NOTES FROM "INTRODUCTION TO RANDOM MATRICES" BY ANDERSON, GUIONNET AND ZEITOUNI

Definition 1. Let $\{Z_{i,j}\}_{1\leq i<j}$ and $\{Y_i\}_{1\leq i}$ be collections of i.i.d. mean zero random variables. Suppose morevoer that $\mathbf{E}(Z_{i,j}^2)=1$ and that both Y and Z have finite moments of every order. For any $n\in\mathbb{N}$ now, we can make an $n\times n$ random symmetric matrix by putting the Z's above the diagonal, and the Y's on the diagonal (below the diagonal is the same as above, since the matrix is symetric) This is:

$$X_n=\frac{1}{\sqrt{n}}\begin{bmatrix} Y_1 & Z_{1,2} & Z_{1,3} & & Z_{1,n} \\ Z_{1,2} & Y_2 & Z_{2,3} & \dots & Z_{2,n} \\ Z_{1,3} & Z_{2,3} & Y_3 & & Z_{3,n} \\ & \vdots & & \ddots & \\ Z_{1,n} & Z_{2,n} & Z_{3,n} & & Y_n \end{bmatrix}$$

This is called a *Wigner Matrix*. If Z and Y have a Gaussian distribution, we call it a *Gaussian Wigner Matrix*.

Remark 2. We will be interested in the *eigenvalues* of the Wigner matrix, as there are some interesting convergence results to be had as $n\to\infty$.

Definition 3. Let X_n be a Wigner matrix and let $\lambda_1^n\leq\lambda_2^n\leq\ldots\leq\lambda_n^n$ be its eigenvalues. (They are real because the matrix is symmetric). We define the *empirical distribution* of the eigenvalues as the measure (here δ is the unit mass):

$$L_n=\frac{1}{n}\sum_{i=1}^n\delta_{\lambda_i^n}$$

So that L_n measures the number of eigenvalues in any set. i.e.:

$$L_n\left(a,b\right)=\frac{1}{n}\left|\{i:\lambda_i^n\in(a,b)\}\right|$$

This is something like "the density" of the eigenvalues in (a,b) since n is the total number of eigenvalues.

Definition 4. The *semicircle distribution* is the probability distribution $\sigma(x)\mathrm{d}x$ which is given by a density with respect to the Lebesgue measure on $\mathbb{R}$:

$$\sigma(x)=\frac{1}{\sqrt{2\pi}}\sqrt{4-x^2}\mathbb{1}_{|x|\leq 2}$$

If you were to graph this, you would see it is a semicircle!

Theorem 5. *[Wigner] For a Wigner matrix X_n the empirical measure L_n converges weakly, in probability, to the semicircle distribution. I.e. for any continuous bounded function $f:\mathbb{R}\to\mathbb{R}$ we have:*

$$\mathbf{P}\left(\left|\langle L_n,f\rangle-\langle\sigma,f\rangle\right|>\epsilon\right)\to 0$$

Where $\langle L_n,f\rangle=\int f dL_n$.

Remark 6. The above terminology can be a bit confusing because "converges weakly" in a a probabilisitc setting ordinarily means that for probability measures $\mathbf{P}_n$ and $\mathbf{P}$ means that $\langle\mathbf{P}_n,f\rangle\to\langle\mathbf{P},f\rangle$ for every bounded continuous functions f (There are lots of other equivalent ways to phrase this, e.g, $\mathbf{P}_n(a,b)\to\mathbf{P}(a,b)$ for continuity sets (a,b) of $\mathbf{P}$). However for us the object L_n is a *random* probability measure (i.e. a probability measure valued random variable), so we have to specify in what sense (e.g. in probability, almost surely etc.) we mean $\langle L_n,f\rangle\to\langle\sigma,f\rangle$ when we say "L_n goes to σ weakly". Wigner's theorem tells us that the convergence is in the sense of *convergence in probability.*

Example 7. Here is an example to make sure our heads are screwed on right for this convergence in distribution. Let A_i be a sequence of i.i.d. random variables and let $E_n=\frac{1}{n}\sum_{i=1}^n\delta_{A_i}$ be the *empirical distribution* of the first n random variables. This is a random measure! If $\mathbf{P}$ is the law of the A_i's, one can check that E_n goes to $\mathbf{P}$ weakly, almost surely. This is true since $E_n(a,b)=\frac{1}{n}\left|\{i:A_i\in(a,b)\right|=\frac{1}{n}\sum_{i=1}^n\mathbb{1}_{\{A_i\in(a,b)\}}\overset{\text{a.s.}}{\to}\mathbf{E}\left(\mathbb{1}_{\{A\in(a,b)\}}\right)=\mathbf{P}(a,b)$. The almost sure convergence here is from the strong law of large numbers since the indicator random variables $\mathbb{1}_{\{A_i\in(a,b))}$ are all i.i.d.

Remark 8. We will now set out to prove Wigner's theorem. The proof we give here will be based on combinatorial arguments. We first prove a few facts about the semi-circle law.

Lemma 9. *[Moments of the Semi-Circle Law] Let $m_k = \left\langle \sigma, x^k \right\rangle$ be the $k-$th moment of the semi-circle law. The odd moments vanish, and the even moments are equal to the Catalan numbers. That is:*

$$\begin{aligned} m_{2k+1} &= 0 \\ m_{2k} &= C_k \end{aligned}$$

Where C_k is the $k-$th Catalan number $C_k = \frac{1}{k+1}\binom{2k}{k}$.

Proof. m_{2k+1} is clear since $\sigma(x)x^{k+1}$ is an odd function. To see $m_{2k} = C_k$ look at $\int \sigma(x)x^{2k}\mathrm{d}x$ and do a change of variable to polar coordinates, and then integrate by parts in a clever way to get a recurrence relation. Have:

$$\begin{aligned} m_{2k} &= \int_{-2}^{2} x^{2k}\sigma(x)\mathrm{d}x \\ &= \frac{2\cdot 2^{2k}}{\pi}\int_{-\pi/2}^{\pi/2}\sin^{2k}(\theta)\cos^2(\theta)\mathrm{d}\theta \\ &= \frac{2\cdot 2^{2k}}{\pi}\int_{-\pi/2}^{\pi/2}\sin^{2k}(\theta)\mathrm{d}\theta - \frac{2\cdot 2^{2k}}{\pi}\int_{-\pi/2}^{\pi/2}\sin^{2k+2}(\theta)\mathrm{d}\theta \\ &= \frac{2\cdot 2^{2k}}{\pi}\int_{-\pi/2}^{\pi/2}\sin^{2k}(\theta)\mathrm{d}\theta - \frac{2\cdot 2^{2k}}{\pi}\int_{-\pi/2}^{\pi/2}\left[\sin^{2k+1}(\theta)\right]\left[\sin(\theta)\right]\mathrm{d}\theta \\ &= \frac{2\cdot 2^{2k}}{\pi}\int_{-\pi/2}^{\pi/2}\sin^{2k}(\theta)\mathrm{d}\theta - (2k+1)m_{2k} \end{aligned}$$

The last equality comes from doing integration by parts, differentiating $\sin^{2k+1}$ and integrating sin, and then recognizing $m_{2k} \sim \int \sin^{2k}\cos^2$. From here we get:

$$\begin{aligned} m_{2k} &= \frac{1}{2k+2}\frac{2\cdot 2^{2k}}{\pi}\int_{-\pi/2}^{\pi/2}\sin^{2k}(\theta)\mathrm{d}\theta \\ &= \frac{4(2k-1)}{2k+2}m_{2k-2} \end{aligned}$$

Where we use the same integration by parts trick again to see that $m_{2k-2} \sim \int \sin^{2k} \sim \int \sin^{2k-2}\cos^2$. From this recurrence relationship we can easily prove (by induction for instance) that $m_{2k} = C_k = \frac{1}{k+1}\binom{2k}{k}$ ☐

Remark 10. [About the Catalan Numbers] The Catalan numbers arise in all sort of combinatorial enumeration problems. One is the number of NE-SE paths (the type you normally consider for random walks). The number of paths which have $2k$ total steps, k NE steps, k SE steps and which *never* go under zero is C_k. These are called *Dyck paths*. (You can derive this with the reflection principle). One of the most important properties of the Catalan number is the recurrance:

$$C_k = \sum_{j=1}^{k} C_{k-j}C_{j-1}$$

One can also show that the generating function $\beta(z) = \sum_{k=0}^{\infty} C_k z^k$ is (the convention is $C_0 = 1$):

$$\beta(z) = \frac{1-\sqrt{1-4z}}{2z}$$

This is done by writing $C_k = \sum_{j=1}^{k} C_{k-j}C_{j-1}$ because the Dyck paths of length $2k$ can be divided by the *first time* they touch zero. If this first time is $2j$ then there are C_{j-1} paths possible on the left of the first hitting time and C_{k-j} paths possible on the right. Summing gives the formula above, and then some generating function manipulation yields a quadratic equation $\beta(z) = 1 + z\beta(z)^2$ which we can then solve for $\beta(z)$.

This recurrence property of the Catalan number also lets us see the Catalan numbers appearing in other places, for example the number of rooted planar trees with k edges and $k+1$ vertices, because it is not hard to exhibit the same recurrence for these objects. (A rooted planar tree is a tree with a root and a choice of ordering on the children of every node) (For the rooted planar trees, look at the root and its first child. Let $j-1$ be the number of edges in the subtree coming from the child so there are $k-j$ in the subtree coming from the root which does not contain the child. These subtrees are exactly rooted planar trees, so summing over j now gives the same recurrence relation as above. Another way to see this is a dircet bijection betwen roothed planar trees and Dyck paths. To do this you "explore" around the outside of the tree, and create a Dyck path by taking a step up every time you move down a generation, and a step down every time you move up a generation.)

One can show that *non-crossing* partitions of $\{1,\dots k\}$have the same recurrence and so are counted by C_k as well. This is because if j is the largest element connected to 1 in the partition, then the non-crossing property means that we will induce a non-crossinng partion on $\{1\dots j-1\}$and $\{j+1,\dots,k\}$too. Summing over j gives the same recurrence.

1. First Proof OF Wigner's Theorem

Definition 11. Recall the definition of L_n the empirical distribution of the eigenvalues for the Wigner matrix. Let $\bar{L}_n = \mathbf{E}(L_n)$ be the (non-random) measure on $\mathbb{R}$ given by $\bar{L}_n(a,b) = \mathbf{E}(L_n(a,b))$ or equivalently $\langle \bar{L}_n, f\rangle = \mathbf{E}(\langle L_n, f\rangle)$. Recall that $m_k$was the $k-th$ moment of the semi-circle lawσ. Let $m_k^n = \langle \bar{L}_n, x^k\rangle$ be the k-th moment.

Remark 12. To prove that $L_n \xrightarrow{\mathbf{P}} \sigma$ weakly, we will show that as n gets large, L_n is very close to $\bar{L}_n$ and that $\bar{L}_n$ is very close to σ. To make this precise, we will prove the following two lemmas:

Lemma 13. *For every $k\in\mathbb{N}$:*

$$\begin{aligned}\lim_{n\to\infty} m_k^n &= m_k\\ \lim_{n\to\infty}\langle \bar{L}_n, x^k\rangle &= \langle \sigma, x^k\rangle\end{aligned}$$

Lemma 14. *For every $k\in\mathbb{N}$ and every $\epsilon>0$:*

$$\lim_{n\to\infty}\mathbf{P}\left(\left|\langle L_n, x^k\rangle - \langle \bar{L}_n, x^k\rangle\right| > \epsilon\right) = 0$$

We will now prove Wigner's theorem assuming these two lemmas have been proven, and then we will come back to the proof of these lemmas afterward.

Theorem 15. *[Wigner] For a Wigner matrix X_n the empirical measure L_n converges weakly, in probability, to the semicircle distribution. I.e. for any continuous bounded function $f:\mathbb{R}\to\mathbb{R}$ we have:*

$$\lim_{n\to\infty}\mathbf{P}\left(\left|\langle L_n, f\rangle - \langle \sigma, f\rangle\right| > \epsilon\right) = 0$$

Proof. (Assuming the two lemmas) We will first prove that:

$$\limsup_{n\to\infty}\mathbf{P}\left(\langle L_n, |x|^k 1_{|x|>5}\rangle > \epsilon\right) = 0$$

By Chebyshev inequality, we have:

$$\begin{aligned}\mathbf{P}\left(\langle L_n, |x|^k 1_{|x|>B}\rangle > \epsilon\right) &\le \frac{1}{\epsilon}\mathbf{E}\left(\langle L_n, |x|^k 1_{|x|>B}\rangle\right)\\ &\le \frac{1}{\epsilon}\mathbf{E}\left(\left\langle L_n, \frac{|x|^{2k}}{B^k}1_{|x|>B}\right\rangle\right)\\ &\le \frac{1}{\epsilon B^k}\mathbf{E}\left(\langle L_n, x^{2k}\rangle\right) = \frac{m_{2k}^n}{\epsilon B^k}\end{aligned}$$

Have then, since $\lim_{n\to\infty} m_k^n = m_k$, that:

$$\limsup_{n\to\infty}\mathbf{P}\left(\langle L_n, |x|^k 1_{|x|>B}\rangle > \epsilon\right) \le \limsup_{n\to\infty}\frac{m_{2k}^n}{\epsilon B^k} = \frac{m_{2k}}{\epsilon B^k} = \frac{C_k}{\epsilon B^k}$$

If we choose $B=5$ and then use the simple inequality $C_k \le 4^k$ (C_k is the number of Dyck paths, while $4^n$is the *total* number of NE-SE paths of length $2n$) then we see the right hand side is going to zero as $k\to\infty$. Since $|x|^k 1_{|x|>B}$ is *increasing* in k when $B=5$, we have then that:

$$\begin{aligned}\limsup_{n\to\infty}\mathbf{P}\left(\langle L_n, |x|^k 1_{|x|>5}\rangle > \epsilon\right) &\le \limsup_{n\to\infty}\mathbf{P}\left(\langle L_n, |x|^{k+l} 1_{|x|>5}\rangle > \epsilon\right)\\ &\le \frac{C_{k+l}}{\epsilon 5^{k+l}}\\ &\to 0 \text{ as } l\to\infty\end{aligned}$$

So we conclude that:

$$\limsup_{n\to\infty}\mathbf{P}\left(\langle L_n, |x|^k 1_{|x|>5}\rangle > \epsilon\right) = 0$$

We will now show that this reduces the problem to examining only functions f which are supported in $[-5,5]$. Choosing $k=0$ gives us that $\lim_{n\to\infty}\mathbf{P}\left(\left\langle L_n, K1_{|x|>5}\right\rangle>\epsilon\right)=0$, so for bounded continuos functions f we have:

$$\begin{aligned}
\mathbf{P}\left(\left|\langle L_n,f\rangle-\langle\sigma,f\rangle\right|>\epsilon\right) &= \mathbf{P}\left(\left|\left\langle L_n, f1_{|x|\le 5}+f1_{|x|>5}\right\rangle-\left\langle\sigma, f1_{|x|\le 5}+f1_{|x|>5}\right\rangle\right|>\epsilon\right)\\
&= \mathbf{P}\left(\left|\left(\left\langle L_n, f1_{|x|\le 5}\right\rangle-\left\langle\sigma, f1_{|x|\le 5}\right\rangle\right)+\left(\left\langle L_n, f1_{|x|>5}\right\rangle-\left\langle\sigma, f1_{|x|>5}\right\rangle\right)\right|>\epsilon\right)\\
&\le \mathbf{P}\left(\left|\left\langle L_n, f1_{|x|\le 5}\right\rangle-\left\langle\sigma, f1_{|x|\le 5}\right\rangle\right|>\frac{\epsilon}{2}\right)+\mathbf{P}\left(\left|\left\langle L_n, f1_{|x|>5}\right\rangle-\left\langle\sigma, f1_{|x|>5}\right\rangle\right|>\frac{\epsilon}{2}\right)
\end{aligned}$$

Hence:

$$\begin{aligned}
\mathbf{P}\left(\left|\langle L_n,f\rangle-\langle\sigma,f\rangle\right|>\epsilon\right)-\mathbf{P}\left(\left|\left\langle L_n, f1_{|x|\le 5}\right\rangle-\left\langle\sigma, f1_{|x|\le 5}\right\rangle\right|>\frac{\epsilon}{2}\right) &\le \mathbf{P}\left(\left|\left\langle L_n, f1_{|x|>5}\right\rangle-\left\langle\sigma, f1_{|x|>5}\right\rangle\right|>\frac{\epsilon}{2}\right)\\
&= \mathbf{P}\left(\left|\left\langle L_n, f1_{|x|>5}\right\rangle-0\right|>\frac{\epsilon}{2}\right)\\
&\le \mathbf{P}\left(\left|\left\langle L_n, (\sup f)\,1_{|x|>5}\right\rangle\right|>\frac{\epsilon}{2}\right)\\
&\to 0
\end{aligned}$$

So showing that $\mathbf{P}\left(\left|\left\langle L_n, f1_{|x|\le 5}\right\rangle-\left\langle\sigma, f1_{|x|\le 5}\right\rangle\right|>\frac{\epsilon}{2}\right)\to 0$ is sufficient to show $\mathbf{P}\left(\left|\langle L_n,f\rangle-\langle\sigma,f\rangle\right|>\epsilon\right)\to 0$. This means we can restrict our attention to functions f supported on $[-5,5]$.

Fix such a function f and any $\delta>0$. By the Stone-Weierstrass theorem, we can find a polynomial $Q_\delta(x)=\sum_{i=0}^{L}c_i x^i$ (depending on δ) that approximated f in the sup norm to within $\delta/8$ so that the difference $\Delta=Q_\delta-f$ has:

$$\sup_{|x|\le 5}|\Delta(x)|=\sup_{|x|\le 5}|Q_\delta(x)-f(x)|\le\frac{\delta}{8}$$

Notice that since f is supported in $[-5,5]$ we may write $f=Q_\delta-Q_\delta 1_{|x|>5}+\Delta 1_{|x|<5}$. Have then:

$$\begin{aligned}
\mathbf{P}\left(\left|\langle L_N,f\rangle-\langle\sigma,f\rangle\right|>\delta\right) &= \mathbf{P}\left(\left|\left\langle L_N, Q_\delta-Q_\delta 1_{|x|>5}+\Delta 1_{|x|<5}\right\rangle-\left\langle\sigma, Q_\delta-Q_\delta 1_{|x|>5}+\Delta 1_{|x|<5}\right\rangle\right|>\delta\right)\\
&\le \mathbf{P}\left(\left|\langle L_N,Q_\delta\rangle-\langle\sigma,Q_\delta\rangle\right|+\left|\left\langle L_N,Q_\delta 1_{|x|>5}\right\rangle\right|+\right.\\
&\qquad\left.\left|\left\langle L_N,\Delta 1_{|x|<5}\right\rangle\right|+\left|\left\langle \sigma,Q_\delta 1_{|x|>5}\right\rangle\right|+\left|\left\langle\sigma,\Delta 1_{|x|<5}\right\rangle\right|>\delta\right)\\
&\le \mathbf{P}\left(\left|\langle L_N,Q_\delta\rangle-\langle\sigma,Q_\delta\rangle\right|+\left|\left\langle L_N,Q_\delta 1_{|x|>5}\right\rangle\right|+\right.\\
&\qquad\left.\frac{\delta}{8}+0+\frac{\delta}{8}>\delta\right)\\
&\le \mathbf{P}\left(\left|\langle L_N,Q_\delta\rangle-\left\langle\bar{L}_N,Q_\delta\right\rangle\right|+\left|\left\langle\bar{L}_N,Q_\delta\right\rangle-\langle\sigma,Q_\delta\rangle\right|+\left|\left\langle L_N,Q_\delta 1_{|x|>5}\right\rangle\right|>\frac{3\delta}{4}\right)\\
&\le \mathbf{P}\left(\left|\langle L_N,Q_\delta\rangle-\left\langle\bar{L}_N,Q_\delta\right\rangle\right|>\frac{\delta}{4}\right)+\mathbf{P}\left(\left|\left\langle\bar{L}_N,Q_\delta\right\rangle-\langle\sigma,Q_\delta\rangle\right|>\frac{\delta}{4}\right)+\mathbf{P}\left(\left|\left\langle L_N,Q_\delta 1_{|x|>5}\right\rangle\right|>\frac{\delta}{4}\right)\\
&:= P_1+P_2+P_3
\end{aligned}$$

Since Q_δ is a polynomial, the result of the second lemma 14 on the preceding page tells us that $P_1\to 0$ and the result of the first lemma 13 on the previous page tells us that $P_2\to 0$. We know that $P_3\to 0$ by $\mathbf{P}\left(\left\langle L_n,|x|^k 1_{|x|>5}\right\rangle>\epsilon\right)\to 0$ which we proved at the beginning of the proof, and again since Q_δ is a polynomial. Hence $\mathbf{P}\left(\left|\langle L_N,f\rangle-\langle\sigma,f\rangle\right|>\delta\right)\to 0$ too. □

1.1. **Proof of the Lemmas.** The starting point of the proof of lemma 13 is to notice the following idenity :

$$\begin{aligned}
\left\langle\bar{L}_N,x^k\right\rangle &= \frac{1}{N}\sum\mathbf{E}\left(\lambda_i^k\right)\\
&= \frac{1}{N}\mathbf{E}\left(\mathrm{Tr}\left(X_N^k\right)\right)\\
&= \frac{1}{N}\sum_{i_1,\ldots i_k=1}^{N}\mathbf{E}\left(X_N(i_1,i_2)\cdot X_N(i_2,i_3)\cdot\ldots\cdot X_N(i_{k-1},i_k)\cdot X_N(i_k,i_1)\right)\\
&:= \frac{1}{N}\sum_{i_1,\ldots i_k=1}^{N}\mathbf{E}\left(T^N_{(i_1,i_2,\ldots,i_k)}\right):=\frac{1}{N}\sum_{i_1,\ldots i_k=1}^{N}\bar{T}^N_{(i_1,i_2,\ldots,i_k)}\\
&= \frac{1}{N}\sum_{i_1,\ldots i_k=1}^{N}\mathbf{E}\left(T^N_{\mathbf{i}}\right)=\frac{1}{N}\sum_{i_1,\ldots i_k=1}^{N}\bar{T}^N_{\mathbf{i}}
\end{aligned}$$

Where $\mathbf{i} = (i_1, i_2, \ldots, i_k)$ and $T_{\mathbf{i}}^N$ and $\bar{T}_{\mathbf{i}}^N$ are defined by the above. The proof of the lemma now comes from combinatorial arguements over which $\mathbf{i}$ contribute to the above sum. Indeed since $X_N(i_a, i_b) = Z_{a,b}$ or $Y_{a,b}$ are independent and mean zero, $\mathbf{E}(T_{\mathbf{i}}^N) = 0$ for many choices of $\mathbf{i}$, for example if there is a pair i_a i_b that only appears once in $\mathbf{i}$. With some work, we will see that there are order $N^{k/2+1}$ non-zero terms. We will also see that there are order $N^{k/2}$ terms involving moments of $Z_{a.b}$ or $Y_{a,b}$ higher than or equal to 4. Since all these moments are finite, and they represent a $\frac{1}{N}$ fraction of the sum, these will not contribute in the limit that $N \to \infty$. We will now create some combinatorial objects to investigate this in detail.

Definition 16. Given a set $\mathcal{L}$, an *$\mathcal{L}$-letter* is simply an element $s \in \mathcal{L}$. An *$\mathcal{L}$-word* w is a non-empty finite sequence of $\mathcal{L}$-letter, $s_1 s_2 \ldots s_n$. An $\mathcal{L}$-word is called *closed* if its first and last letters are the same. Two $\mathcal{L}$ words are called *equivalent* if there is a bijection on $\mathcal{L}$ that maps one into the other. We also let $\ell(w) = n$ be the *length* of the word and $\mathrm{wt}(w)$ the *weight* as the number of distinct letters of $\mathcal{L}$, and $\mathrm{supp}(w)$, the *support* of the word, is the set of distinct letters which appear. If $\mathcal{L} = \{1, 2, \ldots, N\}$ we often use the terminology N-word, or if the set $\mathcal{L}$ is clear, we just say "word".

Definition 17. Given any word $w = s_1 s_2 \ldots s_n$, we define the *graph associated with the word* w by $G_w = (V_w, E_w)$ be the graph with $V_w = \mathrm{supp}(w)$ the set of letters appearing in w and with edges $E_w = \{\{s_i, s_{i+1}\} : 1 \leq i \leq n-1\}$. The edge set can be divided into *self-edges* $E_w^s = \{\{u, u\} : u \in V_w\}$ and *connecting edges* $E_w^c = E_w - E_w^s$. Notice that two word are equivalent if and only if the corresponding graphs are isomorphic.

Definition 18. The graph G_w is connected because the word w, when read in order, gives a spanning path. For $e \in E_w$ we let N_e denote the number of time this path crosses the edge e (in any direction).

Remark 19. The tuple $\mathbf{i} = (i_1, \ldots, i_k)$ that appears in the evaluation of $\bar{T}_{\mathbf{i}}^N$ defines $w_{\mathbf{i}} = i_1 i_2 \ldots i_k i_1$ a closed word of length $k+1$ on $\mathcal{L} = \{1, 2, \ldots, N\}$. If we let $\mathrm{wt}_{\mathbf{i}}$ be the weight of this word, then the independence of the entries of the matrix X and the fact they are all identically distributed lets us write (Recall that X is scaled by $\frac{1}{\sqrt{N}}$ and that it is symetric):

$$\bar{T}_{\mathbf{i}}^N = \frac{1}{N^{k/2}} \prod_{e \in E_{\mathbf{i}}^c} \mathbf{E}\left(Z_{1,2}^{N_e}\right) \prod_{e \in E_{\mathbf{i}}^s} \mathbf{E}\left(Y_1^{N_e}\right)$$

Since the Z's and Y's are mean zero, this product is non-zero only if $N_e \geq 2$ for all $e \in E_{\mathbf{i}}$. This forces that $\mathrm{wt}_{\mathbf{i}} \leq \frac{k}{2} + 1$. We also see from this that equivalent words (in the sense of isomorphisms on $\mathcal{L} = \{1, 2, \ldots, N\}$) have the same value for$\bar{T}_{\mathbf{i}}^N$.

Definition 20. Let $S_{k,t}$ denote the set of all closed words of length $k+1$ on $\mathcal{L} = \{1, 2, \ldots, t\}$ and weight equal to t (i.e. every letter in $\{1, 2, \ldots, t\}$ is used at least once) and which have the property that $N_e \geq 2$ for every edge $e \in E_w$. Let $\mathcal{W}_{k,t}$ be the set of representatives of equivalence classes of $S_{k,t}$ under the equivalence of words (which we recall corresponds to isomorphisms of $\mathcal{L}$)

Remark 21. For every every representative $w \in \mathcal{W}_{k,t}$ there are $C_{N,t} = N(N-1)(N-2)\cdots(N-t+1)$ words on the set $\mathcal{L} = \{1, 2, \ldots, N\}$ that are equivalent to w. All these words $\mathbf{i}$ will have the same value for $\bar{T}_{\mathbf{i}}^N$.

Proposition 22. *From these definitions we have that:*

$$\langle \bar{L}_N, x^k \rangle = \sum_{t=1}^{\lfloor k/2 \rfloor + 1} \frac{C_{N.t}}{N^{k/2+1}} \sum_{w \in \mathcal{W}_{k,t}} \prod_{e \in E_w^c} \mathbf{E}\left(Z_{1,2}^{N_e}\right) \prod_{e \in E_w^c} \mathbf{E}\left(Y_1^{N_e}\right)$$

Remark 23. Notice that $|\mathcal{W}_{k,t}| \leq t^k$ is less than the number of closed words of length $k+1$ from $\mathcal{L} = \{1, 2, \ldots, t\}$. For fixed k, this means that in our sum $|\mathcal{W}_{k,t}| \leq t^k \leq (\lfloor k/2 \rfloor + 1)^k = C = cons't$. Since the moments of Z and Y are finite, this means that we have the bound:

$$\bar{T}_{\mathbf{i}}^N \leq C \sum_{t=1}^{\lfloor k/2 \rfloor + 1} \frac{C_{N.t}}{N^{k/2+1}}$$

Since $C_{N.t} = N(N-1)(N-2)\cdots(N-t+1) = O(N^t)$ this logic us that for k odd, $\bar{T}_{\mathbf{i}}^N \to 0$ (since $\lfloor k/2 \rfloor < k/2$, so $\frac{C_{N,t}}{N^{k/2+1}} \to 0$ for every t in our sum) and that for k even, only the term where $t = k/2+1$ survives in the limit $N \to \infty$, and the coefficient in front has $\frac{C_{N.k/2+1}}{N^{k/2+1}} \to 1$ We have then the formulas:

$$\begin{aligned}
\lim_{N \to \infty} \langle \bar{L}_N, x^k \rangle &= 0 \text{ for } k \text{ odd} \\
\lim_{N \to \infty} \langle \bar{L}_N, x^k \rangle &= \sum_{w \in \mathcal{W}_{k,k/2+1}} \prod_{e \in E_w^c} \mathbf{E}\left(Z_{1,2}^{N_e}\right) \prod_{e \in E_w^c} \mathbf{E}\left(Y_1^{N_e}\right) \text{ for } k \text{ even}
\end{aligned}$$

This is the motivation for the definition of a Wigner word:

Definition 24. A closed word w of length $k+1 \geq 1$ is called a Wigner word if either $k=0$ or k is even and w is equivalent to an element of $\mathcal{W}_{k,k/2+1}$. These are the only words that appear in the evaluation of the moments of $\lim_{N\to\infty}\left\langle \bar{L}_N, x^k\right\rangle$.

Proposition 25. *For a Wigner word w, the graph G_w is a tree with no loops. Moreover, every edge $e \in E_w$ has $N_e = 2$.*

Proof. G_w is connected, and has $|V_w| = k/2+1$, so it suffices to prove that $|E_w| = k/2$ to see that G_w is a tree. Indeed, $|E_w| \geq |V_w| - 1 = k/2$ or else G_w cannot be connected, and $|E_w| \leq k/2$ by the condition that $N_e \geq 2$ for each $e \in E_w$ (since $\sum_e N_e = k$ is the length of the path). Hence $|E_w| = k/2$ and it is a tree with no self loops. Moreover, since $2|E_w| = \sum_e N_e = k$, it must be the case that $N_e = 2$ for every edge e. □

Corollary 26. *For k even:*$\lim_{N\to\infty}\left\langle \bar{L}_N, x^k\right\rangle = \left|\mathcal{W}_{k,k/2+1}\right|$

Proof. This follows from the above formula for $\lim_{N\to\infty}\left\langle \bar{L}_N, x^k\right\rangle$, the proposition, and since $\mathbf{E}\left(Z_{1,2}^2\right) = 1$. □

Finally to see the lemma, we establish a bijection between $\mathcal{W}_{k,k/2+1}$ (Wigner words of length $k+1$) and rooted planar trees to establish the lemma.

Lemma 27. *For every $k \in \mathbb{N}$:*

$$\lim_{n\to\infty}\left\langle \bar{L}_n, x^k\right\rangle = \left\langle \sigma, x^k\right\rangle$$

Proof. From our above work, it suffices to create a bijection between $\mathcal{W}_{k,k/2+1}$ and rooted planar trees with $k+1$ vertices. Indeed, for every $w \in \mathcal{W}_{k,k/2+1}$ the graph G_w is a rooted planar tree where the ordering of the children of every node is the order in which they are visited in the reading of w. This is a bijection because if $G_w = G_{\tilde{w}}$ then they must be equilvalent words, so they are the same element from $\mathcal{W}_{k,k/2+1}$. The proof can also be seen by bijection to Dyck paths by using an *exploration process* between Dyck paths and planar rooted trees. □

We now work on the next lemma.

Definition 28. A pair partition of $\{1,\ldots,k\}$ is a partition of $\{1,\ldots,k\}$ where every partition is a pair (i.e. every partition consists of exactly 2 elements) A non-crossing pair partition is a pair partition that is also a non-crossing partition.

Fact 29. *The number of pair partitions of $\{1,\ldots,2k\}$ is C_k, the $k-$th Catalan number. (Compare this to the fact that there are C_k non-crossing partitions of $\{1,\ldots,k\}$)*

Proposition 30. *Given a Wigner word $w = i_1 i_2 \ldots i_{k+1}$of length $k+1$, let Π_w be the partition of $\{1,\ldots,k\}$ generated by the function $f : \{1,\ldots,k\} \to E_w$ by $f(j) = \{i_j, i_{j+1}\}$. (A partition of $\{1,\ldots,k\}$ means that the blocks are defined by $f^{-1}(\{a\})$ as a ranges in the possible target space.) Then the following hold:*

(1) Π_w is a non-crossing pair partition.
(2) Every non-crossing pair partition of $\{1,\ldots,k\}$ is of the form Π_w for some Wigner word w of length $k+1$
(3) If two Wigner words w and $w'$of length $k+1$ satisfy $\Pi_w = \Pi_{w'}$, then w and w' are equivalent.

Proof. □

(1) Because every Wigner word w can be viewed as a walk on the graph $G_w$and each edge is crossed exactly twice in the graph, so$\Pi_w$is a pair partition. Because the graph G_w is a tree, the partition Π_w is non-crossing. (Say $\{a,b\}$and $\{x,y\}$ are partitions, so that $f(a) = f(b) = e$ and $f(x) = f(y) = d$. WOLOG $a < x$. Suppose by contradiction $a < x < b < y$. Then in the walk on G_w, we encontour e first, then d, and then e again. But there is no way to get back to d again since we would have to cross e again. The argument is the same as showing that the exploration process of a tree gives a dyck path.)
(2) Every non-crossing pair partition can be turned into a rooted planar tree by an exploration. Every pair in the partition corresponds to crossing an edge on the tree twice. This rule can be used to recursivly build the tree up. As we've seen before, these trees are in bijection with Wigner words.
(3) In this case the trees they define would be the same, and so the words would be isomorphic.

Remark 31. To prove 14 on page 3, by Chebyshev's inequality it is enough to prove that the variance goes to zero:

$$\lim_{N\to\infty}\left|\mathbf{E}\left(\left\langle L_N, x^k\right\rangle^2\right) - \left\langle \bar{L}_N, x^k\right\rangle^2\right| = \lim_{N\to\infty}\mathbf{Var}\left(\left\langle L_N, x^k\right\rangle\right) = 0$$

Proceeding as in the above computation for the moments of$\bar{L}_N$ going through the trace of the N-th power of the matrix, since $\langle \bar{L}_N, x^k \rangle = \frac{1}{N}\sum_{\mathbf{i}} T_{\mathbf{i}}^N$, we have that:

$$\begin{aligned} \left| \mathbf{E}\left(\langle L_N, x^k \rangle^2 \right) - \langle \bar{L}_N, x^k \rangle^2 \right| &= \mathbf{Var}\left(\langle L_N, x^k \rangle \right) \\ &= \frac{1}{N^2} \sum_{\substack{i_1,\ldots i_k=1 \\ i'_1 \ldots i'_k=1}}^{N} \bar{T}_{\mathbf{i},\mathbf{i}'}^N \end{aligned}$$

With:

$$\begin{aligned} \bar{T}_{\mathbf{i},\mathbf{i}'}^N &= \mathbf{E}\left(T_{\mathbf{i}}^N T_{\mathbf{i}'}^N \right) - \mathbf{E}\left(T_{\mathbf{i}}^N \right) \mathbf{E}\left(T_{\mathbf{i}'}^N \right) \\ &= \mathbf{Cov}\left(T_{\mathbf{i}}^N, T_{\mathbf{i}'}^N \right) \end{aligned}$$

To study this we need to further our setup to account for *pairs* of words. We do this below.

Definition 32. Given a set $\mathcal{L}$, a $\mathcal{L}$-sentence a is a finite sequence of $\mathcal{L}$-words $w_1 w_2 \ldots w_n$ at least one word long. Two $\mathcal{L}$sentences are called equivalent if there is a bijection on $\mathcal{L}$ that maps one to another.

Definition 33. The graph associated with an $\mathcal{L}$-sentence is obtained by piecing together (taking the union) of the graphs associated with the individual words. Notice that this may be disconnected!

Now to evaluate $\bar{T}_{\mathbf{i},\mathbf{i}'}^N$, we notice that the pair $\mathbf{i},\mathbf{i}'$defines a two word sentence $a = w_{\mathbf{i}} w_{\mathbf{i}'}$, and we have:

$$\begin{aligned} \bar{T}_{\mathbf{i},\mathbf{i}'}^N = \frac{1}{N^k} \Bigg[& \prod_{e \in E^c_{a_{\mathbf{i},\mathbf{i}'}}} \mathbf{E}\left(Z_{1,2}^{N_e^a} \right) \prod_{e \in E^s_{a_{\mathbf{i},\mathbf{i}'}}} \mathbf{E}\left(Y_1^{N_e^a} \right) \\ & - \prod_{e \in E^c_{w_{\mathbf{i}}}} \mathbf{E}\left(Z_{1,2}^{N_e^{w_{\mathbf{i}}}} \right) \prod_{e \in E^s_{w_{\mathbf{i}}}} \mathbf{E}\left(Y_1^{N_e^{w_{\mathbf{i}}}} \right) \prod_{e \in E^c_{w_{\mathbf{i}'}}} \mathbf{E}\left(Z_{1,2}^{N_e^{w_{\mathbf{i}'}}} \right) \prod_{e \in E^s_{w_{\mathbf{i}'}}} \mathbf{E}\left(Y_1^{N_e^{w_{\mathbf{i}'}}} \right) \Bigg] \end{aligned}$$

Since these random variables are mean zero, we notice that $\bar{T}_{\mathbf{i},\mathbf{i}'}^N$is zero unless $N_e^a \geq 2$ for all $e \in E_{a_{\mathbf{i},\mathbf{i}'}}$. (It helps here to notice that $N_e^a \geq N_e^w$ for $w = w_{\mathbf{i}}$ or $w = w_{\mathbf{i}'}$). Also, if $E_{w_{\mathbf{i}}} \cap E_{w_{\mathbf{i}'}} = \emptyset$, the graphs $G_{w_{\mathbf{i}}}$ and $G_{w_{\mathbf{i}'}}$ are disjoint and so $\bar{T}_{\mathbf{i},\mathbf{i}'}^N = \mathbf{Cov}\left(T_{\mathbf{i}}^N, T_{\mathbf{i}'}^N \right) = 0$. (This can also be seen directly from the formula above with a bit of care).

As before, to evaluate this we look only at representatives from an equivalence class of the relevant sentences. Define $\mathcal{W}_{k,t}^{(2)}$ to be the set of representatives for equivalence clases of sentences a of weight t (recall weight is the number of distinct letters used) that consist of two closed t-words $w_1 w_2$ each of length $k+1$ and with $N_e^a \geq 2$ for each $e \in E_a$ and with $E_{w_1} \cap E_{w_2} \neq \emptyset$. With this definition we have a formula for $\bar{T}_{\mathbf{i},\mathbf{i}'}^N$ akin to the one for $\bar{T}_{\mathbf{i}}^N$ we had earlier. Here $C_{N,t}$ is again the number of sentences in each equivalence class. Have:

$$\begin{aligned} \mathbf{Var}\left(\langle L_N, x^k \rangle \right) = \sum_{t=1}^{2k} \frac{C_{N,t}}{N^{k+2}} \sum_{a=(w_1,w_2) \in \mathcal{W}_{k,t}^{(2)}} \Bigg[& \prod_{e \in E^c_a} \mathbf{E}\left(Z_{1,2}^{N_e^a} \right) \prod_{e \in E^s_a} \mathbf{E}\left(Y_1^{N_e^a} \right) \\ & - \prod_{e \in E^c_{w_1}} \mathbf{E}\left(Z_{1,2}^{N_e^{w_1}} \right) \prod_{e \in E^s_{w_1}} \mathbf{E}\left(Y_1^{N_e^{w_2}} \right) \prod_{e \in E^c_{w_2}} \mathbf{E}\left(Z_{1,2}^{N_e^{w_2}} \right) \prod_{e \in E^s_{w_2}} \mathbf{E}\left(Y_1^{N_e^{w_2}} \right) \Bigg] \end{aligned}$$

To show that this goes to zero as $N \to \infty$, since the products in square brackets are bounded for fixed k, it suffices to show that $\mathcal{W}_{k,t}^{(2)}$ is empty for $t \geq k+2$, because in this case the coefficient $\frac{C_{N,t}}{N^{k+2}} \to 0$ for $t < k+2$ will take the whole sum to zero in the limit $N \to \infty$. We will actually prove a stronger claim, that $\mathcal{W}_{k,t}^{(2)}$ is empty for $t \geq k+1$, as this result will be useful later.

Proposition 34. $\mathcal{W}_{k,t}^{(2)}$ is empty for $t \geq k+1$.

Proof. Let $a = w_1 w_2 \in \mathcal{W}_{k,t}^{(2)}$. Then G_a is a connected graph (since $E_{w_1} \cap E_{w_2} \neq \emptyset$) with t vertices and at most k edges (since $N_e^a \geq 2$ for $e \in E_a$), which is impossible for $t \geq k+2$ (Since a tree has the minimal number of edges for a fixed number of vertices, and a tree would have $t = k+1$) When $t = k+1$, it must be that G_a is a tree and that $N_e^a = 2$ for every edge e. But the path generated by w_1 starts and ends at the same vertex, so it must visit each edge on its route at least twice. The path generated by w_2 also visits each edge on its route at least twice. Since each edge is visited exactly twice by these two routes ($N_e^a = 2$ for all edges e), it must be that the paths from w_1

and w_2 are disjoint. But then a is disconnected! This contradicts the existence of such an element $a \in \mathcal{W}_{k,t}^{(2)}$ when $t = k+1$. □

Lemma 35. *For every $k \in \mathbb{N}$ and every $\epsilon > 0$:*

$$\lim_{n\to\infty} \mathbf{P}\left(\left|\langle L_n, x^k\rangle - \langle \bar{L}_n, x^k\rangle\right| > \epsilon\right) = 0$$

Proof. By Chebyshev, $\mathbf{P}\left(\left|\langle L_n, x^k\rangle - \langle \bar{L}_n, x^k\rangle\right| > \epsilon\right) \le \frac{1}{\epsilon^2}\mathbf{E}\left(\left|\langle L_n, x^k\rangle - \langle \bar{L}_n, x^k\rangle\right|^2\right) = \frac{1}{\epsilon^2}\mathbf{Var}\left(\langle L_n, x^k\rangle\right)$ so it suffices to show that this goes to zero as $N \to \infty$. We have that:

$$\begin{aligned}\mathbf{Var}\left(\langle L_N, x^k\rangle\right) &= \sum_{t=1}^{2k} \frac{C_{N,t}}{N^{k+2}} \sum_{a=(w_1,w_2)\in\mathcal{W}_{k,t}^{(2)}} \left[\prod_{e\in E_a^c} \mathbf{E}\left(Z_{1,2}^{N_e^a}\right) \prod_{e\in E_a^s} \mathbf{E}\left(Y_1^{N_e^a}\right)\right. \\ &\quad \left. - \prod_{e\in E_{w_1}^c} \mathbf{E}\left(Z_{1,2}^{N_e^{w_1}}\right) \prod_{e\in E_{w_1}^s} \mathbf{E}\left(Y_1^{N_e^{w_2}}\right) \prod_{e\in E_{w_2}^c} \mathbf{E}\left(Z_{1,2}^{N_e^{w_2}}\right) \prod_{e\in E_{w_2}^s} \mathbf{E}\left(Y_1^{N_e^{w_2}}\right)\right]\end{aligned}$$

When $t \le k$ in the sum, the coefficient $\frac{C_{N,t}}{N^{k+2}} \to 0$ and the terms in square brackets are bounded (by something like k^k), so these terms vanish in the limit $N \to \infty$. When $t \ge k+1$, $\mathcal{W}_{k,t}^{(2)}$ is empty, so these terms are always zero. Hence the whole sum vanishes as $N \to \infty$. □

1.2. GOE and the GUE.

Definitions of different letters

(1) $\underline{\beta = 1}$. We use the field $\mathbb{F} = \mathbb{R}$, $\mathcal{H}_N^{(1)} \subset \mathrm{Mat}_N(\mathbb{R})$ is the space of *real symmetric* $N \times N$ matrices ($A^\intercal = A$). $\mathcal{U}_N^{(1)}$ is the set of $N \times N$ *orthogonal* matrices ($A^\intercal A = Id$).
(2) $\underline{\beta = 2}$. We use the field $\mathbb{F} = \mathbb{C}$, $\mathcal{H}_N^{(2)} \subset \mathrm{Mat}_N(\mathbb{C})$ is the space of *complex Hermitian* $N \times N$ matrices ($A^* = A$). $\mathcal{U}_N^{(2)}$ is the set of $N \times N$ *unitary* matrices ($A^* A = Id$).
$\mathcal{D}_N$ is the set of $N \times N$ diagonal matrices with real entries.

Definition 36. (GOE) Let $\{\xi_{i,j}\}_{i,j=1}^{\infty}$ be an i.i.d. family of $N(0,1)$ random variables. Define $P_N^{(1)}$ to be the law of the random $N \times N$ real symmetric matrix with $X_{ii} = \sqrt{2}\xi_{ii}$ on the diagonal, and $X_{ji} = X_{ij} = \xi_{i,j}$ above the diagonal. This is a probability measure on $\mathcal{H}_N^{(1)}$ For example:

$$P_3^{(1)} \overset{D}{\equiv} \begin{bmatrix} \sqrt{2}\xi_{11} & \xi_{12} & \xi_{13} \\ \xi_{12} & \sqrt{2}\xi_{22} & \xi_{23} \\ \xi_{13} & \xi_{23} & \sqrt{2}\xi_{33} \end{bmatrix} \in \mathcal{H}_3^{(1)}$$

A random matrix $X \in \mathcal{H}_N^{(1)}$ with law $P_N^{(1)}$ is said to belong to the *Gaussian Orthogonal Ensemble (GOE)*

Definition 37. (GUE) Let $\{\xi_{i,j}, \eta_{i,j}\}_{i,j=1}^{\infty}$ be an i.i.d. family of $N(0,1)$ random variables. Define $P_N^{(2)}$ to be the law of the random $N \times N$ complex Hermitian matrix with $X_{ii} = \xi_{ii}$ on the diagonal, and $\bar{X}_{ji} = X_{ij} = \frac{1}{\sqrt{2}}(\xi_{i,j} + \imath\eta_{i,j})$ above the diagonal (Here $\imath = \sqrt{-1}$ is the imaginary unit). This is a probability measure on $\mathcal{H}_N^{(2)}$ For example:

$$P_3^{(2)} \overset{D}{\equiv} \begin{bmatrix} \xi_{11} & \frac{\xi_{12}+\imath\eta_{12}}{\sqrt{2}} & \frac{\xi_{13}+\imath\eta_{13}}{\sqrt{2}} \\ \frac{\xi_{12}-\imath\eta_{12}}{\sqrt{2}} & \xi_{22} & \frac{\xi_{23}+\imath\eta_{23}}{\sqrt{2}} \\ \frac{\xi_{13}-\imath\eta_{13}}{\sqrt{2}} & \frac{\xi_{23}-\imath\eta_{23}}{\sqrt{2}} & \xi_{33} \end{bmatrix} \in \mathcal{H}_3^{(2)}$$

A random matrix $X \in \mathcal{H}_N^{(2)}$ with law $P_N^{(2)}$ is said to belong to the *Gaussian Unitary Ensemble (GUE)*

Remark 38. Both the GUE and GOE give examples of Wigner matrices. In particular Wigner's theorem applies and lets us conclude that the empirical distribution of the eigenvalues for the random matrix $\frac{1}{\sqrt{N}}X_N \in \mathcal{H}_N^{(\beta)}$ converges to the semicircle law we found earlier.

Remark 39. What makes the GOE and the GUE special? What do they have to do with orthongonal or unitary matrices? (Recall orthogonal matrices have $X^\intercal X = Id$ while unitary matrices have $X^* X = Id$). The answer is that the built in symmetry of the definition makes it so that the GOE and GUE are invariant under orthogonal matrices and unitary matrices respectively. One way you can see this is to calculate the probability density of the measure $P_N^{(\beta)}$ with respect to the Lebesgue measure. We do this below.

Definition 40. ($\underline{\ell_N^{(\beta)}}$ when$\beta = 1$) Let $\underline{\ell_N^{(1)}}$ be the *Lebesgue measure* on $\mathcal{H}_N^{(1)}$ which is defined by the pullback of the Lebesgue measure on $\mathbb{R}^{N(N+1)/2}$ through the the one-to-one and onto map $\mathcal{H}_N^{(1)} \to \mathbb{R}^{N(N+1)/2}$ defined by taking the on-or-above the diagonal entries of a matrix in $\mathcal{H}_N^{(1)}$ as coordinates in $\mathbb{R}^{N(N+1)/2}$

Definition 41. ($\underline{\ell_N^{(\beta)}}$ when$\beta = 2$) Let $\underline{\ell_N^{(2)}}$ be the *Lebesgue measure* on $\mathcal{H}_N^{(2)}$ which is defined by the pullback of the Lebesgue measure on $\mathbb{R}^{N^2}$ through the the one-to-one and onto map $\mathcal{H}_N^{(2)} \to \mathbb{R}^N \times \mathbb{C}^{N(N-1)/2} \approxeq \mathbb{R}^{N^2}$ defined by taking the on-or-above the diagonal entries of a matrix in $\mathcal{H}_N^{(2)}$ as coordinates in $\mathbb{R}^N \times \mathbb{C}^{N(N-1)/2}$

Definition 42. The probability density (in the sense of Radon-Nikodym derivative) of $P_N^{(\beta)}$ with respect to the Lebesgue measure $\ell_N^{(\beta)}$ is given by:

$$\begin{aligned}\frac{\mathrm{d}P_N^{(1)}}{\mathrm{d}\ell_N^{(1)}}(H) &= 2^{-N/2}(2\pi)^{-N(N+1)/4}\exp\left(-\mathrm{Tr}\left(H^2\right)/4\right)\\ \frac{\mathrm{d}P_N^{(2)}}{\mathrm{d}\ell_N^{(2)}}(H) &= 2^{-N/2}\pi^{-N^2/2}\exp\left(-2\mathrm{Tr}\left(H^2\right)/4\right)\end{aligned}$$

In particular, we notice that the denisty depends only on the trace of the matrix squared.

Proof. This is a straightforward computation using the independence of the entries and the fact that they are Gaussian.

($\beta = 1$)Notice that $\mathrm{Tr}\left(H^2\right) = \mathrm{Tr}\left(H^\intercal H\right) = \sum_{i=1}^N H_{i,i}^2 + 2\sum_{1\le i<j\le N} H_{i,j}^2$. Now, since each $\xi_{i,j}$ is Gaussian and independent, we have that:

$$\begin{aligned}\frac{\mathrm{d}P_N^{(1)}}{\mathrm{d}\ell_N^{(1)}}(H) &= \prod_{i=1}^N \rho\left(\sqrt{2}\xi_{i,i} = H_{i,i}\right)\prod_{1\le i<j\le N}^N \rho\left(\xi_{i,j} = H_{i,j}\right)\\ &= \prod_{i=1}^N \frac{1}{\sqrt{2}}\rho\left(\xi_{i,i} = \frac{H_{i,i}}{\sqrt{2}}\right)\prod_{1\le i<j\le N}^N \rho\left(\xi_{i,j} = H_{i,j}\right)\\ &= \frac{1}{\sqrt{2\pi}^{N(N+1)/2}}\frac{1}{\sqrt{2}^N}\exp -\frac{1}{2}\left(\sum_{i=1}^N \frac{H_{i,i}^2}{2} + \sum_{1\le i<j\le N} H_{i,j}^2\right)\end{aligned}$$

Which gives the desired result when we use the expression for trace.

($\beta = 2$) Firsty, notice that

$$\begin{aligned}\mathrm{Tr}\left(H^2\right) &= \mathrm{Tr}\left(H^* H\right) = \sum_{i=1}^N H_{i,i}^2 + 2\sum_{1\le i<j\le N}\left|H_{i,j}\right|^2\\ &= \sum_{i=1}^N H_{i,i}^2 + 2\sum_{1\le i<j\le N}\mathrm{Re}H_{i,j}^2 + 2\sum_{1\le i<j\le N}\mathrm{Im}H_{i,j}^2\end{aligned}$$

Now, since each $\xi_{i,j}, \eta_{i,j}$ is Gaussian and independent, we have that:

$$\begin{aligned}\frac{\mathrm{d}P_N^{(2)}}{\mathrm{d}\ell_N^{(2)}}(H) &= \prod_{i=1}^N \rho\left(\xi_{i,i} = H_{i,i}\right)\prod_{1\le i<j\le N}^N \rho\left(\frac{\xi_{i,j}}{\sqrt{2}} = \mathrm{Re}H_{i,j}\right)\prod_{1\le i<j\le N}^N \rho\left(\frac{\eta_{i,j}}{\sqrt{2}} = \mathrm{Im}H_{i,j}\right)\\ &= \prod_{i=1}^N \rho\left(\xi_{i,i} = H_{i,i}\right)\prod_{1\le i<j\le N}^N \sqrt{2}\rho\left(\xi_{i,j} = \sqrt{2}\mathrm{Re}H_{i,j}\right)\prod_{1\le i<j\le N}^N \sqrt{2}\rho\left(\eta_{i,j} = \sqrt{2}\mathrm{Im}H_{i,j}\right)\\ &= \frac{\sqrt{2}^{2N(N+1)/2}}{\sqrt{2\pi}^{N^2}}\exp -\frac{1}{2}\left(\sum_{i=1}^N H_{i,i}^2 + \sum_{1\le i<j\le N} 2\mathrm{Re}H_{i,j}^2 + \sum_{1\le i<j\le N} 2\mathrm{Im}H_{i,j}^2\right)\end{aligned}$$

Which gives the desired result when we use the identity for trace. □

Definition 43. For $x_1, x_2, \ldots, x_N \in \mathbb{C}^N$ we define the <u>*Vandermonde determinant*</u> associated with x by:

$$\Delta(x) = \det\left(\{x_i^{j-1}\}_{i,j}\right) = \prod_{i<j}\left(x_j - x_i\right)$$

Theorem 44. *[Joint Distribution of the Eigenvalues of the GOE/GUE] Let $X \in \mathcal{H}_N^{(\beta)}$ be a random matrix with law $P_N^{(\beta)}$, with $\beta = 1, 2$. The joint distribution of the egienvalues $\lambda_1(X) \leq \ldots \leq \lambda_N(X)$ has density with respect to the Lebesgue measure which equals:*

$$N!\bar{C}_N^{(\beta)} 1_{\{x_1 \leq \ldots \leq x_n\}} |\Delta(x)|^\beta \prod_{i=1}^N e^{-\beta x_i^2/4}$$

Here $\bar{C}_N^{(\beta)}$ is a normalizing constant. This constant is:

$$\begin{aligned} N!\bar{C}_N^{(\beta)} &= N! \left(\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} |\Delta(x)|^\beta \prod_{i=1}^N e^{-\beta x_i^2/4} dx_i \right)^{-1} \\ &= (2\pi)^{-N/2} \left(\frac{\beta}{2}\right)^{\beta N(N-1)/4 + N/2} \prod_{j=1}^N \frac{\Gamma(\beta/2)}{\Gamma(j\beta/2)} \end{aligned}$$

Here$\Gamma(s) = \int_0^\infty x^{s-1} e^{-x} dx$ is Euler's Gamma function.

Remark 45. From this we can easily see the density for the *unordered* eigenvalues has density $\mathcal{P}_N^{(\beta)}$ on $\mathbb{R}^N$ with density:

$$\frac{\mathrm{d}\mathcal{P}_N^{(\beta)}}{\mathrm{dLeb}_N} = \bar{C}_N^{(\beta)} |\Delta(x)|^\beta \prod_{i=1}^N e^{-\beta x_i^2/4}$$

Remark 46. A consequence of 44 is that the probability of a repeated eigenvalue is zero (since Δ vanishes if $x_i = x_j$ for any i, j). This means that every eigenspace is one dimensional. Let $v_1, v_2, \ldots, v_n$ be the basis of eigenvectors for the matrix X that is normalized so that the first coordinate of v_i is real and positive and so that $|v_i| = 1$. The invariance of X under arbitrary orthogonal (when $\beta = 1$) or unitary (when $\beta = 2$) transformations means that matrix constructed from the eigenvectors $[v_1, \ldots, v_n]$ is distributed like the *Haar measure* on the set of orthogonal (when $\beta = 1$) or unitary (when $\beta = 2$) matrices. (Recall the Haar measure is the unique measure that is invariant under the action of the matrices, i.e. $\mathrm{d}\mu(U_0 U) = \mathrm{d}\mu(U)$). In particular, any single vector, say v_1, is distributed uniformly on the set of norm one vectors whose first coordinate is real and positive.

Corollary 47. *Let $S_+^{N-1} = \{\mathbf{x} = (x_1, \ldots, x_n) : x_i \in \mathbb{R}, \|\mathbf{x}\|_2 = 1, x_1 > 0\}$. Then v_1 is uniformly distibuted in $S_+^{N-1}$for the GOE or in $S_{\mathbb{C},+}^{N-1} = \{\mathbf{x} = (x_1, \ldots, x_n) : x_1 \in \mathbb{R}, x_i \in \mathbb{C}, i \geq 2, \|\mathbf{x}\|_2 = 1, x_1 > 0\}$for the GUE. Furthermore, $v_1, \ldots, v_n$ is distributed like a sample of the Haar measure on orthogonal or unitary matrices, with each column multiplied by a norm one scalar so each column belongs to S_+^{N-1}(for the GOE) and $S_{\mathbb{C},+}^{N-1}$ (for the GUE).*

Proof. Write $X = UDU^*$. Let T be an orthogonal (for GOE) or unitary (for GUE) matrix distributed like the Haar measure. The consider TXT^*. This has the same eigenvalues as X! Since the law of X depends only on the eigenvalues of X, we know $TXT^* \overset{D}{\equiv} X$. Moreover, since $TU \overset{D}{\equiv} T$ (by definition of the Harr measure), we have then that $X \overset{D}{\equiv} TDT^*$, i.e. the orthogonal (for GOE) or unitary (for GUE) matrix that diagonalizes X is distributed like the Haar measure. The columns of this matrix make a basis of eigenvectors, and multiplying each by a norm one constant normalizes them the way we want. ☐

Remark 48. We will now write an outline of the proof for 44 of the distribution of the eigenvalues for the GOE/GUE. For any $X \in \mathcal{H}_N^{(\beta)}$ (i.e. X is symmetric/Hermitian), write $X = UDU^*$ with $U \in \mathcal{U}_N^{(\beta)}$ (i.e. U is orthogonal/unitary) and $D \in \mathcal{D}_N^{(\beta)}$ (i.e. D diagonal with real entries). Suppose the map $\mathcal{H}_N^{(\beta)} \to \mathcal{U}_N^{(\beta)} \times \mathcal{D}_N$ was a bijection. (It turns out it is not a bijection because of the possibility of repeated eigenvalues; we will make an argument to get around this later). Then one could paramaterize $\mathcal{U}_N^{(\beta)}$ using $\beta N(N-1)/2$ parameters in a smooth way ($N + \beta N(N-1)/2$ real parameters to parametrize $\mathcal{H}_N^{(\beta)}$ and subtract the N degrees of freedom coming from $\mathcal{D}_N$) . An easy computation shows that the Jacobian of the transformation would then be a polynomial of degree $\beta N(N-1)/2$ in the eigenvalues of X with coefficients that are functions of the parametrization for $\mathcal{U}_N^{(\beta)}$. Since the Jacobian must vanish on the set where there are repeated eigenvalues, we know the roots of this polynomial! Symmetry and degree considerations then show that the Jacobian must be proportional to the factor $\Delta(x)^\beta$. Integrating out the parametrization for $\mathcal{U}_N^{(\beta)}$ then gives 44.

To get around the fact that the map $\mathcal{H}_N^{(\beta)} \to \mathcal{U}_N^{(\beta)} \times \mathcal{D}_N$ is not actually a bijection, we have to ignore some measure zero sets. We do this below.

Definition 49. Let $\mathcal{U}_N^{(\beta),g} = \left\{ U \in \mathcal{U}_N^{(\beta)} : \text{every diagonal entry of } U \text{ is a strictly positive real and every entry of } U \text{ is non-zero} \right\}$
We call these "good unitary matrices" or just "good"

Let $\mathcal{D}_N^d = \{D \in \mathcal{D}_N : \text{every entry of } D \text{ is distinct}\}$ We call these "distinct diagonal matrices" or just "distinct".

Let $\mathcal{D}_N^{do} = \{D \in \mathcal{D}_N : \text{every entry of } D \text{ is distinct and they are decreasing as we go down the diagonal}\}$ We call these "distinct ordered diagonal matrices" or just "distinct ordered".

Let $\mathcal{H}_N^{(\beta),dg} = \left\{ H \in \mathcal{H}_N^{(\beta)} : H = UDU^* \text{where } D \in \mathcal{D}_N^d \text{ and } U \in \mathcal{U}_N^{(\beta),g} \right\}$.

Lemma 50. *The set $\mathcal{H}_N^{(\beta)} \backslash \mathcal{H}_N^{(\beta),dg}$ can be thought of as a subset of $\mathbb{R}^{N(N+1)/2}$ or $\mathbb{R}^{N^2}$ as described before. This is a null set with repsect to the Lebesgue measure. Furthermore, the map $\mathcal{D}_N^{do} \times \mathcal{U}_N^{(\beta),g} \to \mathcal{H}_N^{(\beta),dg}$ given by $(D,U) \to UDU^*$ is one-to-one and onto, and the map $\mathcal{D}_N^{do} \times \mathcal{U}_N^{(\beta),g} \to \mathcal{H}_N^{(\beta),dg}$ given by the same map $N!$-to-one.*

Proof. Firstly, notice that for a non-trivial polynomial $p : \mathbb{R}^k \to \mathbb{R}$, the set $\{X : p(X) = 0\} = p^{-1}(\{0\})$ is a closed set and is measure 0 with respect to Lebesgue measure on $\mathbb{R}^k$ (since it has at most k roots). Hence to prove the claim, it is enough to find a $p : \mathcal{H}_N^{(\beta)} \to \mathbb{R}$ which is a polynomial in the entries and so that $p(H) = 0$ for every $H \in \mathcal{H}_N^{(\beta)} \backslash \mathcal{H}_N^{(\beta),dg}$. Let us use the notation that $H^{(i,j)}$is the $N-1 \times N-1$ submatrix obtained from H by deleting the ith column and $j$the row of H .

Claim: Say $X = UDU^*$ for $D \in \mathcal{D}_N^d$ (so that X has distinct eigenvalues) and suppose that X and $X^{(k,k)}$ do not have any eigenvalues in common for $k = 1, 2, \ldots, N$. Then all the entries of U are nonzero.

Corollary: For every $H \in \mathcal{H}_N^{(\beta)} \backslash \mathcal{H}_N^{(\beta),dg}$, either has some repeated eigenvalues OR there is some k so that the $N-1 \times N-1$ submatrix $H^{(k,k)}$ shares an eigenvalue with the matrix H.

Pf of Corr: If H has repeated eigenvalues we are done. Otherwise, H has no repeated eigenvalues. Suppose by contradiction that X and $X^{(k,k)}$ do not have any eigenvalues in common for $k = 1, 2, \ldots, N$. Then by the claim $H = UXU^\star$ will have U with all nonzero entries. But then H has distince eigenvalues, and a good unitary matrix, so $H \in \mathcal{H}_N^{(\beta),dg}$ which is a contradiction.

Pf of Claim: Let λ be an eigenvalue of X , and let $A = X - \lambda I$. Define A^{adj}as the $N \times N$ matrix with $A_{i,j}^{adj} = (-1)^{i+j} \det\left(A^{(i,j)}\right)$.This is the *adjunct* matrix from linear algebra, which can be easily verified to have the property that $AA^{adj} = \det(A)I$. Since λ is an eigenvalue of X, $\det A = 0$ so we have $AA^{adj} = 0$. Now, the dimension of the nullspace of A is the dimension of the $\lambda-$eigenspace for X which is one (since the eigenvalues for X are distinct). Now, since $AA^{adj} = 0$, we know that each column of the adjugate is perpendicular to the row-space of A. Since the row-space of A is $N-1$dimensional, (by rank-nulltiy theorem), every column of A^{adj} is in the 1-dimensional space orthogonal to the row-space of A. Have then that every column of A^{adj} is a scalar multiple of some vector v_λ. Since $Av_\lambda = 0$, we know that v_λ is an eigenvector of X of eigenvalue λ. Since X and $X^{(k,k)}$ have no eigenvalues in common, we know that $\det\left(X^{(k,k)} - \lambda I\right) \neq 0$. Notice that by definition of A, we have $A_{i,i}^{adj} = \det\left(X^{(k,k)} - \lambda I\right) \neq 0$, so the diagonal entries of A^{adj} are all non-zero. Since the columns of A^{adj} are scalar multiples of v_λ it must be that v_λ has all non-zero components! (Or else if $v_\lambda(i) = 0$, the entire i-th row of A^{adj} would be zero, contradiction the above.) Finally, since each column of the diagonalizing matrix U is a scalar mulitple of some v_λ, we know every entry of U is non-zero.

Fact: For any polynomials $p, q : \mathbb{R}^n \to \mathbb{R}$, there is a function $f(p,q)$ which is *polynomial* in the *coefficients of* p, q so that $f(p,q) = 0$ if and only if p and q share a root. f is called the *resultant* of p, q. A corollary is that $f(p,p')$ is a polynomial in the coefficients of p which is zero if and only if p has a repeated root. f is called the *discriminant* of p, q

We are now ready to prove the lemma. Let p be the characteristic polynomial of H and let p_k be the characteristic polynomial of $H^{(k,k)}$. Let P_k be the resultant of p and p_k (This is a polynomial in the coefficients of p, p_k). Since roots of p, p_k correspond to eigenvalues of $H, H^{(k,k)}$, $P_k$is 0 if and only if H and $H^{(k,k)}$ share an eigenvalue. Let P_0 be the discriminant of p (again a polynomial in the coefficients of p) which is 0 if and only if H has a repeated root. Since the coeffcients of $p, p_k$are polynomial in the entries of the matrix H, P_k and P_0 are also polynomial in the entries of H. Let $Q = P_1 \cdot \ldots \cdot P_k \cdot P_0$. This is a polynomial in the entries of H. Moreover by the corollary to the claim, for $H \in \mathcal{H}_N^{(\beta)} \backslash \mathcal{H}_N^{(\beta),dg}$, H either has a repeated root (in which case $P_0 = 0$) or there is some k so that $H^{(k,k)}$shares an eigenvalue with H (in which case $P_k = 0$) . In either case $Q = 0$. So Q works as the polynomial we are looking for!

The one-to-one or $N!$-to-one nature of the map is clear because each eigenspace is of dimension 1 , and the choice of normalization in the definition of a "good" unitary matrix forces it. ☐

Definition 51. let $\mathcal{U}_N^{(\beta),vg} = \left\{ U \in \mathcal{U}_N^{(\beta),g} : \text{all minors of } U \text{ have nonvanishing determinant} \right\}$ be a subset of the good matrices. We call these *very good* unitary matrices. These matrices have a nice parametrization.

Lemma 52. *The map $T:\mathcal{U}_N^{(\beta),vg}\to\mathbb{R}^{\beta N(N-1)/2}$ (where we identify $\mathbb{C}\equiv\mathbb{R}^2$ for $\beta=2$) defined by:*

$$T(U)=\left(\frac{U_{1,2}}{U_{1,1}},\ldots,\frac{U_{1,N}}{U_{1,1}},\frac{U_{2,3}}{U_{2,2}},\ldots,\frac{U_{2,N}}{U_{2,2}},\ldots,\ldots,\frac{U_{N-1,N}}{U_{N-1,N-1}}\right)$$

is one-to-one with a smooth inverse. Furthermore, the set $\left(T(\mathcal{U}_N^{(\beta),vg})\right)^{\complement}$ is closed and has zero Lebesgue measure.
.... I scouted the rest of the proof but didnt write it up.

2. Intro To Fredholm Determinants

Definition 53. A Polish space is a complete metric space which is separable.

Definition 54. Let X be a Polish space and let $\mathcal{B}_X$ be its Borel sigma algebra. For a complex valued measure ν on $(X,\mathcal{B}_X)$ define:

$$\|\nu\|_1=\int_X|x|\,\mathrm{d}\nu(x)$$

We will consider only measures ν with $\|\nu\|<\infty$.

Definition 55. A *kernal* is a Borel measurable, complex-values function $K(x,y)$ defined on $X\times X$ with norm defined by:

$$\|K\|=\sup_{(x,y)\in X\times X}|K(x,y)|$$

The *trace* of a kernal $K(x,y)$ with respect to some measure ν is:

$$\mathrm{Tr}(K)=\int K(x,x)\mathrm{d}\nu(x)$$

Given two kernals $K(x,y)$ and $L(x,y)$ their *composition* $(K\star L)(x,y)$ is another kernal which is defined (with repect some measure ν) as:

$$(K\star L)(x,y)=\int K(x,z)L(z,y)\mathrm{d}\nu(z)$$

(These will be well defined (i.e. the integrals will be finite) as long as $\|K\|$ and $\|\nu\|_1$ are finite)

Proposition 56. *By Fubini $Tr(K\star L)=Tr(L\star K)$ and $(K\star L)\star M=K\star(L\star M)$*

Note 57. WARNING since K is not continuous it might be that $K=K'$almost everywhere but $\mathrm{Tr}(K)=\mathrm{Tr}(K')$ (they could differ on the "diagonal" which is a measure zero set)

Remark 58. If X is the space $\{1,2,\ldots,n\}$ and ν is the counting measure, this feels a lot like a matrix, with Tr being the trace and $\star$ being matrix multiplication. In fact, if we choose n points $x_1,\ldots,x_n$ and $y_1,\ldots y_n$ then $[K(x_i,y_j)]_{i,j}$ is a matrix. It might be interesting to take the determinant of the matrix.

Lemma 59. *Fix $n>0$ for any two kernals $F(x,y)$ and $G(x,y)$ we have:*

$$\left|\det_{i,j=1}^{n}F(x_i,y_j)-\det_{i,j=1}^{n}G(x_i,x_j)\right|\le n^{1+n/2}\|F-G\|\max\left(\|F\|,\|G\|\right)^{n-1}$$

And:

$$\left|\det_{i,j=1}^{n}F(x_i,y_j)\right|\le n^{n/2}\|F\|^n$$

Proof. Define:

$$H_i^{(k)}(x,y)=\begin{cases}G(x,y) & \text{if } i<k\\ F(x,y)-G(x,y) & \text{if } i=k\\ F(x,y) & \text{if } i>k\end{cases}$$

Since determinenats are linear in the rows, we can do some manipulation:

$$\det_{i,j=1}^{n}F(x_i,y_j)-\det_{i,j=1}^{n}G(x_i,x_j)=\sum_{k=1}^{n}\det_{i,j=1}^{n}H_i^{(k)}(x_i,y_j)$$

(This works by rewtriting the vector that appears in the top row of $F(x_i,y_j)$ as $\vec{F}=\left(\vec{F}-\vec{G}\right)+\vec{G}$ recursivly doing this with the top $\vec{F}$ that appears on the left hand side gives us exactly the restult above.) Applying Hadamards

Theorem now (Hadamard: $v_1, \ldots, v_n$ are column vectors of length n with complex entries, then $\det[v_1, \ldots, v_n] \leq \prod_{i=1}^n \sqrt{\bar{v}_i^T v_i} \leq n^{n/2} \prod |v_i|_\infty$.) We get:

$$\left| \det_{i,j=1}^{n} H_i^{(k)}(x_i, y_j) \right| \leq n^{n/2} \|F - G\| \max(\|F\|, \|G\|)^{n-1}$$

Which gives the desired inequality. In the case $G = 0$ the above determinant is 0 and we get the desired inequality on $\|F\|$alone. ☐

Definition 60. For a given kernal K and a measure ν define $\Delta_0 = 1$ as a convention and for $n > 0$:

$$\Delta_n = \Delta_n(K, \nu) = \int \ldots \int \det_{i,j=1}^{n} K(\xi_i, \xi_j) \mathrm{d}\nu(\xi_1) \ldots \mathrm{d}\nu(\xi_n)$$

By the inequality we just proved, $|\Delta_n| \leq (\|\nu\|_1)^n \|K\|^n n^{n/2}$ so the integral is well defined.\

Definition 61. The *Fredholm determinant* associated with the Kernal K is defined as:

$$\Delta(K) = \Delta(K, \nu) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n!} \Delta_n(K, \nu)$$

Remark 62. Here is some motivation for this being called a determinant. Let $f_1(x), \ldots, f_N(x), g_1(x), \ldots, g_N(x)$ be given. Put:

$$K(x, y) = \sum_{i=1}^{n} f_i(x) g_i(y)$$

This is a Kernal, and it will turn out that:

$$\Delta(K) = \det_{i,j=1}^{N} \left(\delta_{ij} - \int f_i(x) g_j(x) \mathrm{d}\nu(x) \right)$$

For this reason sometimes people use the notation that $\Delta(K) = \det(I - K)$.

I will now skip to Lemma 3.2.4 from the book where this expression appears.

3. Hermite Polynomials and the GUE

We are going to prove the following theorem, which tells us that the eigenvalues of the GUE have a probability density given by a Fredholm determinant. Precisely, we are working towards:

Theorem 63. *(Gaudin-Mehta) For any compact set $A \subset \mathbb{R}$:*

$$\begin{aligned} \lim_{N\to\infty} \mathbf{P}\left(\sqrt{N}\lambda_1^N, \ldots \sqrt{N}\lambda_2^N \notin A\right) &= 1 + \Delta(K_{sine}) \\ &= 1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{k} \int_A \ldots \int_A \det_{i,j=1}^{k} K_{sine}(x_i, x_j) \prod_{j=1}^{k} dx_j \end{aligned}$$

where:

$$K_{sine}(x, y) = \begin{cases} \frac{1}{\pi} \frac{\sin(x-y)}{x-y} & x \neq y \\ \frac{1}{\pi} & x = y \end{cases}$$

To begin recall that we calculated the joint distribution of the eigenvalues for the GUE to be (this measure is denoted by $\mathcal{P}_N^{(2)}$):

$$\bar{C}_N^{(2)} |\Delta(x)|^2 \prod_{i=1}^{N} e^{-x_i^2/2}$$

Where $\bar{C}_N^{(2)}$ is a normalizing constant and $\Delta(x)$ is the Vandermonde determinant. For $p \leq N$ let us denote by $\mathcal{P}_{p,N}$ the distribution of p unorderd eigenvalues of the GUE; that is to say the law so that for functions f $\mathbf{E}_{\mathcal{P}_{p,N}}(f(\lambda_1, \ldots, \lambda_p)) = \mathbf{E}_{\mathcal{P}_N^{(2)}}(f(\lambda_1, \ldots, \lambda_p))$. Because the law $\mathcal{P}_N^{(2)}$ is symetric we have that:

$$\mathbf{E}_{\mathcal{P}_{p,N}}(f(\lambda_1, \ldots, \lambda_p)) = \frac{(N-p)!}{N!} \sum_{\sigma \in S_{p,N}} \mathbf{E}_{\mathcal{P}_N^{(2)}}\left(f\left(\lambda_{\sigma(1)}, \ldots, \lambda_{\sigma(p)}\right)\right)$$

Where $S_{p,N}$ is the set of injective maps from $\{1, \ldots, p\}$ to $\{1, \ldots, N\}$. We now define the Hermite polynomials $\mathfrak{H}_n(x)$. These come up in quantum mechanics as the eigen-solutions to the quantum harmonic oscilllator, so they might be somewhat familiar.

Definition 64. The n-th Hermite polynmial $\mathfrak{H}_n(x)$ is defined by:

$$\mathfrak{H}_n(x) := (-1)^n e^{x^2/2}\frac{\mathrm{d}^n}{\mathrm{d}x^n}e^{-x^2/2}$$

One can verify the following properties of the polynomials $\mathfrak{H}_n(x)$, you might rememeber doing this in your quantum mechanics class. In this section $\mathscr{G} = e^{-x^2/2}\mathrm{d}x$ is the Gaussian meausure and $\langle\cdot,\cdot\rangle_{\mathscr{G}}$ is the L^2 inner product with respect to this measure, that is: $\langle f,g\rangle_{\mathscr{G}} = \int f(x)g(x)e^{-x^2/2}\mathrm{d}x = \sqrt{2\pi}\mathbf{E}\left(f(Z)g(Z)\right)$

(1) $\mathfrak{H}_0(x) = 1, \mathfrak{H}_1(x) = x$ and $\mathfrak{H}_{n+1}(x) = x\mathfrak{H}_n(x) - \mathfrak{H}_n'(x)$
(2) $\mathfrak{H}_n(x)$ is a *monic* polynomial of degree n
(3) $\mathfrak{H}_n(x)$ is even when n is even and odd when n is odd.
(4) $\left\langle x, \mathfrak{H}_n^2\right\rangle_{\mathscr{G}} = 0$
(5) $\langle\mathfrak{H}_k,\mathfrak{H}_l\rangle_{\mathscr{G}} = \sqrt{2\pi}k!\delta_{kl}$
(6) $\langle f,\mathfrak{H}_n\rangle_{\mathscr{G}} = 0$ for all polynomials $f(x)$ of degree $< n$
(7) $x\mathfrak{H}_n(x) = \mathfrak{H}_{n+1}(x) + n\mathfrak{H}_{n-1}(x)$ for $n\geq 1$
(8) $\mathfrak{H}'(x) = n\mathfrak{H}_{n-1}(x)$
(9) $\mathfrak{H}_n''(x) - x\mathfrak{H}_n'(x) + n\mathfrak{H}_n(x) = 0$
(10) For $x\neq y$: $\sum_{k=0}^{n-1}\frac{\mathfrak{H}_k(x)\mathfrak{H}_k(y)}{k!} = \frac{(\mathfrak{H}_n(x)\mathfrak{H}_{n-1}(y)-\mathfrak{H}_{n-1}(x)\mathfrak{H}_n(y))}{(n-1)!(x-y)}$

Definition 65. The n-th normalized osciallator wave function is the function:

$$\psi_n(x) = \frac{e^{-x^2/4}\mathfrak{H}_n(x)}{\sqrt{\sqrt{2\pi}n!}}$$

This is normalized so that $\int\psi_k(x)\psi_l(x)\mathrm{d}x = \delta_{kl}$.

Lemma 66. *For any $p\leq N$, the law $\mathcal{P}_{p,N}^{(2)}$ is absolutely continuous with respect to the Lebesgue measure and it has density:*

$$\rho_{p,N}^{(2)}\left(\theta_1,\ldots,\theta_p\right) = \frac{(N-p)!}{N!}\det_{k,l=1}^{p}K^{(N)}\left(\theta_k,\theta_l\right)$$

Where:

$$K^{(N)}(x,y) = \sum_{k=0}^{N-1}\psi_k(x)\psi_k(y)$$

Proof. By the explicit density calculated for the GUE, :

$$\rho_{p,N}^{(2)}(\theta_1,\ldots,\theta_p) = C_{p,N}\int\left|\Delta(\theta_1,\ldots,\theta_p,\zeta_{p+1},\ldots,\zeta_N)\right|^2\prod_{i=1}^{N}e^{-x_i^2/2}\prod_{i=p+1}^{N}\mathrm{d}\zeta_i$$

Now, the fundemental remark of this section is the observation that the Vandermonde determinant can be written in terms of the Hermite polynomials:

$$\Delta(x) = \prod_{1\leq i<j\leq N}(x_j - x_i) = \det_{i,j=1}^{N}x_i^{j-1} = \det_{i,j=1}^{n}\mathfrak{H}_{j-1}(x_i)$$

This works because every $\mathfrak{H}_{j-1}(x_i)$ is a monic polynomial with leading term x_i^{j-1}. The rest of the polynomial does not contribute to the determinant (can be proven by induction) because of 6. $\langle f,\mathfrak{H}_n\rangle_{\mathscr{G}} = 0$ for all polynomials $f(x)$ of degree $< n$ (property 6 above). Using this we have then:

$$\begin{aligned}
\rho_N^{(2)} = \rho_{N,N}^{(2)}(\theta_1,\ldots,\theta_p) &= C_{N,N}\left|\det_{i,j=1}^{n}\mathfrak{H}_{j-1}(\theta_i)\right|\prod_{i=1}^{N}e^{-x_i^2/2}\\
&= \tilde{C}_{N,N}\left|\det_{i,j=1}^{n}\psi_{j-1}(\theta_i)\right|^2\\
&= \tilde{C}_{N,N}\det_{i,j=1}^{n}\left(\sum_{k=0}^{N-1}\psi_k(\theta_i)\psi_k(\theta_j)\right)\\
&= \tilde{C}_{N,N}\det_{i,j=1}^{n}\left(K^{(N)}(\theta_i,\theta_j)\right)
\end{aligned}$$

In the last line we used the fact that $\det(AB) = \det(A)\det(B)$ with $A = B^{\star} = (\psi_{j-1}(\theta_i))_{i,j=1}^{N}$ so $AB^{\star}$ gives the term where $K^{(N)}$ appears. Here $\tilde{C}_{N,N} = \prod_{k=0}^{N-1}(\sqrt{2\pi}k!)\,C_{N,N}$ comes from the normalization constants in the definition of ψ_i . ☐

Lemma 67. *For any square-integrable functions $f_1, \ldots, f_n$ and $g_1, \ldots, g_n$ on the real line, we have:*

$$\begin{aligned}\frac{1}{n!}\int\cdots\int \det_{i,j=1}^{n}\left(\sum_{k=1}^{n} f_k(x_i)g_k(x_j)\right)\prod dx_i &= \frac{1}{n!}\int\cdots\int \det_{i,j=1}^{n} f_i(x_j)\cdot\det_{i,j=1}^{n} g_i(x_j)\prod_{i=1}^{n} dx_i \\ &= \det_{i,j=1}^{n}\int f_i(x)g_j(x)dx\end{aligned}$$

Proof. Use the identity $\det(AB) = \det(A)\det(B)$ applied to the matrix $A = [f_k(x_i)]_{ik}$ and $B = [g_k(x_j)]_{kj}$ so that $AB = [\sum_k f_k(x_i)g_k(x_j)]_{ij}$. This identity gives:

$$\int\cdots\int \det_{i,j=1}^{n}\left(\sum_{k=1}^{n} f_k(x_i)g_k(x_j)\right)\prod_{i=1}^{n}\mathrm{d}x_i = \int\cdots\int \det_{i,j=1}^{n} f_i(x_j)\cdot\det_{i,j=1}^{n} g_i(x_j)\prod_{i=1}^{n}\mathrm{d}x_i$$

Now using the permutation expansion for the determinant, we have:

$$\begin{aligned}\int\cdots\int \det_{i,j=1}^{n} f_i(x_j)\cdot\det_{i,j=1}^{n} g_i(x_j)\prod_{i=1}^{n}\mathrm{d}x_i &= \sum_{\sigma,\tau\in S_n}\mathrm{sgn}(\sigma)\mathrm{sgn}(\tau)\int\cdots\int\prod_{i=1}^{n} f_{\sigma(i)}(x_i)g_{\tau(i)}(x_i)\prod_{i=1}^{n}\mathrm{d}x_i \\ &= \sum_{\sigma,\tau\in S_n}\mathrm{sgn}(\sigma)\mathrm{sgn}(\tau)\prod_{i=1}^{n}\int f_{\sigma(i)}(x)g_{\tau(i)}(x)\mathrm{d}x \\ &= n!\sum_{\sigma\in S_n}\mathrm{sgn}(\sigma)\prod_{i=1}^{n}\int f_{\sigma(i)}(x)g_i(x)\mathrm{d}x \\ &= n!\det_{i,j=1}^{n}\int f_i(x)g_j(x)\mathrm{d}x\end{aligned}$$

Which is the desired result. ☐

Remark 68. If we plug in $f_i = g_i = \psi_{i-1}$ and $n = N$ into the above Lemma, we get that (from the orthogonality of the $\psi_i' s$ that:

$$\begin{aligned}\int \det_{i,j=1}^{N} K^{(N)}(\theta_i,\theta_j)\prod_{i=1}^{N}\mathrm{d}\theta_i &= \int \det_{i,j=1}^{N}\sum_{k=0}^{N-1}\psi_k(\theta_i)\psi_k(\theta_j)\prod\mathrm{d}\theta_i \\ &= N!\det_{i,j=1}^{N}\int\psi_i(\theta)\psi_j(\theta)\mathrm{d}\theta \\ &= N!\det(I_N) = N!\end{aligned}$$

Corollary 69. *The normalizing constant $C_{N,N}$ which appears in the joint distribution for the eigenvalues is:*

$$\begin{aligned}C_{N,N} &= \frac{1}{N!\prod_{k=0}^{N-1}(\sqrt{2\pi}k!)} \\ &= \frac{1}{N!\,(2\pi)^{N/2}\prod_{1}^{N}k!}\end{aligned}$$

Proof. Integrate the density $\rho_{N,N}$to see that $1 = \tilde{C}_{N,N}\int\det_{i,j=1}^{N}K^{(N)}(\theta_i,\theta_j)\prod_{i=1}^{N}\mathrm{d}\theta_i = \tilde{C}_{N,N}\cdot N!$. The result follows from the relationship $\tilde{C}_{N,N} = \prod_{k=0}^{N-1}(\sqrt{2\pi}k!)\,C_{N,N}$. ☐

Corollary 70. *The normalizing constant $\tilde{C}_{p,N} = \frac{(N-p)!}{N!}$*

Proof. Following the above strategy, we will have (For convenience, let $x_i = \theta_i$ if $i \le p$ and $x_i = \zeta_i$ for $i > p$.)

$$\begin{aligned}
\rho_{p,N}^{(2)}(\theta_1,\ldots,\theta_p) &= \tilde{C}_{p,N} \int \left(\det_{i,j=1}^{N} \psi_{j-1}(x_i) \right)^2 \prod_{i=p+1}^{N} \mathrm{d}\zeta_i \\
&= \tilde{C}_{p,N} \sum_{\sigma,\tau \in S_n} \mathrm{sgn}(\sigma)\mathrm{sgn}(\tau) \int \prod_{j=1}^{N} \psi_{\sigma(j)-1}(x_j)\psi_{\tau(j)-1}(x_j) \prod_{i=p+1}^{N} \mathrm{d}\zeta_i \\
&= \tilde{C}_{p,N} \sum_{\sigma,\tau \in S_n} \mathrm{sgn}(\sigma)\mathrm{sgn}(\tau) \prod_{j=1}^{p} \psi_{\sigma(j)-1}(\theta_j)\psi_{\tau(j)-1}(\theta_j) \int \prod_{j=p+1}^{N} \psi_{\sigma(j)-1}(\zeta_j)\psi_{\tau(j)-1}(\zeta_j) \prod_{i=p+1}^{N} \mathrm{d}\zeta_i \\
&= \tilde{C}_{p,N} \sum_{\sigma,\tau \in S_n} \mathrm{sgn}(\sigma)\mathrm{sgn}(\tau) \prod_{j=1}^{p} \psi_{\sigma(j)-1}(\theta_j)\psi_{\tau(j)-1}(\theta_j) \prod_{j=p+1}^{N} \delta_{\sigma(j)\tau(j)}
\end{aligned}$$

This is nonzero only when σ and τ differ only on $\{1, 2 \ldots, p\}$ and in this case the factor at the end is 1. We now divide the sum up based on which elements $\{1, \ldots, p\}$ map to. For $1 \le v_1 < \ldots < v_p \le N$ let $\mathcal{L}(p, v)$ be the bijections from $\{1, \ldots, p\}$ to $\{v_1, \ldots, v_p\}$. We have:

$$\begin{aligned}
\rho_{p,N}^{(2)}(\theta_1,\ldots,\theta_p) &= \tilde{C}_{p,N} \sum_{v_1<\ldots<v_n} \sum_{\sigma,\tau\in\mathcal{L}(p,v)} \mathrm{sgn}(\sigma)\mathrm{sgn}(\tau) \prod_{j=1}^{p} \psi_{\sigma(j)-1}(\theta_j)\psi_{\tau(j)-1}(\theta_j) \\
&= \tilde{C}_{p,N} \sum_{v_1<\ldots<v_n} \left(\det_{i,j=1}^{p} \psi_{v_j-1}(\theta_j) \right)^2
\end{aligned}$$

From here we can integrate both sides and use the lemma to see that $\tilde{C}_{p,N} = \frac{(N-p)!}{N!}$. □

This is now ripe for us to apply the Cauchy-Binet theorem. In our case we use the following "version" of the theorem: Let A be a $p \times N$ matrix and let $C = AA^\star$(this is a $p \times p$ matrix), then $\det C = \sum_{K \in \mathcal{K}_{p,N}} \det A_K \det A_K^\star$where $\mathcal{K}_{p,N}$ is the set of all p element subsets of $\{1, \ldots, N\}$ and A_K is the $p \times p$ matrix which is obtained from A by keeping only the columns in K. This is exactly the set up we have here with $A_{i,j} = \psi_{j-1}(\theta_i)$. Applying this and noticing that $[K(\theta_i, \theta_j)]_{i,j} = [\psi_i(\theta_j)]_{i,j} [\psi_i(\theta_j)]_{i,j}^\star$ from its definition finally gives:

$$\rho_{p,N}^{(2)}(\theta_1,\ldots,\theta_p) = \tilde{C}_{p,N} \det_{i,j=1}^{p} \left(K^{(N)}(\theta_i,\theta_j) \right)$$

Finally, we get to the main result about the GUE as a Fredholm determinant.

Theorem 71. *For any measurable subset A of $\mathbb{R}$:*

$$P_N^{(2)}\left(\bigcap_{i=1}^{N} \{\lambda_i \in A\} \right) = 1 + \sum_{k=1}^{\infty} \frac{(-1)^k}{k!} \int_{A^c} \cdots \int_{A^c} \det_{i,j=1}^{k} K^{(N)}(x_i,x_j) \prod_{i=1}^{k} dx_i$$

Proof. From our previous lemmas:

$$\begin{aligned}
P_N^{(2)}\left(\bigcap_{i=1}^{N} \{\lambda_i \in A\} \right) &= \int_A \cdots \int_A \rho_N^2(\theta_1,\ldots,\theta_p) \prod_{k=1}^{N} \mathrm{d}\theta_k \\
&= \tilde{C}_{N,N} \int_A \cdots \int_A \det_{i,j=1}^{N} K^{(N)}(\theta_i,\theta_j) \prod_{k=1}^{N} \mathrm{d}\theta_k \\
&\quad \tilde{C}_{N,N} \int_A \cdots \int_A \det_{i,j=1}^{N} \left[\sum_{k=0}^{N-1} \psi_k(\theta_i)\psi_k(\theta_j) \right] \prod_{k=1}^{N} \mathrm{d}\theta_k \\
&= \det_{i,j=1}^{N} \left[\int_A \psi_i(x)\psi_j(y)\mathrm{d}x \right] \\
&= \det_{i,j=1}^{N} \left[\delta_{ij} - \int_{A^c} \psi_i(x)\psi_j(y)\mathrm{d}x \right]
\end{aligned}$$

The last line follows by the orthogonality of the ψ's. Now doingsome kind of determinent identity manipulation which I dont quiet see, you get (EDIT: I figured it out! Replace the -1 that appears with an x. Then the determinant is a polynomial in x. By taking derivatives w.r.t to x you get the formula ala Taylor expansion. Finally, put -1 back in for x. See the notes from Lax's functional analysis for this in detail!) :

$$P_N^{(2)}\left(\bigcap_{i=1}^{N}\{\lambda_i \in A\}\right) = 1 + \sum_{k=1}^{N}(-1)^k \sum_{0\le v_1<\ldots<v_k\le N-1} \det_{i,j=1}^{k}\left(\int_{A^c}\psi_i(x)\psi_j(y)\mathrm{d}x\right)$$

Now we just use our lemmas in reverse to rewrite this in terms of $K^{(N)}(\cdot,\cdot)$. Have:

$$\begin{aligned} P_N^{(2)}\left(\bigcap_{i=1}^{N}\{\lambda_i \in A\}\right) &= 1 + \sum_{k=1}^{N}\frac{(-1)^k}{k!}\int_{A^c}\cdots\int_{A^c}\sum_{0\le v_1<\ldots<v_k\le N-1}\left(\det_{i,j=1}^{k}\psi_{v_i}(x)\right)^2\prod \mathrm{d}x_i \\ &= 1 + \sum_{k=1}^{N}\frac{(-1)^k}{k!}\int_{A^c}\cdots\int_{A^c}\sum_{0\le v_1<\ldots<v_k\le N-1}\left(\det_{i,j=1}^{k}K^{(N)}(x_i,x_j)\right)\prod \mathrm{d}x_i \end{aligned}$$

Lastly, we can sum to ∞ instead of N without changing the result, as the rank of the matrix $\left[K^{(N)}(x_i,x_j)\right]_{i,j=1}^{k}$ is at most N because it arises as a product $AA^{\star}$where A is a $k\times N$ matrix of ψ's. $\square$