Proof of Wigner's Semicircular Law by the Stieltjes Transform Method

Laura Veith

Adviser: Professor Jim Morrow

June 11, 2013

1 Introduction

In recent decades, physicists and mathematicians alike have looked toward random matrix theory to help bring insight into systems that involve a multitude of interacting variables. Such systems arise in quantum chaos, thermodynamics, optics, number theory, and encryption, to name a few [4, 15, 18, 20]. Not surprisingly, random matrix theory was popularized in the mid-1950s in direct response to the difficulty of modeling a particularly complex system: the highly excited states of heavy atomic nuclei.

The energy levels of these heavy atomic nuclei are the eigenvalues of a Hamiltonian operator, which, in a sense, encodes the system. However, the complicated structure of the heavy nuclei rendered it impossible to accurately define a Hamiltonian for the system. Eugene Wigner, a theoretical physicist and mathematician who would later win a Nobel Prize for his contributions to nuclear physics, recognized the need to develop a new statistical approach to modeling this system.

The difference between Wigner's scheme and other statistical methods was the absence of a collection of known individual states to average over in order to predict individual states of future systems. As Freeman Dyson, another renowned physicist and mathematician, expressed, this statistical theory would need to "define in a mathematically precise way an ensemble of systems in which all possible laws of interaction are equally probable" [7]. The proposed solution was to create a Hamiltonian out of random elements with certain properties.

Wigner is famous for pioneering this approach. In 1955, he began with $(2N + 1) \times (2N + 1)$ real symmetric matrices (for N large), whose diagonal elements were all zero [25]. The off-diagonal elements had uniform absolute value v with randomly distributed signs. By calculating moments and utilizing perturbation theory, he proved that the distribution of the eigenvalues averaged over all such matrices tended toward a semicircular distribution[1] . In 1958, Wigner relaxed the conditions on the elements of the real symmetric matrices. Independence of entries, uniform variance, and bounded moments comprised the new sufficient conditions [26].

In the 1960s, other mathematicians and physicists explored random matrix theory and contributed to a wider knowledge of the field. Madan Lal Mehta and Michel Gaudin concluded that the matrices could be Hermitian, provided that the elements were randomly and independently distributed with distributions invariant under unitary transformations [10]. Considering the limit as the dimension of properly normalized Hermitian matrices approached infinity, Mehta and Guadin showed that the eigenvalues tended toward a semicircular distribution. Their proof involved Hermite polynomials and properties of the harmonic oscillator wave functions. It was not until the mid-1960s that Vladimir Marchenko and Leonid Pastur suggested a new approach to proving that the eigenvalues of a certain set of matrices tend toward a particular limiting distribution [16]. They proposed finding the Stieltjes transform of the limiting distribution of the eigenvalues and then using an inversion formula to arrive at the desired result.

[1]This is the first manifestation of Wigner's semicircular law, which we will define later.

Over the years, as random matrix theory has developed, new sufficient conditions on the matrix elements and innovative proof techniques have surfaced. This paper is an exposition of Wigner's semicircular law. In brief, the law in its current form states that the distribution of eigenvalues of random Hermitian matrices tends toward a semicircular distribution, independent of the underlying probability measure. Our goal is to explore the proof of this law utilizing the Stieltjes transform.

2 Preliminaries

2.1 Important Definitions

Definition 2.1.1 (Wigner Matrix)**.** A Wigner matrix A is a random Hermitian matrix. The entries $[a_{ij}]$ above the main diagonal are complex-valued, independent, and identically distributed with mean 0 and variance 1, and $a_{ji} = \overline{a_{ij}}$. The entries on the main diagonal $[a_{ii}]$ are real-valued, independent, and identically distributed with mean 0 and variance 1. Furthermore, all entries are uniformly bounded in modulus by K, and the joint distribution of the entries $[a_{ij}]$ is absolutely continuous with respect to Lebesgue measure.

For the result in this paper, we want to work with a normalized Hermitian matrix. The proper normalization of an $n \times n$ Wigner matrix M_n is $\frac{1}{\sqrt{n}}$ [23]. Note that normalizing M_n contracts its eigenvalues by a factor of $\frac{1}{\sqrt{n}}$ but does not change their order.[2]

Definition 2.1.2 (Empirical Spectral Distribution)**.** The empirical spectral distribution (ESD) of the normalized matrix $\frac{M_n}{\sqrt{n}}$ is

$$\mu_n := \frac{1}{n} \sum_{j=1}^{n} \delta_{\lambda_j/\sqrt{n}},$$

where λ_j for $j = 1, \ldots, n$ are the eigenvalues of M_n.

A distribution of particular interest is the distribution μ_{sc} of the semicircular law, which is given by

$$\mu_{sc} := \frac{1}{2\pi} \sqrt{4 - x^2},$$

for $|x| \leq 2$.

2.2 The Stieltjes Transform

Definition 2.2.1 (Stieltjes Transform)**.** The Stieltjes transform of an arbitrary measure μ is defined by

$$s_\mu(z) := \int_{\mathbb{R}} \frac{1}{x - z} d\mu(x)$$

for all $z \in \mathbb{C}$ outside the support of μ.

We will denote by s_μ, s_n, and s_{sc} the Stieltjes transforms of the distributions μ, μ_n, and μ_{sc}, respectively.

3 Statement of the Theorem

Theorem 3.0.1 (Wigner's Semicircular Law)**.** *Let M_n be a sequence of $n \times n$ Wigner matrices. Let μ_n be the empirical spectral distribution of $\frac{M_n}{\sqrt{n}}$. Let μ_{sc} be the distribution of the semicircular law. Then, μ_n converges to μ_{sc} weakly almost surely.*

[2]It is worth noting at this point that the eigenvalues of these matrices are most likely distinct, as are the eigenvalues of any square matrix (Theorem A.1.1).

We will prove this theorem in three steps:

1. For any fixed z in the upper half plane, $|s_n(z,\omega) - \mathbb{E}[s_n(z)]|$ converges to zero almost surely.
2. For any fixed z in the upper half plane, $s_n(z,\omega)$ converges to $s_{sc}(z)$.
3. Outside of a null set, $s_n(z,\omega)$ converges to $s_{sc}(z)$ for all z in the upper half plane.

This will be sufficient to prove Wigner's semicircular law given the following theorem.

Theorem 3.0.2. *Let μ_k be a sequence of random probability measures on the real line, and let μ be a deterministic probability measure. Then μ_k converges weakly almost surely to μ if and only if $s_{\mu_k}(z)$ converges almost surely to $s_\mu(z)$ for every z in the upper half-plane.*

Proof. Suppose μ_k converges to μ weakly almost surely. That is, $\int_{\mathbb{R}} f(x)d\mu_k(x)$ converges to $\int_{\mathbb{R}} f(x)d\mu(x)$ for all continuous, bounded functions f.

Consider the function $f(x) = (x-z)^{-1}$ for fixed z with positive imaginary part. Then, as we will prove in Lemma 4.0.1 and Proposition 4.0.3, f is continuous and bounded. Thus, it must be true that

$$\int_{\mathbb{R}} \frac{1}{x-z} d\mu_k(x) \to \int_{\mathbb{R}} \frac{1}{x-z} d\mu(x),$$

which is to say, s_{μ_k} converges to s_μ almost surely for all z in the upper half plane.

Conversely, suppose s_{μ_k} converges to s_μ almost surely for all z in the upper half plane. Let ϕ be a function on the real line with compact support. In order to prove that μ_k converges to μ weakly almost surely, we will first prove that μ_k converges to μ vaguely almost surely by showing that $\int_{\mathbb{R}} \phi(x)d\mu_k(x)$ converges to $\int_{\mathbb{R}} \phi(x)d\mu(x)$ using a series of approximations.

Let $z = a+ib$, and consider the integral

$$T_\phi(z) = \frac{1}{\pi}\Im\left[\int_{\mathbb{R}} \frac{\phi(x)dx}{x-(a+ib)}\right] = \frac{1}{\pi}\Im\left[\int_{\mathbb{R}} \frac{\phi(x)dx}{a-(x+ib)}\right].$$

According to Theorem B.0.3, this converges to $\phi(a)$ as $b \to 0^+$ uniformly in a. Therefore, for all $\varepsilon > 0$, there exists $\delta > 0$ such that

$$|T_\phi(a+ib) - \phi(a)| < \varepsilon$$

whenever $|b| < \delta$. It follows that

$$\begin{aligned}
\left|\int_{\mathbb{R}} T_\phi(a+ib)d\mu_k(a) - \int_{\mathbb{R}} \phi(a)d\mu_k(a)\right| &= \left|\int_{\mathbb{R}} \left(T_\phi(a+ib) - \phi(a)\right) d\mu_k(a)\right| \\
&\leq \int_{\mathbb{R}} |T_\phi(a+ib) - \phi(a)|\, d\mu_k(a) \\
&< \int_{\mathbb{R}} \varepsilon\, d\mu_k(a) \\
&= \varepsilon.
\end{aligned}$$

Note that the same calculation and conclusion can be produced by replacing μ_k with μ. Thus, by the triangle inequality,

$$\begin{aligned}
\left|\int \phi d\mu_k - \int \phi d\mu\right| &\leq \left|\int \phi d\mu_k - \int T_\phi d\mu_k\right| + \left|\int T_\phi d\mu - \int \phi d\mu_k\right| + \left|\int T_\phi d\mu_k - \int T_\phi d\mu\right| \\
&< 2\varepsilon + \left|\int T_\phi d\mu_k - \int T_\phi d\mu\right|. \qquad (1)
\end{aligned}$$

Next, since ϕ has compact support on the line, it is nonzero on a finite interval, which means $T_\phi(a+ib)$ can be approximated by a Riemann sum:

$$\left| T_\phi(a+ib) - \Delta \frac{1}{\pi} \Im \sum_j \frac{\phi(x_j)}{a-(x_j+ib)} \right| < \varepsilon,$$

where the intervals are chosen to be uniform of length Δ. Then, we have

$$\left| \int_{\mathbb{R}} T_\phi(a+ib) d\mu_k(a) - \Delta \frac{1}{\pi} \Im \int_{\mathbb{R}} \sum_j \frac{\phi(x_j)}{a-(x_j+ib)} d\mu_k(a) \right| < 2\varepsilon,$$

with the same inequality for μ. To simplify our notation, define $g(a)$ by

$$g(a) = \Delta \frac{1}{\pi} \Im \sum_j \frac{\phi(x_j)}{a-(x_j+ib)}.$$

Then, by the triangle inequality,

$$\begin{aligned} \left| \int T_\phi d\mu_k - \int T_\phi d\mu \right| &\leq \left| \int T_\phi d\mu_k - \int g d\mu_k \right| + \left| \int g d\mu - \int T_\phi d\mu \right| + \left| \int g d\mu_k - \int g d\mu \right| \\ &< 4\varepsilon + \left| \int g d\mu_k - \int g d\mu \right|. \end{aligned} \tag{2}$$

Lastly, observe that

$$\begin{aligned} \int_{\mathbb{R}} \left(\Delta \frac{1}{\pi} \Im \sum_j \frac{\phi(x_j)}{a-(x_j+ib)} \right) d\mu_k(a) &= \Delta \frac{1}{\pi} \Im \sum_j \phi(x_j) \int_{\mathbb{R}} \frac{1}{a-(x_j+ib)} d\mu_k(a) \\ &= \Delta \frac{1}{\pi} \Im \sum_j \phi(x_j) s_{\mu_k}(x_j+ib), \end{aligned}$$

and note that a similar formula holds for μ. By assumption, $s_{\mu_k}(z)$ converges to $s_\mu(z)$ almost surely for all z in the upper half plane, so for k large,

$$\left| \Delta \frac{1}{\pi} \Im \sum_j \phi(x_j) s_{\mu_k}(x_j+ib) - \Delta \frac{1}{\pi} \Im \sum_j \phi(x_j) s_\mu(x_j+ib) \right| < \varepsilon. \tag{3}$$

Combining Equations (1), (2), and (3), we get

$$\begin{aligned} \left| \int_{\mathbb{R}} \phi(a) d\mu_k(a) - \int_{\mathbb{R}} \phi(a) d\mu(a) \right| &< 2\varepsilon + \left| \int_{\mathbb{R}} T_\phi(a+ib) d\mu_k(a) - \int_{\mathbb{R}} T_\phi(a+ib) d\mu(a) \right| \\ &< 2\varepsilon + 4\varepsilon + \left| \int_{\mathbb{R}} g(a) d\mu_k(a) - \int_{\mathbb{R}} g(a) d\mu(a) \right| \\ &< 2\varepsilon + 4\varepsilon + \varepsilon \\ &= 7\varepsilon. \end{aligned}$$

Therefore, $\int_{\mathbb{R}} \phi(a) d\mu_k(a)$ converges to $\int_{\mathbb{R}} \phi(a) d\mu(a)$ almost surely, which means μ_n converges vaguely almost surely to μ. Now, since μ_k and μ are probability measures on the real line, $\mu_k(\mathbb{R}) = 1$ for all k, and $\mu(\mathbb{R}) = 1$. So it follows that $\mu(\mathbb{R}) = \lim_{k\to\infty} \mu_k(\mathbb{R})$. Then, according to the Portemanteau Theorem [14], μ_k converges to μ weakly almost surely. $\square$

Before proceeding to prove the three steps we outlined previously, we will discuss some of the properties of the Stieltjes transform of a measure.

4 Properties of the Stieltjes Transform

Lemma 4.0.1. *For z away from the real axis, the Stieltjes transform of an arbitrary measure μ has the pointwise bound*

$$|s_\mu(z)| \leq \frac{1}{|\Im(z)|}.$$

Proof. Since

$$\left|\frac{1}{x-z}\right| = \frac{1}{|x-z|} \leq \frac{1}{|\Im(z)|},$$

we have

$$\begin{aligned}
|s_\mu(z)| &= \left|\int_{\mathbb{R}} \frac{1}{x-z} d\mu(x)\right| \\
&\leq \int_{\mathbb{R}} \frac{1}{|\Im(z)|} d\mu(x) \\
&= \frac{1}{|\Im(z)|} \int_{\mathbb{R}} d\mu(x) \\
&= \frac{1}{|\Im(z)|}.
\end{aligned}$$

□

Proposition 4.0.2. *The Stieltjes transform of μ_n has the identity*

$$s_n(z) = \int_{\mathbb{R}} \frac{1}{x-z} d\mu_n(x) = \frac{1}{n} \operatorname{tr}\left(\frac{1}{\sqrt{n}} M_n - zI_n\right)^{-1}. \tag{4}$$

Proof. To see that this is true, let $X_n = \frac{M_n}{\sqrt{n}}$. Let $\lambda_1 \leq \ldots \leq \lambda_n$ be the eigenvalues of M_n. Then the eigenvalues of X_n are

$$\frac{\lambda_1}{\sqrt{n}} \leq \ldots \leq \frac{\lambda_n}{\sqrt{n}}.$$

We want to show that the eigenvalues of $X_n - zI_n$ are $\frac{\lambda_j}{\sqrt{n}} - z$, where $j = 1, \ldots, n$. Let e_j be an eigenvector of X_n. Then,

$$X_n e_j = \frac{\lambda_j}{\sqrt{n}} e_j,$$

which implies

$$\begin{aligned}
(X_n - zI_n)e_j &= X_n e_j - ze_j \\
&= \frac{\lambda_j}{\sqrt{n}} e_j - ze_j \\
&= \left(\frac{\lambda_j}{\sqrt{n}} - z\right) e_j,
\end{aligned} \tag{5}$$

as desired. Next, we want to show that the eigenvalues of $(X_n - zI_n)^{-1}$ are $(\frac{\lambda_j}{\sqrt{n}} - z)^{-1}$. Note that, in addition to being an eigenvector of X_n, e_j is an eigenvector of $(X_n - zI_n)$ by Equation (5). Then, since $Ax = \lambda x$ implies $A^k x = \lambda^k x$ for all $k \in \mathbb{Z}$,

$$(X_n - zI_n)^{-1} e_j = \left(\frac{\lambda_j}{\sqrt{n}} - z\right)^{-1} e_j.$$

Therefore,

$$\int_{\mathbb{R}} \frac{1}{x-z} d\mu_n(x) = \frac{1}{n}\mathrm{tr}\left(X_n - zI_n\right)^{-1} = \frac{1}{n}\mathrm{tr}\left(\frac{1}{\sqrt{n}}M_n - zI_n\right)^{-1}.$$

$\square$

Remark. From this identity, we see that the imaginary part of $s_n(z)$ is positive for z in the upper half plane. Let $z = a + ib$ be a point in the upper half plane. Then,

$$\begin{aligned}\Im\left(s_n(z)\right) &= \Im\left(\frac{1}{n}\mathrm{tr}\left(\frac{1}{\sqrt{n}}M_n - zI_n\right)^{-1}\right) \\ &= \Im\left(\frac{1}{n}\sum_{j=1}^{n}\frac{1}{\frac{\lambda_j}{\sqrt{n}} - z}\right) \\ &= \frac{1}{n}\sum_{j=1}^{n}\frac{b}{\left(\frac{\lambda_j}{\sqrt{n}} - a\right)^2 + b^2}. \end{aligned} \tag{6}$$

This will be important in Section 6.

Next, we want to consider the expansion of the right-hand side of Equation (4). We have

$$\begin{aligned}(X_n - zI_n)^{-1} &= \left[-z\left(-\frac{X_n}{z} + I_n\right)\right]^{-1} \\ &= -\frac{1}{z}\left(I_n - \frac{X_n}{z}\right)^{-1} \\ &= -\frac{1}{z}\left(I + \frac{X_n}{z} + \frac{X_n^2}{z^2} + \ldots + \frac{X_n^k}{z^k} + \ldots\right).\end{aligned}$$

This expansion is true when $||\frac{X_n}{z}|| < 1$, which holds when $|z| > ||X||$. Therefore, when $|z|$ is large,

$$\begin{aligned}s_n(z) &= -\frac{1}{nz}\mathrm{tr}\left[I + \frac{X_n}{z} + \frac{X_n^2}{z^2} + \ldots\right] \\ &= -\frac{1}{nz}\left[n + \frac{1}{z}\mathrm{tr}(X_n) + \frac{1}{z^2}\mathrm{tr}(X_n^2) + \ldots\right].\end{aligned}$$

In terms of the matrix M_n, we have

$$s_n(z) = -\frac{1}{z}\left[1 + \frac{1}{zn}\mathrm{tr}\left(\frac{M_n}{\sqrt{n}}\right) + \frac{1}{z^2 n}\mathrm{tr}\left(\frac{M_n}{\sqrt{n}}\right)^2 + \frac{1}{z^3 n}\mathrm{tr}\left(\frac{M_n}{\sqrt{n}}\right)^3 + \ldots\right].$$

Therefore, for any μ_n,

$$s_n(z) = -\frac{1}{z} - \frac{1}{z^2 n}O(1), \tag{7}$$

since $\mathrm{tr}\left(\frac{M_n}{\sqrt{n}}\right) = O(1)$. Equation (7) will be useful later when we define a recursion relation for a sequence of Stieltjes transforms of measures, but right now we want to prove that $s_n(z)$ is both continuous and analytic.

Proposition 4.0.3. *The Stieltjes transform of μ is continuous at all points z in the upper half plane.*

Proof. Let $z = a + ib$ be a complex number with positive imaginary part. Let $\delta = \frac{1}{2}b$. Let $f_j(x, z_j) = (x - z_j)^{-1}$ so that $\lim_{z_j \to z} f_j(x, z_j) = f(x, z)$. Then, for all z_j within δ of z,

$$|f_j(x, z_j)| = \left| \frac{1}{x - z_j} \right| \leq \frac{1}{\Im(x - z_j)} \leq \frac{1}{\Im(z_j)} \leq \frac{1}{\delta}.$$

Let $g(x, z) = \frac{1}{\delta}$, so

$$\left| \frac{1}{x - z_j} \right| \leq g(x, z).$$

Note that $\int_{\mathbb{R}} g(x, z) d\mu(x) < \infty$ because

$$\begin{aligned} \int_{\mathbb{R}} g(x, z) d\mu(x) &= \int_{\mathbb{R}} \left(\frac{1}{\delta} \right) d\mu(x) \\ &= \frac{1}{\delta} \int_{\mathbb{R}} d\mu(x) \\ &= \frac{1}{\delta}. \end{aligned}$$

Therefore, by the Lebesgue Dominated Convergence Theorem,

$$\lim_{z_j \to z} \int_{\mathbb{R}} \frac{1}{x - z_j} d\mu(x) = \int_{\mathbb{R}} \frac{1}{x - z} d\mu(x),$$

which is to say, $\lim_{z_j \to z} s_\mu(z_j) = s_\mu(z)$. Thus, the Stieltjes transform of μ is continuous at all points z in the upper half plane. □

Proposition 4.0.4. *The Stieltjes transform of μ is analytic at all points z in the upper half plane.*

Proof. Note that $f(x, z) = (x - z)^{-1}$ is analytic for all z in the upper half plane. This means that for any triangle C in the upper half plane,

$$\int_C f(x, z) dz = 0$$

by the Cauchy Integral Theorem.

Let $d\mu(x) = g(x)dx$, where g is a Riemann-integrable function. Then,

$$\int_C s_\mu(z) dz = \int_C \left[\int_{\mathbb{R}} \frac{d\mu(x)}{x - z} \right] dz = \int_C \left[\int_{\mathbb{R}} \frac{g(x)dx}{x - z} \right] dz.$$

By Fubini's Theorem, we can change the order of integration to get

$$\int_C \left[\int_{\mathbb{R}} \frac{g(x)dx}{x - z} \right] dz = \int_{\mathbb{R}} g(x) \left[\int_C \frac{dz}{x - z} \right] dx = 0.$$

Since $s_\mu(z)$ is continuous in the upper half plane, this tells us that $s_\mu(z)$ is analytic in the upper half plane by Morera's Theorem. □

4.1 Stieltjes Transform of the Semicircular Law

Theorem 4.1.1. *The Stieltjes transform of the distribution μ_{sc} of the semicircular law is*

$$s_{sc}(z) = \frac{-z + \sqrt{z^2 - 4}}{2}.$$

Proof. Fix z in the upper half plane. Then,

$$s_{sc}(z) = \int_{\mathbb{R}} \frac{1}{x-z} d\mu_{sc}(x) = \frac{1}{2\pi} \int_{-2}^{2} \frac{1}{x-z} \sqrt{4-x^2} dx.$$

Let $x = 2\cos y$. This gives us

$$\begin{aligned} s_{sc}(z) &= \frac{1}{2\pi} \int_{\pi}^{0} \frac{1}{(2\cos y) - z} \sqrt{4-(2\cos y)^2}(-2\sin y\, dy) \\ &= \frac{1}{\pi} \int_{0}^{\pi} \frac{2}{2\cos y - z} \sin^2 y\, dy \\ &= \frac{1}{\pi} \int_{0}^{2\pi} \frac{1}{2\left(\frac{e^{iy}+e^{-iy}}{2}\right) - z} \left(\frac{e^{iy}-e^{-iy}}{2i}\right)^2 dy. \end{aligned}$$

Let $\zeta = e^{iy}$. Then,

$$\frac{1}{\pi} \int_{0}^{2\pi} \frac{1}{2\left(\frac{e^{iy}+e^{-iy}}{2}\right) - z} \left(\frac{e^{iy}-e^{-iy}}{2i}\right)^2 dy = -\frac{1}{4\pi i} \oint_{|\zeta|=1} \frac{(\zeta^2-1)^2}{\zeta^2(\zeta^2+1-z\zeta)}\, d\zeta. \tag{8}$$

The integrand has three poles: $\zeta_0 = 0$, $\zeta_1 = \frac{z+\sqrt{z^2-4}}{2}$, and $\zeta_1 = \frac{z-\sqrt{z^2-4}}{2}$, where we choose the branch of the square root with positive imaginary part. In order to apply the Residue Theorem to evaluate the integral, we need to determine which of the poles falls inside the unit circle.

We know

$$\sqrt{z} = \operatorname{sign}(\Im(z)) \frac{|z| + z}{\sqrt{2(|z| + \Re(z))}},$$

for any $z \neq 0$. This means

$$\begin{aligned} \Re(\sqrt{z}) &= \operatorname{sign}(\Im(z)) \frac{|z| + \Re(z)}{\sqrt{2(|z| + \Re(z))}} \\ &= \operatorname{sign}(\Im(z)) \frac{\sqrt{|z|^2 - (\Re(z))^2}}{\sqrt{2(|z| - \Re(z))}} \\ &= \operatorname{sign}(\Im(z)) \frac{\sqrt{\Im(z)^2}}{\sqrt{2(|z| - \Re(z))}} \\ &= \frac{\Im(z)}{\sqrt{2(|z| - \Re(z))}}. \end{aligned}$$

Applying this to $\sqrt{z^2-4}$, we find

$$\begin{aligned} \Re(\sqrt{z^2-4}) &= \frac{\Im(z^2-4)}{\sqrt{2(|z^2-4| - \Re(z^2-4))}} \\ &= \frac{2\Re(z)\Im(z)}{\sqrt{2(|z^2-4| - \Re(z^2-4))}}. \end{aligned}$$

Since we have assumed $\Im(z) > 0$, we see that $\Re(\sqrt{z^2-4})$ and the real part of z have the same sign. Thus, $|\zeta_1| > |\zeta_2|$. Furthermore, $\zeta_1\zeta_2 = 1$, so it must be that $|\zeta_1| > 1$ and $|\zeta_2| < 1$. Therefore, in order to evaluate integral (8), we need to calculate the residue of the integrand at ζ_0 and at ζ_2. For ζ_0, we have

$$\operatorname{Res}(\zeta_0) = \frac{4\zeta_0(\zeta_0^2-1)(\zeta_0^2+1-z\zeta_0) - (\zeta_0^2-1)^2(2\zeta_0 - z)}{(\zeta_0^2+1-z\zeta_0)^2} = z.$$

For ζ_2, we have

$$\operatorname{Res}(\zeta_2) = \frac{(\zeta_2^2 - 1)^2}{\zeta_2^2\left(\zeta_2 - \left(\frac{z+\sqrt{z^2-4}}{2}\right)\right)} = -\sqrt{z^2-4}.$$

Thus, by the Residue Theorem,

$$\begin{aligned}
s_{sc}(z) &= -\frac{1}{4\pi i}\oint_{|\zeta|=1} \frac{(\zeta^2-1)^2}{\zeta^2(\zeta^2+1-z\zeta)}\, d\zeta \\
&= 2\pi i\left(-\frac{1}{4\pi i}\left(z - \sqrt{z^2-4}\right)\right) \\
&= \frac{-z+\sqrt{z^2-4}}{2}. \qquad (9)
\end{aligned}$$

$\square$

5 Step One

Theorem 5.0.1. *For fixed z in the upper half of the complex plane, $\left|s_n(z,\omega) - \mathbb{E}[s_n(z)]\right|$ converges almost surely to zero.*

In order to prove this theorem, we need a few calculations and results.

5.1 Recursion Relation

Proposition 5.1.1. *The Stieltjes transform of μ_n satisfies the recursion relation*

$$s_n(z) = s_{n-1}(z) + O\left(\frac{1}{n}\right).$$

Proof. To begin with, note that

$$\begin{aligned}
\sqrt{n(n-1)}s_{n-1}\left(\frac{\sqrt{n}}{\sqrt{n-1}}\, z\right) &= \frac{1}{n-1}\sqrt{n(n-1)}\operatorname{tr}\left(\frac{M_{n-1}}{\sqrt{n-1}} - \frac{\sqrt{n}}{\sqrt{n-1}} zI\right)^{-1} \\
&= \frac{\sqrt{n}}{\sqrt{n-1}}\left(\frac{\sqrt{n}}{\sqrt{n-1}}\right)^{-1}\operatorname{tr}\left(\frac{M_{n-1}}{\sqrt{n}} - zI\right)^{-1} \\
&= \operatorname{tr}\left(\frac{M_{n-1}}{\sqrt{n}} - zI\right)^{-1}. \qquad (10)
\end{aligned}$$

Let $\lambda_j(M_{n-1})$ be the eigenvalues of M_{n-1}, and let $\lambda_j(M_n)$ be the eigenvalues of M_n. Then,

$$\begin{aligned}
\sqrt{n(n-1)}s_{n-1}\left(\frac{\sqrt{n}}{\sqrt{n-1}}\, z\right) - ns_n(z) &= \operatorname{tr}\left(\frac{M_{n-1}}{\sqrt{n}} - zI\right)^{-1} - \operatorname{tr}\left(\frac{M_n}{\sqrt{n}} - zI\right)^{-1} \\
&= \sum_{j=1}^{n-1}\frac{1}{\frac{\lambda_j(M_{n-1})}{\sqrt{n}} - z} - \sum_{j=1}^{n}\frac{1}{\frac{\lambda_j(M_n)}{\sqrt{n}} - z}. \qquad (11)
\end{aligned}$$

We want to show that (11) is bounded. Fix $z = a + ib$ in the upper half plane, and consider the function

$$f(x) = \frac{1}{x-z}.$$

Let $x_j = \frac{\lambda_j(M_n)}{\sqrt{n}}$ and $\tilde{x}_j = \frac{\lambda_j(M_{n-1})}{\sqrt{n}}$. By Cauchy's Interlace Theorem (A.2.1),

$$x_1 \le \tilde{x}_1 \le x_2 \le \ldots \le x_{n-1} \le \tilde{x}_{n-1} \le x_n.$$

For $k = 1, 2, \dots, n-1$, consider $f(\tilde{x}_k) - f(x_k)$. By the Mean Value Theorem,

$$f(\tilde{x}_k) - f(x_k) = f'(\xi_k)\Delta_k,$$

for some $\xi_k \in (x_k, \tilde{x}_k)$, where $\Delta_k = \tilde{x}_k - x_k$. Therefore,

$$\begin{aligned}
\sum_{j=1}^{n-1} \frac{1}{\frac{\lambda_j(M_{n-1})}{\sqrt{n}} - z} - \sum_{j=1}^{n} \frac{1}{\frac{\lambda_j(M_n)}{\sqrt{n}} - z} &= \sum_{j=1}^{n-1} \frac{1}{\tilde{x}_j - z} - \sum_{j=1}^{n} \frac{1}{x_j - z} \\
&= -f(x_n) + \sum_{k=1}^{n-1} f(\tilde{x}_k) - f(x_k) \\
&= -f(x_n) + \sum_{k=1}^{n-1} f'(\xi_k)\Delta_k \\
&\approx -f(x_n) + \int_{\mathbb{R}} f'(\xi) d\xi \\
&= -f(x_n) + \int_{\mathbb{R}} \frac{-1}{(\xi - z)^2} d\xi.
\end{aligned}$$

Then, since $0 < |f(x_n)| \leq \frac{1}{b}$,

$$\begin{aligned}
\left| -f(x_n) + \int_{\mathbb{R}} \frac{-1}{(\xi - z)^2} d\xi \right| &\leq |f(x_n)| + \left| \int_{\mathbb{R}} \frac{-1}{(\xi - z)^2} d\xi \right| \\
&\leq \frac{1}{b} + \int_{\mathbb{R}} \frac{1}{(\xi - a)^2 + b^2} d\xi \\
&= \frac{1}{b} + \frac{1}{b^2} \arctan\left(\frac{x - a}{b} \right) \Bigg|_{x=-\infty}^{\infty} \\
&= \frac{1}{b} + \frac{\pi}{b}.
\end{aligned}$$

Thus,

$$\sum_{j=1}^{n-1} \frac{1}{\frac{\lambda_j(M_{n-1})}{\sqrt{n}} - z} - \sum_{j=1}^{n} \frac{1}{\frac{\lambda_j(M_n)}{\sqrt{n}} - z} = O(1).$$

Then we have

$$\sqrt{n(n-1)} s_{n-1} \left(\frac{\sqrt{n}}{\sqrt{n-1}} z \right) - n s_n(z) = O(1).$$

Dividing by n gives us

$$\sqrt{\frac{n-1}{n}} s_{n-1} \left(\frac{\sqrt{n}}{\sqrt{n-1}} z \right) - s_n(z) = O\left(\frac{1}{n} \right).$$

Lastly, we need to show that we can approximate

$$\sqrt{\frac{n-1}{n}} s_{n-1} \left(\frac{\sqrt{n}}{\sqrt{n-1}} z \right)$$

by $s_{n-1}(z)$. In order to do this, recall Equation (7), our estimate of the Taylor series expansion of $s_n(z)$:

$$s_n(z) = -\frac{1}{z} - \frac{1}{z^2 n} O(1).$$

We also need the Taylor expansion of $\frac{\sqrt{n-1}}{\sqrt{n}}$ about infinity, which is given by

$$\frac{\sqrt{n-1}}{\sqrt{n}} = 1 - \frac{1}{2n} - \frac{1}{8n^2} + \ldots = 1 + O\left(\frac{1}{n}\right).$$

Therefore,

$$\begin{aligned}
\sqrt{\frac{n-1}{n}} s_{n-1}\left(\frac{\sqrt{n}}{\sqrt{n-1}}(z)\right) &= \sqrt{\frac{n-1}{n}}\left[-\frac{\sqrt{n-1}}{\sqrt{n}}\frac{1}{z} - \frac{n-1}{n}\frac{1}{z^2(n-1)}O(1)\right] \\
&= \left[1 + O\left(\frac{1}{n}\right)\right]\left[-\frac{1}{z} - \frac{1}{z^2(n-1)}O(1)\right] \\
&= \left[1 + O\left(\frac{1}{n}\right)\right] s_{n-1}(z) \\
&= s_{n-1}(z) + O\left(\frac{1}{n}\right). \qquad (12)
\end{aligned}$$

This gives us the relation

$$s_{n-1}(z) + O\left(\frac{1}{n}\right) - s_n(z) = O\left(\frac{1}{n}\right),$$

which simplifies to our desired recursion relation,

$$s_n(z) = s_{n-1}(z) + O\left(\frac{1}{n}\right). \qquad (13)$$

□

From recursion relation (13), we see that the most permuting rows or columns of M_n can influence $s_n(z)$, while keeping the matrix Hermitian, is $O\left(\frac{1}{n}\right)$. So we can apply McDiarmid's Inequality (Proposition D.0.1) to show that $s_n(z)$ is concentrated around its mean. For all $\kappa > 0$, consider

$$P\left(\left|s_n(z) - \mathbb{E}[s_n(z)]\right| \geq \frac{\kappa}{\sqrt{n}}\right).$$

By McDiarmid's Inequality,

$$P\left(\left|s_n(z) - \mathbb{E}[s_n(z)]\right| \geq \frac{\kappa}{\sqrt{n}}\right) \leq 2\exp\left(-\frac{2\left(\frac{\kappa}{\sqrt{n}}\right)^2}{\sum_{i=1}^n c_i^2}\right),$$

where c_i is $O\left(\frac{1}{n}\right)$. Therefore,

$$P\left(\left|s_n(z) - \mathbb{E}[s_n(z)]\right| \geq \frac{\kappa}{\sqrt{n}}\right) \leq Ce^{c\kappa^2},$$

for absolute constants $C, c > 0$. We will utilize this inequality in our proof of Theorem 5.0.1.

5.2 Proof of Step One

Proof of Theorem 5.0.1. Fix z and define $f_n(\omega) \geq 0$ by $f_n(\omega) = s_n(z,\omega) - \mathbb{E}[s_n(z)]$. Note that $f_n(\omega)$ converges to zero if and only if $\limsup_{n\to\infty} f_n(\omega) = 0$ because $f_n(\omega) \geq 0$, which means $\liminf_{n\to\infty} f_n(\omega) \geq 0$. Fix ε and let $A_n = \{\omega : f_n(\omega) \geq \varepsilon\}$. If $\omega \in \limsup_{n\to\infty}(A_n)$, then $f_n(\omega) \geq \varepsilon$ infinitely often (that is, for infinitely many n). If $\omega \notin \limsup_{n\to\infty}(A_n)$, then $f_n(\omega) \geq \varepsilon$ only finitely many times, which is to say, for some $N \in \mathbb{N}$, $f_n(\omega) < \varepsilon$ for all $n \geq N$.

We want to be able to say that

$$\limsup_{n\to\infty} A_n = \{\omega : \limsup_{n\to\infty} f_n(\omega) \geq \varepsilon\}. \tag{14}$$

This is allowed because $\limsup_{n\to\infty} f_n(\omega) \geq \varepsilon$ if and only if $f_n(\omega) \geq \varepsilon$ infinitely often, which is equivalent to $\limsup_{n\to\infty} A_n$.

Now, consider inequality (5.1):

$$P\left(\left|s_n(z) - \mathbb{E}[s_n(z)]\right| \geq \frac{\kappa}{\sqrt{n}}\right) = P\left(\omega : f_n(\omega) \geq \frac{\kappa}{\sqrt{n}}\right) \leq Ce^{-c\lambda^2}.$$

Let $\kappa = \varepsilon n^{1/4}$. Then,

$$P\left(\omega : f_n(\omega) \geq \frac{\varepsilon}{n^{1/4}}\right) \leq Ce^{-c\varepsilon^2\sqrt{n}}.$$

By the Integral Test (Appendix E), we have

$$\sum_{n=1}^{\infty} P\left(\omega : f_n(\omega) \geq \frac{\varepsilon}{n^{1/4}}\right) \leq \sum_{n=1}^{\infty} Ce^{-c\varepsilon^2\sqrt{n}} < \infty.$$

So, by the Borel-Cantelli Theorem (C.0.1),

$$P\left(\omega : \omega \in \limsup_{n\to\infty} \left\{\sigma : f_n(\sigma) \geq \frac{\varepsilon}{n^{1/4}}\right\}\right) = 0.$$

Note that

$$\{\omega : f_n(\omega) \geq \varepsilon\} \subset \left\{\sigma : f_n(\sigma) \geq \frac{\varepsilon}{n^{1/4}}\right\}.$$

Thus, it must be that for every $\varepsilon > 0$

$$P\left(\omega : \omega \in \limsup_{n\to\infty} \{\sigma : f_n(\sigma) \geq \varepsilon\}\right) = 0.$$

Now let $\varepsilon = \frac{1}{m}$ for fixed m. So for each m,

$$P\left(\omega : \omega \in \limsup_{n\to\infty} \left\{\sigma : f_n(\sigma) \geq \frac{1}{m}\right\}\right) = P\left(\omega : \omega \in \left\{\sigma : \limsup_{n\to\infty} f_n(\sigma) \geq \frac{1}{m}\right\}\right) = 0$$

by Equation (14). Therefore,

$$P\left(\omega : \omega \in \bigcup_{m=1}^{\infty} \left\{\sigma : \limsup_{n\to\infty} f_n(\sigma) \geq \frac{1}{m}\right\}\right) = 0,$$

as well, because $\left\{\bigcup_{m=1}^{\infty} \frac{1}{m}\right\}$ is a countable set. If $\limsup_{n\to\infty} f_n(\omega) > 0$, then

$$\omega \in \left\{\sigma : \limsup_{n\to\infty} f_n(\sigma) > \frac{1}{m}\right\},$$

for some m. Thus,

$$P\left(\omega : f_n(\omega) \text{ does not converge to zero}\right) = 0,$$

which is to say $f_n(\omega) = \left|s_n(z,\omega) - \mathbb{E}[s_n(z)]\right|$ converges to zero for almost all ω. $\square$

6 Step Two

Theorem 6.0.1. *For any fixed z in the upper half plane, $\mathbb{E}[s_n(z)]$ converges to $s_{sc}(z)$.*

In order to prove this theorem, we wish to derive the formula

$$\mathbb{E}[s_n(z)] = \frac{1}{-z - \mathbb{E}[s_n(z)]} + o(1).$$

We will do this with the help of Proposition 6.1.1.

6.1 Schur Complement Formula

Proposition 6.1.1. *Let A_n be an $n \times n$ matrix, and let A_{n-1} be the top left $(n-1) \times (n-1)$ minor of A_n. Denote the bottom rightmost entry of A_n by a_{nn}. Let $V, Y \in \mathbb{C}^{n-1}$ be the rightmost column of A_n with the last entry removed and the bottom row of A_n with the last entry removed, respectively. That is,*

$$A_n = \begin{bmatrix} A_{n-1} & V \\ Y & a_{nn} \end{bmatrix}.$$

Suppose that A_n and A_{n-1} are both invertible, and let b_{nn} be the bottom rightmost entry of A_n^{-1}. Then,

$$b_{nn} = \frac{1}{a_{nn} - YA_{n-1}^{-1}V}.$$

Proof. Let Z denote the rightmost column of A_n^{-1} with the last entry removed. Let W be the bottom row of A_n^{-1} with the last entry removed, and let B_{n-1} be the top left $(n-1) \times (n-1)$ minor of A_n^{-1}. That is,

$$A_n^{-1} = \begin{bmatrix} B_{n-1} & Z \\ W & b_{nn} \end{bmatrix}.$$

Since $A_n A_n^{-1} = I$, where I is the identity matrix,

$$\begin{bmatrix} Y & a_{nn} \end{bmatrix} \times \begin{bmatrix} Z \\ b_{nn} \end{bmatrix} = Y \cdot Z + a_{nn} b_{nn} = 1. \tag{15}$$

Also,

$$\begin{bmatrix} A_{n-1} & V \end{bmatrix} \times \begin{bmatrix} Z \\ b_{nn} \end{bmatrix} = A_{n-1} Z + b_{nn} V = \vec{0}. \tag{16}$$

Solving Equation (16) for Z we find

$$Z = -b_{nn} A_{n-1}^{-1} V. \tag{17}$$

If we substitute the right-hand side of Equation (17) for Z in Equation (15), we get

$$Y \cdot [-b_{nn} A_{n-1}^{-1} V] + a_{nn} b_{nn} = 1.$$

Solving this for b_{nn}, we find

$$b_{nn} = \frac{1}{a_{nn} - YA_{n-1}^{-1}V}.$$

□

Note that we can write

$$M_n = \begin{bmatrix} M_{n-1} & X \\ X^* & \xi_{nn} \end{bmatrix},$$

where M_{n-1} is the top left $(n-1)\times(n-1)$ minor of M_n, X is the rightmost column of M_n with the last entry removed, and ξ_{nn} is the bottom rightmost element of M_n. Then, consider the particular case of Proposition 6.1.1 where

$$A_n = \frac{1}{\sqrt{n}} M_n - zI_n.$$

This means

$$A_{n-1} = \frac{1}{\sqrt{n}} M_{n-1} - zI_{n-1},$$

$V = \frac{1}{\sqrt{n}} X$, $Y = \frac{1}{\sqrt{n}} X^*$, and $a_{nn} = \frac{1}{\sqrt{n}} \xi_{nn} - z$. This means

$$\begin{aligned} b_{nn} &= \frac{1}{(\frac{1}{\sqrt{n}}\xi_{nn} - z) - \frac{1}{n} X^*(\frac{1}{\sqrt{n}} M_{n-1} - zI_{n-1})^{-1} X} \\ &= \frac{1}{-z - \frac{1}{n} X^*(\frac{1}{\sqrt{n}} M_{n-1} - zI_{n-1})^{-1} X + \frac{1}{\sqrt{n}}\xi_{nn}}. \end{aligned}$$

We will be considering what happens as $n \to \infty$, so we can take $\frac{1}{\sqrt{n}}\xi_{nn} = o(1)$, to get

$$b_{nn} = \frac{1}{-z - \frac{1}{n} X^*(\frac{1}{\sqrt{n}} M_{n-1} - zI_{n-1})^{-1} X + o(1)}. \tag{18}$$

Note that

$$\begin{aligned} \mathbb{E}[s_n(z)] &= \mathbb{E}\left[\frac{1}{n}\mathrm{tr}\left(\frac{1}{\sqrt{n}} M_n - zI_n\right)^{-1}\right] \\ &= \mathbb{E}\left[\left(\frac{1}{\sqrt{n}} M_n - zI_n\right)^{-1}_{nn}\right] \\ &= \mathbb{E}\left[b_{nn}\right], \end{aligned}$$

since the entries on the diagonal are independent and identically distributed. Therefore, if we take expectations of both sides of Equation (18), we get

$$\mathbb{E}[s_n(z)] = \mathbb{E}\left[\frac{1}{-z - \frac{1}{n} X^*(\frac{1}{\sqrt{n}} M_{n-1} - zI_{n-1})^{-1} X + o(1)}\right]. \tag{19}$$

Let $R_{n-1} = (\frac{1}{\sqrt{n}} M_{n-1} - zI_{n-1})^{-1}$. We want to show that

$$\frac{1}{n} X^* R_{n-1} X = \mathbb{E}[s_n(z)] + o(1). \tag{20}$$

In this way, we will be replacing something random with something deterministic, which will enable us to evaluate the expectation on the right-hand side of Equation (19). We will arrive at Equation (20) by justifying the series of approximations

$$\begin{aligned} \frac{1}{n} X^* R_{n-1} X &= \frac{1}{n}\mathrm{tr} R_{n-1} + o(1) \\ &= s_{n-1}(z) + o(1) \\ &= s_n(z) + o(1) \\ &= \mathbb{E}[s_n(z)] + o(1). \end{aligned} \tag{21}$$

6.2 First Approximation

Proposition 6.2.1. *For R_{n-1} defined above,*

$$\frac{1}{n}X^*R_{n-1}X = \frac{1}{n}\mathrm{tr}R_{n-1} + o(1).$$

Before we can prove Proposition 6.2.1, we need a few calculations and results.

Let A be any square matrix. We can write

$$A = \frac{A+A^*}{2} + \frac{A-A^*}{2}.$$

Let $H = \frac{A+A^*}{2}$ and $Q = -i\frac{A-A^*}{2}$. Then,

$$A = H + iQ.$$

Note that H is Hermitian because

$$H^* = \left(\frac{A+A^*}{2}\right)^* = \frac{A^*+A}{2} = H.$$

Similarly, Q is Hermitian because

$$Q^* = \left(-i\;\frac{A-A^*}{2}\right)^* = i\frac{A^*-A}{2} = -i\;\frac{-A^*+A}{2} = Q.$$

Let ξ_j, for $j = 1, \ldots, n$, be the eigenvalues of H with corresponding eigenvectors e_j. By the Spectral Theorem, we can write

$$\begin{aligned} H &= \sum_{j=1}^{n} \xi_j e_j e_j^* \\ &= \sum_{\xi_j>0} \xi_j e_j e_j^* + \sum_{\xi_j<0} \xi_j e_j e_j^*. \end{aligned}$$

Let $P = \sum_{\xi_j>0} \xi_j e_j e_j^*$, and let $N = -\sum_{\xi_j<0} \xi_j e_j e_j^*$. Note that these are positive semi-definite matrices. We can write

$$H = P - N.$$

Similarly for Q, let $\tilde{\xi}_j$, for $j = 1, \ldots, n$, be its eigenvalues with corresponding eigenvectors $\tilde{e}_j$. By the Spectral Theorem,

$$\begin{aligned} Q &= \sum_{j=1}^{n} \tilde{\xi}_j \tilde{e}_j \tilde{e}_j^* \\ &= \sum_{\tilde{\xi}_j>0} \tilde{\xi}_j \tilde{e}_j \tilde{e}_j^* + \sum_{\tilde{\xi}_j<0} \tilde{\xi}_j \tilde{e}_j \tilde{e}_j^*. \end{aligned}$$

Let $\tilde{P} = \sum_{\tilde{\xi}_j>0} \tilde{\xi}_j \tilde{e}_j \tilde{e}_j^*$, and let $\tilde{N} = -\sum_{\tilde{\xi}_j<0} \tilde{\xi}_j \tilde{e}_j \tilde{e}_j^*$. So, we can write

$$Q = \tilde{P} - \tilde{N}.$$

Returning to our matrix A, we have

$$A = P - N + i\tilde{P} - i\tilde{N},$$

a combination of four positive semi-definite matrices.

Let $a = ||A||_{op}$, and normalize the matrices by $\frac{1}{2aK\sqrt{n-1}}$ to get

$$\frac{A}{2aK\sqrt{n-1}} = \frac{P - N + i\tilde{P} - i\tilde{N}}{2aK\sqrt{n-1}}.$$

In order to apply Talagrand's Inequality (Proposition D.0.2), we need to prove that for any positive semi-definite matrix T with $||T||_{op} \leq a$, $F_T(X) = \frac{X^*TX}{2aK\sqrt{n-1}}$ is convex and 1-Lipschitz.[3]

Proposition 6.2.2. *Let T be an $(n-1) \times (n-1)$ positive semi-definite matrix. Let $F_T(X) = \frac{X^*TX}{2aK\sqrt{n-1}}$, where X is an $(n-1)$ column vector with entries uniformly bounded by K. Then, F is convex.*

Proof. Parameterize X by $X = X_0 + tZ$, for some $(n-1)$ column vector $Z \neq 0$. Let $g(t) = F_T(X_0 + tZ)$. We will prove that F_T is convex by showing that $g''(t) \geq 0$ for all $t \in \mathbb{R}$, which will be sufficient due to Theorem C.0.4. We have

$$\begin{aligned} g(t) &= \frac{1}{2aK\sqrt{n-1}}(X_0 + tZ)^*T(X_0 + tZ) \\ &= \frac{1}{2aK\sqrt{n-1}}(X_0^*TX_0 + tZ^*TX_0 + tX_0^*TZ + t^2Z^*TZ) \\ &= F_T(X_0) + \frac{t}{aK\sqrt{n-1}}\Re(X_0^*TZ) + t^2F_T(Z). \end{aligned}$$

Then,

$$g''(t) = 2F_T(Z) > 0.$$

□

By Proposition 6.2.2, replacing T with $P, N, \tilde{P}$, and $\tilde{N}$, respectively, we see that $F_P, F_N, F_{\tilde{P}}$, and $F_{\tilde{N}}$ are convex.

Proposition 6.2.3. *The function $F_T(X) = \frac{X^*TX}{2aK\sqrt{n-1}}$ is 1-Lipschitz.*

Proof. Note that

$$\begin{aligned} \frac{1}{2aK\sqrt{n-1}}[(X_1^* - X_2^*)TX_1 + X_2^*T(X_1 - X_2)] &= \frac{1}{2aK\sqrt{n-1}}(X_1^*TX_1 - X_2^*TX_1 + X_2^*TX_1 - X_2^*TX_2) \\ &= \frac{1}{2aK\sqrt{n-1}}(X_1^*TX_1 - X_2^*TX_2) \\ &= \frac{1}{2aK\sqrt{n-1}}X_1^*TX_1 - \frac{1}{2aK\sqrt{n-1}}X_2^*TX_2 \\ &= F_T(X_1) - F_T(X_2). \end{aligned}$$

Therefore, by the triangle inequality,

$$\begin{aligned} |F_T(X_1) - F_T(X_2)| &= \frac{1}{2aK\sqrt{n-1}}[(X_1^* - X_2^*)TX_1 + X_2^*T(X_1 - X_2)] \\ &\leq \frac{1}{2aK\sqrt{n-1}}\Big[||(X_1^* - X_2^*)TX_1|| + ||X_2^*T(X_1 - X_2)||\Big] \\ &= \frac{1}{2aK\sqrt{n-1}}\Big[||X_1^* - X_2^*|| \cdot ||T|| \cdot ||X_1|| + ||X_2^*|| \cdot ||T|| \cdot ||X_1 - X_2||\Big] \\ &\leq \frac{1}{2aK\sqrt{n-1}}\Big[aK\sqrt{n-1}\,||X_1^* - X_2^*|| + aK\sqrt{n-1}\,||X_1 - X_2||\Big] \\ &= \frac{1}{2aK\sqrt{n-1}}\left[2aK\sqrt{n-1}\,||X_1 - X_2||\right] \\ &= ||X_1 - X_2||. \end{aligned}$$

[3]We consider a function f to be 1-Lipschitz if $|f(x) - f(y)| \leq c|x - y|$, for all x and y and for $c \leq 1$.

Thus, F_T is 1-Lipschitz. $\square$

To see that Proposition 6.2.3 implies $F_P, F_N, F_{\tilde{P}}$, and $F_{\tilde{N}}$ are 1-Lipschitz, observe that the operator norms of $P, N, \tilde{P}$, and $\tilde{N}$ cannot exceed a. This is because

$$||P||_{op} = \sqrt{\max_{\xi_j > 0} \xi_j} \leq ||H||_{op} \leq a, \quad ||N||_{op} = \sqrt{\max_{\xi_j < 0} \xi_j} \leq ||H||_{op} \leq a,$$
$$||\tilde{P}||_{op} = \sqrt{\max_{\tilde{\xi}_j > 0} \tilde{\xi}_j} \leq ||Q||_{op} \leq a, \quad \text{and} \quad ||\tilde{N}||_{op} = \sqrt{\max_{\tilde{\xi}_j < 0} \tilde{\xi}_j} \leq ||Q||_{op} \leq a.$$

So, by replacing T with $P, N, \tilde{P}$, and $\tilde{N}$, respectively, in Proposition 6.2.3, we find that $F_P, F_N, F_{\tilde{P}}$, and $F_{\tilde{N}}$ are 1-Lipschitz.

Now, note that

$$\begin{aligned}\left\{\left|\frac{X^*AX}{2aK\sqrt{n-1}} - \mathbb{E}\left[\frac{X^*AX}{2aK\sqrt{n-1}}\right]\right| \geq \varepsilon\right\} \subseteq &\left\{\left|\frac{X^*PX}{2aK\sqrt{n-1}} - \mathbb{E}\left[\frac{X^*PX}{2aK\sqrt{n-1}}\right]\right| \geq \frac{\varepsilon}{4}\right\} \\ &\cup \left\{\left|\frac{X^*NX}{2aK\sqrt{n-1}} - \mathbb{E}\left[\frac{X^*NX}{2aK\sqrt{n-1}}\right]\right| \geq \frac{\varepsilon}{4}\right\} \\ &\cup \left\{\left|\frac{X^*\tilde{P}X}{2aK\sqrt{n-1}} - \mathbb{E}\left[\frac{X^*\tilde{P}X}{2aK\sqrt{n-1}}\right]\right| \geq \frac{\varepsilon}{4}\right\} \\ &\cup \left\{\left|\frac{X^*\tilde{N}X}{2aK\sqrt{n-1}} - \mathbb{E}\left[\frac{X^*\tilde{N}X}{2aK\sqrt{n-1}}\right]\right| \geq \frac{\varepsilon}{4}\right\}.\end{aligned}$$

By Talagrand's Inequality (Proposition D.0.2),

$$\left\{\left|\frac{X^*PX}{2aK\sqrt{n-1}} - \mathbb{E}\left[\frac{X^*PX}{2aK\sqrt{n-1}}\right]\right| \geq \frac{\varepsilon}{4}\right\} \leq Ce^{-c\frac{\varepsilon^2}{16}},$$

for absolute constants $c, C > 0$. This inequality holds for $N, \tilde{P}$, and $\tilde{N}$, as well. Therefore,

$$\left\{\left|\frac{X^*AX}{2aK\sqrt{n-1}} - \mathbb{E}\left[\frac{X^*AX}{2aK\sqrt{n-1}}\right]\right| \geq \varepsilon\right\} \leq 4Ce^{-c\frac{\varepsilon^2}{16}}.$$

This can be written

$$\left\{\left|\frac{1}{n}X^*AX - \mathbb{E}\left[\frac{1}{n}X^*AX\right]\right| \geq \varepsilon\frac{2aK\sqrt{n-1}}{n}\right\} \leq 4Ce^{-c\frac{\varepsilon^2}{16}}.$$

Since

$$\frac{\sqrt{n-1}}{n} < \frac{\sqrt{n}}{n} = \frac{1}{\sqrt{n}},$$

we have

$$\begin{aligned}4Ce^{-c\frac{\varepsilon^2}{16}} &\geq P\left(\left|\frac{1}{n}X^*AX - \mathbb{E}\left[\frac{1}{n}X^*AX\right]\right| \geq \varepsilon\frac{2aK\sqrt{n-1}}{n}\right) \\ &\geq P\left(\left|\frac{1}{n}X^*AX - \mathbb{E}\left[\frac{1}{n}X^*AX\right]\right| \geq \varepsilon\frac{2aK}{\sqrt{n}}\right).\end{aligned}$$

Let $\varepsilon = n^{1/4}$. Then,

$$P\left(\left|\frac{1}{n}X^*AX - \mathbb{E}\left[\frac{1}{n}X^*AX\right]\right| \geq \frac{2aK}{n^{1/4}}\right) \leq 4Ce^{-c\frac{\sqrt{n}}{16}}. \tag{22}$$

Now we are ready to prove Proposition 6.2.1.

Proof of Proposition 6.2.1. Let $B = \left\{ \left| \frac{1}{n} X^* A X - \mathbb{E}\left[\frac{1}{n} X^* A X\right] \right| \geq \frac{2aK}{n^{1/4}} \right\}$. Recall that A is deterministic, not random, so

$$\begin{aligned}
\mathbb{E}\left[\frac{1}{n} X_n^* A X_n\right] &= \frac{1}{n} \sum_{i,j=n-1}^{n-1} \mathbb{E}[\overline{\xi}_{in} a_{ij} \xi_{nj}] \\
&= \frac{1}{n} \sum_{i,j=1}^{n-1} a_{ij} \mathbb{E}[\overline{\xi}_{in} \xi_{nj}] \\
&= \frac{1}{n} \sum_{i,j=1}^{n-1} a_{ij} \delta_{ij} \\
&= \frac{1}{n} \mathrm{tr} A.
\end{aligned}$$

Thus, we can write

$$B = \left\{ \left| \frac{1}{n} X^* A X - \frac{1}{n} \mathrm{tr} A \right| \geq \frac{2aK}{n^{1/4}} \right\}. \tag{23}$$

Now consider what happens when we replace the deterministic matrix A in (23) with the random variable R_{n-1}. Let $S = \left\{ \left| \frac{1}{n} X^* R_{n-1} X - \frac{1}{n} \mathrm{tr} R_{n-1} \right| \geq \frac{2aK}{n^{1/4}} \right\}$. We want to estimate $P(S)$, and we will do so with the help of Example 5.1.5 from Rick Durrett's book [6], *Probability: Theory and Examples*:

> **Example 5.1.5** Suppose X and Y are independent. Let φ be a function with $\mathbb{E}[|\varphi(X,Y)|] < \infty$ and let $g(y) = \mathbb{E}[\varphi(X,y)]$. Then $\mathbb{E}[\varphi(X,Y)|Y] = g(Y)$.

To be consistent with Durrett's example, let $X = X$ and $Y = R_{n-1}$. For deterministic x and r, define φ by

$$\varphi(x,r) = \chi\left(\left| \frac{1}{n} x^* r x - \frac{1}{n} \mathrm{tr}(r) \right| \geq \frac{2aK}{n^{1/4}} \right),$$

where χ is the characteristic function. Then,

$$g(r) = \mathbb{E}[\varphi(X,r)] = \int_\Omega \varphi\left(X(\omega), r\right) d\omega.$$

The random variables X and R_{n-1} are independent, and it is easy to check that $\mathbb{E}[|\varphi(X,R_{n-1})|] \leq 1$, so

$$g(R_{n-1}) = \mathbb{E}[\varphi(X,R_{n-1})|R_{n-1}]$$

according to the example. Observe that, since r is deterministic,

$$g(r) = P\left(\left| \frac{1}{n} X^* r X - \frac{1}{n} \mathrm{tr}(r) \right| \geq \frac{2aK}{n^{1/4}} \right) \leq 4Ce^{-c\frac{\sqrt{n}}{16}}$$

by inequality (22). Now, since

$$P(S) = \mathbb{E}[\varphi(X,R_{n-1})] = \mathbb{E}\Big[\mathbb{E}[\varphi(X,R_{n-1})|R_{n-1}]\Big] = \mathbb{E}[g(R_{n-1})],$$

we have

$$\begin{aligned}
P(S) &= \int_\Omega g(R_{n-1}) d\omega \\
&= \left| \int_\Omega g(R_{n-1}) d\omega \right| \\
&\leq \int_\Omega |g(R_{n-1})| d\omega \\
&\leq \int_\Omega 4Ce^{-c\frac{\sqrt{n}}{16}} d\omega \\
&= 4Ce^{-c\frac{\sqrt{n}}{16}} \int_\Omega d\omega \\
&= 4Ce^{-c\frac{\sqrt{n}}{16}}.
\end{aligned}$$

Therefore,

$$P\left(\left| \frac{1}{n} X^* R_{n-1} X - \frac{1}{n} \mathrm{tr} R_{n-1} \right| \geq \frac{2aK}{n^{1/4}} \right) \leq 4Ce^{-c\frac{\sqrt{n}}{16}},$$

which means

$$\frac{1}{n} X^* R_{n-1} X = \frac{1}{n} \mathrm{tr} R_{n-1} = o(1).$$

$\square$

6.3 Justification of the Remaining Equalities

By Equations (10) and (12),

$$\begin{aligned}
\frac{1}{n} \mathrm{tr} R_{n-1} &= \frac{1}{n} \left[\sqrt{n(n-1)} s_{n-1} \left(\frac{\sqrt{n}}{\sqrt{n-1}} z \right) \right] \\
&= s_{n-1}(z) + O\left(\frac{1}{n} \right).
\end{aligned}$$

According to Equation (13), $s_{n-1}(z) = s_n(z) + O\left(\frac{1}{n}\right)$. Lastly, we showed in Theorem 5.0.1, that $s_n(z) = \mathbb{E}[s_n(z)] + o(1)$. Thus, we have justified all steps leading to

$$\frac{1}{n} X^* R_{n-1} X = \mathbb{E}[s_n(z)] + o(1).$$

6.4 Proof of Theorem 6.0.1

By replacing $\frac{1}{n} X^* R_{n-1} X$ with $\mathbb{E}[s_n(z)] + o(1)$ in Equation (19), we get

$$\mathbb{E}[s_n(z)] = \mathbb{E}\left[\frac{1}{-z - \mathbb{E}[s_n(z)] + o(1)} \right].$$

Since z is fixed,

$$\mathbb{E}[s_n(z)] = \frac{1}{-z - \mathbb{E}[s_n(z)] + o(1)} + o(1). \tag{24}$$

The denominator on the right-hand side is bounded, so we can rewrite Equation (24) as

$$\mathbb{E}[s_n(z)] = \frac{1}{-z - \mathbb{E}[s_n(z)]} + o(1), \tag{25}$$

because

$$\frac{1}{-z-\mathbb{E}[s_n(z)]+o(1)}-\frac{1}{-z-\mathbb{E}[s_n(z)]}=\frac{o(1)}{\big(-z-\mathbb{E}[s_n(z)]+o(1)\big)\big(-z-\mathbb{E}[s_n(z)]\big)}=o(1).$$

Proof of Theorem 6.0.1. Equation (25) tells us that if $L=\lim_{n\to\infty}\mathbb{E}[s_n(z)]$ exists, it must satisfy the fixed point formula

$$L=\frac{1}{-z-L}.$$

Since each $s_n(z)$ has positive imaginary part by Equation (6), L must have positive imaginary part. Therefore, if the limit exists,

$$L=\frac{-z+\sqrt{z^2-4}}{2} \tag{26}$$

because we have chosen the convention that we take the branch of the square root with positive imaginary part. Note that by Theorem 4.1.1,

$$s_{sc}(z)=\frac{-z+\sqrt{z^2-4}}{2},$$

so if the limit exists,

$$\lim_{n\to\infty}\mathbb{E}[s_n(z)]=s_{sc}(z).$$

In order to prove that the limit does exist, let $W=\{z:\Im(z)>\delta\}$, and let $f_n=\mathbb{E}[s_n]$. This function is analytic by Proposition B.0.2 for all $z\in W$, and $|f_n(z)|$ is uniformly bounded by $\frac{1}{\delta}$ according to Equation (28) for all $z\in W$. Let g_{n_k} be any subsequence of f_n. Since $|g_{n_k}|<\frac{1}{\delta}$, there exists a subsequence $g_{n_{k_j}}$ that converges uniformly on compact subsets of W to a function G by Montel's Theorem (C.0.2). However, since $\{g_{n_{k_j}}\}\subseteq\{f_n\}$, we have

$$\lim_{n\to\infty}g_{n_{k_j}}=L$$

for all $z\in W$. Thus, it must be that $g_{n_k}(z)$ converges to $s_{sc}(z)$ for all $z\in W$.

We claim that this implies f_n converges to s_{sc} for all $z\in W$. Assume, for contradiction, that f_n does not converge to s_{sc} on W. This means, for some $\varepsilon>0$, there does not exist an N such that $n\geq N$ implies $||f_n-s_{sc}||_\infty<\varepsilon$. Thus, there exists a sequence n_k that goes to infinity such that $||f_{n_k}-s_{sc}||_\infty\geq\varepsilon$. However, we have shown that every subsequence of f_n has a subsequence that converges on W to s_{sc}. That is, for j large, $||f_{n_{k_j}}-s_{sc}||_\infty<\varepsilon$, where $f_{n_{k_j}}$ is a subsequence of f_{n_k}. Thus, we have a contradiction, as this would imply $||f_{n_k}-s_{sc}||_\infty<\varepsilon$, for large k. So, it must be that f_n converges to s_{sc} on W.

Since this holds for all $\delta>0$, $\mathbb{E}[s_n(z)]$ converges to $s_{sc}(z)$ for fixed z in the upper half plane.

□

7 Step Three

Theorem 7.0.1. *Outside of a null set, $s_n(z,\omega)$ converges to $s_{sc}(z)$ for every z in the upper half plane.*

Proof. Theorem 5.0.1 and Theorem 6.0.1 have shown that for a fixed z in the upper half of the complex plane, $s_n(z,\omega)$ converges to $s_{sc}(z)$ almost surely. This means that for each z in the upper half plane, there exists a null set N_z such that for all $\omega\in N_z^c$,

$$s_n(z,\omega)\to s_{sc}(z).$$

Let $\mathbb{C}_{\mathbb{Q}}=\{z_j\}_{j=1}^{\infty}$ be the set of all points in the upper half plane with rational coordinates. Then, let $N=\cup N_{z_j}$. The set N is a null set because it is the union of countably many sets, all of which have measure zero. So,

$$s_n(z,\omega)\to s_{sc}(z)$$

for all $\omega \in N^c$ and all $z \in \mathbb{C}_{\mathbb{Q}}$. Fix $m > 0$, and let $\mathbb{C}_m = \{z : \Im(z) > \frac{1}{m}\}$. Therefore, by the elementary bound (Lemma 4.0.1),

$$|s_n(z)| \leq \frac{1}{|\Im(z)|} < m,$$

for all $z \in \mathbb{C}_m$. We can now apply Vitali's Theorem (C.0.3) to conclude that

$$s_n(z, \omega) \to s_{sc}(z)$$

for all $\omega \in N^c$ and all $z \in \mathbb{C}_m$. Since this holds for all $m > 0$, we see

$$s_n(z, \omega) \to s_{sc}(z)$$

for all $\omega \in N^c$ and all z in the upper half of the complex plane. □

8 Proof of the Semicircular Law

We have everything we need to prove Wigner's semicircular law.

Proof of Theorem 3.0.1. By Theorem 7.0.1, we know that $s_n(z)$ converges to $s_{sc}(z)$ almost surely for every z in the upper half plane. Then, according to Theorem 3.0.2, this means μ_n converges to μ_{sc} weakly almost surely. □

Appendix A Eigenvalues

A.1 Eigenvalues of Matrices are Most Likely Distinct

Theorem A.1.1. *Let A be any square matrix. Assume that the joint distribution of its elements $[a_{ij}]$ is absolutely continuous with respect to Lebesgue measure. Then, with probability 1, the eigenvalues of A are distinct.*

Proof. Let p be a polynomial defined by

$$\begin{aligned} p(x) &= \det(xI - A) \\ &= x^n + (-1)\mathrm{tr}(A)x^{n-1} + \ldots + (-1)^n \det(A) \\ &= x^n + \sigma_1 x^{n-1} + \sigma_2 x^{n-2} + \ldots + \sigma_{n-1}x + \sigma_n. \end{aligned}$$

Let $r_1, \ldots r_n$ be the roots (not necessarily distinct) of the polynomial p. Then,

$$(r_1 - r_2)^2(r_1 - r_3)^2(r_1 - r_4)^2 \ldots (r_2 - r_3)^2 \ldots (r_{n-1} - r_n)^2 = g(\sigma_1, \sigma_2, \ldots, \sigma_n).$$

We will refer to the function g as the discriminant D. It is written in terms of the elementary symmetric functions $\sigma_1, \ldots, \sigma_n$, which themselves are determined by the elements $[a_{ij}]$ of the matrix A. Note that if any of the roots of p are not distinct, then the discriminant is zero. In order to prove that the eigenvalues are distinct with probability 1, we will show that the zero locus of p has measure zero.

We know

$$\begin{vmatrix} 1 & r_1 \\ 1 & r_2 \end{vmatrix} = r_2 - r_1$$

and

$$\begin{vmatrix} 1 & r_1 & r_1^2 \\ 1 & r_2 & r_2^2 \\ 1 & r_3 & r_3^2 \end{vmatrix} = (r_2 - r_1)(r_3 - r_2)(r_3 - r_1).$$

Therefore, by induction, we see that

$$\begin{vmatrix} 1 & r_1 & \dots & r_1^{n-1} \\ 1 & r_2 & \dots & r_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & r_n & \dots & r_n^{n-1} \end{vmatrix} = \prod_{i<j}^{n} (r_j - r_i).$$

However, we want to find a formula for

$$\prod_{i<j}^{n} (r_j - r_i)^2$$

using determinants. Since A is a square matrix,

$$(\det(A))^2 = \det(A)\det(A) = \det(A)\det(A^T) = \det(A^T A).$$

So,

$$\begin{aligned} \det(A^T A) &= \det\left(\begin{bmatrix} 1 & 1 & \dots & 1 \\ r_1 & r_2 & \dots & r_n^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ r_1^{n-1} & r_2^{n-1} & \dots & r_n^{n-1} \end{bmatrix} \begin{bmatrix} 1 & r_1 & \dots & r_1^{n-1} \\ 1 & r_2 & \dots & r_2^{n-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & r_n & \dots & r_n^{n-1} \end{bmatrix}\right) \\ &= \begin{vmatrix} n & \sum r & \dots & \sum r^{n-1} \\ \sum r & \sum r^2 & \dots & \sum r^n \\ \vdots & \vdots & \ddots & \vdots \\ \sum r^{n-1} & \sum r^n & \dots & \sum r^{2n-1} \end{vmatrix} \\ &= \prod_{i<j}^{n} (r_j - r_i)^2, \end{aligned}$$

a polynomial in power sums. Now, we want to express this product in terms of the elementary symmetric polynomials:

$$\sigma_j(n) = \sum_{k_1<k_2<\ldots<k_j}^{n} \left(\prod_{i=1}^{j} r_{k_i} \right).$$

Note that they can be written as

$$\sigma_j(n) = r_i \frac{\partial}{\partial r_i} (\sigma_j(n)) + \frac{\partial}{\partial r_i} (\sigma_{j+1}(n)).$$

Let $B_\ell = \sum_{i=1}^{n} r_i^\ell$. Then,

$$\begin{aligned}
\sum_{j=0}^{m-1}(-1)^j B_{m-j}\sigma_j(n) &= \sum_{j=0}^{m-1}(-1)^j \left[\sum_{i=1}^{n} r_i^{m-j}\right]\left[r_i \frac{\partial}{\partial r_i}(\sigma_j(n)) + \frac{\partial}{\partial r_i}(\sigma_{j+1}(n))\right] \\
&= \sum_{i=1}^{n}\sum_{j=0}^{m-1}\left[(-1)^j r_i^{m-j}\left(r_i \frac{\partial}{\partial r_i}(\sigma_j(n)) + \frac{\partial}{\partial r_i}(\sigma_{j+1}(n))\right)\right] \\
&= \sum_{i=1}^{n}\left[r_i^m\left(r_i \frac{\partial}{\partial r_i}(\sigma_0(n)) + \frac{\partial}{\partial r_i}(\sigma_1(n))\right) - r_i^{m-1}\left(r_i \frac{\partial}{\partial r_i}(\sigma_1(n)) + \frac{\partial}{\partial r_i}(\sigma_2(n))\right)\right. \\
&\qquad \left. + \ldots + (-1)^{m-1} r_i \left(r_i \frac{\partial}{\partial r_i}(\sigma_{m-1}(n)) + \frac{\partial}{\partial r_i}(\sigma_m(n))\right)\right] \\
&= \sum_{i=1}^{n}\left[r_i^{m+1}\frac{\partial}{\partial r_i}(\sigma_0(n)) + \left(r_i^m \frac{\partial}{\partial r_i}(\sigma_1(n)) - r_i^m \frac{\partial}{\partial r_i}(\sigma_1(n))\right)\right. \\
&\qquad \left. + \ldots + \left((-1)^{m-2} r_i^2 \frac{\partial}{\partial r_i}(\sigma_{m-1}(n)) + (-1)^{m-1} r_i^2 \frac{\partial}{\partial r_i}(\sigma_{m-1}(n))\right) + (-1)^{m-1} r_i \frac{\partial}{\partial r_i}(\sigma_m(n))\right] \\
&= \sum_{i=1}^{n}(-1)^{m-1} r_i \frac{\partial}{\partial r_i}(\sigma_m(n)) \\
&= (-1)^{m-1} m(\sigma_m(n)), \qquad (27)
\end{aligned}$$

where the final equality follows from Euler's theorem for homogeneous functions. This gives us Newton's identities:

$$\sigma_m = -\frac{1}{m}\sum_{j=0}^{m-1}(-1)^{m-j} B_{m-j}\sigma_j.$$

So we have

$$\begin{vmatrix} n & B_1 & \ldots & B_{n-1} \\ B_1 & B_2 & \ldots & B_n \\ \vdots & \vdots & \ddots & \vdots \\ B_{n-1} & B_n & \ldots & B_{2n-1} \end{vmatrix} = \prod_{i<j}^{n}(r_j - r_i)^2.$$

From Equation (27), we find that

$$\begin{aligned}
B_1 &= \sigma_1 \\
B_2 &= \sigma_1^2 - 2\sigma_2 \\
B_3 &= \sigma_1^3 - 3\sigma_1\sigma_2 + 3\sigma_3 \\
B_4 &= \sigma_1^4 - 4\sigma_1^2\sigma_2 + 4\sigma_1\sigma_3 + 2\sigma_2^2 - 4\sigma_4 \\
&\vdots
\end{aligned}$$

Thus,

$$\begin{vmatrix} n & \sigma_1 & \sigma_1^2 - 2\sigma_2 & \ldots \\ \sigma_1 & \sigma_1^2 - 2\sigma_2 & \sigma_1^3 - 3\sigma_1\sigma_2 + 3\sigma_3 & \ldots \\ \sigma_1^2 - 2\sigma_2 & \sigma_1^3 - 3\sigma_1\sigma_2 + 3\sigma_3 & \sigma_1^4 - 4\sigma_1^2\sigma_2 + 4\sigma_1\sigma_3 + 2\sigma_2^2 - 4\sigma_4 & \ldots \\ \vdots & \vdots & \vdots & \ddots \end{vmatrix} = \prod_{i<j}^{n}(r_j - r_i)^2.$$

This is the discriminant D we wanted to find. It is written in terms of the elementary symmetric functions $\sigma_1, \ldots, \sigma_n$, which themselves are determined by the entries $[a_{ij}]$ of the matrix A. Note that if any of the roots of p are not distinct, then $\prod_{i<j}^{n}(r_j - r_i)^2 = 0$, which is a polynomial in $\mathbb{R}^n$. The zero locus of this polynomial has measure zero. Therefore, we can say the eigenvalues of A are distinct with probability 1. $\square$

A.2 Cauchy's Interlace Theorem

Theorem A.2.1. *Let A_n be an $n \times n$ Hermitian matrix. Write*

$$A_n = \begin{bmatrix} B_{n-1} & X \\ X^* & a_{nn} \end{bmatrix},$$

where B_{n-1} is the top left $(n-1) \times (n-1)$ principal submatrix of A_n, X is the rightmost $(n-1)$ column vector of A_n, X^* is the complex conjugate transpose of X, and a_{nn} is the bottom rightmost entry of A_n. Note that this means B_{n-1} is also a Hermitian matrix. Let $\alpha_1 \leq \ldots \leq \alpha_n$ be the eigenvalues of A_n, and let $\beta_1 \leq \ldots \leq \beta_{n-1}$ be the eigenvalues of B_{n-1}. Then,

$$\alpha_k \leq \beta_k \leq \alpha_{k+1}.$$

Proof. Let u_i, for $i = 1, \ldots, n$, and v_j, for $j = 1, \ldots, n-1$ be the eigenvectors of A_n and B_{n-1}, respectively. Note that $u_i^* u_j = \delta_{ij}$, where δ_{ij} is the Kronecker delta, because A_n is Hermitian. Similarly, $v_i^* v_j = \delta_{ij}$. Let

$$w_i = \begin{bmatrix} v_i \\ 0 \end{bmatrix}.$$

Let $1 \leq k \leq n-1$. Denote the span of $u_k, \ldots, u_n$ by S_1 and the span of $w_1, \ldots, w_k$ by S_2. Note that $\dim(S_1) = n - k + 1$ and $\dim(S_2) = k$. By the elementary formula

$$\dim(S_1 \cap S_2) = \dim(S_1) + \dim(S_2) - \dim(S_1 + S_2),$$

we know that $\dim(S_1 \cap S_2) > 0$, since $\dim(S_1 + S_2) \leq n$. Thus, there exists $y \in S_1 \cap S_2$ such that $y^* y = 1$ and $\alpha_k \leq y^* A_n y \leq \beta_k$. This is because y is a linear combination of the eigenvectors $u_k, \ldots, u_n$ of A_n and a linear combination of the eigenvectors $w_1, \ldots, w_k$ of B_{n-1}. Thus,

$$\alpha_k \leq y^* A_n y \leq \alpha_n.$$

Also,

$$y^* A_n y = y^* B_{n-1} y$$

and

$$\beta_1 \leq y^* B_{n-1} y \leq \beta_k$$

imply

$$\beta_1 \leq y^* A_n y \leq \beta_k.$$

Therefore,

$$\alpha_k \leq y^* A_n y \leq \beta_k.$$

Now we will consider

$$-A_n = \begin{bmatrix} -B_{n-1} & -X \\ -X^* & -a_{nn} \end{bmatrix}.$$

Then, the eigenvalues of $-A_n$ are $\tilde{\alpha}_1 \geq \ldots \geq \tilde{\alpha}_n$, where $\tilde{\alpha}_i = -\alpha_i$. Similarly, the eigenvalues of $-B_{n-1}$ are $\tilde{\beta}_1 \geq \ldots \geq \tilde{\beta}_n$, where $\tilde{\beta}_i = -\beta_i$. The eigenvectors are still u_i and v_j for $-A_n$ and $-B_{n-1}$, respectively. Then, let $\tilde{S}_1$ be the span of $u_1, \ldots, u_{k+1}$, and let $\tilde{S}_2$ be the span of $w_k, \ldots, w_n$. As before, there must exist some $z \in \tilde{S}_1 \cap \tilde{S}_2$ such that $z^* z = 1$ and

$$\tilde{\alpha}_1 \geq z^*(-A_n) z \geq \tilde{\alpha}_{k+1},$$

which means $z^* A_n z \leq -\tilde{\alpha}_{k+1} = \alpha_{k+1}$. Similarly,

$$z^*(-A_n) z = z^*(-B_{n-1}) z,$$

which leads to the inequality

$$\tilde{\beta}_k \geq z^*(-B_n) z \geq \tilde{\beta}_n.$$

Thus, $z^* B_{n-1} z \geq -\tilde{\beta}_k = \beta_k$. Therefore, $\beta_k \leq z^* A_n z \leq \alpha_{k+1}$. Putting everything together, we have

$$\alpha_k \leq \beta_k \leq \alpha_{k+1}.$$

□

Appendix B Properties of the Expectation of the Stieltjes Transform of a Measure

Proposition B.0.1. *The expectation of the Stieltjes transform of μ is continuous at all points z in the upper half plane.*

Proof. Fix $\delta > 0$ and let $W = \{z : \Im(z) > \delta\}$. Recall that $s_\mu(z, \omega)$ is continuous for all z with $\Im(z) > 0$ according to Proposition 4.0.3. By Lemma 4.0.1, we know $|s_\mu(z, \omega)| < \frac{1}{\delta}$ for all $z \in W$. This implies $|\mathbb{E}[s_\mu(z)]| < \frac{1}{\delta}$ because

$$\begin{aligned} \Big|\mathbb{E}[s_\mu(z)]\Big| &= \left|\int_\Omega s_\mu(z, \omega) d\omega\right| \\ &\leq \int_\Omega |s_\mu(z, \omega)| d\omega \\ &< \int_\Omega \frac{1}{\delta} d\omega \\ &= \frac{1}{\delta} \int_\Omega d\omega \\ &= \frac{1}{\delta}. \end{aligned} \tag{28}$$

For any $z_j \in W$, let $f_j(z_j) = \mathbb{E}[s_\mu(z_j)]$ so that $\lim_{z_j \to z} f_j(z_j) = f(z)$. Let $g(z, \omega) = \frac{1}{\delta}$, so

$$\Big|\mathbb{E}[s_\mu(z_j)]\Big| < g(z, \omega).$$

Note that $\int_\Omega g(z, \omega) d\omega < \infty$ because

$$\begin{aligned} \int_\Omega g(z, \omega) d\omega &= \int_\Omega \left(\frac{1}{\delta}\right) d\omega \\ &= \frac{1}{\delta} \int_\Omega d\omega \\ &= \frac{1}{\delta}. \end{aligned}$$

Therefore, by the Lebesgue Dominated Convergence Theorem,

$$\lim_{z_j \to z} \int_\Omega s_\mu(z_j, \omega) d\omega = \int_\Omega s_\mu(z, \omega) d\omega,$$

which is to say, $\lim_{z_j \to z} \mathbb{E}[s_\mu(z_j)] = \mathbb{E}[s_\mu(z)]$. Since this holds for any $\delta > 0$, we can say that the expectation of the Stieltjes transform of μ is continuous at all points z in the upper half plane. $\square$

Proposition B.0.2. *For all z in the upper half plane, $\mathbb{E}[s_\mu(z)]$ is analytic.*

Proof. Proposition 4.0.4 tells us that $s_\mu(z, \omega)$ is analytic for all z in the upper half plane. Thus, for any triangle C in the upper half plane,

$$\int_C s_\mu(z, \omega) dz = 0$$

by the Cauchy Integral Theorem.

Then,

$$\int_C \mathbb{E}[s_\mu(z)] dz = \int_C \left[\int_\Omega s_\mu(z, \omega) d\omega\right] dz.$$

By Fubini's Theorem, we can change the order of integration to get

$$\int_C \left[\int_\Omega s_\mu(z,\omega)d\omega\right] dz = \int_\Omega \left[\int_C s_\mu(z,\omega)dz\right] d\omega = 0.$$

Since $\mathbb{E}[s_\mu(z)]$ is continuous in the upper half plane according to Proposition B.0.1, this tells us that $\mathbb{E}[s_\mu(z)]$ is analytic in the upper half plane by Morera's Theorem. □

Theorem B.0.3. *(Poisson Kernel) Let φ be a continuous function on $\mathbb{R}$ with compact support. Let $T_\varphi(z) = \frac{1}{\pi}\Im\left[\int_\mathbb{R} \frac{\varphi(x)dx}{x-z}\right]$, and let $z = a+ib$. Then,*

$$\lim_{z\to a_0} T_\varphi(z) = \varphi(a_0).$$

Proof. Since $T_\varphi(z) = \frac{1}{\pi}\Im\left[\int_\mathbb{R} \frac{\varphi(x)dx}{x-z}\right]$, we have

$$T_\varphi(z) = \frac{1}{\pi}\int_\mathbb{R} \frac{b\varphi(x)dx}{(x-a)^2+b^2} = \int_\mathbb{R} \frac{1}{\pi}\frac{b}{(x-a)^2+b^2}\varphi(x)dx = \int_\mathbb{R} P(z-x)\varphi(x)dx,$$

where P is the Poisson kernel. Choose $\delta > 0$ such that $|\varphi(a) - \varphi(a_0)| < \varepsilon$, for some $\varepsilon > 0$, whenever $|a - a_0| < \delta$. Let $\psi = \psi_1 + \psi_2$, where

$$\psi_1 = \begin{cases} \varphi - \varphi(a_0) & |a-a_0| < \delta \\ 0 & \text{otherwise} \end{cases}$$

and

$$\psi_2 = \begin{cases} 0 & |a-a_0| < \delta \\ \varphi - \varphi(a_0) & \text{otherwise} \end{cases}.$$

Thus, we want to show that

$$\lim_{z\to a_0} T_\psi(z) = 0.$$

Notice that $|\psi_1| < \varepsilon$. Then,

$$T_{-\varepsilon} < T_{\psi_1} < T_\varepsilon,$$

which means $-\varepsilon < T_{\psi_1} < \varepsilon$, so $|T_{\psi_1}| < \varepsilon$.

Next, we need to show that

$$\lim_{z\to a_0} T_{\psi_2}(z) = 0.$$

Since,

$$T_{\psi_2}(z) = \int_{|a-x|\geq\delta} \frac{b\psi_2(x)dx}{(x-a)^2+b^2},$$

the integrand is continuous in x, a, and b. Thus, we can pass the limit inside the integral to get

$$\lim_{z\to a_0} T_{\psi_2}(z) = \int_{|a-x|\geq\delta} \lim_{z\to a_0} \frac{b\psi_2(x)dx}{(x-a)^2+b^2} = 0,$$

because $b \to 0^+$. Therefore,

$$\lim_{z\to a_0} T_\psi(z) = 0,$$

so

$$\lim_{z\to a_0} T_\varphi(z) = \varphi(a_0).$$

□

Appendix C General Theorems

Theorem C.0.1 (Borel-Cantelli Theorem)**.** *Let A_n be a sequence of events in the sample space Ω.*

(i) *If*

$$\sum_{n=1}^{\infty} P(A_n) < \infty,$$

then $P(\limsup_{n\to\infty}(A_n)) = 0$.

(ii) *Suppose each event A_j is independent. If*

$$\sum_{j=1}^{\infty} P(A_j) = \infty,$$

then $P(\limsup_{j\to\infty}(A_j)) = 1$.

Proof. (i) Suppose

$$\sum_{n=1}^{\infty} P(A_n) < \infty.$$

Thus, we know

$$\lim_{k\to\infty} \sum_{n=k}^{\infty} P(A_n) = 0.$$

Let $B_1 = \bigcup_{n=1}^{\infty} A_n$, $B_2 = \bigcup_{n=2}^{\infty} A_n$, and so on. Note that

$$\limsup_{n\to\infty}(A_n) \in B_1 \cap B_2 \cap B_3 \cap \ldots \cap B_\ell \in B_\ell.$$

Since $B_\ell = \bigcup_{n=\ell}^{\infty} A_n$,

$$P(B_\ell) = P(A_\ell \cup A_{\ell+1} \cup \ldots) \leq \sum_{n=\ell}^{\infty} P(A_n).$$

Therefore,

$$P(\limsup_{n\to\infty}(A_n)) \leq P(B_\ell) \leq \sum_{n=\ell}^{\infty} P(A_n).$$

Taking the limit as $n \to \infty$, we get

$$\lim_{n\to\infty} P(\limsup_{n\to\infty}(A_n)) \leq \lim_{\ell\to\infty} \sum_{n=\ell}^{\infty} P(A_n) = 0.$$

(ii) Let the set $S = \limsup_{j\to\infty}(A_j)$. We will show that $P(S^C) = 0$. We know

$$S^C = \left(\bigcap_{k=1}^{\infty} \bigcup_{j=k}^{\infty} A_j \right)^C = \bigcup_{k=1}^{\infty} \bigcap_{j=k}^{\infty} A_j^C.$$

In order to prove that $P(S^C) = 0$, we will prove that $P\left(\bigcap_{j=k}^{\infty} A_j^C\right) = 0$, for each k. Fix $m > 0$ and consider

$$P\left(\bigcap_{j=k}^{k+m} A_j^C \right).$$

Since the events are independent,

$$P\left(\bigcap_{j=k}^{k+m} A_j^C\right) = \prod_{j=k}^{k+m} P(A_j^C).$$

Then,

$$P(A_j^C) = 1 - P(A_j) \le e^{-P(A_j)}.$$

This is because $0 \le P(A_j) \le 1$ and the Taylor series expansion of e^{-x} is given by

$$\begin{aligned} e^{-x} &= 1 - x + \frac{x^2}{2} - \frac{x^3}{6} + \dots \\ &= 1 - x + (c)\frac{x^2}{2}, \end{aligned}$$

where $0 < c$ by the Mean Value Theorem. So,

$$\prod_{j=k}^{k+m} P(A_j^C) \le \prod_{j=k}^{k+m} e^{-P(A_j)} = e^{-\sum_{j=k}^{m+k} P(A_j)}.$$

Taking the limit as $m \to \infty$, $\sum_{j=k}^{m+k} P(A_j) \to \infty$, by assumption. Thus,

$$\lim_{m\to\infty} \prod_{j=k}^{k+m} e^{-P(A_j)} = e^{-\sum_{j=k}^{m+k} P(A_j)} = 0.$$

□

Theorem C.0.2 (Montel's Theorem)**.** *Suppose $f_n \in \mathcal{O}(W)$ is a sequence of analytic functions. Suppose for some $M > 0$, $|f_n(z)| \le M$ for all n and all $z \in W$. Then there exists a subsequence f_{n_k} which converges uniformly on all compact subsets of W to a function $f \in \mathcal{O}(W)$.*

Proof. Choose $z_0 \in W$ and consider the disk $D_r = \{z : |z - z_0| < r\} \subseteq W$, for some $r > 0$. Define the disk D_ρ by

$$D_\rho = \{z : |z - z_0| \le \rho < r\}.$$

Let $\delta = r - \rho$. Then we have,

$$|f_n'(z)| = \left|\frac{1}{2\pi i}\int_{|\zeta - z| = \frac{\delta}{2}} \frac{f_n(\zeta)}{(\zeta - z)^2} d\zeta\right| \le \frac{1}{2\pi}\pi\delta\frac{|f_n(z)|}{(\frac{\delta}{2})^2} = \frac{2|f_n(z)|}{\delta} \le \frac{2M}{r-\rho}.$$

Now, let $R = \{z \in W : z \text{ has rational coordinates}\}$. Thus, we can enumerate the points in this set: $R = \{z_j\}_{j=1}^{\infty}$. Fix $z_1 \in R \cap D_\rho$. Since $|f_n(z_1)|$ is bounded by M, there exists a convergent subsequence $f_{n_k}(z_1)$ by the Bolzano-Weierstrass Theorem. To simplify notation, let $f_{n_1} = f_{11}$, $f_{n_2} = f_{22}$, and so on. Fix $z_2 \in R \cap D_\rho$. Then, since $|f_{1n}(z_2)| \le M$, there exists a convergent subsequence $f_{1n_k}(z_2)$. Let $f_{21} = f_{1n_1}$, $f_{22} = f_{1n_2}$, $f_{23} = f_{1n_3}$, and so on. We continue this pattern for all $z_j \in R \cap D_\rho$.

Let $g_k = f_{kk}$, and consider the subsequence of f_n:

$$g_1, g_2, g_3, \dots .$$

Then $g_n(z_j)$ converges for all $z_j \in R$, because for $n \ge k$, the limit as $n \to \infty$ of $g_n(z_k)$ exists. Let g be the function to which g_n converges. By the Fundamental Theorem of Calculus,

$$|g_n(z_1) - g_n(z_2)| = \left|\int_{z_1}^{z_2} g_n'(\zeta) d\zeta\right| \le \frac{2M}{r-\rho}|z_2 - z_1|.$$

Let $C = \frac{2M}{r-\rho}$ for simplicity. Let $K \subset W$ be compact. Fix $\varepsilon > 0$, and let $\gamma = \frac{\varepsilon}{6C}$. Cover K with open disks D_γ of radius γ centered around each $z_j \in R \cap K$. Since K is compact, finitely many of these disks form a subcover of K. Thus, for every $z \in K$, there exists a $z_j \in R \cap K$ such that $|z - z_j| < 2\gamma$. Also, we know that for any z_j at the center of a disk in the finite subcover, $|g_n(z_j) - g_m(z_j)| \leq \frac{\varepsilon}{3}$ when $n, m \geq N_j$, a positive integer. Let $N = \max\{N_j\}$. Thus, $|g_n(z_j) - g_m(z_j)| \leq \frac{\varepsilon}{3}$ when $n, m \geq N$. Therefore, we have

$$\begin{aligned}|g_n(z) - g_m(z)| &\leq C|z - z_j| + |g_n(z_j) - g_m(z_j)| + C|z_j - z| \\ &< C\left(\frac{2\varepsilon}{6C}\right) + \frac{\varepsilon}{3} + C\left(\frac{2\varepsilon}{6C}\right) \\ &= \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} \\ &= \varepsilon\end{aligned}$$

when $n, m \geq N$. Thus, there exists a subsequence of f_n (namely g_n) which converges uniformly on all compact subsets of W to g.

Also, since each g_n is analytic, g is analytic. To see this, let Γ be any triangle in W. Then, by the Cauchy Integral Theorem,

$$\int_\Gamma g_n(\zeta) d\zeta = 0,$$

for all n. So, by Morera's Theorem, we have

$$\lim_{n\to\infty} \int_\Gamma g_n(\zeta) d\zeta = \int_\Gamma \lim_{n\to\infty} g_n(\zeta) d\zeta = \int_\Gamma g(\zeta) d\zeta = 0.$$

Note that we can bring the limit inside the integral because we have proved uniform convergence of g_n to g on all compact subsets of W. Thus, g is analytic in W. $\square$

Theorem C.0.3 (Vitali's Theorem)**.** *Let f_n, for $n = 1, 2, \ldots$, be analytic functions in $\mathcal{O}(W)$, satisfying $|f_n(z)| \leq M$ for every n and every $z \in W$, and $f_n(z)$ converges as $n \to \infty$ for each z in a subset of W having a limit point in W. Then there exists a function $\tilde{F}$ analytic in W for which $f_n(z)$ converges to $f(z)$ uniformly for all $z \in W$.*

Proof. Let $S \subset W$ be a set with a limit point in W. Let F be the function to which $f_n(z)$ converges for all $z \in S$. Let g_{n_k} be any subsequence of f_n. Then, since $|g_{n_k}| \leq M$, there exists a subsequence $g_{n_{k_j}}$ that converges uniformly on compact subsets of W to $G \in \mathcal{O}(W)$ by Montel's Theorem. However, since $f_n(z)$ converges to $F(z)$ on S, $g_{n_{k_j}}(z)$ converges to $F(z)$ on S. So, $G(z) = F(z)$ on S.

Similarly, let h_{n_k} be any subsequence of f_n. Since $|h_{n_k}| \leq M$, there exists a subsequence $h_{n_{k_j}}$ that converges uniformly on compact subets of W to $H \in \mathcal{O}(W)$ by Montel's Theorem. However, since $f_n(z)$ converges to $F(z)$ on S, $h_{n_{k_j}}(z)$ converges to $F(z)$ on S. So, $H(z) = F(z) = G(z)$ on S. Since H and G are analytic functions that agree on S, $H(z) = G(z)$ for all $z \in W$ by the Identity Theorem. We will call this function $\tilde{F}$. Thus, every subsequence of f_n has a subsequence that converges uniformly on compact sets of W to $\tilde{F}$.

Now, we will prove that f_n converges uniformly on compact sets to $\tilde{F}$. Let $K \subset W$ be compact. Assume, for contradiction, that f_n does not converge to $\tilde{F}$ on K. This means, for some $\varepsilon > 0$, there does not exist an N such that $n \geq N$ implies $||f_n - \tilde{F}||_\infty < \varepsilon$. Thus, there exists a sequence n_k that goes to infinity such that $||f_{n_k} - \tilde{F}||_\infty \geq \varepsilon$. However, we have shown that every subsequence of f_n has a subsequence that converges uniformly on K to $\tilde{F}$. That is, for j large, $||f_{n_{k_j}} - \tilde{F}||_\infty < \varepsilon$, where $f_{n_{k_j}}$ is a subsequence of f_{n_k}. Thus, we have a contradiction, as this would imply $||f_{n_k} - \tilde{F}||_\infty < \varepsilon$, for large k. So, it must be that f_n converges uniformly on compact sets of W to the analytic function $\tilde{F}$. $\square$

Theorem C.0.4. *If $f : \mathbb{R} \to \mathbb{R}$ and $f''(x_0) \geq 0$ for all x_0, then f is convex.*

Proof. Let $s + t = 1$, where $s, t > 0$. Let $x_0 = sx + ty$. We will prove that f is convex by showing $f(x_0) \leq sf(x) + tf(y)$. By Taylor's Theorem and the Mean Value Theorem,

$$\begin{aligned}
sf(x) + tf(y) &= s\left[f(x_0) + f'(x_0)(x - x_0) + \frac{f''(\xi)}{2}(x - x_0)^2\right] + t\left[f(x_0) + f'(x_0)(y - x_0) + \frac{f''(\eta)}{2}(y - x_0)^2\right] \\
&= (s + t)f(x_0) + f'(x_0)\left[s(x - x_0) + t(y - x_0)\right] + \frac{1}{2}\left[f''(\xi)(x - x_0)^2 + f''(\eta)(y - x_0)^2\right] \\
&\geq f(x_0) + f'(x_0)\left[sx + ty - (s + t)x_0\right] \\
&= f(x_0) + f'(x_0)[x_0 - x_0] \\
&= f(x_0).
\end{aligned}$$

So, $sf(x) + tf(y) \geq f(x_0)$. □

Appendix D Inequalities

The following two inequalities were originally proved by Colin McDiarmid [17] and Michel Talagrand [22], respectively, but in this paper we use Terrence Tao's statements of the inequalities [23].

Proposition D.0.1 (McDiarmid's Inequality)**.** *Suppose $X_1, \ldots, X_n$ are independent random variables taking values in ranges $R_1, \ldots, R_n$. Let $F : R_1 \times \cdots \times R_n \to \mathbb{C}$ be a function with the property that if we freeze all but the i^{th} coordinate of $F(x_1, \ldots, x_n)$ for some $1 \leq i \leq n$, then F only fluctuates by most $c_i > 0$. Thus,*

$$|F(x_1, \ldots, x_{i-1}, x_i, x_{i+1}, \ldots, x_n) - F(x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_n)| \leq c_i,$$

for all $x_j \in X_j$ and $x_i' \in X_i$ for $1 \leq j \leq n$. Then, for any $\kappa > 0$, we have

$$P(|F(X) - \mathbb{E}[F(X)]| \geq \kappa\sigma) \leq Ce^{-c\kappa^2},$$

for some absolute constants $C, c > 0$, where $\sigma^2 := \sum_{i=1}^n c_i^2$.

Proposition D.0.2 (Talagrand's Inequality)**.** *Let $K > 0$, and suppose $X_1, \ldots, X_n$ are independent complex variables with $|X_i| \leq K$ for all $1 \leq i \leq n$. Let $F : \mathbb{C}^n \to \mathbb{R}$ be a 1-Lipschitz convex function. Then, for any ε, we have*

$$P(|F(X) - \mathbb{M}[F(X)]| \geq \varepsilon K) \leq Ce^{-c\varepsilon^2},$$

and

$$P(|F(X) - \mathbb{E}[F(X)]| \geq \varepsilon K) \leq Ce^{-c\varepsilon^2},$$

for some absolute constants $C, c > 0$, where $\mathbb{M}[F(X)]$ is the median of $F(X)$.

Appendix E Miscellaneous

E.1 Integral Test

Let $0 < r < 1$. We want to show that

$$\int_1^\infty r^{\sqrt{x}} dx$$

converges. Let $u = \sqrt{x}$, which means $2u\, du = dx$. Then,

$$\int_1^\infty r^{\sqrt{x}} dx = \int_1^\infty 2ur^u du.$$

Applying integration by parts once, we have

$$\begin{aligned}\int_1^\infty 2ur^u du &= \frac{2ur^u}{\log(r)}\bigg|_{u=1}^{\infty} - \frac{2r^u}{\log^2(r)}\bigg|_{u=1}^{\infty} \\ &= \left(0 - \frac{2r}{\log(r)}\right) - \left(0 - \frac{2r}{\log^2(r)}\right) \\ &= \frac{2r}{\log^2(r)}\Big(1 - \log(r)\Big) < \infty.\end{aligned}$$

Therefore, by the Integral Test,

$$\sum_{n=1}^{\infty} r^{\sqrt{n}} < \infty \quad (0 < r < 1).$$

References

[1] Bai, Z. and Silverstein J. *Spectral Analysis of Large Dimensional Random Matrices*. Second Edition. (2010). Springer Series in Statistics. Springer, New York, NY.

[2] Baker, Jr., G. A. *A New Derivation of Newton's Identities and Their Application to the Calculation of eigenvalues of a Matrix*. Journal of the Society for Industrial and Applied Mathematics. **7**, no. 2, (1959). 143-148.

[3] Cook, N. *Two Proofs of Wigner's Semicircular Law*. (2012). http://www.math.ucla.edu/ nickcook/semicircle.pdf

[4] Diaconis, P. *Patterns in Eigenvalues: The 70th Josiah Willard Gibbs Lecture*. Bulletin of the American Mathematical Society. **40**, no. 2, (2003). 155-178.

[5] Diaconis, P. *What is a Random Matrix?*. Notices of the American Mathematical Society. **52**, no. 11, (2005). 1348-1349.

[6] Durrett, R. *Probability: Theory and Examples*. Fourth Edition. Cambridge University Press. (2010). New York, NY. 191-193.

[7] Dyson, F. J. *Statistical Theory of the Energy Levels of Complex Systems*. Journal of Mathematical Physics. **3**, no. 1, (1962). 140-156.

[8] Edelman, A. and Rao, N. R. *Random Matrix Theory*. Actua Numerica. **14**, (2005). 233-297.

[9] Feier, A. R. *Two Proofs of Wigner's Semicircular Law*. (2012). http://www.math.harvard.edu/theses/senior/feier/feier.pdf

[10] Gaudin, M. and Mehta, M. L. *On the Density of Eigenvalues of a Random Matrix*. Nuclear Physics. **18**, (1960). 420-427.

[11] Ikebe, Y., Inagaki, T., and Miyamoto, S. *The Monotonicity Theorem, Cauchy's Interlace Theorem, and the Courant-Fischer Theorem*. The American Mathematical Monthly. **94**, no. 4, (1987). 353-354.

[12] Izenman, A. J. *Introduction to Random Matrix Theory*. (2008). http://astro.temple.edu/ alan/MMST/IntroRMT.PDF

[13] Janson, S. *Resultant and Discriminant of Polynomials*. (2007). http://www2.math.uu.se/ svante/papers/sjN5.pdf

[14] Klenke, A. *Probability Theory: A Comprehensive Course.* (2008). Springer, London, UK. 245-268.

[15] Laaksonen, L. *Quantum Chaos and the Riemann Hypothesis.* (2011). http://uclmaths.org/images/a/a8/Laaksonen-qchaos-18Jan.pdf

[16] Marchenko, V. A. and Pastur, L. A. *Distribution of Eigenvalues for Some Sets of Random Matrices.* Math. USSR Sb.**1**, no. 4, (1967). 457-483.

[17] McDiarmid, C. *On the Method of Bounded Differences.* Surveys in Combinatorics. (1989). Cambridge University Press, Cambridge, UK. 148-188.

[18] Mosk, A. P. and Vellekoop, I. M. *Universal Optimal Transmission of Light Through Disordered Materials.* Physical Review Letters. **101**, no. 12, (2008).

[19] Miller, S. J. and Takloo-Bighash, R. *An Invitation to Modern Number Theory.* (2006). Princeton University Press, Princeton, NJ. 359-403.

[20] Pastur, L. A. *On the Spectrum of Random Matrices.* Theoretical and Mathematical Physics. **10**, no. 1, (1973). 1-68.

[21] Pastur, L. A. *Spectra of Random Self Adjoint Operators.* Russian Mathematical Surveys. **28**, no. 1, (1972). 67-74.

[22] Talagrand, M. *Concentration of Measure and Isoperimetric Inequalities in Product Spaces.* Publications Mathématiques de l'IHÉS. **81**, no. 1, (1995). 73-205.

[23] Tao, T. *Topics in Random Matrix Theory.* American Mathematical Society. (2012). Providence, RI.

[24] Titchmarsh, E. C. *The Theory of Functions.* Second Edition. Oxford University Press. (1976). New York, NY. 168-172

[25] Wigner, E. P. *Characteristic Vectors of Bordered Matrices with Infinite Dimensions.* Annals of Mathematics. **62**, no. 3, (1955). 548-564.

[26] Wigner, E. P. *On the Distribution of the Roots of Certain Symmetric Matrices.* Annals of Mathematics. **67**, no. 2, (1958). 325-327.

[27] Wigner, E. P. *Random Matrices in Physics.* SIAM Review. **9**, no. 1, (1967). 1-23.