2 Wigner Semicircle Law

Recall Wigner semicircle law, introduced in Theorem 1.5.

Theorem (Wigner semicircle law)**.** *Let μ_{GOE} be the empirical measure of GOE. Then, μ_{GOE} converges almost surely to the Wigner semicircle distribution μ_{sc} whose density is defined to be*

$$\mu_{sc}(x)\mathrm{d}x := \frac{1}{2\pi}\sqrt{(4-x^2)_+}\mathrm{d}x. \tag{2.1}$$

In this chapter, we prove Wigner semicircle law in two different ways. The first one by the moment counting is almost the same as the original proof by Wigner. The second one is based on the Stieltjes transform, which is widely used in the random matrix theory.

2.1 Proof by moment counting

We prove a weaker version of Wigner semicircle law, which asserts the convergence in probability instead of the almost sure convergence. Our goal is to show that

$$\mathbb{P}(|\langle \mu, f\rangle - \langle \mu_{sc}, f\rangle| > \epsilon) \to 0$$

for any $f \in C_b(\mathbb{R})$ and for any $\epsilon > 0$. By Weierstrass approximation theorem, f can be approximated by a polynomial. Thus, it suffices to prove that

$$\langle \mu, x^k\rangle \to \langle \mu_{sc}, x^k\rangle.$$

By an explicit calculation, we can find that the $2k$-th and $(2k-1)$-st moments satisfy

$$m_{2k} = C_k, \quad m_{2k-1} = 0, \tag{2.2}$$

where C_k is the k-th Catalan number defined by

$$C_k = \frac{1}{k+1}\binom{2k}{k} = \frac{(2k)!}{(k+1)!k!}. \tag{2.3}$$

Thus, it only remains to prove that

$$\langle \mu, x^k\rangle = \frac{1}{N}\operatorname{tr} H^k \to \begin{cases} C_{k/2} & \text{if } k \text{ is even,} \\ 0 & \text{if } k \text{ is odd.} \end{cases} \tag{2.4}$$

We first prove the following lemma.

Lemma 2.1. *After taking expectation, we have*

$$\mathbb{E}\langle \mu, x^k\rangle \to \begin{cases} C_{k/2} & \text{if } k \text{ is even,} \\ 0 & \text{if } k \text{ is odd.} \end{cases} \tag{2.5}$$

Proof. Consider

$$\mathbb{E}\langle \mu, x^k\rangle = \frac{1}{N}\sum_{i_1,i_2,\cdots,i_k} \mathbb{E}\left[H_{i_1i_2}H_{i_2i_3}\cdots H_{i_{k-1}i_k}H_{i_ki_1}\right]. \tag{2.6}$$

Since each entry in H is independent (up to the constraint on the symmetry), if $\{i_j i_{j+1}\} \neq \{i_{j'} i_{j'+1}\}$ for some $1 \le j \le k$ and all $j' \neq j$, then the summand vanishes. (Here, we let $i_{k+1} = i_1$.). For given $i_1, i_2, \cdots, i_k$, consider a graph $G_{i_1\cdots i_k}$ with vertices $(i_1, i_2, \cdots, i_k)$ and edges $(i_1i_2, i_2i_3, \cdots, i_ki_1)$. From the argument above, it suffices to consider the graphs where vertices i_j and i_{j+1} are connected at least twice. In particular, if the graph $G_{i_1\cdots i_k}$ contains more than $(k/2)+1$ distinct vertices, the contribution from the graph to the sum vanishes.

We now estimate $\mathbb{E}\left[H_{i_1i_2}H_{i_2i_3}\cdots H_{i_{k-1}i_k}H_{i_ki_1}\right]$. Each entry of H contains the factor $N^{-1/2}$, hence the expectation contains the factor $N^{-k/2}$. If we have r distinct vertices in the graph, then, we have $N(N-1)\cdots(N-r+1)$ different choices of labeling the vertices using $1, 2, \cdots, N$. Together with the factor N^{-1} in front of the summation, we conclude that the conribution from the graph $G_{i_1\cdots i_k}$ to the sum is 0 unless $r \geq (k/2)+1$.

So far, we have seen that the only case we need to consider is $r = (k/2)+1$, in particular, the sum converges to 0 when k is odd. When k is even, the graphs we want to consider are doubly connected trees, which can be identified with ordered trees, with $(k/2)+1$ vertices. The number of different ordered trees with $(k/2)+1$ vertices is $C_{k/2}$, thus

$$\frac{1}{N}\sum_{i_1,i_2,\cdots,i_k}\mathbb{E}\left[H_{i_1i_2}H_{i_2i_3}\cdots H_{i_{k-1}i_k}H_{i_ki_1}\right]\to C_{k/2} \tag{2.7}$$

when k is even. □

In order to complete the proof of the semicircle law, it is enough to prove that

$$\mathbb{E}\left[\langle\mu,x^k\rangle^2\right]-\left[\mathbb{E}\langle\mu,x^k\rangle\right]^2\to 0 \tag{2.8}$$

and use Chebyshev inequality.

Problem 1. *Prove* (2.8).

2.2 Stieltjes transform

We briefly study the basic properties of Stieltjes transform.

Definition 2.2. For a finite measure ν on the real line, the Stieltjes transform of ν, m_ν is defined by

$$m_\nu(z) := \int_{\mathbb{R}}\frac{\nu(\mathrm{d}x)}{x-z} \tag{2.9}$$

for $z\in\mathbb{C}\backslash\mathbb{R}$.

If we let $z = E+\mathrm{i}\eta$, then we have the following trivial properties.

Lemma 2.3. *Let m_ν be the Stieltjes transform of a finite measure ν on the real line.*

1. *If $\eta>0$, then $\operatorname{Im} m_\nu(z)>0$.*
2. *If $\nu(\mathbb{R})=C_\nu$, then $|m_\nu(z)|\leq C\eta^{-1}$.*

The Stieltjes transform m_ν is (complex) analytic on $\mathbb{C}\backslash\mathbb{R}$. If we restrict $z\in\mathbb{C}^+$, after multiplying by $1/\pi$, the imaginary part of $m_\nu(z)$ can be considered as the Poisson kernel on the upper half plane. From this observation, we find the following formula for the Stieltjes inversion.

Lemma 2.4 (Stieltjes inversion)**.** *Assume that ν is a probability measure without any point mass (atom). Then,*

$$\frac{1}{\pi}\operatorname{Im} m_\nu(\cdot+\mathrm{i}\eta)\to\nu$$

vaguely as $\eta\searrow 0$. In particular, for any interval I,

$$\nu(I)=\frac{1}{\pi}\lim_{\eta\searrow 0}\int_I \operatorname{Im} m_\nu(E+\mathrm{i}\eta)\mathrm{d}E. \tag{2.10}$$

As we can see from Lemma 2.4, we can recover the measure from its Stieltjes transform. Thus, given a sequence of measures, it is natural to consider the Stieltjes transform of the measures in order to find its limit. The following lemma supports this technique.

Lemma 2.5. *Let (ν_i) be a sequence of random probability measures. Then, $m_{\nu_i}(z)\to m_\nu(z)$ almost surely (in probability) for some deterministic probability measure ν if and only if $\nu_i\to\nu$ almost surely (in probability) for every z in the upper half plane.*

The Stieltjes transform of a measure ν contains information on the moments of ν. A formal calculation shows that

$$S_\mu(z) = \int \frac{\mu(\mathrm{d}x)}{x-z} = -\frac{1}{z}\sum_{k=0}^{\infty}\frac{1}{z^k}\langle x^k, \mu\rangle.$$

For the semicircle distribution, if we consider

$$g(z) := \sum_{k=0}^{\infty} m_k z^k,$$

then, for $|z| < 1/2$, $g(z)$ is analytic. Moreover, we find that

$$g(z) = 1 + z^2[g(z)]^2 = \frac{1-\sqrt{1-4z^2}}{2z^2}.$$

We also have for $|z| < 1/2$ that

$$g(z) = \sum_{k=0}^{\infty} z^k \int x^{2k}\mu_{sc}(x)\mathrm{d}x.$$

We thus obtain the Stieltjes transform of the semicircle measure

$$m_{sc} = -\frac{g(z^{-1})}{z} = \frac{-z+\sqrt{z^2-4}}{2}. \tag{2.11}$$

2.3 Local semicircle law

In order to understand the spectrum of a given operator, we use the resolvent, or the Green's function defined as follows:

Definition 2.6 (Resolvent(Green's function))**.** For a real symmetric (complex Hermitian) matrix H, the resolvent G of H is defined by

$$G(z) := (H - zI)^{-1}. \tag{2.12}$$

The idea of using Stieltjes transform to prove the semicircle law based on the normalized trace of the resolvent,

$$m(z) := \frac{1}{N}\operatorname{Tr} G(z) = \frac{1}{N}\sum_{j=1}^{N} G_{jj}(z). \tag{2.13}$$

Recall that the empirical measure μ was defined by

$$\mu = \frac{1}{N}\sum_{j=1}^{N}\delta_{\lambda_j}.$$

If we calculate the Stieltjes transform of μ, then we get

$$m_\mu(z) = \int \frac{\mu(\mathrm{d}x)}{x-z} = \frac{1}{N}\sum_{j=1}^{N}\frac{1}{\lambda_j - z} = \frac{1}{N}\operatorname{Tr}(H-zI)^{-1} = m(z).$$

Thus, the semicircle law can be attained once we prove that $m \to m_{sc}$ as $N \to \infty$.

To estimate the difference $|m - m_{sc}|$, we use the following lemmas:

Lemma 2.7 (Schur complement formula)**.** *Let $H^{(i)}$ be the submatrix of H attained by removing the i-th row and the column of H. (The indices are not changed.) If we let $G^{(i)}$ be the resolvent of $H^{(i)}$, then we have*

$$G_{ii} = \frac{1}{h_{ii} - z - \sum_{s,t}^{(i)} h_{is}G_{st}^{(i)}h_{ti}}.$$

Lemma 2.8 (Large deviation estimate/Hanson–Wright inequality)**.** *Let (a_i) and (b_i) be centered and independent complex random variables with variance σ^2 and having subexponential decay*

$$\mathbb{P}\left(|a_i| \geq x\sigma\right) \leq C_0 e^{-x^{1/\theta}}, \qquad \mathbb{P}\left(|b_i| \geq x\sigma\right) \leq C_0 e^{-x^{1/\theta}},$$

for some positive constant C_0 and $\theta > 1$. Then there exist constants a_0, A_0 and C, depending on θ and C_0, such that for $a_0 \leq \xi \leq A_0 \log\log N$, and $\varphi_N = (\log N)^C$,

$$\mathbb{P}\left(\left|\sum_{i=1}^{N} A_i a_i\right| \geq (\varphi_N)^{\xi}\sigma\left(\sum_{i=1}^{N}|A_i|^2\right)^{1/2}\right) \leq e^{-(\log N)^{\xi}}, \tag{2.14}$$

$$\mathbb{P}\left(\left|\sum_{i=1}^{N} \overline{a_i} B_{ii} a_i - \sum_{i=1}^{N}\sigma^2 B_{ii}\right| \geq (\varphi_N)^{\xi}\sigma^2\left(\sum_{i=1}^{N}|B_{ii}|^2\right)^{1/2}\right) \leq e^{-(\log N)^{\xi}}, \tag{2.15}$$

$$\mathbb{P}\left(\left|\sum_{i\neq j}^{N} \overline{a_i} B_{ij} a_j\right| \geq (\varphi_N)^{\xi}\sigma^2\left(\sum_{i\neq j}|B_{ij}|^2\right)^{1/2}\right) \leq e^{-(\log N)^{\xi}}. \tag{2.16}$$

From Schur complement formula and the large deviation estimate, together with the fact that $h_{ii} \sim N^{-1/2}$, we find that

$$G_{ii} \simeq \frac{1}{-z - N^{-1}\sum_j^{(i)} G_{jj}^{(i)}} \simeq \frac{1}{-z - m(z)}.$$

In particular,

$$m(z) \simeq \frac{1}{-z - m(z)}.$$

Since $m_{sc}(z)$ is the solution to the equation

$$m_{sc}(z) = \frac{1}{-z - m_{sc}(z)},$$

we can estimate the difference $|m(z) - m_{sc}(z)|$ in terms of N.

For any fixed $z = E + \mathrm{i}\eta$, the best result known is that

$$|m(z) - m_{sc}(z)| \lesssim \frac{1}{N\eta}, \tag{2.17}$$

which is enough to prove the semicircle law in the limit $N \to \infty$. (It is believed to be optimal.) If we choose $\eta \ll 1$, we can obtain more information on the distribution of eigenvalues. Since (2.17) describes the local behaviour of the eigenvalues, it is often called the local semicircle law. The local law has many important consequences including the complete delocalization of eigenvectors and the rigidity of eigenvalues, which will be discussed in the following chapters.

2.4 Marchenko–Pastur law

We briefly explain the idea of the proof of Marchenko–Pastur law, Theorem 1.6. Recall that a sample covariance matrix is of the form

$$S_N = \frac{1}{N} X X^T$$

for an $P \times N$ i.i.d. random matrix X. Related to S_N is a matrix $\widetilde{S}_N = \frac{1}{N} X^T X$. Denote by $m(z)$ and $\widetilde{m}(z)$ the normalized traces of the resolvents of S_N and $\widetilde{S}_N$, respectively, i.e.,

$$m(z) = \frac{1}{P}\sum_{i=1}^{P}\frac{1}{\lambda_i - z}, \qquad \widetilde{m}(z) = \frac{1}{P}\sum_{i=1}^{P}\frac{1}{\lambda_i - z} - \frac{N-P}{Pz} = m(z) - \frac{N-P}{Pz}.$$

Instead of directly applying Schur complement formula to the resolvent of S_N, we consider the linearization of S_N, defined as

$$H(z) = \left(\begin{array}{c|c} -zI & X/\sqrt{N} \\ \hline X^T/\sqrt{N} & -I \end{array} \right).$$

Set $R \equiv R(z) = H^{-1}(z)$. By Schur complement formula,

$$\begin{aligned} m(z) &= \frac{1}{P}\sum_{i=1}^{P} R_{ii} = \frac{1}{P}\sum_{i=1}^{P}\left(-z - \frac{1}{N}\sum_{\alpha,\beta=P+1}^{P+N} X_{i\alpha}R^{(i)}_{\alpha\beta}X_{\beta i} \right)^{-1} \\ &\simeq \left(-z - \frac{1}{N}\sum_{\alpha=P+1}^{P+N} R_{\alpha\alpha} \right)^{-1} = \left(-z - \frac{1}{N}\sum_{\alpha=P+1}^{P+N} \Big(-I + \frac{X^T X}{zN} \Big)^{-1}_{\alpha\alpha} \right)^{-1} \\ &= \frac{1}{z}\left(-1 - \frac{P\widetilde{m}(z)}{N} \right)^{-1} = \frac{1}{z}\left(-1 - \frac{Pm(z)}{N} - \frac{P}{Nz} + \frac{1}{z} \right)^{-1}. \end{aligned}$$

It turns out that m_{MP}, the Stieltjes transform of the Marchenko–Pastur law in Theorem 1.6, solves the equation

$$zdm_{MP}(z)^2 + (z - 1 + d)m_{MP}(z) + 1 = 0.$$

This leads us to the local Marchenko–Pastur law, which in turn proves the Marchenko–Pastur law.