RANDOM MATRIX THEORY

TEODORO FIELDS COLLIN

ABSTRACT. This paper proves several important results in Random Matrix Theory, the study of matrices with random entries. All of these results focus on the distribution of the eigenvalues of these matrices. After the development of enough measure theoretic probability to prove the results, we provide a proof of the limiting distribution of the eigenvalues of a large class of random matrices, called Wigner matrices. We then find the distribution of the eigenvalues for matrices of a fixed size in the Gaussian Unitary Ensemble (a subset of Wigner matrices). Finally, we investigate the limiting distribution of the eigenvalues of these matrices and allude to the Tracy-Widom distribution.

CONTENTS

1. Introduction

The objective of this paper is to prove or sketch a few important results in Random Matrix Theory. In the interest of accessibility and completeness, a short introduction to measure theoretic probability is included. If the reader is already familiar with this topic, then this section can skipped.

In section 3, we examine Wigner matrices and show that as their size increases to infinity, the distribution of the average eigenvalue converges to a semicircle distribution. This result is called Wigner's theorem. We present a combinatorial proof of this fact and we eschew many combinatorial details that are easy to verify. Eugene Wigner, a physicist, used his result to describe the energy levels of atomic nuclei, which happen to be the eigenvalues of certain Hermitian operators. By finding the limiting distribution of the average eigenvalue of a certain class of matrices, Wigner approximated part of the required distribution. See [5] or [4] for more information.

In section 4, we find the density of a vector of the eigenvalues of a very specific class of matrices, called the Gaussian Unitary Ensemble (GUE). This section can be divided into two parts. In the first, we employ several implementations of classic techniques for finding density functions to find the density modulo a constant. In the second, we find the full form of the density by evaluating the constant. To evaluate the constant, we prove Selberg's Integral formula by proving many technical lemmas regarding integrals over monic polynomials. This result has many applications, but the broadest is perhaps that to Principle Component Analysis (PCA), a common statistical technique in the social sciences. In PCA, one takes sample data with many variables and summarizes it with a sample covariance matrix. If one visualizes the data as a point cloud in Euclidean space, the eigenvalues of the sample covariance matrix estimate the amount of variance along each axis. This information is then used to eliminate variables that don't produce much variance. Thus, knowing the distribution of the eigenvalues of the sample covariance matrix helps determine if certain variables should be removed in the same way that knowing the distributions of sample averages helps determine if there is a difference between population averages. See [3] for more information.

In section 5, we investigate what happens to the distribution of the eigenvalues of a matrix in the GUE as the size of the matrix goes to infinity. This section is a brief sketch; the section takes the theory of Fredholm determinants as a black box. We conclude section 5 by teasing the limiting distribution of the largest eigenvalue, the Tracy-Widom distribution. This final tease has many applications (including PCA since the largest eigenvalue is particularly significant in a high dimensional data), but it is also notable because it is much easier to work with than the other limiting result in section 5. See [4] or [3] for more information.

2. Measure Theoretic Probability Concepts

We assume a fair amount of measure theory and we state without proof several important theorems in a development of probability from the perspective of measure theory. We eschew many details. For a thorough treatment of both measure theory and measure theoretic probability, we direct the reader to [1] or [7].

2.1. **Setting.**

Definition 2.1. Given a measure space (Ω, B_Ω) (a pair consisting of a set and a σ-algebra), a **probability measure** $\mathbb{P}$ on the space is a measure such that

$$\mathbb{P}(\Omega) = 1.$$

Definition 2.2. Given a measure space (Ω, B_Ω) and a probability measure $\mathbb{P}$, we call $(\Omega, B_\Omega, \mathbb{P})$ a **probability space**.

Remarks 2.3. Throughout this section:

- $(\Omega, B_\Omega, \mathbb{P})$ is a probability space and $(H, \mathcal{H})$ is a measure space. The measure ν is always defined on $(H, \mathcal{H})$.
- We may occasionally refer to a σ-algebra B_Ω or $\mathcal{H}$ as the set of $\mathbb{P}$ or ν measurable functions.
- We denote Lebesgue measure as μ.

2.2. Random Variables and Related Topics.

Definition 2.4. A function $X : \Omega \to H$ is a **random variable** if it is measurable. For all of the random variables that we care about, $H = \mathbb{R}^k$.

Definitions 2.5. Given a random variable X as before, the **probability distribution** of X is a measure $F_X : \Omega \to [0, 1]$ such that

$$F_X(A) = \mathbb{P}(X^{-1}(A))$$

holds for every $A \in \mathcal{H}$.

We interpret this as the probability that X takes a value in the set A.

Definition 2.6. If τ is also a measure on the measure space $(H, \mathcal{H})$, then τ is **absolutely continous** with respect to ν if for every $A \in \mathcal{H}$, $\nu(A) = 0$ implies that $\tau(A) = 0$.

Definition 2.7. The measure ν is **σ-finite** if H is a countable union of sets with finite ν measure.

Theorem 2.8 (Radon−Nikodym). *If a σ-finite measure τ is absolutely continous with respect to ν and ν is σ-finite, then there exists a non-negative measurable function $\frac{d\tau}{d\nu} \in \mathcal{H}$ such that*

$$\tau(A) = \int_A \frac{d\tau}{d\nu}(x)d\nu(x) \tag{2.9}$$

holds for every set $A \in \mathcal{H}$.

Remark 2.10. This theorem indicates that in a weak sense $\frac{d\tau}{d\nu}$ is the derivative of τ. This derivative is often called the Radon−Nikodym derivative. Additionally, it furnishes us with the concept of a probability density function. If f_x is the Radon−Nikodym derivative of F_x with respect to μ, then it is the **probability density function** of the random variable X. It is clear that

$$\int f_x d\mu = F_x(\Omega) = 1. \tag{2.11}$$

Definition 2.12. If $(F, \mathcal{F})$ is measure space and $f : H \to F$ is a measurable function, then $f(X)$ is a random variable, referred to as **a function of a random variable**.

Given some random variable and its probability distribution, it is not always easy to find the density. However, if the random variable can be expressed as a function of a random variable that we know the density of, it is sometimes easier to find the density. This is illustrated by the following result.

Theorem 2.13. *Given f and X as before, if f_x is a continous density function with respect to Lebesgue measure and f is a continuously differentiable bijection such that the Jacobian of the density $J_f(x) \neq 0$ for all $x \in H$, the density function of $f(X)$ is*

$$f_x(f^{-1}(x))|J_f(x)|.$$

We discuss why Theorem 2.13 is true in Remark 4.26.

2.3. **Expectations and Moments.**

Definition 2.14. The **expectation of a random variable** X is

$$\mathbb{E}[X] := \int_\Omega X dP. \tag{2.15}$$

Remark 2.16. We observe that the expectation of a random variable inherits several useful properties from the integral. The most important of these is linearity.

Definition 2.17. The **k-th moment** of a random variable is $\mathbb{E}[X^k]$, assuming that the function x^k is measurable in the relevant measure space and that the integral exists.

We now state two useful inequalities. The first is called Markov's inequality.

Theorem 2.18. *If Z is a random variable with distribution $\mathbb{P}$ and $a > 0$ and $\mathbb{E}[|Z|] < \infty$, then*

$$\mathbb{P}(|Z| \geq a) \leq \frac{\mathbb{E}[|Z|]}{a}.$$

The second is called Chebyshev's inequality.

Theorem 2.19. *If Z is a random variable with distribution $\mathbb{P}$ and $a > 0$ and $\mathbb{E}[|Z|] < \infty$, then*

$$\mathbb{P}(|Z - \mathbb{E}[Z]| \geq a) \leq \frac{\mathbb{E}[(Z - \mathbb{E}[Z])^2]}{a^2}.$$

2.4. **Joint Distributions and Independence.**

Definitions 2.20. Given random variables $X_1, \dots, X_n$ taking values in $(H_1, \mathcal{H}_1), \dots, (H_n, \mathcal{H}_n)$ with distributions $F_{X_1}, \dots, F_{X_n}$, we say that the distribution of the vector $(X_1, \dots, X_n)$ is the **joint distribution** of the random variables $X_1, \dots, X_n$. We write the joint distribution of $X_1, \dots, X_n$ as $F_{X_1,\dots,X_n}$.

If $F_{X_1,\dots,X_n}$ has an associated density, then we call $f_{X_1,\dots,X_n}$ the **joint density** and for each random variable X_i for $i = 1, \dots, n$ we can refer to f_{X_i} as a **marginal density**.

Given a joint density, the following theorem illustrates how to find a marginal density.

Theorem 2.21. *Given $f_{X_1,\dots,X_n}$, we can obtain f_{X_i} with the following formula:*

$$f_{X_i} = \int_{H_1 \times \cdots \times H_{i-1} \times H_{i+1} \times \cdots \times H_n} f_{X_1,\dots,X_n} d\mu(x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n). \tag{2.22}$$

Definition 2.23. We say the random variables $X_1, \dots, X_n$ taking values in $(H_1, \mathcal{H}_1), \dots, (H_n, \mathcal{H}_n)$ with distributions F_{X_i} are **independent** if

$$F_{X_1,\dots,X_n}(A) = F_1(A_1)F_2(A_2)\cdots F_n(A_n) \tag{2.24}$$

holds for every measurable subset A of $H_1 \times \cdots \times H_n$ where $A = A_1 \times \cdots \times A_n$.

We state one important consequence of the independence of these variables.

Theorem 2.25. *If $X_1, \dots, X_n$ are independent and all take values in the same space, then*

$$\mathbb{E}[X_1 \cdots X_n] = \mathbb{E}[X_1] \cdots \mathbb{E}[X_n].$$

2.5. **Measure Valued Random Variables.** Eschewing a large amount of topological detail, we define measure valued random variables.

Definitions 2.26. Recall our usual probability space $(\Omega, B_\Omega, \mathbb{P})$ and our measure space $(H, \mathcal{H})$. Define $M(H, \mathcal{H})$ as the **space of all probability measures** on $(H, \mathcal{H})$. A measurable map $\delta \colon \Omega \to M(H, \mathcal{H})$ is a **measure valued random variable**.

2.6. **Notions of Convergence.** We require three important notions of convergence.

Definition 2.27. Given a sequence of random variables $\{X_n\}$, we say that a sequence of random variables **converges in probability** to a deterministic value x if

$$\lim_{n\to\infty} \mathbb{P}(|X_n - x| > \epsilon) = 0$$

for any $\epsilon > 0$.

Definition 2.28. Given a sequence of measures $\{\nu_n\}$, we say that it **converges weakly** to some measure ν if for every continous bounded function f on the relevant measure space

$$\int f d\nu_n \to \int f d\nu.$$

We can combine these two notions to talk about convergence of sequences of random measures.

Definition 2.29. A sequence of random probability measures $\{\nu_n\}$ **converges weakly in probability** to a deterministic measure ν if for every continous bounded function f

$$\lim_{n\to\infty} P\left(\left|\int f d\nu_n - \int f d\nu\right| > \epsilon\right) = 0$$

holds for every $\epsilon > 0$.

3. Wigner's Theorem

In this section, we find the limiting case of the distribution of the average eigenvalue of a very general class of matrices. This distribution has a nice form and is easy to compute. The proof relies heavily on combinatorics and in the interest of space, I eschew some combinatorial details that are easy to verify.

Definition 3.1. Let $\{Z_{i,j}\}$ be a sequence of independent random variables with the same distribution with all first moments 0 and all second moments 1. Let $\{Y_i\}$ be a sequence of independent random variables with the same distribution with all first moments 0. Assume all Z and Y are independent of each other. Lastly, assume all moments of both sets of random variables are finite. A **Wigner matrix** is a symmetric matrix X_N of size N such that

$$X_N(i,j) = X_N(j,i) = \begin{cases} Z_{i,j}/\sqrt{N} & \text{if } i < j \\ Y_i/\sqrt{N} & \text{if } i = j. \end{cases}$$

Definition 3.2. Given a Wigner matrix X_N with eigenvalues $\lambda_{1,N} \leq \cdots \leq \lambda_{N,N}$, we define the **empirical distribution** L_N as

$$L_N(A) := \frac{1}{N}\sum_{i=1}^{N} 1_{\lambda_{i,N}}(A)$$

where A is a measurable subset of $\mathbb{R}$ and

$$1_{\lambda_{i,N}}(A) = \begin{cases} 1 & \text{if } \lambda_{i,N} \in A \\ 0 & \text{otherwise.} \end{cases}$$

The goal of this section is the prove the following result, called Wigner's Theorem in honor of Eugene Wigner who first proved it in 1955.

Theorem 3.3. *The empirical distribution converges weakly in probability to the probability measure with density σ given by*

$$\sigma(x) = \frac{1}{2\pi}\sqrt{4-x^2} \cdot 1_{|x|\leq 2} \tag{3.4}$$

where $1_{|x|\leq 2}$ is the usual indicator function.

3.1. Preliminary Reductions.

We note an important set of numbers.

Definition 3.5. The **Catalan numbers** are given by

$$C_k = \frac{1}{k+1}\binom{2k}{k}. \tag{3.6}$$

We omit the following calculation in the interest of space.

Theorem 3.7. *If m_k is the k-th moment of the semicircle distribution, then $m_{2k} = C_k$ while $m_{2k+1} = 0$.*

To prove Theorem 3.3, it will be useful to have two lemmas. The first relates the moments of the semicircle distribution to the expected values of the moments of the empirical measure.

Lemma 3.8. *If $m_k^N := E\int x^k dL_N$ then for every natural number k, we have*

$$\lim_{N\to\infty} m_k^N = m_k. \tag{3.9}$$

The second lemma relates the moments of L_N to the expected moments m_k^N.

Lemma 3.10. *For every natural number k and $\epsilon > 0$, we have*

$$\lim_{N\to\infty} \mathbb{P}\left(\left|\int x^k dL_N - m_k^N\right| > \epsilon\right) = 0. \tag{3.11}$$

In both of these lemmas, it is important to note that $\int x^k dL_N$ is a real valued random variable. We sketch the proof of the first lemma and comment on the proof of the second.

3.2. **Proof of the Convergence of the Moments of the Empirical Distribution.** We dedicate this subsection to the proof of the first lemma. We need a whole new section because we need to introduce a large amount of combinatorial machinery.

3.2.1. *Combinatorial Preliminaries.*

Definition 3.12. A **Dyck path** of length 2k is a sequence of integers S_i such that $S_1 = S_{2k} = 0$ and $|S_{t+1} - S_t| = 1$ and $S_t > 0$ for all $t = 2, \ldots, 2k-1$. This is path from 0 to 0 that only takes a step of one integer at a time and stays in the positive plane.

We magically count the number of paths.

Lemma 3.13. *The number of Dyck path of length* $2k$ *is* C_k.

Definitions 3.14. An **N-word** is a finite number sequence of letters $s_1 \cdots s_n$ where each $s_i \in \{1, \ldots, N\}$. Let w_1, w_2 be N-words.

We say w_1 is closed if $w_1 = s_1 \ldots s_n$ and $s_1 = s_n$ We write $w_1 \sim w_2$ if there is a bijection on $\{1, \ldots, N\}$ that maps one word to the other. The length of w_1 is the number of letters. The weight of w_1 is the number of unique letters. The support of w_1 is the set of letters appearing in w_1.

We also associate a graph with each word $w_1 = s_1 \ldots s_n$. The set of vertices of the graph is the support of the graph. The set of edges is the set $E_{w_1} = \{\{s_i, s_{i+1}\} \colon i = 1, \ldots, n-1\}$. The graph of the word is clearly connected as the word defines a path that traverses every edge. The subset of E_{w_1} where $s_i = s_{i+1}$ is called the set of self-edges, denoted $E^s_{w_1}$. The set of self-edges could also be called the set of edges that are loops. The set of connecting edges, denoted $E^c_{w_1}$, is the set of all other edges. We let N^w_e be the number of times an edge appears in the graph of the word w. Lastly, we note that if $w_1 \sim w_2$ then the graphs of both words must be isomorphic, thus sharing all the above properties.

Definition 3.15. Denote $\mathcal{W}_{k,t}$ as the set of representatives for equivalence classes of closed t-words of length $k+1$ and weight t with $N^w_e \geq 2$ for all $e \in N^s_w$.

We note one such set of particular interest.

Definition 3.16. If $w \in \mathcal{W}_{k,k/2+1}$ then w is a **Wigner word**.

Two lemmas are needed to elucidate the relevant structure of the Wigner words.

Lemma 3.17. *If* $w \in \mathcal{W}_{k,k/2+1}$ *then the representative graph is a tree and* $N^w_e = 2$ *for every edge.*

Proof. Call (V_w, E_w) the graph associated with the word. Since the weight is $k/2+1$, $|V_w| = k/2 + 1$. Since the graph is connected and there are $k/2 + 1$ edges, $|E_w| \geq k/2$. Since each edge must appear at least twice and there are at most k edges, $|E_w| \leq k/2$. Thus $k/2 = |E_w| = |V_w| - 1$. This proves both claims. □

Lemma 3.18.

$$|\mathcal{W}_{k,k/2+1}| = C_{k/2}. \tag{3.19}$$

Proof. In the interest of space, we do not give a full proof. Essentially we construct a bijection between representatives of equivalence classes of Wigner words and Dyck paths. Then we appeal to Lemma 3.13. The construction of the bijection can easily be seen via the selection of representatives so we briefly describe that. We select a representative $w = s_1 \dots s_{k+1}$ so that we set $s_1 = s_{k+1} = 1$ and so that we set $s_i = s_{i-2}$ or $s_i = \max(s_1, \dots, s_{i-1}) + 1$ for each i where $1 < i < k+1$. This structure ensures each new letter can be seen as a single increment or decrement in a Dyck path and suggests an inverse process by which we may build a representative from a path.

If this biection is given, all that is needed is a proof of the uniqueness and existence of the representative. □

How all of this combinatorial machinery comes to bear on the proof of Lemma 3.8 is made clear by the following lemma.

Lemma 3.20. *For every natural number k and a Wigner matrix X_N, we have that*

$$E\int x^k dL_N = \frac{1}{N}\sum_{i_1,\dots,i_k=1}^{N} EX_N(i_1,i_2)X_N(i_2,i_3)\dots X_N(i_{k-1},i_k)X_N(i_k,i_i).$$

Proof. The value of $E\int x^k dL_N$ is essentially the average of the k-th powers of the eigenvalues of X_N. This is also $E\frac{1}{N}trX_N^k$. The result follows by the definitions of trace and matrix multiplication. □

3.2.2. *Proof of Lemma 3.8.* We are now prepared to sketch the proof of Lemma 3.8.

Proof. First, set $T_{\mathbf{i}}^N = EX_N(i_1,i_2)X_N(i_2,i_3)\dots X_N(i_{k-1},i_k)X_N(i_k,i_i)$. We claim that by mapping the pattern of accesses of random variables in the matrix X_N to a word $w_{\mathbf{i}} = i_1 \dots i_k i_1$ and by comparing Definition 3.2 and Definitions 3.14, we can show that

$$T_{\mathbf{i}}^N = \frac{1}{N^{k/2}}\prod_{e\in E^c_{w_{\mathbf{i}}}}\mathbb{E}[Z_{1,2}^{N_e^{w_{\mathbf{i}}}}]\prod_{e\in E^s_{w_{\mathbf{i}}}}\mathbb{E}[Y_1^{N_e^{w_{\mathbf{i}}}}]. \tag{3.21}$$

Additionally, we can see that if two patterns of accesses generate equivalent words then their $T_{\mathbf{i}}^N$ are equivalent. It thus makes sense to sum over representatives of equivalence classes (Definition 3.15). To do this, note that an N word of weight k has $\binom{N}{k}k! = (N-k)!$ words in its equivalence class. We also note that maximum possible weight is $\lfloor k/2+1 \rfloor$. We can now see that

$$E\int x^k dL_N = \sum_{t=1}^{\lfloor k/2+1\rfloor}\frac{(N-t)!}{N^{k/2+1}}\sum_{w\in\mathcal{W}_{k,t}}\prod_{e\in E^c_{w_{\mathbf{i}}}}\mathbb{E}[Z_{1,2}^{N_e^{w_{\mathbf{i}}}}]\prod_{e\in E^s_{w_{\mathbf{i}}}}\mathbb{E}[Y_1^{N_e^{w_{\mathbf{i}}}}]. \tag{3.22}$$

Since the moments of Y_1 and $Z_{2,3}$ are all finite and since the number of N words is clearly finite, we can treat the inner sum as some constant. Turning to the analysis of $\frac{(N-t)!}{N^{k/2+1}}$, we find that unless $t \geq k/2+1$ we have that $\frac{(N-t)!}{N^{k/2+1}} \to 0$ as

$N \to \infty$ for every summand. Thus if k is odd, then the entire sum goes to 0 as no sum will have $t \geq k/2+1$. This corresponds to the odd moments of the semicircle distribution. If k is even, then we get $t = k/2+1$ in the last summand, proving that

$$\lim_{N\to\infty} E\int x^k dL_N = \sum_{w\in\mathcal{W}_{k,k/2+1}} \prod_{e\in E^c_{w_\mathbf{i}}} \mathbb{E}[Z_{1,2}^{N_e^{w_\mathbf{i}}}] \prod_{e\in E^s_{w_\mathbf{i}}} \mathbb{E}[Y_1^{N_e^{w_\mathbf{i}}}] = \sum_{w\in\mathcal{W}_{k,k/2+1}} 1 = |W_{k,k/2+1}|.$$

The second equality follows easily from Lemma 3.17. Now, by Lemma 3.18 the even moments converge to the Catalan numbers, which are the even moments of the semicircle distribution. □

Remark 3.23. The proof of Lemma 3.10 follows in a similar manner to that of Lemma 3.8. We use Chebyshev's inequality (Theorem 2.19) to reduce it to an analysis of an expectation and apply the same tactics but using sentences (unions of words) instead of just words.

3.3. **Proof of Wigner's Theorem.**

Proof. Fix a test function f and an $\epsilon > 0$. By the Weierstrass Approximation Theorem, there is a polynomial $Q(x)$ such that

$$\sup_{|x|\leq 5} |Q(x) - f(x)| \leq \epsilon. \tag{3.24}$$

We have that

$$\begin{aligned}\int f dL_N - \int f\sigma &= \int Q(t)dL_N - E\int Q(t)dL_N \\ &\quad + E\int Q(t)dL_N - \int Q(t)\sigma \\ &\quad - \int Q(t)1_{|x|>5}dL_N\end{aligned}$$

where the last term exists to eliminate the error caused by σ having support $[-2,2]$ while L_N has an unknown support and Q only approximates f on $[-5,5]$. Via use of the triangle inequality and properties of probability measures, we have that

$$\begin{aligned}\mathbb{P}\left(\left|\int f dL_N - \int f\sigma\right| > \epsilon\right) &\leq \mathbb{P}\left(\left|\int Q(t)dL_N - E\int Q(t)dL_N\right| > \epsilon\right) \\ &\quad + \mathbb{P}\left(\left|E\int Q(t)dL_N - \int Q(t)\sigma\right| > \epsilon\right) \\ &\quad + \mathbb{P}\left(\left|\int Q(t)1_{|x|>5}dL_N\right| > \epsilon\right).\end{aligned}$$

That the first two terms go to 0 as $N \to \infty$ follows from Lemma 3.10 and Lemma 3.8 respectively. This follows because expectations of polynomial functions of random variables are essentially linear combinations of moments. We turn to the last term. Via Markov's inequality (Theorem 2.18), we have that

$$\mathbb{P}\left(\left|\int |x|^k 1_{|x|>5}dL_N\right| > \epsilon\right) \leq \frac{1}{\epsilon}\int |x^k|1_{|x|>5}dL_N \leq \frac{\int x^{2k}dL_N}{\epsilon 5^k}.$$

Via Lemma 3.8, we have that

$$\limsup_{N\to\infty} \mathbb{P}\left(\left|\int |x|^k 1_{|x|>5} dL_N\right| > \epsilon\right) \leq \frac{\int |x|^{2k}\sigma}{\epsilon 5^k} = \frac{C_k}{\epsilon 5^k}.$$

Using a bound on the binomial coefficient we can see that $C_k \leq 4^k$. Thus, we get that

$$\limsup_{N\to\infty} \mathbb{P}\left(\left|\int |x|^k 1_{|x|>5} dL_N\right| > \epsilon\right) \leq \frac{1}{\epsilon}(\frac{4}{5})^k. \tag{3.25}$$

Now we note that in the standard proof of Weierstrass Approximation Theorem, the degree of the polynomial $Q(t)$ goes to infinity as $N \to \infty$. We also note that since we only consider $|x| > 5$, the last term of the polynomial times some constant dominates the value of $Q(t)$. Using these two facts, we are justified in applying (3.25) to see that $\mathbb{P}\left(\left|\int Q(t) 1_{|x|>5} dL_N\right| > \epsilon\right) \to 0$ as $N \to \infty$. □

Remark 3.26. In this section we presented a combinatorial method for proving Wigner's theorem. This was the original method employed by Wigner, but it is not the only method. Wigner's theorem can instead be proven using techniques from complex analysis. This alternative is well detailed in [10].

4. Joint Density of the Eigenvalues of the GUE

In this section, we find the joint density of the eigenvalues of a much smaller set of random matrices, called the Gaussian Unitary Ensemble (GUE). In contrast to the previous section, we find the density for a matrix of any size rather than just in the limiting case. We are able to find the joint density of the eigenvalues for each N because we constrain the GUE enough to find the density of a random matrix in the GUE and use this to find the density of eigenvalues.

4.1. **Definition of the GUE and Density of the Matrices.**

Definition 4.1. Let $\{X_{i,j}\}$ and $\{Y_{i,j}\}$ be sequences of independent standard normal random variables. Each $X_{i,j}$ is independent with each $Y_{i,j}$. A random matrix $\{Z_{i,j}\}_{i,j=1}^N$ is a member of the **Gaussian Unitary Ensemble** if $Z_{i,i} = X_{i,i}$ and $Z_{i,j} = \dfrac{X_{i,j} + iY_{i,j}}{\sqrt{2}}$ and $Z_{j,i} = \dfrac{X_{i,j} - iY_{i,j}}{\sqrt{2}}$. If this is the case, we write $\{Z_{i,j}\}_{i,j=1}^N \in \mathcal{H}_N^2$.

The principle objective of this section is to prove the following.

Theorem 4.2. *The joint density of the eigenvalues $\lambda_1 \leq \cdots \leq \lambda_N$ with respect Lebesgue measure on $\mathbb{R}^N$ of a random matrix $X \in \mathcal{H}_N^2$ is*

$$N! C_N^2 1_{\lambda_1 \leq \cdots \leq \lambda_N} \Delta(\lambda_1, \cdots, \lambda_N)^2 \prod_{i=1}^N e^{\frac{-\lambda_i^2}{2}} \tag{4.3}$$

where $N!C_n$ is a normalization constant with the value

$$N! C_n = (2\pi)^{-N/2} \prod_{j=1}^N \frac{1}{\Gamma(j)} \tag{4.4}$$

and where $1_{\lambda_1 \leq \cdots \leq \lambda_N}(H)$ is 1 if and only if the eigenvalues of H are increasingly ordered and finally where $\Delta(\lambda_1, \cdots, \lambda_N)$ is the Vandermonde determinant of the vector of eigenvalues of H.

Remark 4.5. We will often treat a random matrix as simply a vector of the random variables that are needed to determine the matrix. In the case of $X \in \mathcal{H}_N^2$, we need N standard normal random variables for the diagonal and $N(N-1)$ for the upper off diagonal entries. Since the matrix is clearly Hermitian, this is sufficient to represent the matrix. Thus, when we say the density of a matrix of the GUE, we mean the joint density of the N variables on the diagonal and the $N(N-1)$ variables on the upper off diagonal.

We now find the distribution of the GUE itself.

Theorem 4.6. *If $\frac{dP_N^2}{d\mu}(H)$ is the density of $H \in \mathcal{H}_N^2$ with respect to Lebesgue measure on $\mathbb{R}^{N^2}$, we have that*

$$\frac{dP_N^2}{d\mu}(H) = (2)^{-N/2}(\pi)^{-N^2/2}e^{-\mathrm{Tr}H^2/2}. \tag{4.7}$$

Proof. Due to the $\frac{1}{\sqrt{2}}$ factor in the off diagonals, the standard properties of normal variables imply that the off diagonal variables have mean 0 and variance $\frac{1}{2}$. These have density $\frac{1}{\sqrt{\pi}}e^{-x^2}$. The diagonal entries have density $\frac{1}{\sqrt{2\pi}}e^{-x^2/2}$. Since all variables involved are independent, the density of X is

$$\begin{aligned}\frac{dP_N^2}{d\mu}(H) &= \prod_{i=1}^{N} \frac{1}{\sqrt{2\pi}}e^{-H_{i,i}^2/2} \prod_{1 \leq i < j \leq N} \frac{1}{\sqrt{\pi}}e^{-H_{i,j}^2} \\ &= (2)^{-N/2}(\pi)^{-N^2/2}e^{\sum_{i=1}^{N} -H_{i,i}^2/2 + \sum_{1 \leq i < j \leq N} -H_{i,j}^2}.\end{aligned}$$

Since $\mathrm{Tr}H^2 = \sum_{i=1}^{N} |H_{i,i}|^2 + 2\sum_{1 \leq i < j \leq N} |H_{i,j}|^2$, we have the desired result. □

Remark 4.8. Since the trace of the square of a matrix is the sum of the squares of the eigenvalues, the density of the matrix is determined only by the eigenvalues of the matrix.

4.2. **Construction of a Map.** Theorem 4.6 suggests that we ought to find the joint density of the eigenvalues by way of Theorem 2.13 because a vector of eigenvalues is a function of a matrix. However, this function is not nice enough for Theorem 2.13 so we instead map to the spectral decomposition of the matrix and use Theorem 2.21 to find the joint density of the eigenvalues. However, even this is not nice enough for Theorem 2.13. However, it turns out that it nice enough almost everywhere and this is sufficient for our purposes. This subsection is dedicated to the construction of the relevant map and set that the map is nice enough on.

4.2.1. *Throwing Away Sets of $\mathcal{H}_N^2$.*

Definitions 4.9. We state definitions that help us define the subsets we want to keep.

We call U_N the set of unitary matrices. We call D_N the set of diagonal matrices. We call U_N^g the set of good matrices. A matrix is good if it is unitary and its diagonals are strictly positive real and all other entries are nonzero. We call U_N^{vg} the set of very good matrices. A matrix is very good if it is good and all minors have nonzero determinants. We call D_N^d the set of distinct diagonal matrices. A matrix is distinct diagonal if each diagonal entry is distinct. We call D_N^{do} the set of decreasingly ordered distinct diagonal matrices. A distinct diagonal is decreasingly ordered if $D_{i,i} > D_{i+1,i+1}$ for $i = 1, \ldots, n-1$.

Before we make the next definition we note that if $X \in \mathcal{H}_N^2$, then X is Hermitian and so by spectral decomposition we know that X has a decomposition $UDU^\star$ where $\star$ indicates conjugate transpose and $U \in U_N$ and $D \in D_N$.

Definitions 4.10. We can now define the subsets we wish to keep.

We write $X \in \mathcal{H}_N^{2,dg}$ if $X = UDU^\star$ where $U \in U_N^g$ and $D \in D_N^d$. We write $X \in \mathcal{H}_N^{2,vg}$ if $X = UDU^\star$ where $U \in U_N^{vg}$ and $D \in D_N^d$.

We now prove a lemma that we will appeal to in the analysis of these sets.

Lemma 4.11. *A set $A \subset \mathbb{R}^n$ has Lebesgue measure zero and is closed if there exists a non-vanishing polynomial p of n variables such that $\boldsymbol{x} \in A$ if and only if $p(\boldsymbol{x}) = 0$.*

Proof. By hypothesis, $\{\mathbf{x} \colon p(\mathbf{x}) = 0\} = A$. The case $n = 1$ is clear as the set A must be finite. For the case $n > 1$, we use Fubini's Theorem and the base case:

$$\begin{aligned} \mu(\{\mathbf{x} \colon p(\mathbf{x}) = 0\}) &= \int_{\mathbb{R}^n} 1_{p(\mathbf{x})=0} d\mu\mathbf{x} \\ &= \int \cdots \int \int 1_{p(x_1,\ldots,x_n)=0} d\mu(x_1) d\mu(x_2) \cdots d\mu(x_n) \\ &= \int \cdots \int 0 d\mu(x_2) \cdots d\mu(x_n) = 0. \end{aligned}$$

That A is closed follows from it being the inverse image of $\{0\}$ under a continous function. □

We now begin to throw things out by proving a lemma that we will use to construct a polynomial. We note that given a matrix X we call $X^{(k,k)}$ the result of deleting the k-th column and k-th row from the matrix X.

Lemma 4.12. *For $X \in \mathcal{H}_N^2$, we have that $X \notin \mathcal{H}_N^{2,dg}$ if and only if X has a duplicate eigenvalue or for some k the matrix X and the matrix $X^{(k,k)}$ share an eigenvalue.*

Proof. Write $X = UDU^\star$ via spectral decomposition. First assume that $U \in U_N^g$. Clearly, if X has a duplicate eigenvalue then $D \notin D_N^d$ so $X \notin \mathcal{H}_N^{2,dg}$ but if X does not have a duplicate eigenvalue then $D \in D_N^d$ so $X \in \mathcal{H}_N^{2,dg}$.

Now, assume that $D \in D_N^d$. Let λ be an eigenvalue of X that we will pick later. Set $A = X - \lambda I$. Regardless of the eigenvalue, the null space of A has dimension 1 because the distinctness of the eigenvalues means the null space of A is spanned by a single eigenvector of X. Call the eigenvector in question v_λ. Take the adjoint of A, denoting it A^{adj} and see that $AA^{adj} = \det(A)I = 0$ as $\det X - \lambda I = 0$ by the definition of λ. Hence, the columns of A^{adj} are in the null space of A and can be written as scalar multiples of v_λ.

Now, suppose that X and $X^{(k,k)}$ share an eigenvalue λ. The definition of the adjoint tells us that some entry of A is 0. By the above, we have that some entry of v_λ is 0. Since v_λ is a column of U, we have that some entry of U is zero so then $U \notin U_N^g$. Starting from $U \notin U_N^g$, we have that at least one entry of one eigenvector of X is zero. Suppose this eigenvector is v_λ; we then know that some entry of A is 0. By the definition of the adjoint there is some k such that X and $X^{(k,k)}$ share the eigenvalue λ. □

Lemma 4.13. *The Lebesgue measure of* $\mathcal{H}_N^2 \setminus \mathcal{H}_N^{2,dg}$ *is zero.*

Proof. We show that there exists a polynomial p such that $p(X) = 0$ if and only if $X \in \mathcal{H}_N^2 \setminus \mathcal{H}_N^{2,dg}$. This will complete the proof by Lemma 4.11.

Let $p_0(X)$ be the discriminate of the characteristic polynomial of X. By definition, this is zero if and only if there is a repeated eigenvalue. For k such that $1 \le k \le N$, let p_k be the resultant the characteristic polynomials of X and $X^{(k,k)}$. By definition, $\prod_{i=1}^{N} p_i(X)$ is only zero if some X and $X^{(k,k)}$ share an eigenvalue for some k. Let $p(X) = \prod_{i=0}^{N} p_i(X)$. By Lemma 4.12, we have that $p(X) = 0$ if and only if $X \in \mathcal{H}_N^2 \setminus \mathcal{H}_N^{2,dg}$. □

Lemma 4.14. *The Lebesgue measure of* $\mathcal{H}_N^2 \setminus \mathcal{H}_N^{2,vg}$ *is zero.*

We define a subset of $\mathcal{H}_N^{2,vg}$ purely for the purpose of this proof.

Definition 4.15. A matrix $X = (UDU^\star) \in \mathcal{H}_N^{2,vg}$ is in $\mathcal{H}_N^{2,dvg}$ if D is strongly distinct, meaning that for any integer $r = 1, \dots, N-1$ and ordered r-indices I, J generated from the set $\{1, \dots, N\}$, it holds that $\prod_{i \in I} D_{i,i} \neq \prod_{j \in J} D_{j,j}$.

Proof of Lemma 4.14. It is sufficient to show that $\mathcal{H}_N^2 \setminus \mathcal{H}_N^{2,dvg}$ has Lebesgue measure 0. Select $X \in \mathcal{H}_N^2$ Given r, I, and J as in Definition 4.15, define

$$\big(\wedge X\big)_{IJ}^{r} := \det_{i,j}^{r} X_{I_i, J_j}$$

and define a matrix $\wedge^r X$ indexed by r-indices I and J. By the Cauchy-Binet Theorem, we have that if $X = UDU^\star$, $\wedge^r X = (\wedge^r U)(\wedge^r D)(\wedge^r U^\star)$. If U is not very good, then for some r at least one entry of $\wedge^r U$ must be 0. Take the product of all entries of $\wedge^r U$ to get a polynomial $p_1(X)$ that is zero if and only if U has a minor with non-vanishing determinant. Alternatively, if U is very good, but D is not strongly distinct then there is a duplicate in $\wedge^r D$, meaning $\wedge^r X$ has duplicate eigenvalues. Take the resultant of the characteristic polynomial of $\wedge^r X$ and call it $p_2(X)$. Finally, repeat the argument given in the proof of Lemma 4.13 that used Lemma 4.12 to generate polynomials that test for a non-good unitary matrix. Multiply these polynomials and p_1 and p_2 to arrive at a polynomial that is zero if and only if $X \in \mathcal{H}_N^2 \setminus \mathcal{H}_N^{2,dvg}$. Appeal to Lemma 4.11 to complete the proof. □

4.2.2. *Defining the required map.* We are now in a position to define the map that we need to perform change of variables. We start with some subsidiary maps.

Lemma 4.16. *The map $g(U,D) = UDU^\star$ with $g\colon (U_N^g, D_n^{do}) \to \mathcal{H}_N^{2,dg}$ is a bijection. If D_n^{do} is replaced with D_n^d the map becomes $N!$-to-one.*

Proof. That the map is onto is clear from the definition of $\mathcal{H}^{2,dg}$. To see that it is one to one observe that we cannot permute the eigenvalues or eigenvectors (i.e the rows of D or U). The distinct and ordered condition ensure this for the eigenvalues. The strictly positive real diagonal entries condition ensures this condition for the eigenvectors. The $N!$ occurs because of all possible permutations of the eigenvalues. □

Lemma 4.17. *The map $T\colon U_N^{vg} \to \mathbb{R}^{N(N-1)}$ defined by*

$$T(U) = (\frac{U_{1,2}}{U_{1,1}}, \dots, \frac{U_{1,N}}{U_{1,1}}, \frac{U_{2,3}}{U_{2,2}}, \dots, \frac{U_{2,N}}{U_{2,2}}, \dots, \frac{U_{N-1,N}}{U_{N-1,N-1}}) \tag{4.18}$$

has a one-to-one smooth inverse.

Proof. This is a proof by construction. The essential idea is that on each row we already know up to a multiple the value of every entry to the right of the diagonal and the very good condition allows us to find the rest. Additionally, the smoothness of the map will be evident in the construction.

For starters, since we know that $\sum_{i=1}^{N} |U_{1,i}|^2 = 1$, we have that $U_{1,1}^{-2} = 1 + \sum_{j=2}^{N} \frac{|U_{1,j}|^2}{|U_{1,1}|^2}$. Observe that it is clear that $U_{1,1}$ is smoothly determined and that this construction would not be possible if we did not require that $U_{i,i}$ be strictly positive real.

Now, suppose that for any $1 < i_0 < N$ we know the values of the entries of U for all rows $1 \leq i \leq i_0$. We can solve the following system of linear equations:

$$\begin{bmatrix} U_{1,1} & \cdots & U_{1,i_0} \\ U_{2,1} & \cdots & U_{2,i_0} \\ \cdots & \cdots & \cdots \\ U_{i_0,1} & \cdots & U_{i_0,i_0} \end{bmatrix} Z = - \begin{bmatrix} U_{1,i_0+1} + \sum_{i=i_0+2}^{N} U_{1,i}(\frac{U_{i_0+1,i}}{U_{i_0+1,i_0+1}})^\star \\ U_{2,i_0+1} + \sum_{i=i_0+2}^{N} U_{2,i}(\frac{U_{i_0+1,i}}{U_{i_0+1,i_0+1}})^\star \\ \cdots \\ U_{i_0,i_0+1} + \sum_{i=i_0+2}^{N} U_{i_0,i}(\frac{U_{i_0+1,i}}{U_{i_0+1,i_0+1}})^\star \end{bmatrix}$$

We note several items. The matrix on the left represents information we were given inductively. The vector on the right mostly comes from the unitary constraint on U and information from the map. Since the left matrix is a minor of U, the very good condition ensures that Z has a unique solution.

Then if we set

$$U_{i_0+1,i_0+1}^{-2} = 1 + \sum_{k=1}^{i_0} |Z_k|^2 + \sum_{i=i_0+2}^{N} |\frac{U_{i_0+1,i}}{U_{i_0+1,i_0+1}}|^2$$

it becomes clear that $U_{i_0+1,j} = Z_j^\star U_{i_0+1,i_0+1}$ for $1 \leq j \leq i_0$ as this is the choice that ensures the matrix is in U_N^{vg}. We observe that this fills in all the missing information for the row $i_0 + 1$ and that it is clear that entries are determined in a smooth manner.

The very last row follows directly from the unitary constraint. □

We prove a lemma regarding the image of $T(U)$ before continuing on to define the needed map.

Lemma 4.19. *The set* $\mathbb{R}^{N(N-1)} \setminus T(U_N^{vg})$ *is closed and has zero Lebesgue measure.*

Proof. Observe that the set is the vanishing set of polynomials generated by the application of the very good property in the proof of Lemma 4.17. Thus applying Lemma 4.11 to the product of these polynomials completes the proof. □

Finally we have our map. It is $\bar{T}(z, \lambda) \colon T(U_N^{vg}) \times \mathbb{R}^n \to \mathcal{H}_N^{2,vg}$ and it is defined by

$$\bar{T}(z, \lambda) = T^{-1}(z) D(\lambda) T^{-1}(z)^\star \tag{4.20}$$

where $D(\lambda)$ is the diagonal matrix generated from the vector of eigenvalues λ so that the i-th eigenvalue provided is the i-th entry of the diagonal of $D(\lambda)$.

4.2.3. *Properties of the Map and the Proof of the Density up to Normalization Constant.*

Lemma 4.21. *The map* $\bar{T}$ *is smooth,* $N!$*-to-1 on a set of full Lebesgue measure, and locally one to one on another set of full Lebesgue measure.*

Proof. Smoothness follows from Lemma 4.17. Combine this with Lemma 4.16 and Lemma 4.13 and Lemma 4.14 to conclude the rest. □

Lemma 4.22. *The Jacobian of* $\bar{T}$ *has the form*

$$g(z)\Delta(\lambda)^2 \tag{4.23}$$

for some continous function $g(z)$.

Proof. Let $W = T^{-1}(z)$. Write $\bar{T} = WDW^\star$ where $W^\star W = I$. If $d\bar{T}$ is the matrix of differentials corresponding to the derivative of $\bar{T}$, then multiple applications of the product rule yields that $d\bar{T} = (dW)DW^\star + W(dD)W^\star + WD(dW^\star)$. Another application gives us that $0 = dI = d(W^\star W) = (dW^\star)W + W^\star(dW)$. This gives us $(dW^\star)W = -W^\star(dW)$. Via substitution and the properties of unitary matrices, we get that

$$W^\star(d\bar{T})W = W^\star(dW)D - DW^\star(dW) + (dD) \tag{4.24}$$

We can write (4.24) as $A = CD - DC + dD$. Thus, for $j > k$ we have that

$$A_{j,k} = \sum_{i=1}^{N} C_{j,i}D_{i,k} - D_{j,i}C_{i,k}. \tag{4.25}$$

Since D is a diagonal matrix of eigenvalues $\lambda_1, \ldots, \lambda_N$, we have that $A_{j,k} = C_{j,k}\lambda_k - \lambda_j C_{j,k}$. This computation shows how the term $(\lambda_k - \lambda_j)$ appears in the determinant of A twice. If we observe that W and $W^\star$ do not depend on any λ_i for $i = 1, \ldots, N$, it is clear that the two appearances of $(\lambda_k - \lambda_j)$ must come from the determinant of $(d\bar{T})$ (the Jacobian of $\bar{T}$). Thus the Jacobian of $\bar{T}$ is zero if $\lambda_k = \lambda_j$ for $j > k$

and it is clear that for the same j and k, $(\lambda_k - \lambda_j)^2$ is a factor of the Jacobian of $\bar{T}$. This shows that $\Delta(\lambda)^2$ is a factor of the Jacobian of $\bar{T}$. A simple analysis of the degree of the Jacobian shows that this will factor all λ variables out, leaving only terms that depend on z. □

Remark 4.26. We can now complete the proof Theorem 4.2 modulo the value of the normalization constant. The proof is a computation using the change of variables formula. But before we carry it out, it is important to understand why it works.

First, why is Theorem 2.13 true? It is a corollary of the traditional change of variables theorem, but change of variables only holds on open sets whereas as a density function must integrate to the value of the probability measure on all measurable sets. Obviously some functions will integrate to the correct value on some measurable subsets but not all. We get around this conceptual issue by recalling that if probability measures are equal on an π-system of sets (a non-empty collection of sets closed under finite intersection), then they are equal on the σ-algebra. Hence, Theorem 2.13 works because if we can find the density on an π-system consisting only of open sets, then it automatically extends to showing that the resulting density works for all measurable sets.

With this in mind, we can see why our change of variables modulo closed sets of measure zero still works. What we are implicitly doing via this change of variables is saying this density calculates the right values on a collection of open sets created by taking all open sets and subtracting the closed sets of measure zero specified in Lemmas 4.19, 4.14, and 4.13. Moreover, these sets are enough to show this is the density that works for all measurable sets.

Proof of Theorem 4.6 modulo normalization. We start with (4.7):

$$\frac{dP_N^2}{d\mu}(H) = (2)^{-N/2}(\pi)^{-N^2/2}e^{-trH^2/2} = (2)^{-N/2}(\pi)^{-N^2/2}\prod_{i=1}^{N} e^{-\lambda_i^2(H)/2}$$

and we multiply it by $1_{\lambda_1 \leq \cdots \leq \lambda_N}(H)$. Since we are working only on the sets where the eigenvalues are ordered, this does not affect the value of (4.7). We apply Theorem 2.13, justifying it with Lemma 4.21 and Remark 4.26, finding

$$N!\frac{dP_N^2}{d\mu}(\bar{T}(z,\lambda))1_{\lambda_1 \leq \cdots \leq \lambda_N}(\bar{T}(z,\lambda))|J_{\bar{T}}(z,\lambda)|$$

where $J_{\bar{T}}$ is the Jacobian of $\bar{T}$. The $N!$ emerges from the extension to the alternative map in Lemma 4.21. It is not necessary to do this as constants do not matter in this process, but factoring the $N!$ out of the normalization constant makes the eventual evaluation of it easier.

Since the indicator and original density only depend on the eigenvalues, we reduce to

$$N!\frac{dP_N^2}{d\mu}(\lambda)1_{\lambda_1 \leq \cdots \leq \lambda_N}(\lambda)|J_{\bar{T}}(z,\lambda)|.$$

Factoring the Jacobian via Lemma 4.22 yields a nice form of the joint density of z and λ (the density of a vector in $\mathbb{R}^{N^2}$):

$$N!|g(z)|\frac{dP_N^2}{d\mu}(\lambda)1_{\lambda_1 \leq \cdots \leq \lambda_N}(\lambda)\Delta(\lambda)^2.$$

Via Theorem 2.21, we get the density of λ:

$$N!\Big(\int_{T(U_N^{vg})}|g(z)|\Big)\frac{dP_N^2}{d\mu}(\lambda)1_{\lambda_1\leq\cdots\leq\lambda_N}(\lambda)\Delta(\lambda)^2.$$

We mark the normalization constant and allow the constants in (4.7) to be absorbed into it:

$$N!C_N^2 1_{\lambda_1\leq\cdots\leq\lambda_N}(\lambda)\Delta(\lambda)^2\prod_{i=1}^N e^{-\lambda_i^2/2}.$$

This is (4.3) and completes the proof of Theorem 4.2 up to the value of the normalization constant,(4.4), which we deal with next. □

4.3. **Evaluation of the Normalization Constant and Selberg's Integral Formula.** Recalling Remark 2.10, we know that

$$\int_{\mathbb{R}^N} N!C_N^2 1_{\lambda_1\leq\cdots\leq\lambda_N}(\lambda)\Delta(\lambda)^2\prod_{i=1}^N e^{-\lambda_i^2/2}=1.$$

So we have that

$$N!C_N^2=N!\Big(\int_{\mathbb{R}^N}1_{\lambda_1\leq\cdots\leq\lambda_N}(\lambda)\Delta(\lambda)^2\prod_{i=1}^N e^{-\lambda_i^2/2}\Big)^{-1}.$$

Hence the proof of (4.4) is the evaluation of the integral on the right. It turns out that (4.4) follows from the $c=1$ case of the following formula.

Corollary 4.27. *For all positive real numbers c we have*

$$\frac{1}{N!}\int_{\mathbb{R}^N}|\Delta(x)|^{2c}\prod_{i=1}^N e^{-x_i^2/2}dx=(2\pi)^{N/2}\prod_{j=0}^{N-1}\frac{\Gamma((j+1)c)}{\Gamma(c)}. \tag{4.28}$$

We omit the proof of this corollary because it follows in a tedious but straightforward fashion from Selberg's Integral Formula.

Theorem 4.29 (Selberg's Integral Formula)**.** *For all positive real numbers a,b,c we have*

$$\frac{1}{N!}\int_0^1\cdots\int_0^1|\Delta(x)|^{2c}\prod_{i=1}^N x_i^{a-1}(1-x_i)^{b-1}dx_i=\prod_{j=0}^{N-1}\frac{\Gamma(a+jc)\Gamma(b+jc)\Gamma((j+1)c)}{\Gamma(a+b+(N-j-1)c)\Gamma(c)}.$$

We denote the left side $S_N(a,b,c)$. We spend the rest of this section proving Theorem 4.29. The basic strategy will be to transform the integral into an integral over polynomials and then develop tools for evaluating the aforementioned integrals.

4.3.1. *Integrals over Polynomials.*

Definitions 4.30. We set up a number of definitions regarding polynomials.

- We call the space of monic polynomials of degree n with n distinct real roots D_n. Given $I\subset\mathbb{R}$, D_nI is the subset of $P\in D_n$ with roots in I.
- We call the space of increasingly ordered n tuples of real numbers $\bar{D}_n$.
- Given a polynomial $P\in D_n$ with roots $\alpha_1<\cdots<\alpha_n$. Let $\tau_0(P)=1$ and for $n-k$ where $0<k\leq n$ let

$$\tau_k(P)=\sum_{1\leq i_1<\cdots<i_k\leq i_n}\alpha_{i_1}\cdots\alpha_{i_n}. \tag{4.31}$$

- Define $y(P)\colon D_n \to \mathbb{R}^n$ as $y(P) = (\tau_1(P), \cdots, \tau_n(P))$ and $P_x(t)\colon \mathbb{R}^n \to D_n$ as $P_x(t) = t^n + \sum_{i=1}^{n}(-1)^i x_{n-i}t^{n-i}$. These maps invert each other.
- Given a polynomial $P \in D_n$ with roots $\alpha_1 < \cdots < \alpha_n$, set $\Delta(P) = \Delta(\alpha_1, \cdots, \alpha_n)$.
- We define a measure L_n on $A \subset D_n$ as the Lebesgue measure of $y(A)$.

All of these definitions allow us to integrate over sets of polynomials, which we can treat as sets of coefficients. It is therefore of interest to investigate the nature of a transform from an integral over roots (points of $\bar{D}_n$) to polynomials (vectors of coefficients or equivalently D_n). We state without proof a lemma regarding the Jacobian of the transform.

Lemma 4.32. *For $k, l = 1, \ldots, n$ and $\alpha \in \bar{D}_n$, define $\tau_{k,l} = \dfrac{\partial \tau_k}{\partial \alpha_l}$. We have that*

$$|\det_{k,l}^{n} \tau_{k,l}| = \Delta(\alpha). \tag{4.33}$$

We can now provide a change of variables lemma that follows easily from our definitions and the last lemma.

Lemma 4.34. *For a non-negative L_n measurable function f on D_n, we have that*

$$\int_{D_n} f dL_n = \int_{\bar{D}_n} f(\prod_{i=1}^{n}(t-\alpha_i))\Delta(\alpha)d\alpha. \tag{4.35}$$

If we take $f(P) = |P(0)|^{a-1}|P(1)|^{b-1}(\Delta(P))^{2c-1}$ and use (4.35) we get that

$$S_N(a,b,c) = \frac{1}{N!}\int_{D_N(0,1)} |P(0)|^{a-1}|P(1)|^{b-1}(\Delta(P))^{2c-1}. \tag{4.36}$$

We have now reduced Theorem 4.29 to showing that right side of (4.36) is equal to the right side of Selberg's Integral Formula.

4.3.2. *Computational Lemmas.* To evaluate the right side of (4.36), we examine a new set of polynomials and develop lemmas to evaluate these sets.

Definition 4.37. Given a polynomial $P \in \mathcal{D}_n$ with roots $\alpha_1 < \cdots < \alpha_n$ and a polynomial $Q \in \mathcal{D}_{n+1}$ with roots $\beta_1 < \cdots < \beta_{n+1}$, we say that $(P,Q) \in \mathcal{E}_n$ if $\alpha_i \in (\beta_i, \beta_{i+1})$.

We prove a lemma that relates P and Q in the above definition in a non-obvious way.

Lemma 4.38. *Fix such a Q and real numbers $\gamma_1, \ldots, \gamma_n$. If $P(t)$ is the unique polynomial of degree n with at most n distinct real roots such that*

$$\frac{P(t)}{Q(t)} = \sum_{i=1}^{n+1} \frac{\gamma_i}{t-\beta_i} \tag{4.39}$$

then the following statements are equivalent:

- $(P,Q) \in \mathcal{E}_n$
- $\min \gamma_1, \ldots, \gamma_{n+1} > 0$ *and* $\sum_{i=1}^{n+1} \gamma_i = 1$

Proof. First from (4.39), we have that

$$P(t) = \sum_{i=1}^{n+1} \frac{\gamma_i}{t-\beta_i} Q(t). \tag{4.40}$$

From (4.40) alone we can derive one part of the equivalence. We have that $P(t)$ is monic if and only if the γ_i sum to one as each $Q(t)/(t-\beta_i)$ is monic.

Note that as $t \to \beta_j$ for $j = 1, \ldots, n+1$ in (4.40) we get $\frac{0}{0}$ so we can apply L'Hopitale's rule to find that $\gamma_i = P(\beta_i)/Q'(\beta_i)$. Using this, we can prove the interlacing of the roots of P with those of Q is equivalent to $\gamma_i > 0$ for all i.

In both cases, we know that since $Q(t) \in D_{n+1}$ has $n+1$ distinct real roots, it must switch signs at its roots.

Starting with the interlacing of P with Q, we infer that the $P(\beta_i)$ alternate signs. A similar analysis applies to the $Q'(\beta_i)$ because the derivative will have zeros at local maximums and minimums that occur between the roots β_i. Since both the $Q'(\beta_i)$ and the $P(\beta_i)$ are nonzero and alternate signs, concluding that $\min \gamma_1, \ldots, \gamma_{n+1} > 0$ amounts to showing $\gamma_{n+1} > 0$. This is clear because $P(t)/Q'(t)$ is continous in $[\beta_{n+1}, \infty)$ and by the monic property of the polynomials tends to $\frac{1}{n+1}$.

Starting with $P(\beta_i)/Q'(\beta_i) = \gamma_i > 0$ for all i, we can deduce that since $Q'(\beta_i)$ alternates signs, we have that P must alternate signs $n+1$ times. By the intermediate value property, P must have n distinct roots and they must be interlaced with Q's roots. □

We now prove a lemma that gives us an explicit formula for the measure of certain subsets of D_n.

Lemma 4.41. *For fixed $Q \in \mathcal{D}_{n+1}$ with roots $\beta_1 < \cdots < \beta_{n+1}$*

$$L_n(\{P \colon (P,Q) \in \mathcal{E}_n\}) = \frac{|\Delta(Q)|}{n!} = \frac{1}{n!} \prod_{j=1}^{n+1} |Q'(\beta_j)|^{1/2}. \tag{4.42}$$

Proof. By definition, the L_n measure of $\{P \colon (P,Q) \in \mathcal{E}_n\}$ is the Lebesgue measure of $\{x \in \mathbb{R}^n \colon (P_x, Q) \in \mathcal{E}_n\}$. By Lemma 4.38, the former set is the simplex formed by the polynomials $Q_j = Q(t)/(t-\beta_j)$ for $j = 1, \ldots, n+1$. Thus, the latter set can be viewed as the interior of the simplex formed by points Q_j translated into euclidean space. In other words, A is the simplex formed by the points $(\tau_{2,j}(Q), \ldots, \tau_{n+1,j}(Q))$ for $j = 1, \ldots, n+1$. Using the typical formula for the volume of a simplex and Lemma 4.32, we get the first equality in (4.42). The second equality can be proved by induction on n. □

We now prove an explicit formula for an integral over polynomials.

Lemma 4.43. *For fixed positive real numbers $s_1 \ldots, s_{n+1}$ and for fixed $Q \in \mathcal{D}_{n+1}$ with roots $\beta_1 < \cdots < \beta_{n+1}$, we have:*

$$\int_{\{P \in \mathcal{D}_n \colon (P,Q) \in \mathcal{E}_n\}} \prod_{i=1}^{n+1} |P(\beta_i)|^{s_i - 1} dL_n(P) = \frac{\prod_{i=1}^{n+1} |Q'(\beta_i)|^{s_i - 1/2} \Gamma(s_i)}{\Gamma(\sum_{i=1}^{n+1} s_i)}. \tag{4.44}$$

Proof. We first prove that

$$\int_{\{P \in \mathcal{D}_n \colon (P,Q) \in \mathcal{E}_n\}} \prod_{i=1}^{n+1} |\frac{P(\beta_i)}{Q'(\beta_i)}|^{s_i - 1} = \frac{\Gamma(s_i)}{\Gamma(\sum_{i=1}^{n+1} s_i)}. \tag{4.45}$$

Picking $\gamma_i = P(\beta_i)/Q'(\beta_i)$ for $i = 1, \ldots, n+1$, this equation follows from Lemma 4.38 and the following well known integral:

$$\int_{\{(x_1,\ldots,x_N):\, x_i\in(0,1),\sum x_i=1\}} \prod_{i=1}^{N} x_i^{s_i-1} = \frac{\prod_{i=1}^{N}\Gamma(s_i)}{\Gamma(\sum_{i=1}^{N} s_i)}. \tag{4.46}$$

Equation (4.45) is actually (4.44) times a constant $C = (\prod_{i=1}^{n+1} Q'(\beta_i))^{1/2}$ on the right side. If we take $s_i = 1$ for $i = 1, \ldots, n$, equation (4.42) tells us that the left side of (4.45) is just $\frac{C}{(n)!}$ and the right side is $\frac{1}{n!}$. Thus $C = 1$ and since C is independent of the choice of s_i, equation (4.44) follows. □

4.3.3. *Proof of Selberg's Integral Formula.* We now have proven enough computational lemmas to prove Selberg's integral formula by induction.

Proof. The case $N = 1$ is the well known Beta integral:

$$S_1(a,b,c) = \int_0^1 x^{a-1}(1-x)^{b-1} = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}. \tag{4.47}$$

To prove the inductive case, we examine the following double integral:

$$K_N(a,b,c) := \int_{\mathcal{E}_N(0,1)} |Q(0)|^{a-1}|Q(1)|^{b-1}|R(P,Q)|^{c-1} dL_N(P)dL_{N+1}(Q) \tag{4.48}$$

where $R(P,Q)$ is the resultant, which in this case satisfies

$$R(P,Q) = \prod_{i=1}^{N} Q(\alpha_i) = \prod_{j=1}^{N+1} P(\beta_j). \tag{4.49}$$

We evaluate (4.48) via two different iterated integrals, which we can do via Fubini's Theorem. Integrating with respect to P then with respect to Q we have

$$K_N(a,b,c) = \int_{\mathcal{D}_{N+1}(0,1)} |Q(0)|^{a-1}|Q(1)|^{b-1} \int_{\{P:\,(P,Q)\in\mathcal{E}_N\}} |R(P,Q)|^{c-1}. \tag{4.50}$$

The inner integral is evaluated via Lemma 4.43 with $s_i = c$ for all i:

$$\begin{aligned}\int_{\{P:\,(P,Q)\in\mathcal{E}_N\}} |R(P,Q)|^{c-1} dL_N(P) &= \int_{\{P:\,(P,Q)\in\mathcal{E}_N\}} \prod_{i=1}^{N+1} |P(\beta_i)|^{c-1} dL_N(P) \\ &= \frac{\prod_{i=1}^{N+1} |Q'(\beta_i)|^{c-1/2}\Gamma(c)^{N+1}}{\Gamma((N+1)c)}.\end{aligned}$$

We evaluate the outer integral via an application of Lemma (4.42):

$$\begin{aligned} K_N(a,b,c) &= \frac{\Gamma(c)^{N+1}}{\Gamma((N+1)c)} \int_{\mathcal{D}_{N+1}(0,1)} |Q(0)|^{a-1}|Q(1)|^{b-1} \prod_{i=1}^{N+1} |Q'(\beta_i)|^{c-1/2} \\ &= \frac{\Gamma(c)^{N+1}}{\Gamma((N+1)c)} \int_{\mathcal{D}_{N+1}(0,1)} |Q(0)|^{a-1}|Q(1)|^{b-1} \prod_{i=1}^{N+1} |Q'(\beta_i)|^{\frac{2c-1}{2}} \\ &= \frac{\Gamma(c)^{N+1}}{\Gamma((N+1)c)} \int_{\mathcal{D}_{N+1}(0,1)} |Q(0)|^{a-1}|Q(1)|^{b-1} \frac{|\Delta(Q)|^{2c-1}}{(N+1)!} \\ &= \frac{\Gamma(c)^{N+1}}{\Gamma((N+1)c)} S_{N+1}(a,b,c). \end{aligned}$$

Evaluating $K_N(a,b,c)$ with respect to Q then with respect to P gives us

$$K_N(a,b,c) = \int_{\mathcal{D}_N(0,1)} \int_{\{Q\colon (Q,P)\in\mathcal{E}_{N+2}\}} |Q(0)|^{a-1}|Q(1)|^{b-1}|R(P,Q)|^{c-1}. \tag{4.51}$$

We evaluate the inner integral of (4.51) first and the key realization is that since $R(P,Q) = \prod_{i=1}^{N} Q(\alpha_i)$, we can treat the integrand as an instance of Lemma 4.43 if we take P in (4.44) to be Q and Q to be $S(t) = (t-1)tP(t)$. Taking $s_i = c$ for $i = 1,\ldots,N$ and $s_{N+1} = a$ and $s_{N+2} = b$, we have

$$K_N(a,b,c) = \frac{\Gamma(a)\Gamma(b)\Gamma(c)^N}{\Gamma(a+b+Nc)} \int_{\mathcal{D}_N(0,1)} |S'(0)|^{a-1/2}|S'(1)|^{b-1/2}|R(P,S')|^{c-1/2} dL_N(P).$$

Using the formula for the resultant and Lemma (4.42), we have

$$\begin{aligned} K_N(a,b,c) &= \frac{\Gamma(a)\Gamma(b)\Gamma(c)^N}{\Gamma(a+b+Nc)} \int_{\mathcal{D}_N(0,1)} |S'(0)|^{a+c-1}|S'(1)|^{b+c-1} \frac{\Delta(S')^{c-1/2}}{N!} \\ &= \frac{\Gamma(a)\Gamma(b)\Gamma(c)^N}{\Gamma(a+b+Nc)} S_N(a+c,b+c,c). \end{aligned}$$

Combining the two ways of evaluating $K_N(a,b,c)$ gives us the relation

$$S_{N+1}(a,b,c) = S_N(a+c,b+c,c)\frac{\Gamma(a)\Gamma(b)\Gamma((N+1)c)}{\Gamma(c)\Gamma(a+b+Nc)}. \tag{4.52}$$

and this is enough to prove the inductive step that completes the proof of Selberg's integral formula. □

Remark 4.53. Besides being involved in the proof, equations (4.47) and (4.46) are related to Selberg's integral formula in another way. The first equation is clearly the formula that Selberg's integral formula generalizes, but it is also the formula for the normalization constant used in the Beta distribution. The Beta distribution is generalized by the Dirichlet distribution and the normalization constant for the Dirichlet distribution is given by (4.46).

With Selberg's integral formula established, we have finished this section.

5. Sketch of the Limiting Distribution of the GUE and Tracy-Widom

Wigner's Theorem establishes the limiting distribution of the average eigenvalue of a general class of random matrices. The study of the GUE yielded a formula for the joint distribution of the eigenvalues of a matrix of the GUE of a fixed size. In this section, we sketch a proof of the limiting case of the joint distribution of the eigenvalues of the GUE and briefly examine the limiting distribution of the largest eigenvalue of the GUE, called the Tracy-Widom distribution. In order to do this we briefly discuss the Fredholm Determinant and a few special functions.

Definitions 5.1. The **Airy function** is given by

$$Ai(x) := \frac{1}{2\pi i}\int_C e^{y^3/3-xy}dy \tag{5.2}$$

where C is the path starting at $-\infty$ with argument $-\frac{\pi}{2}$ and ending at ∞ with argument $\frac{\pi}{2}$. The **Airy kernel** is given by

$$A(x,y) := \frac{Ai(x)Ai'(y) - Ai'(x)Ai(y)}{x-y}. \tag{5.3}$$

We are now prepared to state the main result concerning the limiting joint distribution of the eigenvalues of the GUE.

Theorem 5.4. *For all $-\infty < t \leq t' \leq \infty$, we have that*

$$\lim_{N\to\infty} P\left(N^{2/3}(\frac{\lambda_{i,N}}{\sqrt{N}} - 2) \notin [t,t'], i = 1,\dots,N\right) = \tag{5.5}$$
$$1 + \sum_{k=1}^{\infty}\frac{(-1)^k}{k!}\int_t^{t'}\cdots\int_t^{t'}\det_{i,j=1}^{k} A(x_i,x_k)\prod_{a=1}^{k}dx_a.$$

5.1. Reformulation of the Density of the GUE.

In order to prove Theorem 5.4, it is necessary to rewrite the result of Theorem 4.2 so that it is more amenable to the theory of Fredholm Determinants. We introduce some special functions that are used to do this.

Definitions 5.6. The n-th **Hermite polynomial** is given by

$$\mathcal{N}_n(x) := (-1)^n e^{-x^2/2}\frac{d^n}{dx^n}e^{-x^2/2}. \tag{5.7}$$

The n-th **normalized oscillator wave function** is given by

$$\phi_n(x) := \frac{e^{-x^2/4}\mathcal{N}_n(x)}{\sqrt{n!\sqrt{2\pi}}}. \tag{5.8}$$

We observe that a Hermite polynomial is monic and that $\int \phi_k\phi_l$ is 1 if $k = l$ and 0 otherwise.

We can now state the rewrite.

Lemma 5.9. *The joint density with respect to Lebesgue measure of the eigenvalues of the GUE is given by*

$$p_N(\lambda_1,\dots,\lambda_N) = \frac{1}{N!}\det_{k,l=1}^{N} K^N(\lambda_k,\lambda_l) \tag{5.10}$$

where

$$K^N(x,y) = \sum_{k=0}^{N-1} \phi_k(x)\phi_k(y). \tag{5.11}$$

Proof. We start with the known form of the joint density, given by

$$p_N(\lambda_1,\ldots,\lambda_N) = (2\pi)^{-N/2}\prod_{j=1}^{N}\frac{1}{\Gamma(j)}\Delta(\lambda_1,\cdots,\lambda_N)^2\prod_{i=1}^{N}e^{\frac{-\lambda_i^2}{2}}. \tag{5.12}$$

We observe that monic property of the Hermite polynomials in order to see that $\Delta(\lambda_1,\cdots,\lambda_n) = \det_{i,j=1}^{N}\mathcal{N}_{j-1}(\lambda_i)$. Substituting this into (5.12) and then absorbing the normalization constant gives us that

$$\begin{aligned} p_N(\lambda_1,\ldots,\lambda_N) &= (2\pi)^{-N/2}\prod_{j=1}^{N}\frac{1}{\Gamma(j)}(\det_{i,j=1}^{N}\mathcal{N}_{j-1}(\lambda_i))^2\prod_{i=1}^{N}e^{\frac{-\lambda_i^2}{2}} \\ &= \frac{1}{N!}(\det_{i,j=1}^{N}\phi_{j-1}(\lambda_i))^2\prod_{i=1}^{N}e^{\frac{-\lambda_i^2}{2}}. \end{aligned}$$

Applying the usual product determinant rule and the definition of matrix multiplication completes the proof. □

By assuming certain properties of square integrable functions, we can now prove the following lemma that sets up the use of Fredholm Determinant.

Lemma 5.13. *For any measurable subset A of $\mathbb{R}$, we have that*

$$\mathbb{P}(\cap_{i=1}^{N}\{\lambda_{i,N}\in A\}) = 1 + \sum_{k=1}^{\infty}\frac{(-1)^k}{k!}\int_{A^c}\cdots\int_{A^c}\det_{i,j=1}^{k}K^N(x_i,x_j)dx. \tag{5.14}$$

However, in the interest of space and time we do not provide a proof, but only note that it follows from Lemma 5.9, the properties of ϕ_k, and the following formula that holds for square integrable real functions $f_1,\ldots,f_n,g_1,\ldots,g_n$:

$$\frac{1}{n!}\int\cdots\int\det_{i,j=1}^{n}(\sum_{k=1}^{n}f_k(x_i)g_k(x_j)) = \det_{i,j=1}^{n}\int f_i(x)g_j(x)dx.$$

5.2. **The Fredholm Determinant.** We define the Fredholm Determinant of a kernel and note one useful result.

Definition 5.15. A **kernel** on $\mathbb{R}$ is a measurable complex valued function $K(x,y)$ defined on $\mathbb{R}^2$ such that

$$\sup_{(x,y)\in\mathbb{R}^2}|K(x,y)| < \infty.$$

The **Fredholm Determinant** of K is defined as

$$\Delta(K) := \sum_{n=0}^{\infty}\frac{(-1)^n}{n!}\int\cdots\int\det_{i,j=1}^{n}K(x_i,x_j)d\mu(x_1,\ldots,x_n). \tag{5.16}$$

The one result that we require is noted.

Lemma 5.17. *Given a sequence of kernels $A_n(x,y)$ and another kernel $A(x,y)$, if $\sup_{(x,y)\in\mathbb{R}^2}|A_n(x,y)-A(x,y)|\to 0$ then $\Delta(A_n)\to\Delta(A)$.*

5.3. **The Limiting Distribution and Tracy-Widom.** The right side of (5.14) is clearly $1+\Delta(K^N)$. We recognize that the right side of (5.5) is $1+\Delta(A)$. Thus, Lemma 5.17 tells us that the proof of Theorem 5.4 modulo normalization amounts to showing that $K^n \to A$. We now investigate the normalization as the main component of an incomplete proof.

Proof of Theorem 5.4. Via (5.14), we have that the left side of (5.5) is equal to

$$\lim_{N\to\infty} 1+\sum_{k=1}^{\infty}\frac{(-1)^k}{k!}\int_t^{t'}\cdots\int_t^{t'}\det_{i,j=1}^{k}\frac{1}{N^{1/6}}K^N(2\sqrt{N}+\frac{x_i}{n^{1/6}},2\sqrt{N}+\frac{x_j}{N^{1/6}})dx.$$

We have just plugged in the normalization that is attached to the eigenvalues in (5.5) and we set $A^c=[t,t']$. We now move the normalization inside the kernel by setting

$$\Phi_N(x) := N^{1/12}\phi_N(2\sqrt{N}+\frac{x}{N^{1/6}}) \tag{5.18}$$

and then setting

$$A^N(x,y) := \sum_{k=0}^{N-1}\Phi_k(x)\Phi_k(y). \tag{5.19}$$

The objective is now to show that $A^N\to A$ as $N\to\infty$. Two identities accomplish this and we state them without proof. First, $A^N(x,y)$ can be written to look like the Airy kernel:

$$A^N(x,y) = \frac{\Phi_N(x)\Phi_N'(y)-\Phi_N(y)\Phi_N'(x)}{x-y}-\frac{1}{2N^{1/3}}\Phi_N(x)\Phi_N(y). \tag{5.20}$$

This reduces the problem to showing that $\Phi_N\to Ai(x)$ and $\Phi_N'(x)\to Ai'(x)$. These two together constitute the second identity. We do not show this, but refer the reader to the method of steepest descent and [2]. □

This argument establishes the limiting result for the joint distribution of the eigenvalues of the GUE, but its real value is that it leads to the following result regarding the limiting distribution of the largest eigenvalue of the GUE.

Theorem 5.21 (Tracy-Widom)**.**

$$\lim_{N\to\infty} P\left(N^{2/3}(\frac{\lambda_{N,N}}{\sqrt{N}}-2)\le t\right) = \exp(-\int_t^{\infty}(x-t)q(x)^2dx) \tag{5.22}$$

where q satisfies

$$q'' = tq+2q^3, q(t)/Ai(t)\to 1 \text{ as } t\to\infty. \tag{5.23}$$

We note that (5.23) is the Painlevè II differential equation and the item on the right of (5.22) is the Tracy-Widom distribution. We observe that this yields a much more wieldy formula than the previous result. We also note that largest eigenvalue is often of applied and theoretical interest. See [8] for more information.

6. Acknowledgments

This paper would not have been possible without my mentor, Jonathan Wang, who suggested the topic to me and provided a substantial amount of advice and aid. Many of the results and proofs of this paper are elaborations, informed by my own understanding, of those given in [2], which is a useful encyclopedia of results and proofs in Random Matrix Theory. Occasionally alternative formulations of these results and proofs were useful and in this regard I enjoyed [6] and [9]. I would also like to thank all the faculty involved in the REU. In particular, I would like to thank Professor Peter May for creating, organizing, and hosting the REU.

References

[1] Erhan Cinlar Probablity and Stochastics Springer
[2] Grew W. Anderson,Alice Guionnet, Ofer Zeitouni An Introduction to Random Matrices Cambridge University Press
[3] Craig A. Tracy and Harold Widom The Distributions of Random Matrix Theory and their Applications https://www.math.ucdavis.edu/ tracy/talks/SITE7.pdf
[4] Gemot Akemann, Jinho Balk and Philippe Di Francesco The Oxford Handbook of Random Matrix Theory Oxford University Press
[5] Yi-Kai Liu Statistical Behavior of the Eigenvalues of Random Matrices
[6] Terrance Tao Topics in Random Matrix Theory https://terrytao.files.wordpress.com/2011/02/matrix-book.pdf
[7] Achim Klenke Probability Theory: A Comprehensive Course Springer
[8] Sergey V. Lototsky Tracy-Widom Distribution and its Applications http://www-bcf.usc.edu/ lototsky/Math625-14/Week15-3.pdf
[9] Todd Kemp MATH 247A: INTRODUCTION TO RANDOM MATRIX THEORY http://www.math.ucsd.edu/ tkemp/247A/247A.Notes.pdf
[10] Laura Veith Proof of Wigner's Semicircular Law by the Stieltjes Transform Method https://www.math.washington.edu/ morrow/papers/laura-thesis.pdf