Wigner's semicircle law part II

November 18, 2024

1 Last time

Last time we spent time to formalise the problem and came up with the following statement:

Theorem 1.1 (Wigner semi-circle law)**.** *Let $X = X_N$ be a $N \times N$ symmetric matrix with i.i.d entries upper diagonally - i.e. a Wigner matrix with eigenvalues $\lambda_1, \dots, \lambda_N$. Moreover assume that the entries have zero mean, unit variance and satisfy $\mathbb{E}|X_{i,j}|^3 < \infty$.*

Then the empirical eigenvalue distribution $\mu_N := \frac{1}{N}\sum_{i=1}^N \delta_{\lambda_i/\sqrt{N}}$ of $X/\sqrt{N}$ converges in the sense of weak convergence of measures, in probability to the semi-circle law μ_{sc} with density proportional to $\sqrt{4-x^2}$.

We discussed the scaling: this can be seen from the fact that $\mathbb{E}(\sum_{i=1}^N \lambda_i^2) = \mathbb{E}\sum X_{i,j}^2 = N$ and saw that even the simpler problem, the convergence of the expected measures $\tilde{\mu}_N := \mathbb{E}\mu_N$ is non-trivial.

We spent also some time in trying to figure out how to address the convergence - which observables to choose - and ended up going for the moment method, as we have that $\mu_n(x^k) = \frac{1}{N}Tr(X^k)$, which seems accessible.

And then we showed how to apply it to obtain a weak ex-changeability statement:

Lemma 1.2. *Suppose the entries of two matrices X, Y are bounded. Then for every $k \in \mathbb{N}$ we have that as $N \to \infty$*

$$|\mathbb{E}\frac{1}{N}(Tr(X^k) - Tr(Y^k))| \to 0.$$

Today we will see how :

- similar calculations help us prove the Wigner semi-circle law first in expectation for bounded entries.
- how to improve the convergence to that in probability
- how to remove the criteria on bounded entries

But first let's do a correction.

1.1 The operator norm of the matrix

Last time I was rushing a bit and said that the operator norm of X_N should be bounded when the entries are bounded.

This is not true: consider for example the matrix with all entries equal to 1.

However, it is morally true in the case of Wigner matrices, meaning that it is true with overwhelmingly large probability:

Exercise 1.1. *Let X be a Wigner $N \times N$ matrix with mean-zero, uniformly bounded (wlog, by one) entries, i.e., $(X_{ij})_{i \leq j \leq N}$ are independent with $\mathbb{E}[X_{ij}] = 0$ and $|X_{ij}| \leq 1$ almost surely and $X_{ji} = \bar{X}_{ij}$. Show that there exists $C, c > 0$ (absolute constants independent of N) such that $\mathbb{P}[\|X\|_{op} / \sqrt{N} \geq t] \leq e^{-cNt^2}$ for all $t \geq C$.*

The proofs are not entirely straightforward in fact! A possible proof is sketched on the example sheet, another one follows from concentration of measure results together with a bound on the expectation. That in turn follows from our computations to follow.

2 Proof of the semi-circle law

Recall that

$$Tr(X^k) := \sum_{i_1, i_2, \dots, i_k} X_{i_1 i_2} \cdots X_{i_k i_1}$$

and that $\mathbb{E}Tr(X^k) = 0$ whenever some entry is used only once because of the i.i.d. condition.

Thus each entry $X_{ij} = X_{ji}$ has to appear at least twice. Let us see that it has to appear exactly twice.

Claim 2.1. *The only terms in*

$$\mathbb{E}\frac{1}{N} Tr((X/\sqrt{N})^k) = N^{-1-k/2} \sum_{i_1, i_2, \dots, i_k} \mathbb{E}(X_{i_1 i_2} \cdots X_{i_k i_1})$$

that survive in the limit use exactly $k/2$ different entries, each with multiplicity two.

Moreover, we can assume that there are no diagonal terms.

Proof. Indeed, suppose we take l different entries. Let us calculate how many terms there are of this type in the sum when entry multiplicities come with $d_1, \dots d_l$ with $\sum d_i = k$ and $d_i \geq 2$, and in particular $l \leq k/2$. As we have $O(N^2)$ choices for the first entry and $O(N)$ for each next new one, we see that there are at most $N^{k+1-\sum_{i=1}^{l}(d_i - 1)} = N^{l+1}$ summands. But this means that their contribution to $\mathbb{E}\frac{1}{N} Tr((X/\sqrt{N})^k)$ is $O(N^{-1-k/2} N^{l+1})$ and we conclude that there have to be $l = k/2$ different terms, which means that each $d_i = 2$. Moreover, we also see directly that k has to be even for the moment to survive in the limit - indeed otherwise for any l the sum of the terms with l different entries is negligible as $N \to \infty$ and as l varies over $1 \dots k$, this means that the whole sum is negligible.

Finally, to see the claim about diagonal terms notice that choosing at least one diagonal term reduces the number of possible sums by a factor of N^{-1}. $\square$

In particular as the variance of each entry is 1 by assumption , we have that

$$\lim_{N\to\infty} \mathbb{E}\frac{1}{N}Tr((X/\sqrt{N})^k) = N^{-1-k/2}\#\{\text{summands with } l = k/2\}$$

. Thus we have reduced the problem to a counting question and we will next find a combinatorial representation of each term $X_{i_1 i_2}\cdots X_{i_k i_1}$ to do this counting.

First notice that the number of possibilities for assigning to each index i_j a number from $\{1,\dots,N\}$ is $N(N-1)\cdots(N-m+1)$, where m is the number of different indexes. For the terms to survive, this has to be exactly equal to $m = k/2+1$ and then this number of possibilities is exactly $N(N-1)\dots(N-k/2)$ so that $N^{-1-k/2}N(N-1)\dots(N-k/2)\to 1$ as $N\to\infty$.

Thus we only need to count the number of words $i_1 i_2 i_3 \dots i_k i_{k+1}$ with the constraints that

- $i_{k+1} = i_1$
- For each j, there is some l such that $\{i_j, i_{j+1}\} = \{i_l, i_{l+1}\}$
- There are $k/2+1$ different indexes that we can take to be the set $\{1,2,\dots,k/2+1\}$ in the order of appearance.

Lemma 2.2. *The collection of words $i_1 i_2 i_3 \dots i_k i_{k+1}$ with the constraints above is in one-to-one correspondence with pairs (T,P) where T is a tree on the vertex set $\{1,2,\dots,k/2+1\}$, P is a cycle on this graph such that it starts from the vertex 1, it traverses each edge twice and each vertex i appears before $i+1$.*

Proof. In one direction, we define the vertices of the graph by different indexes i_j and the edges by $\{i_j, i_{j+1}\}$. As there are $k/2+1$ vertices and $k/2$ edges, the graph is a tree. Further the cycle is given by following the word $i_1 i_2 \dots i_k$ and returning to i_1.

In the other direction we let $i_1 \dots i_k i_{k+1}$ just be equal to the vertices visited along the cycle. □

Counting the pairs of trees and cycles with the conditions above is made simple by the following observation:

Lemma 2.3. *The pairs (T,P) described in the previous lemma are in one-to-one correspondence with walks of k steps starting and ending at 0, going either up or down by 1 at each step and staying non-negative throughout.*

Proof. Given the tree and the cycle, we observe that the distance from the vertex 1 throughout the cycle gives such a walk.

In the other direction, given the walk we can define the tree geometrically: we draw the graph of the walk, then put glue under the graph and smash it together. I leave it to you to figure out the formal statement. ;) □

Naturally, one also set up this bijection directly with the words, but this bijection between trees and paths is too nice to leave out. It now remains to count the number of such walks.

To do this we use the reflection principle. First, the number of the above-mentioned paths is equal to the number of all paths starting from 0 and ending at 0 minus the number of paths that start from 0, end at 0 and visit -1. But by reflection principle this latter is equal to the number of paths starting from 0 and ending at -2.

Thus we obtain $\binom{k}{k/2} - \binom{k}{k/2-1}$, which equals to $\frac{1}{k/2+1}\binom{k}{k/2}$ for even k (recall that odd k gave zero anyways).

This matches the moments of the semi-circle law (see exercise sheet) and hence we have proved the convergence in expectation for matrices of bounded entries. Let us now discuss how to enhance this to convergence in probability and to relax the boundedness assumption.

3 Extensions

3.1 Convergence in probability

To prove convergence in probability, it suffices to show that $\mathrm{Var}(\frac{1}{N}Tr((X/\sqrt{N})^k)$ goes to zero as $N \to \infty$. To do this we can apply Efron-Stein inequality together with the computations of last time. This will be on the example sheet.

3.2 Removing the boundedness assumption

To remove the boundedness assumption, we use the Hofmann-Wielandt inequality (Exercise 3 on the sheet of this week).

Indeed, using this inequality we see that for any two symmetric matrices X, Y of the same size and for all $t > 0$

$$|F_{\mu_X}(t) - F_{\mu_Y}(t+\epsilon)| \leq \frac{1}{N\epsilon^2}\|X-Y\|_F^2$$

where F_{μ_X} denotes the cumulative distribution function of the corresponding eigenvalue distribution. In particular $\sup_{t\in\mathbb{R}} |F_{\mu_X}(t) - F_{\mu_Y}(t+\epsilon)| \leq \frac{1}{N\epsilon^2}\|X-Y\|_F^2$

But now by our assumption, for every $\delta > 0$ we can choose $C > 0$ such that

$$\mathbb{E}\|X - X1_{|X_{i,j}<C|\forall i,j}\|_F^2 < \delta N^2.$$

This means that if we denote by $Y_N := X_N 1_{|X_{i,j}<C|\forall i,j}$, we have that $\|Y_N/\sqrt{N} - X_N/\sqrt{N}\|_F^2 \leq N\delta$ and thus for every $\epsilon > 0$, we can choose $\delta > 0$ (and thus $C > 0$ large enough) such that

$$\mathbb{E}(\sup_{t\in\mathbb{R}} |F_{\mu_{Y_N/\sqrt{N}}}(t) - F_{\mu_{Y_N/\sqrt{N}}}(t+\epsilon)|) \leq \epsilon.$$

But this means it suffices to study the case of bounded matrices as by increasing C we can refind the original $F_{\mu_{X_N/\sqrt{N}}}$.