WIGNER'S SEMICIRCLE LAW

MEHMET OZAN BAYKAN

Abstract. In this paper, we present a combinatorial proof of the weak, in probability convergence of empirical spectral density of $N \times N$ Wigner matrices to Wigner's Semicircle Law in the limit $N \to \infty$. As a prequel, we develop the measure theoretic notions required to grasp the mathematical objects and notions in probability and random matrix theory.

Introduction

Wigner's Semicircle Law, as first described by Eugene P. Wigner in his seminal work [9] describes the limiting distribution the eigenvalues of a large class of random matrices with certaiin conditions of symmetry, independence and moment bounds. The motivation for the problem for Wigner arose in his studies of wave functions of quantum mehcanical systems, yet, the result have proven to be mathematically significant in probability theory. The result has been important in the algebraically-rich formulation of probabaility theory under the topic of *free probabilty.* For further discussion of free probability and the technical aspects thereof, the reader should consult [11].

In this paper, we aim to explore this universality result by first developing a rigorous grounding in the language of measure theory. The exposition of basics of measure theory is for two purposes, first, to develop the necessary tools and objects to define notions of convergence required to formulate Wigner's result, and second, to build intuition on how to apporach probability theory in a rigorous way. The first section is reserved for the developing the background in measure and probability. The second and main section presents the result, and provides a combinatorial approach to proving Wigner's Semicirlce Law.

1. Preliminary Results and Notation

1.1. Measure and Probability.

Definition 1.1. Let X be a set. A collection $\mathcal{A} \subseteq \mathcal{P}(X)$ of subsets of X is called a *σ-algebra* if,

(i) $X \in \mathcal{A}$
(ii) If $A \in \mathcal{A}$, then $A^{\mathsf{c}} \in \mathcal{A}$, i.e $\mathcal{A}$ is closed under complements.
(iii) If $\{A_i\}$ is an infinite collection of sets in $\mathcal{A}$, then $(\bigcup_{i=1}^{\infty} A_i) \in \mathcal{A}$, i.e, $\mathcal{A}$ is closed under arbitrary unions.

Proposition 1.2. *Let X be a set, and $\{\mathcal{A}_i\}_{i \in I}$ be an arbitrary collection of σ-algebras on X indexed by an arbitrary index set I. Then, $\bigcap_i \mathcal{A}_i$ is a also σ-algebra.*

Proof of Proposition 1.2, [4]. Let $\mathcal{A} := \bigcap_i \mathcal{A}_i$. First, we observe that $X \in \mathcal{A}$, since by definition $X \in \mathcal{A}_i$ for all $i \in I$. Now suppose $A \in \mathcal{A}$, then, by the fact that $A \in \mathcal{A}_i$ for all $i \in I$ and that $\mathcal{A}_i$ are σ-algebras, $A^{\mathsf{c}} \in \mathcal{A}$. Finally, let $\{A_k\}$ be a sequence of sets in $\mathcal{A}$, hence $\{A_k\}$ is in every $\mathcal{A}_i$, hence $\bigcap_k \mathcal{A}_k$ is in every $\mathcal{A}_i$, hence in $\mathcal{A}$. Therefore, $\mathcal{A}$ is a σ-algebra. □

Proposition 1.2 allows us to define the σ-algebra generated by a collection of subsets through the following corollary.

Corollary 1.3. *Let X be a set, and $\mathcal{F} \subseteq \mathcal{P}(X)$ a collection of subset of X. Then, there exists a unique smallest σ-algebra on X that includes $\mathcal{F}$. We denote $\sigma(\mathcal{F})$, and call it* σ-algebra generated by $\mathcal{F}$.

Proof of Corollary 1.3, [4]. Let $\mathcal{S}$ be the collection of σ-algebras on X that include $\mathcal{F}$. $\mathcal{S}$ in non-empty, since $\mathcal{P}(X)$ is a σ-algebra on X for any X. Let $\sigma(\mathcal{F}) := \bigcap_{\mathcal{A} \in \mathcal{S}} S$. By Proposition 1.2, $\sigma(\mathcal{F})$ is a σ-algebra, and by definition, is included in every σ-algebra that includes $\mathcal{F}$, hence, is the smallest such σ-algebra. Moreover, if $\sigma'(\mathcal{F})$ is also the smallest such σ-algebra, then $\sigma(\mathcal{F}) \subseteq \sigma'(\mathcal{F})$ and $\sigma'(\mathcal{F}) \subseteq \sigma(\mathcal{F})$, hence $\sigma(\mathcal{F}) = \sigma'(\mathcal{F})$, thus $\sigma(\mathcal{F})$ is unique. □

Definition 1.4. The *Borel σ-algebra*, $\mathcal{B}(\mathbb{R}^d)$ is the σ-algebra on $\mathbb{R}^d$ generated by the open sets in $\mathbb{R}^d$ with respect to the usual (metric) topology on $\mathbb{R}^d$, i.e $\mathcal{B}(\mathbb{R}^d) := \sigma(\{U \subseteq \mathbb{R}^d : U \text{ is open}\})$

Definition 1.5. Let X be a set, and $\mathcal{A}$ be a σ-algebra on X. A function $\mu : \mathcal{A} \to [0, +\infty]$ is a *measure*, if

(i) $\mu(\emptyset) = 0$
(ii) If $\{A_i\}$ is an infinite collection of disjoint sets in $\mathcal{A}$, then $\mu\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i)$, i.e, μ is countably additive.

Definition 1.6. Let X be a set, and $\mathcal{A}$ be a σ-algebra on X, then the tuple $(X, \mathcal{A})$ is called a *measurable space*. If $\mu : \mathcal{A} \to \mathbb{R}$ is a measure on $(X, \mathcal{A})$, then triplet $(X, \mathcal{A}, \mu)$ is called a *measure space*.

A measure on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is called a *Borel measure* on $\mathbb{R}^d$. In the following section, we construct the Lebesgue measure, the canonical Borel measure. For this construction, we first define outer measures, which will lead us to the notion of measurability and construction of measures for any measurable space. For our purposes, it suffices to construct the Lebesgue measure on $\mathbb{R}$, hence, we will limit ourselves to this case. For a more comprehensive construction, the reader may consult [4], Chatper 1.

It is appropriate to define probability spaces here, since as Tao notes in [11], "At a purely formal level, one could call probability theory the study of measure spaces with total measure one.". In this spirit, a *probability space* is a measure space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\mathbb{P}(\Omega) = 1$. In this context, we refer to elements of Ω as *elementary outcomes*, and set in $\mathcal{F}$ *events*. Hence, for any $A \in \mathcal{F}$, $\mathbb{P}(A)$ is called the *probability of A*

Definition 1.7. Let X be a set. A function $\mu^* : \mathcal{P}(X) \to [0, +\infty]$ is an *outer measure*, if

(i) $\mu(\emptyset) = 0$
(ii) If $A \subseteq B \subseteq X$, then $\mu^*(A) \leq \mu^*(B)$, i.e μ^* is non-decreasing.
(ii) If $\{A_i\}$ is an infinite collection of subsets of X, then $\mu^*\left(\bigcup_{i=1}^{\infty} A_i\right) \leq \sum_{i=1}^{\infty} \mu^*(A_i)$, i.e, μ^* is countably subadditive.

Definition 1.8. Let $A \subseteq \mathbb{R}$, and let $\mathcal{C}_A := \{\{(a_i, b_i)\}_{i=1}^{\infty} : A \subseteq \bigcup_i (a_i, b_i)\}$ the set of all infinite sequence of open bounded intervals such that A is contained in their union. The *Lebesgue outer measure* on $\mathbb{R}$, $\lambda^* : \mathcal{P}(\mathbb{R}) \to [0, +\infty]$ is defined as

$$\lambda^*(A) := \inf\{\sum_i (b_i - a_i) : \{(a_i, b_i)\}_{i=1}^{\infty} \in \mathcal{C}_A\}$$

Proposition 1.9. *The Lebesgue outer measure λ^* is an outer measure, and assigns to each subinterval of $\mathbb{R}$ its length.*

Proof of Proposition 1.9, [4] *(with slight modifications).* First, we show that λ^* is indeed an outer measure. $\lambda^*(\emptyset) = 0$ follows from the fact that for any $\epsilon > 0$, there exists an infinite sequence of bounded open intervals $\{(a_i, b_i)\}_{i=1}^{\infty}$ such that $\sum_i^{\infty} (b_i - a_i) < \epsilon$, e.g, letting $\{(a_i, b_i)\}_{i=1}^{\infty}$ such that $a_i = 0$ for all $i \in \mathbb{N}$, and $b_i := (-1)^{i+1}(\frac{\epsilon}{\lfloor \frac{i+1}{2} \rfloor})$, and observing that $0 < \sum_{i=1}^{k} (b_i - a_i) \leq \epsilon/k < \epsilon$ for any $k \in \mathbb{N}$. Since every interval includes the empty set by definition, we have $\lambda^*(\emptyset) = 0$.

To show monotonicity, we observe that if $A \subseteq B \subseteq \mathbb{R}$, then any sequence of intervals that cover A also cover B, hence $\lambda^*(A) \leq \lambda^*(B)$

To show countable subadditivity, let $A_{n_{n=1}}^{\infty}$ be a sequence of subsets of $\mathbb{R}$. If $\sum_{n=1}^{\infty} A_n = +\infty$, the result follows. Suppose $\sum_{n=1}^{\infty} A_n < +\infty$, and let $\epsilon < 0$. For each $n \in \mathbb{N}$, let $\{(a_{n,i}, b_{n,i})\}_{i=1}^{\infty}$ a sequence that covers A_n and satisfies,

$$\sum_{i=1}^{\infty} (b_{n,i} - a_{n,i}) < \lambda^*(A_n) + \frac{\epsilon}{2^n}$$

By the definition of λ^* and the Approximation Property of Infimum, such a sequence of intervals surely exists. Suppose we combine this collection of sequences into one sequence $\{(a_j, b_j)\}$. The particular way we do this is irrelevant for this proof, and this may be achieved by any technique for the enumeration of $\mathbb{N} \times \mathbb{N}$. Now, observe that the new sequence, by definition, covers the union $\bigcup_n A_n$, hence,

$$\lambda^*(\cup_n A_n) \leq \sum_j (b_j, a_j) < \sum_n (\lambda^*(A_n) + \frac{\epsilon}{2^n}) = \sum_n \lambda^*(A_n) + \epsilon$$

which implies $\lambda^*(\cup_n A_n) \leq \sum_n \lambda^*(A_n)$ as $\epsilon > 0$ is arbitrary. Therefore, λ^* is an outer measure on $(\mathbb{R}, \mathcal{B}\mathbb{R})$

To compute the value of λ^* on subintervals of $\mathbb{R}$, let $a < b$, and consider the closed interval $[a, b]$. Next, we consider the sequence of intervals we used to show $\lambda^*(\emptyset) = 0$, but with $(a_0, b_0) = (a - \epsilon/4, b + \epsilon/4)$. Then, $\{a_i, b_i\}_{i=0}^{\infty}$ includes $[a, b]$, and $\sum_{i=0}^{k} (b_i - a_i) < (b - a) + \epsilon$ for $k \in \mathbb{N}, k > 4$, hence $\lambda^*([a, b]) \leq (b - a)$. For the reverse inequality, we use the Heine-Borel Theorem to see that for any sequence of bounded open intervals $\{(a_i, b_i)\}$ that cover $[a, b]$, there exists $n \in \mathbb{N}$ such that $[a, b] \subseteq \bigcup_{i=1}^{n} (a_i, b_i)$. By induction on n, one can show that $b - a \leq \sum_{i=1}^{n} (b_i - a_i)$. Hence, $b - a \leq \sum_{i=1}^{\infty} (b_i - a_i)$ for any such sequence, therefore, $\lambda^*([a, b]) = b - a$..

To see that the result holds for any subinterval I, we may use the monotonicity of outer measures and the fact that any interval may be approximated by a sequence of closed intervals that cover and approach I to show that intervals with identical endpoints have the same length under λ^*. Lastly, since any unbounded interval includes arbitrarily long closed intervals, unbounded intervals have infinite Lebesgue outer measure. □

Definition 1.10. Let X be a set and μ^* an outer measure on X. A subset B of X is μ^**-measurable* if for any $A \subseteq X$, $\mu^*(A) = \mu^*(A \cap B) + \mu^*(A \cap B^c)$

Here, we outline how to use the tools we have defined up to this point to define the Lebesgue measure on $\mathbb{R}^d$. Using this definition of measurability and the Lebesgue outer measure, one may construct the σ-algebra $\mathcal{M}_{\lambda^*}$ defined to be the σ-algebra generated by λ^*-measurable subset of $\mathbb{R}^d$, and show that the restriction of λ^* to $\mathcal{M}_{\lambda^*}$, which we denote by λ from now on, is indeed a measure on $(\mathbb{R}^d, \mathcal{M}_{\lambda^*})$. Moreover, one can show that $\mathcal{B}(\mathbb{R}^d) \subseteq \mathcal{M}_{\lambda^*}$, i.e, that all Borel subsets in $\mathbb{R}^d$ are Lebesgue measurable. However, for our purposes, it will suffice to work with $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \lambda)$, so we will not be characterizing $\mathcal{M}_{\lambda^*}$. Note that $\mathcal{B}(\mathbb{R}^d) \neq \mathcal{M}_{\lambda^*}$. For this discussion, the curious reader may refer to [4] Chatper 1.3.

Now, we are ready to introduce the concept of measurable functions, which will be essential in how we define random variables on probability spaces.

Definition 1.11. Let $(X, \mathcal{A})$ and $(Y, \mathcal{B})$ be measurable spaces. A function $f : Y \to X$ is called *measurable* if for any $A \in \mathcal{A}, \{y \in Y : y \in f(A)\} \in \mathcal{B}$.

Let $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty, \infty\}$. Then, consider the measurable spaces $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d)), (\bar{\mathbb{R}}, \mathcal{B}(\bar{\mathbb{R}}))$, and a function $f : \mathbb{R}^d \to \bar{\mathbb{R}}$. (Notice that the only difference between $\mathcal{B}(\mathbb{R})$ and $\mathcal{B}(\bar{\mathbb{R}})$ is that intervals of the form $(a, \infty]$ and $[-\infty, a)$ are also in $\mathcal{B}(\bar{\mathbb{R}})$. By this alteration, the expression $f(x) \leq t$ in the expression below refers to a measurable set in $\bar{\mathbb{R}}$.) With respect to the definition above, we may express the measurability condition on f as,

$$\forall t \in \mathbb{R}, \{x \in \mathbb{R}^d : f(x) \leq t\} \in \mathcal{B}(\mathbb{R}^d) \tag{1.1}$$

Expectedly, such functions are called *Borel measurable*. Note that we may also have considered the preimage under $f(x) \geq t$, which is equivalent to our expression since σ-algebras are closed under complements. For the rest of the text, we will be mostly dealing with Borel measurable functions, hence, from now on we will use the term measurable functions to refer to Borel measurable functions, except where explicitly stated.

In the context of probability spaces, the definition of measurable functions underlies the definition of random variables.

Definition 1.12. A *random variable* X is an $\mathbb{R}$-valued measurable function on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, $X : \Omega \to \mathbb{R}$, such that for every Borel set $B \subseteq \mathbb{R}$,

$$X^{-1}(B) = \{\omega \in \Omega : X(\omega) \in B\} \in \mathcal{F}.$$

An $\mathbb{R}$-valued random variable assigns each event a Borel subset on the real line. Here, we may interpret, for some $F \in \mathcal{F}$ in the domain of X, $X(F) = B$ as the representation of the random event as a subset of the real line.

Definition 1.13. Let X be a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. The measure defined by $\mathbb{P}X^{-1} : \mathbb{R} \to \mathbb{R}$, $\mathbb{P}X^{-1}(A) = \mathbb{P}(X^{-1}(A))$ on $(\Omega, \mathcal{F})$ is called the *distribution (or law) of* X, and is denoted $\mathbb{P}X^{-1} = P_X$.

Next, we state two propositions that describe the measure-theoretic relation between cumulative distribution functions and distributions of a random variable. Note that a measure μ on $(X, \mathcal{A})$ is *finite* if $\mu(X) < \infty$.

Proposition 1.14. *Let μ be a finite measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, and let $F_\mu : \mathbb{R} \to \mathbb{R}$ be defined by $F_\mu(x) := \mu((-\infty, x])$. Then, F_μ is bounded, nondecreasing and right-continuous, and satisfies* $\lim_{x \to -\infty} F_\mu(x) = 0$

Proposition 1.15. *For each bounded, non-decreasing, right-continuous function F : $\mathbb{R} \to \mathbb{R}$ that satisfies $\lim_{x \to -\infty} F_\mu(x) = 0$, there is a unique finite measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that for all $x \in \mathbb{R}$, $F(x) = \mu((-\infty, x])$*

Proof of 1.14 and 1.15. See [4] □

By Propositions 1.15 and 1.14, for any random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$, the distribution $P_X : \mathbb{R} \to \mathbb{R}$ is a finite measure, since $\mathbb{P}(\Omega) = 1$, thus gives rise to the *cumulative distribution function* $F_X : \mathbb{R} \to$ of the random variable X, defined as,

$$F_X(x) := P_X((-\infty, x]) = \mathbb{P}(X \le x)$$

Now, we define the Lebesgue integral and what it means to calculate the Lebesgue integral of a measurable function with respect to that measure. The following concise treatment is from [5]. The details of the construction of the Lebesgue integral are omitted for brevity. For the full treatment, one may consult [4], Chatper 2.3.

In what follows, let $(X, \mathcal{A})$ be a measurable space, and μ some measure on $(X, \mathcal{A})$. A function $f : X \to \bar{\mathbb{R}}$ is *simple* if it attains only finitely many values. A simple function is $\mathcal{A}$-measurable if for all distinct the finite values $a_1, ..., a_n \in \bar{\mathbb{R}}$ f attains, the sets $A_i := \{x \in X : f(x) = a_i\}$ lie in $\mathcal{A}$. Notice that A_i's are disjoint by definition.

Let $s\mathcal{A}$ denote the collection of all simple $\mathcal{A}$-measurable $\bar{\mathbb{R}}$-valued functions, and $s\mathcal{A}_+$ all non-negative simple $\bar{\mathbb{R}}$-valued functions. Notice that f may be represented as $f = \sum_{i=1}^n a_i \chi_{A_i}$, where χ_{A_i} is the indicator function of A_i. Moreover, we have $m\mathcal{A}_+$, the collection of all non-negative $[0, \infty]$-valued $\mathcal{A}$-measurable functions, and $m\mathcal{A}_+$, the collection of all $\bar{\mathbb{R}}$-valued $\mathcal{A}$-measurable functions.

Definition 1.16 ([5])**.** Let $(X, \mathcal{A}, \mu)$ be a measure space, and define $\int f d\mu$ for $f \in m\mathcal{A}$ by the following line of construction.

(i) First, define $\mu_0(\chi_A) := \mu(A)$ for $A \in \mathcal{A}$ notice that μ_0 is merely a placeholder for the integral restricted to simple functions.

(ii) Let $f \in s\mathcal{A}_+$ with representation $f = \sum_{i=1}^n a_i \chi_{A_i}, n \in \mathbb{N}, a_i \in [0, \infty]$ for all $i \in \{1...n\}$, and define

$$\mu_0'(f) := \sum_{i=1}^n a_i \mu_0(\chi_{A_i})$$

(iii) Let $f \in m\mathcal{A}_+, f : A \to [0, \infty]$ and define

$$\int f d\mu := \sup\{\mu_0'(g) : g \in s\mathcal{A}_+, \forall x \in A, g(x) \le f(x)\}$$

(iv) Let $f \in m\mathcal{A}, f : A \to \bar{\mathbb{R}}$ and for $x \in A$ define $f_+(x) := max\{0, f(x)\} \in m\mathcal{A}_+, f_-(x) := -min\{0, f(x)\} \in m\mathcal{A}_+$. Then, if either $\int f_+ d\mu < \infty, \int f - d\mu <$

∞, we define the *Lebesgue integral with respect to* μ as,

$$\int f d\mu := \int f_+ d\mu - \int f_- d\mu$$

If $|\int f d\mu| < \infty$, then we say f is μ*-integrable*.

Lebesgue integral allows us to define the integral of functions with respect to measures in the language of measure theory and generalizes the Riemann Integral for non-continuous functions. For our purposes in the study of probability measures, the Lebesgue integral will serve us to define the density and moments of random variables by measure theoretic tools and analyze the convergence of measures using the convergence of moments. First, we define the density of a random variable.

Definition 1.17. Let $f : \mathbb{R} \to \mathbb{R}$ be a non-negative Borel measurable function such that

$$\int f d\lambda = 1$$

Then,

$$P(A) := \int_A f d\lambda$$

defines a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Therein, we call f the *density* of the distribution P.

We will use the relation in Definition 1.17 when we define Wigner's Semicircle Law through its density.

Next, we have to define the expectation of a random variable under a given measure.

Definition 1.18. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a measure space, and $X : \Omega \to \mathbb{R}$ be a real-valued random variable. The *expectation* of X is defined as,

$$\mathbf{E}(X) = \int X d\mathbb{P}$$

In this setting, we define the *k-th moment* of X to be simply the expectation of the k-th power of X, i.e,

$$\mathbf{E}(X^k) = \int X^k d\mathbb{P}$$

For skeptical readers, the following proposition shows that powers of a random variable, hence the moments are indeed well-defined,

Proposition 1.19. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$*, and* $(X, \mathcal{A}), (Y, \mathcal{B})$ *be two measurable spaces. Let* $X : \Omega \to X$ *be a random variable,* $f : X \to Y$ *be a* $\mathcal{B}$*-measurable function. Then, the composition* $f(X) : \Omega \to Y$ *is* $\mathcal{B}$*-measurable, hence a random variable.*

Proof of Proposition 1.19, [5]. Let $B \in \mathcal{B}$. Since f is $\mathcal{B}$-measurable, $f^{-1}(B) \in \mathcal{A}$. Since X is $\mathcal{F}$-measurable, it follows that, for any $B \in \mathcal{B}$, $(f(X))^{-1}(B) = X^{-1}(f_{-1}(B)) \in \mathcal{F}$, implying that $f(X)$ is $\mathcal{F}$-measurable, hence a random variable. □

Notice that, in the context of $\mathbb{R}$-valued random variables on some $(\Omega, \mathcal{F}, \mathbb{P})$, Proposition 1.19 implies that any $\mathbb{R}$-valued Borel measurable function of an $\mathbb{R}$-valued random variable is again a random variable. Furthermore, by the topological characterization of continuity on $\mathbb{R}^d$ with the metric topology, we may recall that for any continuous function f on $\mathbb{R}^d$ and for any open subset $U \subseteq \mathbb{R}^d$, the preimage $f^{-1}(B)$ of B, is open and that $\mathcal{B}(\mathbb{R}^d)$ is generated by open sets in $\mathbb{R}^d$ to deduce that any continuous function is Borel measurable. (A detailed exposition of this argument in measure-theoretic terms may be found in [4] Proposition 2.1.9). Hence, Proposition 1.19 applies to polynomial functions on $\mathbb{R}^d$, implying that moments of $\mathbb{R}^d$-valued random variables are Borel measurable, hence random variables.

One of the central pieces of our proof of Wigner's Semicircle Law is the *method of moments*. To aid our understanding of the analytical argument, we now define some notions of converge of measurable functions and measures to see what we will be aiming for in the proof. We provide these notions first in purely measure-theoretic terms, so that we may better see the underlying mathematical structure while we are dealing with probabilistic objects and results.

We now define the notion of when a probability distribution is *determined by its moments*. The following formal expression of this notion as adapted from [2] with modifications. In relation to previous definitions, the reader may conceive the role of μ in the following statements analogous to $\mathbb{P}X^{-1} = P_X$, the distribution of a random variable X with under the probability measure $\mathbb{P}$.

Definition 1.20. A probability measure μ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ is determined by its moments if:

(i) $\int x^k d\mu(x) < \infty$ for all $k \in \mathbb{N}$
(ii) μ is unique, in the sense that if for some probability measure ν, $\int (x^k) d\mu(x) = \int (x^k) d\nu(x)$, then $\mu = \nu$.

Definition 1.21. Let μ be a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\alpha_k = \int x^k d\mu(x) < \infty$ for all $k \in \mathbb{N}$. If the power series $\sum_k \alpha_k \frac{r^k}{k!}$ has non-zero radius of convergence, then μ is determined by its moments.

Proof of 1.21. See [2], p. 388 □

Now, we are ready to define the notions of convergence relevant to this exposition.

1.2. **Convergence.**

Definition 1.22. Let f and $f_1, f_2, ...$ be $\bar{\mathbb{R}}$-valued functions on the measure space $(X, \mathcal{A}, \mu)$. f_n is said to converge to f *in measure* if, for any $\epsilon > 0$,

$$\lim_n \mu(\{x \in X : |f_n(x) - f(x)| > \epsilon\}) = 0$$

If the measure space $(X, \mathcal{A}, \mu)$ in Definition 1.22 is a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, then we call this notion convergence *in probability*. For the following, let $\mathcal{C}_b(\mathbb{R}^d)$ denote the set of bounded continuous functions on $\mathbb{R}^d$

Definition 1.23. Let μ be a probability measure, and $\{\mu_i\}$ be a sequence of probability measures on $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. $\{\mu_i\}$ converges to μ *in distribution*, or converges *weakly* if for all $f \in \mathcal{C}_b(\mathbb{R}^d)$,

$$\int f d\mu = \lim_n \int f d\mu_n$$

1.3. **Random Matrix.** Simply put, a random matrix is a matrix-valued random variable. In this treatment, however, we will limit ourselves to a specific type of random matrices called Wigner matrices.

Definition 1.24. Let $\{Y_i\}_{i\geq 1}$ and $\{Z_{ij}\}_{1\leq i<j}$ be two collections of real-valued, zero-mean, i.i.d. random variables. Furthermore, suppose that, for all i,j, $E[Z_{ij}^2]=1$ and for each $k\in\mathbb{N}$,

$$\max(E[Z_{ij}^k],E[Y_i^k])<\infty$$

Consider an $n\times n$ symmetric matrix M_n whose entries are given by:

$$\begin{cases} M_n(i,i)=Y_i \\ M_n(j,i)=Z_{ij}=M_n(i,j) \end{cases}$$

The matrix M_n is called a (real symmetric) *Wigner matrix*.

Here, we introduce the Gaussian Orthogonal Ensemble as a special type of Wigner matrices. Although we will not explicitly use the GOE in the following treatment, it belongs to a very important class of random matrix ensembles (a concept analogous to distributions) due to (i) the independence of its entries, and (ii) the rotational invariance of such matrices under rotation by normal operators. One may find the a discussion of the importance of these notions in the general context of random matrix theory in [8].

Example 1.25. *Gaussian Orthogonal Ensemble.* Let M_n be a Wigner matrix. With notation as before, if $Y_i\sim\mathcal{N}(0,2)$ and $Z_{ij}\sim\mathcal{N}(0,1)$, then the resulting random matrix distribution given the name *Gaussian Orthogonal Ensemble*

Definition 1.26. The *spectrum* of a matrix A, denoted as $\sigma(A)$, is the set of eigenvalues of A. In other words, $\sigma(A)=\{\lambda:\det(A-\lambda I)=0\}$, where I is the identity matrix.

The spectra of matrices are of both mathematical and physical importance in many areas, e.g. in determining principle components of physical systems or the time evolution of dynamical systems, and here they will serve to characterize the distributions of ensembles of random matrices.

For example, the joint pdf of elements of Gaussian Ensembles can be conveniently expressed in terms of the trace of a matrix, which renders working with properties of spectra an important tool in this context.

As random matrices M_n can be thought of as linear operators, we can talk about its operator norm, defined as

$$\|M_n\|_{\mathrm{op}}=\min\{c\geq 0:\|M_nv\|_2\leq c\|v\|_2,\ \text{for all } v\in\mathbb{R}^n\},$$

For reasons that are beyond our scope, the operator norm of a Wigner matrix of size n is typically $O(\sqrt{n})$. Thus, when we are concerned about limiting cases of various statistics about Wigner matrices, we will prefer to work with normalized matrices $X_n:=\frac{M_n}{\sqrt{n}}$.

Before presenting Wigner's Semicircle Law, we should define the empirical spectral density as a probability measure on $(\mathbb{R},\mathcal{B}(\mathbb{R}))$. The ESD expresses the probability of finding an eigenvalue $\lambda_i(X_n)$ in a given Borel set in $\mathbb{R}$. As we deal with real symmetric matrices, indeed, for all $i\in\{1...n\}$, $\lambda_i(X_n)\in\mathbb{R}$.

Definition 1.27. Let X_N be a normalized Wigner matrix. The empirical spectral distribution (ESD), $\mu_{X_n} : \mathbb{R} \to [0,1]$ is defined as,

$$\mu_{X_n} := \frac{1}{n}\sum_{j=1}^{n} \delta_{\lambda_j(X_n)},$$

To check that the ESD is indeed a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, fix some Wigner matrix X_n. First, notice that $\mu_{X_n}(\emptyset) = 0$, by definition. Moreover, let $\{A_k\}$ be a sequence of disjoint Borel sets in $\mathbb{R}$, and $A = \bigcup_k A_k$ Since for all $i \in \{1...n\}$, $\lambda_i(X_n)$ is in at most one of the sets in $\{A_k\}$, we must have that $\mu_{X_n}(A) = \sum_k \mu_{X_n}(A_k)$.

The justification for the argument for σ-additivity is as follows. For a Wigner Matrix X_n, the matrix is real symmetric, hence is diagonalizable by an orthogonal matrix in $\mathbb{R}^{n\times n}$. Hence, each eigenvalue $\lambda_i(X_n)$ has multiplicity 1, and since $\mathbb{R}$ is Hausdorff with the metric topology, we may find disjoint open intervals $A_i \subseteq \mathbb{R}$ such that, $\lambda_j(X_n) \in A_i$ if and only if $i = j$. So it is impossible that an eigenvalue is in two disjoint Borel sets.

By Proposition 1.14, we have the corresponding cumulative distribution function,

$$F_{\mu_{X_n}}(x) := \frac{1}{n}\sum_{j=1}^{n} \mathbb{1}_{\{\lambda_j(X_n)<x\}}$$

It is important to recognize the dual character of the ESD. The particular measure defined by the ESD depends on the matrix X_n, hence, as such, the ESD is in fact a *random* probability measure on the real line. In other words, the ESD is itself a random variable, with values in the space of probability measures $\mathbf{Pr}(\mathbb{R})$ on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$.

However, what we wish to show is that the ESD converges (weakly, in probability) to Wigner's Semicircle Law independent of the specific realization of X_n. To this end, our first step is to define the *expected* ESD, for which we require the Riesz Representation Theorem. Unfortunately, a comprehensive treatment of topics surrounding the Riesz Representation Theorem is out of our scope since we don't have the space to introduce the required functional analytic tools, but the curious reader may consult [4] Chapter 7, where the theorem is stated in Theorem 7.2.8. However, we state the theorem here for completeness. $\mathcal{C}_c(X)$ denotes the set of $\mathbb{R}$-valued, continuous and compactly supported functions on X.

Theorem 1.28. *Let X be a locally compact Hausdorff space, and let I be a positive linear functional on $\mathcal{C}_c(X)$. Then, there is a unique Borel measure μ on X such that,*

$$I(f) = \int f d\mu$$

Proof of Theorem 1.28. See [4] Theorem 7.2.8 □

The converse of Theorem 1.28 also holds in the sense that if μ is a Borel measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, then μ defines a bounded linear functional via the map $f \mapsto \int f d\mu$. Hence, Borel measures on the real line are in 1-1 correspondence with elements of the dual space $\mathcal{C}_c(X)^*$. From here on, one may consider the weak-* topology on $\mathcal{C}_c(X)^*$, and show that the notion of weak convergence of measures we have defined is equivalent to weak-* convergence of linear fucntionals in the Banach space $\mathcal{C}_c(X)^*$. The curious reader may consult [10] for a detailed view into this construction.

For our purposes, however, the expectation of the ESD will have a particularly simple form, through some elementary linear algebra, that we may analyze with the probabilistic definitions we have set up.

We are ready to define the deterministic ESD via this duality. Notice that the expectation $\mathbf{E}\mu_n$ is in $\mathbf{Pr}(\mathbb{R})$, and the expectation is with respect to the random matrix X_n that determines the measure. The rigorous treatment of this beyond our scope, again, the reader may consult [10] for a treatment in the context of Wigner's Semicircle Law, or [5] Section 3.2.2 where the author rigorously constructs a suitable metrizable topology under which one may define the convergence of measures in $\mathbf{Pr}(\mathbb{R})$.

Definition 1.29. The *deterministic ESD* is defined as,

$$\bar{\mu}_n := \mathbf{E}\mu_n = \mathbf{E}\frac{1}{n}\sum_{j=1}^{n}\delta_{\lambda_j(X_n)}$$

such that for any $f \in \mathcal{C}_c(\mathbb{R})$

$$\int f d\mathbf{E}\mu_n = \mathbf{E}\int f d\mu_n$$

For the last bit of notation, we will here on denote $\langle \mu, \phi \rangle := \int_{\mathbb{R}} \phi(x)\, d\mu(x)$ for any $\phi \in \mathcal{C}_c(\mathbb{R})$.

2. Wigner's Semicircle Law

Figure 1 below displays the intuition behind Wigner's Semicircle Law. The histograms were made by generating normalized Wigner Matrices of size N, as noted on the top of the figures and histogramming their eigenvalues with 50 bins, which is merely a crude representation of the ESD on the given matrix. The red graph is Wigner's Semicircle Law. Notice how nicely the distribution approaches the semicircle law as N grows. More importantly, these graphs have been generated by eigenvalues of only a single normalized Wigner Matrix of the given size. That we did not have to accumulate eigenvalues of a sample of Wigner Matrices to arrive at this illustration clearly emphasizes the intuition behind the fact that spectral distributions of samples of Wigner matrices converge to the semicircle distribution in the infinite-dimension limit regardless of the underlying matrix X_n, which underlies the strong universality of the result.

For intuition, one may contrast this behavior with the limiting behavior of the mean of i.i.d samples, which is formalized in the usual Central Limit Theorem for real-valued random variables. Notice how the role of a random sample of real-valued variables is analogous to that of a sample Wigner matrix, and how the role of the mean of the given sample is analogous to the eigenvalue distribution of a sample Wigner matrix. Further rigorous discussion of Wigner's Semicircle Law as the *free probability* analogue of the usual Central Limit Theorem may be found in [11].

Now we state Wigner's Semicircle Law, the main result of this text. Here, we abuse notation use the notation μ_n instead of μ_{X_n}, and have the dependence on the Wigner matrix X_n implicitly.

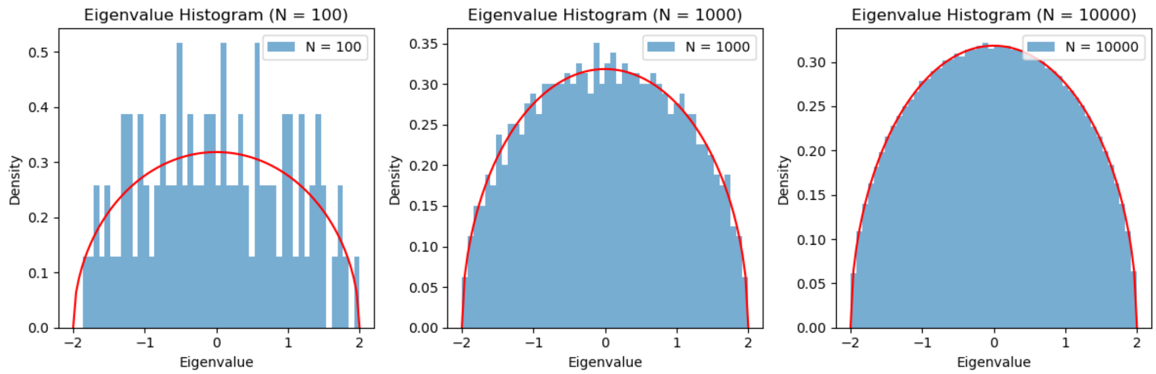

FIGURE 1. Histogram for eigenvalue distributions of Wigner matrices of different sizes

Wigner's Semicircle Law. *Let X_n be a Wigner matrix. Then $\mu_n \to \sigma$ weakly, in probability, where σ is the semicircle law $\sigma : \mathbb{R} \to \mathbb{R}$ is defined as,*

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

Note that, with respect to Definition 1.17, we have defined the semicircle distribution as a measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ through its density function $f(x) = \frac{1}{2\pi}\sqrt{(4-x^2)}\mathbb{1}_{|x|<2}$. In this statement, we followed the notation in [11]. To interpret this statement in the language we have developed, it would have been more accurate to represent this by the Lebesgue measure as,

$$\sigma(I) = \int_I \frac{1}{2\pi}\sqrt{(4-x^2)}\mathbb{1}_{|x|<2}d\lambda(x)$$

for $I \in \mathcal{B}(\mathbb{R})$. Nevertheless, no confusion will occur regarding this notation, as long as the reader refers to this explanation if any occurs.

From now on, σ will always refer to the Wigner semicircle law. Our proof of Wigner's Semicircle Law will depend on three main lemmas. We shall first prove them, and present the proof of the theorem afterwards. The proofs followed those given in [1], [7], [11], [9], with modifications and explanations given where it was deemed necessary given the scope of this paper. Although there exist purely analytical arguments in the literature that are also better suited for further topics in random matrices and free probability, the combinatorial proof at hand is very similar to the proof devised by Wigner in [9] for the special case of random matrices with Bernoulli distributed elements, and is a profound demonstration of Wigner's remarks in his 1960 paper titled *The Unreasonable Effectiveness of Mathematics in the Natural Sciences* that "the concepts of mathematics are not chosen for their conceptual simplicity - even sequences of pairs of numbers are far from being the simplest concepts - but for their amenability to clever manipulations and to striking, brilliant arguments". In this spirit, we begin the proof with the following lemma which clarifies the calculation of the moments of the semicircle law and serves as a guide for the rest of the argument.

Lemma 2.1. *Moments of $\sigma(dx)$ are given by,*

$$\langle \sigma, x^k \rangle = \begin{cases} 0 & \text{if } k \text{ is odd} \\ C_{k/2} & \text{if } k \text{ is even} \end{cases} \tag{2.1}$$

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

Proof of Lemma 2.1, [1], [10]. Let $k \in \mathbb{N}$ be odd. Indeed, by the symmetry of σ around $x = 0$,

$$\int_{-2}^{2} x^k \sigma(x) dx = 0$$

Now let $k \in \mathbb{N}$. Notice that the Catalan numbers satisfy the following recursion formula,

$$\frac{C_{k+1}}{C_k} = \frac{\frac{(2k+2)!}{(k+2)(k+1)!(k+1)!}}{\frac{(2k)!}{(k+1)(k)!(k)!}} = \frac{4k+2}{k+2}$$

Moreover, observe that,

$$C_0 = 1 = \int_{-2}^{2} \sigma(x) dx$$

where the second equality follows by simply considering the area of the semicircle with radius 2. Now we show that for any $k \in \mathbb{N}$, $\langle \sigma, x^{2k} \rangle = C_k$ by induction on k. As we have shown the base case $k = 0$ above, we assume,

$$\int_{-2}^{2} x^{2k} \sigma(x) dx = C_k$$

Observe that it suffices to show the recurrence relation between Catalan numbers for the moments.

First, we make the change of variables $x = 2sin(u), dx = 2cos(u)$, to express the integral in a convenient form, and get:

$$\begin{aligned} C_k &= \frac{1}{2\pi} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} 2^{2k+2} sin^{2k}(u) cos^2(u) du \\ &= \frac{2^{2k+2}}{2\pi} \left(\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} sin^{2k}(u) du - \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} sin^{2k+2}(u) du \right) \\ &= \frac{2^{2k+2}}{2\pi} \left(\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} sin^{2k}(u) du - \frac{2k+1}{2k+2} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} sin^{2k}(u) du \right) \\ &= \frac{2^{2k+2}}{2\pi} \frac{1}{2k+2} \left(\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} sin^{2k}(u) du \right) \end{aligned}$$

The third inequality follows from,

$$\begin{aligned}\int sin^n(x)dx &= -sin^{n-1}(x)cos(x) + \int (n-1)sin^{n-2}(x)cos^2(x)dx \\ &= -sin^{n-1}(x)cos(x) + \int (n-1)sin^{n-2}(x)cos^2(x)dx \\ &= -sin^{n-1}(x)cos(x) + \int (n-1)sin^{n-2}(x)dx - \int (n-1)sin^n(x)dx\end{aligned}$$

by rearranging the terms, where in the third equality we have used the same change of variables as above.

Now we evaluate $\langle \sigma, x^{2(k+1)} \rangle$. To avoid repetition, we assume that the abovementioned formula and change of variables have been applied to the integral.

$$\begin{aligned}\langle \sigma, x^{2(k+1)} \rangle &= \frac{2^{2k+2}}{2\pi} \left(\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} sin^{2k+2}(u)du - \frac{2k+3}{2k+4} \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} sin^{2k+2}(u)du \right) \\ &= \frac{2^{2k+2}}{2\pi} \frac{1}{2k+4} \left(\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} sin^{2k+2}(u)du \right) \\ &= \frac{2^{2k+2}}{2\pi} \frac{2k+1}{(2k+4)(2k+2)} \left(\int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} sin^{2k}(u)du \right) \\ &= \frac{4k+2}{k+2} C_k \\ &= C_{k+1}\end{aligned}$$

□

The following lemma shows that all moments of the deterministic ESD converge to the semicircle distribution thereof.

Lemma 2.2. *For any positive* $k \in \mathbb{N}$, $\lim_{n\to\infty} \langle \bar{\mu_n}, x^k \rangle = \langle \sigma, x^k \rangle$

Before proving this lemma, we present some observations and define the combinatorial tools necessary to arrive at the conclusion. For convenience, we largely adopt the "ad hoc" language developed in [7].

First, we observe an important identity. Notice that, for a Wigner Matrix X_n, the matrix is real symmetric, hence is diagonalizable by an orthogonal matrix in $\mathbb{R}^{n\times n}$. Hence, each eigenvalue $\lambda_i(X_n)$ has multiplicity 1. Since singletons are Lebesgue measurable, let $A_i = \{\lambda_i(X_n)\}$. Then $\lambda_j(X_n) \in A_i$ if and only if $i = j$, and for all $i \in \{1...n\}$, $\mu_n(A_i) = \frac{1}{n}$ Consider the map $X : x \mapsto x^k \chi_{A_i}$. Notice that X is a random variable on $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu_n)$ by the $\mathcal{B}(\mathbb{R})$-measurability of the indicator function and the polynomial, and X is simple for any $k, n \in \mathbb{N}$. X here is the random variable taking on the values k^{th} powers of eigenvalues of X_n. Therefore, we may express the Lebesgue integral of the map with respect to the measure μ_n as,

$$\int_{\mathbb{R}} x^k \chi_{A_i} d\mu_n = \sum_{i=1}^{n} \lambda_i^k(X_n)\mu_n(A_i) = \frac{1}{n} \sum_{i=1}^{n} \lambda_i^k(X_n) = \frac{1}{n} \operatorname{Tr}(X_n^k)$$

where the last equality follows again from the diagonalizability of X_n. From here, one may use the line of defintions in Definition 1.16, and conclude that

$$\langle \mu_n, x^k \rangle = \frac{1}{n} \left[\operatorname{Tr} X_n^k \right] \tag{2.2}$$

Furthermore, by using the duality in Definition 1.29, one may express

$$\langle \bar{\mu_n}, x^k \rangle = \frac{1}{n} \mathbf{E} \left[\operatorname{Tr} X_n^k \right] \tag{2.3}$$

$$= \frac{1}{n} \sum_{i_1 \ldots i_k = 1}^{n} \mathbf{E} \left[X_n(i_1, i_2) \ldots X_n(i_k, i_1) \right] \tag{2.4}$$

where in the last equality is by [9], where $X_n(i_i, i_j)$ denotes the ij^{th} element of X_n.

Now let $\mathbf{i} = i_1 i_2 \ldots i_k i_1$ for $i_j \in \{1 \ldots n\}$. It is clear that the set of such sequence are in 1-1 correspondence with the terms in the sum in 2, and defines a *closed*, i.e starts and ends on the same vertex, *connected*, i.e there exists a path from i_j, i_k for all $i_j, i_k \in \{1 \ldots n\}$, *path* on the set of vertices $\{i_1 i_2 \ldots i_k\}$ where edges are represented by consecutive indices. As such, the set of such sequences are 1-1 correspondence also with paths of length k on the set of vertices $\{i_1 i_2 \ldots i_k\}$. It is important to observe that, since elements of X_n are defined to be zero-mean and i.i.d, all summands that correspond to a sequence $\mathbf{i}$ with any edge traversed only once will equal zero. Hence, for summands that are non-zero, there must be at most $k/2$ distinct edges, thus at most $k/2 + 1$ distinct vertices.

Next, we define the *weight* $w_{\mathbf{i}}$ of the sequence $\mathbf{i}$ to be the number of distinct vertices in $\mathbf{i}$. Hence, the abovementioned fact may be expressed as, a summand in 2 is non-zero if and only if $w_{\mathbf{i}} \leq k/2 + 1$. Moreover, let us say two sequences $\mathbf{i}$ and $\mathbf{j}$ are *equivalent* if there exists a bijection on $\{1 \ldots n\}$ that maps $i_l \to j_l$ for all $l \in \{1 \ldots n\}$. As such, equivalent sequences must have the same weight because the bijection preserves the structure of the edges. Recalling that all elements are i.i.d, this leads us to the conclusion that summands with equivalent sequences $\mathbf{i}$ also have the value. Our last observation in this regard is to say that, under this equivalence, the number of equivalence classes depends on k, since every class may be represented by a sequence $\{i_1 i_2 \ldots i_k\}$ such that $i_j \in \{1 \ldots k\}$ for all $j \in \{1 \ldots k\}$.

Now we are ready to tackle the proof of Lemma 2.2.

Proof of 2.2, [1], [7], [9]. Notice that, for any $\mathbf{i} = i_1 i_2 \ldots i_k i_1$ with $w_{\mathbf{i}} = r$, since the equivalence relation we defined preserves the structure of the edges, we have ${}^nP_k = n(n-1) \ldots (n-r+1) \leq n^r$ in the equivalence class of $\mathbf{i}$. Since the matrix X_n is normalized with $1/\sqrt{n}$, and since all elements have bounded moments of all orders, we must have that, for any sequence equivalent to $\mathbf{i}$,

$$\frac{1}{n} \sum_{i_1 \ldots i_k = 1}^{n} \mathbf{E} \left[X_n(i_1, i_2) \ldots X_n(i_k, i_1) \right] = \mathcal{O}\left(\frac{1}{n} \frac{1}{\sqrt{n}^k}\right)$$

which means to express that the sum on the right-hand side is bounded by $C \frac{1}{\sqrt{n}^k}$, where $C \in \mathbb{R}$ is some constant. Consequently, for any equivalence class with weight

$r < k/2+1$ the total contribution of all elements of the equivalence class is of order $\mathcal{O}(n^r \frac{1}{n}\frac{1}{\sqrt{n}^k})$. Since the number of equivalence classes do not depend on n as we have shown, and since $lim_{n\to\infty} n^r \frac{1}{n}\frac{1}{\sqrt{n}^k} = 0$, all terms whose sequence of indices belong to an equivalence class with weight strictly smaller than $k/2+1$ vanish in the limit $n \to \infty$.

Next, we look at equivalence class of sequences with weight $r = k/2+1$. Notice that, as was hinted in the Lemma 2.1, this case is impossible when k is odd, hence, for all odd k, the sum vanishes.

For k even, the condition $w_{\mathbf{i}} = k/2+1$ implies that all such $\mathbf{i}$ are connected paths on $k/2+1$ vertices k distinct edges, and thus contains no cycles. Such a graph structure is called a *tree*. Notice that, such $\mathbf{i}$ must describe a path on the tree that traverses each edge twice, each in opposite directions. As such, no consequent elements of the sequence are the same, hence all matrix elements in the corresponding summand are off-diagonal, and the summand involves $k/4$ pairs of symmetric elements, and the multiplication of each pair is the variance of the corresponding normalized real valued random variable $Z_{ij}/\sqrt(n)$, for which $\mathbf{E}[(Z_{ij}/\sqrt(n))^2] = 1/n$. Combining all these observations, we deduce that, for any $\mathbf{i} = i_1 i_2 ... i_k i_1$ with $w_{\mathbf{i}} = k/2+1$,

$$\frac{1}{n}\sum_{i_1...i_k=1}^{n} \mathbf{E}\left[X_n(i_1,i_2)...X_n(i_k,i_1)\right] = \frac{1}{n}\frac{1}{\sqrt{n}^k}$$

Now we find the number of such sequences. Such sequences are called *non-crossing* (One may imagine the tree structure with vertices corresponding to evenly spaced points on a circle plus the centre, and deduce why they are called so). Moreover, call an edge in such a sequence *free* if it is traversed the first time and *repetitive* if it is traversed for the second time. In this sense, we define the *type sequence* of a non-crossing path $\mathbf{i}$ of length k to be the sequence $\mathbf{t} = t_1...t_k t_1$ such that $t_j = f - r$, where $f = \#$ of free steps up to t_j and $r = \#$ of free steps up to t_j. As such, a type sequence always starts with 1 and ends with 0, with consecutive terms differing by 1.

More importantly, type sequences are conserved among equivalence classes of non-crossing sequences, since the edge structure is conserved, thus there are ${}^nP_k = \mathcal{O}(n^{k/2+1})$ many sequences with the same type sequence. Putting all these observations together, and letting $\mathcal{P}_{k,k/2+1}$ be the set of equivalence classes of paths with length k and weight k/2 + 1, and $\mathcal{T}_k$ be the set of type sequences of representatives of each equivalence class in $\mathcal{P}_{k,k/2+1}$

$$\langle \bar{\mu_n}, x^k \rangle = |\mathcal{P}_{k,k/2+1}| = |\mathcal{T}_k|$$

For the final step, observe that any type sequence in $\mathcal{T}_{k,k/2+1}$ has length k/2. Moreover, for any type sequence in $\mathcal{T}_{k,k/2+1}$ with no 0 occuring before the last term, we may simply remove the first and the last element and subtract 1 from the remaining elements to arrive at a type sequence in $\mathcal{T}_{k-2}$, The converse is also true with the inverse manipulations. Hence, letting $\mathcal{T}_k^j$ denote the set of type sequences of length k with first 0 appearing at the j-th element, we get that $|\mathcal{T}_k^k| = |\mathcal{T}_{k-2}|$. Generalizing this observation by considering the possible positions of the first 0 in a type sequence of length k, one may realize that the right-hand and left-hand sides of the first 0 in the type sequence of length k are also type sequences, and conclude that $|\mathcal{T}_k^j| = |\mathcal{T}_{k-j}||\mathcal{T}_{j-1}|$, and that the following formula holds,

$$|\mathcal{T}_k| = \sum_{i=1}^{k} |\mathcal{T}_{k-i}||\mathcal{T}_{i-1}| \tag{2.5}$$

Setting $|\mathcal{T}_0| := 1$, we observe that this is exactly the reccurence relation satisfied by the Catalan numbers (one may use a generating fucntion argument to arrive at this fact). Hence, the conclusion follows. $\square$

In the next and final lemma, we show that all moments of the ESD converge to the moments of the deterministic ESD in probability. Note that the notation is consistent with the notation for ESD and deterministic ESD as defined in the end of the Preliminaries section.

Lemma 2.3. *For any positive $k \in \mathbb{N}$ and any $\epsilon \in \mathbb{R}, \epsilon > 0$,.*

$$\lim_{n\to\infty} \mathbb{P}(|\langle \mu_n, x^k\rangle - \langle \bar{\mu_n}, x^k\rangle| < \epsilon) = 0$$

Before we move on with the proof, we present two important inequalities that we will utilize in the following sections.

Theorem 2.4 (*Markov's Inequality*)**.** *Let X be a non-negative $\mathbb{R}$-valued random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, $\epsilon > 0$, and $k \in \mathbb{N}$. Then,*

$$\mathbb{P}\left[|X| \geq \epsilon\right] \leq \frac{1}{\epsilon^k}\mathbf{E}\left[|X|^k\right]$$

Proof of 2.4. See [2], p. 80. $\square$

As a corollary, we have the following:

Corollary 2.5 (*Chebyshev's Inequality*)**.** *Let X be a non-negative $\mathbb{R}$-valued random variable on $(\Omega, \mathcal{F}, \mathbb{P})$, $\epsilon > 0$, and $k \in \mathbb{N}$. Then,*

$$\mathbb{P}\left[|X - \mathbf{E}[X]| \geq \epsilon\right] \leq \frac{1}{\epsilon^2}\mathbf{E}\left[|X - \mathbf{E}[X]|^2\right] = \frac{1}{\epsilon^2}\left[\mathbf{E}\left[X^2\right] - \mathbf{E}[X]^2\right] = \frac{1}{\epsilon^2}Var(X)$$

Proof of Lemma 2.3, [1], [7]. By Chebyshev's Inequality, we have,

$$\mathbb{P}(|\langle \mu_n, x^k\rangle - \langle \bar{\mu_n}, x^k\rangle| < \epsilon) \leq \frac{1}{\epsilon^2}|\mathbf{E}(\left[\langle \mu_n, x^k\rangle^2\right] - (\langle \mu_n, x^k\rangle)^2|$$

So it suffices to show that,

$$\lim_{n\to\infty} |\mathbf{E}(\left[\langle \mu_n, x^k\rangle^2\right] - (\langle \mu_n, x^k\rangle)^2| = 0$$

Again, we rewrite the moment using the diagonalizability of X_n, using the same notation as in the proof of 2.2,

$$\begin{aligned} |\mathbf{E}(\left[\langle \mu_n, x^k\rangle^2\right] - (\langle \mu_n, x^k\rangle)^2| &= \frac{1}{n^2}\left(\mathbf{E}[\operatorname{Tr} X_n^k]^2 - \left(\mathbf{E}\left[\operatorname{Tr} X_n^k\right]\right)^2\right) \\ &= \frac{1}{n^2}\sum_{\mathbf{i},\mathbf{i}'}\left(\mathbf{E}[X_n(\mathbf{i})X_n(\mathbf{i}')] - \mathbf{E}\left[X_n(\mathbf{i})\right]\mathbf{E}\left[X_n(\mathbf{i}')\right]\right) \end{aligned}$$

where $X_n(\mathbf{i}) := X_n(i_1, i_2)...X_n(i_k, i_1)$. Similar to the argument in the proof of 2.2, we consider the graph generated by $\mathbf{i}, \mathbf{i}'$), giving rise to the vertex set $V_{\mathbf{i},\mathbf{i}'} = \{i_1, i_2...i_k\} \cup$

$\{i'_1, i'_2...i'_k\}$ and likewise the edge set $E_{\mathbf{i},\mathbf{i}'} = \{i_1i_2, i_2i_3...i_{k-1}i_k\} \cup \{i'_1i'_2, i'_2i'_3...i'_{k-1}i'_k\}$. In this setting, we define the *weight* of $(\mathbf{i}, \mathbf{i}')$ as $|V_{\mathbf{i},\mathbf{i}'}|$. Moreover, we define $(\mathbf{i}, \mathbf{i}'), (\mathbf{j}, \mathbf{j}')$ to be *equivalent* if there is a bijection on $\{1...n\}$ that maps elements with the same subindices. Moreover, again, refering to the independence and bounded moments of the elements of the normalized matrix X_n, we see that summands with equivalent sequences are equal, and if the summand corresponding to $(\mathbf{i}, \mathbf{i}')$ is non-zero, then $E_{\mathbf{i},\mathbf{i}'}$ must have each edge at least twice. However, the converse is true only if $\mathbf{i}$ and $\mathbf{i}'$ have some edge in common. Otherwise, independence of the matrix elements implies that the summand is zero. Let us call $\mathbf{i}$ and $\mathbf{i}'$ *non-zero* if it satisfies these two conditions.

Suppose $(\mathbf{i}, \mathbf{i}')$ is such that $|V_{\mathbf{i},\mathbf{i}'}| \leq k + 1$. Then we must have ${}^nP_k \leq n^{k+1}$ pairs equivalent to $(\mathbf{i}, \mathbf{i}')$, and each such summand is $\mathcal{O}(\frac{1}{n^{k+1}\frac{1}{n}})$. Hence, the contribution of each equivalence class with representative $(\mathbf{i}, \mathbf{i}')$ such that $|V_{\mathbf{i},\mathbf{i}'}| \leq k+1$ is $\mathcal{O}(\frac{1}{n})$. Since the number of equivalence classes depend on k, we deduce that the contribution of all summands that belong to some equivalence class with $|V_{\mathbf{i},\mathbf{i}'}| \leq k + 1$ vanishes at the limit $n \to \infty$.

Moreover, notice that each non-zero pair $(\mathbf{i}, \mathbf{i}')$ must generate a connected graph with at most k unique edges, for if not, then the corresponding summand is zero due to independence of matrix elements. If $|V_{\mathbf{i},\mathbf{i}'}| > k+1$, then $|E_{\mathbf{i},\mathbf{i}'}| > k$ by the connectedness of the generated graph, hence, all such sequences produce zero summands.

Now we tackle the case $|V_{\mathbf{i},\mathbf{i}'}| = k + 1$. A similar reasoning as before shows that the graph generatd by $(\mathbf{i}, \mathbf{i}')$ is a tree, and each edge is traversed twice in each direction. From here, we observe that each such edge couple must reside in either $\mathbf{i}$ or $\mathbf{i}'$, hence $\mathbf{i}$ and $\mathbf{i}'$ must have disjoint edge sets, implying that the corresponding summand is zero by independence.

Hence, the only non-zero contributions in the sum are from pairs $(\mathbf{i}, \mathbf{i}')$ such that $|V_{\mathbf{i},\mathbf{i}'}| \leq k + 1$, which is $\mathcal{O}(\frac{1}{n})$ in total. The result follows. ☐

Now, we have all the tools necessary to proceed with the proof of Wigner's Semicircle Law.

Proof of Wigner's Semicircle Law [1], [7]. We wish to show that, for any $f \in \mathcal{C}_b(\mathbb{R})$,

$$\lim_{n\to\infty} \mathbb{P}\left(|\langle \mu_n, x^k\rangle - \langle \sigma, x^k\rangle| > \epsilon\right) = 0. \tag{2.6}$$

The proof will utilize the Weierstrass Approximation Theorem to express $f \in \mathcal{C}_b(\mathbb{R})$ in terms of a polynomial, upon which we may apply our results about the convergence of the moments and arrive at the result. To this end, we first show that we may assume f to be compactly supported on a closed interval around zero. We do this by showing that any moment of the ESD is supported around a compact interval around zero. Let $B \in \mathbb{R}$. Applying Chebyshev's Inequality,

$$\mathbb{P}\left(\langle \mu_n, |x|^k \mathbb{1}_{|x|>B}\rangle > \epsilon\right) \leq \frac{1}{\epsilon}\mathbf{E}\left[\langle \mu_n, |x|^k \mathbb{1}_{|x|>B}\rangle\right] \leq \frac{\langle \bar{\mu_n}, |x|^k \mathbb{1}_{|x|>B}\rangle}{\epsilon B^k}$$

By Lemma 2.2, we look at $\limsup$ of the left hand side as $n \to \infty$. Since the limit in Lemma 2.2 is deterministic, using $\limsup$ instead of the limit just as well allows us to replace the moment of the deterministic ESD with the moment of σ and introducing $\frac{x^k}{B^K}$ inside the integral:

$$\limsup_{n\to\infty} \mathbb{P}\left(\langle \mu_n, |x|^k \mathbb{1}_{|x|>B}\rangle > \epsilon\right) \leq \frac{\langle \sigma, |x|^{2k}\mathbb{1}_{|x|>B}\rangle}{\epsilon B^k} \leq \frac{4^k}{\epsilon B^k} \tag{2.7}$$

where the last inequality follows from the easily verifiable fact that $C_k \leq 4^k$ for the k^{th} Catalan number C_k. Thus, for any $B > 4$, the right hand side is decreasing in k, whereas the left-hand side is bounded below by 0 and is nondecreasing in k for $|x| > B > 4$:

$$\limsup_{n\to\infty} \mathbb{P}\left(\langle \mu_n, |x|^k \mathbb{1}_{|x|>B}\rangle > \epsilon\right) = 0 \tag{2.8}$$

Hence, we may suppose that f is supported on $[-B, B]$ for $B > 4$.

Consequently, we employ the Weierstrass Approximation Theorem for some continuous f, let $\delta > 0$ and let $p_\delta(x) = \sum_{i=0}^{\infty} a_i x^i$ be a polynomial such that

$$sup_{\{x:|x|<B\}} |p_\delta(x) - f(x)| \leq \frac{\delta}{8} \tag{2.9}$$

By applying the Triangle Inequality and linearity of the Lebesgue integral:

$$\begin{aligned}
|\langle \mu_n, f\rangle - \langle \sigma, f\rangle| &\leq |\langle \mu_n, f - p_\delta\rangle - \langle \sigma, f - p_\delta\rangle| + |\langle \mu_n, p_\delta\rangle - \langle \sigma, p_\delta\rangle| \\
&\leq |\langle \mu_n, (f - p_\delta)\mathbb{1}_{|x|\leq 5}\rangle| + |\langle \sigma, (p_\delta - f)\mathbb{1}_{|x|>5}\rangle| \\
&\quad + |\langle \mu_n, p_\delta \mathbb{1}_{|x|>5}\rangle| + |\langle \mu_n, p_\delta\rangle - \langle \sigma, p_\delta\rangle| \\
&\leq \frac{\delta}{4} + |\langle \mu_n, p_\delta \mathbb{1}_{|x|>5}\rangle| + |\langle \mu_n, p_\delta\rangle - \langle \sigma, p_\delta\rangle|
\end{aligned}$$

Applying the triangle inequality and by the monotonicity of the measure $\mathbb{P}$,

$$\begin{aligned}
\mathbb{P}(|\langle \mu_n, f\rangle - \langle \sigma, f\rangle| > \delta) \leq\ & \mathbb{P}(|\langle \mu_n, p_\delta 1_{|x|>5}\rangle| > 3\delta/4) \\
&+ \mathbb{P}(|\langle \bar{\mu}_n, p_\delta\rangle - \langle \sigma, p_\delta\rangle| > 3\delta/4) \\
&+ \mathbb{P}(|\langle \bar{\mu}_n, p_\delta\rangle - \langle \mu_n, p_\delta\rangle| > 3\delta/4)
\end{aligned}$$

By 2.9, the first term vanishes as $n \to \infty$. By Lemma 2.2, the second term vanishes as $n \to \infty$. By Lemma 2.3, the third term vanishes as $n \to \infty$. Hence, 2.6 holds. Therefore, we conclude $\mu_n \to \sigma$ weakly, in probability. □

3. Final Remarks and Further Directions

It is curious how Wigner's Semicircle Law is related to the Central Limit Theorem for i.i.d distributed real-valued random variables. To better understand the subtleties in the analogy between the two results, the reader should see the proof of Wigner's Semicircle Law by the Stieltjes transform, which relates to probability measures in a similar way the characteristic function defned via Fourier transform relates to real-valued random variables. Furthermore, the cognizant reader should have picked up some of the gaps in our narrative that have been made to limit the scope. The most curious of these is the definiton of weak convergnece in the language of Banach spaces, and preliminary result that would facilitate a proof of the Semicircle Law through such tools. For a discussion of this topic, reader should consult [11], [5].

Although physical applications of random matrixt theory usually require stronger reuslts, Wignee's Semicircle Law is a great soruce of intuition in order to understand such more advanced results. For a friendly yet technical introduction to some famous results and concepts in random matrix theory that pertains to physics, the reader may consult [8]. Similarly, [6] takes a numerical analysis point-of-view with describing important applications. For applciations of tools of random matrix theory in financial mathematics, one should see [3].

References

[1] G.W Anderson, A. Guionnet, O. Zeitouni, *An Introduction to Random Matrix Theory*, Cambridge University Press, 2011.
[2] P. Billingsly, *Probability and Measure*, John Wiley and Sons, New York, NY, 1995.
[3] J.P. Bouchaud, M. Potters Financial Applications of Random Matrix Theory: a short review, *arXiv Preprint*, **79**(2009)
[4] D. Cohn, *Measure Theory*, Theta, Bucharest 2004.
[5] A. Dembo Probability Theory STAT310/MATH230 Lecture Notes, Stanford University, 2021.
[6] A. Edelman ,N.R Rao Random Matrix Theory *Acta Numerica*, (2005), 1-65.
[7] A. R. Feier, *Methods of Proof in Random Matrix Theory*, BSc Thesis, Harvard University, Cambridge, MA, 2012.
[8] G. Livan, M. Novaes, P. Vivo Introduction to Random Matrices Theory and Practice, *arXiv Preprint*, (2007)
[9] E.P. Wigner On the Distribution of the Roots of Certain Symmetric Matrices,*Annals of Mathematics*, **62**(3)(2009), pp. 548-564.
[10] N. Sweeney and X. Zhang, REU Paper, University of Chicago, 2022, (https://math.uchicago.edu/ may/REU2022/REUPapers/Sweeney.pdf)
[11] T. Tao *Topics in Random Matrix Theory*, AMS, 2012.

Department of Mathematics, Bilkent University, 06800 Bilkent, Ankara, Turkey
Email address: ozan.baykan@ug.bilkent.edu.tr