Then

$$E\langle x^k, D_N\rangle = \sum_{u=1}^{\lfloor k/2\rfloor+1} \frac{\psi_{N,u}}{N^{k/2+1}} \sum_{w\in\mathbb{W}_{k,u}} \prod_{e\in E_w^c} E\tilde{Z}_{1,2}^{N_w^e} \prod_{e\in E_w^s} E\tilde{Y}_{1,1}^{N_w^e}.$$

The moments of $Z_{1,2}$ and $Y_{1,1}$ are all finite by assumption. $|\mathbb{W}_{k,u}| \le u^k \le k^k$ because there are u vertices for each of the k edges to end at in the N-word of weight u and length $k+1$. Also, $\psi_{N,u} \le N^u$. From these inequalities if k is odd, then $\lfloor k/2\rfloor + 1 = k/2 - 1/2$, so

$$\lim_{N\to\infty} E\langle x^k, D_N\rangle = 0.$$

For k even, $\lfloor k/2\rfloor + 1 = k/2 + 1$, so

$$\lim_{N\to\infty} E\langle x^k, D_N\rangle = p^{\frac{k}{2}} \sum_{w\in\mathbb{W}_{k,k/2+1}} \prod_{e\in E_w^c} E\tilde{Z}_{1,2}^{N_w^e} \prod_{e\in E_w^s} E\tilde{Y}_{1,1}^{N_w^e}.$$

Any $w \in \mathbb{W}_{k,k/2+1}$ is connected and $|V_w| = k/2+1$. This tells us that $|E_w| \ge k/2$. Because $N_e^w \ge 2$ for all $e \in E_w$, $|E_w| \le k/2$, which tells us that $|E_w| = k/2$. This tells us that $|N_e^w| = 2$ for all $e \in |V_w|$.

A tree is a connected graph with no cycles. Any graph G with $|E_G| = |V_G| - 1$ is a tree, so each $w \in \mathbb{W}_{k,k/2+1}$ is a tree. Because these w have no cycles, $E_w^s = \emptyset$. This tells us that for even k

$$\lim_{N\to\infty} E\langle x^k, D_N\rangle = p^{\frac{k}{2}} |\mathbb{W}_{k,k/2+1}|.$$

A Bernoulli walk of length ℓ is an integer valued sequence $\{B_n\}_{n=1}^{\ell}$ with $|B_i - B_{i-1}| = 1$. A Dyck path of length ℓ is a non-negative Bernoulli walk of length ℓ with $B_1 = B_\ell = 0$. To complete the proof and as shown in [1] on page 15, for k even we construct a bijection between $\mathbb{W}_{k,k/2+1}$ and the Dyck paths of length k. Also, shown in [1] on page 8, there are $C_{\frac{k}{2}}$ Dyck paths of length k for k even. This completes the proof of condition (i)

Proof of Condition (ii): By Chebyshev's inequality,

$$\mathbb{P}\left(\left|\langle x^k, D_N\rangle - E\langle x^k, D_N\rangle\right| > \epsilon\right) \le \frac{1}{\epsilon^2}\left(E[\langle x^k, D_N\rangle^2] - E[\langle x^k, D_N\rangle]^2\right).$$

We will show that

$$\lim_{N\to\infty} E(\langle x^k, D_N\rangle^2) - E\big(\langle x^k, D_N\rangle\big)^2 = 0.$$

Using the same definitions for $\boldsymbol{i}$ and $T_{\boldsymbol{i}}^N$ as in the proof of Lemma 2.1,

$$E(\langle x^k, D_N\rangle^2) - E\big(\langle x^k, D_N\rangle\big)^2 = \frac{1}{N^2}\sum_{\boldsymbol{i},\boldsymbol{i}'} ET_{\boldsymbol{i}}^N T_{\boldsymbol{i}'}^N - ET_{\boldsymbol{i}}^N ET_{\boldsymbol{i}'}^N.$$

An N-sentence is a finite sequence of N-words $r = a_1, a_s \ldots, a_n$. The support of the sentence supp$r = \cup_{j=1}^{n}$ supp(a_j), and the weight wt$(r) = |\text{supp}(r)|$. Let each word $a_i = b_1^i b_2^i \ldots b_{\ell(a_i)}^i$. Then the undirected graph generated by sentence s G_s has vertices $V_r =$ supp(r) and edges $E_r = \left\{\{b_k^i, b_{k+1}^i\} : k = 1,2, \ldots, \ell(a_i); i = 1,2, \ldots, n\right\}$. The self edges of the graph generated by r $E_r^s = \{e \in E_r : e = \{u,u\}, u \in V_r\}$, and the connecting edges $E_r^c = E_r - E_r^s$. Note that the graph need not be connected. Let N_r^e be the number of times the graph G_r crosses edge e.

With these definitions and notation, we denote by $r_{\boldsymbol{i},\boldsymbol{i}'}$ the two-word N-sentence $w_{\boldsymbol{i}}, w_{\boldsymbol{i}'}$. Then

$$\begin{aligned} &ET_{\boldsymbol{i}}^N T_{\boldsymbol{i}'}^N - ET_{\boldsymbol{i}}^N T_{\boldsymbol{i}'}^N \\ &= \frac{1}{N^k}\left[\prod_{e\in E^c_{r_{i,i'}}} E\tilde{Z}_{1,2}^{N^e_{r_{i,i'}}} \prod_{e\in E^s_{r_{i,i'}}} E\tilde{Y}_{1,1}^{N^e_{r_{i,i'}}}\right. \\ &\left. - \prod_{e\in E^c_{w_{\boldsymbol{i}}}} E\tilde{Z}_{1,2}^{N^e_{w_{\boldsymbol{i}}}} \prod_{e\in E^s_{w_{\boldsymbol{i}}}} E\tilde{Y}_{1,1}^{N^e_{w_{\boldsymbol{i}}}} \prod_{e\in E^c_{w_{\boldsymbol{i}'}}} E\tilde{Z}_{1,2}^{N^e_{w_{\boldsymbol{i}'}}} \prod_{e\in E^s_{w_{\boldsymbol{i}'}}} E\tilde{Y}_{1,1}^{N^e_{w_{\boldsymbol{i}'}}}\right]. \end{aligned}$$

If $N^e_{r_{i,i'}} = 1$ for any $e \in E_{r_{i,i'}}$, then $ET_{\boldsymbol{i}}^N T_{\boldsymbol{i}'}^N - ET_{\boldsymbol{i}}^N T_{\boldsymbol{i}'}^N = 0$. If $E_{w_i} \cap E_{w_{i'}} = \emptyset$, then $ET_{\boldsymbol{i}}^N T_{\boldsymbol{i}'}^N - ET_{\boldsymbol{i}}^N T_{\boldsymbol{i}'}^N = 0$. Then we only need to consider $r_{\boldsymbol{i},\boldsymbol{i}'}$ where $N^e_{r_{i,i'}} \geq 2$ for all $e \in E_{r_{i,i'}}$ and $E_{w_i} \cap E_{w_{i'}} \neq \emptyset$.

Two sentences $r_{\boldsymbol{i},\boldsymbol{i}'}$ and $r_{\boldsymbol{j},\boldsymbol{j}'}$ are equivalent if there is a bijection from one to the other. Therefore, they generate the same graph, and

$$ET_{\boldsymbol{j}}^N T_{\boldsymbol{j}'}^N - ET_{\boldsymbol{j}}^N T_{\boldsymbol{j}'}^N = ET_{\boldsymbol{j}}^N T_{\boldsymbol{j}'}^N - ET_{\boldsymbol{j}}^N T_{\boldsymbol{j}'}^N.$$

There are exactly $\psi_{N,u}$ N-sentences equivalent to any given N-sentences of weight u. Let $\mathbb{W}_{k,u}^2$ be the set of representatives for the equivalence classes of N-sentences r of weight

u consisting of two closed N-words w_1, w_2 of length $k+1$ with $N_r^e \geq 2$ for all $e \in E_r$ and $E_{w_1} \cap E_{w_2} \neq \emptyset$. Then

$$E(\langle x^k, D_N \rangle^2) - E\big(\langle x^k, D_N \rangle\big)^2$$
$$= \sum_{u=1}^{2k} \frac{\psi_{N,u}}{N^{k+2}} \sum_{r=(w_1,w_2)\in \mathbb{W}_{k,u}^2} \left[\prod_{e \in E_{r_{i,i'}}^c} E\tilde{Z}_{1,2}^{N_r^e} \prod_{e \in E_{r_{i,i'}}^s} E\tilde{Y}_{1,1}^{N_r^e} \right.$$
$$\left. - \prod_{e \in E_{w_i}^c} E\tilde{Z}_{1,2}^{N_{w_i}^e} \prod_{e \in E_{w_i}^s} E\tilde{Y}_{1,1}^{N_{w_i}^e} \prod_{e \in E_{w_{i'}}^c} E\tilde{Z}_{1,2}^{N_{w_{i'}}^e} \prod_{e \in E_{w_{i'}}^s} E\tilde{Y}_{1,1}^{N_{w_{i'}}^e} \right].$$

$\frac{\psi_{N,u}}{N^{k+2}} \leq N^{u-k-2}$. In the limit we only need to consider $u \geq k+2$. For each $r \in \mathbb{W}_{k,u}^2$ $E_{w_1} \cap E_{w_2} \neq \emptyset$, so the graph G_r is a connected graph with u vertices. The graph has at most k edges since $N_r^e \geq 2$. It is clearly impossible to have a connect graph with $u \geq k+2$ vertices and at most k edges, so $\mathbb{W}_{k,u}^2 = \emptyset$. Thus,

$$\lim_{n\to\infty} E(\langle x^k, D_N \rangle^2) - E\big(\langle x^k, D_N \rangle\big)^2 = 0.$$

2.3. Proof of Theorem 1.3

The proof of condition (ii) in for Theorem 1.2 in Section 2.2 still applies in this situation. For the proof of Lemma 1.2, we know that if k is odd, then

$$\lim_{N\to\infty} E\langle x^k, D_N \rangle = 0.$$

For k even, we see that,

$$\lim_{N\to\infty} \left| E\langle x^k, D_N \rangle - E\langle x^k, \boldsymbol{\sigma_1} \rangle \right| \leq \lim_{N\to\infty} \left| E\langle x^k, L_N \rangle - E\langle x^k, D_N \rangle \right| + \lim_{N\to\infty} \left| E\langle x^k, L_N \rangle - E\langle x^k, \boldsymbol{\sigma_1} \rangle \right|.$$

The second term on the right-hand side goes to zero by Wigner's Semicircle Law. Then we only need to consider the first term.

$$\left| E\langle x^k, L_N \rangle - E\langle x^k, D_N \rangle \right| = \left| \frac{1}{N^{k/2+1}} \sum_{i_1,..,i_k=1} [1 - 1_N(i_1,i_2) \dots 1_N(i_k,i_1)] E X_N(i_1,i_2) \dots X_N(i_k,i_1) \right|.$$

By the proof presented in Section 2.3, when ignoring the $[1 - 1_N(i_1, i_2) \dots 1_N(i_k, i_1)]$ part of each term, we know that there are an order $C_{\frac{k}{2}} N^{k/2+1}$ terms $EX_N(i_1, i_2) \dots X_N(i_k, i_1)$ in the summation that survive when N goes to ∞, and they are all equal to 1. For each term we refer to the $X(i_j, i_{j+1})$ as slot j, with $X(i_k, i_1)$ being slot k. Then entry $X_N(s, t)$ appears each slot j an order of $C_{\frac{k}{2}} N^{k/2-1}$ times. Then each entry appears in at most $k \cdot C_{\frac{k}{2}} \cdot N^{k/2-1}$.

We see that $[1 - 1_N(i_1, i_2) \dots 1_N(i_k, i_1)] = 1$ whenever at least one of the $1_N(i_j i_{j+1})$ is equal to zero. These are the distortions, and there are $\theta_j(1_N)$ of them. Because the $\theta_j(1_N)$ distortions appear in at most $k \cdot C_{\frac{k}{2}} \cdot N^{k/2-1}$ terms,

$$\left| E\langle x^k, L_N\rangle - E\langle x^k, D_N\rangle \right| \le \theta(1_N) \cdot k \cdot C_{\frac{k}{2}} \cdot \frac{N^{k/2-1}}{N^{k/2+1}} = \frac{\theta_j(1_N)}{N^2} \to 0.$$

This completes the proof of conditions (i), so Theorem 1.3 holds.

3. The General Case of Theorem 1.2

Theorem 1.2 assumes that $\theta_j(1_N) = pN$ for all $j \in \mathbb{N}$. We have also studied when this is not the case, and $\theta_j(1_N)$ can be anything for each j. We are unable to prove any strong theorems in this case but are able to find some moments for the distribution of the eigenvalues. By the proof of Theorem 1.2, we still know that the odd moments are zero when taking the limit as N goes to infinity. For the even moments we know that

$$\lim_{N\to\infty} E\langle x^2, D_N\rangle = \lim_{N\to\infty} \sum_{i_1, i_2 = 1}^{N} 1_N(i_1, i_2).$$

$$\lim_{N\to\infty} E\langle x^4, D_N\rangle = \lim_{N\to\infty} 2 \sum_{i_1, i_2, i_3 = 1}^{N} 1_N(i_1, i_2) 1_N(i_2, i_3) = \sum_{j=1}^{N} \theta_j^2.$$

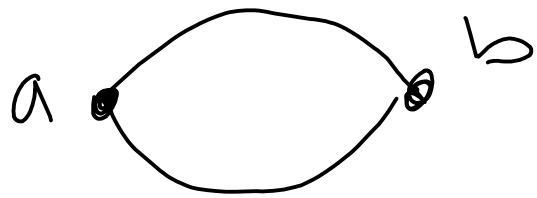
$$\lim_{N\to\infty} E\langle x^6, D_N\rangle = \lim_{N\to\infty}\left[2\sum_{i_1,i_2,i_3=1}^{N} 1_N(i_1,i_2)1_N(i_2,i_3)1_N(i_2,i_4)\right.$$

$$\left.+\,3\sum_{i_1,i_2,i_3,i_4=1}^{N} 1_N(i_1,i_2)1_N(i_2,i_3)1_N(i_3,i_4)\right]$$

$$= \lim_{N\to\infty}\left[2\sum_{j=1}^{N}\theta_j^2 + 3\sum_{i,j,=1}^{N}\theta_i 1_N(i,j)\theta_j\right].$$

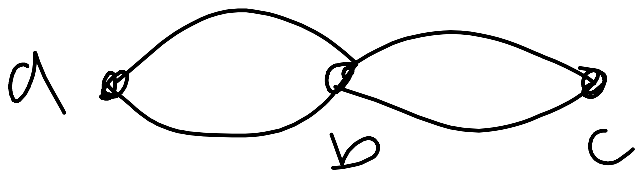
We can find similar expressions for higher moments. We find these expressions by writing out all the possible graphs of closed words length $\frac{k}{2} + 1$ where each edge in the graph are crossed twice, and then counting how many times each graph can be generated by a word. From the proof of Theorem 1.2, these are the only terms that are contribute to the sum in the limit. The graphs and counts for the for the moment equations shown above are listed on the next page.

Closed words of length 3 and weight 2:



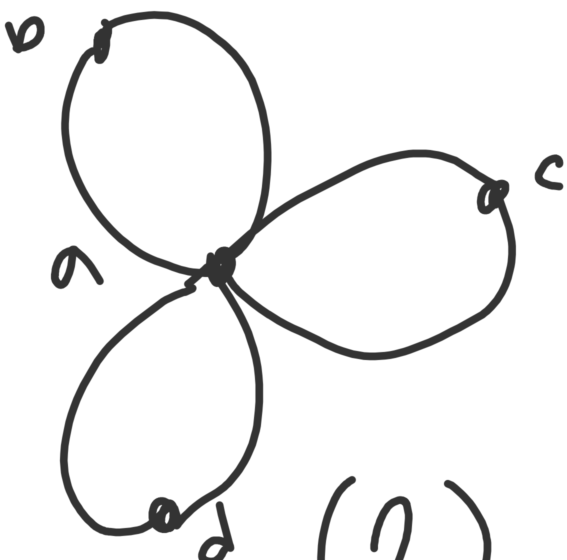
(1)

Closed words of length 5 and weight 3:

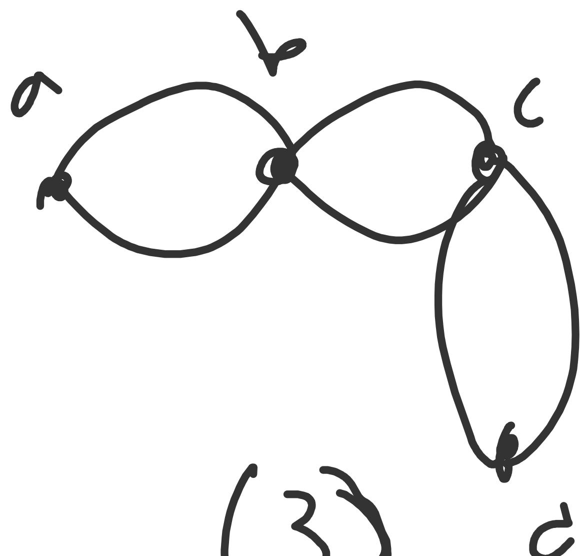


(2)

Closed words of length 7 and weight 4:



(2)



(3)

References

[1] G. Anderson, A. Guionnet, O. Zeitouni, *An Introduction to Random Matrices*, Chapter 2 (2010)

[2] O. Beckwith, V. Luo, S. Miller, K. Shen, N. Triantafillou, *Distribution of Eigenvalues of Weighted, Structured Matrix Ensembles*, Integers **15** (2015), paper A21, 28 pages.

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

[4] P. Burkhardt, P. Cohen, J. Dewitt, M. Hlavacek, S. Miller, C. Sprunger, Y.N. Truong Vu, R. Van Peski, K. Yang, *Random Matrix Ensembles with Split Limiting Behavior*, Random Matrices: Theory and Applications (2018) **7** No. 3.