Introduction to Random Matrix Theory and Applications

Def: Let $\{Z_{ij}\}_{1\le i\le j\le n}$ be independent random variables, such that

- $\mathbb{E}[Z_{ij}]=0$ for all $i\le j$
- $\mathbb{E}[Z_{ij}^2]=1$ for all $i<j$. (No requirement for $i=j$).
- $\mathbb{E}[|Z_{ij}|^k]<C_k$ for a constant C_k and all $i\le j$.

The symmetric matrix $X\in\mathbb{R}^{n\times n}$ with

$$X_{ij}=\begin{cases} Z_{ij}/\sqrt{n} & i\le j \\ Z_{ji}/\sqrt{n} & i>j \end{cases}$$

is called a <u>Wigner matrix</u>.

Def: If $Z_{ij}\sim\mathcal{N}(0,1)$ for $i<j$ and $Z_{ii}\sim\mathcal{N}(0,2)$ for $i=j$, then X is called the <u>Gaussian Orthogonal Ensemble</u> (GOE). We write $X\sim GOE(n)$.

Prop: If $X\sim GOE(n)$ and $O\in\mathbb{R}^{n\times n}$ is any orthogonal matrix, then also $O^TXO\sim GOE(n)$.

Proof: Let $W\in\mathbb{R}^{n\times n}$ be an <u>asymmetric</u> matrix, with $W_{ij}\overset{iid}{\sim}\mathcal{N}(0,1)$ for all $i,j\in\{1,\dots,n\}$. Then $X=\frac{W+W^T}{\sqrt{2n}}$ is a symmetric matrix with Gaussian entries, satisfying:

- $\{X_{ij}\}_{i\le j}$ are independent.
- $\mathbb{E}[X_{ij}]=\frac{1}{\sqrt{2n}}\mathbb{E}[W_{ij}+W_{ji}]=0.$

- $\mathrm{Var}[X_{ij}] = \frac{1}{2n} \mathrm{Var}[W_{ij} + W_{ji}] = \begin{cases} \frac{1}{n} & i<j \\ \frac{2}{n} & i=j. \end{cases}$

Thus $X \sim GOE(n)$. Let $\widetilde{W} = O^T W O$. Then $O^T X O = \frac{\widetilde{W} + \widetilde{W}^T}{\sqrt{2n}}$,

and it suffices to show $W \overset{L}{=} \widetilde{W}$. For this, note:

- $\sum_{ij} W_{ij}^2 = \mathrm{Tr}\, W^T W = \mathrm{Tr}\, \widetilde{W}^T \widetilde{W} = \sum_{ij} \widetilde{W}_{ij}^2$.

  So $W \mapsto O^T W O$ is an isometry from $\mathbb{R}^{n \times n}$ to itself, with Jacobian 1.

- The joint density of entries of W is

$$\prod_{ij} \frac{1}{\sqrt{2\pi}} e^{-\frac{W_{ij}^2}{2}} = \left(\frac{1}{\sqrt{2\pi}}\right)^{n^2} e^{-\frac{1}{2}\sum_{ij} W_{ij}^2}$$

$$= \left(\frac{1}{\sqrt{2\pi}}\right)^{n^2} e^{-\frac{1}{2}\sum_{ij} \widetilde{W}_{ij}^2} = \prod_{ij} \frac{1}{\sqrt{2\pi}} e^{-\frac{\widetilde{W}_{ij}^2}{2}}.$$

Thus $\prod_{ij} \frac{1}{\sqrt{2\pi}} e^{-\frac{\widetilde{W}_{ij}^2}{2}}$ is the joint density of entries of $\widetilde{W}$, i.e. $\widetilde{W}_{ij} \overset{iid}{\sim} N(0,1)$.

Corollary: Let $X \sim GOE(n)$. Let $X = V \Lambda V^T$ be its spectral decomposition, where $\Lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_n)$ is ordered s.t. $\lambda_1 \geq \dots \geq \lambda_n$. Then for any orthogonal matrix $O \in \mathbb{R}^{n \times n}$, $O^T V \overset{L}{=} V$.

Def: For $X \in \mathbb{R}^{n \times n}$ symmetric, its <u>empirical spectral distribution (ESD)</u> is the discrete probability distribution

$$\mu_n = \frac{1}{n} \sum_{i=1}^{n} \delta_{\lambda_i(X)},$$

where $\lambda_1(X), \dots, \lambda_n(X)$ are the (real) eigenvalues of X.

Thm: Let $X \in \mathbb{R}^{n \times n}$ be Wigner. Then almost surely as $n \to \infty$, its ESD μ_n converges weakly to the <u>semicircle law</u> μ on $[-2, 2]$. This is the continuous distribution with density

$$f(x) = \frac{1}{2\pi} \sqrt{4 - x^2} \cdot \mathbb{1}\{x \in [-2,2]\}.$$

<u>Probability measures, weak convergence, and the moment method</u>

Idea: We want a way of describing when two probability measures μ_n and μ are "close." We often do this by showing that

$$\left| \int f(x)\, d\mu_n(x) - \int f(x)\, d\mu(x) \right| \text{ is small,}$$

for a sufficiently large class of test functions f.

In random matrix theory, we want to study a class of test functions f such that

- Sufficiently large to approximate any other "nice" function.
- $\int f(x)\, d\mu_n(x) = \frac{1}{n} \sum_{i=1}^{n} f(\lambda_i(X))$ can be described simply in terms of the entries of X.

Today: We'll take polynomials, generated by $f(x) = x^k$.

Then $\int f(x)\, d\mu_n(x) = \frac{1}{n} \sum_{i=1}^{n} \lambda_i(X)^k = \frac{1}{n} \mathrm{Tr}\, X^k$.

This is called the <u>moment method</u>.

Next week: We'll take rational functions, generated by $f(x) = \frac{1}{x - z}$ for $z \in \mathbb{C}$. Then $\int f(x)\, d\mu_n(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\lambda_i(X) - z} = \frac{1}{n} \mathrm{Tr}\, (X - zI)^{-1}$.

This is called the <u>Stieltjes transform / resolvent method</u>.

Def: Let $\mathcal{P}$ be the space of (Borel) probability measures on $\mathbb{R}$. $\mu_n \in \mathcal{P}$ converges <u>weakly</u> to $\mu \in \mathcal{P}$ if, for any continuous, bounded function $f: \mathbb{R} \to \mathbb{R}$,

$$\int f(x)\, d\mu_n(x) \to \int f(x)\, d\mu(x) \quad \text{as } n \to \infty.$$

Facts:

- This topology of weak convergence is metrizable: There is a metric $d(\cdot,\cdot)$ on $\mathcal{P}$ such that $\mu_n \to \mu$ weakly if and only if $d(\mu_n, \mu) \to 0$.

  One such choice is the Lévy distance

  $$d(\mu_n, \mu) = \inf \{ \varepsilon > 0 : F_n(x-\varepsilon) - \varepsilon \le F(x) \le F_n(x+\varepsilon) + \varepsilon \}$$

  where F_n, F are the CDFs of μ_n, μ.

- "$\mu_n \to \mu$ weakly" is equivalent to "$F_n(x) \to F(x)$ at every x where F is continuous."

- The open sets of $(\mathcal{P}, d)$ generate a Borel σ-algebra on $\mathcal{P}$.

  For a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, a <u>random measure</u> μ_n is a Borel-measurable map $\Omega \mapsto \mathcal{P}$. We'll typically ignore Ω and just write $\mu_n \in \mathcal{P}$ (in the same way we write $X \in \mathbb{R}$ for a r.v. X).

  Then $\mu_n \to \mu$ <u>weakly a.s.</u> if $d(\mu_n, \mu) \to 0$ a.s. as $n \to \infty$.

Prop (Moment Method): Suppose $\int x^k d\mu_n(x) \to \int x^k d\mu(x)$ as $n \to \infty$ for each fixed integer $k \ge 0$, and $\mu \in \mathcal{P}$ has bounded support. Then $\mu_n \to \mu$ weakly.

Proof: Let the support of μ be contained in $[-M, M]$. Then for any $j, k \ge 0$,

$$\int x^{2j}\, \mathbb{1}\{|x| > M+1\}\, d\mu_n(x) \le \int \frac{x^{2j+2k}}{(M+1)^{2k}}\, d\mu_n(x)$$

$$\to \frac{\int x^{2j+2k} d\mu(x)}{(M+1)^{2k}} \le \left(\frac{M}{M+1}\right)^{2k} \cdot M^{2j}.$$

Since k is arbitrary, this shows $\int x^{2j}\, \mathbb{1}\{|x| > M+1\}\, d\mu_n(x) \to 0$ for any fixed j. Now let $f(x)$ be any continuous bounded function. Fix $\varepsilon > 0$. By the Weierstrass approximation theorem, there is a polynomial $p(x)$ s.t.

$$\sup_{|x| \le M+1} |p(x) - f(x)| \le \varepsilon.$$

Then

$$\left| \int f(x)\, d\mu_n(x) - \int f(x)\, d\mu(x) \right|$$

$$\le \left| \int p(x)\, d\mu_n(x) - \int p(x)\, d\mu(x) \right| + \int |p(x) - f(x)| \cdot d\mu(x) + \int |p(x) - f(x)|\, d\mu_n(x).$$

1. $\int p(x)\, d\mu_n(x) - \int p(x)\, d\mu(x) \to 0.$
2. $\int |p(x) - f(x)|\, d\mu(x) = \int |p(x) - f(x)|\, \mathbb{1}\{|x| \le M+1\}\, d\mu(x) \le \varepsilon.$
3. $\int |p(x) - f(x)|\, d\mu_n(x) = \underbrace{\int |p(x) - f(x)|\, \mathbb{1}\{|x| \le M+1\}\, d\mu_n(x)}_{\le \varepsilon} + \underbrace{\int |p(x) - f(x)|\, \mathbb{1}\{|x| > M+1\}\, d\mu_n(x)}_{\to 0}.$

Since $\varepsilon > 0$ is arbitrary, this shows $\int f(x)\, d\mu_n(x) \to \int f(x)\, d\mu(x)$.

Note: Some condition on μ is necessary, otherwise it may not be uniquely defined by its moments.

Cor: If $\mu_n \in \mathcal{P}$ is random, $\int x^k d\mu_n(x) \to \int x^k d\mu(x)$ a.s. for all k, and μ has bounded support, then $\mu_n \to \mu$ weakly a.s.

Moments of the semicircle law

Q: Let μ be the semicircle law on $[-2,2]$. What is $m_k \equiv \int_{-2}^{2} x^k \, d\mu(x)$?

Prop: Let $C_k = \frac{\binom{2k}{k}}{k+1} = \frac{(2k)!}{k!\,(k+1)!}$. Then $m_{2k} = C_k$, $m_{2k+1} = 0$.

Proof: Odd moments are 0 by symmetry. Integration by parts shows

$$m_{2k} = \frac{1}{2\pi}\int_{-2}^{2} x^{2k}\sqrt{4-x^2}\,dx = \frac{2k(2k-1)}{k(k+1)} m_{2k-2},$$ and even moments follow.

Remark: C_k is the k^{th} Catalan number. These satisfy a recursion

$$C_k = \sum_{i=0}^{k-1} C_i C_{k-1-i}.$$ Richard Stanley, Enumerative Combinatorics: Vol 2

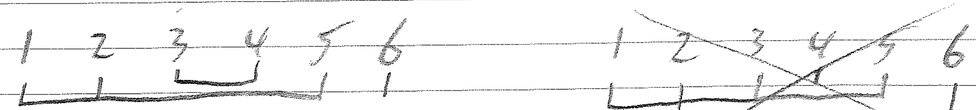
lists 66 different things counted by C_k. For example:

- Number of Dyck words of length $2k$:

  (((())))   ()(()())   (())(())   ~~)()(~~

  Indeed, look at where first (is paired by). If, inside this pairing, is a Dyck word of length $2i$, then after this pairing is a Dyck word of length $2(k-1-i)$. This yields $C_k = \sum_{i=0}^{k-1} C_i C_{k-1-i}$.

- Number of non-crossing partitions of $\{1,\dots,k\}$:



  Indeed, look at largest number j in the set containing 1. Removing j induces a non-crossing partition of $\{1,\dots,j-1\}$ and of $\{j+1,\dots,k\}$.

  Again we get $C_k = \sum_{j=1}^{k} C_{j-1} C_{k-j} = \sum_{i=0}^{k-1} C_i C_{k-1-i}$.

Lemma: Let X be a Wigner matrix, with ESD μ_n. For any fixed $k \geq 0$,

$$\mathbb{E}\left[\int x^k \, d\mu_n(x)\right] \to m_k.$$

Proof: Write

$$\int x^k \, d\mu_n(x) = \frac{1}{n}\sum_{i=1}^{n} \lambda_i(X)^k = \frac{1}{n}\sum_{i=1}^{n}\lambda_i(X^k) = \frac{1}{n}\operatorname{Tr} X^k.$$

Note: The (i,j) entry of X^2 is $\sum_{k=1}^{n} X_{ik}X_{kj}$.

The (i,j) entry of X^3 is $\sum_{l=1}^{n}(X^2)_{il}X_{lj} = \sum_{k,l=1}^{n} X_{ik}X_{kl}X_{lj}$.

etc...

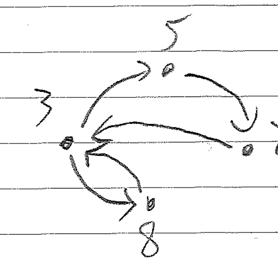
The (i_1,i_1) diagonal entry of X^k is

$$\sum_{i_2,\dots,i_k=1}^{n} X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_{k-1} i_k} X_{i_k i_1}.$$

Then $$\frac{1}{n}\operatorname{Tr} X^k = \frac{1}{n}\sum_{i_1,\dots,i_k=1}^{n} X_{i_1 i_2}\cdots X_{i_{k-1} i_k} X_{i_k i_1}.$$

Each term of the sum is indexed by $(i_1, i_2, \dots, i_k)$. Associate to this a walk on the graph w/ vertices $\{1,\dots,n\}$ starting at i_1 and traversing the edges $(i_1,i_2), (i_2,i_3), \dots, (i_k,i_1)$.

$(3, 5, 2, 3, 8) \Longrightarrow$



The complete, undirected graph on $\{1,\dots,n\}$ has $\frac{n(n+1)}{2}$ distinct edges, including self-loops. These correspond to the $\frac{n(n+1)}{2}$ independent entries of X.

For any walk $(i_1, \dots, i_k)$, if $\mathcal{E}$ is the set of distinct edges traversed by this walk, and $k(e)$ is the number of times edge $e \in \mathcal{E}$ is traversed, then by independence

$$\mathbb{E}[X_{i_1 i_2} X_{i_2 i_3} \cdots X_{i_k i_1}] = \prod_{e \in \mathcal{E}} \mathbb{E}[X_e^{k(e)}]$$

where $X_e \equiv X_{ij}$ if $e = \{i, j\}$.

We divide the set of all walks $(i_1, i_2, \dots, i_k)$ into three cases:

① Some edge e is traversed by this walk only once, i.e. $k(e) = 1$.

Then $\mathbb{E}[X_{i_1 i_2} \cdots X_{i_k i_1}] = 0$, since $\mathbb{E}[X_e] = 0$.

For all other walks, the number of distinct edges traversed is at most:

$$\begin{cases} k/2 & k \text{ even} \\ \frac{k-1}{2} & k \text{ odd.} \end{cases}$$

Then the number of distinct vertices visited is at most $k/2 + 1$, if k is even, or $\frac{k-1}{2} + 1 = \frac{k+1}{2}$ if k is odd.

② The number of vertices visited is at most $\frac{k+1}{2}$.

How many such walks can there be? We can construct any such walk by first picking $\frac{k+1}{2}$ vertices from $\{1, \dots, n\}$, then picking one of these vertices for each $i_1, \dots, i_k$. So

$$\# \text{ such walks} \le \binom{n}{\frac{k+1}{2}} \cdot \left(\frac{k+1}{2}\right)^k \le n^{\frac{k+1}{2}} \cdot C_k$$

for a constant $C_k > 0$.

Recall $X_{ij} = \frac{1}{\sqrt{n}} Z_{ij}$ where Z_{ij} has bounded moments of all orders.

Then, since $\sum_{e \in \mathcal{E}} k(e) = k$, we have $\left|\prod_{e \in \mathcal{E}} \mathbb{E}[X_e^{k(e)}]\right| \le \frac{1}{n^{k/2}} \cdot C_k'$ for some constant $C_k' > 0$, and any walk.

Then $\left|\mathbb{E}\left[\frac{1}{n} \sum_{\text{Case ② walks}} X_{i_1 i_2} \cdots X_{i_k i_1}\right]\right|$

$$\le \frac{1}{n} \cdot n^{\frac{k+1}{2}} \cdot C_k \cdot \frac{1}{n^{k/2}} \cdot C_k' \le \frac{C_k C_k'}{\sqrt{n}}.$$

③ Each edge e is traversed at least twice, k is even, and the number of visited vertices is exactly $\frac{k}{2} + 1$. Then:

- Each edge is traversed exactly twice
- The graph of traversed edges is a tree w/ $\frac{k}{2}$ edges and no self-loops.

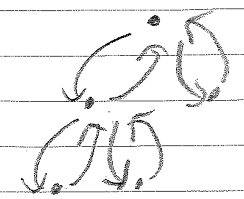
For each such walk, $k(e) = 2 \ \forall e \in \mathcal{E}$, and $\mathbb{E}[X_e^2] = \frac{1}{n}$, so

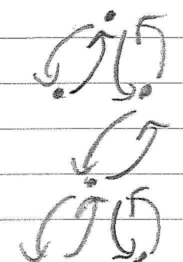
$$\prod_{e \in \mathcal{E}} \mathbb{E}[X_e^{k(e)}] = \frac{1}{n^{k/2}}.$$

To count the number of such walks: Call two walks $(i_1, \dots, i_k)$ and $(i_1', \dots, i_k')$ equivalent if some permutation of the labels $\{1, \dots, n\}$ maps one to the other. This divides the set of such walks into equivalence classes, where walks in the same class have the same "shape".

- Within each equivalence class, the walk is specified by the labels of the visited vertices, in the order they are visited. Thus there are exactly $n(n-1)(n-2) \cdots (n - \frac{k}{2})$ walks per class.
- Since the traversed graph is a tree, to specify an equivalence

class, we must specify for each of the k steps of the walk whether (a) we visit a new vertex, or (b) we return to its parent. This is in 1-to-1 correspondence with Dyck words of length k, where (a) $\leftrightarrow$ (and (b) $\leftrightarrow$):

 $\iff$ (()())()

 $\iff$ ()((()()))

The number of equivalence classes is exactly the Catalan number $C_{k/2}$. So

$$\mathbb{E}\Big[\frac{1}{n}\sum_{\text{Case ③ walks}} X_{i_1 i_2}\cdots X_{i_k i_1}\Big] = \frac{1}{n}\cdot C_{k/2}\cdot n(n-1)\cdots\left(n-\tfrac{k}{2}\right)\cdot\frac{1}{n^{k/2}}.$$

Putting the three cases together:

$$\mathbb{E}\Big[\frac{1}{n}\mathrm{Tr}\,X^k\Big] = \underbrace{\mathbb{E}\Big[\frac{1}{n}\sum_{\text{Case ① walks}}\Big]}_{=0} + \underbrace{\mathbb{E}\Big[\frac{1}{n}\sum_{\text{Case ② walks}}\Big]}_{\to 0 \text{ as } n\to\infty.} + \underbrace{\mathbb{E}\Big[\frac{1}{n}\sum_{\text{Case ③ walks}}\Big]}_{\to C_{k/2} \text{ for even } k.}$$

Thus for every fixed k,

$$\mathbb{E}\Big[\int x^k\, d\mu_n(x)\Big] = \mathbb{E}\Big[\frac{1}{n}\mathrm{Tr}\,X^k\Big] \to m_k = \int x^k\, d\mu(x).$$

<u>Problem 1.</u> Let $X \in \mathbb{R}^{n\times n}$ be symmetric, with $X_{ij} = \frac{1}{\sqrt{n}} Z_{ij}$ where:

- $Z_{ij} = Z_{ji}$, and $\{Z_{ij}\}_{i\le j}$ are independent
- $\mathbb{E}[Z_{ij}] = 0$ and $\mathbb{E}[|Z_{ij}|^k] < C_k$ for all i,j and some constants C_k.
- There is a bounded, continuous function
$$s : [0,1]\times[0,1] \to [0,\infty)$$
satisfying $s(x,y) = s(y,x)$, so that
$$\mathbb{E}[Z_{ij}^2] = s\left(\tfrac{i}{n}, \tfrac{j}{n}\right).$$

Let μ_n be the ESD of X.

(a) Suppose, for each fixed y, that $\int s(x,y)\,dx = 1$. Show for every $k \ge 0$ that
$$\mathbb{E}\Big[\int x^k\, d\mu_n(x)\Big] \to \int x^k\, d\mu(x)$$
where μ is still the semi-circle law.

(b) More generally, for a Dyck word w, consider all outermost parentheses and let $w_1, \dots, w_k$ be the Dyck words contained in those parentheses. E.g.,

$$w = (\underbrace{()()}_{w_1})(\underbrace{()}_{w_2})$$

$$w = (\underbrace{\ }_{w_1})(\underbrace{()}_{w_2})(\underbrace{()(())}_{w_3})$$

Define a function $\Phi(w,x)$ recursively by:

$$\Phi(w,x) = \begin{cases} 1 & w = \phi \\ \prod_{i=1}^{k} \int s(x,y)\, \Phi(w_i, y)\, dy & \text{otherwise.} \end{cases}$$

Show for every even $k \geq 0$, that

$$\mathbb{E}\left[\int x^k \, d\mu_n(x)\right] \to \sum_{\substack{\text{Dyck words } w \\ \text{of length } k}} \int \Phi(w,x)\, dx,$$

and that the limit is 0 for odd k.