RANDOM MATRIX THEORY AND THE SEMICIRCLE LAW

ROGER VAN PESKI

0.1. **Intro.** [1]

Classical random matrix theory over $\mathbb{C}$, while now a full-fledged field of its own, arguably grew out of physicist Eugene Wigner's attempts in the 1950s to model the energy levels of heavy nuclei using eigenvalues of random matrices. In this setting, existing atomic physics predicted that the energy levels of such nuclei were given by eigenvalues of a corresponding infinite-dimensional Hermitian Hamiltonian operator, but this proved too difficult to analyze. Wigner's approach was to model this operator by a random matrix with independent Gaussian entries subject to a symmetry restriction so that the matrix would be symmetric/Hermitian, and study the eigenvalue distribution of such random matrices as a model. There are also many connections to the zeros of L-functions, for which a very readable source is [1].

Definition 1. *A* random matrix *is a matrix-valued random variable.*

Here we should review a bit of terminology. You know what a random variable is, and a reasonable random variable X has a density function f such that $\mathbb{E}[g(X)] = \int_{\mathbb{R}} g(x)f(x)dx$ for reasonable g. Using f, we define a *measure* μ on $\mathbb{R}$ by saying $\mu(S) = \int_S f(x)dx$ for reasonable subsets $S \subset \mathbb{R}$. I will use random variables, their densities, and the induced measure pretty interchangeably.

Wigner studied the following ensemble.

Definition 2. *The $N \times N$* Gaussian orthogonal ensemble (GOE) *is the random matrix ensemble given by*

$$M = \begin{pmatrix} X_{0,0} & X_{0,1} & \cdots & X_{0,N-1} \\ X_{1,0} & X_{1,1} & \cdots & X_{1,N-1} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N-1,0} & X_{N-1,1} & \cdots & X_{N-1,N-1} \end{pmatrix}, \tag{1}$$

with the $X_{\ell,j}$ are random variables defined as follows. For $\ell < j$ $X_{\ell,j}$ are iid Gaussians with mean 0 and variance 1, and $X_{j,\ell} = X_{\ell,j}$ (i.e. the ensemble is symmetric). Furthermore, $X_{j,j}$ are iid Gaussians with mean 0 and variance 2.

M has the important property that for any orthogonal matrix $P \in O(N)$, PMP^{-1} has the same distribution.

Example 0.1. Another ensemble one might study is

$$M = \begin{pmatrix} X_{0,0} & X_{0,1} & \cdots & X_{0,N-1} \\ X_{1,0} & X_{1,1} & \cdots & X_{1,N-1} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N-1,0} & X_{N-1,1} & \cdots & X_{N-1,N-1} \end{pmatrix}, \tag{2}$$

Date: December 13, 2021.

[1]Small parts of these notes are shamelessly copied from my senior thesis, and the treatment of the main result draws from Ioanna Dimitriu's "Moment Method I" course at the 2018 Michigan Random Matrix Summer School.

where $X_{i,j} = X_{j,i}$ are iid and $X_{i,j} = \pm 1$ with probability $1/2$ each way.

We care about the eigenvalues of these ensembles. The largest eigenvalue is a real random variable, and the vector of all eigenvalues ordered by size is an $\mathbb{R}^N$-valued random variable. In some ways, the theory of eigenvalues of random matrices mimics that of real random variables, and the following distribution plays the role of the Gaussian.

Definition 3. *The* semicircle law *is the probability measure*[2] *on* $\mathbb{R}$ *with density given by*

$$\nu(x) = \begin{cases} \frac{2}{\pi}\sqrt{1-(x/2)^2} & -2 \leq x \leq 2 \\ 0 & else \end{cases} . \tag{3}$$

Strictly speaking this is a semi-ellipse, not a semicircle.

Wigner showed that as $N \to \infty$, the eigenvalue distribution of the GOE converges to the semicircle law. I'm not making this precise yet, but the idea is that if you sample a random matrix from the GOE and then pick one of its eigenvalues at random, the resulting distribution will depend on N and converge to the semicircle as $N \to \infty$. It was quickly recognized that random matrices exhibit *universality*, i.e. ensembles with differently distributed entries have same limiting eigenvalue distribution. The main result of this series is in this direction.

Definition 4. *The* empirical spectral measure *of a fixed* $N \times N$ *Hermitian matrix* A *with eigenvalues* $\lambda_1, \dots, \lambda_N$ *is the measure with density*

$$\mu_A(x) = \frac{1}{N}\sum_{i=1}^{N} \delta(x - \frac{\lambda_i}{\sqrt{N}}),$$

i.e. the uniform discrete probability measure on $\{\lambda_1, \dots, \lambda_N\}$.

Want to show convergence of ESD to ν. Do this by means of moments.

0.2. **Moment Preliminaries.**

Definition 5. *The* m^{th} *moment of a measure* μ *on* $\mathbb{R}$ *is*

$$\mu^{(m)} := \int_{\mathbb{R}} x^m d\mu.$$

These are kind of like Taylor coefficients, and characterize the measure in some situations.

Proposition 0.1 (Carleman's condition)**.** Let μ be a measure on $\mathbb{R}$ with all moments finite and

$$\sum_{m=1}^{\infty} (\mu^{(2m)})^{-\frac{1}{2n}}$$

diverges. Then μ is the only measure with this sequence of moments.

Note that this means the moments do not grow too fast, because the exponent is negative and we ask for divergence. We can show convergence of measures by convergence of their moments.

[2]What is a probability measure? Just think of it as assigning numbers between 0 and 1 to every reasonable subset of $\mathbb{R}$ which satisfy obvious properties, e.g. $\mu(A \sqcup B) = \mu(A) + \mu(B)$.

Proposition 0.2 (Moment convergence theorem). Let $X_1, X_2, \ldots$ and X be random variables with moments $\mu_n^{(m)} := \mathbb{E}[X_n^m]$ and $\mu^{(m)} := \mathbb{E}[X^m]$. Then if $\mu_n^{(m)} \to \mu^{(m)}$ as $n \to \infty$ and the moments of X uniquely characterize it (by, e.g. Carleman's condition), we have that $X_n \to X$ in distribution.

Exercise 1. Prove that the moments of μ uniquely characterize it under the stronger assumption that μ has compact support. Hint: Stone-Weierstrass theorem.

Example 0.2. The moments of the Gaussian are

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-x^2/2} x^m dx = \begin{cases} 0 & m \text{ odd} \\ (m-1)!! & m \text{ even} \end{cases}.$$

You can compute this by knowing that the zeroth moment is 1 (this is kind of tricky) and integrating by parts to reduce other moments to this case (exercise, not too hard).

Exercise 2 (More integration tricks). Prove that the moments of the semicircle are

$$\int_{-2}^{2} \frac{2}{\pi} \sqrt{1-(x/2)^2} = \begin{cases} 0 & m \text{ odd} \\ C_{m/2} & m \text{ even} \end{cases}$$

where $C_{m/2} = \frac{1}{m/2+1}\binom{m}{m/2}$ is the $(m/2)^{th}$ Catalan number.

Hint: trig sub and the Catalan recurrence.

Remark 0.1 ((Warning: contains material I don't know well)). $(2m-1)!!$ is the number of ways to partition a set of $2m$ elements into pairs. C_m is the number of 'noncrossing partitions' of a $2m$-element set. This suggests a notion of a q-weighted set partition, i.e.

$$F_{2m}(q) = \sum_{\Pi \in P(2m)} q^{n(\Pi)}$$

where $P(2m)$ is the set of all partitions of $\{1, \ldots, 2m\}$ into pairs and for any such partition Π, $n(\Pi)$ is the number of crossings. Then $F_{2m}(1) = (2m-1)!!$ and $F_{2m}(0) = C_m$ by above. It is also true that $F_{2m}(-1) = 1$, so $F_{2m}(-1)$ is the $2m^{th}$ moment of the random variable X which takes values ± 1 with probability $1/2$ each. There is real structure here which I will now handwave: these three cases of $q = 1, 0, -1$ correspond to classical probability, 'free probability' and 'boolean probability'. Free probability is a theory of noncommuting random variables with applications to random matrices, and together these three have been shown to correspond to the *only* notions of independence of 'algebraic probability spaces'.

0.3. **Main result.** Back to random matrix theory.

Definition 6. *Let $(X_n)_{n \geq 1}$ and X be random variables. Then $X_n \to X$* in probability *($X_n \xrightarrow{p} X$) if for all $\epsilon > 0$, $\lim_{n\to\infty} \Pr(|X - X_n| > \epsilon) = 0$.*

We can now state the main result.

Theorem 0.1. Let $M_1, M_2, \ldots$ be a sequence of random matrices

$$M_N = \begin{pmatrix} X_{0,0} & X_{0,1} & \cdots & X_{0,N-1} \\ X_{1,0} & X_{1,1} & \cdots & X_{1,N-1} \\ \vdots & \vdots & \ddots & \vdots \\ X_{N-1,0} & X_{N-1,1} & \cdots & X_{N-1,N-1} \end{pmatrix}, \tag{4}$$

where $X_{i,j} = X_{j,i}$, the $X_{i,j}$'s are any iid random variables with mean 0 and finite higher moments, and for $i \neq j$ the $X_{i,j}$'s have variance 1. Then the moments of the empirical spectral measures converge in probability to that of the semicircle, $\mu_{M_N}^{(m)} \xrightarrow{p} \nu^{(m)}$ as $N \to \infty$.

Proof. The idea is to show that $\mathbb{E}[\mu_{M_N}^{(m)}] \to \nu^{(m)}$ (this is just convergence of real numbers in the usual sense) and then show $\mathbb{E}[(\mu_{M_N}^{(m)} - \mathbb{E}[\mu_{M_N}^{(m)}])^2] \to 0$, which ensures the moments stay close to their expectations. Both parts have a similar proof.

Recall $\mathrm{Tr}(A^k) = \sum_i \lambda_i^k$ for any matrix A. Since trace is sum of diagonal elements, we further have

$$\mathrm{Tr}(A^k) = \sum_{1 \leq i_1, \ldots, i_k \leq N} a_{i_1,i_2} a_{i_2,i_3} \cdots a_{i_k,i_1}.$$

Now, we have that

$$\mu_{M_N}^{(m)} = \int_{\mathbb{R}} x^m d\mu_{M_N} = \frac{1}{N} \sum_{i=1}^{N} \left(\frac{\lambda_i}{\sqrt{N}} \right)^m = \frac{1}{N^{1+m/2}} \mathrm{Tr}(M_N^m) = \frac{1}{N^{1+m/2}} \sum_{1 \leq i_1, \ldots, i_m \leq N} X_{i_1,i_2} X_{i_2,i_3} \cdots X_{i_m,i_1}.$$

Let $I = (i_1, \ldots, i_m)$ and consider the graph G_I with vertices $V_I = \{i_1, \ldots, i_m\}$ (note some of the i_j may be repeated) and edges $E_I = \{(i_1, i_2), (i_2, i_3), \ldots, (i_m, i_1)\}$. I defines a walk on G_I which starts at i_1 and returns to i_1. If any of the $X_{i_j,i_{j+1}}$ in $X_{i_1,i_2} X_{i_2,i_3} \cdots X_{i_k,i_1}$ appears only once, then $\mathbb{E}[X_{i_1,i_2} X_{i_2,i_3} \cdots X_{i_k,i_1}] = 0$ because $X_{i_j,i_{j+1}}$ is mean zero; hence the walk I must traverse each edge at least twice. Now express

$$\frac{1}{N^{1+m/2}} \sum_{1 \leq i_1, \ldots, i_m \leq N} \mathbb{E}[X_{i_1,i_2} X_{i_2,i_3} \cdots X_{i_m,i_1}]$$

as

$$\frac{1}{N^{1+m/2}} \sum_{I, G_I} \sum_{\text{indexings of } G_I} \mathbb{E}[X_{i_1,i_2} X_{i_2,i_3} \cdots X_{i_m,i_1}].$$

where the first sum is over graphs G on $\leq m$ vertices (allowing self-loops) with a base point and a walk of length m starting and ending at the base point and traversing all edges at least twice, and the second sum is over all ways to assign indices in $\{1, \ldots, N\}$ to the vertices. The first sum is finite, and the terms $\mathbb{E}[X_{i_1,i_2} X_{i_2,i_3} \cdots X_{i_m,i_1}]$ are finite and bounded by a constant dependent on m. Hence this sum is

$$\frac{1}{N^{1+m/2}} \sum_{I, G_I} O(N^{|V_I|}),$$

so only the graphs with most vertices contribute in the $N \to \infty$ limit. We note that $|E_I| \leq m/2$ since each edge is traversed twice, and thus $|V_I| \leq m/2 + 1$ with the maximum achieved iff G_I is a tree. If m is odd, then $|V_I| \leq m/2 + 1$ is an integer and hence $\frac{1}{N^{1+m/2}} O(N^{|V_I|}) = O(1/N^{1/2})$ goes to zero, so all odd moments are zero. Now compute even moments.

Since trees maximize number of vertices, the only contribution is from terms where G_I is a tree and I traverses each edge exactly twice. These are counted by the Catalan number $C_{m/2}$, and the term $\mathbb{E}[X_{i_1,i_2} X_{i_2,i_3} \cdots X_{i_m,i_1}]$ corresponding to such a (I, G_I) pair is 1 because all $X_{i,j}$ have variance 1. Hence

$$\lim_{N \to \infty} \frac{1}{N^{1+m/2}} \sum_{I, G_I} \sum_{\text{indexings of } G_I} \mathbb{E}[X_{i_1,i_2} X_{i_2,i_3} \cdots X_{i_m,i_1}] = C_{m/2}.$$

We have

$$\mathbb{E}[(\mu_{M_N}^{(m)} - \mathbb{E}[\mu_{M_N}^{(m)}])^2] \tag{5}$$

$$= \mathbb{E}[(\mu_{M_N}^{(m)})^2] - \mathbb{E}[\mu_{M_N}^{(m)}]^2 \tag{6}$$

$$= \frac{1}{N^{2+m}} \sum_{I,J,G_{I\cup J}} \sum_{\text{indexings of } G_{I\cup J}} \mathrm{Cov}(X_{i_1,i_2}\cdots X_{i_m,i_1}, X_{j_1j_2}\cdots X_{j_mj_1}) \tag{7}$$

where

$$\mathrm{Cov}(X_{i_1,i_2}\cdots X_{i_m,i_1}, X_{j_1j_2}\cdots X_{j_mj_1}) \tag{8}$$

$$:= (\mathbb{E}[(X_{i_1,i_2}\cdots X_{i_m,i_1})(X_{j_1j_2}\cdots X_{j_mj_1})] - \mathbb{E}[X_{i_1,i_2}\cdots X_{i_m,i_1}]\mathbb{E}[X_{j_1j_2}\cdots X_{j_mj_1}]) \tag{9}$$

Similarly to before, we need to find which I, J, G_I, G_J have

$$\mathbb{E}[(X_{i_1,i_2}\cdots X_{i_m,i_1})(X_{j_1j_2}\cdots X_{j_mj_1})] - \mathbb{E}[X_{i_1,i_2}\cdots X_{i_m,i_1}]\mathbb{E}[X_{j_1j_2}\cdots X_{j_mj_1}]$$

nonzero. The only way this term can be nonzero is if there is at least one pair $X_{i_k,i_{k+1}}$ and $X_{j_l,j_{l+1}}$ which are equal, for if they are all independent then it makes no difference whether or not they are inside the same expectation. Each edge of $G_{I\cup J}$ (note that when I take this union I am identifying along edges) must be covered at least twice, hence $|E_{I\cup J}| \le m$. Thus

$$|V_{I\cup J}| \le |E_{I\cup J}| + 1 \le m+1.$$

Now, $\mathrm{Cov}(X_{i_1,i_2}\cdots X_{i_m,i_1}, X_{j_1j_2}\cdots X_{j_mj_1})$ is bounded by a constant dependent only on m, and by the above we have

$$\sum_{\text{indexings of } G_{I\cup J}} \mathrm{Cov}(X_{i_1,i_2}\cdots X_{i_m,i_1}, X_{j_1j_2}\cdots X_{j_mj_1}) = O(N^{m+1}).$$

Since $\sum_{I,J,G_{I\cup J}}$ is finite, this shows

$$\frac{1}{N^{2+m}} \sum_{I,J,G_{I\cup J}} \sum_{\text{indexings of } G_{I\cup J}} \mathrm{Cov}(X_{i_1,i_2}\cdots X_{i_m,i_1}, X_{j_1j_2}\cdots X_{j_mj_1}) \to 0$$

as $N \to \infty$, proving that $\mathrm{Var}(\mu_{M_N}^{(m)}) \to 0$ as $N \to \infty$.

Lemma 0.1 (Chebyshev's inequality)**.** For a random variable X with mean μ and variance σ, and any $\epsilon > 0$, we have $\Pr(|X-\mu| \ge \epsilon) \le \frac{\sigma^2}{\epsilon^2}$.

Proof.

$$\mathbb{E}[\mathbb{1}_{|X-\mu|^2\ge\epsilon^2}] \le \mathbb{E}[\frac{|X-\mu|^2}{\epsilon^2}\mathbb{1}_{|X-\mu|^2\ge\epsilon^2}] \tag{10}$$

$$\le \mathbb{E}[\frac{|X-\mu|^2}{\epsilon^2}] \tag{11}$$

$$= \frac{\sigma^2}{\epsilon^2}. \tag{12}$$

□

Thus

$$\Pr(|\mu_{M_N}^{(m)} - \nu^{(m)}| \le \epsilon) \le \frac{\mathbb{E}[(\mu_{M_N}^{(m)} - \mathbb{E}[\mu_{M_N}^{(m)}])^2]}{\epsilon^2}$$

goes to 0 as $N \to \infty$, so $\mu_{M_N}^{(m)} \xrightarrow{p} \nu^{(m)}$, completing the proof. □

References

[1] F. W. Firk and S. J. Miller. Nuclei, primes and the random matrix connection. *Symmetry*, 1(1):64–105, 2009.