数理解析研究所講究録
第1904巻 2014年 105-122

Universality in Random Matrix Theory

Ji Oon Lee

Department of Mathematical Sciences, KAIST

1 Introduction

Let $X_1, X_2, \cdots, X_N$ be independent, identically distributed random variables. Assume that $\mathbb{E}X = 0$ and $\mathbb{E}X^2 = 1$. Then, their sum $S_N := X_1 + X_2 + \cdots + X_N$ satisfies

$$\frac{S_N}{\sqrt{N}} \to N(0,1) \qquad \text{(in distribution)}, \tag{1.1}$$

which is the Central Limit Theorem. In principle, all the random variables $X_1, X_2, \cdots, X_N$ can be of order 1, hence $S_N \sim 1$ as well, but the probability of having such a rare event is incredibly small. We can even estimate the bound on the probability for the rare event from the large deviation principle.

A similar phenomenon happens when we form a large matrix from i.i.d. random variables and consider the distribution of its eigenvalues. The simplest example for the random matrix model is the Gaussian Orthogonal Ensemble (GOE).

Definition 1.1 (Gaussian Orthogonal Ensemble (GOE)). An $N \times N$ random matrix $H = (h_{ij})$ is said to be the Gaussian Orthogonal Ensemble if its entries h_{ij} are real Gaussian random variables, independent up to the symmetry constraint $h_{ij} = h_{ji}$, satisfying

1. $\mathbb{E}h_{ij} = 0$,
2. $\mathbb{E}|h_{ij}|^2 = N^{-1}$ for $i \neq j$, and
3. $\mathbb{E}|h_{ii}|^2 = 2N^{-1}$.

Similarly, we can define a model for complex Hermitian matrices, which is known as Gaussian Unitary Ensemble (GUE).

Definition 1.2 (Gaussian Unitary Ensemble (GUE)). An $N \times N$ random matrix $H = (h_{ij})$ is said to be the Gaussian Unitary Ensemble if its entries h_{ij} are complex Gaussian random variables, independent up to the symmetry constraint $h_{ij} = \overline{h_{ji}}$, satisfying

1. $\mathbb{E}h_{ij} = 0$,
2. $\mathbb{E}|h_{ij}|^2 = N^{-1}$, and
3. $\mathbb{E}(h_{ij})^2 = 0$ if $i \neq j$.

In order to introduce results on GOE or GUE analogous to CLT, we consider the empirical spectral distribution of a random matrix.

Definition 1.3. Let $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$ be the eigenvalues of H. Then, the empirical (spectral) measure of H is defined by

$$\mu(H) := \frac{1}{N}\sum_{j=1}^{N} \delta_{\lambda_j}. \tag{1.2}$$

Random matrices such as GOE or GUE were first introduced by Wigner [17] for the spectral analysis of heavy nuclei. Later, Gaudin, Mehta, and Dyson [10, 11, 3, 4] proved important results on Gaussian ensembles such as the bulk universality results. While the early works focused on the properties of the eigenvalue distribution of the Gaussian ensembles, the recent developments on the random matrix theory showed that in fact most of the properties hold for a larger class of Wigner matrices as well [7, 14]. The most basic yet the most well-known such a phenomena is the semicircle law proved by Wigner [17]. We first consider the GOE case.

Theorem 1.4. *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*

$$\mu_{sc}(x)\mathrm{d}x := \frac{1}{2\pi}\sqrt{(4-x^2)_+}\mathrm{d}x. \tag{1.3}$$

Wigner semicircle law holds for a much larger class of random matrices. The most notable examples are Wigner matrix ensembles.

Definition 1.5. An $N \times N$ random matrix $H = (h_{ij})$ is said to be a (symmetric / Hermitian) Wigner matrix if its entries h_{ij} are (real / complex) random variables satisfying

1. The upper right entries $h_{ij}(i < j)$ are i.i.d. with $\mathbb{E}h_{ij} = 0$, and $\mathbb{E}|h_{ij}|^2 = N^{-1}$.
2. The diagonal entries h_{ii} are i.i.d. with $\mathbb{E}h_{ii} = 0$ and $\mathbb{E}|h_{ii}|^2 = CN^{-1}$ for some constant C.
3. In addition, for the Hermitian case, $\mathbb{E}(h_{ij})^2 = 0$.

The bulk universality concerns the behavior of eigenvalues in the bulk of the spectrum. Consider the following k-point correlation function of the eigenvalues.

Definition 1.6 (k-point correlation function)**.** Let $p_N(x_1, x_2, \cdots, x_N)$ be the symmetric density function of the eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_N$. For any $k = 1, 2, \cdots, N$, we define the k-point correlation function of the eigenvalues by

$$p_N^{(k)}(x_1, x_2, \cdots, x_k) := \int p_N(x_1, x_2, \cdots, x_N)\mathrm{d}x_{k+1}\cdots\mathrm{d}x_N. \tag{1.4}$$

For *GUE*, it is proved by Gaudin and Mehta [10, 11] that $p_N^{(k)}$ has a determinantal structure, i.e., there exists a function K_N such that

$$p_N^{(k)}(x_1, x_2, \cdots, x_k) = \det[K_N(x_i, x_j)]_{1\le i,j\le k}. \tag{1.5}$$

Later, Dyson found the formula for K_N which becomes

$$K_N(x + \frac{y_1}{N\mu_{sc}(x)}, x + \frac{y_2}{N\mu_{sc}(x)}) = \mu_{sc}(x)K(y_1, y_2), \tag{1.6}$$

where K is the function

$$K(y_1, y_2) := \frac{\sin(\pi(y_1 - y_2))}{\pi(y_1 - y_2)}, \tag{1.7}$$

which is called the Dyson sine kernel.

As in the Wigner semicircle law, the limiting distribution for the correlation function is universal, under suitable conditions, in the sense that it does not depend on the specific matrix model. There is, however, one crucial difference between the universality of the correlation function and that of the semi-circle law, which is that the correlation function may depend on whether the given random matrix model is an orthogonal ensemble or a unitary ensemble. The Dyson sine kernel can describe the limiting distribution for unitary ensembles but not for orthogonal ensembles. Including this feature, we use the term **bulk universality** that asserts the fact that the limiting distribution for the correlation function depends only on the symmetry class.

The bulk universality also implies that the limiting distribution of the 'gap', the distance between adjacent eigenvalues, is also universal. In particular, since the adjacent eigenvalues in GUE or GOE tend to keep distance from each other, we can also find that the adjacent eigenvalues in many random matrix models. This is called 'level repulsion'. Note that this also tells the eigenvalues of random matrix models are strongly correlated, unlike i.i.d. random variables.

The dyson sine kernel, and in general the determinantal structure of the correlation functions, appear in many other fields of mathematics. For example, if we consider the non-trivial zeros of the Riemann zeta function on the critical line, then after a proper normalization, the imaginary parts of the zeros behave like the eigenvalues of GUE.

Near the edge of the spectrum, i.e., for the eigenvalues very near 2 (or -2), the limiting distribution for the correlation function has different asymptotics. In this case, K_N converges to a different kernel, which is usually called the Airy kernel. However, there is another formula, proved by Tracy and Widom [15], for the limiting distribution for the largest eigenvalue.

Definition 1.7 (Tracy-Widom distribution). Let $q(t)$ be a solution to the Painlevé II equation

$$q'' = tq + 2q^3$$

satisfying

$$q(t) \sim Ai(t) \text{ as } t \to \infty.$$

The function F_2 is a distribution function defined by

$$F_2(t) = \exp\left(-\int_t^\infty (x-t)q(x)^2 \mathrm{d}x\right). \tag{1.8}$$

The function F_2 is called the Tracy-Widom distribution.

It is proved by Tracy and Widom that the cumulative distribution function of the rescaled largest eigenvalue of GUE

$$\mathbb{P}[N^{2/3}(\lambda_N - 2) \le t] \to F_2(t)$$

in distribution (vague convergence). In particular, the size of the fluctuation of the largest eigenvalue of GUE around 2 is of order $N^{-2/3}$.

The subindex 2 in F_2 again denotes the dependence on the symmetry class [16]. For orthogonal ensembles, the limiting distribution function is different from F_2, and it is usually called F_1. As in the bulk universality, under suitable conditions, the limiting distribution for the largest eigenvalue only depends on the symmetry class. This fact is what we call the **edge universality**.

The Tracy-Widom distribution appears in many other applications. For example, if we consider L_N, the length of the longest increasing subsequence of a random permutation on $\{1, 2, \cdots, N\}$, then $L_N/\sqrt{N}$ behaves as the largest eigenvalue of GUE, hence the limiting distribution is given by Tracy-Widom.

The universality results, bulk and edge, provide a very important strategy in the random matrix theory. When a certain property of the eigenvalues of a given random matrix model is hard to achieve, we can first consider the property we want to attain in GOE or GUE. Then, by proving a universality result, by comparing the limiting distribution of some eigenvalues of the given model with that of GOE or GUE, we can find a desired property in the given random matrix. This is very powerful because there are additional tools working only on GUE or GOE due to some symmetry or some explicit formulas.

Although omitted in this article, there are many other random matrix models that exhibit similar behaviours as Wigner matrices. Among important random matrix models are the Wishart random matrix (sample covariance matrix), which is of the form XX^* for an $M \times N$ random matrix X with i.i.d. entries, and the Circular Ensembles (COE / CUE). In some cases, a random matrix ensemble is defined by a probability distribution on the set of matrices, such as

$$P(M)\mathrm{d}M = \frac{1}{Z_N} e^{-\operatorname{tr} Q(M)} \mathrm{d}M, \tag{1.9}$$

where dM is the Lebesgue measure on the independent entries of M, Q is a real-valued function satisfying $Q(x) \to \infty$ as $x \to \pm\infty$, and Z_N is the normalization constant (partition function) given by

$$Z_N = \int e^{-\operatorname{tr} Q(M)} dM. \tag{1.10}$$

The set $\{M\}$ can be a set of all real symmetric matrices or a set of all Hermitian matrices. The former is known as 'orthogonal ensembles' and the latter 'unitary ensembles'. As the names suggest, GOE and GUE belong to orthogonal ensembles and unitary ensembles, respectively, where the 'potential' Q becomes

$$Q(x) = \frac{\beta N}{4} x^2 \tag{1.11}$$

with $\beta = 1$ for orthogonal ensembles and $\beta = 2$ for unitary ensembles. In the invariant ensemble, although a typical choice for β is $\beta = 1$ or $\beta = 2$, it is also possible to use different β other than 1 or 2. (The choices $\beta = 1$, 2, and 4 are called 'classical'.) In this case, the bulk universality means that the limiting distribution depends only on β under suitable conditions on Q.

Throughout the article, the limit, O, o, $\gg$, $\ll$, and other similar notations will always refer to the limit $N \to \infty$ unless noticed otherwise. In some cases the notation $A \lesssim B$ will be used, where it means that

$$A \leq (\log N)^{C \log\log N} B$$

for some constant C independent of N. The notation $A \sim B$ means that

$$C^{-1}|B| \leq |A| \leq C|B|$$

for some constant C independent of N.

2 Wigner semicircle law

In this chapter, we prove Wigner semicircle law for Wigner matrices in two different ways. The first one by the moment counting is almost the same as the original proof by Wigner, and it requires some combinatorial arguments. The second one uses the Stieltjes transform, which provides a nice tool in the random matrix theory in general.

2.1 Proof by moment counting

In the original work of Wigner, a combinatorial proof by moment counting was used. In order to prove the semicircle law, it suffices 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

$$m_{2k} = C_k, \quad m_{2k-1} = 0, \tag{2.1}$$

where C_k is the k-th Catalan number defined by

$$C_k = \frac{1}{k+1}\binom{2k}{k} = \frac{(2k)!}{(k+1)!k!}. \tag{2.2}$$

Thus, it only remains to prove that

$$\langle \mu, x^k \rangle = \frac{1}{N} \operatorname{tr} H^k \to \begin{cases} C_{k/2} & \text{if } k \text{ is even,} \\ 0 & \text{if } k \text{ is odd.} \end{cases} \tag{2.3}$$

We first prove the following lemma.

Lemma 2.1. *After taking expectation, we have*

$$\mathbb{E}\langle \mu, x^k \rangle \to \begin{cases} C_{k/2} & \textit{if } k \textit{ is even,} \\ 0 & \textit{if } k \textit{ is odd.} \end{cases} \tag{2.4}$$

Proof. Consider

$$\mathbb{E}\langle \mu, x^k \rangle = \frac{1}{N} \sum_{i_1, i_2, \cdots, i_k} \mathbb{E}\left[H_{i_1 i_2} H_{i_2 i_3} \cdots H_{i_{k-1} i_k} H_{i_k i_1} \right]. \tag{2.5}$$

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 $1 \leq j \leq k$ 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 $(i_1 i_2, i_2 i_3, \cdots, i_k i_1)$. 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 $\mathbb{E}\left[H_{i_1 i_2} H_{i_2 i_3} \cdots H_{i_{k-1} i_k} H_{i_k i_1} \right]$. 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

$$\frac{1}{N} \sum_{i_1, i_2, \cdots, i_k} \mathbb{E}\left[H_{i_1 i_2} H_{i_2 i_3} \cdots H_{i_{k-1} i_k} H_{i_k i_1} \right] \to C_{k/2} \tag{2.6}$$

when k is even. □

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

$$\mathbb{E}\left[\langle \mu, x^k \rangle^2 \right] - \left[\mathbb{E}\langle \mu, x^k \rangle \right]^2 \to 0 \tag{2.7}$$

and use Chebyshev inequality. It requires more graph counting, which we omit here.

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

$$m_\nu(z) := \int_{\mathbb{R}} \frac{\nu(dx)}{x - z} \tag{2.8}$$

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

If we let $z = E + i\eta$, 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 $\mathbb{C}\backslash\mathbb{R}$. It is also possible to 'invert' the Stieltjes transform.

Lemma 2.4. *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,

$$\nu(I) = \frac{1}{\pi}\lim_{\eta\searrow 0}\int_I \operatorname{Im} m_\nu(E + \mathrm{i}\eta)\mathrm{d}E. \tag{2.9}$$

As we can see from the above lemma, we can recover information on 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 the 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

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

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

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

We thus obtain the Stieltjes transform of the semicircle measure

$$m_{sc} = -\frac{1}{z}g(\frac{1}{z}) = \frac{-z + \sqrt{z^2 - 4}}{2}. \tag{2.10}$$

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

$$G(z) := (H - zI)^{-1}. \tag{2.11}$$

The idea of using Stieltjes transform to prove the semicircle law based on the averaged Green's function

$$m(z) := \frac{1}{N} \operatorname{tr} G(z) = \frac{1}{N} \sum_{j=1}^{N} G_{jj}(z). \tag{2.12}$$

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 formula:

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*

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

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

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

(See [6] for more detail.) 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 [8] is that

$$|m(z) - m_{sc}(z)| \lesssim \frac{1}{N\eta}, \tag{2.13}$$

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. Because Equation (2.13) 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.

3 Green Function Comparison

3.1 Lindeberg replacement strategy

Suppose that we want to prove the central limit theorem. For simplicity, assume that (Y_i) are i.i.d. real random variables with $\mathbb{E}Y_i = 0$, $\mathbb{E}|Y_i|^2 = 1$, and $\mathbb{E}|Y_i|^p < C_p$ for any finite p. Let

$$W_N = \frac{Y_1 + Y_2 + \cdots + Y_N}{\sqrt{N}}.$$

Suppose that (X_i) are i.i.d. real Gaussian random variables with $\mathbb{E}X_i = 0$ and $\mathbb{E}|X_i|^2 = 1$. We also let

$$Z_N = \frac{X_1 + X_2 + \cdots + X_N}{\sqrt{N}}.$$

Notice that Z_N is a centered Gaussian with variance 1, which can be easily checked by a Fourier transform method. Define

$$W_N^{(i)} = \frac{X_1 + \cdots + X_i + Y_{i+1} + \cdots + Y_N}{\sqrt{N}},$$

in particular, $W_N^{(0)} = W_N$ and $W_N^{(N)} = Z_N$. We now consider the m-th moment of W_N. Since (Y_i) are i.i.d. with $\mathbb{E}Y_i = 0$, we find that

$$\mathbb{E}(W_N)^m = \sum_{i_1, i_2, \cdots, i_k} \mathbb{E}\left(\frac{Y_{i_1}}{\sqrt{N}}\right)^{r_1}\left(\frac{Y_{i_2}}{\sqrt{N}}\right)^{r_2}\cdots\left(\frac{Y_{i_k}}{\sqrt{N}}\right)^{r_k} \tag{3.1}$$

with $i_1, i_2, \cdots, i_k > 0$ and $i_1 + i_2 + \cdots + i_k = m$.

We now consider $\mathbb{E}[(W_N^{(j)})^m - (W_N^{(j-1)})^m]$. from the argument above, we can easily find that

$$\mathbb{E}[(W_N^{(j)})^m - (W_N^{(j+1)})^m] = \sum_{s=2}^{m} C_s \mathbb{E}\left[\left(\frac{X_j}{\sqrt{N}}\right)^s - \left(\frac{Y_j}{\sqrt{N}}\right)^s\right] \tag{3.2}$$

for some finite constants $C_2, C_3, \cdots$. (They may depend on m, but we omit the detail.) Since Y_j and X_j have matching second moment, we find that $\mathbb{E}[(W_N^{(j)})^m - (W_N^{(j+1)})^m] = O(N^{-3/2})$. Applying the argument N-times, we find from the telescoping series

$$\mathbb{E}[(W_N)^m - (Z_N)^m] = \sum_{j=0}^{N-1} \mathbb{E}[(W_N^{(j)})^m - (W_N^{(j+1)})^m]$$

that $\mathbb{E}[(W_N)^m - (Z_N)^m] = O(N^{-1/2})$. This proves the central limit theorem.

The advantage of using Lindeberg replacement strategy in proving CLT is that it does not require the exact calculation on the moments of W_N. Instead, the moments are simply compared with well-known objects - the moments of the Gaussian - and the proof can be done by controlling the error terms in the comparison. Lindeberg replacement strategy has many important applications in random matrix theory, most notably, the Green function comparison for the universality.

3.2 Eigenvalue counting and Green function

For $E_1 < E_2$, let

$$\mathcal{N}(E_1, E_2) = |\{\lambda_j : E_1 \le \lambda_j < E_2\}|,$$

i.e., the number of eigenvalues in the interval $[E_1, E_2)$. If we consider the characteristic function $\chi = \chi_{[E_1, E_2)}$, we have

$$\mathcal{N}(E_1, E_2) = \operatorname{tr} \chi(H).$$

Since χ is not a continuous function, we want to approximate it by a smooth function. We use the Poisson kernel in the upper half plane, i.e.,

$$P_\eta(x) = \frac{\eta}{\pi(x^2+\eta^2)}.$$

Notice that $P_\eta * \chi$ is a smooth function for any $\eta > 0$.

Using $P_\eta * \chi$ instead of χ, we find

$$\mathcal{N}(E_1, E_2) = \operatorname{tr}\chi(H) \simeq \operatorname{tr}(P_\eta * \chi)(H) = \operatorname{tr}\int_{E_1}^{E_2} P_\eta(H-E)\mathrm{d}E. \tag{3.3}$$

Applying the identity

$$P_\eta(x) = \frac{1}{\pi}\operatorname{Im}\frac{1}{x-\mathrm{i}\eta},$$

we get

$$\mathcal{N}(E_1, E_2) \simeq \frac{1}{\pi}\int_{E_1}^{E_2} \operatorname{tr}\operatorname{Im}\frac{1}{H-E-\mathrm{i}\eta}\mathrm{d}E = \frac{N}{\pi}\int_{E_1}^{E_2} \operatorname{Im} m(E+\mathrm{i}\eta)\mathrm{d}E. \tag{3.4}$$

Roughly speaking, this approximation shows that the estimate on the difference of the Green functions of two different Wigner matrices may imply the universality of eigenvalue distribution.

3.3 Four moment theorem and Green function comparison theorem

Definition 3.1 (Moment matching). For two independent real random variables X and Y, if

$$\mathbb{E}X^k = \mathbb{E}Y^k$$

for $k = 1, 2, \cdots, p$, we say that the first p moments of X and Y match. For complex random variables, we use the condition

$$\mathbb{E}\overline{X}^j X^k = \mathbb{E}\overline{Y}^j Y^k$$

for $j, k = 0, 1, 2, \cdots, p$ with $j + k \leq p$, then we say that the first p moments of X and Y match.

If we have some moment matching condition on the entries of two Wigner matrices H and H', then the eigenvalue distributions of them can be easily bounded. The following theorem by Tao and Vu [14] shows that matching four moments is always sufficient.

Theorem 3.2 (Four moment theorem). *Let H and H' be Wigner matrices. Assume that the first four moments of h_{ij} and h'_{ij} match, the first two moments of h_{ii} and h_{ij} match, and the subexponential decay condition holds for the single entry distributions. Then, there exists a small positive constant c_0 such that, for any sufficiently small positive ϵ, for any function $F : \mathbb{R}^k \to \mathbb{R}$ satisfying $|\nabla^j F| \leq N^\epsilon$ for $j \leq 5$, and for any selection of k-tuple of indices $i_1, i_2, \cdots, i_k \in [\epsilon N, (1-\epsilon)N]$, we have*

$$\left|\mathbb{E}F(\lambda_{i_1}, \lambda_{i_2}, \cdots, \lambda_{i_k}) - \mathbb{E}F(\lambda'_{i_1}, \lambda'_{i_2}, \cdots, \lambda'_{i_k})\right| \leq N^{-c_0} \tag{3.5}$$

for any sufficiently large N.

In particular, if a real symmetric Wigner matrix H satisfies that

$$\mathbb{E}x_{ij} = 0, \quad \mathbb{E}x_{ij}^2 = 1, \quad \mathbb{E}x_{ij}^3 = 0, \quad \mathbb{E}x_{ij}^4 = 3, \quad Ex_{ii} = 0, \quad \mathbb{E}x_{ii}^2 = 2,$$

then we always have universality. In order to find the statistics of eigenvalues around a fixed point, instead of considering the eigenvalues with fixed indices, the following theorem can be used [5].

Theorem 3.3 (Green function comparison theorem). *Suppose that H^v and H^w are Wigner matrices satisfying the subexponential decay condition and the first four moments of H^v and H^w match. Then, there exist a positive constant $\epsilon > 0$ such that, for any $\eta \in [N^{-1-\epsilon}, N]$ and for any z_1, z_2 with $\operatorname{Im} z_1, \operatorname{Im} z_2 = \pm\eta$, we have*

$$\lim_{N\to\infty} \left[\mathbb{E}\operatorname{tr} G^v(z_1)\operatorname{tr} G^v(z_2) - \mathbb{E}\operatorname{tr} G^w(z_1)\operatorname{tr} G^2(z_2)\right] = 0, \tag{3.6}$$

where G^v and G^w denote the Green functions of H^v and H^w, respectively.

Both of the theorems can be applied to prove the universality, provided that a nice matrix H' exists.

Idea of proof of Green function comparison theorem. Suppose that we want to estimate $|\mathbb{E}(m - m')|$, where m and m' denote the averaged Green functions of H and H', respectively. Let $\gamma = N(N+1)/2$. Adopting the Lindeberg replacement strategy, we consider a sequence of Wigner matrices $\{H_0, H_1, \cdots, H_\gamma\}$, where $H_0 = H$, $H_\gamma = H'$, and H_k and H_{k+1} differ at most in two matrix entries. We now consider a telescoping series

$$\mathbb{E}m - \mathbb{E}m' = \frac{1}{N}\sum_{k=1}^{\gamma}\left[\mathbb{E}\operatorname{tr}(H_{k-1} - z)^{-1} - \mathbb{E}\operatorname{tr}(H_k - z)^{-1}\right]. \tag{3.7}$$

Suppose that H_k and H_{k-1} differ only at the (i,j)-position. (It can be a diagonal element, i.e., it is allowed to have $i = j$.) Let Q be a matrix having zero entries at the (i,j)-position and all the other entries are equal to those of H_k and H_{k+1}. We let

$$H_k = Q + V, \quad H_{k-1} = Q + W.$$

We now expand the resolvents of H_k and H_{k-1}. Let

$$R = (Q - z)^{-1}, \quad S = (H_k - z)^{-1}, \quad T = (H_{k-1} - z)^{-1}.$$

Then, we have an identity

$$S = R - RVS. \tag{3.8}$$

Applying (3.8) successively, we also find that

$$S = R - RVR + RVRVR - (RV)^3R + (RV)^4R - (RV)^5S \tag{3.9}$$

We then calculate the expectation of S_{aa}. From (3.9), we find that

$$\mathbb{E}S_{aa} = \mathbb{E}R_{aa} - \mathbb{E}(RVR)_{aa} + \cdots - [(RV)^5S]_{aa}.$$

Consider the second term in the above expansion. Since the entries of V are all zero except V_{ij} and V_{ji}, we have

$$(RVR)_{aa} = R_{ai}V_{ij}R_{ja} + R_{aj}V_{ji}R_{ia}.$$

However, Q is independent of V, thus R is also independent of V. Thus, by taking partial expectation with respect to V_{ij}, we find that

$$\mathbb{E}(RVR)_{aa} = \mathbb{E}V_{ij}\cdot\mathbb{E}[R_{ai}R_{ja}] + \mathbb{E}V_{ji}\cdot\mathbb{E}[R_{aj}R_{ia}] = 0.$$

The third term in the expansion (3.9) does not vanish. Nevertheless, we can decouple V and R by using the same idea as above, so we find that

$$\mathbb{E}(RVRVR)_{aa} = \mathbb{E}[V_{ij}V_{ji}]\cdot\mathbb{E}[R_{ai}R_{jj}R_{ia}] + \cdots.$$

We can also consider the expansion of T, and in the expansion, we find a term

$$\mathbb{E}(RWRWR)_{aa} = \mathbb{E}[W_{ij}W_{ji}]\cdot\mathbb{E}[R_{ai}R_{jj}R_{ia}] + \cdots,$$

which will exactly cancel the term above in the difference $\mathbb{E}S_{aa} - \mathbb{E}T_{aa}$. In this way, from the four moment matching condition, we find that

$$\mathbb{E}S_{aa} - \mathbb{E}T_{aa} = -\mathbb{E}[(RV)^5 S]_{aa} + \mathbb{E}[(RW)^5 T]_{aa}.$$

From the local law, we can attain a priori estimates $|S_{ii}| < C$, $|S_{ij}| = o(1)$, and similar results for R and T. From those estimates, we find that

$$\mathbb{E}S_{aa} - \mathbb{E}T_{aa} = O(N^{-5/2}).$$

Since we have $O(N^2)$ terms in the telescoping series (3.7), we conclude that

$$\mathbb{E}m - \mathbb{E}m' = O(N^{-1/2}).$$

□

Remark 3.4. The four moment theorem as well as the Green function comparison theorem can be extend to the edge of the spectrum.

Remark 3.5. The fourth moment matching condition (and the two moment matching condition for the diagonal entries) can be weakened to $\mathbb{E}(h_{ij})^4 - \mathbb{E}(h'_{ij})^4 = O(N^{-\delta})$ for some $\delta > 0$.

In the proof of the bulk universality and the edge universality, the Green function comparison theorem is applied to compare the eigenvalue distribution of a given Wigner matrix and a reference matrix such as Gaussian ensembles. In some situations, we even consider a chain of Wigner matrices and compare the Green functions of them one pair at a time.

4 Dyson sine kernel and the bulk universality

4.1 Joint density of Gaussian unitary ensemble

Recall that we defined GUE as a Wigner matrix with Gaussian random variables. With a slight abuse of notation, we may write it using the Gaussian measures such as

$$\mathrm{d}h_{ii} = C_1^{(N)} e^{-Nh_{ii}^2/2} \mathrm{d}M_{ii}, \quad \mathrm{d}h_{ij}^R = C_2^{(N)} e^{-N(h_{ij}^R)^2} \mathrm{d}M_{ij}^R, \quad \mathrm{d}h_{ij}^I = C_2^{(N)} e^{-N(h_{ij}^I)^2} \mathrm{d}M_{ij}^I,$$

where $\mathrm{d}M_{ii}$, $\mathrm{d}M_{ij}^R$, and $\mathrm{d}M_{ij}^I$ are Lebesgue measures on $\mathbb{R}$. (Here, R and I stand for 'real' and 'imaginary', respectively.) If we consider the probability measure corresponding to GUE defined on the N^2-dimensional space of Hermitian matrices, then we obtain that

$$\begin{aligned} \mathrm{d}H &= \prod_{i=1}^N \mathrm{d}h_{ii} \prod_{i<j} \mathrm{d}h_{ij}^R \mathrm{d}h_{ij}^I = C_N \exp\left(-\frac{N}{2} \sum_{i,j=1}^N |h_{ij}|^2\right) \prod_{i=1}^N \mathrm{d}M_{ii} \prod_{i<j} \mathrm{d}M_{ij}^R \mathrm{d}M_{ij}^I \\ &= C_N \exp\left(-\frac{N}{2} \operatorname{tr} H^2\right) \prod_{i=1}^N \mathrm{d}M_{ii} \prod_{i<j} \mathrm{d}M_{ij}^R \mathrm{d}M_{ij}^I \end{aligned} \tag{4.1}$$

For the GUE matrix, we can diagonalize it and have $H = U\Lambda U^\dagger$ for a unitary matrix U and a real diagonal matrix $\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_N)$. Notice that there is arbitrariness in the choice of U, which corresponds to the phase factor, i.e., for any $T = diag(e^{\mathrm{i}\theta_1}, e^{\mathrm{i}\theta_2}, \cdots, e^{\mathrm{i}\theta_N})$, we have $T\Lambda T^\dagger = \Lambda$. Thus, we consider the quotient group $\tilde{\mathrm{U}} = \mathrm{U}/T$, where U is the unitary group of order N, and let $\mathrm{d}\tilde{\mathrm{U}}$ be the Haar measure on $\tilde{\mathrm{U}}$. We now have

$$\mathrm{d}H = C_N \exp\left(-\frac{N}{2} \operatorname{tr} H^2\right) \mathcal{J}(H) \mathrm{d}\Lambda \mathrm{d}\,\mathrm{U} = C_N \exp\left(-\frac{N}{2} \sum_{i=1}^N \lambda_i^2\right) |\mathcal{J}(H)| \mathrm{d}\Lambda \mathrm{d}\,\mathrm{U} \tag{4.2}$$

where $\mathcal{J}(H)$ is the Jacobian and $\mathrm{d}\Lambda = \mathrm{d}\lambda_1 \mathrm{d}\lambda_2 \cdots \mathrm{d}\lambda_N$ is the Lebesgue measure on $\mathbb{R}^N$. A calculation shows that

$$\mathcal{J}(H) = \prod_{i<j} (\lambda_i - \lambda_j)^2, \tag{4.3}$$

hence we find the formula

$$\mathrm{d}H = \frac{1}{Z_N} \exp\left(-\frac{N}{2} \sum_{i=1}^{N} \lambda_i^2\right) \prod_{i<j} (\lambda_i - \lambda_j)^2 \mathrm{d}\Lambda \mathrm{d}\,\mathrm{U}\,. \tag{4.4}$$

In particular, for any function $f(H) = f(\lambda_1, \lambda_2, \cdots, \lambda_N)$ invariant under $H \to UHU^\dagger$, we have

$$\int f(H)\mathrm{d}H = \frac{1}{Z_N} \int_{\mathbb{R}^N} f(\lambda_1, \lambda_2, \cdots, \lambda_N) \exp\left(-\frac{N}{2} \sum_{i=1}^{N} \lambda_i^2\right) \prod_{i<j} (\lambda_i - \lambda_j)^2 \mathrm{d}\Lambda. \tag{4.5}$$

The expression $\prod_{i<j}(x_i - x_j)$ is known as the van der Monde determinant; using the determinant, we can rewrite it as

$$\prod_{i<j} (x_i - x_j) = (-1)^{n(n-1)/2} \det \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_N \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{N-1} & x_2^{N-1} & \cdots & x_N^{N-1} \end{pmatrix}. \tag{4.6}$$

4.2 Orthogonal polynomial

Consider a family of polynomials $\{f_k(x)\}$, where f_k is a degree k monic polynomial with the following property:

$$\int_{\mathbb{R}} f_j(x) f_k(x) e^{-Q(x)} \mathrm{d}x = c_i c_j \delta_{ij}$$

for some function Q and for some constants $c_1, c_2, \cdots$. There are many 'classical' orthogonal polynomials, e.g., Jacobi polynomials, Laguerre polynomials, Hermite polynomials, etc., depending on the choice of Q. When $Q(x) = x^2/2$, we can use Hermite polynomials defined as follows:

Definition 4.1 (Hermite polynomials). The n-th Hermite polynomial H_n is define by

$$H_n(x) = (-1)^n e^{x^2/2} \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-x^2/2}. \tag{4.7}$$

We also define the n-th normalized oscillator wavefunction ψ_n by

$$\psi_n(x) = \frac{e^{-x^2/4}}{\sqrt{n!\sqrt{2\pi}}} H_n(x). \tag{4.8}$$

Remark 4.2. Often in literature, the choice of $Q(x) = x^2$ is used when defining Hermite polynomials, which gives "physicists' Hermite polynomials".

Notice we chose the constant in the definition of ψ_n to ensure that

$$\int_{\mathbb{R}} \psi_m(x) \psi_n(x) \mathrm{d}x = \delta_{mn}. \tag{4.9}$$

Recall the van der Monde determinant from the last section. Since H_n is monic, we also have

$$\det \begin{pmatrix} 1 & 1 & \cdots & 1 \\ x_1 & x_2 & \cdots & x_N \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{N-1} & x_2^{N-1} & \cdots & x_N^{N-1} \end{pmatrix} = \det \begin{pmatrix} H_0(x_1) & H_0(x_2) & \cdots & H_0(x_N) \\ H_1(x_1) & H_1(x_2) & \cdots & H_1(x_N) \\ \vdots & \vdots & \ddots & \vdots \\ H_{N-1}(x_1) & H_{N-1}(x_2) & \cdots & H_{N-1}(x_N) \end{pmatrix}.$$

Thus, if we let

$$K_N(x,y) := \sum_{k=1}^{N} \psi_{k-1}(x)\psi_{k-1}(y), \tag{4.10}$$

then we obtain that

$$\exp\left(-\sum_{i=1}^{N}\frac{x_i^2}{2}\right)\prod_{i<j}(x_i-x_j)^2 = C\left[\det \psi_{i-1}(x_j)\right]^2 = C\det K_N(x_i,x_j).$$

We remark that

$$\int K_N(x,x)\mathrm{d}x = N. \tag{4.11}$$

We also have the following properties:

Lemma 4.3. *For ψ_j and K_N defined above, we have*

1. *(Reproducing kernel property)*

$$\int K_N(x,y)K_N(y,z)\mathrm{d}y = K_N(x,z).$$

2. *(Christoffel-Darboux formula)*

$$K_N(x,y) = \sqrt{N}\left[\frac{\psi_N(x)\psi_{N-1}(y)-\psi_N(y)\psi_{N-1}(x)}{x-y}\right].$$

Moreover, if $x=y$, we have

$$K_N(x,x) = \sqrt{N}\left(\psi_N'(x)\psi_{N-1}(x)-\psi_{N-1}'(x)\psi_N(x)\right).$$

3. *(Asymptotic behaviour)*

$$\lim_{N\to\infty} N^{1/4}\psi_n(N^{-1/2}t) = \frac{1}{\sqrt{\pi}}\cos\left(t-\frac{n\pi}{2}\right)$$

locally uniformly, where $n = N-\ell$ for a fixed ℓ.

4.3 Dyson sine kernel

We first consider the determinantal structure of the correlation functions. If we let

$$\begin{aligned} p_N^{(k)}(x_1,x_2,\cdots,x_k) &= \frac{1}{Z}\int \exp\left(-\frac{1}{2}\sum_{i=1}^{N}x_i^2\right)\prod_{i<j}(x_i-x_j)^2\mathrm{d}x_{k+1}\cdots\mathrm{d}x_N \\ &= C\int \det K_N(x_i,x_j)\mathrm{d}x_{k+1}\cdots\mathrm{d}x_N, \end{aligned}$$

then the reproducing kernel property shows that the determinantal structure holds for $p_N^{(k)}$. For example, when $N=3$, we have

$$\begin{aligned} &\int \det K_N(x_i,x_j)\mathrm{d}x_3 \\ &= \int \mathrm{d}x_3\left[K_N(x_3,x_1)\begin{vmatrix}K_N(x_1,x_2) & K_N(x_1,x_3)\\ K_N(x_2,x_2) & K_N(x_2,x_3)\end{vmatrix} - K_N(x_3,x_2)\begin{vmatrix}K_N(x_1,x_1) & K_N(x_1,x_3)\\ K_N(x_2,x_1) & K_N(x_2,x_3)\end{vmatrix}\right] \\ &\quad + \int \mathrm{d}x_3 K_N(x_3,x_3)\begin{vmatrix}K_N(x_1,x_1) & K_N(x_1,x_2)\\ K_N(x_2,x_1) & K_N(x_2,x_2)\end{vmatrix} \\ &= \begin{vmatrix}K_N(x_1,x_2) & K_N(x_1,x_1)\\ K_N(x_2,x_2) & K_N(x_2,x_1)\end{vmatrix} - \begin{vmatrix}K_N(x_1,x_1) & K_N(x_1,x_2)\\ K_N(x_2,x_1) & K_N(x_2,x_2)\end{vmatrix} + N\begin{vmatrix}K_N(x_1,x_1) & K_N(x_1,x_2)\\ K_N(x_2,x_1) & K_N(x_2,x_2)\end{vmatrix} \\ &= \begin{vmatrix}K_N(x_1,x_1) & K_N(x_1,x_2)\\ K_N(x_2,x_1) & K_N(x_2,x_2)\end{vmatrix}. \end{aligned}$$

Collecting the tools we developed in the last section, we can prove the formula established by Dyson. We first rescale the formulas we have and find that

$$\exp\left(-\frac{N}{2}\sum_{i=1}^{N}x_i^2\right)\prod_{i<j}(x_i-x_j)^2 = N^{-\frac{N(N-1)}{2}}\exp\left(-\frac{N}{2}\sum_{i=1}^{N}x_i^2\right)\prod_{i<j}(\sqrt{N}x_i-\sqrt{N}x_j)^2$$
$$= CN^{-\frac{N(N-1)}{2}}\det K_N(\sqrt{N}x_i,\sqrt{N}x_j).$$

We then use the Christoffel-Darboux formula to get

$$K_N(\sqrt{N}x_i,\sqrt{N}x_j) = \frac{\psi_N(\sqrt{N}x_i)\psi_{N-1}(\sqrt{N}x_j)-\psi_N(\sqrt{N}x_j)\psi_{N-1}(\sqrt{N}x_i)}{x_i-x_j}.$$

When x_i and x_j are near zero, we let

$$x_i = \frac{\pi y_i}{N}, \quad x_j = \frac{\pi y_j}{N}.$$

Using the property on the asymptotic behaviour of ψ_N, we obtain

$$K_N(\sqrt{N}x_i,\sqrt{N}x_j)$$
$$\to \frac{\sqrt{N}}{\pi}\frac{\cos(\pi y_i - N\pi/2)\cos(\pi y_j-(N-1)\pi/2)-\cos(\pi y_j-N\pi/2)\cos(\pi y_i-(N-1)\pi/2)}{\pi\sqrt{N}(x_i-x_j)}$$
$$= \sqrt{N}\frac{\sin(\pi(y_i-y_j))}{\pi(y_i-y_j)}.$$

In particular, if we let

$$K(y_i,y_j) = \frac{\sin(\pi(y_i-y_j))}{\pi(y_i-y_j)}, \tag{4.12}$$

then we find that

$$p_N^{(k)}(y_1,y_2,\cdots,y_k) \to C\det\left[K(x_i,x_j)\right]_{i,j=1}^{k},$$

for y_i near zero. This proves the Dyson sine kernel formula near zero.

For general eigenvalues, we use an asymptotic formula

$$\lim_{N\to\infty} N^{1/4}\psi_n\left(\sqrt{N}x+\frac{t}{\sqrt{N}\mu_{sc}(x)}\right) = \mu_{sc}(x)\cos\left(t-\frac{n\pi}{2}\right)$$

and proceed as the proof for the eigenvalues near zero.

In order to fix the constant, we may use the following lemma.

Lemma 4.4 (Integrating out lemma). *For any $m \le N$, we have*

$$\int \det\left[K_N(x_i,x_j)\right]_{i,j=1}^{m} dx_m = (N-m+1)\det\left[K_N(x_i,x_j)\right]_{i,j=1}^{m-1}. \tag{4.13}$$

Instead of tracking the constants, we can easily find that

$$\frac{1}{N!}\det K_N(x_i,x_j)$$

defines a probability measure. After applying the integrating out lemma $(N-k)$-times, we find that

$$p_N(x_1,x_2,\cdots,x_N) \to \frac{(N-k)!}{N!}\det\left[K(x_i,x_j)\right]_{i,j=1}^{k}, \tag{4.14}$$

since the right hand side defines a probability measure.

5 Edge universality and Tracy-Widom distribution

5.1 Tracy-Widom distribution

Recall that the Tracy-Widom distribution is defined by

$$F_2(t) = \exp\left(-\int_t^\infty (x-t)q(x)^2 \mathrm{d}x\right) \tag{5.1}$$

where $q(t)$ is a solution to the Painlevé II equation

$$q'' = tq + 2q^3$$

satisfying

$$q(t) \sim Ai(t) \text{ as } t \to \infty.$$

The first result on the distribution of the largest eigenvalue was based on the Fredholm determinant as in the gap probability [9]. Considering an interval (t, ∞), we can see that

$$\lim_{N\to\infty} \mathbb{P}\left(N^{-2/3}(\lambda_N - 2) \le t\right) = 1 + \sum_{s=1}^{\infty} \frac{(-z)^s}{s!} \int_{(t,\infty)^s} \det[K(x_i, x_j)]_{i,j=1}^s \mathrm{d}x_1 \cdots \mathrm{d}x_s. \tag{5.2}$$

Later, it was shown by Tracy and Widom that the above limit becomes F_2. For a different ensemble, the limiting distribution is different, though it is also called the Tracy-Widom distribution. For example,

$$F_1(s) = \exp\left(-\frac{1}{2}\int_s^\infty q(x)dx\right)(F_2(s))^{1/2},$$

which is the Tracy-Widom distribution for GOE.

Using the tools as Green function comparison, we can prove that the edge universality holds for a very general class of Wigner matrices, e.g., a Wigner matrix with subexponential decay

$$\mathbb{P}\left(|\sqrt{N}h_{ij}| \ge x\right) \le C_0 e^{-x^{1/\theta}}$$

for some positive constant C_0 and $\theta > 1$. Notice that this conditions guarantees that the p-th moment of the random variable $x_{ij} = \sqrt{N}h_{ij}$ is finite for any fixed p. On the other hand, if the p-th moment of x_{ij} is infinite for some $p > 0$, then the Wigner matrix may exhibit a totally different behaviour. For example, it is known that the Wigner semicircle law fails if the variance of x_{ij} is infinite. When the variance is finite but the 4-th moment is infinite, the largest eigenvalue can become larger than 2 although the semicircle law holds as we can see from the following theorem [2]:

Theorem 5.1. *Let λ_N be the largest eigenvalue of a Wigner matrix H and let $x_{ij} = \sqrt{N}h_{ij}$. Then, $\lambda_N \to 2$ as $N \to \infty$ almost surely (with probability 1) if and only if $\mathbb{E}|x_{ij}|^4 < \infty$ for $i \neq j$.*

5.2 Poisson statistics for the largest eigenvalue

If the entries of the Wigner matrix in question have heavy tails, the edge universality may fail, i.e., the limiting distribution for the largest eigenvalue differs from the Tracy-Widom distribution [13]. More precisely, let $x_{ij} = \sqrt{N}h_{ij}$. If the decay of x_{ij} is given by

$$\mathbb{P}(|x_{ij}| > s) \sim s^{-\alpha}$$

for $\alpha < 4$, the largest eigenvalue has Poisson statistics [1]. Notice that the fourth moment of $|x_{ij}|$ is infinite in this case. For simplicity, we consider real symmetric cases only in the following.

When $\alpha < 4$, setting $\delta = 4 - \alpha$, we have

$$\mathbb{P}(|h_{ij}| > N^{\frac{\delta}{4\alpha}}) = \mathbb{P}(|x_{ij}| > N^{\frac{1}{2}+\frac{\delta}{4\alpha}}) \sim N^{-\frac{\alpha}{2}-\frac{\delta}{4}} = N^{-2+\frac{\delta}{2}} \gg N^{-2+\frac{\delta}{4}}.$$

Thus, the probability of having at least one entry bigger than $N^{\frac{\delta}{4\alpha}}$ is

$$1 - \left(\mathbb{P}(|h_{ij}| \le N^{\frac{\delta}{4\alpha}})\right)^{N(N+1)/2} \ge 1 - \left(1 - N^{-2+\frac{\delta}{4}}\right)^{N(N+1)/2} \ge 1 - Ce^{-N^c},$$

and we find that there exists with high probability an entry bigger than $N^{\frac{\delta}{4\alpha}}$ in H.

When $h_{ij} > N^\epsilon$ for some $\epsilon > 0$, we consider $\mathbf{v} = (v_1, v_2, \cdots, v_N)^T \in \mathbb{R}^N$ such that

$$v_i = v_j = \frac{1}{\sqrt{2}}, \quad v_k = 0 \ \text{ if } k \neq i, j.$$

(If $i = j$, we let $v_i = 1$ and all the other entries of $\mathbf{v}$ vanish.) Since $\|v\|_2 = 1$ and

$$\langle \mathbf{v}, H\mathbf{v} \rangle = \frac{1}{2}(H_{ij} + H_{ji}) = H_{ij} > N^\epsilon.$$

Thus, by using the minimax principle, we find that $\lambda_N > N^\epsilon$.

In general, when $\alpha < 4$, the limiting distribution of the largest eigenvalue is solely determined by the order statistics of $\{|h_{ij}|\}_{1 \le i < j \le N}$, i.e., the distribution of the largest one among $(N(N+1)/2)$ i.i.d. random variables $\{|h_{ij}|\}_{1 \le i < j \le N}$. (This becomes the Frechét distribution whose p.d.f. is given by $e^{-x^{-c}}$.) For the k-largest eigenvalue, by using the interlacing property of the eigenvalues, we can prove a similar statement, i.e., the limiting distribution of them is given by the order statistics. It is also known that the such an order statistics corresponds to an inhomogeneous Poisson random point process.

5.3 A necessary and sufficient condition for the edge universality

For a Wigner matrix with heavy tails, the following theorem gives a necessary and sufficient condition for the edge universality [12].

Theorem 5.2. *Let λ_N be the largest eigenvalue of a Wigner matrix H and let $x_{ij} = \sqrt{N} h_{ij}$. Then, the distribution of the rescaled eigenvalue $N^{2/3}(\lambda_N - 2)$ converges to the Tracy-Widom distribution in probability as $N \to \infty$ if and only if*

$$\lim_{s \to \infty} s^4 \mathbb{P}(|x_{ij}| \ge s) = 0$$

for $i \neq j$.

Notice that the condition is weaker than the existence of 4-th moment. For example, if the density of $|x_{ij}|$ decays as $|x|^{-5}(|\log x|^{-1})$, then the 4-th moment does not exists since

$$\int_s^\infty \frac{x^4}{x^5 \log x} \mathrm{d}x = \int_{\log s}^\infty \frac{\mathrm{d}t}{t} = \infty.$$

However, it satisfies the condition in the above theorem, since

$$\lim_{s \to \infty} s^4 \mathbb{P}(|x_{ij}| \ge s) = \lim_{s \to \infty} s^4 \int_s^\infty \frac{1}{x^5 \log x} \mathrm{d}x = \lim_{s \to \infty} \frac{s^{-5}(\log s)^{-1}}{-4s^{-5}} = 0.$$

We remark that the condition in Theorem 5.2 does not contradicts the conclusion in Theorem 5.1 because of the different notion of the convergence.

In order to prove the necessity, one may argue as in the previous section, i.e., the probability of having an element bigger than 4 is quite large if the condition fails. The proof of the sufficiency requires the Green

function comparison theorem. The Green function comparison at the edge features a different scaling of η. Recall that

$$\mathcal{N}(E_1, E_2) \simeq \frac{1}{\pi}\int_{E_1}^{E_2} \mathrm{tr}\,\mathrm{Im}\,\frac{1}{H - E - \mathrm{i}\eta}\mathrm{d}E = \frac{N}{\pi}\int_{E_1}^{E_2} \mathrm{Im}\, m(E + \mathrm{i}\eta)\mathrm{d}E.$$

To see the fluctuation of the largest eigenvalue, we let E_1 and E_2 satisfy $|E_1 - 2|, |E_2 - 2| \sim N^{-2/3}$. Then, the fluctuation of m at this scale can be estimated by using $\eta \sim N^{-2/3}$. In this scaling, the off-diagonal term in the resolvent expansion becomes $G_{ij} \sim N^{-1/3}$, which gives an extra converging factor in the Green function comparison.

References

[1] Auffinger, A., Ben Arous, G., Péché, S.: *Poisson Convergence for the Largest Eigenvalues of Heavy Tailed Random Matrices*, Ann. Inst. Henri Poincaré - Probab. Stat. **45**, 589-610 (2009).

[2] Bai, Z. D., Yin, Y. Q.: *Necessary and Sufficient Conditions for Almost Sure Convergence of the Largest Eigenvalue of a Wigner Matrix*, Ann. of Probab. **16**, 1729-1741 (1988).

[3] Dyson, F. J.: *Statistical Theory of Energy Levels of Complex Systems, I, II, and III.*, J. Math. Phys. **3**, 140-156, 157-165, 166-175 (1962).

[4] Dyson, F. J.: *Correlations between Eigenvalues of a Random Matrix*, Commun. Math. Phys. **19**, 235-250, (1970).

[5] Erdős, L.: *Universality of Wigner random matrices: a Survey of Recent Results*, arXiv:1004.0861v2 (2010).

[6] Erdős, L., Schlein, B., Yau, H.-T.: *Semicircle Law on Short Scales and Delocalization of Eigenvectors for Wigner Random Matrices*, Ann. Probab. **37**, 815-852 (2009).

[7] Erdős, L., Schlein, B., Yau, H.-T.: *Universality of Random Matrices and Local Relaxation flow*, Invent. Math. **185**, 75-119 (2011).

[8] Erdős, L., Yau, H.-T., Yin, J.: *Rigidity of Eigenvalues of Generalized Wigner Matrices*, Adv. Math. **229**, 1435-1515 (2012).

[9] Forrester, P.: *The Spectral Edge of Random Matrix Ensembles*, Nucl. Phys. B **402**, 709-728 (1993).

[10] Mehta, M. L.: *On the Statistical Properties of the Level-Spacings in Nuclear Spectra*, Nuclear Phys. **18**, 395-419 (1960).

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

[12] Lee, J. O., Yin, J.: *A Necessary and Sufficient Condition for Edge Universality of Wigner Matrices*, To appear in Duke Math. J., arXiv:1206.2251 (2012).

[13] Soshnikov, A.: *Poisson Statistics for the Largest Eigenvalue of Wigner Random Matrices with Heavy Tails*, Elect. Comm. in Probab. **9**, 82-91 (2004).

[14] Tao, T., Vu, V.: *Random Matrices: Universality of the Local Eigenvalue Statistics*, Acta Math, **206**, 127-204 (2011).

[15] Tracy, C., Widom, H.: *Level-Spacing Distributions and the Airy Kernel*, Commun. Math. Phys. **159**, 151-174 (1994).

[16] Tracy, C., Widom, H.: *On Orthogonal and Symplectic Matrix Ensembles*, Commun. Math. Phys. **177**, 727-754 (1996).

[17] Wigner, E.: *Characteristic vectors of bordered matrices with infinite dimensions*, Ann. Math. **62**, 548-564 (1955).