

Marchenko-Pastur Law

Chi Li

1 Statement of main theorem

Definition 1.1 (Marchenko-Pastur Distribution)

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

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

Theorem 1.2 (Marchenko-Pastur)

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

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

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

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

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

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

weakly in probability.

2 Structure of proof

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

Lemma 2.1 (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 [2]. The proof is mainly computational and relies on Vandermonde's identity¹.

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


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

at zero.

Lemma 2.2 (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 2.3 (Moments of Eigenvalues)

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

For each $k \geq 1$, We have

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

as $n, m \rightarrow \infty$.

Lemma 2.4 (Concentration)

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

$$\mathbb{P}\left(\left|\int x^k dL_n - \mathbb{E} \int x^k dL_n\right| > \epsilon\right) \rightarrow 0.$$

Proof of Theorem 1.2 assuming the previous four Lemmas. Let $\mu \stackrel{\text{def}}{=} \tilde{\mu} + \left(1 - \frac{1}{y}\right)_+ \delta_0$. Then all k -th moments of L_n converge to the k -th moments of μ .

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

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

First, by Markov,

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


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

Taking limsup,

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

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

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

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. 

3 Calculation of moments of L_n

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

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

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

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

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

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

Proposition 3.1


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 is 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

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

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

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

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

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

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 3.2

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 [1], 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:

- (up, up) \rightarrow (down, down)
- (up, down) \rightarrow (up, down)
- (down, up) \rightarrow (down, up)
- (down, down) \rightarrow (up, 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

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



4 Concentration inequality

Proof of Lemma 2.4. Let $\epsilon > 0$. We apply Chebyshev’s inequality to

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

where

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

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

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

so the contribution from these terms is

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

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



References

- [1] Arnab Ganguly. *Lecture Notes in Random Matrices - Marchenko-Pastur Law*.
- [2] Jack W. Silverstein Zhidong Bai. *Spectral Analysis of Large Dimensional Random Matrices*. Springer New York, NY, 2009.