REU PAPER

NICK SWEENEY
XINYI ZHANG

Abstract. Wigner's semicircle law captures a striking phenomenon in random matrix theory: The eigenvalues of an $N \times N$ real symmetric or Hermitian random matrix accumulate about the origin (in a probabilistic context) in a semicircular fashion when N is large. In this paper, we state multiple versions of Wigner's semicircle law and prove them using both analytic and combinatorial approaches. We contribute proofs of fundamental results to clarify equivalence between alternative characterizations of the convergence as $N \to \infty$ (and equivalence in some definitions involved).

Contents

Date: August 2022.

1. Introduction

An intriguing feature in probability theory is universality, where the underlying structure of random variables becomes irrelevant when the sources of randomness adhere to sufficient conditions. The well-known Central Limit Theorem is a perfect example, where a distribution of increasingly large random samples converges to the Gaussian distribution regardless of the original distribution of the sample. In this paper, we study an equally astounding universality result in random matrix theory known as Wigner's semicircle law.

To formalize the theorem, one defines a sequence of random measures ρ_N that encode the eigenvalues of each $N \times N$ Wigner random matrix, such that as $N \to \infty$, these random measures converge to a measure whose density takes the shape of a radius-2 semicircle. In our paper, we will first set up all of the required machinery, so as to establish what the space of these measures will "look" like. Once that has been done, we will be ready to prove (slightly different versions of) Wigner's semicircle law in two distinct ways: one mainly using methods from mathematical analysis, and the other finding utility in combinatorics.

In the analytic proof of Wigner's semicircle law, we establish that convergence in probability of a sequence of random probability measures of $\mathbb{R}$ can be deduced from convergence in probability of their Stieltjes transforms given any fixed input variable z in the complex upper half-plane. By use of a Lipschitz continuity property for the Stieltjes transforms of ρ_N, we establish furthermore that to prove Wigner's semicircle law, it is sufficient to look only at the expectations of the Stieltjes transforms of ρ_N. By showing that the limit of these expectations is given by the same quadratic polynomial in z as is the Stieltjes transform of μ_W, we complete a proof of Wigner's semicircle law. For this analytic proof, we closely follow the 2012 notes for Fraydoun Rezakhanlou's Lectures on Random Matrices [6], with additional guidance from Terence Tao's *Topics in Random Matrix Theory* [9], although we provide original proofs of foundational results throughout the setup for the theorem.

For the combinatorial approach, we examine the expectation of the trace in tandem with combinatorial objects such as the non-crossing pairings, and we discover that both possess the same inherent recursive structures. We use this connection to prove the convergence of moments in expectation. For this part of the proof, we largely follow the lecture notes by Roland Speicher [8]. Then, we use measure theory to show that given certain constraints, convergence in moments implies convergence in measures. Eventually, we go back to graph theory to prove the almost sure convergence of the random empirical measures of eigenvalues to the semicircular measure.

In this paper, we presume some basic knowledge of probability, combinatorics, real analysis, and complex analysis. We will also re-emphasize some of the important definitions and propositions in these fields relevant to our discussion along the way.

2. Wigner's Semicircle Law: Setting Up the Theorem

In our text, we assume a basic understanding of measure theory and of measure-theoretic probability theory. (The reader who lacks such background may find [1] and [2] helpful.) All definitions in this section and the next are credited to Terence Tao's *Topics in Random Matrix Theory* [9] unless otherwise specified. All proofs in this section are original.

We will begin our analysis of Wigner's semicircle law by introducing the main objects of the theorem. Wigner's semicircle law will describe a convergence in probability of random measures, which calls for consideration of random variables whose values are not numbers but instead probability measures of the real line. To each outcome in the sample space of an *i.i.d.* (real symmetric or Hermitian) random matrix, we will assign a probability measure on $\mathbb{R}$ that marks the locations of its (real) eigenvalues, which under the right conditions will converge in probability as $N \to \infty$ to a measure whose density (with respect to Lebesgue measure) makes a semicircle above the x-axis.

The goal of this section is to make precise the definitions that are needed for the theorem. We will denote $\mathbf{Pr}(\mathbb{R})$ to be the set of regular probability measures on $\mathbb{R}$ with the Borel σ-algebra. The "regularity" condition means that $\forall E \subset \mathbb{R}$ measurable,

$$\mu(E) = \inf\{\mu(V) : E \subset V, V \text{ is open}\} = \sup\{\mu(K) : K \subset E, K \text{ is compact}\}$$

We will only be concerned with such well-behaved measures. To handle these measures as random objects and talk about their convergence in probability, we will require machinery from functional analysis, a suitable characterization of $\mathbf{Pr}(\mathbb{R})$ as a metric space, and an ability to generalize the typical definition of convergence in probability to this space in a well-defined manner.

We state the following definition of a random variable:

Definition 2.1. A *random variable* X is a measurable function $X : \Omega \to M$ from a sample space Ω (i.e. a measure space with a probability measure) to a σ-compact metric space M with the Borel σ-algebra.

By *σ-compact*, we mean that M is a countable union of compact sets, i.e. there are sets $K_n \subset M$ such that K_n is compact $\forall n$, and $\bigcup_{n=1}^{\infty} K_n = M$.

For a random variable $X : \Omega \to M$, we say that X "attains values" in M. Wigner's semicircle law will concern us with convergence in probability of random variables that attain values in $\mathbf{Pr}(\mathbb{R})$.

A random matrix can then be defined as follows:

Definition 2.1. A *random matrix* is a matrix whose entries are random variables, all of which share a sample space and attain values in the same space.

It is equivalent to say that a *random matrix* is a random variable that attains values in a space of matrices $M_{m\times n}(A)$, where A is a σ-compact metric space. It is instructive to check that these two definitions are the same.

In the most commonplace probabilistic scenario, we deal with random variables whose values are numbers, and we say that $X_n \to X$ in probability as $n \to \infty$ if $\lim_{n\to\infty} \mathbb{P}(|X_n - X| \geq \varepsilon) = 0$ for any $\varepsilon > 0$, where $\mathbb{P}$ denotes probability on the sample space. We generalize this definition as follows:

Definition 2.2. Let $\{X_n\}_{n\in\mathbb{N}}$ be a sequence of random variables which share a sample space and attain values in a σ-compact metric space M. Let X be another such random variable. We say that X_n *converges in probability* to X if $\forall \varepsilon > 0$, we have

$$\mathbb{P}\left(d(X_n, X) \geq \varepsilon\right) \to 0, \ as\, n \to \infty$$

where d is the metric on M.

An equivalent definition is that for any $\varepsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(d(X_n, X) < \varepsilon\right) = 1$$

One might ask if the probability $\mathbb{P}\left(d(X_n, X) \geq \varepsilon\right)$ in this definition is well-defined in the sense that $\{\omega : d\left(X_n(\omega), X(\omega)\right) \geq \varepsilon\} \subset \Omega$ is a measurable subset of the sample space. This is indeed the case because M is a σ-compact metric space. We sketch a proof below:

Theorem 2.3. *The probability $\mathbb{P}\left(d(X_n, X) \geq \varepsilon\right)$ in Definition 2.3 is well-defined.*

Proof. The metric $d : M \times M \to \mathbb{R}^{\geq 0}$ is a continuous function, so d is Borel measurable. Since M is a σ-compact metric space, $M \times M$ is a σ-compact metric space when we define the metric d_2 such that

$$d_2\left((x_1, y_1), (x_2, y_2)\right) := \max\{d(x_1, x_2), d(y_1, y_2)\}$$

In the map $\omega \mapsto (X_n(\omega), X(\omega))$ from the sample space to $M \times M$, we have that every open ball in this metric has a measurable preimage.

Since $M \times M$ is a σ-compact metric space, it is second-countable with a countable basis consisting of open balls, so that every open set in $M \times M$ has a measurable preimage. Thus, the map $\omega \mapsto (X_n(\omega), X(\omega))$ is measurable. Thus, the probability $\mathbb{P}\left(d(X_n, X) \geq \varepsilon\right)$ measures a preimage of a Borel set on a composition of measurable functions, i.e. a measurable set, so this probability is well-defined in the metric d. □

In order to define convergence in probability, we have fixed a metric d on the space of values. We will soon need to consider the case when a topological space is *metrizable* but does not necessarily come paired with a metric. More on this will come.

To study convergence in probability $\mu_n \to \mu$ of the random probability measures in $\mathbf{Pr}(\mathbb{R})$, we will need to assign a metric to the set $\mathbf{Pr}(\mathbb{R})$. We will now set up for the definition of a metric under which $\mathbf{Pr}(\mathbb{R})$ is a σ-compact metric space. Recall that we need such a metric in order to handle random variables that take values in $\mathbf{Pr}(\mathbb{R})$.

We say a complex measure μ is regular if its total variation $|\mu|$ is regular. We have the following famous result from analysis [7]:

Theorem 2.4 (Riesz representation theorem for continuous compactly supported functions)**.** *Let X be a locally compact Hausdorff space. For every $\phi \in C_C(X)^*$, there is a unique complex regular Borel measure μ such that*

$$\phi(f) := \int f \, d\mu$$

where additionally,

$$\|\phi\| = |\mu|(X)$$

In the Riesz representation theorem, $C_C(X)$ is the space of complex-valued continuous compactly supported functions of X (where $C_C(X)$ has the supremum norm), while $C_C(X)^*$ is its continuous dual space (with the operator norm). As a converse to this result, each complex regular Borel measure μ of X defines a bounded (i.e. continuous) linear functional via the map $f \mapsto \int f \, d\mu$. Thus, the complex regular Borel measures are not only determined by the functions in $C_C(X)$, but are

in one-to-one correspondence with the continuous linear functionals of $C_C(X)$ via the maps given by integration.

Thus, we can interpret the measures in $\mathbf{Pr}(\mathbb{R})$ as elements of $C_C(\mathbb{R})^*$. If we denote $C_b(\mathbb{R})$ to be the larger space that consists of complex-valued continuous bounded functions of $\mathbb{R}$, then the closure of $C_C(\mathbb{R})$ in $C_b(\mathbb{R})$ is the Banach space $C_0(\mathbb{R})$, where $C_0(\mathbb{R})$ denotes the space of complex-valued continuous functions vanishing at infinity with the supremum norm. $C_C(\mathbb{R})$ and $C_0(\mathbb{R})$ have the same dual space, and $C_0(\mathbb{R})^*$ can be substituted for $C_C(\mathbb{R})^*$ in the Riesz representation theorem.

One has from functional analysis that the weak-* topology on the (continuous) dual space $C_0(\mathbb{R})^*$ is the coarsest topology for which the map $\mu \mapsto \int f\, d\mu$ is continuous for any $f \in C_0(\mathbb{R})$, so that weak-* convergence $\mu_n \to \mu$ of measures $\mu_n \in C_0(\mathbb{R})^*$ occurs if and only if

$$\int f\, d\mu_n \to \int f\, d\mu$$

as $n \to \infty$ for any $f \in C_0(\mathbb{R})$. This topology on the measures will suit us.

Definition 2.5. The *vague topology* is defined to be the weak-* topology on $C_0(\mathbb{R})^*$.

Here, points in $C_0(\mathbb{R})^*$ are topologically distinguished only to the extent that their values at individual functions in $C_0(\mathbb{R})$ are topologically distinguished in the ground field $\mathbb{C}$. This gives a more *vague* notion of "space" in $C_0(\mathbb{R})^*$ than is present under the $C_0(\mathbb{R})^*$ operator norm, which induces a less coarse topology. Since the standard topology in $C_0(\mathbb{R})^*$ induced by the operator norm has less compact sets (and therefore less σ-compact sets) than the vague topology does, this alternative, more "vague" topology will be better suited to our needs.

We endow $\mathbf{Pr}(\mathbb{R})$ with the *vague topology* by taking it to be a (topological) subspace of $C_0(\mathbb{R})^*$ with the vague topology.

A useful property of the Banach space $C_0(\mathbb{R})$ is that it is separable (i.e. it has a countable dense subset). To prove this, the Stone-Weierstrass theorem is used: one can find a dense sequence by enumerating all polynomials with coefficients whose real and imaginary parts are rational, cutting them off at 0 outside of a compact interval, and repeating countably many times for successively larger intervals. One has a standard result from functional analysis [3]:

Theorem 2.6. *Suppose E is a separable Banach space. Then, any norm-bounded subset B of E^* is weak-* metrizable. Choosing $\{v_n\}_{n\in\mathbb{N}} \subset E$ to be a dense subset of the unit ball of E, the metric*

$$d^*(\phi_1, \phi_2) := \sum_{n=1}^{\infty} 2^{-n} |\phi_1(v_n) - \phi_2(v_n)|$$

induces the weak- topology on B.*

The probability measures of $\mathbf{Pr}(\mathbb{R}) \subset C_0(\mathbb{R})^*$ are linear functionals with norm 1. By the Banach-Alaoglu theorem from functional analysis, any closed ball with respect to the operator norm centered at the origin in $C_0(\mathbb{R})^*$ is compact in the vague topology. Thus, any norm-bounded subset of $C_0(\mathbb{R})^*$ that is vaguely closed is vaguely compact. In $\mathbf{Pr}(\mathbb{R})$, one sees that the point measures $\mu_N = \delta_N$ converge in the vague topology to 0 as $N \to \infty$; thus, while $\mathbf{Pr}(\mathbb{R})$ is norm-bounded in $C_0(\mathbb{R})^*$, it is not vaguely closed (where compact implies closed in a metrizable

space). However, $\mathbf{Pr}(\mathbb{R})$ is vaguely σ-compact. The vague σ-compactness of $\mathbf{Pr}(\mathbb{R})$, along with its vague metrizability, is what will allow us to handle random variables that take values in $\mathbf{Pr}(\mathbb{R})$ and to look at their convergence.

While Theorem 2.3 establishes that $\mathbf{Pr}(\mathbb{R})$ with the vague topology is metrizable, we do not have a canonical choice of metric. This is because the numerical distances assigned by d^* in Theorem 2.3 depend upon the dense sequence of points $\{v_n\}_{n\in\mathbb{N}}$ in the unit ball of E, which is arbitrarily chosen. Given random measures μ_n, μ with values in $\mathbf{Pr}(\mathbb{R})$, we want convergence in probability $\mathbb{P}\left(d^*(\mu_N, \mu) \geq \varepsilon\right) \to 0$ for all $\varepsilon > 0$ to be meaningful, which brings us to a question: Given two metrics d and d' that induce the same topology, is convergence in probability equivalent between the two metrics? Does convergence in one imply convergence in the other? The answer is yes, and we prove this here:

Theorem 2.7 (Metric invariance of convergence in probability)**.** *Let M be a σ-compact space under the metrics d, d', which induce the same topology. Let $\{X_n\}_{n\in\mathbb{N}}$ and X be random variables from a sample space to M. Then, $X_n \to X$ in probability with respect to d if and only if $X_n \to X$ in probability with respect to d', i.e., convergence in probability is always invariant under the choice of metric.*

Proof. Suppose that d, d' are two metrics that induce the σ-compact space M. Suppose that $\{X_n\}_{n\in\mathbb{N}}, X$ are random variables taking values in M such that $X_n \to X$ in probability with respect to d'. Our aim is to show that $X_n \to X$ in probability with respect to d. Namely, given $\varepsilon > 0$, we aim to show that

$$\mathbb{P}\left(d(X_n, X) \geq \varepsilon\right) \to 0$$

as $n \to \infty$. The idea will be to show that $\mathbb{P}\left(d(X_n, X) \geq \varepsilon\right)$ is bounded by a value close to $\mathbb{P}\left(d'(X_n, X) \geq \delta\right)$ for some $\delta > 0$.

Choose sets $K_n \subset M$ such that $\bigcup_{n\in\mathbb{N}} K_n = M$, where each K_n is compact. We are safe to assume that $K_1 \subset K_2 \subset \cdots$, as a finite union of compact sets is compact. Then,

$$\mathbb{P}(X \in K_n) \to 1$$

as $n \to \infty$, so given an additional parameter $\varepsilon' > 0$, we can choose $r \in \mathbb{N}$ large enough that $\mathbb{P}(X \in K_r) > 1 - \varepsilon'$, i.e., $\mathbb{P}(X \notin K_r) < \varepsilon'$.

For every point $x \in K_r$, choose $\delta_x > 0$ such that $B^{d'}_{\delta_x}(x) \subset B^d_\varepsilon(x)$, where $B^d_\varepsilon(x)$ denotes the open ball centered at x of radius ε with respect to the metric d. Choose a finite subcover of

$$\bigcup_{x\in K_r} B^{d'}_{\frac{\delta_x}{2}}(x) \supset K_r$$

and denote it

$$B^{d'}_{\frac{\delta_{x_1}}{2}}(x_1) \cup \cdots \cup B^{d'}_{\frac{\delta_{x_k}}{2}}(x_k)$$

Let $\delta = \min\{\delta_{x_1}, ..., \delta_{x_k}\}$. Now, suppose $d'(y, z) < \frac{\delta}{2}$, where in particular, $z \in K_r$. Then, there exists x_i from the finite subcover such that $d'(z, x_i) < \frac{\delta_{x_i}}{2}$, so that

$$d'(y, x_i) \leq d'(y, z) + d'(z, x_i) < \delta_{x_i}$$

which gives $y \in B^{d'}_{\delta_{x_i}}(x_i) \Rightarrow y \in B^d_\varepsilon(x_i)$. Thus,

$$d(y, z) \leq d(y, x_i) + d(z, x_i) < \varepsilon + \varepsilon = 2\varepsilon$$

and we see that $d'(y, z) < \frac{\delta}{2}$ implies $d(y, z) < 2\varepsilon$ when $z \in K_r$.

Now, we have

$$\mathbb{P}\left(d'(X_n, X) < \frac{\delta}{2}\right) \leq \mathbb{P}\left(d(X_n, X) < 2\varepsilon, X \in K_r\right) + \mathbb{P}(X \notin K_r)$$
$$< \mathbb{P}\left(d(X_n, X) < 2\varepsilon\right) + \varepsilon'$$

Equivalently,

$$\mathbb{P}\left(d(X_n, X) \geq 2\varepsilon\right) < \mathbb{P}\left(d(X_n, X) \geq \frac{\delta}{2}\right) + \varepsilon'$$

which means that $\limsup_{n\to\infty} \mathbb{P}\left(d(X_n, X) \geq 2\varepsilon\right) \leq \varepsilon'$. Since $\varepsilon' > 0$ was arbitrary, we therefore have

$$\mathbb{P}\left(d(X_n, X) \geq 2\varepsilon\right) \to 0$$

as $n \to \infty$. Since $\varepsilon > 0$ was arbitrary, we may conclude that $X_n \to X$ in probability with respect to d if $X_n \to X$ in probability with respect to d'. Therefore, convergence in probability is invariant under the choice of metric on the space of values. □

By this result, convergence in probability is in fact well-defined for any *σ-compact metrizable space.*

A useful characterization to work with for convergence in probability of random measures $\mu_n \to \mu$ in $\mathbf{Pr}(\mathbb{R})$ will be that $\int f\, d\mu_n \to \int f\, d\mu$ in probability for every $f \in C_C(\mathbb{R})$. In fact, this is equivalent to the notion of convergence in probability that we have defined, and we establish this fact here:

Theorem 2.8. *Suppose B is a norm-bounded, weak-* σ-compact subset of E^*, where E is a separable Banach space. For random variables whose space of values is B, the following are equivalent:*

(1) $X_n \to X$ in probability.

(2) $X_n(v) \to X(v)$ in probability for all v in some dense subset of E.

Proof. We first show that (1) $\Rightarrow$ (2). Choose a sequence $\{v_k\}_{k\in\mathbb{N}}$ dense in the unit ball of E, and define

$$d^*(\phi_1, \phi_2) := \sum_{k=1}^{\infty} 2^{-k} |\phi_1(v_k) - \phi_2(v_k)|$$

Suppose that $\exists v_k$ such that $X_n(v_k) \not\to X(v_k)$ in probability as $n \to \infty$, so that $\exists \varepsilon > 0$ such that

$$\mathbb{P}(|X_n(v_k) - X(v_k)| \geq \varepsilon) \not\to 0$$

as $n \to \infty$. Then, whenever $|X_n(v_k) - X(v_k)| \geq \varepsilon$, we have that $d^*(X_n, X) \geq 2^{-k}\varepsilon$, where then

$$\mathbb{P}\left(d^*(X_n, X) \geq 2^{-k}\varepsilon\right) \geq \mathbb{P}\left(|X_n(v_k) - X(v_k)| \geq \varepsilon\right)$$

for all n. Thus, $\mathbb{P}\left(d^*(X_n, X) \geq 2^{-k}\varepsilon\right) \not\to 0$, and $X_n \not\to X$ in probability.

We now have that $X_n \to X$ in probability implies

$$X_n(v_k) \to X(v_k)$$

in probability for all v_k in the sequence. By enumerating all rational scalar multiplies of points in $\{v_k\}_{k\in\mathbb{N}}$, we get a sequence dense in E. Noting that

$$\mathbb{P}\left(|X_n(v) - X(v)| \geq \varepsilon\right) = \mathbb{P}\left(|X_n(rv) - X(rv)| \geq r\varepsilon\right)$$

for any $r > 0$, we see if $X_n \to X$ in probability that $X_n(v) \to X(v)$ for all v in a dense subset of E. Thus, (1) implies (2).

Now, we aim to show that (2) implies (1). Suppose that $X_n \not\to X$ in probability. Suppose that $D \subset E$ is some dense subset and that $\exists v \in E$ such that

$$X_n(v) \not\to X(v)$$

in probability. Choose $C > 0$ such that $C \geq \sup_{\phi \in B} \|\phi\|$. For this C, we can choose $w \in D$ such that $\|w - v\| < \frac{\varepsilon}{4C}$, where ε is chosen such that

$$\mathbb{P}\left(|X_n(v) - X(v)| \geq \varepsilon\right) \not\to 0$$

Then, whenever $|X_n(v) - X(v)| \geq \varepsilon$, we have

$$|X_n(v-w)| + |X_n(w) - X(w)| + |X(v-w)| \geq \varepsilon$$

which implies

$$|X_n(w) - X(w)| > \varepsilon - \frac{\varepsilon}{4} - \frac{\varepsilon}{4} = \frac{\varepsilon}{2}$$

Thus, $\mathbb{P}\left(|X_n(v) - X(v)| \geq \varepsilon\right) \not\to 0$ implies $\mathbb{P}\left(|X_n(w) - X(w)| \geq \frac{\varepsilon}{2}\right) \not\to 0$. To prove that (2) implies (1), it therefore suffices to show that $X_n(v) \not\to X(v)$ in probability for some $v \in E$.

For $\varepsilon > 0$ such that $\mathbb{P}\left(d^*(X_n, X) \geq \varepsilon\right) \not\to 0$, where d^* is defined according to a sequence $\{v_k\}_{k \in \mathbb{N}}$ dense in the unit ball of E, we can choose $r \in \mathbb{N}$ large enough to have

$$\sum_{k=r}^{\infty} 2^{-k} |X_n(v_k) - X(v_k)| < \sum_{k=r}^{\infty} 2^{-k} \cdot 2C = 2^{-(r-2)} C < \frac{\varepsilon}{2}$$

regardless of n or the outcome in the sample space. Then, whenever $d^*(X_n, X) \geq \varepsilon$, we have

$$\sum_{k=1}^{r-1} 2^{-k} |X_n(v_k) - X(v_k)| > \frac{\varepsilon}{2}$$

in which case $\exists k \in \{1, ..., r-1\}$ such that $|X_n(\frac{v_k}{2^k}) - X(\frac{v_k}{2^k})| > \frac{\varepsilon}{2(r-1)}$.

Since $\mathbb{P}\left(d^*(X_n, X) \geq \varepsilon\right) \not\to 0$, $\exists \varepsilon' > 0$ such that $\mathbb{P}\left(d^*(X_n, X) \geq \varepsilon\right) \geq \varepsilon'$ *i.o.* (infinitely often) with respect to n. Therefore,

$$\begin{aligned}\sum_{k=1}^{r-1} \mathbb{P}\left(|X_n(\frac{v_k}{2^k}) - X(\frac{v_k}{2^k})| \geq \frac{\varepsilon}{2(r-1)}\right) &\geq \mathbb{P}\left(\bigcup_{k=1}^{r-1} \{|X_n(\frac{v_k}{2^k}) - X(\frac{v_k}{2^k})| \geq \frac{\varepsilon}{2(r-1)}\}\right) \\ &\geq \mathbb{P}\left(d^*(X_n, X) \geq \varepsilon\right) \geq \varepsilon'\end{aligned}$$

i.o. with respect to n. Thus, from the pigeonhole principle, $\exists k \in \{1, ..., r-1\}$ such that

$$\mathbb{P}\left(|X_n(\frac{v_k}{2^k}) - X(\frac{v_k}{2^k})| \geq \frac{\varepsilon}{2(r-1)}\right) \geq \frac{\varepsilon'}{r-1}$$

i.o. with respect to n, so $\frac{v_k}{2^k} \in E$ is such that $X_n(\frac{v_k}{2^k}) \not\to X(\frac{v_k}{2^k})$ in probability. Thus, (2) implies (1), and the theorem is proven. □

In Wigner's semicircle law, we look at convergence in probability of random variables $\rho_N \to \mu_W$ whose values are measures, where μ_W is a random variable constant at a measure with a semicircular density

$$\mu_W(dx) := \frac{1}{2\pi} \sqrt{(4 - x^2)^+} dx$$

where we use the notation

$$f(x)^+ := \begin{cases} f(x) & \text{if } f(x) > 0 \\ 0 & \text{otherwise} \end{cases}$$

We now establish that the following are equivalent:

Theorem 2.9. *Let $\{\mu_N\}_{N\in\mathbb{N}}$ be a sequence of random probability measures in $\mathbf{Pr}(\mathbb{R})$, and let $\mu_W(dx) = \frac{1}{2\pi}\sqrt{(4-x^2)^+}\,dx$. The following are equivalent:*

(1) $\mu_N \to \mu_W$ vaguely in probability

(2) $\int f\,d\mu_N \to \int f\,d\mu_W$ in probability $\forall f \in C_C(\mathbb{R})$

(3) $\mu_N(I) \to \mu_W(I)$ in probability for any bounded interval $I \subset \mathbb{R}$

For condition (3), we provide additionally that the map $\omega \mapsto \mu_N(\omega)(I)$ from the sample space to $\mathbb{R}$ is measurable for all N and for any bounded interval $I \subset \mathbb{R}$.

Proof. Equivalence of (1) and (2) follows from the proof of Theorem 2.5. To establish the equivalence of (1) and (2) with (3), one makes use of the fact that the measure μ_W is absolutely continuous with respect to Lebesgue measure. □

3. Wigner's Semicircle Law: An Analytic Approach

We are now ready for a statement of the theorem. For the scope of this section, we limit ourselves to the case of Gaussian-distributed random variables. However, further analytic arguments can be used to extend this result to the general real symmetric case, or even to the more general case of Hermitian random matrices (whose entries are complex-valued). We credit this statement of the theorem to Fraydoun Rezakhanlou's Lectures on Random Matrices [6]:

Theorem 3.1 (Wigner's semicircle law, convergence in probability)**.** *Let $(h_{ij})_{i<j}$ be i.i.d. Gaussian random variables (of the same sample space) where j ranges over the natural numbers. Let (h_{ii}) be another set of i.i.d. Gaussian random variables (of that same sample space), indexed by $i \in \mathbb{N}$. We assume that*

$$\mathbb{E}h_{ij} = 0,\ \mathrm{Var}(h_{ij}) = 1$$

for all $i < j$, and

$$\mathbb{E}h_{ii} = 0,\ \mathrm{Var}(h_{ii}) = 2$$

Define $h_{ji} = h_{ij}$ for all i, j. Let H_N be the real symmetric random matrix with entries $(\frac{h_{ij}}{\sqrt{N}})_{1\le i,j\le N}$. We define the random measure

$$\rho_N = \frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i}$$

where $\lambda_1 \le \cdots \le \lambda_N$ are the eigenvalues of H_N.

Then, $\rho_N \to \mu_W$ in probability in the vague topology, where

$$\mu_W(dx) = \frac{1}{2\pi}\sqrt{(4-x^2)^+}\,dx$$

gives the density of μ_W with respect to Lebesgue measure.

Here, δ_x denotes the point measure for which the singleton $\{x\}$ has measure 1 (while the complement of $\{x\}$ has measure zero). The scaling factor $\frac{1}{\sqrt{N}}$ yields Gaussian random matrices with $\mathrm{Var}(\frac{h_{ij}}{\sqrt{N}}) = \frac{1}{N}$ (off-diagonal) or $\mathrm{Var}(\frac{h_{ii}}{\sqrt{N}}) = \frac{2}{N}$ (on-diagonal), through which by Wigner's semicircle law, the eigenvalues are revealed

to concentrate semicircularly as we send $N \to \infty$. We take care to notice that since H_N is a real symmetric random matrix, it is in all outcomes a self-adjoint linear operator and is therefore diagonalizable with real eigenvalues.

We make the necessary observation that ρ_N indeed defines a random variable (with values in $\mathbf{Pr}(\mathbb{R})$). The random matrix H_N is a measurable function of the sample space, while if we take ρ_N to be a function of the space of $M_{N\times N}(\mathbb{R})$ symmetric matrices, we can see that it is continuous. The vague topology on $\mathbf{Pr}(\mathbb{R})$ has a basis given by finite intersections of preimages of open sets on maps $\mu \mapsto \int f\, d\mu$ for any $f \in C_0(\mathbb{R})$, and when a matrix is changed by a small amount, its eigenvalues are changed by a small amount, so that vaguely open sets have open preimages in the space of matrices. As such, ρ_N is continuous as a function of the space of $M_{N\times N}(\mathbb{R})$ symmetric matrices and is therefore Borel measurable, and thus, ρ_N defines a measurable function of the sample space and is a random variable.

In this section of the paper, we will prove Wigner's semicircle law by using methods from analysis. An essential tool to this end will be the Stieltjes transform, wherein to determine vague convergence of a sequence of probability measures, it suffices to check convergence of the integral for a certain subset of $C_0(\mathbb{R})$ functions. We will contribute a proof of a key result known as the Stieltjes continuity theorem for convergence in probability, whose statement we credit to Terence Tao's *Topics in Random Matrix Theory* [9]. Then, we will prove Wigner's semicircle law by closely following the arguments found in the 2012 notes for Fraydoun Rezakhanlou's Lectures on Random Matrices [6]. The reader is encouraged to access the two aforementioned sources for a greater coverage of this remarkable topic.

We define the Stieltjes transform as follows [6]:

Definition 3.2. Let $\mu \in \mathbf{Pr}(\mathbb{R})$. The *Stieltjes transform* of μ is the function $S_\mu : (\mathbb{C} \setminus \mathbb{R}) \to \mathbb{C}$ defined by

$$S_\mu(z) = \int \frac{1}{x-z}\, \mu(dx)$$

For a measure $\mu \in \mathbf{Pr}(\mathbb{R})$, we can observe that this function S_μ is analytic in the upper and lower half-planes using the limit definition of the complex derivative and the bounded convergence theorem. Writing $z = a + ib$ and letting a vary, we have

$$|S_\mu(a_1+ib) - S_\mu(a_2+ib)| \le \int |\frac{a_1-a_2}{(x-a_1-ib)(x-a_2-ib)}|\, \mu(dx) \le \frac{1}{b^2}|a_1-a_2|$$

so that $S_\mu(z)$ is Lipschitz continuous when the imaginary part of z is fixed (or bounded away from 0). Note that a Lipschitz constant of $\frac{1}{b^2}$ is shared by functions $a \mapsto S_\mu(a+ib)$ for all possible choices of μ. We will make use of this fact in the coming proofs.

In the next lemma,

$$\frac{1}{\pi}\operatorname{Im} S_\mu(a+ib)\, da$$

denotes the measure that has density $\frac{1}{\pi}\operatorname{Im} S_\mu(a+ib)$ with respect to Lebesgue measure in a. Here, we strengthen a result established in Rezakhanlou's notes (by proving a uniformity condition over probability measures) [6]:

Lemma 3.3. *Given μ, we have*

$$\lim_{b\to 0}\frac{1}{\pi}\operatorname{Im} S_\mu(a+ib)\,da=\mu$$

in the vague topology. If some $f\in C_0(\mathbb{R})$ is fixed while the measure μ is allowed to vary, then the convergence

$$|\int f\frac{1}{\pi}\operatorname{Im} S_\mu(a+ib)\,da-\int f\,d\mu|\to 0$$

as $b\to\infty$ is uniform over μ.

Proof. It suffices to show the uniform convergence for when $f\in C_0(\mathbb{R})$ is Lipschitz continuous, since Lipschitz functions are dense in $C_0(\mathbb{R})$. Let $f\in C_0(\mathbb{R})$ be Lipschitz. For any μ, we have

$$\begin{aligned}\int f\frac{1}{\pi}\operatorname{Im} S_\mu(a+ib)\,da&=\int f\frac{1}{\pi}\int \operatorname{Im}\left[\frac{1}{x-a-ib}\right]\mu(dx)\,da\\&=\int f\frac{1}{\pi}\int \operatorname{Im}\left[\frac{x-a}{(x-a)^2+b^2}+i\frac{b}{(x-a)^2+b^2}\right]\mu(dx)\,da\\&=\int\int f(a)\frac{1}{\pi}\frac{b}{(x-a)^2+b^2}\,\mu(dx)\,da=\int\int f(a)\frac{1}{\pi}\frac{b}{(x-a)^2+b^2}da\,\mu(dx)\\&=\int (f*C_b)\,d\mu\end{aligned}$$

where we use the notation C_b for the Cauchy density

$$C_b(x):=\frac{1}{b}C_1(\frac{x}{b})=\frac{1}{\pi}\frac{b}{x^2+b^2}$$

Here, we denote

$$C_1(x):=\frac{1}{\pi}\frac{1}{x^2+1}$$

Note that an antiderivative of $C_b(x)$ is given by $\frac{1}{\pi}\arctan(\frac{x}{b})$. Fubini's theorem could be applied above because f is bounded, while the integral of $|\frac{1}{\pi}\frac{b}{(x-a)^2+b^2}|$ with respect to da is 1 for any x, so that the integrand is integrable with respect to $\mu(dx)\times da$.

Choose $C>Lip(f)$, where $Lip(f)$ denotes *the* Lipschitz constant of f (i.e. the infimum of its Lipschitz constants). Let $M>0$ be an upper bound to $|f|$. Let $\varepsilon>0$. Choose N large enough that $|\int C_1(a)\,da-\int_{-N}^{N}C_1(a)\,da|<\frac{\varepsilon}{4M}$. Now, suppose that $b\in\mathbb{R}\setminus\{0\}$ is such that $|b|N<\frac{\varepsilon}{2C}$, i.e. $|b|<\frac{\varepsilon}{2CN}$. Then, for all x, we have

$$\begin{aligned}|(f*C_b)(x)-f(x)|&=|\int\frac{1}{\pi}\frac{b}{a^2+b^2}f(x-a)\,da-\int\frac{1}{\pi}\frac{b}{a^2+b^2}f(x)\,da|\\&\le\int_{(-\infty,-|b|N)\cup(|b|N,\infty)}|f(x-a)-f(x)|\frac{1}{\pi}\frac{|b|}{a^2+b^2}\,da\\&\quad+\int_{-|b|N}^{|b|N}|f(x-a)-f(x)|\frac{1}{\pi}\frac{|b|}{a^2+b^2}\,da\\&<2M\frac{\varepsilon}{4M}+C\frac{\varepsilon}{2C}=\varepsilon\end{aligned}$$

Thus, $|b| < \frac{\varepsilon}{2CN}$ implies $\|(f * C_b) - f\| < \varepsilon$, so that

$$|\int f \frac{1}{\pi} \operatorname{Im} S_\mu(a+ib)\, da - \int f\, d\mu| = |\int ((f * C_b) - f)\ d\mu| < \varepsilon$$

This inequality holds regardless of μ, and we conclude for any $f \in C_0(\mathbb{R})$ that

$$|\int f \frac{1}{\pi} \operatorname{Im} S_\mu(a+ib)\, da - \int f\, d\mu| \to 0$$

as $b \to 0$ uniformly over μ, and

$$\frac{1}{\pi} \operatorname{Im} S_\mu(a+ib)\, da \to \mu$$

vaguely as $b \to 0$. □

We note the following immediate corollary:

Corollary 3.4. *Given μ, the values of S_μ on the upper half-plane uniquely determine the measure.*

We now state the Stieltjes continuity theorem for convergence in probability and provide a proof. Our statement of the theorem comes from [9], although our proof is original:

Theorem 3.5 (Stieltjes continuity theorem, convergence in probability)**.** *Suppose that $\{\mu_n\}_{n\in\mathbb{N}}, \mu$ are random measures with a shared sample space. Then,*

$$\mu_n \to \mu$$

vaguely in probability if and only if

$$S_{\mu_n}(z) \to S_\mu(z)$$

in probability for every z in the upper half-plane.

Proof. The "only if" direction is clear from the definition of the Stieltjes transform and from the proof of Theorem 2.5. Suppose, for the converse, that the random measures $\{\mu_n\}_{n\in\mathbb{N}}, \mu$ are such that

$$S_{\mu_n}(z) \to S_\mu(z)$$

in probability for every z in the upper half-plane. We aim to show that

$$\mu_n \to \mu$$

vaguely in probability. For this, it suffices to show that

$$\int f\, d\mu_n \to \int f\, d\mu$$

in probability for every Lipschitz $f \in C_C(\mathbb{R})$, since the space of Lipschitz functions is dense in $C_0(\mathbb{R})$.

Let $f \in C_C(\mathbb{R})$ be Lipschitz. Let $\varepsilon > 0$. We aim to show that

$$\mathbb{P}\left(|\int f\, d\mu_n - \int f\, d\mu| \geq \varepsilon\right) \to 0$$

as $n \to \infty$. Fix $b > 0$ small enough that

$$|\int f \frac{1}{\pi} \operatorname{Im} S_\nu(a+ib)\, da - \int f\, d\nu| < \frac{\varepsilon}{3}$$

for any probability measure ν. Let $I \subset \mathbb{R}$ be a bounded interval containing the support of f, and let m denote Lebesgue measure on $\mathbb{R}$, so that $m(I) < \infty$.

For any ν, the function $a \mapsto f(a)\frac{1}{\pi}\operatorname{Im} S_\nu(a+ib)$ is a product of bounded and Lipschitz functions and is therefore Lipschitz. Since the functions

$$a \mapsto \frac{1}{\pi}\operatorname{Im} S_\nu(a+ib)$$

are uniformly bounded over ν by $\frac{1}{\pi b}$, and since they share a Lipschitz constant of $\frac{1}{\pi b^2}$ for all ν, we have that the functions

$$a \mapsto f(a)\frac{1}{\pi}\operatorname{Im} S_\nu(a+ib)$$

indexed over ν share a common Lipschitz constant over all ν.

Choose some

$$C > \sup_\nu Lip\left(a \mapsto f(a)\frac{1}{\pi}\operatorname{Im} S_\nu(a+ib)\right)$$

We can partition I into finitely many intervals where each has length less than $\frac{\varepsilon}{9m(I)C}$. Choose points $a_1 < \cdots < a_k$ where one a_j is chosen from each interval. Let $M > 0$ be an upper bound to $|f|$, and suppose that ν_1, ν_2 are two measures such that

$$|S_{\nu_1}(a_j+ib) - S_{\nu_2}(a_j+ib)| < \frac{\pi\varepsilon}{9m(I)M}$$

for all $1 \le j \le k$. Then, for any $a \in I$, we can choose a_j such that $|a - a_j| < \frac{\varepsilon}{9m(I)C}$. We may then notice that

$$\begin{aligned}
&|f(a)\frac{1}{\pi}\operatorname{Im} S_{\nu_1}(a+ib) - f(a)\frac{1}{\pi}\operatorname{Im} S_{\nu_2}(a+ib)| \\
&\quad \le |f(a)\frac{1}{\pi}\operatorname{Im} S_{\nu_1}(a+ib) - f(a_j)\frac{1}{\pi}\operatorname{Im} S_{\nu_1}(a_j+ib)| \\
&\qquad + |f(a_j)||\frac{1}{\pi}\operatorname{Im} S_{\nu_1}(a_j+ib) - \frac{1}{\pi}\operatorname{Im} S_{\nu_2}(a_j+ib)| \\
&\qquad + |f(a_j)\frac{1}{\pi}\operatorname{Im} S_{\nu_2}(a_j+ib) - f(a)\frac{1}{\pi}\operatorname{Im} S_{\nu_2}(a+ib)| \\
&\quad < C\frac{\varepsilon}{9m(I)C} + M\frac{\varepsilon}{9m(I)M} + C\frac{\varepsilon}{9m(I)C} = \frac{\varepsilon}{3m(I)}
\end{aligned}$$

and thus,

$$|\int f\frac{1}{\pi}\operatorname{Im} S_{\nu_1}(a+ib)\,da - \int f\frac{1}{\pi}\operatorname{Im} S_{\nu_2}(a+ib)\,da| < \frac{\varepsilon}{3m(I)}m(I) = \frac{\varepsilon}{3}$$

Therefore, given n and some outcome in the sample space such that

$$|S_{\mu_n}(a_j+ib) - S_\mu(a_j+ib)| < \frac{\pi\varepsilon}{9m(I)M}$$

for all $1 \leq j \leq k$, we have

$$\begin{aligned}|\int f\,d\mu_n - \int f\,d\mu| \leq |\int f\,d\mu_n - \int f\frac{1}{\pi}\operatorname{Im} S_{\mu_n}(a+ib)\,da| \\ + |\int f\frac{1}{\pi}\operatorname{Im} S_{\mu_n}(a+ib)\,da - \int f\frac{1}{\pi}\operatorname{Im} S_{\mu}(a+ib)\,da| \\ + |\int f\frac{1}{\pi}\operatorname{Im} S_{\mu}(a+ib)\,da - \int f\,d\mu| \\ < \frac{\varepsilon}{3}+\frac{\varepsilon}{3}+\frac{\varepsilon}{3} = \varepsilon\end{aligned}$$

We now conclude for all n that

$$\begin{aligned}&\mathbb{P}\left(\bigcap_{j=1}^{k}\{|S_{\mu_n}(a_j+ib) - S_\mu(a_j+ib)| < \frac{\pi\varepsilon}{9m(I)M}\}\right) \\ &\leq \mathbb{P}\left(|\int f\,d\mu_n - \int f\,d\mu| < \varepsilon\right)\end{aligned}$$

Equivalently,

$$\begin{aligned}\mathbb{P}\left(|\int f\,d\mu_n - \int f\,d\mu| \geq \varepsilon\right) &\leq \mathbb{P}\left(\bigcup_{j=1}^{k}\{|S_{\mu_n}(a_j+ib) - S_\mu(a_j+ib)| \geq \frac{\pi\varepsilon}{9m(I)M}\}\right) \\ &\leq \sum_{j=1}^{k}\mathbb{P}\left(|S_{\mu_n}(a_j+ib) - S_\mu(a_j+ib)| \geq \frac{\pi\varepsilon}{9m(I)M}\right)\end{aligned}$$

Since

$$\sum_{j=1}^{k}\mathbb{P}\left(|S_{\mu_n}(a_j+ib) - S_\mu(a_j+ib)| \geq \frac{\pi\varepsilon}{9m(I)M}\right) \to 0$$

as $n \to \infty$, we therefore have

$$\mathbb{P}\left(|\int f\,d\mu_n - \int f\,d\mu| \geq \varepsilon\right) \to 0$$

as $n \to \infty$. This concludes the proof of the theorem. □

For Wigner's semicircle law, we consider the Stieltjes transform of

$$\rho_N = \frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i}$$

which for each N and each outcome is given by

$$S_{\rho_N}(z) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{\lambda_i - z}$$

for eigenvalues $\lambda_1 \leq \cdots \leq \lambda_N$. Given N and an outcome in the sample space, $S_{\rho_N}(z)$ is equal to the *normalized trace* (i.e., $1/N$ times the trace) of the resolvent of the $N \times N$ matrix H_N, where for a self-adjoint matrix A, we define

$$R_A(z) := (A - z)^{-1}$$

to be the *resolvent* of A [6]. (We denote $z := zI$ when we write $(A - z)^{-1}$.)

An essential ingredient in the proof of Wigner's semicircle law will be the fact that when $z \in \mathbb{C} \setminus \mathbb{R}$ and N are fixed, the normalized trace of the resolvent will be a Lipschitz function of the space of self-adjoint (i.e., Hermitian) $N \times N$ matrices. We credit the following theorem and arguments to Rezakhanlou's notes [6].

Theorem 3.6. *Fix $z \in \mathbb{C} \setminus \mathbb{R}$. As a function of the space of $N \times N$ self-adjoint complex matrices (with the $\mathbb{R}^{N^2}$ Euclidean norm), the function $A \mapsto \frac{1}{N} \mathbf{Tr}\, R_A(z)$ is Lipschitz continuous, with*

$$Lip\left(A \mapsto \frac{1}{N} \mathbf{Tr}\, R_A(z)\right) \leq \frac{|\operatorname{Im} z|^{-2}}{\sqrt{N}}$$

A proof of this makes use of the following lemma:

Lemma 3.7. *Let A, A' be two $N \times N$ self-adjoint complex matrices whose respective eigenvalues are ordered $\lambda_1^A \leq \cdots \leq \lambda_N^A$ and $\lambda_1^{A'} \leq \cdots \leq \lambda_N^{A'}$. Then,*

$$\sum_{i=1}^{N} (\lambda_i^A - \lambda_i^{A'})^2 \leq \|A - A'\|^2 = \mathbf{Tr}\left((A - A')^2\right)$$

Proof. The equality $\|A - A'\|^2 = \mathbf{Tr}\left((A - A')^2\right)$ is because $A - A'$ is Hermitian. Since $\sum_{i=1}^{N} (\lambda_i^A)^2 = \mathbf{Tr}\, A^2$ and $\sum_{i=1}^{N} (\lambda_i^{A'})^2 = \mathbf{Tr}(A')^2$, a proof of this lemma amounts to showing that

$$\mathbf{Tr}(AA' + A'A) \leq 2 \sum_{i=1}^{N} \lambda_i^A \lambda_i^{A'}$$

It suffices for us to show WLOG that $\mathbf{Tr}\, AA' \leq \sum_{i=1}^{N} \lambda_i^A \lambda_i^{A'}$, where indeed, one finds that $\mathbf{Tr}\, AA'$ is real. Assuming without loss of generality that A is diagonal with entries $A_{ii} = \lambda_i^A$, we can diagonalize the Hermitian matrix A' by some orthonormal matrix $U = (u_{ij})_{1 \leq i,j \leq N}$ such that $(UA'U^*)_{ii} = \lambda_i^{A'}$ $\forall i$, where U^* denotes the conjugate transpose; then,

$$\mathbf{Tr}\, AA' = \mathbf{Tr}\, AU^*(UA'U^*)U = \sum_{1 \leq i,j \leq N} \lambda_i^A \lambda_j^{A'} |u_{ij}|^2$$

We note that the matrix $(|u_{ij}|^2)_{1 \leq i,j \leq N}$ is a nonnegative matrix and is doubly stochastic, i.e., each of its rows and columns sums to 1. Let us suppose that $|u_{11}|^2 < 1$. Then, there exist $1 \leq k, \ell \leq N$ such that $|u_{k1}|^2 > 0$ and $|u_{1\ell}|^2 > 0$. Let $r = \min\{|u_{k1}|^2, |u_{1\ell}|^2\}$, and let $V = (v_{ij})_{1 \leq i,j \leq N}$ denote the matrix obtained from $(|u_{ij}|^2)_{1 \leq i,j \leq N}$ by subtracting r from the $k1$ and 1ℓ entries and adding r to the 11 and $k\ell$ entries. This results in a doubly stochastic matrix whose 11 entry has been made strictly greater and which has one more 0 entry outside of its diagonal. Here, we have

$$\sum_{1 \leq i,j \leq N} \lambda_i^A \lambda_j^{A'} v_{ij} - \sum_{1 \leq i,j \leq N} \lambda_i^A \lambda_j^{A'} |u_{ij}|^2 = r(\lambda_1^A - \lambda_k^A)(\lambda_1^{A'} - \lambda_\ell^{A'}) \geq 0$$

We can repeat this process until the 11 entry becomes 1, while all non-diagonal entries in the 1st row and 1st column are zero. Proceeding for every ii entry as i increases to N, we utlimately arrive at the identity matrix, and we find

$$\mathbf{Tr}\, AA' = \sum_{1 \leq i,j \leq N} \lambda_i^A \lambda_j^{A'} |u_{ij}|^2 \leq \sum_{i=1}^{N} \lambda_i^A \lambda_i^{A'}$$

The proof of the lemma is complete. □

We may now prove Theorem 3.3. If A and A' are two self-adjoint $N \times N$ matrices whose respective eigenvalues are ordered $\lambda_1^A \leq \cdots \leq \lambda_N^A$ and $\lambda_1^{A'} \leq \cdots \leq \lambda_N^{A'}$, then fixing $z \in \mathbb{C} \setminus \mathbb{R}$, we need only observe that

$$\begin{aligned}
|\frac{1}{N}\,\mathbf{Tr}\,R_A(z) - \frac{1}{N}\,\mathbf{Tr}\,R_{A'}(z)| &= |\frac{1}{N}\sum_{i=1}^{N}(\lambda_i^A - z)^{-1} + \frac{1}{N}\sum_{i=1}^{N}(\lambda_i^{A'} - z)^{-1}| \\
&\leq \frac{1}{N}\sum_{i=1}^{N}\frac{|\lambda_i^{A'} - \lambda_i^A|}{|(\lambda_i^A - z)(\lambda_i^{A'} - z)|} \leq \frac{1}{\sqrt{N}}|\operatorname{Im} z|^{-2}\left(\frac{1}{\sqrt{N}}\sum_{i=1}^{N}|\lambda_i^A - \lambda_i^{A'}|\right) \\
&\leq \frac{|\operatorname{Im} z|^{-2}}{\sqrt{N}}\left(\sum_{i=1}^{N}|\lambda_i^A - \lambda_i^{A'}|^2\right)^{\frac{1}{2}} \leq \frac{|\operatorname{Im} z|^{-2}}{\sqrt{N}}\|A - A'\|
\end{aligned}$$

This proves the theorem. We have used the inequality $\frac{\|x\|_1}{\sqrt{N}} \leq \|x\|_2$ on Euclidean vectors $x \in \mathbb{R}^N$; this equality can be obtained by applying the Cauchy-Schwarz inequality to vectors $(|x_1|, ..., |x_N|)$ and $(\frac{1}{\sqrt{N}}, ..., \frac{1}{\sqrt{N}})$. Hence, if we fix $z \in \mathbb{C} \setminus \mathbb{R}$, then we can interpret the Stieltjes transform $S_{\rho_N}(z)$ as a Lipschitz function of the space of $N \times N$ real symmetric matrices attained by the random matrix H_N, whose Lipschitz constant is bounded in proportion to $\frac{1}{\sqrt{N}}$.

The Lipschitz continuity that characterizes $S_{\rho_N}(z)$ will be used in service of a concentration inequality. To prepare this, we introduce the *Logarithmic Sobolev Inequality*, or *LSI*, whose definition comes from [6]:

Definition 3.8. Let μ be a probability measure of a Euclidean space $\mathbb{R}^M$. We say that the measure μ is $LSI(a)$ if for any $f : \mathbb{R}^M \to \mathbb{R}$ that is a probability density with respect to μ, we have

$$\int f \log f \, d\mu \leq a \int |\nabla \sqrt{f}|^2 \, d\mu$$

(when f is differentiable).

A useful property of $LSI(a)$ measures is as follows [6]:

Lemma 3.9 (Herbst's)**.** *Let μ be an $LSI(a)$ probability measure of $\mathbb{R}^M$. If*

$$\int F \, d\mu = 0$$

for a Lipschitz function $F : \mathbb{R}^M \to \mathbb{R}$, then for all $t \geq 0$, we have

$$\int e^{tF} \, d\mu \leq \exp\left(\frac{1}{4}at^2 Lip(F)^2\right)$$

Proof. We again go by the steps found in [6]. We first assume that $F \in C^1$ and that F is bounded. Let $Z(t) = \int e^{tF} \, d\mu$. Then, since F is bounded, we have by the bounded convergence theorem that

$$Z'(t) = \lim_{h \to 0} \frac{\int e^{(t+h)F} \, d\mu - \int e^{tF} \, d\mu}{h} = \int \lim_{h \to 0} \frac{e^{(t+h)F} - e^{tF}}{h} \, d\mu = \int F e^{tF} \, d\mu$$

Choose $t > 0$, and let $f(x) = \frac{e^{tF(x)}}{Z(t)}$, so that f is a probability density function with respect to μ. Then,

$$f \log f = \frac{e^{tF}}{Z(t)} \left(tF - \log Z(t)\right)$$

while

$$\frac{\partial}{\partial x_i} \sqrt{f} = \frac{t\left(\frac{\partial}{\partial x_i} F(x)\right) e^{\frac{1}{2}tF}}{2\sqrt{Z(t)}}$$

for all $1 \leq i \leq M$, so that since μ is $LSI(a)$, we have

$$t\frac{Z'(t)}{Z(t)} - \log Z(t) \leq \frac{a}{4Z(t)} \int t^2 e^{tF} |\nabla F|^2 \, d\mu \leq \frac{at^2}{4} Lip(F)^2$$

From here, we see that

$$\frac{d}{dt} \frac{\log Z(t)}{t} = \frac{1}{t} \frac{Z'(t)}{Z(t)} - \frac{1}{t^2} \log Z(t) \leq \frac{a}{4} Lip(F)^2$$

for $t > 0$. Meanwhile, we have by L'Hôpital's rule and the bounded convergence theorem that

$$\lim_{t \to 0^+} \frac{\log Z(t)}{t} = \lim_{t \to 0^+} \frac{Z'(t)}{Z(t)} = \int F \, d\mu = 0$$

Thus, we have

$$\log \int e^{tF} \, d\mu \leq \frac{1}{4} a t^2 Lip(F)^2$$

for all $t > 0$ by the fundamental theorem of calculus, and this gives the result for when F is differentiable and bounded.

We extend to the case when F is bounded and Lipschitz but not necessarily differentiable by approximating it uniformly by smooth functions and applying the bounded convergence theorem. Finally, the result for all real-valued and Lipschitz F is obtained by taking truncations of F and applying the monotone convergence theorem. □

We add to the above proof by showing that the uniform approximation of Lipschitz functions by smooth functions is valid:

Lemma 3.10 (Lipschitz approximation by smooth functions)**.** *Let $f : \mathbb{R}^M \to \mathbb{R}$ be Lipschitz continuous. Then, for all $\varepsilon > 0$, there exists $g \in C^\infty(\mathbb{R}^M)$ such that*

$$|f(x) - g(x)| < \varepsilon$$

for all x, and

$$Lip(g) \leq Lip(f)$$

Proof. Let $\rho_1 : \mathbb{R}^M \to \mathbb{R}$ be a nonnegative mollifier, and denote $\rho_\delta(x) = \frac{1}{\delta}\rho_1(\frac{x}{\delta})$. Let $\varepsilon > 0$. Choose $C > Lip(f)$. Choose $\delta > 0$ such that $\operatorname{supp} \rho_\delta \subset B_{\frac{\varepsilon}{C}}(0)$, where $B_{\frac{\varepsilon}{C}}(0) \subset \mathbb{R}^M$ denotes the open ball of radius $\frac{\varepsilon}{C}$ centered at 0. Then, $f * \rho_\delta : \mathbb{R}^M \to \mathbb{R}$ is a smooth function, and for all $x \in \mathbb{R}^M$, we have

$$|(f * \rho_\delta)(x) - f(x)| = |\int \rho_\delta(y) f(x-y) \, dy - \int \rho_\delta(y) f(x) \, dy| < C\frac{\varepsilon}{C} = \varepsilon$$

Given $x_1, x_2 \in \mathbb{R}^M$, we have

$$|(f * \rho_\delta)(x_1) - (f * \rho_\delta)(x_2)| = |\int \rho_\delta(y) f(x_1 - y)\, dy - \int \rho_\delta(y) f(x_2 - y)\, dy|$$
$$\leq \int \rho_\delta(y) |f(x_1 - y) - f(x_2 - y)|\, dy \leq Lip(f)|x_1 - x_2|$$

Thus, $Lip(f * \rho_\delta) \leq Lip(f)$. □

We have established already that when $z \in \mathbb{C} \setminus \mathbb{R}$ is fixed, $S_{\rho_N}(z)$ is Lipschitz when we interpret it as a function of the space of $N \times N$ self-adjoint matrices. Of course, we are actually working in a probabilistic setting under which ρ_N for each N is a random measure, so that given N and z, the Stieltjes transform $S_{\rho_N}(z)$ is most accurately understood as a complex-valued random variable of our sample space. Indeed, since ρ_N is a random variable, and since the map $\mu \mapsto \int \frac{1}{x-z}\, d\mu$ is continuous by definition of the vague topology, we have that $S_{\rho_N}(z)$ as a function of the sample space is a composition of measurable functions and therefore defines a random variable. Furthermore, since $S_{\rho_N}(z)$ is uniformly bounded over the sample space by $|\operatorname{Im} z|^{-1}$, it has a well-defined expectation.

We will proceed from here by establishing a concentration inequality, after which we will use an LSI property to place a bound on

$$\mathbb{P}\left(|S_{\rho_N}(z) - \mathbf{E}\, S_{\rho_N}(z)| \geq \varepsilon\right)$$

given $\varepsilon > 0$. A further analysis of $\mathbf{E}\, S_{\rho_N}(z)$ will then establish convergence to the Stieltjes transform of $\mu_W(dx) = \frac{1}{2\pi}\sqrt{(4 - x^2)^+}\, dx$, the semicircular measure, and combining all results will complete a proof of the Gaussian case for Wigner's semicircle law.

Now, we modify a statement in [6] to obtain a concentration inequality as follows:

Lemma 3.11. *Let μ be an $LSI(a)$ probability measure of $\mathbb{R}^M$, and let $F : \mathbb{R}^M \to \mathbb{C}$ be a μ-integrable and Lipschitz continuous function. Then, for $r \geq 0$, we have*

$$\mu\left(|F - \int F\, d\mu| \geq r\right) \leq 4 \exp\left(-\frac{r^2}{2aLip(F)^2}\right)$$

Proof. First, suppose F is real-valued. Let $t \geq 0$. Since $r \mapsto e^{tr}$ is a nonnegative and nondecreasing function, we have

$$e^{tr}\mu\left(F - \int F\, d\mu \geq r\right) = \int e^{tr} \mathbb{1}_{\{F - \int F\, d\mu \geq r\}}\, d\mu \leq \int e^{t(F - \int F\, d\mu)}\, d\mu$$

Since μ is a probability measure, $F - \int F\, d\mu$ has an integral of 0 with respect to μ. As the sum of a constant and a Lipschitz continuous function, $F - \int F\, d\mu$ is Lipschitz continuous with

$$Lip\left(F - \int F\, d\mu\right) = Lip(F)$$

Thus, from Herbst's lemma, we get

$$\mu\left(F - \int F\, d\mu \geq r\right) \leq e^{-tr} \int e^{t(F - \int F\, d\mu)}\, d\mu \leq \exp\left(\frac{1}{4}at^2 Lip(F)^2 - tr\right)$$
$$= \exp\left(t\left(\frac{1}{4}aLip(F)^2 t - r\right)\right)$$

By optimizing this inequality at $t = \frac{2r}{aLip(F)^2}$, we get

$$\mu\left(F - \int F\,d\mu \geq r\right) \leq \exp\left(\frac{r^2}{aLip(F)^2} - \frac{2r^2}{aLip(F)^2}\right) = \exp\left(-\frac{r^2}{aLip(F)^2}\right)$$

Then, by replacing F with $-F$ in the above arguments, we find in the same way that

$$\mu\left(\int F\,d\mu - F \geq r\right) \leq \exp\left(-\frac{r^2}{aLip(F)^2}\right)$$

and this establishes the inequality

$$\mu\left(|F - \int F\,d\mu| \geq r\right) \leq 2\exp\left(-\frac{r^2}{aLip(F)^2}\right)$$

for when F is Lipschitz, μ-integrable, and real-valued.

Now, let F instead be complex-valued. By applying our result for the real case to the real and imaginary parts of $F - \int F\,d\mu$, we establish

$$\begin{aligned}
&\mu\left(|F - \int F\,d\mu| \geq r\right) \\
&\leq \mu\left(|\operatorname{Re} F - \int \operatorname{Re} F\,d\mu| \geq \frac{r}{\sqrt{2}}\right) + \mu\left(|\operatorname{Im} F - \int \operatorname{Im} F\,d\mu| \geq \frac{r}{\sqrt{2}}\right) \\
&\leq 4\exp\left(-\frac{r^2}{2aLip(F)^2}\right)
\end{aligned}$$

□

In establishing an LSI property for $S_{\rho_N}(z)$, we will make use of the following result stated in [6]:

Theorem 3.12. *Let γ be a probability measure of $\mathbb{R}^M$ with density $\gamma(dx) = e^{-V(x)}\,dx$ (with respect to Lebesgue measure), where $V(x) \in C^2(\mathbb{R}^M)$. Then, μ is $LSI(4c)$ if we have*

$$D^2V(x) \geq c^{-1}I$$

for all $x \in \mathbb{R}^M$.

Here, $D^2V(x)$ is the Hessian matrix, and $D^2V(x) \geq c^{-1}I$ denotes that

$$D^2V(x) - c^{-1}I$$

is positive-semidefinite, i.e.,

$$\langle v, \left(D^2V(x) - c^{-1}I\right)v\rangle \geq 0$$

for any M-dimensional vector v, where $\langle\cdot,\cdot\rangle$ denotes the inner product.

To carry out a proof of the above theorem, one uses the Markov semigroup of $L^2(\gamma)$ linear operators, which has the generator $L = \Delta - \nabla V \cdot \nabla$. Although we forego a proof here, the reader is encouraged to see the proof given in [6], which uses various facts and identities pertaining to Markov semigroups. To those interested in learning about these and related topics, we also recommend Joe Neeman's Notes on Markov Diffusions as a resource [4].

To prove the Lipschitz continuity property

$$Lip\left(A \mapsto \frac{1}{N}\,\mathbf{Tr}\,R_A(z)\right) \leq \frac{|\operatorname{Im} z|^{-2}}{\sqrt{N}}$$

for the Stieltjes transform of ρ_N, we used the standard $\mathbb{R}^{N^2}$ norm on the space of matrices. To make use of our LSI results in regard to $S_{\rho_N}(z)$, it will be fitting to consider an equivalent norm on the space of $N \times N$ self-adjoint matrices given by

$$\|(h_{ij})_{1\le i,j\le N}\|' = \left(\sum_{1\le i\le j\le N} |h_{ij}|^2\right)^{\frac{1}{2}}$$

In essence, we are allowing ourselves to interpret the real symmetric random matrix H_N as an $\mathbb{R}^{\frac{N(N+1)}{2}}$ random vector with no repeated components. This gives us the advantage of dealing only with entries that are independent as random variables. Note that this norm will be greater than or equal to $\frac{1}{2}$ times the standard matrix norm, so that our Lipschitz constant becomes

$$Lip\left(A \mapsto \frac{1}{N}\,\mathbf{Tr}\,R_A(z)\right) \le 2\frac{|\operatorname{Im} z|^{-2}}{\sqrt{N}}$$

Let $M = \frac{N(N+1)}{2}$, and relabel $H_N = (h_1, ..., h_M)$. With the above norm, the space of $N \times N$ real symmetric matrices is in isometry with $\mathbb{R}^M$, while for any M-dimensional rectangle

$$A_1 \times A_2 \times \cdots \times A_M$$

of open sets $A_i \subset \mathbb{R}$, we have

$$\begin{aligned}\mathbb{P}\left(H_N \in A_1 \times \cdots \times A_M\right) &= \prod_{i=1}^{M} \mathbb{P}\left(h_i \in A_i\right)\\ &= \frac{1}{(2\pi)^{\frac{M}{2}} \prod_{i=1}^{M} \sigma_i} \int_{A_1\times\cdots\times A_M} e^{-\frac{1}{2}\left(\frac{x_1}{\sigma_1}\right)^2 - \cdots - \frac{1}{2}\left(\frac{x_M}{\sigma_M}\right)^2}\, dx_1 \ldots dx_M\end{aligned}$$

where $\sigma_i \in \{\sqrt{\frac{1}{N}}, \sqrt{\frac{2}{N}}\}$ for each i. Since rectangles of the form $A_1 \times \cdots \times A_M$ are a π-system (i.e. are closed under finite intersection), and since they generate the Borel σ-algebra on $\mathbb{R}^M$ (see the argument used to prove Theorem 2.1), this allows us to conclude that the distribution of H_N on $\mathbb{R}^M$ (i.e. the pushforward measure $(H_N)_*(\mathbb{P})$) has a probability density of $\frac{1}{(2\pi)^{\frac{M}{2}} \prod_{i=1}^{M} \sigma_i} e^{-\left(\frac{1}{2}(\frac{x_1}{\sigma_1})^2 + \cdots + \frac{1}{2}(\frac{x_M}{\sigma_M})^2\right)}$ with respect to Lebesgue measure [2].

Pushing forward the probability $\mathbb{P}$, we have

$$\mathbf{E}\, S_{\rho_N}(z) := \int S_{\rho_N}(z)\, d\mathbb{P} = \int F\, d(H_N)_*(\mathbb{P})$$

when we denote $F(A) := \frac{1}{N}\,\mathbf{Tr}\,R_A(z)$ on the space of real symmetric matrices. Now, since the function

$$V(x) = \frac{1}{2}\left(\frac{x_1}{\sigma_1}\right)^2 + \cdots + \frac{1}{2}\left(\frac{x_M}{\sigma_M}\right)^2$$

satisfies $D^2V \ge (\frac{2}{N})^{-1} I$, we may apply Theorem 3.4 to establish that the measure $(H_N)_*(\mathbb{P})$ is $LSI(\frac{8}{N})$.

Using this LSI result and Lemma 3.5, we now have the bound

$$\mathbb{P}\left(|S_{\rho_N}(z) - \mathbf{E}\, S_{\rho_N}(z)| \ge \varepsilon\right) \le 4\exp\left(-\frac{1}{64} N^2 |\operatorname{Im} z|^4 \varepsilon^2\right)$$

for any $\varepsilon > 0$. Thus, given $z \in \mathbb{C} \setminus \mathbb{R}$, $|S_{\rho_N}(z) - \mathbf{E}\, S_{\rho_N}(z)|$ must converge in probability to 0. It is this bound that will allow our proof to be completed. Note that if we will be able to prove that $\mathbf{E}\, S_{\rho_N}(z)$ converges to the Stieltjes transform of μ_W, then since

$$|S_{\rho_N}(z) - S_{\mu_W}(z)| \leq |S_{\rho_N}(z) - \mathbf{E}\, S_{\rho_N}(z)| + |\,\mathbf{E}\, S_{\rho_N}(z) - S_{\mu_W}(z)|$$

implies $|S_{\rho_N}(z) - \mathbf{E}\, S_{\rho_N}(z)| \geq \frac{\varepsilon}{2}$ or $|\,\mathbf{E}\, S_{\rho_N}(z) - S_{\mu_W}(z)| \geq \frac{\varepsilon}{2}$ in the case that $|S_{\rho_N}(z) - S_{\mu_W}(z)| \geq \varepsilon$, we will have

$$\begin{aligned}
&\mathbb{P}\left(|S_{\rho_N}(z) - S_{\mu_W}(z)| \geq \varepsilon\right) \\
&\leq \mathbb{P}\left(|S_{\rho_N}(z) - \mathbf{E}\, S_{\rho_N}(z)| \geq \frac{\varepsilon}{2}\right) + \mathbb{P}\left(|\,\mathbf{E}\, S_{\rho_N}(z) - S_{\mu_W}(z)| \geq \frac{\varepsilon}{2}\right) \to 0
\end{aligned}$$

as $N \to \infty$, completing a proof of Wigner's semicircle law. It will suffice, then, to show convergence of $\mathbf{E}\, S_{\rho_N}(z)$ for all z in the upper half-plane and to then show that the convergence in question is indeed to the Stieltjes transform of the semicircular measure $\mu_W(dx) = \frac{1}{2\pi}\sqrt{(4-x^2)^+}\, dx$. We will follow the steps found in Rezakhalou's notes [6] to finish the proof.

Fix N and z. Let $s_N(z) = \mathbf{E}\, S_{\rho_N}(z)$. Denote $H_N = (h_{ij})_{1 \leq i,j \leq N}$, and denote

$$(r_{ij})_{1 \leq i,j \leq N} = R_{H_N}(z) = (H_N - z)^{-1}$$

The equality $(H_N - z) + z = H_N$ gives

$$(H_N - z)^{-1} + z^{-1} = z^{-1} H_N (H_N - z)^{-1}$$

Taking the normalized trace of both sides then gives

$$s_N(z) = -z^{-1} + z^{-1}\, \mathbf{E}\left(N^{-1}\, \mathbf{Tr}\, H_N (H_N - z)^{-1}\right) = -z^{-1} + z^{-1} N^{-1} \sum_{i,j} \mathbf{E}\, h_{ij} r_{ij}$$

We have used that H_N is symmetric. Note that $\mathbf{E}\, r_{ij}$ for each (i,j) exist:: each entry r_{ij} can be written as a sum of products of entries in $H_N - z$ divided by $\det(H_N - z)$ (whose absolute value is bounded below by $|\operatorname{Im} z|^N$), and since all moments of a Gaussian-distributed random variable exist, integrability of r_{ij} over the sample space follow from repeated application of the Cauchy-Schwarz inequality.

By term-wise product rule, we have the equality

$$\frac{\partial R_{H_N}(z)}{\partial h_{ij}}(H_N - z) + R_{H_N}(z)\frac{\partial H_N}{\partial h_{ij}} = 0$$

for each (i,j) due to $R_{H_N}(z)(H_N - z) = I$, so that

$$\frac{\partial R_{H_N}(z)}{\partial h_{ij}} = -R_{H_N}(z)\left(\mathbb{1}_{\{(i,j)\} \cup \{(j,i)\}}\right)_{k,\ell} R_{H_N}(z)$$

when we take $R_{H_N}(z)$ to be a function of the independent entries $(h_1, ..., h_M)$ of H_N. Meanwhile, by pushing forward and pulling back the probability $\mathbb{P}$, we can assume WLOG that $h_{ij} = h_1$ and write

$$\begin{aligned}
\mathbf{E}\, h_{ij} r_{ij} &= \frac{1}{(2\pi)^{\frac{M}{2}} \prod_{i=1}^M \sigma_i} \int h_1 r_{ij} e^{-\frac{1}{2}(\frac{h_1}{\sigma_1})^2} e^{-\frac{1}{2}(\frac{h_2}{\sigma_2})^2 - \cdots - \frac{1}{2}(\frac{h_M}{\sigma_M})^2}\, dh_1\, dh_2 \ldots dh_M \\
&= \frac{1}{(2\pi)^{\frac{M}{2}} \prod_{i=1}^M \sigma_i} \int \sigma_1^2 \frac{\partial r_{ij}}{\partial h_1} e^{-\frac{1}{2}(\frac{h_1}{\sigma_1})^2} e^{-\frac{1}{2}(\frac{h_2}{\sigma_2})^2 - \cdots - \frac{1}{2}(\frac{h_M}{\sigma_M})^2}\, dh_1\, dh_2 \ldots dh_M \\
&= \sigma_1^2\, \mathbf{E}\, \frac{\partial r_{ij}}{\partial h_{ij}}
\end{aligned}$$

Here, we have used integration by parts.

Thus, we may establish the equalities

$$\begin{aligned} s_N(z) &= -z^{-1} + z^{-1}N^{-1}\sum_{i,j}\mathbf{E}\, h_{ij}r_{ij} \\ &= -z^{-1} + z^{-1}N^{-1}\sum_{i\neq j} N^{-1}\,\mathbf{E}\,\frac{\partial r_{ij}}{\partial h_{ij}} + z^{-1}N^{-1}\sum_i 2N^{-1}\,\mathbf{E}\,\frac{\partial r_{ii}}{\partial h_{ii}} \\ &= -z^{-1} - z^{-1}N^{-2}\sum_{i,j}\mathbf{E}\left(r_{ii}r_{jj} + r_{ij}^2\right) \end{aligned}$$

Since the inverse of a symmetric matrix is symmetric, we have

$$\mathbf{E}\sum_{i,j} r_{ij}^2 = \mathbf{E}\,\mathbf{Tr}(H_N - z)^{-2}$$

Meanwhile, we have

$$\mathbf{E}\, N^{-2}\sum_{i,j} r_{ii}r_{jj} = \mathbf{E}\left(N^{-1}\,\mathbf{Tr}\, R_{H_N}(z)\right)^2 = \mathbf{E}\, S_{\rho_N}(z)^2$$

Our next step will be to write

$$s_N(z) = -z^{-1} - z^{-1}s_N(z)^2 + e_1 + e_2$$

with two error terms e_1 and e_2, such that

$$\begin{aligned} e_1 &= z^{-1}s_N(z)^2 - z^{-1}\,\mathbf{E}\, S_{\rho_N}(z)^2 = -z^{-1}\,\mathbf{E}\left(S_{\rho_N}(z) - s_N(z)\right)^2 \\ e_2 &= -z^{-1}N^{-2}\,\mathbf{E}\,\mathbf{Tr}(H_N - z)^{-2} \end{aligned}$$

For the first error term, we can use the rule

$$\mathbf{E}\,|X| = \int_0^\infty \mathbb{P}\left(|X| \geq t\right)\, dt$$

to see that

$$\begin{aligned} |e_1| &\leq |z^{-1}|\int_0^\infty \mathbb{P}\left(|S_{\rho_N}(z) - s_N(z)| \geq \sqrt{t}\right)\, dt \\ &\leq |z^{-1}|\int_0^\infty 4\exp\left(-\frac{1}{64}N^2|\operatorname{Im} z|^4 t\right)\, dt = |z^{-1}|256N^{-2}|\operatorname{Im} z|^{-4} \to 0 \end{aligned}$$

as $N \to \infty$. Meanwhile, by looking at the eigenvalues of $R_{H_N}(z)$, we see that

$$|e_2| \leq |z^{-1}|N^{-2}\cdot N|\operatorname{Im} z|^{-2} \to 0$$

as $N \to \infty$.

Thus, we have

$$s_N(z)^2 + zs_N(z) + 1 \to 0$$

Writing $s_N(z)^2 + zs_N(z) + 1 = c_N$, we have

$$s_N(z) = \frac{-z \pm \sqrt{z^2 - 4(1 - c_N)}}{2}$$

where $c_N \to 0$. If we enforce that $\operatorname{Im} z > 0$, then we see that

$$\operatorname{Im}\frac{-z - \sqrt{z^2 - 4(1 - c_N)}}{2} < 0$$

when we denote $\sqrt{z^2 - 4(1 - c_N)}$ to be the square root with positive imaginary part. In fact, since $\operatorname{Im} \frac{1}{\lambda_i - z} > 0$ for any real number λ_i when $\operatorname{Im} z > 0$, we must have $\operatorname{Im} s_N(z) > 0$ for all N, which reduces us to the case

$$s_N(z) = \frac{-z + \sqrt{z^2 - 4(1 - c_N)}}{2}$$

Therefore, we can conclude that

$$s_N(z) \to \frac{1}{2}\left(-z + \sqrt{z_2 - 4}\right)$$

as $N \to \infty$ when z lies in the upper half-plane, where $\operatorname{Im} \sqrt{z^2 - 4} > 0$.

It now remains for us to show that

$$\frac{1}{2}(-z + \sqrt{z^2 - 4})$$

gives the Stieltjes transform of

$$\mu_W(dx) = \frac{1}{2\pi}\sqrt{(4 - x^2)^+}\, dx$$

for z in the upper half-plane. Indeed, since the map $\frac{1}{x-z} \mapsto \lim_{N\to\infty} \mathbf{E} \int \frac{1}{x-z}\, d\rho_N$ is well-defined for all z in the upper half-plane, and since $f \mapsto \mathbf{E} \int f\, d\rho_N$ is a linear functional of $C_0(\mathbb{R})$ that is norm-bounded by 1 for each N, we have that

$$f \mapsto \lim_{N\to\infty} \mathbf{E} \int f\, d\rho_N$$

is a (norm-)bounded linear functional of the vector subspace of $C_0(\mathbb{R})$ consisting of all finite linear combinations of functions $\frac{1}{x-z}$ for z in the upper half-plane. By the analytic Hahn-Banach theorem, this can be extended to a linear functional defined on all of $C_0(\mathbb{R})$, so that by the Riesz representation theorem, there exists some measure whose Stieltjes transform is $\frac{1}{2}\left(-z + \sqrt{z^2 - 4}\right)$ on the upper half-plane. By Lemma 3.1, this measure is uniquely determined, and it now suffices for us to check the limit of

$$\frac{1}{\pi} \operatorname{Im}\left(\frac{-(x + i\varepsilon) + \sqrt{(x + i\varepsilon)^2 - 4}}{2}\right) dx$$

as $\varepsilon \to 0^+$ and to see that it is μ_W.

Letting $\sqrt{(x^2 - \varepsilon^2 - 4) + 2x\varepsilon i} = a + ib$ for real numbers a, b, we can write

$$a^2 - b^2 = x^2 - \varepsilon^2 - 4$$
$$2ab = 2x\varepsilon$$

As such, we have

$$b^4 + b^2(a^2 - b^2) - a^2b^2 = b^4 + (x^2 - \varepsilon^2 - 4)b^2 - x^2\varepsilon^2 = 0$$

so that

$$b^2 = \frac{-(x^2 - \varepsilon^2 - 4) + \sqrt{(x^2 - \varepsilon^2 - 4)^2 + 4x^2\varepsilon^2}}{2}$$

(Note that we can rule out $b^2 \neq \frac{-(x^2-\varepsilon^2-4)^2 - \sqrt{(x^2-\varepsilon^2-4)+4x^2\varepsilon^2}}{2}$, as $b^2 \geq 0$.) Since we have

$$\operatorname{Im}\left(\frac{-z + \sqrt{z^2 - 4}}{2}\right) > 0$$

we may now establish

$$b = \sqrt{\frac{-(x^2-\varepsilon^2-4)+\sqrt{(x^2-\varepsilon^2-4)^2+4x^2\varepsilon^2}}{2}}$$

To determine vague convergence of the measure $\frac{1}{\pi}\operatorname{Im}\left(\frac{-(x+i\varepsilon)+\sqrt{(x+i\varepsilon)^2-4}}{2}\right)dx$ as $\varepsilon \to 0^+$, it suffices for us to check convergence of the integral of any $f \in C_C(\mathbb{R})$. (Recall that $C_C(\mathbb{R})$ is dense in $C_0(\mathbb{R})$.) We now see that

$$\begin{aligned}\int f \frac{1}{\pi}\operatorname{Im}\left(\frac{-(x+i\varepsilon)+\sqrt{(x+i\varepsilon)^2-4}}{2}\right)dx \\ = \int f \frac{1}{2\pi}\left(-\varepsilon + \sqrt{\frac{-(x^2-\varepsilon^2-4)+\sqrt{(x^2-\varepsilon^2-4)^2+4x^2\varepsilon^2}}{2}}\right)dx\end{aligned}$$

We note that

$$\begin{aligned}\sqrt{\frac{-(x^2-\varepsilon^2-4)+\sqrt{(x^2-\varepsilon^2-4)^2+4x^2\varepsilon^2}}{2}} &\leq \sqrt{\frac{2|x^2-\varepsilon^2-4|}{2}} + \sqrt{\frac{\sqrt{4x^2\varepsilon^2}}{2}} \\ &= \sqrt{|x^2-\varepsilon^2-4|} + |x|\varepsilon\end{aligned}$$

so that the integrand is uniformly bounded (on the support of f) over ε near 0. For $x \in [-2,2]$, we have $4-x^2 \geq 0$, so that

$$\sqrt{\frac{-(x^2-\varepsilon^2-4)+\sqrt{(x^2-\varepsilon^2-4)^2+4x^2\varepsilon^2}}{2}} \to \sqrt{\frac{2(4-x^2)}{2}} = \sqrt{4-x^2}$$

as $\varepsilon \to 0^+$. For $|x| > 2$, we have $4-x^2 < 0$, so that

$$\sqrt{\frac{-(x^2-\varepsilon^2-4)+\sqrt{(x^2-\varepsilon^2-4)^2+4x^2\varepsilon^2}}{2}} \to 0$$

as $\varepsilon \to 0^+$. Thus, we may apply the dominated convergence theorem and conclude

$$\int f \frac{1}{\pi}\operatorname{Im}\left(\frac{-(x+i\varepsilon)+\sqrt{(x+i\varepsilon)^2-4}}{2}\right)dx \to \int f \frac{1}{2\pi}\sqrt{(4-x^2)^+}\,dx$$

as $\varepsilon \to 0^+$.

We now conclude that the random measures

$$\rho_N = \frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i}$$

converge vaguely in probability to

$$\mu_W(dx) = \frac{1}{2\pi}\sqrt{(4-x^2)^+}\,dx$$

as $N \to \infty$. The analytic proof of Wigner's semicircle law is complete.

We have now proven Wigner's semicircle law for vague convergence in probability. (A stronger notion of convergence is *almost-sure weak convergence* of the random measures ρ_N to μ_W, and we will see in the coming sections that this holds too.)

4. Wigner's Semicircle Law: A Combinatorial Approach

In this section, we will provide a combinatorial perspective for proving the Wigner Semicircle Law. Once we establish the connections between certain combinatorial objects, pairings and graphs for example, and the expectation of the moments of the empirical measures, we can fully utilize the knowledge we have about the structures and the properties of these combinatorial objects and transfer them into useful information about the expectation.

For the purpose of proving the almost sure weak convergence of the empirical measures of the eigenvalues to the semicircular measure, we will divide our proofs into three main parts.

In the first step, we will introduce various combinatorial objects as a means to prove the convergence of the expectations of all the moments of the empirical measure to all the moments of the semicircular measure, i.e., for all $m \in \mathbb{N}$,

$$\mathbb{E}[\frac{1}{N}\sum_{i=1}^{N}\lambda_i^m] \to \int_{\mathbb{R}} x^m d\mu_W(x)$$

where $\lambda_1, \cdots, \lambda_N$ are eigenvalues of A_N and μ_W is the semicircular measure. Here, we will enlarge the scope of $N \times N$ random matrices A_N to not only Gaussian Unitary Ensembles but also Wigner matrices.

Our second step is then to show that for certain types of probability measures, the convergence of all the moments implies the weak convergence of measures. In particular, the semicircular measure satisfies that certain criteria.

Because the empirical measures of the eigenvalues are not deterministic but random, we need to bridge the convergence of moments in expectation to the convergence of moments in probability in the last step. Here, we will again appeal to the graph theory to show that the variances of the moments are indeed bounded. We therefore use Chebyshev's inequality and the Borel-Cantelli lemma to prove the almost sure convergence.

4.1. **Convergence of Moments in Expectation.**

4.1.1. *Catalan Numbers and the Semicircle measure.*

Given the explicit expression of the semicircular measure, it is not difficult to directly compute its moments. However, it is rather non-trivial to realize the recursive structure of the moments. To put it more straightforwardly, we will introduce the Catalan numbers and define such recursive structure explicitly and show that the moments of semicircular measure correspond exactly to the Catalan numbers.

Definition 4.1. The Catalan numbers $(C_k)_{k\geq 0}$ are given by

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

with $C_0 = 1$.

Lemma 4.2. *For $k \geq 1$ and $C_0 = 1$:*

$$C_k = \sum_{l=0}^{k-1} C_l C_{k-l-1}.$$

Therefore, Catalan numbers are uniquely determined by the recursion above and the initial value $C_0 = 1$. Additionally, the moments of semicircular measure μ_W are exactly given by the Catalan numbers:

$$\frac{1}{2\pi}\int_{-2}^{2} x^n\sqrt{4-x^2}dx = \begin{cases} 0, & n = 2k+1 \\ C_k, & n = 2k \end{cases}$$

Proof. See Appendix A.2. □

4.1.2. *Wick Formula.*

As we have emphasized in the previous section, the moments of the Semicircular measure have a recursive structure represented through the Catalan numbers. In this section, we will see that the moments of Gaussian random variables, which are closely linked to the expectation of moments of the eigenvalues' empirical measure, also encompass the same recursive structure represented through the pairings. Particularly, we shall use the concept of pairings to prove the Wick Formula. It will greatly facilitate our computation of the expectation of moments of empirical measures for Gaussian Unitary Ensembles.

Proposition 4.3. *Let X be a standard Gaussian random variable. The moments of X are of the form*

$$\mathbb{E}[X^n] = \frac{1}{2\pi}\int_{\mathbb{R}} t^n e^{-\frac{t^2}{2}}dt = \begin{cases} 0, & n \text{ odd}, \\ (n-1)!!, & n \text{ even} \end{cases}$$

where $(2n-1)!! = 1\cdot 3\cdots(2n-1)$.

Proof. The odd case is obvious since the Gaussian distribution is symmetric. The even case can be proved by induction and integration by part. A.3. □

The moments of the standard Gaussian random variable actually correspond to the pairing of sets which we define as follows:

Definition 4.4. For any natural number $n \in \mathbb{N}$, we let $[n] = \{1, \cdots, n\}$. A pairing π of $[n]$ is a decomposition of $[n]$ into disjoint subsets of size 2. The set of all pairings of $[n]$ is denoted by

$$P_2(n) = \{\pi | \pi \text{ is a pairing of } [n]\}.$$

Proposition 4.5. *We have*

$$\#P_2(n) = \begin{cases} 0, & n \text{ odd}, \\ (n-1)!!, & n \text{ even}. \end{cases}$$

Thus, for a standard Gaussian variable X, we have $\mathbb{E}[X^n] = \#P_2(n)$.

Proof. To prove the proposition, we need to count elements in $\#P_2(n)$ recursively. For any element $\pi \in \#P_2(n)$, it must pair the first element in $[n]$ with one of the other $n-1$ elements. Then we can view π as an element in $\#P_2(n-2)$ for the other $n-2$ elements. Hence we have

$$\#P_2(n) = (n-1)\cdot\#P_2(n-2).$$

Along with the initial condition $\#P_2(2) = 1$, this proves our proposition. □

Remark 4.6. For independent Gaussian random variables X and Y, we have that

$$\mathbb{E}[X^n Y^m] = \mathbb{E}[X^n] \cdot \mathbb{E}[Y^m].$$

This gives us a combinatorial description for the mixed moments:

$$\begin{aligned}\mathbb{E}[X^n Y^m] &= \mathbb{E}[X^n] \cdot \mathbb{E}[Y^m] \\ &= \#\{\text{pairings of } X \cdots X\} \cdot \#\{\text{pairings of } Y \cdots Y\} \\ &= \#\{\text{pairings of } X \cdots XY \cdots Y \text{ which connect X with X and Y with Y}\}.\end{aligned}$$

Now consider $x_1, \cdots, x_n \in \{X, Y\}$. We still have

$$\mathbb{E}[x_1 \cdots x_n] = \#\{\text{pairings of } X \cdots XY \cdots Y \text{ which connect X with X and Y with Y}\}.$$

We can read the information whether $x_i = x_j$ or $x_i \neq x_j$ from the corresponding second moment

$$\mathbb{E}[x_i x_j] = \begin{cases} \mathbb{E}[x_i^2] = 1, & x_i = x_j \\ \mathbb{E}[x_i]\mathbb{E}[x_j], & x_i \neq x_j. \end{cases}$$

Thus we have,

$$\mathbb{E}[x_1 \cdots x_n] = \sum_{\pi \in P_2(n)} \prod_{(i,j) \in \pi} \mathbb{E}[x_i x_j].$$

We can generalize the two independent variables to finitely many independent variables and derive the Wick formula.

Theorem 4.7. *Let $Y_1, \cdots, Y_p$ be independent standard Gaussian random variables and consider $x_1, \cdots, x_n \in \{Y_1, \cdots, Y_p\}$. We have the following the Wick formula*

$$\mathbb{E}[x_1 \cdots x_n] = \sum_{\pi \in P_2(n)} \prod_{(i,j) \in \pi} \mathbb{E}[x_i x_j].$$

Remark 4.8. Note that this formula is linear in x_i hence it remains valid if we replace the x_i by linear combinations of the x_j. In particular we can generalize the Wick Formula to complex Gaussian variables.

Definition 4.9. A standard complex Gaussian random variable Z is of the form

$$Z = \frac{X + iY}{\sqrt{2}}$$

where X and Y are independent standard real Gaussian variables.

Remark 4.10. Let $Z = \frac{X+iY}{\sqrt{2}}$, then $\bar{Z} = \frac{X-iY}{\sqrt{2}}$. Through simple calculation, we get that $\mathbb{E}[Z] = \mathbb{E}[\bar{Z}] = 0$, $\mathbb{E}[Z^2] = \mathbb{E}[\bar{Z}^2] = 0$, and $\mathbb{E}[|Z^2|] = \mathbb{E}[Z\bar{Z}] = 1$. Hence, for $z_1, z_2 \in \{Z, \bar{Z}\}$ and π a pairing of z_1 and z_2, we have

$$\mathbb{E}[z_1 z_2] = \begin{cases} 1, & \pi \text{ connects } Z \text{ with } \bar{Z}, \\ 0, & \pi \text{ connects } Z \text{ with } Z \text{ or } \bar{Z} \text{ with } \bar{Z}. \end{cases}$$

Theorem 4.11. *Let $Z_1, \cdots, Z_p$ be independent standard complex Gaussian random variables and consider $z_1, \cdots, z_n \in \{Z_1, \bar{Z}_1, \cdots, Z_p, \bar{Z}_p\}$. Then we have the Wick formula*

$$\begin{aligned}\mathbb{E}[z_1, \cdots, z_n] &= \sum_{\pi \in P_2(n)} \mathbb{E}_\pi[z_1, \cdots, z_n] \\ &= \#\{\textit{pairings of } [n] \textit{ which connect } Z_i \textit{ with } \bar{Z}_i\}.\end{aligned}$$

4.1.3. *Gaussian Unitary Ensemble.*

Definition 4.12. A Gaussian Unitary Ensemble (GUE) is a random matrix of the form $A_N = \frac{1}{\sqrt{N}}(a_{ij})_{i,j=1}^N$ where

(1) $A_N = A_N^*$, i.e., $a_{ij} = \bar{a}_{ji}$ for all i, j.
(2) $\{a_{ij} | i \geq j\}$ are independent.
(3) a_{ij} is a standard Gaussian random variable, which is complex for $i \neq j$ and real for $i = j$. [9]

Now we can apply the Wick formula to calculate the expectation of m-th moments of the empirical measures of eigenvalues, which is the same as the trace of A_N^m divided by N. We will use $\mathrm{tr}(A_N^m)$ to denote the normalized trace, $\frac{1}{N}\mathbf{Tr}(A_N^m)$. In this case, we will also identify a pairing $\pi \in P_2(m)$ with a permutation $\pi \in S_m$ by

$$(k, l) \in \pi \leftrightarrow \pi(k) = l, \pi(l) = k.$$

Now we have

$$\begin{aligned}
\mathbb{E}[\mathrm{tr}(A_N^m)] &= \frac{1}{N^{m/2+1}} \sum_{i(1),\cdots,i(m)=1}^{N} \mathbb{E}[a_{i(1)i(2)} a_{i(2)i(3)} \cdots a_{i(m)i(1)}] \\
&= \frac{1}{N^{m/2+1}} \sum_{i(1),\cdots,i(m)=1}^{N} \sum_{\pi \in P_2(m)} \prod_{(k,l)\in\pi} \mathbb{E}[a_{i(k)i(k+1)} a_{i(l)i(l+1)}] \\
&= \frac{1}{N^{m/2+1}} \sum_{\pi \in P_2(m)} \sum_{i(1),\cdots,i(m)=1}^{N} \prod_k [i(k) = i(\pi(k)+1)].
\end{aligned}$$

The last equality is because $\prod_{(k,l)\in\pi} \mathbb{E}[a_{i(k)i(k+1)} a_{i(l)i(l+1)}] = 1$ if and only if $\mathbb{E}[a_{i(k)i(k+1)} a_{i(l)i(l+1)}] = 1$ for all $(k, l) \in \pi$ if and only if $i(k) = i(\pi(k) + 1)$ for all k. We can view $\pi(k) + 1$ as $\gamma\pi(k)$ where $\gamma = (1, 2, \cdots, m) \in S_m$. Hence, $\prod_k [i(k) = i(\pi(k) + 1)]$ will not vanish if and only if $i : [m] \to [N]$ is constant on the cycles of $\gamma\pi$.

Hence, letting $\#\gamma\pi$ denote the number of cycles in $\gamma\pi$, we get

$$\mathbb{E}[\mathrm{tr}(A_N^m)] = \sum_{\pi \in P_2(2)} N^{\#\gamma\pi - \frac{m}{2} - 1}.$$

Now the only obstacle is to compare the number of cycles in $\gamma\pi$ for each $\pi \in P_2(m)$ with $\frac{m}{2} - 1$. To do so, we will introduce this concept of non-crossing pairings.

4.1.4. *Non-crossing Pairings.*

Definition 4.13. A pairing $\pi \in P_2(m)$ is non-crossing (NC) if there are no pairs (i, k) and (j, l) in π with $i < j < k < l$. We define

$$NC_2(m) = \{\pi \in P_2(m) | \pi \text{ is non-crossing}\}.$$

Remark 4.14. The NC-pairings have a recursive structure.

(1) The pair of $\pi \in NC_2(2k)$ that contains 1 must have the form $(1, 2l)$ and the remaining pairs can only pair within $\{2, \cdots, 2l - 1\}$ or within $\{2l + 1, \cdots, 2l\}$.

(2) Repeating the process above, we know there must exist a pair of the form $(i, i+1)$ with $1 \leq i \leq 2k-1$. Removing this pair gives a NC-pairing of $2k-2$ points. Through iterated removal of neighboring pairs, NC-pairings can be reduced to the empty set.

Proposition 4.15. *Let m be an even integer and let $\pi \in P_2(m)$ which we identify with a permutation $\pi \in S_m$. As before, $\gamma = (1, 2, \cdots, m) \in S_m$. Then we have*

(1) $\#(\gamma\pi) - \frac{m}{2} - 1 \leq 0$ for all $\pi \in P_2(m)$.
(2) $\#(\gamma\pi) - \frac{m}{2} - 1 = 0$ if and only if $\pi \in NC_2(m)$.

Proof. To count the number of cycles in $\gamma\pi$, let's first look at the cycles of length 1 in $\gamma\pi$, i.e., fixed points of $\gamma\pi$. Suppose i is a fixed point of $\gamma\pi$, then $\pi(i) = \gamma^{-1}(i) = i-1$. Because π is a pairing, $(i-1, i) \in \pi$. Conversely, if $(i-1, i) \in \pi$, then i is a fixed point of $\gamma\pi$. Note that $i-1$ is not a fixed point in $\gamma\pi$, thus it must reside in some cycle with length at least 2. Consequently, we can remove neighboring pairs $(i-1, i)$ in π which correspond to reducing the number of cycles in $\gamma\pi$ by 1. For NC-pairings, we can repeat the above process until there is only one pair $(1, 2)$ left. In this case, both 1 and 2 are fixed points of $\gamma\pi$. Hence we know that for NC-pairings

$$\#(\gamma\pi) - \frac{m}{2} - 1 \leq 0 \text{ for all } \pi \in P_2(m).$$

On the other hand, for non NC-pairing $\pi \in P_2(m)$, we keep doing this process of removing neighboring cycles until it is no longer possible. Then for the remaining $\bar{\pi} \in P_2(k)$ for some $0 < k \leq m$, we know that $\gamma\bar{\pi}$ contain no more fixed points. Thus, all cycles have at least 2 elements.

$$\#(\gamma\pi) = \frac{m-k}{2} + \#(\gamma\bar{\pi}) \leq \frac{m-k}{2} + \frac{k}{2} < \frac{m}{2} + 1.$$

□

Now we can finally prove the averaged version of the semicircle law for GUE.

Theorem 4.16. *Let A_N be a GUE $N \times N$ random matrix. Then we have for all $m \in \mathbb{N}$:*

$$\lim_{N\to\infty} \mathbb{E}[\mathrm{tr}(A_N^m)] = \frac{1}{2\pi}\int_{-2}^{2} x^m \sqrt{4-x^2}dx.$$

Proof. When m is odd, both sides give zero. For $m = 2k$, we have

$$\lim_{N\to\infty} \mathbb{E}[\mathrm{tr}(A_N^m)] = \sum_{\pi\in P_2(2)} \lim_{N\to\infty} N^{\#\gamma\pi-k-1} = \sum_{\pi\in NC_2(2k)} 1 = \#NC_2(2k).$$

Since the moments of the semicircle are given by the Catalan numbers, it remains to show that $\#NC_2(2k) = C_k$. This is essentially due to the same recursive structure of NC-pairings and Catalan numbers. Let $d_k = \#NC_2(2k)$. We can identify $\pi \in NC_2(2k)$ with $\{(1, 2l)\} \cup \pi_0 \cup \pi_1$ where $l \in \{1, \cdots, k\}$ and $\pi_0 \in NC_2(2l-2)$ and $\pi_1 \in NC_2(2k-2l)$. Hence we have

$$d_k = \sum_{l=1}^{k} d_{l-1}d_{k-l} = \sum_{l=0}^{k-1} d_l d_{k-l-1}$$

where $d_0 = 1$. This is exactly the recursion for the Catalan numbers where $d_k = C_k$ for all $k \in \mathbb{N}$. □

In fact, the structure of the exact expectation is far richer than the structure of NC-pairings although NC-pairings are the only dominating factor in large scale limit. We can state the expectation of the trace more explicitly and such expression depends on the minimal genus, denoted by $g(\pi)$, of a surface on which the corresponding pairing π can be drawn without crossings.

$$\mathbb{E}[\mathrm{tr}(A_N^m)] = \sum_{\pi \in P_2(m)} N^{-2g(\pi)}$$

where NC-pairings have minimal genus $g = 0$. For readers who are interested in the connection between the expectation and the topological recursion, more details can be found in [8].

4.1.5. *Wigner Matrices.*

Not only does Wigner's Semicircle Law apply to the Gaussian Unitary Ensembles, but we could also prove it for the general symmetric Wigner matrices. The independent and identical distributions of entries will still be required, but we will relax the constraint on the Gaussian distribution and allow arbitrary distribution. Challenges are presented because the Wick formula could no longer be applied here, but it turns out that we can examine the problem under the lens of graph theory and prove it using rather straightforward techniques.

Definition 4.17. Let μ be a probability distribution on the real line. A Wigner random matrix with respect to μ is defined as $A_N = \frac{1}{\sqrt{N}}(a_{ij})_{i,j=1}^N$ where the following conditions are satisfied:

(1) $a_{ij} = a_{ji} \in \mathbb{R}$ for all i, j.
(2) $\{a_{ij} | i \geq j\}$ are independent.
(3) a_{ij} has distribution μ for all i, j.

We will further assume that all the moments of μ exist and the first moment vanishes. We will also normalize the second moment to 1. Here we have followed the definition from [8] where the Wigner matrix is real symmetric instead of complex Hermitian.

Now, for a Wigner matrix A_N which satisfy the above assumptions, we will show that

$$\mathbb{E}[\mathrm{tr}(A_N^m)] = \frac{1}{N^{1+\frac{m}{2}}} \sum_{i_1, \cdots, i_m = 1}^{N} \mathbb{E}[a_{i_1 i_2} \cdots a_{i_m i_1}] = \frac{1}{N^{1+\frac{m}{2}}} \sum_{\sigma \in P(m)} \sum_{\substack{i:[m] \to [N], \\ \ker\, i = \sigma}} \mathbb{E}[\sigma].$$

The first equality is given by unpacking the definition while the second equality requires a few definitions first.

Definition 4.18. A partition σ of $[n]$ is a decomposition of $[n]$ into disjoint, non-empty subsets of arbitrary size. The elements of σ which are subsets of $[n]$ are called blocks. We use $P(n)$ to denote the set of all partitions of $[n]$.

Definition 4.19. For a multi-index $i = (i_1, \cdots, i_m) \in [N]^m$, we can identify i with a function $i : [m] \to [N]$ via $i(k) = i_k$. Thus, we define the kernel of i as the following:

$$ker(i) = \{i^{-1}(1), \cdots, i^{-1}(N)\}$$

where empty sets are discarded.

It is easy to notice that for $i = (i_1, \cdots, i_m)$ and $j = (j_1, \cdots, j_m)$ where $ker(i) = ker(j)$, we have

$$\mathbb{E}[a_{i_1 i_2} \cdots a_{i_m i_1}] = \mathbb{E}[a_{j_1 j_2} \cdots a_{j_m j_1}].$$

For $ker(i) = \sigma$, let us denote the above value by

$$\mathbb{E}[\sigma] = \mathbb{E}[a_{i_1 i_2} \cdots a_{i_m i_1}].$$

As a result, we can rewrite the expectation as

$$\mathbb{E}[\mathrm{tr}(A_N^m)] = \frac{1}{N^{1+\frac{m}{2}}} \sum_{\sigma \in P(m)} \mathbb{E}[\sigma] \times \#\{i : [m] \to [N] | ker(i) = \sigma\}.$$

It is easy to see that

$$\#\{i : [m] \to [N] | ker(i) = \sigma\} = N(N-1) \cdots (N - \#\sigma + 1)$$

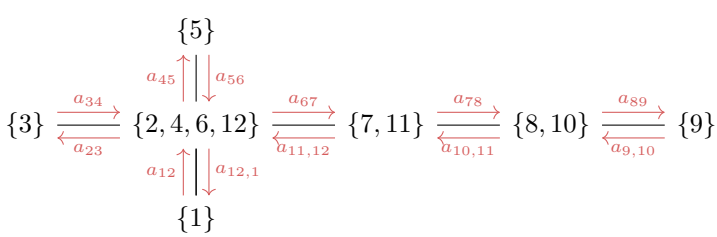
where $\#\sigma$ is the number of blocks in σ.

Thus it is only left to know the value of $\mathbb{E}[\sigma]$. To do so, we need to use graph theory and construct $\sigma \in P(m)$ as a graph G_σ.

Definition 4.20. Let $\sigma = \{V_1, \cdots, V_k\} \in P(m)$. Define the corresponding graph G_σ as follows. The vertices of G_σ are the blocks V_p of σ. There is an edge connecting V_p and V_q if there is an $r \in [m]$ such that $r \in V_p$ and $r + 1 (\mathrm{mod}\ m) \in V_q$.

Remark 4.21. G_σ is always connected. Another way of viewing G_σ is to start with a graph with $V = \{1, 2, \cdots, m\}$ and $E = \{(1,2), (2,3), \cdots, (m-1, m), (m, 1)\}$ and then identify vertices if they belong to the same blocks of σ. After erasing multiple edges, we will get G_σ.

Example 4.22. For example, let $\sigma = \{\{1\}, \{2,4,6,12\}, \{3\}, \{5\}, \{7,11\}, \{8,10\}, \{9\}\}$. Then G_σ can be drawn as the following



For $ker(i) = \sigma$, the term $a_{i_1 i_2} \cdots a_{i_m i_1}$ corresponds now to a walk in G_σ along the edges with steps

$$i_1 \to i_2 \to \cdots \to i_m \to i_1$$

which is represented in red arrows in the example above.

Different edges in G_σ correspond to independent random variables. By definition of G_σ, every edge will be traversed at least once in our walk. If one edge of the G_σ is only traversed once in the walk, then $\mathbb{E}[\sigma] = 0$ since the expectation will have a vanishing first moment factor. Therefore, for $\mathbb{E}[\sigma] \neq 0$, every edge must be traversed at least twice, i.e.

$$\#E \leq \frac{m}{2}.$$

where $\#E$ denotes the number of edges of G_σ.

Since we know that $\#\sigma = \#V$ where $\#V$ is the number of vertices of G_σ and the expression of $\#\{i : [m] \to [N] | ker(i) = \sigma\}$ is based on $\#V$, we need to connect

the dots between the number of vertices and the number of edges. We appeal to the well-known theorem of graph theory:

Theorem 4.23. *Let $G = (V, E)$ be a connected finite graph with vertices V and edges E (loops are allowed but multiple edges are not). Then we have*

$$\#V \leq \#E + 1.$$

The equality holds if and only if G is a tree, i.e., a connected graph without cycles.

Proof. A simple induction on the number of vertices suffices. □

Now we are ready to show the averaged version of the Wigner's semicircle law for Wigner matrices.

Theorem 4.24. *Let A_N be a Wigner matrix with respect to the distribution μ such that all the moments of μ exist. The first moment of μ is 0 and the second moment is 1. Then for all $m \in \mathbb{N}$, we have*

$$\lim_{N\to\infty} \mathbb{E}[\mathrm{tr}(A_N^m)] = \frac{1}{2\pi}\int_{-2}^{2} x^m \sqrt{4-x^2}dx.$$

Proof. We have already shown that

$$\begin{aligned}\lim_{N\to\infty} \mathbb{E}[\mathrm{tr}(A_N^m)] &= \lim_{N\to\infty} \sum_{\sigma\in P(m)} \mathbb{E}[\sigma] N(N-1)\cdots(N-\#V(G_\sigma)+1)N^{\frac{m}{2}-1} \\ &= \sum_{\sigma\in P(m)} \mathbb{E}[\sigma] \lim_{N\to\infty} N^{\#V(G_\sigma)-\frac{m}{2}-1}.\end{aligned}$$

where $\#V(G_\sigma)$ is the number of vertices of G_σ. For $\mathbb{E}[\sigma] \neq 0$, we need the following inequality to be satisfied

$$\#E(G_\sigma) \leq \frac{m}{2}$$

where $\#E(G_\sigma)$ is the number of edges of G_σ. By Theorem 4.23, this means that

$$\#V(G_\sigma) \leq \#E(G_\sigma) + 1 \leq \frac{m}{2} + 1.$$

For $N^{\#V(G_\sigma)-\frac{m}{2}-1}$ to be non-vanishing in the large N limit, we need $\#V(G_\sigma) - \frac{m}{2} - 1 = 0$. Because $P(m)$ is finite and $\mathbb{E}[\sigma]$ is finite for any $\sigma \in P(m)$, we know that $\mathbb{E}[\sigma]$ is bounded above for any $\sigma \in P(m)$. Thus the only non-vanishing terms in the sum are those terms that

$$\#V(G_\sigma) = \#E(G_\sigma) + 1 = \frac{m}{2} + 1. \tag{1}$$

The first equality implies that G_σ is a tree and the second equality implies that every edge is traversed exactly twice and in opposite directions (because the graph is a tree). Thus m must be even. Let $m = 2k$. For such $\sigma \in P(2k)$, we know that σ is a product of purely second moments and thus equals 1. Thus,

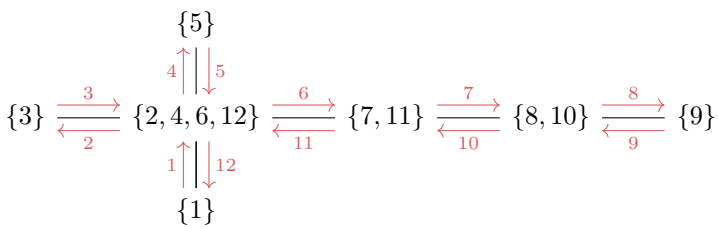
$$\lim_{N\to\infty} \mathbb{E}[\mathrm{tr}(A_N^m)] = \sum_{\substack{\sigma\in P(m)\\ G_\sigma \text{ is a tree}}} \mathbb{E}[\sigma] = \#\{\sigma \in P(m) | G_\sigma \text{ is a tree}, \#E(G_\sigma) = \frac{m}{2}\}.$$

Surprisingly, the number of $\sigma \in P(2k)$ such that G_σ is a tree and $\#E(G_\sigma) = k$ is exactly equal to the number of non-crossing pairings $\pi \in NC_2(2k)$ and therefore, the Catalan number C_k. □

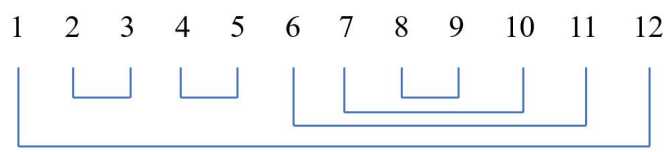
Lemma 4.25. *Let $A = \{\sigma \in P(2k) | G_\sigma$ is a tree, $\#E(G_\sigma) = k\}$ and $B = NC_2(2k)$. Then, there exists a bijection between A and B as sets.*

Proof. Although this proof is quite long, the basic idea behind is rather simple and can be well demonstrated through a picture. Consider the example we have stated before where $\sigma = \{\{1\}, \{2, 4, 6, 12\}, \{3\}, \{5\}, \{7, 11\}, \{8, 10\}, \{9\}\}$.

The corresponding tree can be drawn as the following:

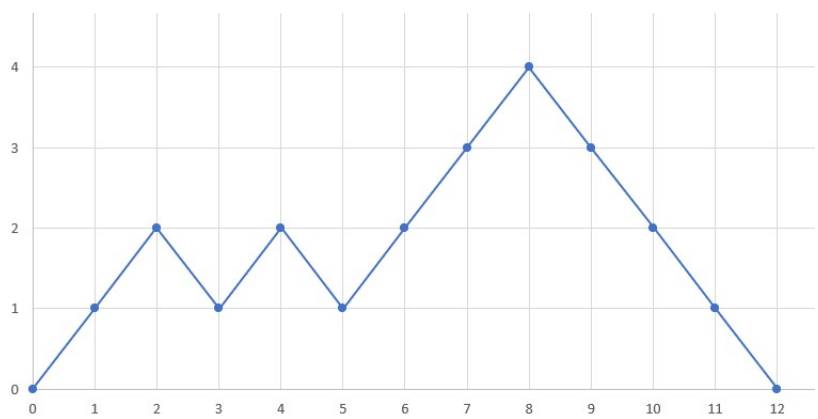


Note that instead of using a_{ij} to label the walk as above, I have used i not only for notation simplicity but also to connect back to the pairings. The tree will correspond to the following pairing of numbers.



Essentially, we will just pair the two numbers which correspond to the the walk on the same edge, but in opposite directions.

To convert the non-crossing pairing back to the tree, we shall first construct the walk as the following:



We can view the walk as a function $f : [0, 2k] \to [0, \infty)$ where it takes integer values to integer values (other values are given by linear interpolation) and $f(0) = f(2k) = 0$. The idea is that if i is connected to a number on its right, then $f(i) = f(i-1)+1$ and if i is connected to a number on its left, then $f(i) = f(i-1)-1$.

Now we can recover the tree from the walk by identifying all the points that can be connected by a horizontal line under the graph where the vertices are just the lattice points after identification.

Given the clear pictorial intuition, now the only work left is to actually construct the bijection. We will define maps $F : A \to B$ and $G : B \to A$ and show that they are inverses of each other.

Let $\sigma \in P(2k)$ such that G_σ is a tree and $\#E(G_\sigma) = k$. The latter condition means that for $i = (i_1, i_2, \cdots, i_{2k})$ such that $ker(i) = \sigma$, the walk

$$i_1 \to i_2 \to \cdots \to i_{2k} \to i_1$$

is a walk that traverse all the edges exactly twice and in opposite directions. We will label the edge connecting i_j and i_{j+1} as j for $1 \leq j \leq 2k-1$ and the edge connecting i_{2k} and i_1 as $2k$. Each edge will be labeled exactly twice. This gives a pairing $\pi_\sigma \in P_2(2k)$ associated to σ: For any $i, j \in [2k]$, they are paired to each other if they are the labels for the same edge. The pairing π_σ is non-crossing because of the fact that every tree has a leaf (a vertex that has only one edge) and removing a leaf (the vertex and the edge) from a tree does not destroy the tree structure. Let v be a leaf and e be the only edge connecting v. Then e must have labels j and $j+1$ for some $j \in [2k]$. We can remove this leaf (both v and e) from the graph G_σ and remove the pairing j and $j+1$ from π_σ. Repeating this process will show that π_σ is non-crossing. We define $F : A \to B$ by $F(\sigma) = \pi_\sigma$. Because the construction of the pairing does not use any information particular to i, it is independent of the multi-index as long as the kernel of the multi-index is σ and F is therefore well-defined.

On the other hand, given a non-crossing pairing $\pi \in P_2(2k)$, we now want to find a $\sigma \in P(2k)$ that corresponds to it. Let $x = (x_1, \cdots, x_{2k}, x_{2k+1})$. For any $i, j \in [2k]$ such that they are paired to each other in π, we identify x_i and x_{j+1}. If both x_i and x_k are identified with x_j, then we will consider them to all have the same value. If x_i is not identified with x_j, we will consider them to have different values. Because π is non-crossing, we will eventually identify x_1 with x_{2k+1}. Let $x_\pi = (x_1, \cdots, x_{2k})$ after the identification. We can then define $G : B \to A$ by $G(\pi) = ker(x_\pi)$. If we consider the graph $G = (V, E)$ where $V = \{1, 2, \cdots, 2k+1\}$ and $E = \{(1, 2), \cdots, (2k+1, 1)\}$ and identify vertices i and j if x_i is identified with x_j, this is exactly the graph $G_{ker(x_\pi)}$. We have k pairings in π and each pairing reduces the vertices of G by 1. Thus, $G_{ker(x_\pi)}$ has exactly $k+1$ vertices and is therefore a tree due to the argument above.

Now it is only left to show that $G \circ F(\sigma) = \sigma$ for all $\sigma \in A$ and $F \circ G(\pi) = \pi$ for all $\pi \in B$. Let $\sigma \in P(2k)$ be such that G_σ is a tree and $i = (i_1, \cdots, i_{2k})$ be such that $ker(i) = \sigma$. We will show that $ker(F(\sigma)) = ker(x_{\pi_\sigma}) = \sigma$. Note that m and n will only be paired in π_σ if they are the labels for the same edge, meaning that the edge connecting i_m and i_{m+1} is the same as the edge connecting i_{n+1} and i_n (Recall that each edge is traversed exactly twice and in opposite directions.) This means that m and n will only be paired if i_m is identified with i_{n+1} and i_n is identified with i_{m+1}. However, if m and n are paired, $x_{\pi_\sigma m}$ will be identified with $x_{\pi_\sigma n+1}$ and $x_{\pi_\sigma n}$ will be identified with $x_{\pi_\sigma m+1}$. Therefore, $ker(x_{\pi_\sigma}) = ker(i) = \sigma$. The argument for the other direction is essentially the same. □

4.2. **From Convergence of Moments to Convergence of Measures.**

As suggested by the section title, we will prove that for probability measures with all the moments finite and determined by their moments, the convergence of all the moments implies the convergence of measures.

Before anything else, we will define the characteristic function of a probability measure and use that to specify the criteria for a measure to be determined by its moments.

Definition 4.26. We define the characteristic function of a probability measure μ on the real line as

$$\varphi(t) = \int_{\mathbb{R}} e^{itx}\mu(dx) \text{ for } t \in \mathbb{R}.$$

Remark 4.27. There is a one-to-one correspondence between the characteristic functions and the cumulative distribution functions using the inversion formula (382, [2]). Thus, the characteristic function uniquely determines the distribution.

Definition 4.28. A probability measure μ on $\mathbb{R}$ is determined by its moments if

(1) all moments $\int t^k d\mu(t) < \infty$ exist ($k \in \mathbb{N}$).
(2) μ is the only probability measure with those moments: if ν is a probability measure and $\int t^k dv(t) = \int t^k d\mu(t)$ for all $k \in \mathbb{N}$, then $v = \mu$.

Theorem 4.29. *Let μ be a probability measure on $\mathbb{R}$ such that all of its moments exist, i.e., $\alpha_k = \int_{\mathbb{R}} x^k \mu(dx) \leq \infty$ for all $k \in \mathbb{N}$. If the power series $\sum_{k=0} \frac{\alpha_k r^k}{k!}$ has positive radius of convergence, then μ is determined by its moments.*

Proof. See [2] page 388. □

Corollary 4.30. *The semicircle measure μ_W is determined by its moments.*

Proof. We have already known the moments of μ_W:

$$\int_{-2}^{2} x^n \sqrt{4-x^2}dx = \begin{cases} 0, & n = 2k+1 \\ 2\pi C_k, & n = 2k. \end{cases}$$

Let $\alpha_n = \int_{-2}^{2} x^n \sqrt{4-x^2}dx$. It suffices to show $\sum_{n=0} \frac{\alpha_n r^n}{n!}$ has a positive radius of convergence. Since

$$\sum_{n=0} \frac{\alpha_n r^n}{n!} = 2\pi \sum_{n=0} \frac{C_n r^{2n}}{(2n)!} = 2\pi \sum_{n=0} \frac{r^n}{n!n!(n+1)},$$

the series clearly converges for $r < 1$. □

To prove that for the semicircular measure, the convergence in all moments implies the weak convergence of the measures, we still need some measure theoretical knowledge to justify the theorem.

Definition 4.31. Let μ and $(\mu_N)_N$ be probability measures on the real line. We say that μ_N converges weakly to μ, denoted by $\mu_N \xrightarrow{w} \mu$, if

$$\int_{\mathbb{R}} f(t)d\mu_N(t) \to \int_{\mathbb{R}} f(t)d\mu(t)$$

for all bounded continuous functions f on $\mathbb{R}$.

Lemma 4.32. *Let μ and $(\mu_N)_N$ be probability measures on $\mathbb{R}$, then $\mu_N \xrightarrow{w} \mu$ if and only if each subsequence of $(\mu_N)_N$ contains a further subsequence that converges to μ.*

Proof. Since the convergence of the whole sequence easily implies the convergence of subsequence, we only need to consider the other direction. We will prove it by a contrapositive argument. Assume that μ_N does not converge to μ. Then there exists some bounded continuous function f on $\mathbb{R}$ such that $\int_{\mathbb{R}} f d\mu_N$ does not converge to $\int_{\mathbb{R}} f d\mu$. This means that there exists a subsequence (μ_{N_k}) of (μ_N) such that $|\int_{\mathbb{R}} f d\mu_{N_k} - \int_{\mathbb{R}} f d\mu| > \epsilon$ for all $k \in \mathbb{N}$ and for some $\epsilon > 0$. Clearly, (μ_{N_k}) cannot contain any subsequence that converges to μ. □

Definition 4.33. A family of probability measures on $\mathbb{R}$, Γ, is tight if for any $\epsilon > 0$, there exists a compact set $K \subset \mathbb{R}$ such that $\mu(K) > 1 - \epsilon$ for all $\mu \in \Gamma$.

Now we are finally ready to go from convergence of moments to convergence of measures.

Theorem 4.34. *Let μ and $(\mu_N)_N$ be probability measures on $\mathbb{R}$ for which all moments exist. Assume that μ is determined by its moments. Assume furthermore that we have convergence of moments, i.e.,*

$$\lim_{N\to\infty} \int t^k d\mu_N(t) = \int t^k d\mu(t)$$

for all $k \in \mathbb{N}$. Then we have weak convergence $\mu_N \xrightarrow{w} \mu$.

Proof. Firstly, we will show that convergence of moments implies that (μ_N) is tight, which gives us a weakly convergent subseqeunce μ_{N_m} that converges to some ν by Prokhorov's Theorem (See [5]).

By Chebyshev's inequality, we have

$$\mu_N(\mathbb{R} \setminus [-R, R]) \leq \frac{\int t^2 d\mu_N(t)}{R^2}.$$

For any $\epsilon > 0$, we can find $M \in \mathbb{N}$ such that $|\int t^2 d\mu_m(t) - \int t^2 d\mu(t)| < 1$ for all $m \geq M$. Let $c = \max\{\int t^2 d\mu_1(t), \cdots, \int t^2 d\mu_{M-1}(t)\}$ and $d = \int t^2 d\mu(t)$. Choose $R \in \mathbb{R}$ such that $R^2 > \epsilon(c + d + 1)$. Then,

$$\mu_N(\mathbb{R} \setminus [-R, R]) < \epsilon.$$

for all $N \in \mathbb{N}$. Thus, (μ_N) is tight.

Also, convergence of moments implies that they are uniformly integrable, which implies that the moments of this subsequence also converge to the moments of ν.

To show the uniform integrability, we can choose random variables $\{X_N\}_{N\in\mathbb{N}}, X$ with a shared sample space such that X_N has the distribution μ_N, and X has the distribution ν. Note that because all moments converge, we have $\sup_N \int |x^k|^2 d\mu_N(x) = c < \infty$. thus $\mathbb{1}_{|X^k|>R}|X^k|R \leq |X^k|^2$, then

$$\int_{\mathbb{R}} \mathbb{1}_{|x^k|>R}|x^k| d\mu_N(x) \leq \frac{1}{R} \sup_N \int_{\mathbb{R}} |x^k|^2 d\mu_N(x) \leq \frac{c}{R}.$$

Because we can make R arbitrarily big, (X_N^k) are uniformly integrable for all $k \in \mathbb{N}$.

Then let μ_0 be the probability measure on the sample space. Let $\varepsilon > 0$. Choose $r > 0$ such that $\int_{\{|X_N^k|>r\}} |X_N^k| \, d\mu_0 < \frac{\varepsilon}{6}$ for all N, and $\int_{\{|X^k|>r\}} |X^k| \, d\mu_0 < \frac{\varepsilon}{6}$. Then, if f_r is the truncation function such that

$$f_r(t) = \begin{cases} t^k & \text{if } |t^k| \leq r \\ t^k \cdot \frac{r}{|t^k|} & \text{if } |t^k| > r, \end{cases}$$

we have

$$|\int X_N^k \, d\mu_0 - \int f(X_N) \, d\mu_0| \leq \int_{\{|X_N^k|>r\}} |X_N^k| \, d\mu_0 + \int_{\{|X_N^k|>r\}} r \, d\mu_0 < \frac{\varepsilon}{3}$$

The same goes when we substitute X_N with X.

Since $X_N \to X$ weakly, we have $\int f(X_N) \, d\mu_0 \to \int f(X) \, d\mu_0$. Thus, for all N large enough, we have $|\int X_N^k \, d\mu_0 - \int X^k \, d\mu_0| < \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon$. Thus, $\int X_N^k \, d\mu_0 \to \int X^k \, d\mu_0$, which means that

$$\int x^k \, d\mu_N \to \int x^k \, d\nu$$

so all moments converge.

But the moments of the subsequence also converge to moment of μ which means that μ and ν have the same moments. Since μ is determined by its moments, we have that $v = \mu$. Because we can repeat the above process for any subsequence of (μ_N), we know that each subsequence contains a further subsequence that converges weakly to μ. Applying Lemma 4.32, we get that the whole sequence must converge weakly to μ. □

4.3. **Almost Sure Weak Convergence of Measures.**

Finally, we can prove the Wigner Semicircle law in terms of almost sure weak convergence of measures. To do so, we will prove a variance bound for the moments of the empirical measures and use Chebyshev's inequality and the Borel-Cantelli lemma.

Proposition 4.35. *Let A_N be a Wigner matrix with respect to the distribution μ where all moments of μ exist. In particular, the first moment of μ vanishes and the second moment of μ is 1. Then,*

$$Var(\mathrm{tr}(A_N^m)) < \frac{B_m}{N^2}$$

where B_m is a constant that depends only on m.

Proof. Let's first rewrite the variance in terms of the entries of the matrix.

$$\begin{aligned} Var(\mathrm{tr}(A_N^m)) &= \mathbb{E}[\mathrm{tr}(A_N^m)^2] - (\mathbb{E}[\mathrm{tr}(A_N^m)])^2 \\ &= N^{-2-m}\mathbb{E}[\sum_{i\in[n]^k} Y_i \sum_{j\in[n]^k} Y_j] - N^{-2-m}\mathbb{E}[\sum_{i\in[n]^k} Y_i]^2 \\ &= N^{-2-m} \sum_{i,j\in[n]^k} [\mathbb{E}(Y_iY_j) - \mathbb{E}(Y_i)\mathbb{E}(Y_j)]. \end{aligned}$$

Both i and j are multi-indexes, i.e. $i = (i_1, \cdots, i_m)$ and $j = (j_1, \cdots, j_m)$. We use Y_i to denote $a_{i_1i_2} \cdots a_{i_mi_1}$. Let $G_{i,j} = G_{ker(i)} \cup G_{ker(j)}$. By unions of graphs, we mean that the vertices of the new graph are the union of the vertices of the existing graphs and the edges of the new graph are the union of the edges of the existing graphs. Let w_i and w_j, respectively, denote the walks on $G_{i,j}$ as

$$i_1 \to i_2 \to \cdots \to i_m \to i_1,$$

$$j_1 \to j_2 \to \cdots \to j_m \to j_1.$$

The tuple $(G_{i,j}, w_i, w_j)$ completely determines the value of $\mathbb{E}[Y_iY_j]$ and since we can recover $G_{ker(i)}$ and $G_{ker(j)}$ from $(G_{i,j}, w_i, w_j)$, the value $\mathbb{E}[Y_i]\mathbb{E}[Y_j]$ is also

determined by $(G_{i,j}, w_i, w_j)$. Let us denote the value $\mathbb{E}[Y_iY_j] - \mathbb{E}[Y_i]\mathbb{E}[Y_j]$ by $\pi(G_{i,j}, w_i, w_j)$.

Since $G_{ker(i)}$ and $G_{ker(j)}$ are all connected, the only way that $G_{i,j}$ is not connected is that $\{i_1, \cdots, i_m\} \cap \{j_1, \cdots, j_m\} = \varnothing$. In this case, Y_i and Y_j are therefore independent, the term $\mathbb{E}[Y_iY_j] - \mathbb{E}[Y_i]\mathbb{E}[Y_j]$ vanishes. Therefore, we only need to consider the connected graphs. Let's use $G_{m,m}$ to denote the set of connected graphs where the number of vertices less or equal to $2m$, together with two walks each of length m whose union covers the graph.

We can now rewrite the variance as

$$Var(\mathrm{tr}(A_N^m))$$
$$= \frac{1}{N^{2+m}} \sum_{(G,w_1,w_2)\in G_{m,m}} \pi(G, w_1, w_2) \times \#\{(i,j) \in [n]^{2m} | (G_{i,j}, w_i, w_j) = (G, w_1, w_2)\}.$$

Let $E^s_{i,j}$ denote the set of self-connecting edges in $G_{i,j}$ and $E^c_{i,j}$ denote the set of edges connecting two different vertices in $G_{i,j}$. Similarly, E^s_i and E^c_i for sets of self-connecting edges and other edges in $G_{ker(i)}$. Furthermore, for each edge e in $G_{i,j}$, let $d(e)$ denotes the number of times e is traversed by either w_i or w_j. For edge e in $G_{ker(i)}$, $d(e)$ means the number of times e is traversed by w_i. Now

$$\begin{aligned}\mathbb{E}[Y_iY_j] - \mathbb{E}[Y_i]\mathbb{E}[Y_j] = &\prod_{e\in E^s_{i,j}} \mathbb{E}[a_{11}^{d(e)}] \prod_{e\in E^c_{i,j}} \mathbb{E}[a_{12}^{d(e)}] \\ &- \prod_{e\in E^s_i} \mathbb{E}[a_{11}^{d(e)}] \prod_{e\in E^c_i} \mathbb{E}[a_{12}^{d(e)}] \prod_{e\in E^s_j} \mathbb{E}[a_{11}^{d(e)}] \prod_{e\in E^c_j} \mathbb{E}[a_{12}^{d(e)}].\end{aligned}$$

The key observation is that

$$\sum_{e\in G_{i,j}} d(e) = \sum_{e\in G_{ker(i)}} d(e) + \sum_{e\in G_{ker(j)}} d(e) = 2m.$$

The sum of all exponents in $\mathbb{E}[Y_iY_j]$ or $\mathbb{E}[Y_i]\mathbb{E}[Y_j]$ is exactly $2m$. Thus, the value of $\mathbb{E}[Y_iY_j] - \mathbb{E}[Y_i]\mathbb{E}[Y_j]$ is completely determined by the integer partitions of $2m$ that we assign to each exponent. Since there are only finitely many such integer partitions of $2m$, we can find a constant C_m that only depends on m but not N such that

$$\mathbb{E}[Y_iY_j] - \mathbb{E}[Y_i]\mathbb{E}[Y_j] \le C_m$$

for any $i, j \in [n]^m$.

For $\pi(G_{i,j}, w_i, w_j)$ to be non-vanishing, we need every edge of the graph $G_{i,j}$ to be traversed by w_i or w_j at least twice. By construction, every edge is traversed at least once by w_i or w_j. If one edge is traversed only once in $G_{i,j}$, then $\mathbb{E}[Y_iY_j] = 0$. It also implies that one edge in $G_{ker(i)}$ or $G_{ker(j)}$ is traversed only once and hence the product $\mathbb{E}[Y_i]\mathbb{E}[Y_j] = 0$. If each edge is traversed at least twice in $G_{i,j}$ by w_i or w_j, it means that $G_{i,j}$ could only have at most m edges and at most $m+1$ vertices by Theorem 4.23. We will further show that for $\pi(G_{i,j}, w_i, w_j)$ to be non-vanishing, the number of vertices of $G_{i,j}$ cannot be equal to $m+1$. If so, equation 1 must hold and $G_{i,j}$ must be a tree and every edge of $G_{i,j}$ must be traversed exactly twice. Because $G_{i,j}$ is a tree, $G_{ker(i)}$ and $G_{ker(j)}$ are also trees as sub-graphs of a tree. Thus each edge in $G_{ker(i)}$ or $G_{ker(j)}$ is traversed by w_i or w_j at least an even number of times since w_i and w_j must return to their starting points and there are no loops. This means that every edge of $G_{i,j}$ must be traversed by only w_i twice or w_j twice. If e is traversed by w_i once and w_j once, then e in $G_{ker(i)}$ is traversed an

odd number of times. Consequently, $G_{ker(i)}$ and $G_{ker(j)}$ share no edge in common, i.e.,

$$\{\{i_1,i_2\},\{i_2,i_3\},\cdots,\{i_m,i_1\}\}\cap\{\{j_1,j_2\},\{j_2,j_3\},\cdots,\{j_m,j_1\}\}=\varnothing.$$

Therefore, Y_i and Y_j are independent, hence $\pi(G_{i,j},w_i,w_j)=0$.

Now we can finally bound the variance:

$$\begin{aligned}
&Var(\mathrm{tr}(A_N^m))\\
&=\sum_{\substack{(G,w_1,w_2)\in G_{m,m}\\ \#V(G)\le m}}\pi(G,w_1,w_2)\frac{\#\{(i,j)\in[n]^{2m}|(G_{i,j},w_i,w_j)=(G,w_1,w_2)}{N^{2+m}}\\
&\le C_m\#G_{m,m}\frac{N(N-1)\cdots(N-m+1)}{N^{2+m}}\\
&\le\frac{C_m\#G_{m,m}}{N^2}
\end{aligned}$$

where $\#G_{m,m}$ and C_m both depends only on m and not on N. Let $B_m = C_m\#G_{m,m}$. We have therefore proven the proposition. □

Remark 4.36. The above proof also works for Gaussian Unitary Ensembles. Hence the same variance bound applies to GUE matrices as well.

Theorem 4.37. *(Wigner's semicircle law for Wigner and GUE matrices) Let A_N be a Wigner $N\times N$ matrix or a $N\times N$ GUE matrix. Let μ_N be the empirical measures of the eigenvalues of A_N. Then μ_N converges to μ_W weakly and almost surely.*

Proof. To show this theorem, we will mainly use the result from Theorem 4.34 and the Chebyshev's inequality. The only subtlety here is that the probability measures in Theorem 4.34 are deterministic instead of random. For random probability measures, we need to treat them as random variables that take values as probability measures, i.e. measurable functions from a general sample space Ω to the space of probability measures. Thus, for any $a\in\Omega$, $\mu_N(a)$ is a normal, deterministic probability measure. However, nothing in the following expression depends on a or A_N^m:

$$\int t^m\mathbb{E}[d\mu_N(a)](t)=\mathbb{E}[\int t^m(d\mu_N(a))(t)]=\mathbb{E}[\mathrm{tr}(A_N^m)]=\sum_{\pi\in P_2(m)}N^{\#\gamma\pi-\frac{m}{2}-1}.$$

In other words, the convergence of $\int t^m\mathbb{E}[d\mu_N(a)](t)$ to $\int t^m(d\mu_W(a))(t)=\int t^m d\mu_W(t)$ holds for all $a\in\Omega$. Now we can apply Theorem 4.34 and get that $\mathbb{E}[\mu_N(a)]$ converges to μ_W weakly for all $a\in\Omega$. We have shown above that this convergence implies the convergence of $\int fd\mathbb{E}[\mu_N(a)]$ to $\int fd\mu_W$ for all bounded, continuous functions f and all $a\in\Omega$. Therefore, we know that for bounded, continuous functions f

$$\begin{aligned}
&\mathbb{P}\left\{a\in\Omega|\left|\int fd\mu_N(a)-\int fd\mu_W\right|>\epsilon\right\}\\
&\le\mathbb{P}\left\{a\in\Omega|\left|\int fd\mu_N(a)-\int f\mathbb{E}[d\mu_N(a)]\right|>\frac{\epsilon}{2}\right\}\\
&+\mathbb{P}\left\{a\in\Omega|\left|\int fd\mu_W-\int f\mathbb{E}[d\mu_N(a)]\right|>\frac{\epsilon}{2}\right\}
\end{aligned}$$

Applying Proposition 4.35 and Chebyshev's inequality, we can bound the first term as following:

$$\mathbb{P}\left\{a \in \Omega \middle| \left|\int f d\mu_N(a) - \int f \mathbb{E}[d\mu_N(a)]\right| > \frac{\epsilon}{2}\right\} \leq \frac{4B_m}{N^2\epsilon^2}.$$

For the second term, the convergence of $\int f d\mathbb{E}[\mu_N(a)] \to \int f d\mu_W$ does not depend on $a \in \Omega$. There exists a $M \in \mathbb{N}$ such that $\left|\int f d\mu_W - \int f \mathbb{E}[d\mu_N(a)]\right| \leq \frac{\epsilon}{2}$ for all a and $N \geq M$, thus

$$\sum_{N=1}^{\infty} \mathbb{P}\left\{a \in \Omega \middle| \left|\int f d\mu_W - \int f \mathbb{E}[d\mu_N(a)]\right| > \frac{\epsilon}{2}\right\} \leq M < \infty.$$

We then can apply the Borel-Cantelli lemma since

$$\sum_{N=1}^{\infty} \mathbb{P}\left\{a \in \Omega \middle| \left|\int f d\mu_N(a) - \int f d\mu_W\right| > \epsilon\right\} < \infty$$

and conclude that the convergence of $\int f d\mu_N \to \int f d\mu_W$ is almost surely. Consequently, the convergence of $\mu_N \to \mu_W$ is also almost surely.

□

5. Acknowledgement

First of all, we would like to thank Professor Peter May for making everything possible during such a difficult period in the pandemic. We would like to thank Wenqi Li and Samanda Hu for their kind help and support as always. We would like also to thank Fraydoun Rezakhanlou, Roland Speicher, and Terence Tao whose texts we learned from and adapted to write this paper. Lastly, it's our pleasure to thank our mentor, Professor Alisa Knizel, for being so inspiring, patient and encouraging throughout the program. This project would never have been possible without her.

References

[1] Richard F. Bass. *Real Analysis for Graduate Students, Version 4.2.* 2020.
[2] Patrick Billingsley. *Probability and Measure.* John Wiley and Sons, second edition, 1986.
[3] H. Brezis. *Functional Analysis, Sobolev Spaces and Partial Differential Equations.* Universitext. Springer New York, 2010.
[4] Joe Neeman. Notes on markov diffusions, 2016.
[5] K. Parthasarathy. Probability measure on metric spaces. *Journal of the American Statistical Association*, 63, 09 1968.
[6] Fraydoun Rezakhanlou. Lectures on random matrices - university of california, berkeley, 2012.
[7] Walter Rudin. *Real and Complex Analysis, 3rd Ed.* McGraw-Hill, Inc., USA, 1987.
[8] Roland Speicher. Lecture notes on "random matrices", 2020.
[9] Terence Tao. *Topics in random matrix theory.* 2011.

Appendix A.

Lemma A.1. *For any probability measure μ on $\mathbb{R}$, there exists only countably many singletons as subsets of $\mathbb{R}$ that have non-zero measures.*

Proof. Let $A_n = \{a \in \mathbb{R} | \mu(\{a\}) > \frac{1}{n}\}$. Because μ is a probability measure, the cardinality of A_n must be bounded by n. Thus, $S = \bigcup_{n=1}^{\infty} A_n$ is countable. □

Lemma A.2. *For $k \geq 1$ and $C_0 = 1$:*

$$C_k = \sum_{l=0}^{k-1} C_l C_{k-l-1}.$$

Therefore Catalan numbers are uniquely determined by the recursion above and the initial value $C_0 = 1$. Additionally, the moments of semicircular measure μ_W are exactly given by the Catalan numbers:

$$\frac{1}{2\pi}\int_{-2}^{2} x^n \sqrt{4-x^2}dx = \begin{cases} 0, & n = 2k+1 \\ C_k, & n = 2k \end{cases}$$

Proof. Let's first define c_k for all k by the recursion and the initial condition that $c_0 = 1$.

$$c_k = \sum_{l=0}^{k-1} c_l c_{k-1-l}$$

These numbers c_k are uniquely defined by the recursion and the initial condition. We will show that these numbers are exactly the Catalan numbers.

First, let's consider the generating function

$$f(x) = \sum_{k=0}^{\infty} c_k x^k.$$

Observe the following relation:

$$f(x) = 1 + xf(x)^2.$$

This is because

$$\begin{aligned} 1 + xf(x)^2 &= 1 + \sum_{k=0}^{\infty}[\sum_{l=0}^{k} c_l c_{k-l}]x^{k+1} \\ &= 1 + \sum_{k=1}^{\infty}[\sum_{l=0}^{k-1} c_l c_{k-l-1}]x^k \\ &= 1 + \sum_{k=1}^{\infty} c_k x^k \\ &= f(x). \end{aligned}$$

Therefore, by quadratic formula we know that $f(x) = \frac{1 \pm \sqrt{1-4x}}{2x}$. However,

$$C_0 = \lim_{x \to 0} f(x) = \lim_{x \to 0} \frac{\pm 2(1-4x)^{-1/2}}{2} = 1,$$

which forces us to choose the negative square root. Hence, $f(x) = \frac{1-\sqrt{1-4x}}{2x}$. To calculate the coefficient of each term in the power series, we need to apply the binomial theorem to $\sqrt{1-4x}$:

$$\begin{aligned}
(1-4x)^{1/2} &= 1 + \sum_{n\geq 1} \binom{1/2}{n} (-4x)^n \\
&= 1 + \sum_{n\geq 1} \frac{(\frac{1}{2})(-\frac{1}{2})(-\frac{3}{2})\cdots(-\frac{2n-3}{2})}{n!}(-4x)^n \\
&= 1 + \sum_{n\geq 1} (-1)^{n-1} \frac{(2n-3)!!}{2^n n!}(-4x)^n \\
&= 1 - \sum_{n\geq 1} \frac{2^n (2n-3)!!}{n!} x^n \\
&= 1 - 2\sum_{n\geq 1} \frac{2^{n-1}(n-1)! \prod_{k=1}^{n-1}(2k-1)}{n(n-1)!(n-1)!} x^n \\
&= 1 - 2\sum_{n\geq 1} \frac{\prod k = 1^{n-1}(2k) \prod_{k=1}^{n-1}(2k-1)}{n(n-1)!(n-1)!} x^n \\
&= 1 - 2\sum_{n\geq 1} \frac{(2n-2)!}{n(n-1)!(n-1)!} x^n \\
&= 1 - 2\sum_{n\geq 1} \frac{1}{n}\binom{2n-2}{n-1} x^n.
\end{aligned}$$

Thus,

$$\begin{aligned}
f(x) &= \frac{1 - 1 + 2\sum_{n\geq 1} \frac{1}{n}\binom{2n-2}{n-1} x^n}{2x} \\
&= \sum_{n\geq 1} \frac{1}{n}\binom{2n-2}{n-1} x^{n-1} \\
&= \sum_{n\geq 0} \frac{1}{n+1}\binom{2n}{n} x^n.
\end{aligned}$$

and we can conclude that $c_n = \frac{1}{n+1}\binom{2n}{n}$.

Secondly,

$$\frac{1}{2\pi}\int_{-2}^{2} \sqrt{4-x^2}\,dx = 1$$

can be easily shown by calculating the area of a semicircle. Moreover, the fact that odd moments of the measure vanish is also obvious because

$$\begin{aligned}
\frac{1}{2\pi}\int_{-2}^{2} x^{2k+1}\sqrt{4-x^2}\,dx &= \frac{1}{2\pi}\int_{-2}^{0} x^{2k+1}\sqrt{4-x^2}\,dx + \frac{1}{2\pi}\int_{0}^{2} x^{2k+1}\sqrt{4-x^2}\,dx \\
&= \frac{1}{2\pi}\int_{-2}^{0} x^{2k+1}\sqrt{4-x^2}\,dx - \frac{1}{2\pi}\int_{-2}^{0} x^{2k+1}\sqrt{4-x^2}\,dx \\
&= 0
\end{aligned}$$

Now, let's show that

$$\frac{1}{2\pi}\int_{-2}^{2} x^{2k}\sqrt{4-x^2}dx = C_k.$$

for all $k \in \mathbb{Z}$.

We will show it by induction where the base case is discussed above. But before we proceed, let's make a simple observation of the Catalan numbers:

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

$$C_{k+1} = \frac{4k+2}{k+2}C_k.$$

Now we assume that

$$C_k = \frac{1}{2\pi}\int_{-2}^{2} x^{2k}\sqrt{4-x^2}dx,$$

and we will show that

$$\frac{1}{2\pi}\int_{-2}^{2} x^{2k+2}\sqrt{4-x^2}dx = \frac{4k+2}{k+2}C_k = C_{k+1}$$

We will first do a change of variables. Let $x = 2\sin(u); dx = 2\cos(u)$. Then,

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

where we have used an integral reduction for the third equality: for any integer $n \geq 2$, we have that

$$\int \sin^n x dx = -\frac{1}{n}\sin^{n-1}x\cos x + \frac{n-1}{n}\int \sin^{n-2}x dx.$$

This can be shown by integration by parts and since we are evaluating at $\frac{\pi}{2}$ and $-\frac{\pi}{2}$ where the cosine vanishes, we will only consider the second term here.

Let $u = \sin^{n-1}x$ and $v' = \sin x$. Then $u' = (n-1)\sin^{n-2}x\cos x$ and $v = -\cos x$. We have

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

Simple re-arrangement will give us the reduction formula.

Now we can finally conclude that

$$\begin{aligned}
&\frac{1}{2\pi}\int_{-\pi/2}^{\pi/2} 2^{2k+2}\sin^{2k+2}(u)4\cos^2(u)du \\
&= \frac{2^{2k+2}}{2\pi}[\int_{-\pi/2}^{\pi/2}\sin^{2k+2}(u)du - \int_{-\pi/2}^{\pi/2}\sin^{2k+4}(u)du] \\
&= \frac{2^{2k+4}}{2\pi}[\int_{-\pi/2}^{\pi/2}\sin^{2k+2}(u)du - \frac{2k+3}{2k+4}\int_{-\pi/2}^{\pi/2}\sin^{2k+4}(u)du] \\
&= \frac{2^{2k+4}}{2\pi}\frac{1}{2k+4}\int_{-\pi/2}^{\pi/2}\sin^{2k+2}(u)du \\
&= \frac{2^{2k+4}}{2\pi}\frac{2k+1}{(2k+4)(2k+2)}\int_{-\pi/2}^{\pi/2}\sin^{2k}(u)du \\
&= \frac{4k+2}{k+2}C_k \\
&= C_{k+1}.
\end{aligned}$$

□

Proposition A.3. *Let X be a standard Gaussian random variable. The moments of X are of the form*

$$\mathbb{E}[X^n] = \frac{1}{2\pi}\int_{\mathbb{R}} t^n e^{-\frac{t^2}{2}}dt = \begin{cases} 0, & n \text{ odd}, \\ (n-1)!!, & n \text{ even} \end{cases}$$

where $(2n-1)!! = 1\cdot 3\cdots(2n-1)$.

Proof. The odd case is obvious since the Gaussian distribution is symmetric. The even case can be proved by induction and integration by part. For $n = 0$, the integrand is the probability density thus the integral equals to 1. For $n = 2$, we know that the integral is exactly the variance of a standard Gaussian hence equals to 1. Assume

$$\frac{1}{2\pi}\int_{\mathbb{R}} t^{2k}e^{-\frac{t^2}{2}}dt = (2k-1)!!$$

for some $k \in \mathbb{Z}$. We will do a change of variable. Let $u = t^{2k+1}$, $v' = te^{-\frac{t^2}{2}}$. Then $u' = (2k+1)t^{2k}$ and $v = -e^{-\frac{t^2}{2}}$. Now

$$\begin{aligned}
&\frac{1}{2\pi}\int_{\infty}^{\infty} t^{2k+2}e^{-\frac{t^2}{2}}dt \\
&= \frac{1}{2\pi}[-t^{2k+1}e^{-\frac{t^2}{2}}]_{-\infty}^{\infty} + \frac{1}{2\pi}\int_{\infty}^{\infty}(2k+1)t^{2k}e^{-\frac{t^2}{2}}dt \\
&= 0 + (2k+1)(2k-1)!! \\
&= (2k+1)!!
\end{aligned}$$

□

Theorem A.4. *Let (μ_n) be a sequence of probability measures on $\mathbb{R}$ that converges vaguely to the probability measure ν. Then (μ_n) converges to ν weakly.*

Proof. The proof boils down to proving the family of probability measures (μ_n) is tight and showing that tightness together with vague convergence implies weak convergence.

Let us first show (μ_n) is tight, i.e., for any $\epsilon > 0$, there exists an $R \in \mathbb{R}$ such that $\mu_n([-R, R]) > 1 - \epsilon$ for all n.

Because ν is a probability measure, we have

$$\nu(\mathbb{R}) = \lim_{N\to\infty} \nu([-N, N]) = 1.$$

Thus, there exists a $\tilde{N} > 0$ such that $\nu([-\tilde{N}, \tilde{N}]) > 1 - \frac{\epsilon}{4}$. Additionally, given Lemma A.1, we can find a $N > \tilde{N}$ such that

$$\nu(\{N\}) = \lim_{\delta\to 0} \nu([N, N + \delta]) = 0.$$

Thus, there exists a $\delta > 0$ such that $\nu([-N - \delta, -N] \cup [N, N + \delta]) < \frac{\epsilon}{4}$. Let

$$f(x) = \begin{cases} 1, & -N \le x \le N \\ \frac{1}{\delta}(N + \delta - x), & N \le x \le N + \delta \\ \frac{1}{\delta}(N + \delta + x), & -N - \delta \le x \le -N \\ 0, & otherwise. \end{cases}$$

Since f is a continuous function that vanishes at infinity and μ_n converges to ν vaguely, we know that $\int f(x)d\mu_n(x) \to \int f(x)d\nu(x)$. Thus, there exists a $M > 0$ such that for all $n \ge M$ we have

$$|\int f(x)d\mu_n(x) - \int f(x)d\nu(x)| < \frac{\epsilon}{4}.$$

Because

$$|\int f(x)d\nu(x) - \nu([-N, N])| < \nu([-N - \delta, -N] \cup [N, N + \delta]) < \frac{\epsilon}{4},$$

we have that for all $n \ge M$

$$|\int f(x)d\mu_n(x) - \nu([-N, N])| < \frac{\epsilon}{2}.$$

Therefore, for all $n \ge M$,

$$\mu_n([-N - 1, N + 1]) > \int f(x)d\mu_n(x) > 1 - \frac{3}{4}\epsilon.$$

Since there are only finitely many probability measures in the family left and for each of them we could find a R such that $\mu_i([-R, R]) > 1 - \epsilon$, taking the maximum of such Rs and $N + 1$ will be enough to show that the family is tight.

Now we are ready to show that μ_n converges to ν weakly. Let f be a bounded continuous function on $\mathbb{R}$ and $M > 0$ be the constant that f is bounded by. We want to show that for any $\epsilon > 0$, there exists a $N > 0$ such that $|\int f(x)d\mu_n(x) - \int f(x)d\nu(x)| < \epsilon$. Because (μ_n) is tight, we can find a $R > 0$ such that $\mu_n([-R, R]) > 1 - \frac{\epsilon}{9M}$ for all n and $\nu([-R, R]) > 1 - \frac{\epsilon}{9M}$. Let

$$g(x) = \begin{cases} f(x), & -R \le x \le R \\ (R + 1)f(R) - f(R)x, & R \le x \le R + 1 \\ (R + 1)f(-R) + f(-R)x, & -R - 1 \le x \le -R \\ 0, & otherwise. \end{cases}$$

Then g is a continuous function that vanishes at infinity. We know that $\int g(x)d\mu_n(x)$ converges to $\int g(x)d\nu(x)$. There exists a $k > 0$ such that for all $n \geq k$ we have

$$|\int g(x)d\mu_n(x) - \int g(x)d\nu(x)| < \frac{\epsilon}{9}.$$

Moreover, we know that $g(x) - f(x) = 0$ for all $x \in [-R, R]$ and $|g(x) - f(x)| < 4M$ for all $x \in [\infty, -R] \cup [R, \infty]$. Hence,

$$\begin{aligned} |\int_{\mathbb{R}} g(x) - f(x)d\mu_n(x)| &= |\int_{[\infty,-R]\cup[R,\infty]} g(x) - f(x)d\mu_n(x)| \\ &< 4M\mu_n([\infty, -R] \cup [R, \infty]) \\ &< \frac{4\epsilon}{9} \text{ for all } n. \end{aligned}$$

Similarly, $|\int_{\mathbb{R}} g(x) - f(x)d\nu(x)| < \frac{4\epsilon}{9}$. Therefore, we have $|\int_{\mathbb{R}} f(x)d\mu_n(x) - \int_{\mathbb{R}} f(x)d\nu(x)| < \epsilon$ for all $n \geq k$. $\square$