Random Matrix Theory and its Applications

MAS583 - Spring 2020

1 Introduction

The random matrix theory was first introduced by Wishart for a multivariate statistics and Wigner for the spectral analysis of heavy nuclei. Later, Gaudin, Mehta, and Dyson 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 random matrices as well.

1.1 Random matrices

A random matrix is a matrix whose entries are random variables. Alternatively, a random matrix is a matrix-valued random variable. These two points of view can be checked from the following classical example.

Definition 1.1 (Gaussian Orthogonal Ensemble (GOE))**.** An $N \times N$ random matrix $H = (H_{ij})$ is said to be a 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. $\mathbb{E}[(H_{ii})^2] = 2N^{-1}$.

Definition 1.1 shows how to define a random matrix as a matrix with random entries. We remark that the normalization $\mathbb{E}[(H_{ij})^2] = N^{-1}$ is to ensure that the typical eigenvalues of a GOE matrix H is of order 1. To see this, we can compute the Frobenius norm of H and find that

$$\|H\|_F^2 = \sum_{i,j=1}^{N} (H_{ij})^2 = N + o(N),$$

which is consistent with that H has N eigenvalues of order 1.

We can introduce a matrix-valued random variable that coincides with the GOE. Since each off-diagonal random variables of a GOE matrix is a Gaussian, the probability density function of a GOE is given by

$$\begin{aligned}
&\prod_{i<j} \frac{1}{\sqrt{2\pi N}} e^{-N(H_{ij})^2/2} \mathrm{d}H_{ij} \prod_{k=1}^{N} \frac{1}{\sqrt{4\pi N}} e^{-N(H_{kk})^2/4} \mathrm{d}H_{kk} \\
&=: Z_N^{-1} \prod_{i<j} e^{-N(H_{ij})^2/2} \mathrm{d}H_{ij} \prod_{k=1}^{N} e^{-N(H_{kk})^2/4} \mathrm{d}H_{kk},
\end{aligned}$$

where Z_N is a normalization constant (partition function) and $\mathrm{d}H_{ij}$ and $\mathrm{d}H_{kk}$ are Lebesgue measures on $\mathbb{R}$. The density function is then

$$Z_N^{-1} \exp\left[-\frac{N}{4} \sum_{i,j=1}^{N} (H_{ij})^2 \right] = Z_N^{-1} \exp\left[-\frac{N}{4} \operatorname{Tr} H^2 \right]. \tag{1.1}$$

We can see that the density depends only on the trace of H^2, and hence the density does not change under an orthogonal transformation O^THO for an $N \times N$ orthogonal matrix. Note also that the normalization $\mathbb{E}[(H_{ii})^2] = 2N^{-1}$ is natural in this point of view.

How can we generalize the GOE? If we use Definition 1.1, it is obvious that we should change the random variables in the matrix. A matrix generalized in this way is called a Wigner matrix. We use the following definition for it.

Definition 1.2 (Wigner matrix)**.** An $N \times N$ symmetric random matrix $H = (H_{ij})$ is said to be a real Wigner matrix if its entries H_{ij} are real 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 $\mathbb{E}[(H_{ii})^2] = 2N^{-1}$ for some constant C.

If we consider the density in (1.1), we can generalize a GOE by defining a probability measure on the space of all $N \times N$ real symmetric matrices whose density is given by

$$Z_N^{-1} \exp\left[-N \ \mathrm{Tr}\, Q(H)\right]$$

for some nice function Q. This matrix-valued random variable is called an orthogonally invariant ensemble or simply an orthogonal ensemble.

Denote by $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ be the eigenvalues of H. We can convert the density function of the GOE, which is defined on $N(N+1)/2$ dimensional space into a density function on its eigenvalues, defined on $\mathbb{R}^N$. Due to the orthogonal invariance, the 'angle' variables do not appear in the density, and by computing the Jacobian, we find that the density is

$$\widetilde{Z}_N^{-1} \exp\left[-N \sum_{i=1}^{N} Q(\lambda_i)\right] \prod_{i<j} |\lambda_i - \lambda_j|.$$

In this definition that contains no matrices anymore, we can think of yet another generalization of the GOE, by introducing a power β in the Vandermonde determinant term $\prod_{i<j} |\lambda_i - \lambda_j|^\beta$. The model is sometimes called a log-gas, and it is also known as a β-ensemble.

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

Definition 1.3 (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 $i \neq j$.

Due to the third condition, and also due to the Itzykson-Zuber integral formula, the GUE is actually simpler to analyze than the GOE in many senses. We also remark that the GUE is a β-ensemble with $\beta = 2$.

In this course, we will mainly consider real symmetric matrices with random variables; other types of random matrices including orthogonal/unitary ensembles, β-ensembles, complex Hermitian matrices, non-square random matrices will only appear briefly.

1.2 Classical ensembles

For historical reasons, some of the most well-known random matrix models are called classical ensembles. The GOE and the GUE are the two most famous examples, and the third classical ensemble involving Gaussians is the Gaussian symplectic ensemble, whose entries are a quaternion Gaussian with a suitable symmetry constraint that guarantees the eigenvalues of a GSE matrix are real.

The following definition is useful in understanding the (global) spectrum of classical ensembles.

Definition 1.4 (Empirical measure)**.** Let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ be the eigenvalues of H. Then, the empirical (spectral) measure of H is defined by

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

The first result in the history of random matrix theory was about the limiting distribution of the empirical measure of a random matrix. For a GOE matrix (and a Wigner matrix in general),

Theorem 1.5 (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*

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

1.2.1 Wishart ensemble

Let (X_{ij}) be i.i.d. Gaussian real random variables with

$$\mathbb{E}[X_{ij}] = 0, \quad \mathbb{E}[X_{ij}^2] = 1$$

for $1 \leq i \leq P$ and $1 \leq j \leq N$. We let

$$S_N = \frac{1}{N}XX^T \in \mathbb{R}^{P\times P}, \tag{1.4}$$

which is called the Wishart ensemble. Random matrices of the form (1.4) appear in the multivariate statistics as 'sample covariance matrices'. It is easy to see that S_N is positive semidefinite, i.e., if we denote by $\lambda_1, \lambda_2, \cdots, \lambda_P$ the eigenvalues of S_N, then $\lambda_i \geq 0$ for all $i = 1, 2, \cdots, P$. We also define the empirical spectral measure of the eigenvalues by

$$\mu(S_N) = \frac{1}{P}\sum_{i=1}^{P} \delta_{\lambda_i}. \tag{1.5}$$

A sample covariance matrix (or Gram matrix) is a generalization of Wishart ensemble, where each X_{ij} is not necessarily a Gaussian. The limiting eigenvalue distribution of the sample covariance matrices is given by the following Marchenko–Pastur law.

Theorem 1.6 (Marchenko–Pastur law)**.** *Assume that $N, P \to \infty$ with the ratio $P/N \to d \in [0, 1]$. Let $a_- = (1-\sqrt{d})^2$ and $a_+ = (1+\sqrt{d})^2$. Then, the empirical measure μ_{S_N} of a Wishart ensemble S_N converges weakly to the Marchenko–Pastur law defined by*

$$\mathrm{d}\mu_{MP(d)}(x) = \frac{1}{2\pi dx}\sqrt{(a_+ - x)(x-a_-)_+}\mathrm{d}x. \tag{1.6}$$

Remark 1.7. Since XX^T and X^TX have the same non-zero eigenvalues, we find for the case $P > N$ that

$$\mu_{S_N} \to (1-\frac{1}{d})\delta_0 + \mu_{MP(d)}.$$

When $P = N$, i.e., $d = 1$, the Marchenko-Pastur law becomes

$$\mathrm{d}\mu_{MP(1)}(x) = \frac{1}{2\pi x}\sqrt{(4x-x^2)_+}\mathrm{d}x = \frac{1}{2\pi}\sqrt{\left(\frac{4}{x}-1\right)_+}\mathrm{d}x, \tag{1.7}$$

which is known as the quarter-circle law (for the eigenvalues). Note that the density diverges as $x \to 0$.

The name 'quarter-circle' originated from the distribution of the singular values. Consider an $N \times N$ random matrix X with i.i.d. entries of mean zero and variance N^{-1}. Since the singular values of X are the square root of the eigenvalues of S_N with $P = N$, by letting $x = y^2$, we find that

$$\mathrm{d}\mu_{sing}(y) = \frac{1}{\pi}\sqrt{4-y^2}\cdot \mathbb{1}_{[0,2]}\mathrm{d}y,$$

which indeed describes a quarter-circle measure.

1.2.2 Ginibre ensemble

Let X_{ij} are i.i.d. complex Gaussian random variables with mean zero and variance N^{-1}. Unlike the Wigner case, if there are no symmetry constraint for an $N \times N$ random matrix $X = (X_{ij})$, the eigenvalues are complex numbers. Such a matrix X is called a Ginibre ensemble. In this case, the empirical measure converges to the uniform measure on the unit disk in the complex plane.

Theorem 1.8 (Full circle law)**.** *Let μ be the empirical spectral measure of a Ginibre ensemble X. Then, μ converges to the uniform measure on the unit disk whose distribution is given by $\frac{1}{\pi}\mathbb{1}_{\{|z|\leq 1\}}$.*

1.2.3 Circular ensemble

The circular unitary ensemble (CUE) is the Haar measure on the unitary group $U(N)$. The circular orthogonal ensemble (COE) and the circular symplectic ensemble (CSE) are defined in a similar manner. It is obvious that every eigenvalue of a circular ensemble is of modulus 1. The limiting distribution of the empirical measure is the uniform measure on the unit circle in the complex plane.

General remarks on the notation

The limit, O, o, $\gg$, $\ll$, and other notations will always refer to the limit $N \to \infty$ unless noticed otherwise. The notation $A \sim B$ means that

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

for some constant C independent of N.

For X and Y, which can be deterministic numbers and/or random variables depending on N, we use the notation $X = \mathcal{O}(Y)$ if for any (small) $\varepsilon > 0$ and (large) $D > 0$ there exists $N_0 \equiv N_0(\varepsilon, D)$ such that $\mathbb{P}(|X| > N^{\varepsilon}|Y|) < N^{-D}$ whenever $N > N_0$.

For an event Ω, we say that Ω holds with high probability if for any (large) $D > 0$ there exists $N_0 \equiv N_0(D)$ such that $\mathbb{P}(\Omega^c) < N^{-D}$ whenever $N > N_0$.