Lecture Notes from the 1st CCDS Summer School on Random Matrix Theory

Arijit Chakrabarty, PhD

Indian Statistical Institute, Kolkata, India
Scribes:
Rakibul Hasan Rajib
Tahmid Hasan Fuad
Rakibul Islam Mithu

June 30 - July 4, 2024

Chapter 1

Basics of Random Matrices

Ginibre Ensemble: A random matrix is a matrix whose entries are random variables. Let $\{X_{ij};\ \ i,j \in \mathbb{N}\}$ be a collection of i.i.d. standard normal random variables. Let G_N be an $N \times N$ matrix with

$$G_N(i,j) = X_{ij}, \qquad 1 \le i,j \le N$$

This random matrix is called a Ginibre ensemble.

Wigner Matrix: Define W_N by

$$W_N(i,j) = X_{i\wedge j, i\vee j} \qquad 1 \le i,j \le N$$

W_N is called Wigner matrix. The Wigner matrix is Hermitian while Ginibre ensemble is not. The upper triangle entries of the Wigner matrix will be i.i.d.

$$\begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{12} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{1n} & X_{2n} & \cdots & X_{nn} \end{bmatrix}$$

Defn. Given $\mu \in \mathbb{R}^p$ and a $p \times p$ **n.n.d**(non-negative definite) matrix Σ, we say a p-variate random vector X, follows $N_p(\mu, \Sigma)$ if $\forall \lambda \in \mathbb{R}^p$

$$\lambda^\top X \sim N\left(\lambda^\top \mu, \lambda^\top \Sigma \lambda\right)$$

Convention. Elements of $\mathbb{R}^p$ are to be thought of as a $p \times 1$ vectors.

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{12} & x_{22} & \cdots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n} & x_{2n} & \cdots & x_{nn} \end{bmatrix}$$

Wishart Matrix: Suppose $X_1, X_2 \ldots, X_n$ are i.i.d from $N_p(\mu, \Sigma)$. Then $\hat{\Sigma} = \frac{1}{n}\sum_{i=1}^{n}\left(X_i - \mu\right)\left(X_i - \mu\right)^\top$ is an estimator of Σ. The matrix $\hat{\Sigma}$ is called the Wishart matrix

Defn. Suppose $\mu, \mu_1, \mu_2, \ldots$ are probability measures on $\mathbb{R}$. We say $\mu_n \Rightarrow \mu$, that is, μ_n converges to μ weakly, if

$$\lim_{n\to\infty} \int f d\mu_n = \int f d\mu$$

for every bounded continuous function $f : \mathbb{R} \to \mathbb{R}$

Defn. Given any probability measure ν on $\mathbb{R}$, there exists a random variable X such that,

$$P(X \in A) = \nu(A) \text{ for all } A.$$

We shall say "X has distribution ν".

Fact. If X has distribution ν, then

$$E[f(x)] = \int f dv = \int f(x)\nu(dx)$$

For random variables $X_1, X_2, \ldots, X$,

$$X_n \Rightarrow X \text{ simply means}$$

$$\lim_{n\to\infty} E\left[f\left(x_n\right)\right] = E[f(x)]$$

for any bounded continuous $f : \mathbb{R} \to \mathbb{R}$.

Fact. (Method of Moment) For Random variables (RV_s) $X, X_1, X_i, \ldots$ having finite moments, assume

$$\lim_{n\to\infty} E\left[X_n^k\right] = E\left[X^k\right], \quad \forall k \in \mathbb{N}.$$

Then $X_n \Rightarrow X$ only if the moments "determine" the distribution X.

Fact. Suppose $\nu, \nu_1, \nu_2 \ldots$ are probability measures with finite moments such that

$$\lim_{n\to\infty} \int x^k \nu_n(dx) = \int x^k \nu(dx), \quad k \in N.$$

Furthermore, assume ν is determined by its moments. Then $\nu_n \Rightarrow \nu, n \to \infty$.
A measure ν is determined by its moments if whenever

$$\int x^k \nu(dx) = \int x^k \mu(dx) \quad \forall k = 1, 2, \ldots \text{ then}$$

$$\nu = \mu.$$

Fact. (Carleman's condition) Suppose $\{m_k\}_{k=1}^{\infty}$ is the moment sequence of a probability measure μ. If

$$\sum_{k=1}^{\infty} m_{2k}^{-1/2k} = \infty$$

then $\{m_k\}$ determines μ.

Fact. If μ is a probability measure such that

$$\int e^{tx} \mu(dx) < \infty \text{ for all } t \in (-1, 1)$$

for some $\varepsilon > 0$, then μ has finite moments which determines μ. [mgf is finite in the neighborhood of μ]

Corollary. If μ is a compactly supported probability measure, then μ is determined by its moments.

Exercise. Show that the standard normal distribution is determined by its moments.

Exercise. (Needs Gamma integrals) Show that for $k = 1, 2, 3, \ldots$

$$\int_{-\infty}^{\infty} x^k \frac{e^{-x^2/2}}{\sqrt{2\pi}} dx = \begin{cases} \frac{k!}{2^{k/2}(k/2)!}, & \text{if } k \text{ is even} \\ 0, & \text{if } k \text{ is odd.} \end{cases}$$

Example, for $k = 4$.

$$\begin{aligned} &\int_{-\infty}^{\infty} x^4 e^{-x^2/2} dx \\ &= 2\int_0^{\infty} x^4 e^{-x^2/2} dx \\ &= \underbrace{2\int_0^{\infty} (2y)^{3/2} e^{-y} dy}_{\text{Gamma integral}} \end{aligned}$$

Let, $y = x^2/2$
$\therefore dy = xdx$
Again, $2y = x^2$
$\therefore (2y)^{3/2} = x^3$

Central limit theorem: Suppose $X_1, X_2, \ldots$ are i.i.d. zero mean RVs with finite variance σ^2. Then as $n \to \infty$

$$\frac{1}{\sqrt{n}}\left(X_1 + X_2 + \cdots + X_n\right) \Rightarrow Z, \text{ where } Z \sim N\left(0, \sigma^2\right)$$

Proof: (under the additional assumption that all moments of X_1 are finite) Let, $S_n = X_1 + X_2 + \ldots + X_n$ clearly, $E\left[S_n\right] = 0$ and $E\left[S_n^2\right] = Var\left[S_n\right] = \sum_{i=1}^{n} Var\left(X_i\right) = n$ Since $X_1, X_2, X_3, \cdots, X_n$ are i.i.d. RVs (Without loss of generality and $\sigma^2 = 1$) We want to compute,

$$\begin{aligned} E\left[S_n^4\right] &= E\left[\left(\sum_{i=1}^{n} X_i\right)^4\right] \\ &= E\left[\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n} X_i X_j X_k X_l\right] \\ &= \sum_{i,j,k,l} E\left(X_i X_j X_k X_l\right) \end{aligned} \tag{1.1}$$

If i, j, k, l are distinct, then

$$E\left(X_i X_j X_k X_l\right) = E\left[X_i\right] E\left[X_j\right] E\left[X_k\right] E\left[X_i\right] = 0$$

In fact, whenever one of i, j, k, l is "isolated", that is, it is distinct from the other three,

$$E\left[X_i X_j X_k X_l\right] = 0$$

In other words, $E\left[X_i X_j X_k X_l\right] = 0$ unless one of the following holds
(I) $i = j = k = l$ (II) $(i = j) \neq (k = l)$
(III) $(i = k) \neq (j = l)$ (IV) $(i = l) \neq (j = k)$

Continuing from 1.1, we write

$$\begin{aligned} E\left(S_n^4\right) &= nE\left(X_1^4\right) + 3n(n-1) \\ E\left[\left(\frac{S_n^4}{\sqrt{n}}\right)^4\right] &= \frac{1}{n^2} E\left[S_n^4\right] \to 3 \text{ an } n \to \infty \end{aligned}$$

To generalize: Let k be a positive even integer. As before,

$$\begin{aligned} E\left[S_n^k\right] &= E\left[\left(\sum_{i=1}^{n} X_i\right)^k\right] \\ &= E\left[\sum_{i_1,\ldots,i_k=1}^{n} \left(X_{i_1} \ldots X_{i_k}\right)\right] \\ &= \sum_{i_1,i_2,\ldots,i_k=1}^{n} E\left(X_{i_1} X_{i_2} \cdots X_{i_k}\right) \end{aligned}$$

Given, $(i_1, \ldots i_k) \in \{1, \ldots, n\}^k$
$E\left[X_{i_1} \ldots X_{i_k}\right] = 0$ if there is any "isolated" index $i_1, i_2 \ldots, i_k$
That is there exists a partition $P_1, P_2, \ldots, P_l$ of $\{1, \ldots, k\}$ such that

$$\begin{aligned} \#P_j \geqslant 2\, (\#P_j \text{ means cardinality of } P_j) \\ P_1 \cup P_2 \cup \ldots \cup P_l = \{1, 2, \ldots, k\} \text{ and } P_1, P_2, \ldots, P_l \text{ are disjoint} \\ i_u = i_v \Leftrightarrow u, v \in P_j \text{ for some j} \end{aligned} \tag{1.2}$$

Thus,

$$\begin{aligned} E\left(S_n^k\right) &= \sum_{P_1,\ldots,P_l} \sum_{\substack{(i_1,\ldots,i_k)\in\{1,\ldots,n\}^k \\ \text{such that } (**)\text{holds}}} E\left[X_1, \ldots X_k\right] \\ &= \sum n(n-1)\cdots(n-l+1) E\left(X_1^{\#P_1}\right) E\left(X_1^{\#P_2}\right) \cdots E\left(X_1^{\#P_l}\right) \end{aligned}$$

Given the partition $P_1, \ldots, P_l$ of $\{1, \ldots, k\}$ with

$$\#P_j \geqslant 2, \quad l \leq k/2$$

Equally holds if and only if $\#P_j = 2$ that is $(P_1, P_2, \cdots, P_l)$ is a pairing of $\{1, 2, \cdots, k\}$
Thus,

$$\begin{aligned} E\left[S_n^k\right] &= \sum_{\substack{P_1, P_2 \ldots, P_{k/2} \\ \text{is a pairing of } \{1, \ldots, k\}}} n(n-1)\cdots(n-k/2+1) + O\left(n^{k/2}\right) \\ &= n(n-1)\cdots(n-k/2+1)\frac{k!}{2^{k/2}(k/2)!} + 0\left(n^{k/2}\right) \end{aligned}$$

Therefore,

$$\lim_{n\to\infty} n^{-k/2} E\left[S_n^k\right] = \frac{k!}{2^{k/2}(k/2)!} + O\left(n^{k/2}\right) \text{ for an even k}$$

Note: It is easier to show that

$$\lim_{n\to\infty} n^{-k/2} E\left[S_n^k\right] = 0 \text{ if k is odd}$$

Therefore, we showed that,

$$\lim_{n\to\infty} E\left[\left(\frac{S_n}{\sqrt{n}}\right)^k\right] = E\left(Z^k\right)$$

for $k = 1, 2, \ldots$, where Z follows standard normal distribution. The method of moment completes the proof.

For an Hermitian matrix A of size $N \times N$ enumerate its eigenvalues in the ascending order by $\lambda_1(A), \ldots, \lambda_N(A)$.
Defn. For an $N \times N$ random matrix W, define its "empirical spectral distribution" or ESD_W by the measure

$$ESD_W(A) = \frac{1}{N}\sum_{i=1}^{N} \delta_{\lambda_i(W)}(A) \text{ for all } A \subseteq \mathbb{R}$$

Here $\delta_x(A) = \begin{cases} 1, & \text{if } x \in A \\ 0, & \text{if } x \neq A \end{cases}$
In other word,

$$\begin{aligned} ESD_W(A) &= \frac{1}{N}\sum 1\left(\lambda_i(W) \in A\right) \\ &= \frac{1}{N}\#\{i : 1 \leq i \leq N,\ \lambda_i(W) \in A\} \end{aligned}$$

Defn. The expected empirical spectral distribution or $EESD$ of W is

$$\begin{aligned} EESD_w(A) &= E\left(ESD_w(A)\right) \\ &= E\left(\frac{1}{N}\sum_{i=1}^{N} 1\left(\lambda_i(w) \in A\right)\right) \\ &= \frac{1}{N}\sum_{i=1}^{N} P(\lambda_i(w) \in A) \end{aligned}$$

In other words, $EESD_W(A)$ is nothing but the average of the distributions of $\lambda_1(w) \ldots, \lambda_n(w)$
In measure theory language,

$$\int f(x) EESD_w(dx) = \frac{1}{N}\sum_{i=1}^{N} E[f(\lambda_i(w))]$$

Let, $\{X_{ij} : 1 \leq i \leq j\}$ be i.i.d RVs with all moments finite. Define a Wigner matrix W_N by

$$W_N(i,j) = \begin{cases} X_{ij}, & \text{if } i \leq j \\ X_{ji}, & \text{if } i > j \end{cases}$$

Our goal is to use the Method of Moment for studying $EESD_{W_N}$
The first moment of $EESD_{W_N}$

$$\begin{aligned}\int_{-\infty}^{\infty} xEESD_{W_N} &= \frac{1}{N}\sum_{i=1}^{N} E[\lambda_i(W_N)] \\ &= \frac{1}{N}E\left(\sum_{i=1}^{N}(\lambda_i(W_N))\right) \\ &= \frac{1}{N}E\left[Tr(W_N)\right] \\ &= \frac{1}{N}E\left(\sum_{i=1}^{N}W_N(i,i)\right) \\ &= \frac{1}{N}E\left(\sum_{i=1}^{N}X_{ii}\right) = 0.\end{aligned}$$

The second moment of $EESD_{W_N}$

$$\begin{aligned}&\int_{-\infty}^{\infty} x^2 EESD_{W_N}(dx) \\ &= \frac{1}{N}E\left(\sum_{i=1}^{N}\lambda_i^2(W_N)\right) \\ &= \frac{1}{N}E\left(\sum_{i=1}^{N}\lambda_i(W_N^2)\right) \\ &= \frac{1}{N}E\left[Tr(W_N^2)\right] \\ &= \frac{1}{N}E\left[\sum_{i=1}^{N}\sum_{j=1}^{N}(W_N(i,j))^2\right] \\ &= \frac{N^2}{N}\sigma^2 = N\sigma^2\end{aligned}$$

where $\sigma^2 = Var(X_{ij}) = E(X_{ij})^2$

$$\begin{aligned}&Tr(W_N^2) \\ &= \sum_{i=1}^{N} W_N^2(i,i) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N} W_N(i,j)W_N(j,i) \\ &= \sum_{i=1}^{N}\sum_{j=1}^{N}(W_N(i,j))^2\end{aligned}$$

As $N\sigma^2$ blows up, we need to scale to get a limit. To get a "finite limit", we scale W_N by $\sqrt{N}$. Look at,

$$\begin{aligned}ESD_{\frac{W_N}{\sqrt{N}}} &= \frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i(\frac{W_N}{\sqrt{N}})} \\ &= \frac{1}{N}\sum_{i=1}^{N}\delta_{\frac{\lambda_i(W_N)}{\sqrt{N}}} \\ \text{and,}\quad EESD_{\frac{W_N}{\sqrt{N}}} &= \int_{-\infty}^{\infty} x^2 ESD_{\frac{W_N}{\sqrt{N}}}(dx) \\ &= \frac{1}{N}\sum_{i=1}^{N}(\frac{\lambda_i}{\sqrt{N}}(W_N))^2 \\ &= \frac{1}{N^2}\sum_{i=1}^{N}\lambda_i^2(W_N)\end{aligned}$$

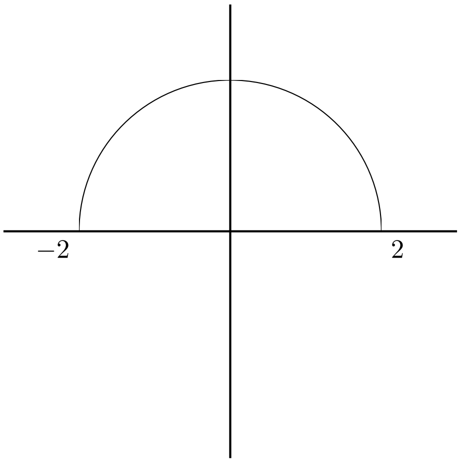
Exercise. Check that,

$$\int_{-\infty}^{\infty} x^2 EESD_{\frac{W_N}{\sqrt{N}}}(dx) = \sigma^2$$

Theorem: (Wigner's Surmise) As $N \to \infty$, $EESD_{\frac{W_N}{\sqrt{N}}} \Rightarrow \mu_{sc}$
where μ_{sc} is the probability measure, whose density is

$$f(x) = \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2}, & -2 \le x \le 2 \\ 0, & \text{Otherwise} \end{cases}$$

Often μ_{sc} is called the semi-circle distribution.



Fourth Moment:
If P is an $N \times N$ matrix, then

$$P^k(i,j) = \sum_{i_1,i_2,\cdots,i_{k-1}=1}^{N} P(i,i_1)P(i_1,i_2)\cdots P(i_{k-1},j)$$

$$\begin{aligned}
\therefore \int_{-\infty}^{\infty} x^4 EESD_{\frac{W_N}{\sqrt{N}}}(dx) &= \frac{1}{N}\sum_{i=1}^{N} E\left[\lambda_i^4\left(\frac{W_N}{\sqrt{N}}\right)\right] \\
&= \frac{1}{N^3}\sum_{i=1}^{N} E\left[\lambda_i^4\left(W_N\right)\right] \\
&= \frac{1}{N^3}\sum_{i=1}^{N} E\left[\lambda_i\left(W_N^4\right)\right] \\
&= \frac{1}{N^3} E\left[Tr\left(W_N^4\right)\right] \\
&= \frac{1}{N^3} E\left(\sum_{i=1}^{N} W_N^4(i,i)\right) \\
&= \frac{1}{N^3} E\left(\sum_{i=1}^{N}\sum_{j=1}^{N}\sum_{k=1}^{N}\sum_{l=1}^{N} W_N(i,j)W_N(j,k)W_N(k,l)W_N(l,i)\right) \\
&= \frac{1}{N^3}\sum_{i,j,k,l=1}^{N}\left(E\left[W_N(i,j)W_N(j,k)W_N(k,l)W_N(l,i)\right]\right) \qquad \left[\text{Here,} \quad W_N(i,j) = X_{\underbrace{i\wedge j}_{min},\underbrace{i\vee j}_{max}}\right] \\
&= \frac{1}{N^3}\sum_{i,j,k,l=1}^{N}\left(E\left[X_{i\wedge j,i\vee j}X_{j\wedge k,j\vee k}X_{k\wedge l,k\vee l}X_{l\wedge i,l\vee i}\right]\right) \\
&= 0 \text{ if one of the } i,j,k,l \text{ is isolated}
\end{aligned}$$

(From the experiment in Central Limit Theory) We know, we need to consider pairing. That is one of the following must hold:

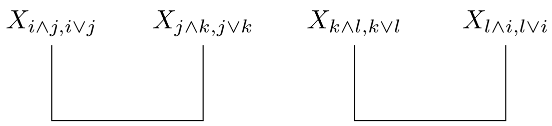
Case1: $\{i,j\} = \{j,k\}$ and $\{k,l\} = \{l,i\}$
Putting $i = l$ ensures both constraints(non-crossing). Approximately $O(N^3)$ many (i,j,k,l) satisfy this.

Case2: $\{i,j\} = \{k,l\}$ and $\{j,k\} = \{l,i\}$
At most $O(N^2)$ choices.

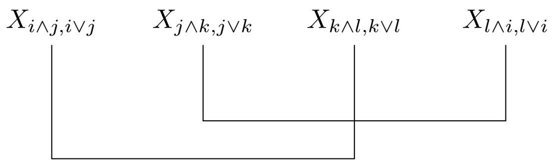
Case3: $\{i,j\} = \{l,i\}$ and $\{j,k\} = \{k,l\}$
Since $j = l$ satisfies both constraints, there are $O(N^3)$ choices.
Therefore ,

$$\lim_{n\to\infty} \frac{1}{N^3} E\left[Tr(W_N^4)\right] = 2$$

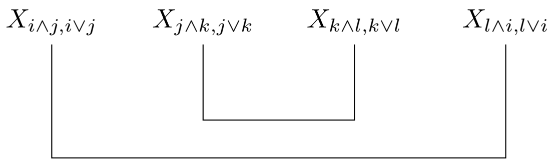
Case 1: ( ) [ ] $\rightarrow$ valid



Case 2: ([ )] $\rightarrow$ not valid



Case 3: ([ ]) $\rightarrow$ valid



1.1 Supplementary Material

Gamma and Beta Integral
Defn: For $\alpha > 0$, $\Gamma(\alpha) = \int_0^\infty e^{-x}x^{\alpha-1}dx$ (Euler's Gamma function)

$$\beta(a,b) = \int_0^1 x^{a-1}(1-x)^{b-1}dx \text{ where } a > 0, b > 0 \quad \text{(Beta function)}$$

Theorem: For $\alpha > 0, \Gamma(\alpha+1) = \alpha\Gamma(\alpha)$
Proof: Integration by parts to get

$$\begin{aligned}\Gamma(\alpha+1) &= \int_0^\infty e^{-x}x^\alpha dx \\ &= (e^{-x})x^\alpha|_0^\infty - \int_0^\infty (-e^{-x})\alpha x^{\alpha-1}dx \\ &= \alpha\int_0^\infty e^{-x}x^{\alpha-1}dx \\ &= \alpha\Gamma(\alpha)\end{aligned}$$

Since $\Gamma(1) = 1$, we get

$$\begin{aligned}&\Gamma(2) = 1.\Gamma(1) = 1 \\ &\Gamma(3) = 2.\Gamma(2) = 2.1 = 2 \\ &\vdots \\ &\Gamma(n+1) = n! \text{ where } n \in \mathbb{R}\end{aligned}$$

Exercise: Calculate $\Gamma(\frac{1}{2})$
Work:

$$\begin{aligned}\Gamma(\frac{1}{2}) &= \int_0^\infty e^{-x}x^{\frac{1}{2}-1}dx & \text{let, } x = \frac{y^2}{2} \\ &= \int_0^\infty e^{-y^2/2}\left(\frac{y^2}{2}\right)^{-1/2} ydy & \Rightarrow dx = ydy \\ &= \sqrt{2}\int_0^\infty e^{-y^2/2}dy = \sqrt{2}\cdot\frac{1}{2}\sqrt{2\pi} \\ &= \sqrt{\pi}\end{aligned}$$

Exercise: Calculate $\Gamma\left(\frac{2k+1}{2}\right)$ for $k \in \mathbb{N}$
Soln: Write $\frac{2k+1}{2} = \frac{2k-1}{2} + 1$

$$\begin{aligned}\Gamma(\frac{2k+1}{2}) &= \frac{2k-1}{2}\Gamma(\frac{2k-1}{2}) \\ &= \frac{2k-1}{2}.\frac{2k-3}{2}\cdots\frac{1}{2}\Gamma(\frac{1}{2}) \\ &= \frac{2k-1}{2}.\frac{2k-3}{2}\cdots\frac{1}{2}.\sqrt{\pi} \\ &= \frac{(2k)!}{2^k.(2.4\cdots.2k)}.\sqrt{\pi} \\ &= \frac{(2k)!}{4^k.k!}.\sqrt{\pi}\end{aligned}$$

Exercise: Calculate the even moments of standard normal.
Soln: Fix $k \in \mathbb{R}$. Then,

$$\begin{aligned}
E[X^{2k}] &= \int_{-\infty}^{\infty} x^{2k} \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx \\
&= \sqrt{\frac{2}{\pi}} \int_0^{\infty} e^{-\frac{x^2}{2}} x^{2k} dx && \text{let, } y = \frac{x^2}{2} \\
&= \sqrt{\frac{2}{\pi}} \int_0^{\infty} e^{-y} (2y)^{\frac{2k-1}{2}} dy && \Rightarrow dy = xdx \\
&= \frac{2^k}{\sqrt{\pi}} \int_0^{\infty} e^{-y} y^{\frac{2k+1}{2}-1} dy \\
&= \frac{2^k}{\sqrt{\pi}} \Gamma\left(\frac{2k+1}{2}\right) \\
&= \frac{2^k}{\sqrt{\pi}} \frac{(2k)!}{4^k k!} \sqrt{\pi} \\
&= \frac{(2k)!}{2^k k!}
\end{aligned}$$

Thus, the $2k$-th moment of the standard normal is $\frac{(2k)!}{2^k (k!)}$

Fact: For $a > 0$ and $b > 0$

$$B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$$

Exercise: Calculate the even moments of the semicircle law (Note: Odd moments vanish)
Soln: For $k \in \mathbb{N}$

$$\begin{aligned}
E\left[X^{2k}\right] &= \frac{1}{2\pi} \int_{-2}^{2} x^{2k} \sqrt{4-x^2} dx && \text{let, } x^2 = 4y \\
&= \frac{1}{\pi} \int_0^1 \left(2y^{\frac{1}{2}}\right)^{2k-1} \sqrt{4-4y} \cdot 2dy && \Rightarrow 2xdx = 4dy \\
&= \frac{2^{2k+1}}{\pi} \int_0^1 y^{\frac{2k-1}{2}} (1-y)^{1/2} dy && \Rightarrow xdx = 2dy \\
&= \frac{2^{2k+1}}{\pi} \cdot B(\frac{2k+1}{2}, \frac{3}{2}) \\
&= \frac{2^{2k+1}}{\pi} \cdot \frac{\Gamma\left(\frac{2k+1}{2}\right)\Gamma(\frac{3}{2})}{\Gamma(k+2)} \\
&= \frac{2^{2k+1}}{\pi} \cdot \frac{\frac{(2k)!}{k!4^k} \cdot \sqrt{\pi} \cdot \frac{1}{2}\sqrt{\pi}}{(k+1)!} \\
&= \frac{(2k)!}{k!(k+1)!}
\end{aligned}$$

Chapter 2

Wigner's Semicircle Law

Theorem: (Wick's formula) If $(G_1, \ldots, G_w\ )$ are $N_k(\mathcal{O}, \Sigma)$, then

$$E\left(G_1, \ldots, G_k\right) = \begin{cases} \sum_{\pi \in G_k} \prod_{(u,v)\in\pi} E(G_u G_v)\textbf{, if } k \text{ is even} \\ 0, \text{ if } k \text{ odd .} \end{cases}$$

For any even number $k, P(k)$ denotes the set of pair partitions of $\{1, \ldots, k\}$
For example, for $k = 4$,

$$P(4) = \{\{(1,2),(3,4)\},\{(1,3),(2,4)\},\{(1,4),(2,3)\}\}$$

Convention: Any element of $P(2k)$ will be denoted by

$$\{(u_1, v_1), \ldots, (u_k, v_k)\} \text{ where } u_1 < \ldots < u_k \text{ and } u_j < v_j \text{ for } j = 1 \ldots k$$

Proof: Denote $G = (G_1, \ldots, G_k)$. Let $Z^{(1)}, Z^{(2)}, \ldots Z_k$ be i.i.d. copies of G. We know Gaussians are symmetric. Symmetry implies,

$$(-G_1, \ldots, -G_k) \stackrel{d}{=} (G_1, \ldots, G_k)$$

$$\text{if k is odd, } -G_1 \ldots G_k \stackrel{d}{=} G_1 \ldots, G_k$$

$$E(G_1, \ldots, G_k) = 0$$

Now assume WLOG; $k = 2m$ for any $m \geq 1$.
Properties of multivariate normal (sum of i.i.d. normal is normal) imply,

$$n^{-1/2}\left(Z^{(1)} + Z^{(2)} + \cdots + Z^{(n)}\right) \stackrel{d}{=} G \text{ for all } n \geq 1$$

Fix n. The above implies,

$$\prod_{j=1}^{2m} G_j \stackrel{d}{=} \prod_{j=1}^{2m} n^{-1/2} \sum_{i=1}^{n} Z_j^{(i)} \qquad Z^{(1)} = (Z_1^{(1)}, \ldots, Z_k^{(1)})$$

$$= n^{-m} \prod_{j=1}^{2m} \sum_{i=1}^{n} Z_j^{(i)}$$

$$= n^{-m} \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_{2m}=1}^{n} \prod_{j=1}^{2m} Z_j^{(i_j)}$$

$$= n^{-m} \sum_{f:\{1,\ldots,2m\}\to\{1,\ldots,n\}} \prod_{j=1}^{2m} Z_j^{(f(j))}$$

Thus,

$$E\left(\prod_{j=1}^{2m} G_j\right) = n^{-m} \sum_{f:\{1,\ldots,2n)\to\{1,\ldots,n\}} E\left(\prod_{j=1}^{2m} Z_j^{(f(j))}\right)$$

Recall, $Z_i^{(f(i))}$ and $Z_j^{(f(j))}$ are independent if $f(i) \neq f(j)$.
Suppose, $i \in Range(f)$, if $\#\{j : f(j) = i\}$ is odd then,

$$E\left(\prod_{j=1}^{2m} z_j^{f(j)}\right) = 0$$

(we proved the for the case k = odd)
Therefore,

$$E\left(\prod_{j=1}^{2m} G_j\right) = n^{-m} \sum_{\substack{f:\{1,\ldots,2m\}\to\{1,\ldots,n\} \\ \text{such that } \#\{j:f(j)=i\} \text{ is even for all } i}} E\left(\prod_{j=1}^{2m} Z_j^{(f(j))}\right)$$

$\therefore$ If f satisfies the above, then

$$\#\ \text{Range } (f) \le m$$

The number of functions $f : \{1, \ldots, 2m\} \to \{1, \ldots, n\}$ with $\#Range(f) \le m-1$ is $O(n^{m-1})$
As $n \to \infty$,

$$\begin{aligned}
\therefore E\left(\prod_{j=1}^{2m} G_j\right) &= O(1) + n^{-m} \sum_{\substack{f:\{1,\ldots,2m\}\to\{1,\ldots,n\} \\ \text{such that } \#\{j:f(j)=i\} \text{ is even for all } i}} E\left(\prod_{j=1}^{2m} Z_j^{(f(j))}\right) \\
&= O(1) + n^{-m} \sum_{\pi\in P(2m)} \sum_{\substack{f:\{1,\ldots,2m\}\to\{1,\ldots,n\} \\ \text{such that } f(u)=f(v) when (u,v)\in\pi}} E\left(\prod_{j=1}^{2m} Z_j^{(f(j))}\right) \\
&= O(1) + n^{-m} \sum_{\pi\in P(2m)} n(n-1)\ldots(n-m+1) \prod_{(u,v)\in\pi} E(G_u G_v)
\end{aligned}$$

As $n \to \infty$,

$$E\left(\prod_{j=1}^{2m} G_j\right) = \sum_{\pi\in P(2m)} \prod_{(u,v)\in\pi} E(G_u G_v)$$

This completes the proof of Wick's formula. ■

Defn: Suppose X and Y are iid from $N(0, \frac{1}{2})$. Define $Z = X + iY$ where $i = \sqrt{-1}$ then Z is said to follow standard CN (complex normal distribution).

Exercise: Calculate $E(Z)$, $E(Z^2)$ and $E(|Z|^2)$.
Soln:

$$\begin{aligned}
&E(Z) = 0 \text{ and } E\left(Z^2\right) = E\left(X^2 - Y^2 + 2iXY\right) = 0 \\
&E\left(|Z|^2\right) = E\left(X^2 + Y^2\right) = 1
\end{aligned}$$

Defn: Let ($Z_{ij} : 1 \le i \le j$) be iid RV_s from standard CN. Define a matrix W_N by

$$W_N(i,j) = \begin{cases} Z_{ij}, & \text{if } i < j \\ \bar{Z}_{ij}, & \text{if } i > j \text{ here, } \bar{Z} = \text{complex conjugate of } Z \\ \sqrt{2}R\left(Z_{ii}\right), & \text{if } i = j \end{cases}$$

Then the random matrix W_d is called a Gaussian Orthogonal Ensemble (GOE).
Exercise: Check that W_N is Hermitian, that is $W_N = W_N^*$. In particular, eigen values of W_N, are real.

Exercise: For $1 \le i, j, k, l \le N$, show that

$$E\left(W_N(i,j)W_N(k,l)\right) = \begin{cases} 1, & \text{if } i = l \text{ and } j = k \\ 0, & \text{otherwise} \end{cases}$$

Denote

$$\delta(u,v) = \begin{cases} 1, & u = v \\ 0, & \text{otherwise.} \end{cases}$$

Therefore,

$$E\left(W_N(i,j)W_N(k,l)\right) = \delta(i,l)\delta(j,k).$$

For fixed $k = 1, 2, \ldots$,

$$\int_{-\infty}^{N} x^k \,\text{EESD}_{\frac{W_N}{\sqrt{N}}}(dx) = \frac{1}{N}\sum_{i=1}^{N} E\left[\lambda_i^k\left(\frac{W_N}{\sqrt{N}}\right)\right]$$

$$= \frac{1}{N^{1+k/2}}\sum_{i=1}^{N} E\left(\lambda_i^k\left(W_N\right)\right)$$

$$= \frac{1}{N^{1+k/2}} E\left(\text{Tr}\left(W_N^k\right)\right)$$

$$= \frac{1}{N^{1+k/2}}\sum_{i_1,i_2,\ldots,i_k=1}^{N} \underbrace{W_N\left(i_1,i_2\right)\cdots W_N\left(i_{k-1},i_k\right)W_N\left(i_k,i_1\right)}_{k\text{ times.}}$$

$$= \frac{1}{N^{1+k/2}}\sum_{i_1,\ldots,i_k=1}^{N} E\left(W_N\left(i_1,i_2\right)\ldots W_N\left(i_k,i_1\right)\right)$$

If k is odd, then this is zero. Assume k is even positive number, then Wick's formula implies that the above equals,

$$\frac{1}{N^{1+k/2}}\sum_{i_1,i_2,\ldots i_k=1}^{N}\sum_{\pi\in P(k)}\prod_{(u,v)\in\pi} E\left(W_N\left(i_u,i_{u+1}\right)W_N(i_v,i_{v+1})\right).$$

For the moment, fix $\pi \in P(k)$. Then,

$$\prod_{(u,v)\in\pi} E\left(W_N\left(i_u,i_{u+1}\right)W_N\left(i_v,i_{v+1}\right)\right) = \prod_{(u,v)\in\pi}\delta(i_u,i_{v+1})\delta(i_{u+1},i_v)$$

Denote, $k = 2m$ and $\pi = \{(u,v_1),\ldots,(u_m,v_m)\}$ following the convention laid down in the beginning. Although π is a pair partition, it can be thought of a function from $\{1,\ldots,2m\} \to \{1,\ldots,2m\}$ with

$$\pi(x) = \begin{cases} v_j, & \text{if } x = u_j \text{ for some } j \\ u_j, & \text{if } x = v_j \text{ for some } j \end{cases}$$

Define $\gamma : \{1,..,2m\} \to \{1,\ldots,2n\}$ by

$$\gamma(j) = \begin{cases} j+1, & \text{if } j \neq 2m \\ 1, & \text{if } j = 2m \end{cases}$$

Thus for $(u,v) \in \pi$

$$\delta\left(i_u,i_{v+1}\right) = \delta\left(i_u,i_{\gamma\pi(u)}\right)$$
$$\text{and } \delta\left(i_{u+1},i_v\right) = \delta\left(i_v,i_{\gamma\pi(v)}\right)$$

Hence,

$$\prod_{j=1}\delta(i_u,i_{v+1})\delta(i_{u+1},i_v) = \prod_{j=1}^{2m}\delta(i_j,i_{\gamma\pi(j)}) = \begin{cases} 1, \text{ if } i_j = \gamma\pi(j), \forall j \\ 0, otherwise \end{cases}$$

Recall that,

$$\sum_{i_1,i_2,\cdots i_k=1}^{N}\prod_{j=1}^{2m}\delta(i_j,i_{\gamma\pi(j)}) \tag{2.1}$$

Exercise: Show that, any permutation is the composition of disjoint cycles.
Suppose, $\gamma\pi = \{s_1,\ldots,s_m\}$ where $S_1,\ldots,s_m$ are disjoint cycles.
Equation (2.1) holds if and only if, $i_u = i_v$ for all $u, v \in S_j$
If $\#(\gamma\pi)$ denotes the numbers of cycles in $\delta\pi$ then

$$\sum_{i_1,\ldots,i_{2m}=1}^{N}\prod_{(u,v)\in\pi}\delta\left(i_u,i_{v+1}\right)\delta\left(i_{u+1},i_v\right) = N^{\#(\gamma\pi)}$$

In this exercise $\#(\gamma\pi) = m$. Thus,

$$\int_{-\infty}^{\infty} x^{2m}\,\text{EESD}_{\frac{W_N}{\sqrt{N}}}(dx) = \sum_{\pi\in P(2n)} N^{\#(\gamma\pi)-1-m}$$

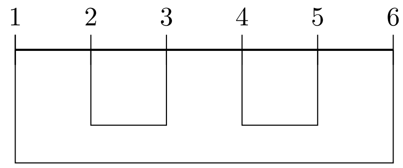
We prove the following theorem, Genus Expansion for $m, N \geqslant 1$. Now as $N \to \infty$ what happens?
Theorem: For all $\pi \in P(2\text{ m})$,

$$\#(\gamma\pi) \leq m+1 \tag{2.2}$$

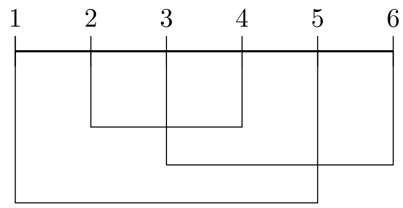
Equality holds if and only if π is a non-crossing pair partition, that is, there do not exist

$$u < v < \omega < z \text{ with } (u,w),(v,z) \in \pi$$

Example of non-crossing pair partition.



Example of crossing pair partition.



Lemma: Suppose, $\pi = \{(u,v_1),\ldots,(u_m,v_m)\}$and $\{w_1,\ldots,w_m\}$ is a cycle of $\gamma\pi$. If,

$$w_1 = \min_{1\leq j\leq m} w_j \tag{2.3}$$

Then, $W_1 \in \{1, u_1+1,\ldots,u_m+1\}$. Thus 2.2 holds. [number of cycles can't exceed $m+1$]

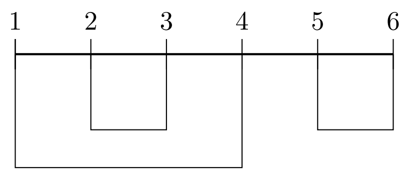
Trivially, $w_1 = \gamma\pi(w_m)$. There are 2 cases which are:
Case 1: $w_m = u_j$ for some j
Case 2: $w_m = v_j$ for some j

In case 1:

$$w_1 = \gamma_\pi(w_m) = \gamma(v_j) = \begin{cases} v_{j+1}, \text{ if } v_j \neq 2m \\ 1, \text{ if } v_j = 2m \end{cases}$$

That, $w_1 = v_{j+1}$ is impossible because then 2.3 would be violated. Thus in this case, necessarily $v_j = 2m$ and hence $w_1 = 1$

In case 2: $w_1 = \gamma\pi(w_m) = \gamma\pi(v_j) = \gamma(u_j) = u_{j+1}$
Thus the claim of the lemma holds.



At least one of the numbers is paired to the next number.

$$\gamma\pi(3) = \gamma(2) \Rightarrow \text{singleton cycle}$$

→ removing one non-crossing pair we get again another non-crossing pair partition.
→ recursively keep removing pairs.

Lemma: Suppose, $\pi \in P(2m)$ and there exists $x \in \{2, 3, \dots, 2m\}$ such that, $(x-1, x) \in \pi$ (pairing of consecutive numbers) Then $\{x\} \to$ singleton cycle in $\gamma\pi$
Furthermore, $\pi' = \pi - \{(x-i, x)\}$
π' is a pairing of $\underbrace{\{1, \dots, 2m\} \backslash \{x-1, x\}}$ and γ' is the cyclic permutation of defined in the obvious way, then

$$\#\,(\gamma'\pi') = \#\,(\gamma\pi) - 1$$

Proof of the theorem: $\#(\gamma\pi) \leq m+1$ has been established. Suppose, π is a non-crossing pair partition. Then there exists $x \in \{2, \dots, 2m\}$ such that $(x-1, x) \in \pi$. By the previous lemma, removal of $(x-1, x)$ from π means, we lose one cycle from $\gamma\pi$. Recursively by deleting $(m-1)$ pairs and hence losing $(m-1)$ cycles, we end up with $\{1, 2\}$. This pair partition pre-multiplied with γ has 2 cycles. This shows,

$$\#(\gamma\pi) = m+1 \qquad \begin{aligned} \gamma\pi(x) &= x \\ \pi(x) &= x-1 \end{aligned}$$

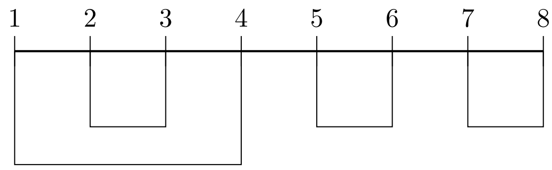
For the converse, assume $\#(\gamma\pi) = m+1$. Assume for the sake of contradiction that π is a crossing-pair partition. If two consecutive elements in π, they can be deleted using the previous lemma at the expense of one cycle in $\gamma\pi$. Inductively, we eventually get $\pi' \in P(2k)$ for some k with $\#(\gamma'\pi') = k+1$ such that there does not exist any $x \in \{2, \dots, 2k\}$ with $(x-1, x) \in \pi'$ Since, $\#(\gamma'\pi') = k+1$, at least two of them are singleton. Say, $\{y\}$ and $\{z\}$. Then, let $x = y \vee z$. Thus $-2 \leq x \leq 2$ and since $\{x\}$ is a cycle in $\gamma'\pi'$, it follows that $(x-1, x) \in \pi$.
This contradiction proves that π is a non-crossing partition.■

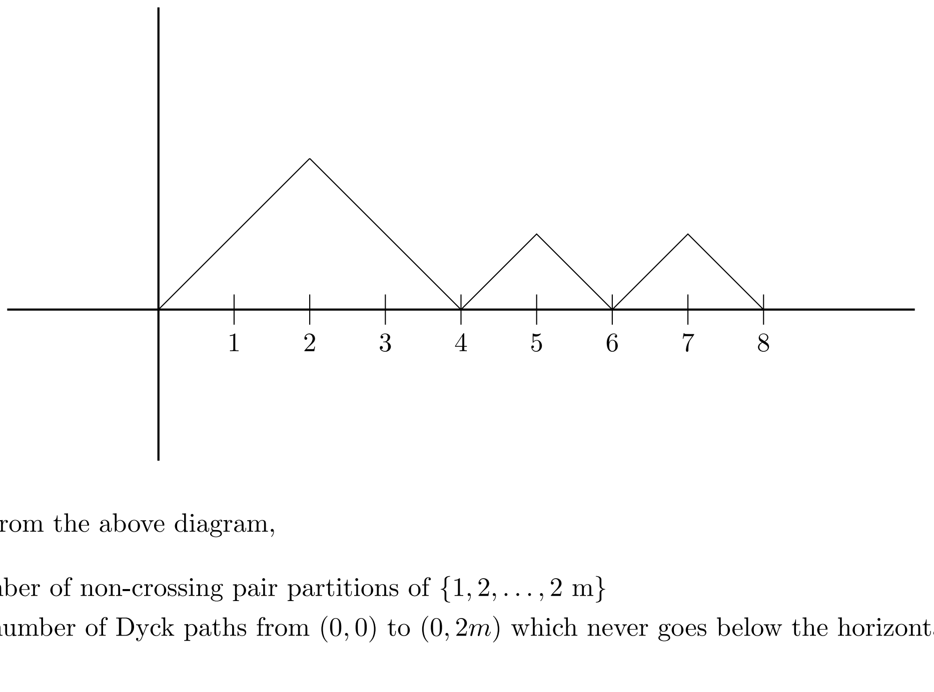
Combining this theorem with Genus expansion, we get

$$\sum N^{\#(8\pi)-m-1)}$$

$$\lim_{N\to\infty} \int x^{2m}\, \mathrm{ESD}_{\frac{W_N}{\sqrt{N}}}(dx) = \#\{\text{number of non-crossing pairing of } \{1, \dots, 2m\}\}$$

Lemma: The number of non-crossing pairings of $\{1, 2, \dots, 2\text{ m}\}$ equals $C_m = \frac{(2m)!}{m!(m+1)!}$. We call C_m the m-th Catalan number.
Proof:





As evident from the above diagram,

the number of non-crossing pair partitions of $\{1, 2, \dots, 2\text{ m}\}$ (2.4)
$=$ the number of Dyck paths from $(0, 0)$ to $(0, 2m)$ which never goes below the horizontal axis.

The 'Reflection principle' implies that

$$\begin{aligned}
&\#\text{Dyck paths from } (0,0) \text{ to } (2n,0) \text{ that touch } -1\\
&= \#\text{ of Dyck paths from } (0,-2) \text{ to } (2m,0)\\
&= \binom{2m}{m-1}.\\
\text{(by (2.4) becomes)} &= \binom{2m}{m} - \binom{2m}{m-1}\\
&= \frac{(2m)!}{m!m!} - \frac{(2m)!}{(m-1)!(m+1)!}\\
&= \quad (m+1-m)\frac{(2m)!}{m!(m+1)}\\
&= \left(\frac{(2m)!}{m!(m+1)!}\right)
\end{aligned}$$

Exercise: Prove this by induction on m.

From yesterday's lecture,

$$\int_{-2}^{2} x^{2m} \frac{1}{2\pi}\sqrt{4-\pi^2}dx = \frac{(2m)!}{m!(m+1)!}$$

Everything put together imply,

$$\lim_{N\to\infty} \int_{-\infty}^{\infty} x^k \,\text{EESD}_{\frac{W_N}{\sqrt{N}}}(dx) = \int_{-\infty}^{\infty} x^k \mu_{sc}(dx) \quad \text{ for all } k.$$

Since the semicircle law is compactly supported, it is determined by its moments. The method of moment proves the following:
As $N \to \infty$,

$$EED_{\frac{W_N}{\sqrt{N}}} \Rightarrow \mu_{sc}$$

where W_N is the $N \times N$, GOE(Gaussian Orthogonal Ensemble)

Universality:
If W_N is of Wigner matrix with iid entries from a zero mean unit variance distribution, then (2.4) holds

Chapter 3

Wishart Matrices

Theorem:
For $z \in \mathbb{C}^+$

$$S_{ESD_A}(z) = \frac{1}{N} T_\gamma \left((A - zI_N)^{-1} \right)$$

Let, $X_{n_1}, \ldots, X_{n_m}$ be iid RVs from $N_{p_1}(0, I_{p_n})$. The subscript "n" will be suppressed.
Define $W_N = \frac{1}{n} \sum_{i=1}^{n} X_i X_i^\top$ is a $p \times p$ Wishart Matrix (Sample equivalent matrix).

Fix $z \in \mathbb{C}^+$. Then

$$\begin{aligned}
S_{ESD_N^N}(z) &= \frac{1}{p} \operatorname{Tr} \left[(\omega_N - z\Psi_P)^{-1} \right] \\
&= \frac{1}{p} \operatorname{Tr} \left[\left(\frac{1}{n} \sum X_i X_i^T - zI_p \right)^{-1} \right] \\
&= \frac{n}{p} \operatorname{Tr} \left[\left(\sum X_i X_i^\top - nzI_P \right)^{-1} \right] \\
&= \frac{n}{p} \operatorname{Tr} \left[\left(\sum_{i=1}^{n-1} X_i X_i^\top - nzI_p + X_n X_n^\top \right)^{-1} \right]
\end{aligned}$$

Denote,

$$\begin{aligned}
B &= \left(\sum_{i=p}^{n} X_p X_p^T - nzI_p \right) \\
A &= \sum_{i=1}^{n-1} X_i X_i^T - nzI_p \\
\therefore B &= A + X_n X_n^\top \\
Thus, \quad S_{ESD_N}(E) &= \frac{n}{p} Tr \left(B^{-1} \right).
\end{aligned}$$

If C is any $n \times n$ invertible matrix, then

$$
\begin{aligned}
&\left(c + xy^\top\right) = \left(I + xy^\top \bar{c}^1\right) c \\
&\therefore (c+xy)^{-1} = c^{-1}\left(I + xy^\top c^{-1}\right)^{-1} \\
&= c^{-1}\left(I - xy^\top c^{-1} + \left(xy^\top c^{-1}\right)^2 - \left(xy^\top c^{-1}\right)^3 + \cdots\right) \\
&= c^{-1}(I - xy^\top c^{-1} + x\underbrace{y^\top c^{-1}}_{\text{scalar}} x\sqrt{c^{-1}} + \cdots) \\
&= c^{-1}\left(I - xy^\top c^{-1} + \left(y^\top c^{-1}x\right)\left(xy^\top c^{-1}\right) - \left(y^\top c^{-1}x\right)^2 xy^\top c^{-1} + \cdots\right) \\
&= \bar{c}^{-1}\left(I - xy^\top \bar{c}^{-1}\left(1 - \left(y^\top c^{-1}x\right) + \left(y^\top c^{-1}x\right)^2 + \cdots\right)\right) \\
&= c^{-1}\left(I - xy^\top \bar{c}^{-1}\frac{1}{1 + y^\top c^{-1}x}\right) \\
&= c^{-1} - \frac{c^{-1}xy^\top c^{-1}}{1 + y^\top c^{-1}x} \qquad \text{(check that this indeed is the inverse)}
\end{aligned}
$$

whenever $y^\top e^{-1}x \neq -1$.

Calculate: $y^\top\left(c + xy^\top\right)^{-1}x = y^\top c^{-1}x - \frac{\left(y^\top c^{-1}x\right)^2}{1+y^\top c^{-1}x} = \frac{y^\top c^{-1}x}{1+y^\top c^{-1}x}$.

Back to the proof:

$$
\begin{aligned}
\therefore X_n^\top B^{-1}X_n &= X_n\left(A + X_nX_n^\top\right)^{-1}X_n \\
&= \frac{X_n^\top A^{-1}X_n}{1 + X_n^\top A^{-1}X_n} \\
&\approx \frac{E\left(X_n^\top A^{-1}X_n\right)}{1 + E\left(X_1^\top A^{-1}X_n\right)}
\end{aligned}
$$

*We could have,

$$
\begin{aligned}
A_k &= \sum_{i \in \{1,\ldots,n\}} X_iX_i^T - nZI_k \\
X_k^T B^{-1}X_k &\approx \frac{E(X_k^T A_k^{-1}X_k)}{1 + E(X_k^T A_k^{-1}X_k)} \\
X_k^T A_k^{-1}X_k &\stackrel{d}{=} X_1^T A_1^{-1}X_1
\end{aligned}
$$

Suppose, $Z \sim N_p(0, I)$
Assume, Λ is pxp real symmetric matrix

$$
\begin{aligned}
&Z^\top \wedge Z,\ \wedge = PDP^\top \\
&Z^\top \wedge Z = Z^\top PDP^\top Z \\
&\qquad = \left(P^\top Z\right)^\top D\left(P^\top Z\right)
\end{aligned}
$$

Since, $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_p)$

and $P^T Z = \begin{pmatrix} v_1 \\ \vdots \\ v_p \end{pmatrix} \sim N(0, I_p)$

$$
Z^\top \wedge Z = \sum_{i=1}^{p} \lambda_i V_i^2
$$

$$
\begin{aligned}
E\left(Z^\top \wedge Z\right) &= \sum \lambda_i E\left(V_i^2\right) = \sum \lambda_i = \operatorname{Tr}(\Lambda) \\
\operatorname{Var}\left(Z^\top \wedge Z\right) &= 2\Sigma\lambda_i^2 = 2\operatorname{Tr}\left(\Lambda^2\right)
\end{aligned}
$$

$X_0 \therefore$ Taking any $X_k^\top B^{-1}X_k$ for $k = 1, \ldots n$,
we get,

$$\frac{E\left(X_n^\top A^{-1} X_n\right)}{1+E\left(X_n^\top A^{-1} X_n\right)} \approx \frac{1}{n} \sum_{k=1}^{n} X_k^\top B^{-1} X_k.$$

$$\begin{aligned} E\left(X_1^\top A^{-1} X_n\right) &= E_A[E\left(X^\top A^{-1} X_n\right) \mid A] \\ &= E\left[\operatorname{Tr}\left(A^{-1}\right)\right] \end{aligned}$$

We know,

$$\begin{aligned} B &= A + X_n X_n^\top \\ &= \sum_{i=1}^{n} X_i X_i^\top - nzI_p \quad + X_n X_n^\top \\ &= \sum X_i X_i^\top = B + nzI_r \end{aligned}$$

Thus,

$$\begin{aligned} \frac{E\left[\operatorname{Tr}(A^{-1})\right]}{1+E\left[\operatorname{Tr}(A^{-1})\right]} &\approx \frac{1}{\lambda} \sum_{k=1}^{n} X_k^T B^{-1} X_k \\ &= \frac{1}{n} \sum_{k=1}^{n} \operatorname{Tr}\left(X_k^\top B^{-1} X_k\right) \\ &= \frac{1}{n} \sum_{k=1}^{n} \operatorname{Tr}\left(B^{-1} X_k X_k^\top\right) \\ &= \frac{1}{n} \operatorname{Tr}\left(\sum_{k=1}^{n} \left(B^{-1} X_k X_k^\top\right)\right) \\ &= \frac{1}{n} \operatorname{Tr}\left(B^{-1} \sum_{k=1}^{n} X_k X_k^\top\right) \\ &= \frac{1}{n} \operatorname{Tr}\left(B^{-1}\left(B + nzI_8\right)\right) \\ &= \frac{p}{n} + z \operatorname{Tr}\left(B^{-1}\right) \end{aligned}$$

Assume that $\frac{p}{n} \Rightarrow \delta(0,1] \longrightarrow$ how does p grows wrt n

$$\frac{\frac{n}{p} E\left[\operatorname{Tr}(A^{-1})\right]}{\frac{n}{p} + \frac{n}{p} E[\operatorname{Tr}(A^{-1})}] = \frac{p}{n} + Z\frac{p}{n}\frac{n}{p}\operatorname{Tr}(B^{-1})$$

$$\begin{aligned} \frac{S_{W_{n-1}}(Z)}{1/\gamma + S_{W_{n-1}}(Z)} &\approx \gamma + Z\gamma S_{W_n}(Z) \\ \frac{S(Z)}{1/\gamma + S(Z)} &\approx \gamma + Z\gamma S(Z). \end{aligned}$$

$$\Rightarrow \frac{S(Z)}{1+\gamma S(Z)} = 1 + ZS(Z) \quad \Rightarrow 1 + (\gamma + Z)S(Z) + \gamma Z (S(Z))^2 = S(Z)$$

$$\therefore S(Z) = \frac{1-\gamma-Z \pm \sqrt{(\gamma+Z-1)^2 - 4\gamma Z}}{2\gamma Z}$$

Exercise:

$$I_m(S(z)) > 0 \implies S(z) = \frac{1-\gamma-z+\sqrt{(z-\gamma_-)(z-\gamma_+)}}{2z}$$

$$\text{where,} \quad \gamma_- = (1-\sqrt{y})^2$$

$$\gamma_+ = (1+\sqrt{y})^2$$

$\therefore$ For $\quad t \in R.$

$$\lim_{t \downarrow 0}(t+it) = \frac{1-\gamma_{-}t\sqrt{(t-\gamma_{-})\,(t-\gamma_{+})}}{2\gamma t}$$

$$\therefore \lim_{t\downarrow 0} \mathrm{I_m}\left(S(t+it) = \begin{cases} \frac{\sqrt{(t-\gamma_{-})(\gamma_{+}-t)}}{2\gamma t} & \gamma \le t \le \gamma_{+} \\ 0, & \text{otherwise.} \end{cases}\right.$$

$\therefore$ We get, $ESD_{N_N} \Rightarrow$ the distribution with density,

$$f(t) = \frac{1}{2\pi\gamma t}\sqrt{(t-\gamma_{-})\,(\gamma_{+}-t)};\ \gamma_{-1} \le t \le \gamma_{+}$$

For $0 < \gamma \le 1$, the Marchenko-Pastur law is the distribution with density.

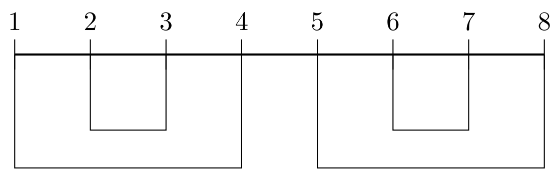
$$\begin{aligned} f(t) = \tfrac{1}{2\pi\gamma_t}\sqrt{(t-\gamma_{-})\,(\gamma_{+}-t)} \\ \gamma_{-} \le t \le \gamma_t \\ \gamma_{-} = (1-\sqrt{\gamma})^2 \\ \gamma_{+} = (1+\sqrt{\gamma})^2. \end{aligned}$$

Stieltjes transform:

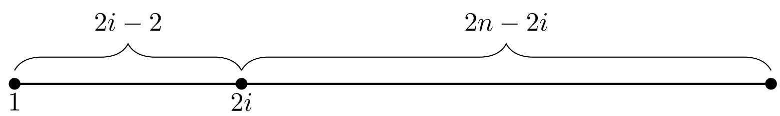
$$\begin{aligned} S_\mu(z) &= \int_R \frac{1}{t-z}\mu(dt) = \int (t-z)^{-1}\mu(dt) \\ &= z^{-1}\int\left(\frac{t}{z}-1\right)^{-1}\mu(dt) \\ &= -z^{-1}\int\left(1-\frac{t}{z}\right)^{-1}\mu(dt) \\ &= -z^{-1}\int\sum_{n=1}^{\infty}\left(\frac{t}{z}\right)^{n}\mu(dt) \\ &= -\sum_{n=0}^{\infty} z^{-n-1}\int_{-\infty}^{\infty} t^n\mu(dt) \end{aligned}$$

the nth moment of μ

Exercise: Obtain a recursive relation for the number of non-crossing pair partitions.
Solution:



Let C_n be the # NCPP of $\{1,\ldots,2n\}$
Suppose, 1 is paired with $2i$ for me $i\{1,\ldots,n\}$



$$C_n = \sum_{i=1}^{n} C_{i-1}C_{n-i}$$

with $C_0 = 1,$

$$C_n = \frac{(2n)!}{(n!)(n+1)!}$$

Exercise:
Show that n left brackets and n right brackets can be arranged in a "legitimate" way in C_n ways.
Solution:
Sliding a counter from the very left, at no points # of left brackets encountered, should not be less than the # of right brackets.. Therefore, there is a one-to-one correspondence between all such arrangements and the Dyck path from $(0,0)$ to $(2n,0)$ that never go below the horizontal line,
Weak Convergence:
Definition: For probability measures $\mu, \mu, \ldots.$. we say $\mu_n \Rightarrow \mu$ if

$$\lim_{n\to\infty} \int f d\mu_n = \int f d\mu$$

for all bounded continuous $f : \mathbb{R} \to \mathbb{R}$

Ques: The CDF of a probability measure μ is

$$F(x) = \mu((-\infty, x]), \quad x \in \mathbb{R}$$

Ques: If $F, F_1, F_2, \cdots$ are $CDFs$ of $\mu, \mu_1, \ldots$ respectively can weak convergence be defined in terms of F_n ?
Ans: Yes. if $\lim_{n\to\infty} F_n(x) = F(x)$ for every x at which F is continuous, then $\mu_n \Rightarrow \mu$.

Helly's selection principle: of $F_1, F_2 \ldots$ are non-decreasing right continuous function, then there exists a subsequm $\{F_{n_u}\}$ of $\{F_n\}$ ad a nonderearing right continuous f sit. $\lim_{n\to\infty} F_{n_k}(x) = f(x)$ for every continuity point x of F.

Levy Continuity theorem: If $\mu, \mu, \ldots$ are probability moments, then $\mu_n \Rightarrow \mu$ iff

$$\lim_{n\to\infty} \phi_n(t) = \phi(t) \text{ for all } t \in \mathbb{R}$$

where $\phi_1(t), \phi_2(t)$ are the characteristic function of $\mu, \mu_1, \mu_2 \cdots$ respectively.

How to prove?
Step 1: The characteristic functions (CHF_s) determine the probability measure (uniqueness).
Step2: $\quad \phi_n(t) = \int e^{itx}\mu(dx)$ the only if part follows from the definition $(\mu_n \Rightarrow \mu) \Rightarrow \phi_n(t) \to \phi$

Proof of "if part"
Assume $\phi_n \to \phi$ is pointwise.
To show $\mu_n \Rightarrow \mu$, it suffice to prove every subsequence of $\{\mu_n\}$ has of further subsequence which converges weakly to μ.
Fix subsequence $\{\mu_{n_k}\}$
Step 3: Use Helly's to get a further subsequence of $\mu_{n_k}, \{\mu_{n_k}\}$ which converges weakly to same probability measure, ν.
Step 4: From step 2, it follows that

$$\phi_{n_{k_l}}(t) \to \phi_r(t), \quad k \to 1$$

Step 5: the assumption (*) ensures

$$\phi \equiv \phi_v$$

It follows, from step 1 , that $\mu = v$.

Chapter 4

Finite Rank Perturbation

Let $\{x_{i,j}, 1 \leq i \leq j\}$ be a collection of iid RVs with mean $\mu > 0$ and variance 1. Construct a Wigner matrix W_N by,

$$W_N(i,j) = \begin{cases} x_{ij}, & \text{if } i \leq j \\ x_{ji}, & \text{if } i > j \end{cases}$$

for all $1 \leq i, j \leq N$.

Q: How does $ESD_{\frac{W_N}{\sqrt{N}}}$ behave for large N?
Ans:

$$E\left(\operatorname{Tr}\left({W_N}^k\right)\right) = \sum_{i_1,\cdots,i_k=1}^{N} E\left[W_N\left(i_1,i_2\right)\cdot \ldots W_N\left(i_k,i_1\right)\right]$$

Proceeding like in the zero mean case is not possible anymore!

Define, $\tilde{W_N} = W_N - \mu 1_N {1_N}^\top$ where, $1_N = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix}_{N\times 1}$

Thus , $\tilde{W_N}$ is a Wigner matrix with zero mean entries.

Q: How to use information about $\tilde{W_N}$ to infer about W_N ?

$$W_N = \widetilde{W_N} + \underbrace{\mu 1_N 1_N^\top}_{\text{Rank } = 1}$$

4.1 Finite-rank Perturbation

Fact: For $N \times N$ Hermitian matrices A and B,

$$\underbrace{\sup_{x\in\mathbb{R}} |F_A(x) - F_B(x)|}_{\|\text{ESD}_A - \text{ESD}_B\|} \leq \frac{1}{N}\operatorname{Rank}(A-B)$$

where F_A and F_B are the CDFs of ESD_A and ESD_B, respectively.
The fact implies,

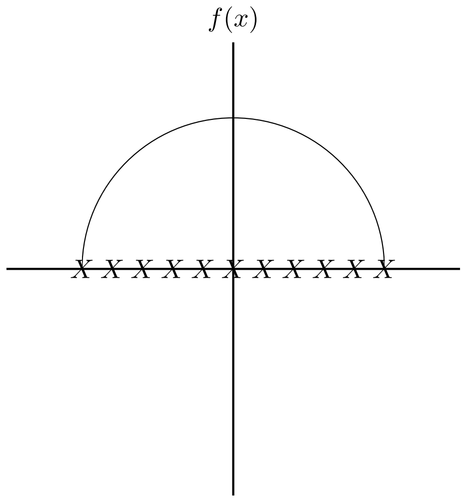
$$\begin{aligned} \left\|\text{ESD}_{W_N} - \text{ESD}_{\tilde{W}_N}\right\| &\leq \frac{1}{N}\operatorname{Rank}\left(W_N - \tilde{W}_N\right) \\ &= \frac{1}{N}\operatorname{Rank}\left(\mu 1_N {1_N}^\top\right) \\ &= \frac{1}{N} \to 0, \quad \text{as} \quad N \to \infty \end{aligned}$$

Since, $\text{ESD}_{\tilde{W}_N} \Rightarrow \mu_{\text{sc}}$ where μ_{sc} is the semicircle law, it follows that,

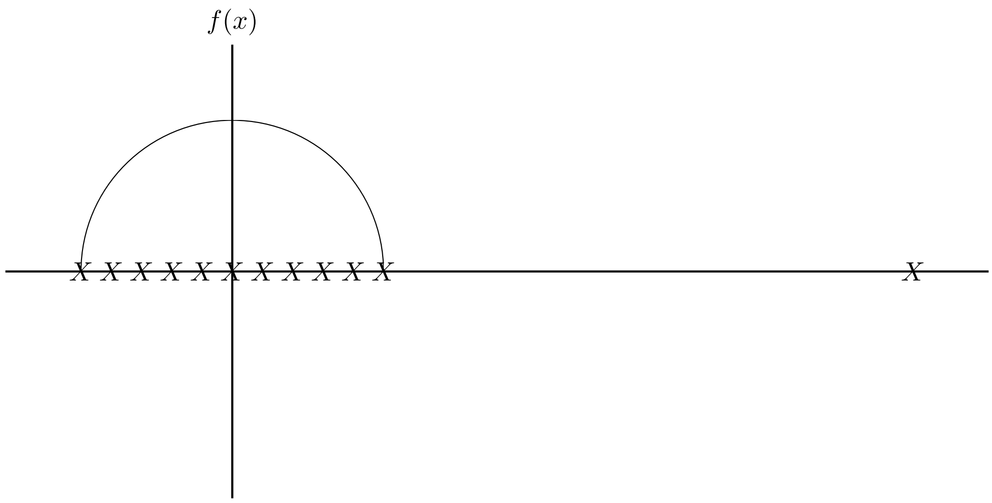
$$\text{E S D}_{W_N} \Rightarrow \mu_{\text{sc}}$$

Q: How does the largest eigenvalue behave?

Ans:



Largest eigenvalue of $\widetilde{W}_N \to 2$



$$W_N = \tilde{W}_N + \mu 1_N 1_N{}^\top$$

Convention: For a Hermitian matrix, denote its eigenvalues by $\lambda_1(A) \geq \lambda_2(A) \geq \cdots \geq \lambda_N(A)$. We want to study $\lambda_1(W_N)$. Recall,

$$W_N = \tilde{W}_N + \mu 1_N 1_N{}^\top$$

The interlacement result implies,

$$\left|\lambda_1(W_N) - \lambda_1\left(\mu 1_N 1_N{}^\top\right)\right| \leq \left\|\tilde{W}_N\right\| \tag{4.1}$$

where for any $N \times N$ Hermitian matrix A,

$$\|A\| = \max_{1 \leq i \leq N} |\lambda_i(A)|$$

Suppose, A is $N \times N$ real symmetric. Then,

$$\lambda_1(A) = \sup_{x \in \mathbb{R}^N : \|x\|=1} x^\top A x$$

$$\begin{aligned} \text{Now, } x^\top A x &= x^\top P D P^\top x \\ &= y^\top D y \qquad (\|y\| = 1) \end{aligned}$$

$$\begin{aligned} \text{Then, } \sup_{x:\|x\|=1} x^\top A x &= \sup_{x \in \mathbb{R}^N : \|x\|=1} x^\top (B + (A - B)) x \\ &\leq \sup x^\top B x + |\sup x^\top (A - B) x| \end{aligned}$$

Therefore,

$$\lambda_1(A) - \lambda_1(B) \leq \|A - B\|$$

Thus (4.1) becomes,

$$|\lambda_1(W_N) - N\mu| \leq \left\|\tilde{W}_N\right\|$$

Dividing throughout by N,

$$\begin{aligned}\left|\frac{\lambda_1(W_N)}{N} - \mu\right| &\leq \frac{1}{N}\left\|\tilde{W}_N\right\| \\ &= \frac{1}{\sqrt{N}} \cdot \frac{\left\|\tilde{W}_N\right\|}{\sqrt{N}} = \frac{2}{\sqrt{N}} \\ \frac{\left\|\tilde{W}_N\right\|}{\sqrt{N}} &\to 2 \\ \therefore \text{As} \quad N \to \infty, \quad &\left|\frac{\lambda_1(W_N)}{N} - \mu\right| \to 0 \\ \Rightarrow \frac{\lambda_1(W_N)}{N} &\to \mu\end{aligned}$$

In other words, the bulk of the eigenvalues of W_N are of the order $\sqrt{N}$, that is,

$$\text{ESD}_{\frac{W_N}{\sqrt{N}}} \Rightarrow \mu_{sc}$$

But the largest eigenvalue is of order N, that is $\frac{\lambda_1(W_N)}{N} \to \mu$ in probability.

Q: How does $\frac{\lambda_1(W_N)}{N}$ fluctuate around μ for large N?
In other words, we want to know if $\left(\frac{\lambda_1(W_N)}{N} - \mu\right)$ can be scaled up to have a non-zero limit.
Ans: Fix N. Let v be the eigenvector of W_N corresponding to the largest eigenvalue of W_n, $\lambda_1(W_N)$ which we will write as λ_1.

That is,

$$\begin{aligned}&W_N v = \lambda_1 v \\ \Rightarrow &\left(\tilde{W}_N + \mu 1_N 1_N{}^\top\right) v = \lambda_1 v \\ \Rightarrow &\mu 1_N \underbrace{\left(1_N{}^\top v\right)}_{\text{scalar}} = \lambda_1 v - \tilde{W}_N v \\ \Rightarrow &\mu\left(1_N{}^\top v\right) 1_N = \left(\lambda_1 I - \tilde{W}_N\right) v \end{aligned} \tag{4.2}$$

Since the eigenvalues of $\tilde{W}_N$ are of the order $\sqrt{N}$ and λ_1 is of order N, $\lambda_1 I_N - \tilde{W}_N$ is invertible with high probability.
The (4.2) implies,

$$v = \mu\left(1_N{}^\top v\right)\left(\lambda_1 I_N - \tilde{W}_N\right)^{-1} 1_N$$

Premultiplying by $1_N{}^\top$, we get

$$\begin{aligned}1_N{}^\top v &= \mu \cdot \left(1_N{}^\top v\right) 1_N{}^\top \left(\lambda_1 I_N - \tilde{W}_N\right)^{-1} 1_N. \\ \Rightarrow 1 &= \mu \cdot 1_N{}^\top \left(\lambda_1 I_N - \tilde{W}_N\right)^{-1} 1_N \\ &= \frac{\mu}{\lambda_1} 1_N{}^\top \left(I_N - \frac{\tilde{W}_N}{\lambda_1}\right)^{-1} 1_N\end{aligned}$$

Thus,

$$\lambda_1 = \mu \cdot 1_N{}^\top \left(1_N - \frac{\tilde{W}_N}{\lambda_1}\right)^{-1} 1_N \tag{4.3}$$

Fact: If $\|A\| < 1$, then $(I-A)^{-1} = \sum_{j=0}^{\infty} A^j$

$$(I-A)\sum_{j=0}^{\infty} A^j = \sum_{j=0}^{\infty} A^j - \sum_{j=1}^{\infty} A^j = I$$

Applying this fact to (4.3), we get,

$$\begin{aligned}\lambda_1 &= \mu \cdot 1_N{}^\top \left(\sum_{j=0}^{\infty} \left(\frac{W_N}{\lambda_1} \right)^j \right) 1_N \\ &= \mu \sum_{j=0}^{\infty} \frac{1_N{}^\top \tilde{W}_N{}^j 1_N}{\lambda_1^j} \\ &= \mu \cdot 1_N{}^\top 1_N + \frac{\mu}{\lambda_1} 1_N{}^\top \tilde{W}_N 1_N + \mu \sum_{j=2}^{\infty} \frac{1_N{}^\top \tilde{W}_N^j 1_N}{\lambda_1^j}\end{aligned}$$

Further,

$$\begin{aligned}1_N{}^\top \tilde{W}_N 1_N &= \sum_{i,j=1}^{N} \tilde{W}_N(i,j) \\ &= \sum_{i=1}^{N} \tilde{X}_{ii} + 2 \sum_{1 \le i < j \le N} \tilde{X}_{ij}\end{aligned}$$

$$\begin{aligned}&\tilde{W_N}(i,j) \\ &= X_{i\wedge j, i \vee j} - \mu \\ &= \tilde{X}_{i\wedge j, i\vee j}\end{aligned}$$

Since $\left\{ \tilde{X}_{ij} : 1 \le i \le j \right\}$ is a collection of iid zero mean RVS, Lindeberg's CLT implies,

$$\frac{1}{N} 1_N{}^T \tilde{W}_N 1_N \Rightarrow N(0,2) \text{ as } N \to \infty$$

Thus, $\frac{\mu}{\lambda_1} 1_N{}^\top \tilde{W}_N 1_N \Rightarrow N(0,2)$ as $N \to \infty$.

$\sum_{j=2}^{\infty} \frac{1_N{}^\top \tilde{W}_N^j 1_N}{\lambda_1^j}$ is concentrated around its expectation.
Thus, $\lambda_1 - E(\lambda_1) \Rightarrow N(0,2)$ as $N \to \infty$
That is, $\lambda_1 (W_N)$ has a Gaussian fluctuation in the limit.

$$\begin{aligned}\lambda_1 &= N\mu + \frac{\mu}{\lambda_1} 1_N{}^\top \tilde{W}_N 1_N + \mu \sum_{j=2}^{\infty} \frac{1_N{}^\top \tilde{W}_N 1_N}{\lambda_1^j} \\ \lambda_1 - E(\lambda_1) &\approx \frac{1}{N} 1_N{}^\top \tilde{W}_N 1_N \Rightarrow N(0,2)\end{aligned}$$

Q: Can the entries of the Wigner-matrix be replaced by independent RVS, having possibly different distributions, with zero mean and variance one, and one still gets the semicircle law in the limit?

CLT: In the CLT, can the summands have a different distribution so that the limit is still normal?

Lindeberg's CLT:
Suppose that for $n = 1, 2, \ldots, n$; $X_{n_1}, \ldots, X_{n_n}$ are independent zero mean RVS with,

$$\lim_{n\to\infty} \sum_{x=1}^{n} \operatorname{Var}(X_{n_i}) = \sigma^2 < \infty.$$

If for all $\varepsilon > 0$,

$$\lim_{n\to\infty} \sum_{i=1}^{n} E\left(X_{n_i}^2 1 \left(|X_{n_i}| > \varepsilon \right) \right) = 0 \quad (\text{ Lindeberg's Condition })$$

then,

$$\sum_{i=1}^{n} x_{n_i} \Rightarrow N\left(0, \sigma^2\right) \text{ as } n \to \infty$$

Usual CLT: Suppose $X_1, X_2, \ldots$ are iid zero mean RVs with variance σ^2. For $n \geq 1$, let

$$X_{n_i} = \frac{1}{\sqrt{n}} X_i, \quad i = 1, \ldots, n$$

It's immediate that $X_{n_1}, \ldots, X_{nn}$ are independent zero mean RVs.
Furthermore,

$$\begin{aligned}\sum_{i=1}^{n} \mathrm{Var}\left(X_{n_i}\right) &= \sum_{i=1}^{n} \mathrm{Var}\left(\frac{X_i}{\sqrt{n}}\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathrm{Var}\left(X_i\right) \\ &= \sigma^2\end{aligned}$$

To check the Lindeberg condition, fix $\varepsilon > 0$ and observe,

$$\begin{aligned}\sum_{i=1}^{n} E\left(X_{n_i}^2 1\left(|X_{n_i}| > 2\right)\right) &= \sum_{i=1}^{n} E\left(\frac{{X_i}^2}{n} 1\left(|X_i| > \sqrt{n}\varepsilon\right)\right) \\ &= E\left(X_i^2 \cdot 1\left(|x_1| > \sqrt{n}\varepsilon\right)\right) \\ &\to 0 \text{ as } n \to \infty \quad (\text{since } X_1 \stackrel{\mathrm{d}}{=} X_2 \stackrel{\mathrm{d}}{=} \cdots \stackrel{\mathrm{d}}{=} X_n)\end{aligned}$$

Now,

$$S_n \stackrel{\mathrm{d}}{\approx} \sum_{i=1}^{n} G_{n_i}$$

Hence,

$$\begin{aligned}S_n &\approx \sum_{i=1}^{d} G_{ni} \sim N\left(0, \sigma_n^2\right) \\ \sigma_n^2 &= \mathrm{Var}\left(S_n\right)\end{aligned}$$

Proof: Let

$$S_n = \sum_{i=1}^{n} X_{n_i}, \quad n \geq 1,$$

$$\sigma_{n_i}^2 = \mathrm{Var}\left(x_{n_i}\right), \quad i = 1, \ldots, n$$

$$\text{and } \sigma_n^2 = \sum_{i=1}^{n} \sigma_i^2$$

Let $G_{ni} : 1 \leq i \leq n$ be a collection of independent RVS, which is independent of X_{n_i} as well, with $G_{ni} \sim N\left(0, \sigma_{ni}^2\right)$.
If we can show that, $S_n \stackrel{\mathrm{d}}{\approx} \sum_{i=1}^{n} G_{ni}$, then the proof would follow because, $\sum_{i=1}^{n} G_{ni} \sim N\left(0, \sigma_n^2\right)$ and $\sigma_n^2 \to \sigma^2$ by assumption.

We shall show that for all bounded function $f : \mathbb{R} \to \mathbb{R}$ which is thrice differentiable and its first three derivatives are bounded,

$$\lim_{n\to\infty} \left| E[f\left(S_n\right)] - E[f\left(\sum_{i=1}^{n} G_{ni}\right)] \right| = 0$$

The above would imply that,

$$\lim_{n\to\infty} E\left[f\left(S_n\right)\right] = E[f(Z)] \quad \text{where } z \sim N\left(0, \sigma^2\right)$$

and hence it would follow that,

$$S_n \Rightarrow z$$

Proof:
Fix $f : \mathbb{R} \to \mathbb{R}$ as above. Taylor's theorem implies,

$$f\left(S_n\right) = f\left(X_{n_1} + \sum_{i=2}^{n} X_{n_i}\right)$$

$$\Rightarrow f\left(S_n\right) = f\left(\sum_{i=2}^{n} X_{n_i}\right) + X_{n_i} f'\left(\sum_{i=2}^{n} X_{n_i}\right) + \frac{1}{2} X_{n_1}^2 f''\left(\sum_{i=2}^{n} X_{n_i}\right) + \frac{1}{3!} X_{n_1}^2 f'''(\xi) \quad \text{for some } \xi$$

Thus,

$$E(f(S_n)) - E(f(\sum_{i=2}^{n} X_{n_i})) = E(X_{n_1} f'(\sum_{i=2}^{n} X_{n_i})) + \frac{1}{2}{\sigma_m}^2 E(f''(\sum_{i=2}^{n} X_{n_i})) + O(E|X_{n_i}|^3)$$

$$\Rightarrow f(S_n) = f(\sum_{i=2}^{n} X_{n_i}) + X_{n_1} f'(\sum_{i=2}^{n} X_{n_i}) + \frac{1}{2}{X_{n_1}}^2 f''(\xi')$$

$$\Rightarrow \left| f(S_n) - f(\sum_{i=2}^{n} X_{n_i}) \right| \leq X_{n_1} f'(\sum_{i=2}^{n} X_{n_i}) + \frac{1}{2}{X_{n_1}}^2 f''(\xi')$$

$$\Rightarrow \left| E\left[f(S_n) - f(\sum_{i=2}^{n} X_{n_i}) - \frac{1}{2}{\sigma_{n_1}}^2 E[f''()] \right] \right| \leq kE\left({X_{n_1}}^2 \wedge |X_{n_1}|^3\right)$$

$$\Rightarrow \left| E\left[f\left(G_{n_1} + \sum_{i=2}^{n} X_{n_i}\right) - f\left(\sum_{i=2}^{n} X_{n_i}\right) \right] \right| \leq kE\left(|G_{n_1}|^3\right) = c{\sigma_{n_1}}^2$$

Combine the two inequalities to get,

$$\left| E\left(f\left(S_n\right)\right) - E\left(f\left(G_m + \sum_{i=2}^{m} X_{n_i}\right)\right) \right| \leq kE\left(X_{n_1}^2 \wedge |X_{n_1}|^3\right) + C\sigma_{n_1}^3$$

Replacing X_{n_i} by G_{n_i}, one at a time, yields,

$$\left| E\left(f\left(S_n\right)\right) - E\left(f\left(\sum_{i=1}^{n} G_{n_i}\right)\right) \right| \leq k\sum_{i=1}^{n}\left(E\left(X_{n_i}^2 \wedge |X_{n_i}|^3\right) + c\sum_{i=1}^{n} \sigma_{n_i}^3\right.$$

Fix $\varepsilon > 0$,

$$\sum_{i=1}^{n} E\left({X_{n_i}}^2 \wedge |X_{n_i}|^3\right) = \sum_{i=1}^{n} E\left({X_{n_i}}^2 \wedge |X_{n_i}|^3\right) 1\left(|X_{n_i}| \leq \varepsilon\right) + \sum_{i=1}^{n} E\left({X_{n_i}}^2 \wedge |X_{n_i}|^3\right) 1\left(|X_{n_i}| > \varepsilon\right)$$

$$\Rightarrow \sum_{i=1}^{n} E\left[{X_{n_i}}^2 \wedge |X_{n_i}|^3 1\left(X_{n_i} \ |\leq \varepsilon\right)\right] \leq \sum_{i=1}^{n} E\underbrace{|X_{n_i}|^3}_{\leq \varepsilon X_{n_i}} 1\left(|X_{n_i}| \leq \varepsilon\right) \leq \varepsilon \sum_{i=1}^{n} E\left(X_{n_i}^2\right) = \varepsilon\sigma_n^2 \to \varepsilon\sigma^2$$

$$\Rightarrow \sum_{i=1}^{n} E\left[\left({X_{n_i}}^2 \wedge |X_{n_i}|^3\right) 1\left(|X_{n_i}| > \varepsilon\right)\right] \leq \sum_{i=1}^{n} \epsilon\left(X_{n_i}^2 + 1\left(|X_{n_i}| > \varepsilon\right)\right) \to 0 \quad \text{(by Lindebarg condition)}$$

To complete the proof, we need to show,

$$\lim_{n\to\infty} \sum_{i=1}^{n} \sigma_{n_i}^3 = 0$$

Fix $\varepsilon > 0$. Then

$$\sum_{i=1}^{n} \sigma_{n_i}^3 = \sum_{i=1}^{n} \sigma_{n_i} \cdot \sigma_{n_i}^2 \leq \left(\max_{1\leq i\leq n} \sigma_{n_i}\right) \underbrace{\sum_{j=1}^{n} {\sigma_{n_j}}^2}_{\sigma^2}$$

If we can show that,

$$\lim_{n\to\infty} \max_{1\leq i\leq n} {\sigma_{n_i}}^2 = 0$$

then we are done.
Proof:
Fix $\varepsilon > 0$,

$$\begin{aligned} \sigma_{n_i}^2 =& E\left({X_{n_i}}^2\right) \\ =& E\left({X_{n_i}}^2 \cdot 1\left(|X_{n_i}| \leq \varepsilon\right)\right) + E\left({X_{n_i}}^2 \cdot 1\left(|X_{n_i}| > \varepsilon\right)\right) \\ \leq& \varepsilon^2 + \epsilon\left({X_{n_i}}^2 1 \cdot \left(|X_{n_i}| > \varepsilon\right)\right) \end{aligned}$$

Therefore,

$$\max_{1\leq 1\leq n} \sigma_{n_i}^2 \leq \varepsilon^2 + \max_{2\leq i\leq n} E\left(X_{n_i} \cdot 1\left(|X_{n_i}| > \varepsilon\right)\right)$$
$$\leq \varepsilon^2 + \underbrace{\sum_{i=1}^{n} E\left(x_{ni}^2 + (|x_{ni}| > \varepsilon)\right)}_{0}$$

The last line follows from Lindeberg condition. Hence we get the complete proof. ■

In Random Matrices, the Lindeberg principle can be applied in a similar way.

$$S(z) = \frac{1}{N} E\left[\mathrm{Tr}\left(W_N - zI_N\right)^{-1}\right]$$

where W_N is an $N \times N$ Wigner matrix (with entries having zero mean & variance one). The entries of W_N can be "replaced" by standard normal RVs, one by one, as Lindeberg CLT.

It can be shown that if (using $[W_N(i,j) = X_{i\wedge j, i\vee j}]$)

$$\lim_{n\to\infty} N^{-2} \sum_{1\leq i\leq j\leq N} E\left(X_{ij}^2 \cdot 1\left(|X_{ij}| > \varepsilon\sqrt{N}\right)\right) = 0 \quad \text{(Pastur's Condition)}$$

then,

$$\frac{1}{N} E\left[\left[\mathrm{Tr}\left((W_N - zI_N)^{-1}\right)\right] - \frac{1}{N} E\left[\mathrm{Tr}\left(\left(\frac{G_N}{\sqrt{N}} - zI_N\right)^{-1}\right]\right.\right. \to 0 \text{ as } \quad N \to \infty$$

where G_N is an $N \times N$ Wigner matrix with entries for standard normal.

Thus it would follow that under Pastur's condition,

$$\mathrm{ESD}_{\frac{W_N}{\sqrt{N}}} \Rightarrow \mu_{\mathrm{sc}}.$$