MATH 510 - topics in Analysis: Random Matrix Theory

Chi Li

Contents

This is a set of notes I took in my final quarter in Northwestern University. I am in the process of digitizing this. Feel free to email me for any suggestions.

0 Brief History of Random Matrix Theory

In 1928 Wishart was looking at the eigenvalues of the a covariance matrix. Let M be an $n \times m$ matrix for m independent observations of n (centered)variables, and consider the covariance matrix MM^T. As we let $n, m \to \infty$ and $n/m \to \alpha$, then the histogram of eigenvalues converges to some distribution (we will make this precise). Around the same time (1930) Eugene Wigner was looking at the energy spectrum of the nuclei of heavy atoms (think emission spectrum). He argues that the behaviour of these atoms are so complex that we might model them as operating under a random Hamiltonian. Under this assumption, he proves the eigenvalues of symmetric (Hermitian) matrices also follow some distribution, known as the semicircle law.

My first encounter with RMT is through analytic number theory. In 1970, Hugh Montgomery was studying the spacing of zeros of the Riemann Zeta function up to height T on the real-half line

$$\#\{\gamma_i - \gamma_{i-1} : b/\log T \le \gamma_i - \gamma_{i-1} \le a/\log T\}$$

Adjusted for the number of zeros $O(T \log T)$, the distribution of the spaces empirically follow the sine kernel

$$\int_a^b 1 - \left(\frac{\sin(\pi x)}{\pi x}\right)^2 dx$$

This is exactly the pair correlation of the eigenvalues of a random Hermitian Matrices. Montgomery and Dyson shared this discovery, and after, this conjecture is known as pair correlation conjecture.

We will see more recent applications of RMT such as Last Passage Percolation (LPP) and the KPZ universality, but we'll get there when we get there.

1 Wigner Semicircle Law

We state Wigner's Law

Theorem 1.1 (Wigner's Semicircle Law)

Let $M_n = \{M_{i,j}\}$ be an $n \times n$ matrix, such that $M_{i,j} = M_{j,i}$ and $M_{i\leq j}$ are IID with

$$\mathbb{E}[M_{i,j}] = 0, \mathbb{E}[M_{i,j}^2] = 1.$$

Let $\lambda_1 \leq ... \leq \lambda_n$ be the eigenvalues of M_n, and

$$L_n \stackrel{\text{def}}{=} \frac{1}{n}\sum_{i=1}^{n} \delta_{\lambda_i/\sqrt{n}}$$

Then as $n \to \infty$,

$$L_n \to \sigma(x) \stackrel{\text{def}}{=} \frac{1}{2\pi}\sqrt{4-x^2}dx$$

weakly almost surely and in L^p.

We shall prove a weaker statement. We assume that $\mathbb{E}[|M_{i,j}|^k] < \infty$, and prove the convergence in probability. That is,

$$\int f dL_n \to \int f d\sigma$$

in probability for any continuous f.

Before we head to the proof, we can generalize the statement for

$$\mathbb{E}[M_{i,j}] = a, \mathbb{E}[M_{i,j}^2] = b.$$

The second is easy, we can just scale M_n uniformly by $\sqrt{b}$ to get variance 1. For the shift, we can isolate $M_{i,j} = a + \tilde{M}_{i,j}$. So that $\tilde{M}_n$ is a matrix with known eigenvalue distribution. Now the $n \times n$ matrix

$$\begin{bmatrix} a & a & ... & a \\ \vdots & \ddots & & \\ a & ... & & a \end{bmatrix}$$

is rank 1, so the resulting matrix gives interlacing of eigenvalues

$$... \leq \tilde{\lambda}_{i-1} \leq \lambda_{i-1} \leq \tilde{\lambda}_i \leq \lambda_i \leq ...$$

So everything is controlled except for possible the top and bottom eigenvalues. This is fine as $\delta_{\lambda_i/\sqrt{n}}/n$ is negligible as $n \to \infty$.

Moment Matching Method

The proof for Wigner's semicircle law is known as the 'moment method' and is deconstructed as follows:

Let $\epsilon > 0$, we want

$$P(|\int f dL_n - \int f\sigma dx| > \epsilon) \to 0.$$

We first show this for polynomials, then apply Weierstrauss for a density argument.

Lemma 1.2 (Moment Matching)

Let $k \geq 1$. We have

$$\mathbb{E}[\int x^k dL_n] \to \int x^k \sigma(x) dx \stackrel{\text{def}}{=} \tilde{M}_k.$$

Next, we have to turn this expected value to statements about probability, thus we need a concentration lemma:

Lemma 1.3 (Concentration)

We have for $k \geq 1$,

$$P(|\int x^k dL_n - \int x^k \sigma dx| > \epsilon) \to 0.$$

Proof of (weaker version of) Semicircle Law with the Lemmas. Let f be continuous and bounded. We approximate f with p_n in the interval $[-B, B]$, for $B \geq 2$ to be determined later, which is the support of $\sigma(x)dx$ Then we have

$$\begin{aligned} &\int f dL_n - \int f \sigma dx \\ =&(\int f d_n - \int p_n dL_n) + (\int p_n dL_n - \int p_n \sigma dx) + (\int p_n \sigma dx - \int f \sigma dx) \end{aligned}$$

The second term goes to 0 in probability by the concentration lemma. In the third term, we can restrict the integration to $[-2, 2]$ and it goes to 0 almost surely as p_n converges to f. We thus need to show that the first term goes to 0 in probability. On the domain $[-B, B]$, this is taken care of by the fact that p_n approximates f uniformly.

Outside $[-B, B]$, we want to show that

$$P(\int |x|^k \mathbb{I}_{|x|>B} dL_n > \epsilon) \to 0$$

for any ϵ. This will take care of the polynomial integral, and will also take care of the integral in f as f is bounded, thus is bounded above by some C.

To show this, we apply Markov's ineqaulity to get

$$\begin{aligned} &P(\int |x|^k \mathbb{I}_{|x|>B} dL_n > \epsilon) \\ \leq&\frac{1}{\epsilon} E\left[\int |x|^k \mathbb{I}_{|x|>B} dL_n\right] \\ \leq&\frac{1}{\epsilon B^k} E\left[\int |x|^{2k} \mathbb{I}_{|x|>B} dL_n\right]. \end{aligned}$$

Taking the limsup as $n \to \infty$,

$$\limsup_{n\to\infty} P(\int |x|^k \mathbb{I}_{|x|>B} dL_n > \epsilon) \leq \frac{\tilde{M}_{2k}}{\epsilon B^k}.$$

This bound works for all k. We will see in the computation that $\tilde{M}_{2k} \leq 4^k$, so taking $B = 5$, we have that the first term is increasing in k but the last term is decreasing in k to 0. Thus the only way this bound works for all k is that the probability also converges to 0.

Remark. *With a little more work we can prove the convergence in L^p/almost sure convergence.*

Moment calculation of distribution

We now complete the lemmas of the moment matching method.

Proposition 1.4

We have

(a) $\tilde{M}_{2k+1} = 0$.

(b) $\tilde{M}_{2k} = \frac{1}{k+1}\binom{2k}{k}$.

The actual computation of this moment of the semicircle law is unenlightening. The more experienced combinatorist will recognize $\tilde{M}_{2k}$ as the k-th Catalan number C_k. There are a few meanings of this C_k. For one, it represents the number of Dyck paths of length $2k$. That is, the number of staircase walks from $(0,0)$ to (k,k) that lie in the upper diagonal. It also represents the number of rooted planar trees with k edges (and $k+1$ nodes). This is because we can identify a tree with its Euler tour as a Dyck path (up if going to a child, right if going to parent).

We give a quick explaination for calculating the number of Dyck paths. The number of paths going from $(0,0)$ to (k,k) without the restriction is $\binom{2k}{k}$. Now we count the number of invalid paths. Suppose a path is invalid, then take the last time it passes the diagonal, then flip the directions of the path right and up. This leads to a path from $(0,0)$ to $(k+1,k-1)$. Similarly, this reflection trick turns every path from $(0,0)$ to $(k+1,k-1)$ to an invalid path from $(0,0)$ to (k,k).

The number of invalid paths is thus $\binom{2k}{k-1}$. So the number of Dyck paths is

$$\binom{2k}{k} - \binom{2k}{k-1} = \frac{1}{k+1}\binom{2k}{k}.$$

More importantly, we have a trivial bound

$$C_k \leq 4^k$$

just by considering the binomial expansion.

We now compute the moments of L_n, which is the key of the proof.

Proof of lemma 1.2. The part

$$\int x^k \sigma(x) dx = \tilde{M}_k$$

is left as an exercise to drain your time on a weekend. We compute the first part

$$\int x^k dL_n.$$

Recall the definition of L_n, which is the discrete measure for the eigenvalues scaled by $1/\sqrt{n}$. This means that the integral of x^k across this measure is equal to

$$\begin{aligned}&\sum_i \frac{1}{n}\left(\frac{\lambda_i}{\sqrt{n}}\right)^k \\ =&n^{-k/2-1}\ \mathrm{Tr}\ \mathrm{M^k}\end{aligned}$$

The trace of M^k can be expanded in a k-sum as follows

$$M_k = \sum_I M_{i_1,i_2} M_{i_2,i_3} ... M_{i_{k-1},i_k} M_{i_k} M_{i_1},$$

where the sum over I runs over all

$$I = i_1 i_2 i_3 ... i_k, \quad i_j = 1, 2, 3, ..., n$$

By the linearity of expectation we just need to know

$$n^{-k/2-1} \sum_I E\left[M_{i_1,i_2} M_{i_2,i_3} ... M_{i_{k-1},i_k} M_{i_k} M_{i_1}\right].$$

Now here comes Wigner's insight: imagine a fully connected undirected graph of n nodes and $n(n+1)/2$ including self edges. Then the edges $(i_1, i_2)...(i_{k,1})$ describe a loop in the graph. Moreover, each edge has to appear at least twice in this loop for the corresponding expectation contribution to be non-zero. Suppose an edge appears only once, then by the independence of each $M_{i,j}$ we can factor $\mathbb{E}[M_{i,j}] = 0$ out of the expectation.

Let us break the expectation into two parts. In the extreme case where the number of nodes involved of the graph is maximal, the nodes and edges of the loop form a tree of $k/2 + 1$ nodes, and I describes an Euler tour. This is known as a Wigner tree. We have argued that there are $C_{k/2}$ of these trees if we do not label the nodes, so up to relabelling of the nodes there are

$$C_{k/2} P^n_{k/2+1} = C_{k/2} n^{k/2+1}(1 + o(1))$$

of these trees. Finally, each of the corresponding expectations factors into some

$$\mathbb{E}[M^2_{i_1,i_2}] ... \mathbb{E}[M^2_{i_l,i_{l+1}}] = 1$$

so the contribution from the trees are

$$n^{-k/1-1} C_{k/2} n^{k/2+1}(1 + o(1)) = C_{k/2}(1 + o(1)).$$

We now consider the case where the number of nodes is less than $k/2 + 1$. This takes care of the remaining k is odd case, and the remaining terms of the even k.

The argument is very simple. Fix a number of nodes $q < k/2 + 1$. Then there are only some finite K_q number of different unlabelled loops with that number of nodes. The number of relabellings of these loops is at most n^q. So the contribution is of the order

$$n^{-k/2-1+q} K_q \to 0$$

as $n \to \infty$.

We now prove the concentration inequality lemma.

Proof of 1.3. There are two ways to approach a concentration inequality. Either use Markov's inequality, or some variant of Markov's inequality. In our case, we use Chebyshev's inequality to get

$$\mathbb{P}(|\int x^k dL_n - \mathbb{E}\int x^k dL_n| > \epsilon) \leq \frac{1}{\epsilon^2} \mathrm{Var} \int \mathrm{x^k dL_n}.$$

Because the expected value converges to that of the semicircle law, showing this goes to zero proves the statement. We have

$$\begin{aligned}\mathrm{Var}\int \mathrm{x^k dL_n} &= \frac{1}{n^{k+2}}(\mathbb{E}[(\mathrm{trM^k})^2]-\mathbb{E}[(\mathrm{trM^k})]^2)\\ &= n^{-k-2}\left(\mathbb{E}[\sum_{I,J}T_IT_J]-\sum_{I,J}\mathbb{E}[T_I]\mathbb{E}[T_J]\right)\end{aligned}$$

where

$$T_I \stackrel{\text{def}}{=} M_{i_1,i_2}...M_{i_k,i_1}$$

and similarly for J for each sequence I, J of length k. By linearity of expectations, we can consider the contribution for each pair I and J separately. If I and J do not share common edges, then the expectation $\mathbb{E}[T_IT_J]$ factors into $\mathbb{E}[T_I]\mathbb{E}[T_J]$ which gives zero contribution. Else, we would have at least one shared edge and (since each edge has at least 2 multiplicity in each of I and J) there can at most be $k/2+k/2-1=k-1$ edges thus k vertices. The number of relabels of these k-vertex graphs is bounded by $O(n^k)$ depending on the size of the higher-order moments which gets dominated by $O(n^{-k-2})$ term and goes to zero as $n\to\infty$.

Example 1.5

Let M be an $n\times m$ matrix of all iid variables of mean 0 variance 1. Try to prove the limit distrbution for eigenvalues of MM^T as $n,m\to\infty$ and $n/m\to\alpha$.
Or whatever, just read the next section.

2 Marchenko-Pastur Distribution

This is an analogous theorem for covariance matrices. If the Semicircle Law is the Gaussian distribution/Central Limit Theorem of random matrix theory, then this is the Poisson distribution of random matrix theory.

Definition 2.1 (Marchenko-Pastur Distribution)

Fix $0<y$. Let $a=(1-\sqrt{y})^2, b=(1+\sqrt{y})^2$. We define the **Marchenko-Pastur Distribution** as

$$\sigma_y(x)\stackrel{\text{def}}{=}\sigma(x)\stackrel{\text{def}}{=}\begin{cases}\frac{1}{2\pi xy}\sqrt{(b-x)(x-a)}, & \mathrm{if a\le x\le b}\\ 0, & \text{otherwise.}\end{cases}$$

Theorem 2.2 (Marchenko-Pastur)

Let M be an $n\times m$ random matrix with i.i.d. entries such that

$$\mathbb{E}[M_{i,j}]=0, \mathbb{E}[M_{i,j}^2]=1, \mathbb{E}[M_{i,j}^k]\le\infty\ \forall k.$$

Let $n/m\to y>0$ as $n\to\infty$, and

$$d\tilde{\mu}\stackrel{\text{def}}{=}\sigma_y(x)dx.$$

Define $\lambda_1, ..., \lambda_n$ to be the eigenvalues of the matrix MM^T counting multiplicity. Then the measure

$$L_n \stackrel{\text{def}}{=} \frac{1}{n} \sum_i \delta_{\lambda_i/m} \to \tilde{\mu} + \max(0, 1 - y^{-1})\delta_0$$

weakly in probability.

Structure of proof

The following lemmas have analogous forms for Wigner's Semicircle Law's moment method proof.

Lemma 2.3 (Moments of Distribution)

We have

$$\int \sigma_y(x)dx = \min(1, y^{-1}).$$

For $k \geq 1$, We have

$$\int x^k \sigma_y(x)dx = \sum_{r=0}^{k-1} \frac{y^r}{r+1} \binom{k}{r} \binom{k-1}{r}.$$

This is Lemma 3.1 from [**?**]. The proof is mainly computational and relies on Vandermonde's identity[1].

Remark. *Adding $(1 - y^{-1})\delta_0$ changes the $0 - th$ moment to 1 and does not affect any of the other moments. There are two ways to see that this point mass term is required. First, this adjustment is needed to make the distribution integrate to 1. Another way to see this is when we have $y > 1$, $n > m$ and there are at least $n - m$ eigenvalues of 0 in MM^T just by considering the nullspace of M^T. This proportion comes up as a point mass of*

$$\frac{1}{n}(n - m) \to 1 - y^{-1}$$

at zero.

Lemma 2.4 (Trivial Bound of moment)

We have

$$\sum_{r=0}^{k-1} \frac{y^r}{r+1} \binom{k}{r} \binom{k-1}{r} \leq \max(y, 1)^k 8^k.$$

Proof. This is a generous bound using $\binom{a}{b} \leq 2^a$.

Lemma 2.5 (Moments of Eigenvalues)

Trivially we have $\mathbb{E}[\sum_i 1/n] = 1$ (with the convention $0^0 = 1$).

[1] I would have tried to solve B5 on the Putnam this year if I had known any of these techniques.

For each $k \geq 1$, We have

$$\mathbb{E}\Big[\int x^k dL_n\Big] \to \sum_{r=0}^{k-1} \frac{y^r}{r+1}\binom{k}{r}\binom{k-1}{r}$$

as $n, m \to \infty$.

Lemma 2.6 (Concentration)

For any $k \geq 1, \epsilon > 0$,

$$\mathbb{P}\Bigg(\Big|\int x^k dL_n - \mathbb{E}\int x^k dL_n\Big| > \epsilon\Bigg) \to 0.$$

Proof of Theorem 2.2 assuming the previous four Lemmas. Let $\mu \overset{\text{def}}{=} \tilde{\mu} + \Big(1 - \frac{1}{y}\Big)_+ \delta_0$. Then all $k-th$ moments of L_n converge to the $k-th$ moments of μ.

Now let $\epsilon > 0$. We want to show

$$\mathbb{P}\left(\Big|\int f dL_n - \int f d\mu\Big| > \epsilon\right) \to 0.$$

First, by Markov,

$$\mathbb{P}\left(\Big|\int |x^k| \mathbb{I}_{x>B} dL_n\Big| > \epsilon\right) \leq \frac{1}{\epsilon}\mathbb{E}\left[\int |x^k|\mathbb{I}_{x>B} dL_n\right] \leq \frac{1}{\epsilon B^k}\mathbb{E}\left[\int |x^{2k}|\mathbb{I}_{x>B} dL_n\right].$$

Taking limsup,

$$\limsup_{n\to\infty} \mathbb{P}\left(\Big|\int |x^k| \mathbb{I}_{x>B} dL_n\Big| > \epsilon\right) \leq \frac{\max(y,1)^k 8^k}{\epsilon B^k}.$$

We take B large enough such that $B > 8(y+1)$. Since the right side is increasing in k and the left side is decreasing in k, and the bound works for all k, the only way this can happen is for

$$\mathbb{P}\left(\Big|\int |x^k| \mathbb{I}_{x>B} dL_n\Big| > \epsilon\right) \to 0.$$

If needed, we extend the range $[-B, B]$ so that it includes (a, b). Applying Weierstrauss approximation for the function f in this compact region will give weak convergence in probability. 🌸

Calculation of moments of L_n

Let $k \geq 1$, and $A \overset{\text{def}}{=} MM^T \overset{\text{def}}{=} [a_{i,j}]$. Then $a_{i,j} = \sum_{l=1}^m M_{i,l}M_{j,l}$. Then we have

$$\begin{aligned}
\mathbb{E}\left[\int x^k dL_n\right] =& \mathbb{E}\Big[\frac{1}{n}\sum_i \frac{\lambda_i^k}{m^k}\Big] \\
=& \frac{1}{nm^k}\mathbb{E}\Big[\mathrm{TrA^k}\Big] \\
=& \frac{1}{nm^k}\mathbb{E}\Big[\sum_I S_I],
\end{aligned}$$

where for each $I = i_1 i_2 ... i_k$ a string of k integers each between 1 to n

$$\begin{aligned} S_I =& a_{i_1,i_2} a_{i_2,i_3} ... a_{i_k,i_1} \\ =& \sum_J M_{i_1,j_1} M_{i_2,j_1} M_{i_2,j_2} M_{i_3,j_2} ... M_{i_k,j_k} M_{i_1,j_k} \\ \stackrel{\text{def}}{=}& \sum_J T_{I,J}, \end{aligned}$$

summed over each $J = j_1 ... j_k$ strings of k integers each between 1 to m inclusive. By linearity of expectations, we have

$$\mathbb{E}\left[\int x^k dL_n\right] = \frac{1}{nm^k} \sum_{I,J} \mathbb{E}\Big[T_{I,J}\Big].$$

Proposition 2.7

In the summation over I and J, we can identify this with a unique bipartite graph with nodes i_l and j_l, and with $2k$ edges $i_l j_l$, $i_l j_{l-1}$. Moreover, as in Wigner's Semicircle Law,

- If at least one edge has multiplicity 1, the contribution in expectation in zero.
- We can consider graphs with exactly $k+1$ nodes. All other graphs have negligible contribution.

Proof. The proof to the first statement is the same. We can factor out the expectation $\mathbb{E}[M_{i,j}] = 0$ by independence.

For the second statement, we remove zero contributions from graphs with multiplicity one edges. The maximum number of nodes that can possibly have non-zero contribution is $k+1$ by the connectedness of the graph (in which case the skeleton is a tree). Fix $N \leq k$. We will show that the contribution of graphs with N vertices have negligible contribution. For each graph, say there are α vertices coming from I and β vertices coming from J, so up to relabelling the contribution in expectation is

$$C(1+o(1))\frac{n^\alpha m^\beta}{nm^k} = C(1+o(1))\frac{y^{a-1}}{m^{k-N+1}} \to 0,$$

where C is some constant counting the number of graphs of α edges from I and β vertices from J.

We therefore just need to compute

$$\frac{1}{nm^k} \sum_{I,J} \mathbb{E}\Big[T_{I,J}\Big]$$

summed over the I, J such that the corresponding graph is a tree and the contribution in expectation is 1. We split the sum across graphs with $l+1$ vertices coming from I and $k-l$ vertices coming from J for each $0 \leq l \leq k$. Up to relabelling we account for the factor of $(1+o(1))n^{l+1}m^{k-l} = (1+o(1))nm^k y^l$.

Thus we can approximate

$$\frac{1}{nm^k} \sum_{I,J} \mathbb{E}\Big[T_{I,J}\Big] \sim \sum_{l=0}^{k-1} y^l \sum_{\substack{\text{trees with fixed labels} \\ |I|=l+1, |J|=k-l}} 1.$$

This is already suggestive of the form of the k-th moment. To finish off the proof of the lemma, it suffices to prove the following final result.

Proposition 2.8

The number of trees (with fixed labels) with $l+1$ vertices in I and $k-l$ vertices in J is equal to

$$\frac{1}{l+1}\binom{k}{l}\binom{k-1}{l}.$$

Proof. This proof is heavily inspired by Arnab Ganguly's Lecture Notes [**?**], with tweaks to shorten the argument and make parallels to Wigner's case. There are two steps to the argument. First we relate each tree to a sequence (similar to Catalan numbers). Then we count the number of such sequences with such property.

We root the tree at a vertex in I. Similar to Wigner's case, we can relate each tree to a sequence of $2k$ "up" and "down" symbols, such that at any point in the sequence the count of "up" symbols is no less than that of "down" symbols. Moreover, if we follow the path of the sequence, we would be at I vertices at any odd position, and J vertices at even positions (before completing the "up" or "down" operation in the position). Since the sequence creates new I vertices exactly when there is an "up" symbol at an even position, the number of "up" symbols at even positions is exactly l. By parity, the number of "down" symbols at odd positions is also exactly l. We now have a one-to-one correspondence with the trees and sequences of $2k$ symbols.

To count the number of such sequences, it is helpful to fix the last symbol to be "down" (or else sequence is invalid anyway). The reason for this choice will be made clear in the reflection argument. The number of sequences of length $2k-1$ with l "up" symbols in even positions and l "down" symbols in odd positions is

$$\binom{k-1}{l}\binom{k}{l}.$$

We now count the number of invalid sequences. For each invalid sequence, find the last time it reaches -1 (i.e. the number of downs is exactly 1 greater than the number of ups). By parity, this corresponds to a "down" symbol at an odd position $2\alpha-1$. We flip the remaining positions in pairs $(2i, 2i+1)$, leaving position $2k$ unchanged, according to the following rule:

- $(\text{up}, \text{up}) \to (\text{down}, \text{down})$
- $(\text{up}, \text{down}) \to (\text{up}, \text{down})$
- $(\text{down}, \text{up}) \to (\text{down}, \text{up})$
- $(\text{down}, \text{down}) \to (\text{up}, \text{up})$

There are a couple observations about this reflection. First, since this subsequence originally connects -1 to 1, the flipped sequence now connects -1 to -3. That is, the original sequence contains two more "up"s than "downs". Next, we consider the number of "up' and "down" symbols in even and odd positions respectively. The second and third lines do not change anything, first line decreases the number of even 'up' symbols and increases the number of odd "down" symbols (vice versa for the fourth line). By a pairing the "up" and "down"s of this subsequence, we must have that this operation increases the number of odd "down" symbols by 1 and decreases even "up" symbols by 1. We thus have $l+1$ "up" symbols split across the k odd positions and $l-1$ "down" symbols split across the $k-1$ even positions.

The number of invalid sequences is thus

$$\binom{k-1}{l-1}\binom{k}{l+1}.$$

Thus the number of valid sequences is

$$\binom{k-1}{l}\binom{k}{l} - \binom{k-1}{l-1}\binom{k}{l+1} = \frac{1}{l+1}\binom{k-1}{l}\binom{k}{l}.$$

Concentration inequality

Proof of Lemma 2.6. Let $\epsilon > 0$. We apply Chebyshev's inequality to

$$\mathbb{P}\left(\Big|\int x^k dL_n - \mathbb{E}\int x^k dL_n\Big| > \epsilon\right) \le \frac{1}{\epsilon^2}\mathrm{Var}\int \mathrm{x^k dL_n},$$

where

$$\begin{aligned}\mathrm{Var}\int \mathrm{x^k dL_n} &= \frac{1}{n^2m^{2k}}\left[\mathbb{E}\Big[\sum_{I,J} T_{I,J}\sum_{I',J'} T_{I',J'}\Big] - \left(\mathbb{E}\Big[\sum_{I,J} T_{I,J}\Big]\right)^2\right] \\ &= \frac{1}{n^2m^{2k}}\sum_{I,J}\sum_{I',J'}\Big[\mathbb{E}\Big[T_{I,J}T_{I',J'}\Big] - \mathbb{E}\Big[T_{I,J}\Big]\mathbb{E}\Big[T_{I',J'}\Big]\Big].\end{aligned}$$

If the skeletons of the graphs of I, J and $I'J'$ do not share edges, then the expectation of the products is the product of expectations by independence. Else they share an edge. This means that the union of the graphs $(I, J), (I', J')$ has at most $(k+1) + (k+1) - 2$ vertices. The number of relabels of these graphs are

$$\le \max(n, m)^{2k},$$

so the contribution from these terms is

$$O((1+y)^{2k}n^{-2})$$

which is negligible as $n \to \infty$.

3 Continuation of Semicircle Law

We can say statements about the top (resp. bottom) eigenvalue of M.

Corollary 3.1: We use the Wigner semicircle law setting in Theorem 1.1. Let $\lambda_{\max}$ be the top eigenvalue of the matrix M. Then for any $\epsilon > 0$ we have

$$\mathbb{P}(\lambda_{\max} < (2-\epsilon)\sqrt{n}) \to 0$$

as $n \to \infty$.

Proof. Suppose not. Then take f be a bump function with support on $[2-\epsilon, 2]$. Then

$$\mathbb{P}(|\int f dL_n - \int f\sigma dx| > \epsilon) \geq \mathbb{P}(\lambda_{\max} < (2-\epsilon)\sqrt{n})$$

for sufficiently small ϵ. This is because conditioned on $\lambda_{\max} < (2-\epsilon)\sqrt{n}$, the integral fdL_n is 0 and σdx is non-zero.

Corollary 3.2: We also have

$$\mathbb{P}(\lambda_{\max} > (2+\epsilon)\sqrt{n}) \to 0.$$

Proof. We use Markov's inequality

$$\begin{aligned}
&\mathbb{P}(\lambda_{\max} > (2+\epsilon)\sqrt{n}) \\
&= \mathbb{P}(\lambda_{\max}^{2k} > (2+\epsilon)^{2k} n^k) \\
&\leq \frac{1}{(2+\epsilon)^{2k}} \mathbb{E}\Big[\Big(\frac{\lambda_{\max}}{\sqrt{n}}\Big)^{2k}\Big] \\
&\leq \frac{N}{(2+\epsilon)^{2k}} \mathbb{E}[\int x^{2k} dL_n] \\
&\leq N \frac{(4)^k + o(1)}{(2+\epsilon)^{2k}}
\end{aligned}$$

This inequality holds for all k, so letting N grow sufficiently slowly as $k \to \infty$ gives the result $\to 0$.

Theorem 3.3 (Bai Yin '88)

$\mathbb{E}[M_{i,j}^4] < \infty$ is a sufficient and necessary condition for

$$\lambda_{\max}/\sqrt{n} \to 2$$

almost surely. (The same is true for Marchenko Pastur distribution, top eigenvalue converges to the edge of the bulk almost surely)

Remark. *To see a non-example, any distribution for $M_{i,j}$ with a heavy polynomial tail.* ***TODO:*** *python code*

Example 3.4

Suppose we have a signal vector

$$X = \{\pm 1\}^n.$$

Its corresponding matrix

$$XX^T$$

has $n-1$ trivial eigenvalues and 1 eigenvalue of $\sqrt{n}$. Now on top this we add a disturbance from a $n \times n$ Wigner matrix M_n to get

$$Y_n = \frac{1}{\sqrt{n}}(\alpha XX^T + M_n),$$

where α is our signal strength. Is it still possible to isolate/estimate X? (Think noise in finance affecting the covariance, can you still get a vector representing the market?)

A simple estimator is to take the top eigenvector of Y. That is,

$$\hat{\Theta}_s(Y_n) \overset{\text{def}}{=} \sqrt{n} \arg \max_{\sigma \in S^{n-1}} \langle y\sigma, \sigma \rangle$$

Theorem 3.5 (Baik-Ben Arous-Péché transition)

Let M also have Gaussian entries $\sim N(0,1)$. Then there is a phase transition for the top eigenvalue and eignevector of Y_n.

- If $\alpha \le 1$, then $\lambda_{\max}(Y_n) \to 2$.
- If $\alpha > 1$ Then $\lambda_{\max}(Y_n) \to \alpha + \alpha^{-1} > 2$.

Moreover, in the regime $\alpha > 1$,

$$\frac{\langle \hat{\Theta}_s(Y_n), X \rangle}{n} \to \sqrt{1 - \frac{1}{\alpha^2}}.$$

This is known as the BBP phase transition.

4 Invariant Ensembles

Let Ω be the space of $N \times N$ symmetric matrices. We construct a measure on this space such that it is invariant under orthogonal/unitary transformations. I.e.

$$P(M) = P(OMO^T)$$

or

$$P(M) = P(UMU^*)$$

for orthogonal O and unitary U respectively.

Definition 4.1

We define the (standard) measure on Ω to be

$$dM \overset{\text{def}}{=} \prod_{i<j} dM_{i,j} dM_{i,j} \prod_i dM_{i,i},$$

where each $dM_{i,j}$ is the standard Lebesgue measure, thus we can define

$$dP(M) = f(M)dM$$

where f is also measureable.

Example 4.2

Notice that the trace is invariant under orthogonal/unitary transformations, so define

$$Q(t) = a_k t^{2k} + \ldots + a_1 t + a_0$$

an even degree polynomial ($a_k > 0$), where a_0 is some normalization constant, and define

$$P(M) = \exp(-\mathrm{Tr}Q(M))dM$$

Example 4.3

We look at a specific example for the unitary ensemble where

$$Q(t) = at^2 + b^t + c$$

A direct calculation would give

$$\begin{aligned}\mathrm{Tr}Q(M) &= a\mathrm{Tr}(M^2) + b\mathrm{T}r(M) + c \\ &= a\sum_{i=1}^{n}\sum_{l=1}^{n} M_{i,l}M_{l,i} + \sum_{i=1}^{n} M_{i,i} + c\end{aligned}$$

so that

$$dP(M) = \exp\Bigg[-(a\sum_{i=1}^{n} M_{i,i}^2 + a\sum_{i<j}\mathrm{Re}(M_{i,j})^2) + \mathrm{Im}(M_{i,j})^2 + b\sum_i M_{i,i} + c\Bigg].$$

If $b = 0$, then this reduces to each of the real and imaginary parts are independent and iid distributed. If we further set $a = 1$, then they are further normal distributed, we call this the Gaussian unitary ensemble (GUE for short). This is a Wigner matrix distribution. It can in fact be proven that the invariant Wigner ensemble is the GUE.

Remark. *If we apply this to orthogonally invariant ensemble, the analogous Wigner case is called the Gaussian orthogonal ensemble. There are higher dimensional Gaussian invariant ensembles such as the sympletic ensemble (GSE).*

The GOE, GUE, and GSE correspond to $\beta = 1, 2, 4$ respectively. We will later explain the definition of β in this context.

Theorem 4.4

We work in the space of Hermitian matrices $M \in \Omega$. Let $\lambda_1, \ldots, \lambda_N$ be the eigenvalues of M. Let $f : \Omega \to \mathbb{R}$ depend only on the eigenvalues of M, and $dP(M) \overset{\text{def}}{=} -\mathrm{Tr}Q(M)dM$ Then

$$\mathbb{E}[f(M)] = \int f(M)dP(M) = \frac{1}{Z_n}\int f(\lambda_1, \ldots, \lambda_N)\exp\left(-\sum_i Q(\lambda_i)\right)\prod_{i<j}(\lambda_i - \lambda_j)^2 d\lambda_1 \ldots d\lambda_N,$$

where Z_n is a noramlization constant. In other words, the density of eigenvalues have joint density

$$\frac{1}{Z_n}\underbrace{\exp\left(\sum Q(\lambda_i)\right)}_{\text{Eigenvalues want to be small}}\underbrace{\prod_{i<j}(\lambda_i-\lambda_j)^2}_{\text{Eigenvalues tend to repel}}$$

. First we consider that eigenvalues are invariant under unitary transformations i.e.

$$\text{tr }\ Q(M)=\text{tr }\ Q(UMU^*)=\sum_i Q(\lambda_i).$$

Thus we decompose

$$M\mapsto(D,\bar{U})$$

where we have D is a diagonal n matrix, $\bar{U}$ represents the class of unitary matrices mod $T^n=(S^1)^n$, and $UDU^*=M$. The mod condition is required, as eigenvectors are defined up to a constant of $e^{i\theta}$. There is a bit of work to how that this is smooth and well defined, but this **should** follow from the implicit function theorem (I have not checked) applied to $(D,U)\mapsto M=UDU^*$.

We hope that

$$dM=\psi(D,U)d\lambda_1...d\lambda_n dU,$$

where $\psi(D,U)$ factorizes separately in the eigenvalues and U. The part in U should integrate to a constant, which we will hide under the Z_n normalization factor. The part in D should be the Vandermonde determinant squared.

We now need to calculate the Jacobian. We use the representation

$$M=(M_{1,1},...,M_{n,n},\operatorname{Re}M_{1,2},\operatorname{Im}M_{1,2},...)$$

and

$$(D,U)=(\underbrace{\lambda_1,...,\lambda_n}_{\text{diagonal entries of }D},\underbrace{p_1,...,p_l}_{l=n^2-n})$$

The calculation is actually a little bit cumbersome, but we have a trick. Notice that only the determinant of the Jacobian is needed, so we fix D_0,U_0, and make the change of variables

$$M\mapsto U_0^*MU_0$$

which does not affect the determinant of the Jacobian, and we calculate the enteries of the Jacobian for the particular value of $M_0=U_0D_0U_0^*$. For the first n entries, we have

$$\frac{\partial}{\partial\lambda_i}U_0^*MU_0=U_0^*U_0\frac{\partial}{\partial\lambda_i}D_0U_0U_0^*=\frac{\partial}{\partial\lambda_i}D_0$$

as U is independent of the eigenvalues.

We also have

$$U_0^*\frac{\partial}{\partial p_k}MU_0=U_0^*\Big(\frac{\partial}{\partial p_k}U\Big)DU^*U_0+U_0^*UD\Big(\frac{\partial}{\partial p_k}U^*\Big)U_0=U_0^*\Big(\frac{\partial}{\partial p_k}U\Big)D+D\Big(\frac{\partial}{\partial p_k}U^*\Big)U_0$$

when evaluated at M.

$$s_k\stackrel{\text{def}}{=}U_0^*\frac{\partial}{\partial p_k}U,$$

since we have

$$0 = \frac{\partial}{\partial p_k} U^* U = \left(\frac{\partial}{\partial p_k} U^*\right) U + U^* \left(\frac{\partial}{\partial p_k} U\right),$$

$$U_0^* \frac{\partial}{\partial p_k} M U_0 = [s_k, D].$$

Thus the whole Jacobian $\frac{\partial M}{\partial(\lambda_k, p_k)}$is in the form

$$\begin{bmatrix} I_n & 0_{n\times l} \\ 0_{l\times n} & \{[s_k, D]\}_{k=1,\dots,l} \end{bmatrix}$$

This we just need to "calculate" the determinant of the bottom right matrix. But we have since D is diagonal, every of the $n^2 - n$ non-zero entries of $[s_k, D]$ is of the form $(\lambda_i - \lambda_j)\alpha$, where α is only dependent of U_0 and k. Therefore each permutation in the determinant sum will have exactly two of each $(\lambda_i - \lambda_j)$ times some constant dependent on U_0. This gives the Vandermonde-style determinant we want.

Remark. *If we apply this to the GOE, we would get the determinant portion to be*

$$\prod_{i<j} |\lambda_i - \lambda_j|^1$$

instead. This power is the β in literature, and the same one mentioned about.

A criticism of the above proof is that we do not know the actual value of Z_n, which is a big problem for applying the result. To compute this value, we first need a few results.

Lemma 4.5 (Vandermonde determinant)

The Vandermonde matrix is given by

$$\begin{bmatrix} 1 & 1 & \dots & 1 \\ x_1 & x_2 & \dots & x_n \\ x_1^2 & x_2^2 & \dots & x_n^2 \\ \vdots & & \ddots & \vdots \\ x_1^{n-1} & x_2^{n-1} & \dots & x_n^{n-1} \end{bmatrix}$$

and has determinant given by

$$\prod_{i<j} (x_i - x_j)$$

The proof of this is by induction. For an intuition,consider the unique factorization of $C[x_1, ..., x_n]$, that $(x_i - x_j)$ are prime factors of the determinant. There are no other primes by the counting the degree of the polynomial in each variable, and finally we pray that the constant is 1.

Lemma 4.6 (Integrate Out)

Let J_N be a $N \times N$ matrix such that

1. $J_{i,j} = f(x_i, x_j)$ for some measurable $f : \mathbb{R}^2 \to \mathbb{C}$,
2. $\int f(x, y) f(y, z) d\mu(y) = f(x, z)$.

Then

$$\int \det J(x_1,...,x_N)d\mu(x_N) = (d-N+1)\det J_{N-1}(x_1,...,x_{N-1}),$$

where

$$d \stackrel{\text{def}}{=} \int f(x,x)d\mu(x).$$

. This is from Anderson, Guionnet and Zeitouni's Introduction to random matrices or Percy Deift's Orthogonal Polynomials and Random Matrices.

We apply the permutation definition of the determinant

$$\begin{aligned}\int \det J_N d\mu x_N &= \int \sum_{\sigma\in S^n} \operatorname{sgn}(\sigma) J_{1,\sigma(1)}...J_{N,\sigma(N)}d\mu \\ &= \int \sum_{\sigma\in S^n} \operatorname{sgn}(\sigma) f(x_1,x_{\sigma(1)})...f(x_N,x_{\sigma(N)})d\mu.\end{aligned}$$

We split the sum in σ into two cases, the first when $\sigma(N)=N$, and the second $\sigma(N)\neq N$. In the first case

$$\sum_{\substack{\sigma\in S^N\\ \sigma(N)=N}} \int \operatorname{sgn}(\sigma) f(x_1,x_{\sigma(1)})...f(x_N,x_N)d\mu(x_N) = d\det J_{N-1},$$

where we identified each permutation as belonging in S^{N-1} in the canonical way which has the same number of inversions, thus preserves the sign.

For the other summation, we let $j \stackrel{\text{def}}{=} \sigma^{-1}(N)$, such that the integrand becomes

$$\sum_{\sigma\in S^n} \operatorname{sgn}(\sigma) \prod_{i\neq j,N} f(x_i,x_{\sigma(i)}) \times f(x_j,x_N)f(x_N,x_{\sigma(N)})$$

The first product in i is constant in x_N, and the second part integrates in x_N to $f(x_j,x_\sigma(N))$. This gives us a total contribution of $-(N-1)\det J_{N-1}$ by looking at the $N-1$ permutations induced by each σ.

Determination of the value of Z_n. We have

$$1 = \mathbb{P}((\lambda_1,...,\lambda_n)\in\mathbb{R}^n) = \frac{1}{Z_n}\int \exp\Big(-\sum_i Q(\lambda_i)\Big)\prod_{i<j}(\lambda_i-\lambda_j)^2 d\lambda_i,$$

so all we just need to compute the value of the integral. Let $\{\pi_i(x)\}_{x\geq 0}$ be a sequence of monic orthogonal polynomials with respect to the measure $\exp(-Q(x))dx$ in $\mathbb{R}$. I.e. the i-th polynomial is of degree i. Then

$$\begin{aligned}
&\exp\Big(-\sum_i Q(\lambda_i)\Big)\prod_{i<j}(\lambda_i-\lambda_j)^2\\
=&\exp\Big(-\sum_i Q(\lambda_i)\Big)\det\begin{bmatrix} 1 & 1 & \dots & 1\\ \lambda_1 & \lambda_2 & \dots & \lambda_n\\ \lambda_1^2 & \lambda_2^2 & \dots & \lambda_n^2\\ \vdots & & \ddots & \vdots\\ \lambda_1^{n-1} & \lambda_2^{n-1} & \dots & \lambda_n^{n-1}\end{bmatrix}^2\\
=&\exp\Big(-\sum_i Q(\lambda_i)\Big)\det\begin{bmatrix} \pi_0(\lambda_1) & \pi_0(\lambda_2) & \dots & \pi_0(\lambda_n)\\ \pi_1(\lambda_1) & \pi_1(\lambda_2) & \dots & \pi_1(\lambda_n)\\ \vdots & & \ddots & \vdots\\ \pi_{n-1}(\lambda_1) & \pi_{n-1}(\lambda_2) & \dots & \pi_{n-1}(\lambda_n)\end{bmatrix}^2
\end{aligned}$$

Let $\phi_j = \pi_j\exp(-Q(x)/2)/c_j$, where c_j is a normalization constant that makes ϕ_j orthonormal functions on $\mathbb{R}$, then we rewrite

$$\exp\Big(-\sum_i Q(\lambda_i)\Big)\prod_{i<j}(\lambda_i-\lambda_j)^2 = \prod_i c_i^2\det\{\phi_{i-1}(\lambda_j)\}^2.$$

The c_j constants are deterministic, so the we only need to compute the determinant.

$$\begin{aligned}
\det\{\phi_{i-1}(\lambda_j)\}^2 &= \det\{\phi_{i-1}(\lambda_j)\}^T\{\phi_{i-1}(\lambda_j)\}\\
&= \det\{\sum_{l=1}^n \phi_{l-1}(\lambda_i)\phi_{l-1}(\lambda_j)\}.
\end{aligned}$$

We now want to apply the integrate out lemma on

$$k(x,y)\stackrel{\text{def}}{=}\sum_{l=1}^n \phi_{l-1}(x)\phi_{l-1}(y)$$

First we verify the property that

$$\int k(x,y)k(y,z)dy = \int\sum_{l=1}^n\phi_{l-1}(x)\phi_{l-1}(y)\sum_{k=1}^n\phi_{k-1}(y)\phi_{k-1}(z)dy = \sum_{l=1}^n\phi_{l-1}(x)\phi_{k-1}(z) = k(x,z)$$

and

$$\int k(x,x)dy = n$$

by orthonormality of the ϕ's. Thus applying the integrate out lemma once will give a factor of 1 the first time. Inductively, we can apply the integrate out lemma on the $n-k$ dimensional space to get factor of $k+1$. This gives a total contribution of $n!$.

Therefore we get

$$Z_n = n!\prod_i c_i^2$$