CPSC 664, Spring 2023 | Scribe: Jinzhao Wu
Feb 21, 2023 | Lecturer: Tim Kunisky

LECTURE 11: Free Probability

From the Kac-Rice Formula, our objective is to comprehend the eigenvalues of $W + D$, where W is the GOE matrix and D denotes the random diagonal matrix. In our previous lecture, we derived the semicircle law for the GOE matrix W and the eigenvalues of D are also clear. In the following, we will present a general method to integrate these findings and elucidate the eigenvalues of the sum of two random matrices.

1 CLT via Renormalization

Before delving into the method, we present an analogous way to understanding the semicircle law through a renormalization approach. Initially, we demonstrate how this approach can be used to prove the Central Limit Theorem.

Theorem 1.1 (Central Limit Theorem)**.** *Consider a sequence of i.i.d. samples* $\{A_1, A_2, \cdots, A_n\}$ *such that* $\mathbb{E}[A_i] = 0$ *and* $\mathsf{Var}[A_i] = 1$*. Define the random variable* S_n *as follows:*

$$S_n \stackrel{def}{=} \frac{A_1 + A_2 + \cdots + A_n}{\sqrt{n}}.$$

Then, in distribution, S_n *converges to a standard guassian distribution, i.e.,*

$$S_n \stackrel{(d)}{\to} N(0,1).$$

In Lecture 7, we established this by computing the moments as n approaches infinity and subsequently solving the moment problem to obtain the limiting distribution. However, in this instance, we will demonstrate the proof using the renormalization approach. We won't give a strict proof here but also provide some high level ideas.

Define M as the set of all probability distributions over $\mathbb{R}$. Let f represent a mapping from M to itself. For any given distribution μ, $f(\mu)$ is defined as the law of $\frac{X_1+X_2}{\sqrt{2}}$, where X_1 and X_2 are i.i.d. random variables drawn from μ.

We first notice that

$$\begin{aligned} f(\text{law of } S_n) &\stackrel{(d)}{=} \frac{S_n + S_n'}{\sqrt{2}} \\ &\stackrel{(d)}{=} \left(\frac{A_1 + A_2 + \cdots + A_n}{\sqrt{n}} + \frac{A_1' + A_2' + \cdots + A_n'}{\sqrt{n}}\right)/\sqrt{2} \\ &\stackrel{(d)}{=} \frac{A_1 + A_2 + \cdots + A_{2n}}{\sqrt{2n}} \\ &\stackrel{(d)}{=} \text{law of } S_{2n} \end{aligned} \tag{1}$$

This implies that as n approaches infinity, the law of S_n always goes along the mapping f. Additionally, observe that $N(0,1)$ serves as a fixed point for the mapping f. This is because that

$$\frac{X_1+X_2}{\sqrt{2}} \overset{(d)}{=} N(0,1) \text{ when } X_1, X_2 \overset{i.i.d.}{\sim} N(0,1)$$

Consequently, the Central Limit Theorem can be interpreted as a fixed-point theorem for the mapping f. Starting from a sufficiently well-behaved distribution, iterating the mapping f will always converge to the fixed point. However, one might question whether $N(0,1)$ is the sole fixed point for this mapping. We will now demonstrate that for any fixed-point distribution μ, μ could be recovered as a Gaussian distribution.

Given any fixed point distribution μ, define m_k as the k-th moment of μ, i.e.,

$$m_k \overset{\text{def}}{=} \mathop{\mathbb{E}}_{X\sim\mu}\left[X^k\right]. \tag{2}$$

As we know that μ is a fixed point distribution of f, this implies that

$$X \overset{(d)}{=} \frac{X_1+X_2}{\sqrt{2}}$$

where X, X_1 and X_2 are i.i.d. samples from μ. Thus, we could see that

$$\begin{aligned} m_k &= \mathop{\mathbb{E}}_{X\sim\mu}\left[X^k\right] \\ &= \mathop{\mathbb{E}}_{X_1,X_2\sim\mu}\left[\left(\frac{X_1+X_2}{\sqrt{2}}\right)^k\right] \\ &= \frac{1}{2^{k/2}}\sum_{j=0}^{k}\binom{k}{j} m_j m_{k-j}. \end{aligned}$$

Rephrasing the equation above, we get the following recursion:

$$m_k = \frac{1}{2^{k/2}-2}\sum_{j=1}^{k-1} m_j m_{k-j}.\ \forall k \neq 2$$

By solving this recursion, we then recover that $N(0,1)$ is the only fixed point given the fact that the variance is only 1.

We can also recover the normal distribution by characteristic function. For any fixed-point distribution u,

$$\begin{aligned} \Phi_u(t) &= \mathop{\mathbb{E}}_{X\sim u} \mathsf{exp}(itX) \\ &= \mathop{\mathbb{E}}_{X_1,X_2\sim u} \mathsf{exp}\left(it\left(X_1+X_2\right)/\sqrt{2}\right) \\ &= \Phi_u\left(\frac{\sqrt{t}}{\sqrt{2}}\right)^2. \end{aligned}$$

Solving this equation, we could also show that the Gaussian distribution is the only valid distribution.

2 Adaption to Random Matrices

We will now adapt the approaches above to the context of random matrices. Let M represent the set of probability distributions over $N \times N$ symmetric matrices, while the function f remains unchanged.

Similarly, we could see that GOE matrix is still a fixed point for the mapping f, i.e.

$$f(\text{GOE}) \overset{(d)}{=} \text{GOE}$$

Our goal is to learn $\lim_{n\to\infty} \frac{1}{N} \operatorname{Tr}\left(W^k\right)$ from the equation above where

$$W_{ij} = W_{ij} \sim N(0, 1 + \delta_{ij}),$$

and $\widehat{W} = \frac{1}{\sqrt{N}} W$ is a normalization version of the GOE matrix. Using the same approach, define m_k as

$$m_k \overset{\text{def}}{=} \mathop{\mathbb{E}}_{W\sim\text{GOE}} \operatorname{Tr}(W^k).$$

We could also get a similar recursion:

$$\begin{aligned} m_k &= \mathop{\mathbb{E}}_{W\sim\text{GOE}} \operatorname{Tr}\left(W^k\right) \\ &= \mathop{\mathbb{E}}_{W_1,W_2\sim\text{GOE}} \operatorname{Tr}\left(\frac{W_1 + W_2}{\sqrt{2}}\right)^k \\ &= \frac{1}{2^{k/2}} \sum_{\substack{a_1+a_2+\cdots a_n=k \\ a_i\geq 1}} \operatorname{Tr}\left(W_1^{a_1} W_2^{a_2} W_1^{a_3} W_2^{a_4} \cdots\right) \end{aligned} \tag{3}$$

However, W_1 and W_2 do not commute with each other, so we cannot do similar operations as the scaler case. We first have a look at a simpler case that we can handle:

Example 2.1. $\mathbb{E}\left[\operatorname{Tr} W_1^a W_2^b\right]$. *This is quite like the scaler case. We could see that*

$$\begin{aligned} \mathbb{E}\left[\operatorname{Tr} W_1^a W_2^b\right] &= \operatorname{Tr}\left(\left(\mathbb{E} W_1^a\right)\left(\mathbb{E} W_2^b\right)\right) \\ &\approx \frac{1}{N}\left(\mathbb{E}\operatorname{Tr} W_1^a\right)\left(\mathbb{E}\operatorname{Tr} W_2^b\right) \\ &\approx C_{a/2} C_{b/2} N^{k/2+1}. \end{aligned}$$

where The last step is based on the path counting argument from our previous lecture.

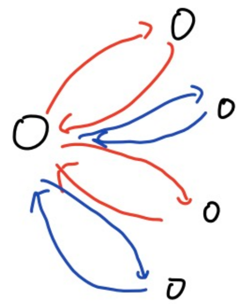
Now let us try to check the case when $k = 4$. Expanding (3), we could see that

$$\begin{aligned} 2N^3 = \frac{1}{4}\Big[&\left(\mathbb{E}\operatorname{Tr} W_1^4 + \mathbb{E}\operatorname{Tr} W_2^4\right) \\ &+ \left(\mathbb{E}\operatorname{Tr} W_1^2 W_2^2 + \mathbb{E}\operatorname{Tr} W_2^2 W_1^2 + \mathbb{E}\operatorname{Tr} W_1 W_2^2 W_1 + \mathbb{E}\operatorname{Tr} W_2 W_1^2 W_2\right) \\ &+ \left(\mathbb{E}\operatorname{Tr} W_1 W_2 W_1 W_2 + \mathbb{E}\operatorname{Tr} W_2 W_1 W_2 W_1\right) \\ &+ \left(\text{terms with odd number of either } W_i\right).\Big] \end{aligned}$$

We could observe that $\mathbb{E}\operatorname{Tr} W_1^4 = \mathbb{E}\operatorname{Tr} W_2^4 \approx 2N^3$, and $\mathbb{E}\operatorname{Tr} W_1^2 W_2^2 = \mathbb{E}\operatorname{Tr} W_2^2 W_1^2 = \mathbb{E}\operatorname{Tr} W_1 W_2^2 W_1 = \mathbb{E}\operatorname{Tr} W_2 W_1^2 W_2 \approx N^3$. It is also clear that the contributions by terms with odd number of either W_i is negligible. This suggests that the contribution from $\mathbb{E}\operatorname{Tr} W_1 W_2 W_1 W_2$ and $\mathbb{E}\operatorname{Tr} W_1 W_2 W_1 W_2$ should be $o(N^3)$. Actually this can be proved by counting colored walks. Let us consider a colored version of the closed walk we encountered last time, where the edges from W_1 are blue and the edges from W_2 are red. As we proved last time, to get a non-zero contribution, every edge must be traversed an even time. This implies that the only contributions are from closed walks within two vertices, which are only $O(n^2)$.

Now we consider a hard case where $k = 8$. In this case, there will be some terms like $\mathbb{E}\operatorname{Tr}(W_1^2 W_2^2 W_1^2 W_2^2)$. Again let us consider counting colored walks. We could see that there could be a valid closed walk within five vertices, which is in the following shape:

Figure 1: A valid closed walk for the term



These observations imply that the contribution of $\mathbb{E}\operatorname{Tr}(W_1^2 W_2^2 W_1^2 W_2^2)$ is $O(N^5) = O(N^{8/2}+1)$, which is not negligible when considering the case where $k = 8$. As a result, this raises the question of when we can obtain $\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\operatorname{Tr}\left(X^{(N)} + Y^{(N)}\right)^k$ from $\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\operatorname{Tr}\left(X^{(N)}\right)^k$ and $\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\operatorname{Tr}\left(Y^{(N)}\right)^k$. In the scalar case, as discussed in Section 1, independence between $X^{(N)}$ and $Y^{(N)}$ is sufficient. However, in the matrix case, mere independence is not adequate to derive this result, as demonstrated above. Consequently, we introduce the concept of asymptotic freeness in the following section, which will prove beneficial in addressing this issue.

3 Asymptotic Freeness

We now introduce the definition of asymptotic freeness.

Definition 3.1. *For any two sequences of two matrices $\left\{X^{(n)}\right\}$ and $\left\{Y^{(n)}\right\}$, they are asymptotically free if and only if for any bunch of polynomials $\{p_1, \cdots, p_l\}$ and $\{q_1, \cdots, q_l\}$ such that*

$$\frac{1}{N}\operatorname{Tr}\left[p_i\left(X^{(N)}\right)\right] \to 0 \text{ and } \frac{1}{N}\operatorname{Tr}\left[q_i\left(Y^{(N)}\right)\right] \to 0 \quad \forall i \in [l],$$

it holds that

$$\frac{1}{N}\mathbb{E}\left[\operatorname{Tr}\left(p_1\left(X^{(N)}\right) q_1\left(Y^{(N)}\right) \cdots p_l\left(X^{(N)}\right) q_l\left(Y^{(N)}\right)\right)\right] \to 0.$$

This condition is actually saying that for any term that forms an interleaving of $X^{(N)}$ and $Y^{(N)}$ where they individually are negligible, the whole term is also negligible.

We could prove that two independent instances drawn from GOE are asymptotically free. Utilizing this definition, we can approach the problem more systematically. Let us now examine some examples.

Example 3.2. *Let $l = 1$, $p_1(t) = t^a - C_{a/2}$ and $q_1(t) = t^b - C_{b/2}$.*

As we proved before, we could easily see that $p_1(t)$ and $q_1(t)$ satisfies the conditions in Definition 3.1. Then, the definition of asymptotic freeness implies that

$$\begin{aligned} 0 &= \lim_{N\to\infty} \frac{1}{N} \mathbb{E}\left[\mathsf{Tr}\left(\left(\widehat{W_1}^a - C_{a/2}I\right)\left(\widehat{W_2}^b - C_{b/2}I\right)\right)\right] \\ &= \lim_{N\to\infty} \frac{1}{N} \mathbb{E}\left[\mathsf{Tr}\,\widehat{W_1}^a\widehat{W_2}^b - C_{a/2}C_{b/2} - C_{a/2}C_{b/2} + C_{a/2}C_{b/2}\right] \end{aligned}$$

This gives us the same result as Example 2.1.

Example 3.3. *Let $l = 2$, and $p_1(t) = p_2(t) = q_1(t) = q_2(t) = t$. Again, it is clear that $p_i(t)$ and $q_i(t)$ satisfies the conditions of asymptotic freeness. Therefore, it holds that*

$$0 = \lim_{N\to\infty} \frac{1}{N} \mathbb{E}\left[\mathsf{Tr}\left(\widehat{W_1}\widehat{W_2}\widehat{W_1}\widehat{W_2}\right)\right],$$

which recovers the proof that shows that the contribution from $\mathbb{E}\,\mathsf{Tr}\,W_1W_2W_1W_2$ is negligible.

Define $\hat{m}_k$ as $\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\,\mathsf{Tr}\,\widehat{W}^k$. Just as Example 3.2, we define similar polynomials $p_i(t)$ and $q_i(t)$. By asymptotically freeness, we could see that

$$0 = \lim_{N\to\infty} \frac{1}{N} \mathbb{E}\,\mathsf{Tr}\left(\left(\widehat{W_1}^{a_1} - C_{a_1/2}I\right)\left(\widehat{W_2}^{b_1} - C_{b_1/2}I\right)\cdots\left(\widehat{W_1}^{a_l} - C_{a_l/2}I\right)\cdots\right).$$

Expanding the equation above, we then get a closed recursion for $\hat{m}_k$.

Finally, we introduce two corollaries that will be useful.

Corollary 3.4. *When $X^{(N)}, Y^{(N)}$ are asymptotically free, then $\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\,\mathsf{Tr}\left(X^{(N)} + Y^{(N)}\right)^k$ is a function of $\left\{\lim_{N\to\infty} \frac{1}{N}\mathbb{E}\,\mathsf{Tr}\left(X^{(N)}\right)^k, \lim_{N\to\infty} \frac{1}{N}\mathbb{E}\,\mathsf{Tr}\left(Y^{(N)}\right)^k\right\}$.*

Corollary 3.5. *Suppose the empirical spectral distribution of $X^{(N)}$ converges to some distribution μ and the empirical spectral distribution of $Y^{(N)}$ converges to ν. If $X^{(N)}$ and $Y^{(N)}$ are asymptotically free, the empirical spectral distribution of $X^{(N)} + Y^{(N)}$ is converges to some distribution determined by μ and ν, and we call it $\mu \boxplus \nu$. This is called free convolution and we will further discuss it in the following lecture.*