Saarland University
Faculty of Mathematics and Computer Science
Department of Mathematics

Random matrices

Lecture notes

Summer 2018

Prof. Dr. Roland Speicher
Dr. Marwa Banna

Disclaimer

These notes are being typed up during the lecture and are **UNOFFICIAL**! They make no claim to be comprehensive or correct. If you find any mistakes please send an email:

ricardo.schnur@math.uni-sb.de

Table of contents

0 Introduction

0.1 Some History

- Hurwitz 1897: "Über die Erzeugung der Invarianten durch Integration"

  (origin of random matrices in mathematics according to P. Diaconis, P. Forrester)

- Wishart 1928: random matrices in statistics for fixed size N

- Wigner 1955: random matrices as statistical models for heavy nuclei, studied in particular asymptotics for $N \to \infty$ ("large N limit")

- Marchenko, Pastur 1967: asymptotics $N \to \infty$ of Wishart matrices

- Since 1960's: random matrices are important tools in physics
  $\to$ quantum chaos
  $\to$ universality
  important work by Mehta, Dyson

- 1967: influential (first) book "Random Matrices" by Mehta

- $\sim$ 1972: relation between statistics of eigenvalues of random matrices and zeros of Riemann ζ-function (Montgomery + Dyson, Odlyzko)

- Since 1990's: random matrices are studied more and more extensively in mathematics
  $\to$ Tracy-Widom distribution of largest eigenvalue
  $\to$ free probability theory
  $\to$ universality of fluctuations
  $\to$ "circular law"
  $\to \cdots$

0.2 What is a random matrix?

Random matrix $A = (a_{ij})_{i,j=1}^{N}$, where entries a_{ij} are chosen randomly (often we require A to be selfadjoint).

Example. Choose $a_{ij} \in \{-1, +1\}$ with $a_{ij} = a_{ji}$ for all i, j. Consider all such matrices and ask for typical or generic behaviour.
[In a more probabilistic language: All allowed matrices have the same probability.]

0.3 Quantity of interest

We are mainly interested in the eigenvalues of the matrices. Consider the situation from above with $a_{ij} \in \{-1, +1\}$ for matrices of different size N.

	matrix	eigenvalues	probability
$N = 1$:			
	$A_1 = (1)$	$+1$	$\frac{1}{2}$
	$A_2 = (-1)$	-1	$\frac{1}{2}$
$N = 2$:	$A_1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$	$0, 2$	$\frac{1}{8}$
	$A_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$	$-\sqrt{2}, \sqrt{2}$	$\frac{1}{8}$
	$A_3 = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$	$0, 2$	$\frac{1}{8}$
	$\vdots$	$\vdots$	$\vdots$
	$A_8 = \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix}$	$-2, 0$	$\frac{1}{8}$

General N: We have $2^{N(N+1)/2}$ matrices, each with probability $2^{-N(N+1)/2}$.
There are always special ones such as

$$A = \begin{pmatrix} 1 & \cdots & 1 \\ \vdots & \ddots & \vdots \\ 1 & \cdots & 1 \end{pmatrix}$$

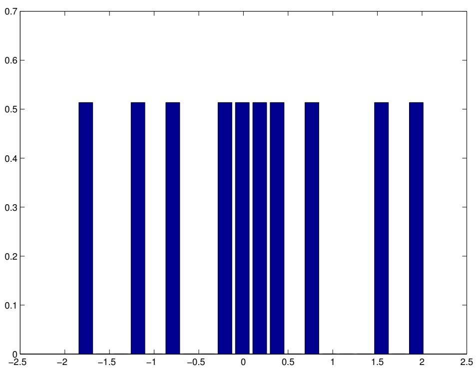
with eigenvalues N and 0 (with multiplicity $N - 1$). They have small probability and are “atypical”.

Question. What is the “typical” behaviour of the eigenvalues?

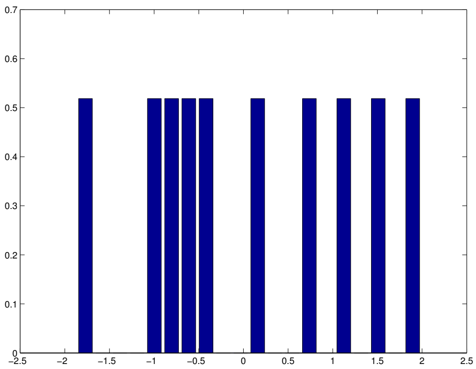
Here is one randomly generated 10×10 matrix

$$\begin{pmatrix} 1 & -1 & -1 & 1 & -1 & 1 & -1 & -1 & -1 & 1 \\ -1 & 1 & -1 & -1 & 1 & 1 & -1 & 1 & 1 & 1 \\ -1 & -1 & 1 & 1 & -1 & 1 & 1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 & -1 & -1 & -1 & 1 \\ -1 & 1 & -1 & 1 & -1 & -1 & -1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 & -1 & 1 & 1 & -1 & 1 & 1 \\ -1 & -1 & 1 & -1 & -1 & 1 & 1 & 1 & -1 & 1 \\ -1 & 1 & 1 & -1 & -1 & -1 & 1 & -1 & -1 & -1 \\ -1 & 1 & -1 & -1 & 1 & 1 & -1 & -1 & 1 & -1 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & -1 & -1 & 1 \end{pmatrix}$$

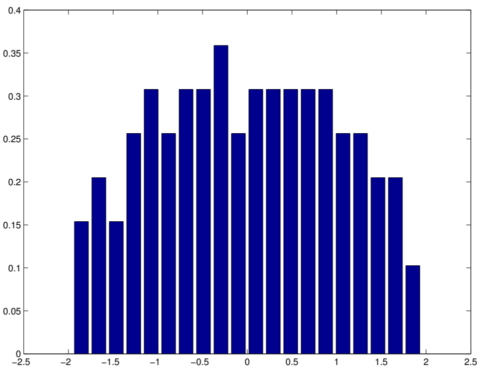
and here is its eigenvalue histogram (with the right scaling, as considered in 0.11)

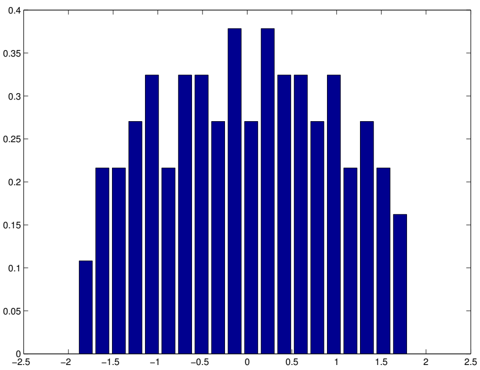


If we create another such 10×10 matrix randomly, then its eigenvalue distribution looks like this

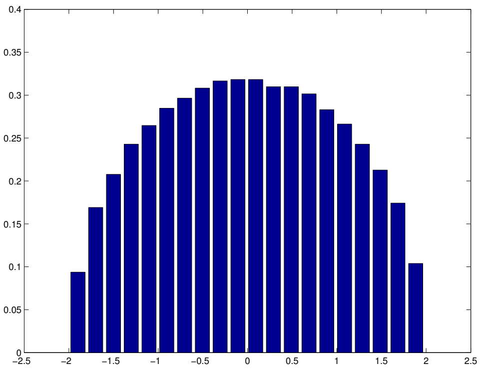


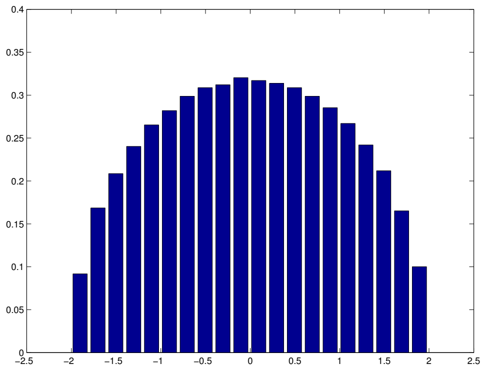
Let us now consider growing size of the matrices. Here are the eigenvalue histograms for two 100×100 matrices ...





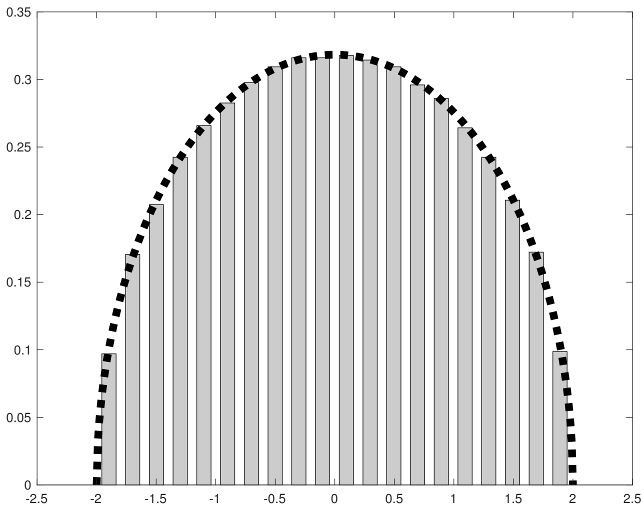
... and here for two 3000×3000 matrices ...





0.4 Wigner's semicircle law

We see that typically the eigenvalue distribution of such a random matrix converges to Wigner's semicircle for $N \to \infty$.



0.5 Universality

This statement is valid much more generally: Choose the a_{ij} not just from $\{-1, +1\}$ but, for example,

- $a_{ij} \in \{1, 2, 3, 4, 5, 6\}$,
- a_{ij} normally (Gauß) distributed,
- a_{ij} distributed according to your favorite distribution,

but still independent (apart from symmetry), then we still have the same result: The eigenvalue distribution still typically converges to a semicircle for $N \to \infty$.

0.6 Concentration phenomena

The (quite amazing) fact that the a priori random eigenvalue distribution is, for $N \to \infty$, not random anymore, but concentrated on one deterministic distribution (namely the semicircle) is an example of the general high-dimensional phenomenon of "measure concentration".

0.7 Example of such a "strange" concentration in high dimensions

In high dimensions the volume of a ball is essentially sitting in the surface:
Denote by $B_r(0)$ the ball of radius r around 0 in $\mathbb{R}^n$ and for $0 < \varepsilon < 1$ consider $B = \{x \in \mathbb{R}^n \mid 1 - \varepsilon \leq \|x\| \leq 1\}$. Then we know that

$$\mathrm{vol}(B_r(0)) = r^n \frac{\pi^{\frac{n}{2}}}{\left(\frac{n}{2} - 1\right)!}$$

and hence

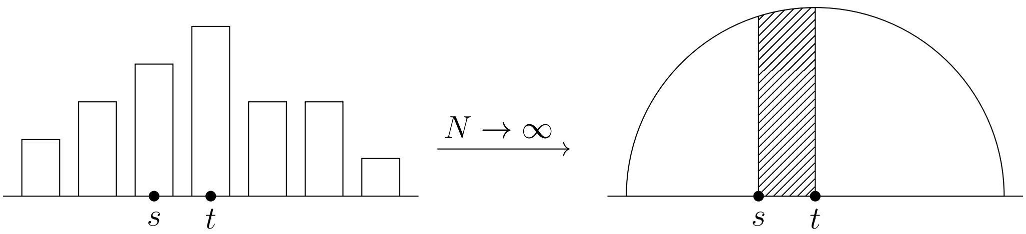
$$\mathrm{vol}(B) = \mathrm{vol}(B_1(0)) - \mathrm{vol}(B_{1-\varepsilon}(0)) = \frac{\pi^{\frac{n}{2}}}{\left(\frac{n}{2} - 1\right)!} \left(1 - (1 - \varepsilon)^n\right).$$

Thus,

$$\frac{\mathrm{vol}(B)}{\mathrm{vol}(B_1(0))} = 1 - (1 - \varepsilon)^n \xrightarrow{n \to \infty} 1.$$

0.8 From histograms to moments

Let $A_N = A = (a_{ij})_{i,j=1}^N$ be our selfadjoint matrix with $a_{ij} = \pm 1$ randomly chosen. Then we typically see for the eigenvalues of A:



This convergence means

$$\frac{\#\{\text{eigenvalues in } [s,t]\}}{N} \xrightarrow{N\to\infty} \int_s^t \mathrm{d}\mu_W = \int_s^t p_W(x)\,\mathrm{d}x,$$

where μ_W is the semicircle distribution, with density p_W.
The left-hand side of this is difficult to calculate directly, but we note that the above statement is the same as

$$\frac{1}{N}\sum_{i=1}^N 1_{[s,t]}(\lambda_i) \xrightarrow{N\to\infty} \int_{\mathbb{R}} 1_{[s,t]}(x)\,\mathrm{d}\mu_W(x), \tag{$\star$}$$

where $\lambda_1,\dots,\lambda_N$ are the eigenvalues of A counted with multiplicity and $1_{[s,t]}$ is the characteristic function of $[s,t]$, i.e,

$$1_{[s,t]}(x) = \begin{cases} 1, & x \in [s,t], \\ 0, & x \notin [s,t]. \end{cases}$$

Hence in ($\star$) we are claiming that

$$\frac{1}{N}\sum_{i=1}^N f(\lambda_i) \xrightarrow{N\to\infty} \int_{\mathbb{R}} f(x)\,\mathrm{d}\mu_W(x)$$

for all $f = 1_{[s,t]}$. It is easier to calculate this for other functions f, in particular, for f of the form $f(x) = x^n$, i.e.,

$$\frac{1}{N}\sum_{i=1}^N \lambda_i^n \xrightarrow{N\to\infty} \int_{\mathbb{R}} x^n\,\mathrm{d}\mu_W(x); \tag{$\star\star$}$$

the latter are the **moments** of μ_W. (Note that μ_W must necessarily be a probability measure.)
We will see later that in our case the validity of ($\star$) for all $f = 1_{[s,t]}$ is equivalent to the validity of ($\star\star$) for all n. Hence we want to show ($\star\star$) for all n.

0.9 What is the advantage of $f(x) = x^n$ over $f = 1_{[s,t]}$?

Note that $A = A^*$ is selfadjoint and hence can be diagonalized, i.e., $A = UDU^*$, where U is unitary and D is diagonal with $d_{ii} = \lambda_i$ for all i. Moreover, we have

$$A^n = (UDU^*)^n = UD^nU^*$$

with

$$D^n = \begin{pmatrix} \lambda_1^n & & \\ & \ddots & \\ & & \lambda_N^n \end{pmatrix},$$

hence

$$\sum_{i=1}^{N} \lambda_i^n = \mathrm{Tr}(D^n) = \mathrm{Tr}(UD^nU^*) = \mathrm{Tr}(A^n)$$

and thus

$$\frac{1}{N}\sum_{i=1}^{N} \lambda_i^n = \frac{1}{N}\,\mathrm{Tr}(A^n).$$

0.10 Notation

We denote by $\mathrm{tr} = \frac{1}{N}\,\mathrm{Tr}$ the **normalized trace** of matrices, i.e.,

$$\mathrm{tr}\left((a_{ij})_{i,j=1}^N\right) = \frac{1}{N}\sum_{i=1}^{N} a_{ii}.$$

So we are claiming that for our matrices we typically have that

$$\mathrm{tr}(A_N^n) \xrightarrow{N\to\infty} \int x^n \,\mathrm{d}\mu_W(x).$$

0.11 Choice of scaling

Note that we need to choose the right scaling in N for the existence of the limit $N \to \infty$. For the case $a_{ij} \in \{\pm 1\}$ with $A_N = A_N^*$ we have

$$\mathrm{tr}(A_N^2) = \frac{1}{N} \sum_{i,j=1}^{N} \underbrace{a_{ij} \underbrace{a_{ji}}_{=a_{ij}}}_{=(\pm 1)^2=1} = \frac{1}{N} N^2 = N.$$

Since this has to converge for $N \to \infty$ we should rescale our matrices

$$A_N \to \frac{1}{\sqrt{N}} A_N,$$

i.e., we consider matrices $A_N = (a_{ij})_{i,j=1}^N$, where $a_{ij} = \pm \frac{1}{\sqrt{N}}$. For this scaling we claim that we typically have that

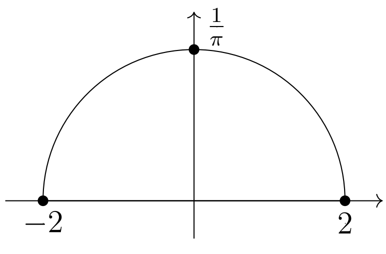
$$\mathrm{tr}(A_N^n) \xrightarrow{N\to\infty} \int x^n \, \mathrm{d}\mu_W(x)$$

for a deterministic probability measure μ_W.

0.12 Definition

(1) The (standard) **semicircular distribution** μ_W is the measure on $[-2, 2]$ with density

$$\mathrm{d}\mu_W(x) = \frac{1}{2\pi} \sqrt{4 - x^2} \, \mathrm{d}x.$$



(2) The **Catalan numbers** $(C_k)_{k\geq 0}$ are given by

$$C_k = \frac{1}{k+1} \binom{2k}{k}.$$

They look like this: $1, 1, 2, 5, 14, 42, 132, \ldots$

0.13 Theorem

Theorem. *(1)* *(i) The Catalan numbers satisfy the following recursion:*

$$C_k = \sum_{l=0}^{k-1} C_l C_{k-l-1} \qquad (k \geq 1)$$

(ii) The Catalan numbers are uniquely determined by this recursion and by the initial value $C_0 = 1$.

(2) The semicircular distribution μ_W is a probability measure, i.e.,

$$\frac{1}{2\pi} \int_{-2}^{2} \sqrt{4 - x^2}\, \mathrm{d}x = 1$$

and its moments are given by

$$\frac{1}{2\pi} \int_{-2}^{2} x^n \sqrt{4 - x^2}\, \mathrm{d}x = \begin{cases} 0, & n \text{ odd}, \\ C_k, & n = 2k \text{ even}. \end{cases}$$

0.14 Type of convergence

So we are still claiming that typically

$$\begin{aligned} \mathrm{tr}(A_N^2) &\to 1, \\ \mathrm{tr}(A_N^4) &\to 2, \\ \mathrm{tr}(A_N^6) &\to 5, \\ \mathrm{tr}(A_N^8) &\to 14, \\ \mathrm{tr}(A_N^{10}) &\to 42, \end{aligned}$$

and so forth. But what do we mean by “typically”? The mathematical expression for this is “almost surely”, but for now let us look on the more intuitive “convergence in probability” for

$$\mathrm{tr}(A_N^{2k}) \to C_k.$$

Denote by Ω_N the set of our considered matrices, that is

$$\Omega_N = \left\{ A_N = \frac{1}{\sqrt{N}} (a_{ij})_{i,j=1}^N \,\middle|\, A_N = A_N^* \text{ and } a_{ij} \in \{\pm 1\} \right\}.$$

Then convergence in probability means that for all $\varepsilon > 0$ we have

$$\frac{\#\left\{A_N \in \Omega_N \,\middle|\, \left|\operatorname{tr}\left(A_N^{2k}\right) - C_k\right| > \varepsilon\right\}}{\#\Omega_N} = P\left(A_N \,\middle|\, \left|\operatorname{tr}\left(A_N^{2k}\right) - C_k\right| > \varepsilon\right) \xrightarrow{N\to\infty} 0. \quad (\star)$$

How can we show $(\star)$?

(1) First show the weaker form of convergence in average, i.e.,

$$\frac{\sum_{A_N\in\Omega_N} \operatorname{tr}\left(A_N^{2k}\right)}{\#\Omega_N} = \mathbb{E}\left[\operatorname{tr}\left(A_N^{2k}\right)\right] \xrightarrow{N\to\infty} C_k.$$

(2) Show that with high probability the derivation from the average will become small as $N \to \infty$.

We will first consider step (1); (2) is a concentration phenomenon and will be treated later.

0.15 Remark

Note that

$$\frac{1}{N}\sum_{i=1}^{N} f(\lambda_i) \xrightarrow{N\to\infty} \int_{\mathbb{R}} f(x)\,\mathrm{d}\mu_W(x) \qquad (\star)$$

is actually a statement on convergence of measures, since

$$\frac{1}{N}\sum_{i=1}^{N} f(\lambda_i) = \int_{\mathbb{R}} f(x)\,\mathrm{d}\mu_N(x)$$

for the **empirical spectral measure** or **density of states**

$$\mu_N = \frac{1}{N}\left(\delta_{\lambda_1} + \cdots + \delta_{\lambda_N}\right),$$

where δ_λ is the **Dirac measure**

$$\delta_\lambda(E) = \begin{cases} 0, & \lambda \notin E, \\ 1, & \lambda \in E. \end{cases}$$

Hence ($\star$) says that

$$\int_{\mathbb{R}} f(x)\,\mathrm{d}\mu_N(x) \xrightarrow{N\to\infty} \int_{\mathbb{R}} f(x)\,\mathrm{d}\mu_W(x).$$

If we require this for sufficiently many f, this is a kind of convergence $\mu_N \to \mu_W$ of measures. We will need to understand such convergence better and develop tools (Cauchy or Stieltjes transform) to deal with them.

1 The Stieltjes transform

1.1 Weak and vague convergence

Denote by $\mathcal{C}_0(\mathbb{R})$ the set of continuous functions on $\mathbb{R}$ vanishing at zero, that is

$$\mathcal{C}_0(\mathbb{R}) = \left\{ f \in \mathcal{C}(\mathbb{R}) \,\middle|\, \lim_{|x|\to\infty} f(x) = 0 \right\}.$$

Also, denote by $\mathcal{C}_b(\mathbb{R})$ the set of continuous, bounded functions on $\mathbb{R}$.

Definition 1.1. (a) We say that μ_n **converges vaguely** to μ if

$$\int f(x)\,\mathrm{d}\mu_n(x) \xrightarrow{n\to\infty} \int f(x)\,\mathrm{d}\mu(x)$$

for every $f \in \mathcal{C}_0(\mathbb{R})$. We write $\mu_n \xrightarrow{v} \mu$.

(b) We say that μ_n **converges weakly** to μ if

$$\int f(x)\,\mathrm{d}\mu_n(x) \xrightarrow{n\to\infty} \int f(x)\,\mathrm{d}\mu(x)$$

for every $f \in \mathcal{C}_b(\mathbb{R})$. We write $\mu_n \xrightarrow{w} \mu$.

Some properties:

(1) If μ_n is a probability measure and $\mu_n \xrightarrow{w} \mu$ then μ is also a probability measure. Indeed,

$$1 = \int_R 1\,\mathrm{d}\mu_n(x) \xrightarrow{n\to\infty} \int_R 1\,\mathrm{d}\mu(x) = \mu(\mathbb{R}).$$

(2) If μ_N is a probability measure and $\mu_n \xrightarrow{v} \mu$ then μ is not necessarily a probability measure. Examples:

(i) For $\mu_n = \delta_n$ and $f \in \mathcal{C}_0(\mathbb{R})$ we have

$$\int f(x)\,\mathrm{d}\delta_n(x) = f(n) \xrightarrow{n\to\infty} 0,$$

hence the vague limit of the δ_n is not a probability measure.

(ii) Let ν be a probability measure, let $\alpha \in (0,1)$ and define

$$\mu_n = (1-\alpha)\nu + \alpha\delta_n.$$

Then (μ_n) is a sequence of probability measures and for $f \in \mathcal{C}_0(\mathbb{R})$ we have

$$\int f \,\mathrm{d}\mu_n = (1-\alpha)\int f \,\mathrm{d}\nu + \alpha f(n) \xrightarrow{n\to\infty} (1-\alpha)\int f \,\mathrm{d}\nu.$$

Hence $\mu_n \xrightarrow{v} \nu_\alpha = (1-\alpha)\nu$ with

$$\nu_\alpha(\mathbb{R}) = \int \mathrm{d}\nu_\alpha = (1-\alpha)\int \mathrm{d}\nu = 1-\alpha < 1.$$

1.2 When does convergence of moments imply weak convergence?

Theorem 1.2. *Let μ be a measure on $\mathbb{R}$. Set $m_k = \int_{\mathbb{R}} x^k \,\mathrm{d}\mu(x)$ for all $k \in \mathbb{N}$. If*

(i) $\limsup_{k\to\infty} \frac{1}{k}\left|m_k\right|^{\frac{1}{k}} < \infty$

or if

(ii) *$m_k < \infty$ for all $k \in \mathbb{N}$ and $\sum_{k=1}^{\infty} (m_{2k})^{-\frac{1}{2k}} = \infty$* ***(Carleman's condition)***

then μ is the only measure on $\mathbb{R}$ with $(m_k)_k$ as its moments. In this case, we say that μ is ***characterized by its moments****.*

Example. If a probability measure μ is compactly supported, i.e., if there exists an $R > 0$ such that $\mu([-R,R]) = 1$, then μ is characterized by its moments, since

$$|m_k| = \left|\int_R^R x^k \,\mathrm{d}\mu(x)\right| \leq R^k \int \mathrm{d}\mu(x) = R^k$$

and hence

$$\frac{1}{k}\left|m_k\right|^{\frac{1}{k}} \leq \frac{1}{k}R \xrightarrow{k\to\infty} 0.$$

Theorem 1.3. *Let $\mu, \mu_1, \mu_2, \ldots$ be probability measures on $\mathbb{R}$. If*

(i) $\int_{\mathbb{R}} x^k \,\mathrm{d}\mu_n(x) \xrightarrow{n\to\infty} \int_{\mathbb{R}} x^k \,\mathrm{d}\mu(x)$ ***(convergence of moments)***

holds for any $k \geq 0$, and

(ii) *μ is characterized by its moments,*

then $\mu_n \xrightarrow{w} \mu$.

For a Wigner matrix, it is enough to consider the moments and then prove convergence to the Catalan numbers in order to conclude that $\mu_n \xrightarrow{w} \mu$, where μ is the semicircle measure.

"This method is combinatorial."

1.3 The Cauchy-Stieltjes transform

In this section, we shall make an introduction to the analytic approach of proving the Wigner theorem via the Cauchy-Stieltjes transform.

Definition 1.4. Let μ be a positive finite measure on $\mathbb{R}$. The **Stieltjes transform** g_μ of μ is defined by

$$g_\mu \colon \mathbb{C}_+ \to \mathbb{C},\ z \mapsto g_\mu(z) = \int_{\mathbb{R}} \frac{1}{\lambda - z} \, \mathrm{d}\mu(\lambda),$$

where $\mathbb{C}_+ = \{z \in \mathbb{C} \mid \operatorname{Im}(z) > 0\}$.

Let $z = x + iy$ with $x, y \in \mathbb{R}$. Then

$$\begin{aligned}\lambda \mapsto \frac{1}{\lambda - z} &= \frac{1}{\lambda - x - iy} \frac{(\lambda - x) + iy}{(\lambda - x) + iy} \\ &= \frac{\lambda - x}{(\lambda - x)^2 + y^2} + i \frac{y}{(\lambda - x)^2 + y^2} \\ &= \operatorname{Re}\left(\frac{1}{\lambda - z}\right) + \operatorname{Im}\left(\frac{1}{\lambda - z}\right)\end{aligned}$$

is continuous and bounded. As μ is finite, $g_\mu(z)$ is well-defined over $\mathbb{C}_+$.

Example. (i) If $\mu = \delta_0$ then $g_\mu(z) = -\frac{1}{z}$ and if $\mu = \delta_x$ then $g_\mu(z) = -\frac{1}{x-z}$.

(ii) Let $\mu_N = \frac{1}{N} \sum_{i=1}^N \delta_{\lambda_i}$, where λ_i are the eigenvalues of a symmetric matrix A. Then

$$g_{\mu_N}(z) = \frac{1}{N} \operatorname{Tr}(A - zI)^{-1} = \operatorname{tr}(A - zI)^{-1},$$

where $(A - zI)^{-1}$ is the so-called **resolvent** of A.

1.4 Weak convergence iff convergence of Stieltjes transform

Theorem 1.5. *Let* $\nu, \nu_1, \nu_2, \ldots$ *be probability measures. Then:*

$$\nu_n \xrightarrow{w} \nu \quad \Leftrightarrow \quad g_{\nu_n}(z) \xrightarrow{n\to\infty} g_\nu(z) \text{ for all } z \in \mathbb{C}_+$$

1.5 Properties of the Stieltjes transform

Proposition 1.6. *Let* μ *be a finite measure. Then*

(i) $|g_\mu(z)| \leq \frac{\mu(\mathbb{R})}{\mathrm{Im}(z)}$ *for any* $z \in \mathbb{C}_+$,

(ii) g_μ *is analytic over* $\mathbb{C}_+$,

(iii) $\mathrm{Im}(g_\mu(z)) > 0$ *for all* $z \in \mathbb{C}_+$,

(iv) $\mathrm{Im}(zg_\mu(z)) \geq 0$ *if* $\operatorname{supp}\mu \subset \mathbb{R}_+$,

(v) $\lim_{y\to\infty} iyg_\mu(iy) = -\mu(\mathbb{R})$.

1.6 The Stieltjes transform seen as a moment generating function

Suppose that μ is compactly supported on $[-R, R]$. Using the geometric series expansion, we have

$$\begin{aligned} g_\mu(z) = \int_{-R}^{R} \frac{1}{\lambda - z} \, \mathrm{d}\mu(\lambda) &= -\int_{-R}^{R} \sum_{n=0}^{\infty} \frac{\lambda^n}{z^{n+1}} \, \mathrm{d}\mu(\lambda) \\ &= -\sum_{n=0}^{\infty} \frac{1}{z^{n+1}} \int_{-R}^{R} \lambda^n \, \mathrm{d}\mu(\lambda) = -\sum_{n=0}^{\infty} \frac{m_n}{z^{n+1}} \end{aligned}$$

for $z \in \mathbb{C}_+$ with $|z| > R$. So in a neighborhood of ∞, g_μ is a power series in $\frac{1}{z}$ whose coefficients are the moments of μ. This can be useful in computing Stieltjes transforms. As the Stieltjes transform is analytic over $\mathbb{C}_+$, we only need to compute it in some open set with an accumulation point to determine its value over $\mathbb{C}_+$.

1.7 Stieltjes transform characterizes measure

Theorem 1.7 (Inversion formula)**.** *Let* μ *be a measure. For all* $a, b \in \mathbb{R}$ *we have*

$$\frac{1}{2}\left[\mu(\{a\}) + \mu(\{b\})\right] + \mu(]a,b[) = \lim_{y \downarrow 0} \frac{1}{\pi} \int_a^b \operatorname{Im} g_\mu(x+iy)\, \mathrm{d}x.$$

Proof. We know that

$$\operatorname{Im} g_\mu(x+iy) = \int_{\mathbb{R}} \operatorname{Im} \frac{1}{\lambda - x - iy}\, \mathrm{d}\mu(\lambda) = \int_{\mathbb{R}} \frac{y}{(\lambda - x)^2 + y}\, \mathrm{d}\mu(\lambda).$$

Thus,

$$\int_a^b \operatorname{Im} g_\mu(x+iy)\, \mathrm{d}x = \int_{\mathbb{R}} \int_a^b \frac{y}{(\lambda - x)^2 + y}\, \mathrm{d}x\, \mathrm{d}\mu(\lambda) = \int_{\mathbb{R}} f(y,\lambda)\, \mathrm{d}\mu(\lambda),$$

where

$$f(y,\lambda) = \arctan \frac{b-\lambda}{y} - \arctan \frac{a-\lambda}{y}.$$

Note that $|f(y,\lambda)| \leq \pi$ for all $y > 0, \lambda \in \mathbb{R}$ and we have $f(y,\lambda) \xrightarrow{y \to 0^+} f(\lambda)$ with

$$f(\lambda) = \begin{cases} 0, & \lambda \notin [a,b], \\ \frac{\pi}{2}, & \lambda \in \{a,b\}, \\ \pi, & \lambda \in (a,b). \end{cases}$$

Then, by the dominated convergence theorem we infer

$$\begin{aligned} \lim_{y \downarrow 0} \frac{1}{\pi} \int_a^b \operatorname{Im} g_\mu(x+iy)\, \mathrm{d}x &= \lim_{y \downarrow 0} \frac{1}{\pi} \int_{\mathbb{R}} f(y,\lambda)\, \mathrm{d}\mu(\lambda) \\ &= \frac{1}{\pi} \int_{\mathbb{R}} f(\lambda)\, \mathrm{d}\mu(\lambda) \\ &= \frac{1}{2}\left(\mu\{a\} + \mu\{b\}\right) + \int_a^b \mathrm{d}\mu(\lambda). \end{aligned}$$

□

2 Gaussian random matrices: Wick formula and combinatorial proof of Wigner's semicircle

We want to prove convergence of our random matrices to the semicircle by showing

$$\mathbb{E}\left[\operatorname{tr} A_N^{2k}\right] \xrightarrow{N\to\infty} C_k.$$

Up to now our matrices were of the form $A_N = \frac{1}{\sqrt{N}}(a_{ij})_{i,j=1}^N$ with $a_{ij} \in \{-1, 1\}$. We now let the entries a_{ij} be Gaussian (normal) random variables.

Definition 2.1. A **standard Gaussian** (or **normal**) random variable X is a real-valued Gaussian random variable with mean 0 and variance 1, i.e., it has distribution

$$\mathbb{P}\left[t_1 \le X \le t_2\right] = \frac{1}{\sqrt{2\pi}} \int_{t_1}^{t_2} e^{-\frac{t^2}{2}}\,\mathrm{d}t$$

and hence its moments are given by

$$\mathbb{E}\left[X^n\right] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} t^n e^{-\frac{t^2}{2}}\,\mathrm{d}t.$$

Proposition 2.2. *The moments of a standard Gaussian random variable are of the form*

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} t^n e^{-\frac{t^2}{2}}\,\mathrm{d}t = \begin{cases} 0, & n \text{ odd,} \\ (n-1)!!, & n \text{ even,} \end{cases}$$

where

$$m!! = m(m-2)(m-4)\cdots 5\cdot 3\cdot 1.$$

Exercise. Show this by partial integration.

Remark 2.3. It is surprising that those integrals evaluate to natural numbers. They actually count interesting combinatorial objects,

$$\mathbb{E}\left[X^{2k}\right] = \#\{\text{pairings of } 2k \text{ elements}\}.$$

Definition 2.4. (i) For a natural number $n \in \mathbb{N}$ we put $[n] = \{1, \dots, n\}$.

(ii) A **pairing** π of $[n]$ is a decomposition of $[n]$ into disjoints subsets of size 2, i.e., $\pi = \{V_1, \dots, V_k\}$ such that for all $i, j = 1, \dots, k$ with $i \neq j$, we have:

- $V_i \subset [n]$
- $\# V_i = 2$
- $V_i \cap V_j = \emptyset$
- $\bigcup_{i=1}^{k} V_i = [n]$

Note that necessarily $k = \frac{n}{2}$.

(iii) The set of all pairings of $[n]$ is denoted by

$$\mathcal{P}_2(n) = \{\pi \mid \pi \text{ is a pairing of } [n]\}.$$

Proposition 2.5. (i) *We have*

$$\#\mathcal{P}_2(n) = \begin{cases} 0, & n \text{ odd}, \\ (n-1)!!, & n \text{ even}. \end{cases}$$

(ii) *Hence for a standard Gaussian variable X we have*

$$\mathbb{E}\left[X^n\right] = \#\mathcal{P}_2(n).$$

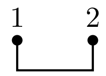
Proof. "(i)" Count elements in $\mathcal{P}_2(n)$ in a recursive way. Choose the pair which contains the element 1, for this we have $n-1$ possibilities. Then we are left with choosing a pairing of the remaining $n-2$ numbers. Hence we have

$$\#\mathcal{P}_2(n) = (n-1) \cdot \#\mathcal{P}_2(n-2).$$

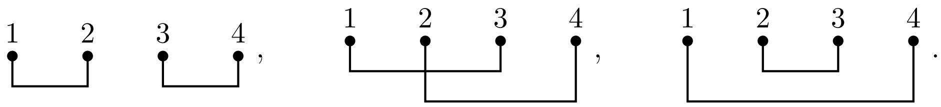
Iterating this and noting that $\#\mathcal{P}_2(1) = 0$ and $\#\mathcal{P}_2(2) = 1$ gives the desired result.

"(ii)" Follows from (i) and Proposition 2.2. □

Example 2.6. Usually we draw our partitions by connecting the elements in each pair. Then $\mathbb{E}[X^2] = 1$ corresponds to the single partition



and $\mathbb{E}[X^4] = 3$ corresponds to the three partitions



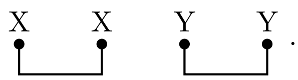
Remark 2.7 (Independent Gaussian random variables). We will have several Gaussian random variables X, Y and have to calculate their joint moments. The random variables are independent, this means that their joint distribution is the product measure of the single distributions,

$$\mathbb{P}[t_1 \leq X \leq t_2, s_1 \leq Y \leq s_2] = \mathbb{P}[t_1 \leq X \leq t_2]\,\mathbb{P}[s_1 \leq Y \leq s_2],$$

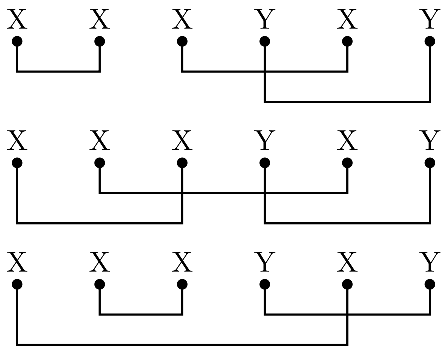
so in particular, for the moments we have

$$\begin{aligned}\mathbb{E}[X^n Y^m] &= \mathbb{E}[X^n]\,\mathbb{E}[Y^m] \\ &= \#\{\text{pairings of } \underbrace{X \cdots X}_{n}\} \cdot \#\{\text{pairings of } \underbrace{Y \cdots Y}_{m}\} \\ &= \#\{\text{pairings of } \underbrace{X \cdots X}_{n}\underbrace{Y \cdots Y}_{m} \text{ which connect } X \text{ with } X \text{ and } Y \text{ with } Y\}.\end{aligned}$$

Example. We have $\mathbb{E}[XXYY] = 1$ since the only possible pairing is



On the other hand, $\mathbb{E}[XXXYXY] = 3$ since we have the following three possible pairings:



Consider $x_1, \dots, x_n \in \{X, Y\}$. Then we still have

$$\mathbb{E}\left[x_1 \dots x_n\right] = \#\{\text{pairings which connect } X \text{ with } X \text{ and } Y \text{ with } Y\}.$$

Can we decide in a more abstract way whether $x_i = x_j$ or $x_i \neq x_j$? Yes, we can read this from the corresponding second moment, since

$$\mathbb{E}\left[x_i x_j\right] = \begin{cases} \mathbb{E}\left[x_i^2\right] = 1, & x_i = x_j, \\ \mathbb{E}\left[x_i\right] \mathbb{E}\left[x_j\right] = 0, & x_i \neq x_j. \end{cases}$$

Hence we have:

$$\mathbb{E}\left[x_1 \dots x_n\right] = \sum_{\pi \in \mathcal{P}_2(n)} \prod_{(i,j) \in \pi} \mathbb{E}\left[x_i x_j\right]$$

Theorem 2.8 (Wick 1950, physics; Isserlis 1918, statistics)**.** *Let $Y_1, \dots, Y_p$ be independent standard Gaussian random variables and consider $x_1, \dots, x_n \in \{Y_1, \dots, Y_p\}$. Then we have the* ***Wick formula***

$$\mathbb{E}\left[x_1 \dots x_n\right] = \sum_{\pi \in \mathcal{P}_2(n)} \mathbb{E}_\pi\left[x_1 \dots x_n\right],$$

where, for $\pi \in \mathcal{P}_2(n)$, we use the notation

$$\mathbb{E}_\pi\left[x_1 \dots x_n\right] = \prod_{(i,j) \in \pi} \mathbb{E}\left[x_i x_j\right].$$

Note that the Wick formula is linear in the x_i, hence it remains valid if we replace the x_i by linear combinations of the x_j. In particular, we can go over to complex Gaussian variables.

Definition 2.9. A **standard complex Gaussian** random variable Z is of the form

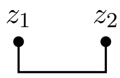
$$Z = \frac{X + iY}{\sqrt{2}},$$

where X and Y are independent standard real Gaussian variables.

Remark 2.10. Let Z be a standard complex Gaussian, i.e., $Z = \frac{X+iY}{\sqrt{2}}$. Then we have

$\overline{Z} = \frac{X-iY}{\sqrt{2}}$ and the first and second moments are given by

$$\begin{aligned}
\mathbb{E}[Z] &= 0,\\
\mathbb{E}\left[\overline{Z}\right] &= 0,\\
\mathbb{E}\left[Z^2\right] &= \mathbb{E}[ZZ] = \frac{1}{2}\left[\mathbb{E}[XX] - \mathbb{E}[YY] + i\left(\mathbb{E}[XY] + \mathbb{E}[YX]\right)\right] = 0,\\
\mathbb{E}\left[\overline{Z}^2\right] &= 0,\\
\mathbb{E}\left[|Z|^2\right] &= \mathbb{E}\left[Z\overline{Z}\right] = \frac{1}{2}\left[\mathbb{E}[XX] + \mathbb{E}[YY] + i\left(\mathbb{E}[YX] - \mathbb{E}[YX]\right)\right] = 1.
\end{aligned}$$

Hence, for $z_1, z_2 \in \{Z, \overline{Z}\}$ and $\pi =$ 

we have

$$\mathbb{E}[z_1 z_2] = \begin{cases} 1, & \pi \text{ connects } Z \text{ with } \overline{Z},\\ 0, & \pi \text{ connects } (Z \text{ with } Z) \text{ or } (\overline{Z} \text{ with } \overline{Z}). \end{cases}$$

Theorem 2.11. *Let $Z_1, \dots, Z_p$ be independent standard complex Gaussian random variables and consider $z_1, \dots, z_n \in \{Z_1, \overline{Z_1}, \dots, Z_p, \overline{Z_p}\}$. Then we have the Wick formula*

$$\begin{aligned}
\mathbb{E}[z_1 \cdots z_n] &= \sum_{\pi \in \mathcal{P}_2(n)} \mathbb{E}_\pi[z_1, \dots, z_n]\\
&= \#\{\textit{pairings of } [n] \textit{ which connect } Z_i \textit{ with } \overline{Z_i}\}.
\end{aligned}$$

Definition 2.12. A **Gaussian random matrix** is of the form $A_N = \frac{1}{\sqrt{N}}\left(a_{ij}\right)_{i,j=1}^N$, where

- $A_N = A_N^*$, i.e., $a_{ij} = \overline{a_{ji}}$ for all i, j,
- $\{a_{ij} \mid i \geq j\}$ are independent,
- each a_{ij} is a standard Gaussian random variable, which is complex for $i \neq j$ and real for $i = j$.

Remark 2.13. (i) More precisely, we should address the above as **selfadjoint** Gaussian random matrices.

(ii) Another common name for those random matrices is **GUE**, which stands for **Gaussian unitary ensemble**. "Unitary" corresponds here to the fact that the entries are complex, since such matrices are invariant under unitary transformations. There are also real and quaternionic versions, **Gaussian orthogonal ensembles (GOE)** and **Gaussian symplectic ensembles (GSE)**.

(iii) Note that we can also express this definition in terms of the Wick formula as

$$\mathbb{E}\left[a_{i(1)j(1)} \cdots a_{i(n)j(n)}\right] = \sum_{\pi \in \mathcal{P}_2(n)} \mathbb{E}_\pi \left[a_{i(1)j(1)}, \ldots, a_{i(n)j(n)}\right]$$

for all n and $1 \leq i(1), j(1), \ldots, i(n), j(n) \leq N$. Furthermore, the second moments are given by

$$\mathbb{E}\left[a_{ij} a_{kl}\right] = \delta_{il}\delta_{jk}.$$

Remark 2.14 (Calculation of $\mathbb{E}\left[\mathrm{tr}(A_N^m)\right]$). For our Gaussian random matrix we want to calculate their moments

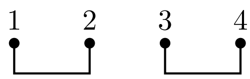
$$\mathbb{E}\left[\mathrm{tr}(A_N^m)\right] = \frac{1}{N}\frac{1}{\sqrt{N}^m} \sum_{i(1),\ldots,i(m)=1}^{N} \mathbb{E}\left[a_{i(1)i(2)} a_{i(2)i(3)} \cdots a_{i(m)i(1)}\right].$$

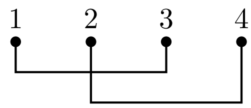
Let us first consider small examples before we treat the general case:

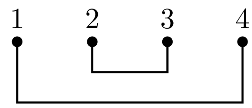
(i)

$$\mathbb{E}\left[\mathrm{tr}(A_N^2)\right] = \frac{1}{N^2} \sum_{i,j=1}^{N} \underbrace{\mathbb{E}\left[a_{ij} a_{ji}\right]}_{=1} = \frac{1}{N^2} N^2 = 1 = C_1$$

(ii) We consider the partitions

$\pi_1 =$ 

, $\pi_2 =$ 

, $\pi_3 =$ 

.

With this, we have

$$\mathbb{E}\left[\mathrm{tr}(A_N^4)\right] = \frac{1}{N^3} \sum_{i,j,k,l=1}^{N} \underbrace{\mathbb{E}\left[a_{ij} a_{jk} a_{kl} a_{li}\right]}_{=\mathbb{E}_{\pi_1}[\ldots] + \mathbb{E}_{\pi_2}[\ldots] + \mathbb{E}_{\pi_3}[\ldots]}$$

and calculate

$$\sum_{i,j,k,l=1}^{N} \mathbb{E}_{\pi_1}\left[a_{ij}, a_{jk}, a_{kl}, a_{li}\right] = \sum_{\substack{i,j,k,l=1 \\ i=k}}^{N} 1 = N^3,$$

$$\sum_{i,j,k,l=1}^{N} \mathbb{E}_{\pi_2}\left[a_{ij}, a_{jk}, a_{kl}, a_{li}\right] = \sum_{\substack{i,j,k,l=1 \\ j=l}}^{N} 1 = N^3,$$

$$\sum_{i,j,k,l=1}^{N} \mathbb{E}_{\pi_3}\left[a_{ij}, a_{jk}, a_{kl}, a_{li}\right] = \sum_{\substack{i,j,k,l=1 \\ i=l, j=k, j=i, k=l}}^{N} 1 = \sum_{i=1}^{N} 1 = N,$$

hence

$$\mathbb{E}\left[\operatorname{tr}(A_N^4)\right] = \frac{1}{N^3}\left(N^3 + N^3 + N\right) = 2 + \frac{1}{N^2},$$

such that

$$\lim_{N\to\infty} \mathbb{E}\left[\operatorname{tr}(A_N^4)\right] = 2 = C_2.$$

(iii) In the general case we have

$$\begin{aligned}\mathbb{E}\left[a_{i(1)i(2)}a_{i(2)i(3)}\cdots a_{i(m)i(1)}\right] &= \sum_{\pi\in\mathcal{P}_2(m)} \mathbb{E}_\pi\left[a_{i(1)i(2)}, a_{i(2)i(3)}, \ldots, a_{i(m)i(1)}\right] \\ &= \sum_{\pi\in\mathcal{P}_2(m)} \prod_{(k,l)\in\pi} \mathbb{E}\left[a_{i(k)i(k+1)}a_{i(l)i(l+1)}\right].\end{aligned}$$

We use the notation $[i = j] = \delta_{ij}$ and, by identifying a pairing π with a permutation $\pi \in S_m$ via

$$(k, l) \in \pi \leftrightarrow \pi(k) = l, \pi(l) = k,$$

find that

$$\begin{aligned}\mathbb{E}\left[\operatorname{tr}(A_N^m)\right] &= \frac{1}{N^{m/2+1}} \sum_{i(1),\ldots,i(m)=1}^{N} \sum_{\pi\in\mathcal{P}_2(m)} \prod_{(k,l)\in\pi} \mathbb{E}\left[a_{i(k)i(k+1)}a_{i(l)i(l+1)}\right] \\ &= \frac{1}{N^{m/2+1}} \sum_{\pi\in\mathcal{P}_2(m)} \sum_{i(1),\ldots,i(m)=1}^{N} \prod_k \Big[i(k) = i(\underbrace{\pi(k)+1}_{\gamma\pi(k)})\Big],\end{aligned}$$

where $\gamma = (1, 2, \ldots, m) \in S_m$ is the shift by 1 modulo m. The above product is different from 0 if and only if

$$i\colon [m] \to [N]$$

is constant on the cycles of $\gamma\pi \in S_m$. Thus,

$$\mathbb{E}\left[\operatorname{tr}(A_N^m)\right] = \frac{1}{N^{m/2+1}} \sum_{\pi\in\mathcal{P}_2(m)} N^{\#(\gamma\pi)},$$

where $\#(\gamma\pi)$ is the number of cycles of the permutation $\gamma\pi$.

Theorem 2.15. *Let A_N be a Gaussian (GUE) random matrix. Then we have for all $m \in \mathbb{N}$,*

$$\mathbb{E}\left[\operatorname{tr}(A_N^m)\right] = \sum_{\pi\in\mathcal{P}_2(m)} N^{\#(\gamma\pi)-\frac{m}{2}-1}.$$

Example 2.16. (i) All odd moments are zero, since $\mathcal{P}_2(2k+1) = \emptyset$.

(ii) If $m = 2$, $\gamma = (1,2)$ and $\pi = (1,2)$ then $\gamma\pi = \mathrm{id} = (1)(2)$ such that $\#(\gamma\pi) = 2$ and

$$\#(\gamma\pi) - \frac{m}{2} - 1 = 0.$$

Thus,

$$\mathbb{E}\left[\operatorname{tr}\left(A_N^2\right)\right] = N^0 = 1.$$

(iii) Let $m = 4$ and $\gamma = (1,2,3,4)$. Then we have

π	$\gamma\pi$	$\#(\gamma\pi) - 3$	contribution
$(1,2)(34)$	$(1,3)(2)(4)$	0	$N^0 = 1$
$(13)(24)$	$(1,4,3,2)$	-2	$N^{-2} = \frac{1}{N^2}$
$(14)(23)$	$(1)(2,4)(3)$	0	$N^0 = 1$

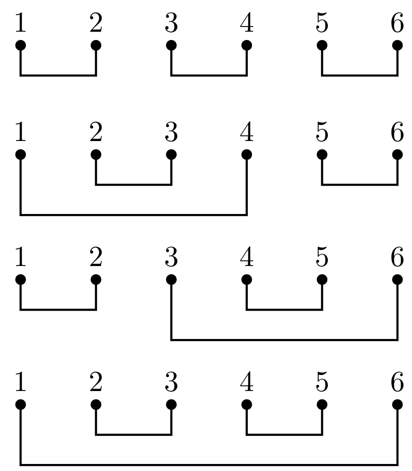
such that

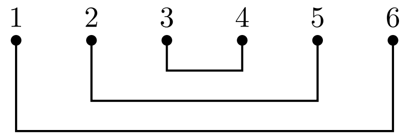
$$\mathbb{E}\left[\operatorname{tr}\left(A_N^4\right)\right] = 2 + \frac{1}{N^2}.$$

(iv) In the same way one can calculate that

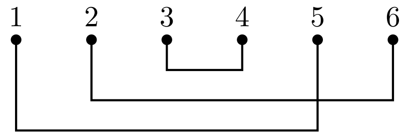
$$\mathbb{E}\left[\operatorname{tr}\left(A_N^6\right)\right] = 5 + 10\frac{1}{N^2},$$
$$\mathbb{E}\left[\operatorname{tr}\left(A_N^8\right)\right] = 14 + 70\frac{1}{N^2} + 21\frac{1}{N^4}.$$

(v) For $m = 6$ the following 5 pairings give contribution N^0:

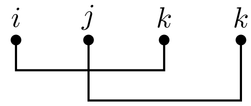




Those are non-crossing pairings, all other pairings $\pi \in \mathcal{P}_2(6)$ have a crossing, e.g.:



Definition 2.17. A pairing $\pi \in \mathcal{P}_2(m)$ is **non-crossing** (**NC**) if there are no pairs (i, k) and (j, l) in π with $i < j < k < l$.

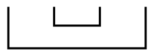
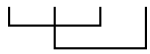


is not allowed!

We put

$$\mathcal{NC}_2(m) = \{\pi \in \mathcal{P}_2(m) \mid \pi \text{ is non-crossing}\}.$$

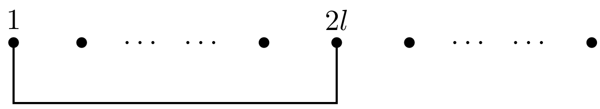
Example 2.18. (i) $\mathcal{NC}_2(2) = \mathcal{P}_2(2) = \{$  $\}$

(ii) $\mathcal{NC}_2(4) = \{$  ,  $\}$ and $\mathcal{P}_2(4) \backslash \mathcal{NC}_2(4) = \{$  $\}$

(iii) The 5 elements of $\mathcal{NC}_2(6)$ are given in Example 2.16 (v), $\mathcal{P}_2(6)$ contains 15 elements.

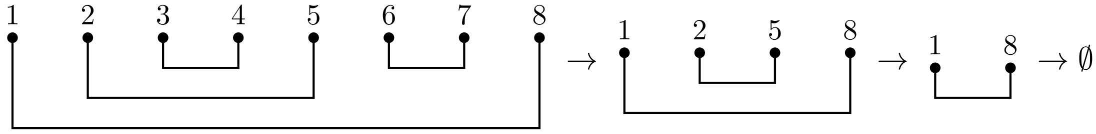
Remark 2.19. Note that NC-pairings have a recursive structure, which usually is crucial for dealing with them.

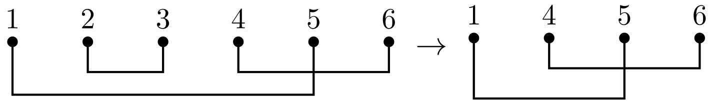
(i) The first pair of $\pi \in \mathcal{NC}_2(2k)$ must necessarily be of the form $(1, 2l)$ and the remaining pairs can only pair within $\{2, \ldots, 2l-1\}$ or within $\{2l+1, \ldots, 2l\}$.



(ii) Iterating this shows that we must find in any $\pi \in \mathcal{NC}_2(2k)$ at least one pair of the form $(i, i+1)$ with $1 \leq i \leq 2k-1$. Removing this pair gives a NC-pairing of $2k-2$ points. This characterizes the NC-pairings as those pairings, which can be reduced to the empty set by iterated removal of pairs, which consist of neighbors.

For example:





no further reduction possible!

Proposition 2.20. *Consider m even and let $\pi \in \mathcal{P}_2(m)$, which we identify with a permutation $\pi \in S_m$. As before, $\gamma = (1, 2, \ldots, m) \in S_m$. Then we have:*

(i) $\#(\gamma\pi) - \frac{m}{2} - 1 \leq 0$ *for all* $\pi \in \mathcal{P}_2(m)$.

(ii) $\#(\gamma\pi) - \frac{m}{2} - 1 = 0$ *if and only if* $\pi \in \mathcal{NC}_2(m)$.

Proof. First we note that a pair $(i, i+1)$ in π corresponds to a fixed point of $\gamma\pi$. More precisely, $i + 1 \xrightarrow{\pi} i \xrightarrow{\gamma} i + 1$ and $i \xrightarrow{\pi} i + 1 \xrightarrow{\gamma} i + 2$. Hence $\gamma\pi$ contains the cycles $(i + 1)$ and $(\ldots, i, i + 2, \ldots)$.

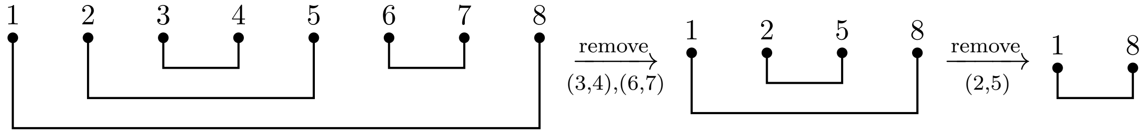
This implication also goes in the other direction: If $\gamma\pi(i+1) = i+1$ and $\gamma\pi(i) = i+2$ then $\pi(i + 1) = \gamma^{-1}(i + 1) = i$ and $\pi(i) = \gamma^{-1}(i + 2) = i + 1$. Hence we have the cycle $(i, i + 1)$ in π.

If we have $(i, i + 1)$ in π, we can remove the points i and $i + 1$, yielding another pairing $\tilde{\pi}$. By doing so, we remove in $\gamma\pi$ the cycle $(i+1)$ and we remove in the cycle $(\ldots, i, i + 2, \ldots)$ the point i, yielding $\gamma\tilde{\pi}$. We reduce m by 2 and $\#(\gamma\pi)$ by 1.

If π is NC we can iterate this until we arrive at $\tilde{\pi}$ with $m = 2$. Then we have $\tilde{\pi} = (1, 2)$ and $\gamma = (1, 2)$ such that $\gamma\tilde{\pi} = (1)(2)$ and $\#(\gamma\tilde{\pi}) = 2$. If $m = 2k$ we did $k - 1$ reductions where we reduced the number of cycles by 1 and at the end we remain with 2 cycles, hence

$$\#(\gamma\pi) = (k - 1) \cdot 1 + 2 = k + 1 = \frac{m}{2} + 1.$$

Here is an example for this:

$\pi =$ 

$$\gamma\pi = (1)(268)(35)(4)(7) \longrightarrow (1)(28)(5) \longrightarrow (1)(8)$$

For a general $\pi \in \mathcal{P}_2(m)$ we remove cycles $(i, i+1)$ as long as possible. If π is crossing we arrive at a pairing $\tilde{\pi}$, where this is not possible anymore. It suffices to show that such a $\tilde{\pi} \in \mathcal{P}_2(m)$ satisfies $\#(\gamma\tilde{\pi}) - \frac{m}{2} < 0$. But since $\tilde{\pi}$ has no cycle $(i, i+1)$, $\gamma\tilde{\pi}$ has no fixed point. Hence each cycle has at least 2 elements, thus

$$\#(\gamma\tilde{\pi}) \leq \frac{m}{2} < \frac{m}{2} + 1.$$

Note that in the above arguments, $(1, m)$ also counts as a pair of neighbors for a $\pi \in \mathcal{P}_2(m)$, in order to have the characterization of fixed points right. Hence, when reducing a general pairing to one without fixed points we have also to remove such cyclic neighbors as long as possible. □

Theorem 2.21 (Wigner's semicircle law for GUE, averaged version)**.** *Let A_N be a Gaussian (GUE) $N \times N$ random matrix. Then we have for all $m \in \mathbb{N}$:*

$$\lim_{N\to\infty} \mathbb{E}\left[\operatorname{tr}\left(A_N^m\right)\right] = \frac{1}{2\pi}\int_{-2}^{2} x^m \sqrt{4 - x^2}\,\mathrm{d}x$$

Proof. This is true for odd m, since then both sides are equal to zero. Consider $m = 2k$ even. Then Theorem 2.15 and Proposition 2.20 show that

$$\lim_{\mathbb{N}\to\infty} \mathbb{E}\left[\operatorname{tr}\left(A_N^m\right)\right] = \sum_{\pi\in\mathcal{P}_2(m)} \lim_{N\to\infty} N^{\#(\gamma\pi)-\frac{m}{2}-1} = \sum_{\pi\in\mathcal{NC}_2(m)} 1 = \#\mathcal{NC}_2(m).$$

Since the moments of the semicircle are given by the Catalan numbers, it remains to see that

$$\#\mathcal{NC}_2(2k) \stackrel{!}{=} C_k.$$

We now count $d_k = \#\mathcal{NC}_2(2k)$ according to the recursive structure of NC-pairings as in 2.19 (i). Namely, we can identify $\pi \in \mathcal{NC}_2(2k)$ with $\{(1, 2l)\} \cup \pi_0 \cup \pi_1$, where $l \in \{1, \dots, 2k-1\}$, $\pi_0 \in \mathcal{NC}_2(2(l-1))$ and $\pi_1 \in \mathcal{NC}_2(2(k-l))$. Hence

$$d_k = \sum_{l=1}^{k} d_{l-1} d_{k-l}, \qquad \text{where } d_0 = 1.$$

This is the recursion for the Catalan numbers, whence $d_k = C_k$ for all $k \in \mathbb{N}$. □

Remark 2.22. (i) One can refine

$$\#(\gamma\pi) - \frac{m}{2} - 1 \leq 0$$

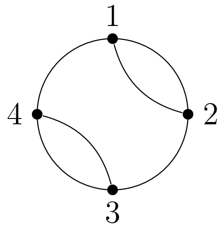
to

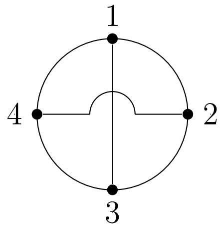
$$\#(\gamma\pi) - \frac{m}{2} - 1 = -2g(\pi)$$

for $g(\pi) \in \mathbb{N}_0$. This g has the meaning that it is the minimal genus of a surface on which π can be drawn without crossings. NC pairings are also called **planar**, they correspond to $g = 0$. Theorem 2.15 is usually addressed as **genus expansion**,

$$\mathbb{E}\left[\operatorname{tr}(A_N^m)\right] = \sum_{\pi \in \mathcal{P}_2(m)} N^{-2g(\pi)}.$$

(ii) For example, $(1,2)(3,4) \in \mathcal{NC}_2(4)$ has $g = 0$, but the crossing pairing $(1,3)(2,4) \in \mathcal{P}_2(4)$ has genus $g = 1$. It has a crossing in the plane but this can be avoided on a torus.





(iii) If we denote

$$\varepsilon_g(k) = \#\left\{\pi \in \mathcal{P}_2(2k) \mid \pi \text{ has genus } g\right\}$$

then the genus expansion can be written as

$$\mathbb{E}\left[\operatorname{tr}(A_N^{2k})\right] = \sum_{g \geq 0} \varepsilon_g(k) N^{-2g}.$$

We know that

$$\varepsilon_g(0) = C_k = \frac{1}{k+1}\binom{2k}{k},$$

but what about the $\varepsilon_g(k)$ for $g > 0$? There does not exist an explicit formula for them, but Harer and Zagier have shown in 1986 that

$$\varepsilon_g(k) = \frac{(2k)!}{(k+1)!(k-2g)!} \cdot \lambda_g(k),$$

where $\lambda_g(k)$ is the coefficient of x^{2g} in

$$\left(\frac{\frac{x}{2}}{\tanh \frac{x}{2}}\right)^{k+1}.$$

We will come back later to this statement of Harer and Zagier.

3 Wigner matrices: Combinatorial proof of Wigner's semicircle law

Definition 3.1. Let μ be a probability distribution on $\mathbb{R}$. A corresponding **Wigner random matrix** is of the form $A_N = \frac{1}{\sqrt{N}} (a_{ij})_{i,j=1}^N$, where

- $A_N = A_N^*$, i.e., $a_{ij} = a_{ji}$ for all i, j,
- $\{a_{ij} \mid i \geq j\}$ are independent,
- each a_{ij} has distribution μ.

Remark 3.2. (i) In our combinatorial setting we will assume that all moments of μ exists, that the first moment is 0 and the second moment will be normalized to 1. In an analytic setting one can deal with more general situations: Usually only the existence of the second moment is needed and one can also allow non-vanishing mean.

(ii) Often one also allows different distributions for the diagonal and for the off-diagonal entries.

(iii) Even more general, one can give up the assumption of identical distribution of all entries and replace this by uniform bounds on the moments.

(iv) Consider a Wigner matrix $A_N = \frac{1}{\sqrt{N}} (a_{ij})_{i,j=1}^N$, where μ has all moments and

$$\int_{\mathbb{R}} x \, \mathrm{d}\mu(x) = 0, \qquad \int_{\mathbb{R}} x^2 \, \mathrm{d}\mu(x) = 1.$$

Then

$$\begin{aligned}
\mathbb{E}\left[\operatorname{tr}(A_N^m)\right] &= \frac{1}{N^{1+\frac{m}{2}}} \sum_{i_1,\dots,i_m=1}^{N} \mathbb{E}\left[a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_m i_1}\right] \\
&= \frac{1}{N^{1+\frac{m}{2}}} \sum_{\sigma \in \mathcal{P}(m)} \sum_{\substack{i \colon [m] \to [N] \\ \ker i = \sigma}} \mathbb{E}\left[\sigma\right].
\end{aligned}$$

Definition 3.3. (i) A **partition** σ of $[n]$ is a decomposition of $[n]$ into disjoint, non-empty subsets (of arbitrary size), i.e., $\sigma = \{V_1, \dots, V_k\}$, where

- $V_i \subset [n]$ for all i,
- $V_i \neq \emptyset$ for all i,
- $V_i \cap V_j = \emptyset$ for all $i \neq j$,
- $\bigcup_{i=1}^k V_i = [n]$.

The V_i are called **blocks** of σ. The set of all partitions of $[n]$ is denoted by

$$\mathcal{P}(n) = \{\sigma \,|\, \sigma \text{ is a partition of } [n]\}\,.$$

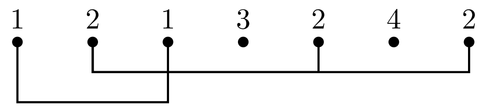
(ii) For a multiindex $i = (i_1, \dots, i_m)$ we denote by $\ker i$ its **kernel**, this is the partition $\sigma \in \mathcal{P}(m)$ such that we have $i_k = i_l$ if and only if k and l are in the same block of σ.

If we identify i with a function $i\colon [m] \to [N]$ via $i(k) = i_k$ then we can also write

$$\ker i = \left\{i^{-1}(1), i^{-1}(2), \dots, i^{-1}(N)\right\},$$

where we discard all empty sets.

Example 3.4. For $i = (1, 2, 1, 3, 2, 4, 2)$ we have



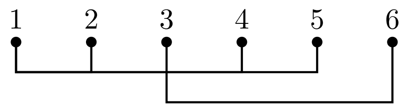
such that

$$\ker i = \{(1, 3), (2, 5, 7), (4), (6)\} \in \mathcal{P}(7).$$

Remark 3.5. The relevance of this kernel in our setting is the following:
For $i = (i_1, \dots, i_m)$ and $j = (j_1, \dots, j_m)$ with $\ker i = \ker j$ we have

$$\mathbb{E}\left[a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_m i_1}\right] = \mathbb{E}\left[a_{j_1 j_2} a_{j_2 j_3} \cdots a_{j_m j_1}\right].$$

For example, for $i = (1, 1, 2, 1, 1, 2)$ and $j = (2, 2, 7, 2, 2, 7)$ we have

$\ker i =$  $= \ker j$

such that

$$\mathbb{E}\left[a_{11} a_{12} a_{21} a_{11} a_{12} a_{21}\right] = \mathbb{E}\left[a_{11}^2\right] \mathbb{E}\left[a_{12}^4\right] = \mathbb{E}\left[a_{22}^2\right] \mathbb{E}\left[a_{27}^4\right] = \mathbb{E}\left[a_{22} a_{27} a_{72} a_{22} a_{27} a_{72}\right].$$

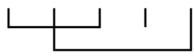
We denote this common value by

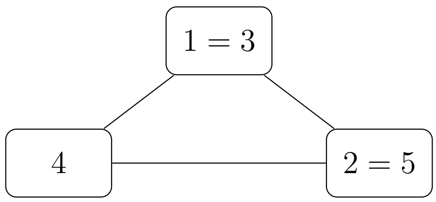
$$\mathbb{E}[\sigma] = \mathbb{E}[a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_m i_1}]$$

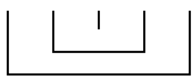
if $\ker i = \sigma$. Thus we get:

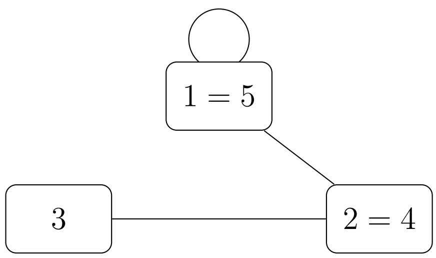
$$\mathbb{E}[\operatorname{tr}(A_N^m)] = \frac{1}{N^{1+\frac{m}{2}}} \sum_{\sigma \in \mathcal{P}(m)} \mathbb{E}[\sigma] \cdot \#\{i \colon [m] \to [N] \mid \ker i = \sigma\}$$

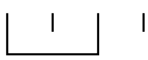
Definition 3.6. For $\sigma = \{V_1, \ldots, V_k\} \in \mathcal{P}(m)$ we define a corresponding graph $\mathcal{G}_\sigma$ as follows: The vertices of $\mathcal{G}_\sigma$ are given by the blocks V_p of σ and there is an edge betwees V_p and V_q if there is an $r \in [m]$ such that, modulo m, $r \in V_p$ and $r+1 \in V_q$. Another way of saying this is that we start with a graph with vertices $1, 2, \ldots, m$ and edges $(1,2), (2,3), (3,4), \ldots, (m-1, m), (m, 1)$ and then identify vertices according to the blocks of σ. We keep loops, but erase multiple edges.

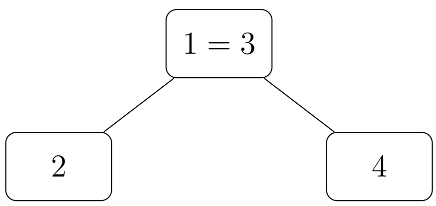
Example 3.7. (i) $\sigma = \{(1,3),(2,5),(4)\} =$ 

$\mathcal{G}_\sigma =$ 

(ii) $\sigma = \{(1,5),(2,4),(3)\} =$ 

$\mathcal{G}_\sigma =$ 

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

$\mathcal{G}_\sigma =$ 

The term $\mathbb{E}[a_{i_1 i_2} a_{i_2 i_3} \cdots a_{i_m i_1}]$ corresponds to a walk in $\mathcal{G}_\sigma$, with $\sigma = \ker i$, along the edges with steps

$$i_1 \to i_2 \to i_3 \to \cdots \to i_m \to i_1.$$

Hence we are using each edge in $\mathcal{G}_\sigma$ at least once. Note that different edges in $\mathcal{G}_\sigma$ correspond to independent random variables. Hence, if we use an edge only once in our walk, then $\mathbb{E}[\sigma] = 0$, because the expectation factorizes into a product with one factor being the first moment of a_{ij}, which is assumed to be zero. Thus, every edge must be used at least twice, but this implies

$$\#\text{ edges in } \mathcal{G}_\sigma \leq \frac{\#\text{ steps in the walk}}{2} = \frac{m}{2}.$$

Since the number of i with the same kernel is

$$\#\{i\colon [m] \to [N] \mid \ker i = \sigma\} = N(N-1)(N-2)\cdots(N-\#\sigma+1),$$

where $\#\sigma$ is the number of blocks in σ, we finally get

$$\mathbb{E}[\operatorname{tr}(A_N^m)] = \frac{1}{N^{1+\frac{m}{2}}} \sum_{\substack{\sigma\in\mathcal{P}(m) \\ \#E(\mathcal{G}_\sigma)\leq\frac{m}{2}}} \mathbb{E}[\sigma] \underbrace{N(N-1)(N-2)\cdots(N-\#\sigma+1)}_{\sim N^{\#\sigma}\text{ for } N\to\infty}. \qquad (\star)$$

We will now use the following well-known basic result from graph theory.

Proposition 3.8. *Let $\mathcal{G} = (V, E)$ be a connected finite graph with vertices V and edges E. Note that we allow loops and multi-edges. Then we have that*

$$\#V \leq \#E + 1$$

*and we have equality if and only if $\mathcal{G}$ is a **tree**, i.e., a connected graph without cycles.*

Theorem 3.9 (Wigner's semicircle law for Wigner matrices, averaged version)**.** *Let A_N be a Wigner matrix corresponding to μ, which has all moments, with mean 0 and second moment 1. Then we have for all $m \in \mathbb{N}$:*

$$\lim_{N\to\infty} \mathbb{E}[\operatorname{tr}(A_N^m)] = \frac{1}{2\pi} \int_{-2}^{2} x^m \sqrt{4-x^2}\,\mathrm{d}x.$$

Proof. From $(\star)$ we get

$$\lim_{N\to\infty} \mathbb{E}[\operatorname{tr}(A_N^m)] = \sum_{\sigma\in\mathcal{P}(m)} \mathbb{E}[\sigma]\, N^{\#V(\mathcal{G}_\sigma)-\frac{m}{2}-1}.$$

In order to have $\mathbb{E}[\sigma] \neq 0$, we can restrict to σ with $\#E(\mathcal{G}_\sigma) \leq \frac{m}{2}$, which by Proposition 3.8 implies that

$$\#V(\mathcal{G}_\sigma) \leq \#E(\mathcal{G}_\sigma) + 1 \leq \frac{m}{2} + 1.$$

Hence all terms converge and the only contribution in the limit $N \to \infty$ comes from those σ, where we have equality, i.e.,

$$\#V(\mathcal{G}_\sigma) = \#E(\mathcal{G}_\sigma) + 1 = \frac{m}{2} + 1.$$

Thus, $\mathcal{G}_\sigma$ must be a tree and in our walk we use each edge exactly twice (necessarily in opposite directions). For such a σ we have $\mathbb{E}[\sigma] = 1$, such that

$$\lim_{N\to\infty} \mathbb{E}\left[\operatorname{tr}\left(A_N^m\right)\right] = \#\left\{\sigma \in \mathcal{P}(m) \,|\, \mathcal{G}_\sigma \text{ is a tree}\right\}.$$

We will check in an exercise that the latter number is also counted by the Catalan numbers. □

4 Semicircle law for GOE via the resolvent method

Recall that a GOE is a symmetric matrix $A_N = \frac{1}{\sqrt{N}} \left(a_{ij}\right)_{i,j=1}^{N}$ such that $\left(a_{ij}\right)_{j \leq i}$ are independent identically distributed real Gaussian random variables with mean 0 and second moments σ^2. Furthermore, the empirical spectral measure was given by

$$\mu_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i},$$

where $\lambda_1, \ldots, \lambda_N$ are the eigenvalues of A_N, and we know that

$$\mu_N \xrightarrow{w} \mu \qquad \text{(almost surely)}, \tag{$\star$}$$

where μ is the semicircle with density

$$\frac{1}{2\pi\sigma^2} \sqrt{4\sigma^2 - x^2}\, \mathrm{d}x.$$

In this chapter, we shall prove ($\star$) via the resolvent method (Stieltjes transform). This is equivalent to proving that for all $z \in \mathbb{C}_+$,

$$g_{\mu_N}(z) \xrightarrow{N \to \infty} g(z) \qquad \text{(almost surely)},$$

where g is the Stieltjes transform of μ given by

$$g(z) = \frac{-z + \sqrt{z^2 - 4\sigma^2}}{2\sigma^2}.$$

By Problem 5, g satisfies the quadratic equation

$$g(z) = -\frac{1}{z} - \frac{1}{z}\sigma^2 \left(g(z)\right)^2.$$

It is typical for such problems to do the following:

Step 1: (Omitted for now) Prove that, almost surely, $g_{A_N}(z)$ asymptotically behaves as $\mathbb{E}\left[g_{A_N}(z)\right]$, i.e, for all $z \in \mathbb{C}_+$,

$$\left|g_{A_N}(z) - \mathbb{E}\left[g_{A_N}(z)\right]\right| \xrightarrow{N\to\infty} 0 \qquad \text{(almost surely)},$$

via concentration inequalities.

Step 2: (Aim of this chapter) Prove for all $z \in \mathbb{C}_+$ that

$$\mathbb{E}\left[g_{A_N}(z)\right] \xrightarrow{N\to\infty} g(z).$$

For a selfadjoint matrix A_N,

$$g_{A_N}(z) = \frac{1}{N} \operatorname{Tr} \left(A_N - zI\right)^{-1} = \operatorname{tr} \left(A_N - zI\right)^{-1}.$$

We will look at the matrix A_N as a matrix-valued mapping of its entries, i.e., we define the mapping

$$A\colon \mathbb{R}^m \to M_N(\mathbb{R}),\ a \mapsto A(a) = A_N.$$

Denote by

$$G(a) = \left(A(a) - zI\right)^{-1}$$

the resolvent matrix G of A.

Lemma 4.1.

$$\frac{\partial G}{\partial a_{ij}}(a) = -G(a)\left(\frac{\partial A}{\partial a_{ij}}(a)\right)G(a)$$

Proof. Differentiate both sides of

$$G(A - zI) = I$$

to obtain

$$0 = \frac{\partial}{\partial a_{ij}}\left[G(A - zI)\right] = \frac{\partial G}{\partial a_{ij}}(A - zI) + G\frac{\partial A}{\partial a_{ij}}$$

such that

$$\frac{\partial G}{\partial a_{ij}}(A - zI) = -G\frac{\partial A}{\partial a_{ij}}$$

and hence

$$\frac{\partial G}{\partial a_{ij}} = -G\frac{\partial A}{\partial a_{ij}}G.$$

□

In particular, we have the Wigner matrix

$$A\colon \mathbb{R}^{N(N+1)/2} \to M_N(\mathbb{R}),\ a = (a_{ij})_{1\le j\le i\le N} \mapsto A(a),$$

where

$$[A(a)]_{ij} = \frac{1}{\sqrt{N}} \begin{cases} a_{ij}, & j \le i, \\ a_{ji}, & j \ge i, \end{cases}$$

and for any i, j, k,

$$\left[\frac{\partial G}{\partial a_{kk}}\right]_{ij} = -\frac{1}{\sqrt{N}} G_{ik}G_{kj}$$

and for $k \neq l$,

$$\left[\frac{\partial G}{\partial a_{kl}}\right]_{ij} = -\frac{1}{\sqrt{N}} \left(G_{ik}G_{lj} + G_{il}G_{kj}\right).$$

Definition 4.2. A $k \times 1$-random vector X is said to be **absolutely continuous** if the set of its values is continuous and for any $[a,b] = [a_1,b_1] \times \cdots [a_k,b_k]$,

$$\mathbb{P}\left[X \in [a,b]\right] = \int_{a_1}^{b_1} \cdots \int_{a_k}^{b_k} f_X(x_1,\ldots,x_k)\,\mathrm{d}x_1 \cdots \mathrm{d}x_k.$$

The function $f_X\colon \mathbb{R}^k \to [0,\infty)$ is called the **joint probability density function** of X. Furthermore, we set

$$\mathbb{E}\left[X\right] = \begin{pmatrix} \mathbb{E}\left[X_1\right] \\ \vdots \\ \mathbb{E}\left[X_k\right] \end{pmatrix}$$

and define the **covariance matrix** Σ by

$$\Sigma = \mathbb{E}\left[\left(X - \mathbb{E}\left[X\right]\right)\left(X - \mathbb{E}\left[X\right]\right)^T\right] = \mathbb{E}\left[XX^T\right] - \mathbb{E}\left[X\right]\mathbb{E}\left[X\right]^T.$$

Notice that Σ is positive semi-definite with rank 1 and

$$(\Sigma)_{ij} = \mathbb{E}\left[\left(X_i - \mathbb{E}\left[X_i\right]\right)\left(X_j - \mathbb{E}\left[X_j\right]\right)^T\right] = \mathrm{Cov}(X_i, X_j).$$

Definition 4.3. (i) We say that an absolutely continuous $k \times 1$-random vector X is a **multivariate Gaussian vector** with mean $\mathbb{E}\left[X\right] = \mu$ and covariance matrix Σ if its joint probability density f_X is given by

$$f_X(x) = \frac{1}{(2\pi)^{\frac{k}{2}}\left|\Sigma\right|^{\frac{1}{2}}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right).$$

We write $X \sim \mathcal{N}(\mu, \Sigma)$.

(ii) If $\mu = 0$ and $\Sigma = I$ we say that X is a **standard Gaussian vector**.

(iii) If $\mu = 0$ and $\Sigma = I\sigma^2$ we set

$$\mathrm{d}\gamma_X^{\sigma^2} = \frac{1}{(2\pi\sigma^2)^{\frac{k}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^k x_i^2}\,\mathrm{d}x_1 \cdots \mathrm{d}x_k$$

and

$$\mathrm{d}\gamma_{X_i}^{\sigma^2} = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x_i^2}{2\sigma^2}}\,\mathrm{d}x_i.$$

Lemma 4.4. *X is a Standard Gaussian vector if and only if $X_1, \ldots, X_k$ are independently identically distributed with $X_1 \sim \mathcal{N}(0,1)$.*

Proof. We have

$$\begin{aligned}
\mathrm{d}\gamma_X^{\sigma^2} &= \frac{1}{(2\pi\sigma^2)^{\frac{k}{2}}} e^{-\frac{1}{2\sigma^2}\sum_{i=1}^k x_i^2}\,\mathrm{d}x_1 \cdots \mathrm{d}x_k \\
&= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x_1^2}{2\sigma^2}}\,\mathrm{d}x_1 \cdots \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x_k^2}{2\sigma^2}}\,\mathrm{d}x_k \\
&= \mathrm{d}\gamma_{X_1}^{\sigma^2} \cdots \mathrm{d}\gamma_{X_k}^{\sigma^2}
\end{aligned}$$

such that

$$\begin{aligned}
\mathbb{P}\left[a_1 \le X_1 \le b_1, \ldots, a_k \le X_k \le b_k\right] &= \int_a^b \mathrm{d}\gamma_X^{\sigma^2} \\
&= \int_{a_1}^{b_1} \cdots \int_{a_k}^{b_k} \mathrm{d}\gamma_{X_1}^{\sigma^2} \cdots \mathrm{d}\gamma_{X_k}^{\sigma^2} \\
&= \mathbb{P}\left[a_1 \le X_1 \le b_1\right] \cdots \mathbb{P}\left[a_k \le X_k \le b_k\right].
\end{aligned}$$

□

Proposition 4.5 (Stein's identity, "Gaussian integration by parts")**.** *Let X be a Gaussian vector with $X \sim \mathcal{N}(0, \sigma^2 I)$. Let $h\colon \mathbb{R}^k \to \mathbb{R}$ be a function such that for each $i \in \{1, \ldots, k\}$, $\partial_i h$ is is continuous almost everywhere and $\mathbb{E}\left[|\partial_i h|\right] < \infty$. Then*

$$\mathbb{E}\left[X_i h(X_1, \ldots, X_k)\right] = \sigma^2 \mathbb{E}\left[\partial_i h(X_1, \ldots, X_k)\right].$$

Proof. For a Gaussian random variable $Z \sim \mathcal{N}(0, \sigma^2)$ and a continuously differentiable function g with $\mathbb{E}[|g'(Z)|] < \infty$, we have, by an easy integration by parts,

$$\mathbb{E}[Zg(Z)] = \sigma^2 \mathbb{E}[g'(Z)]. \tag{$\star$}$$

So as $\Sigma = \sigma^2 I$, the $X_i \sim \mathcal{N}(0, \sigma^2)$ are independently identically distributed and we calculate

$$\begin{aligned}
\mathbb{E}[X_i h(X_1, \ldots, X_k)] &= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} x_i h(x_1, \ldots, x_k) \, \mathrm{d}\gamma_X^{\sigma^2}(x_1, \ldots, x_k) \\
&= \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \prod_{\substack{j=1 \\ j \neq i}}^{k} \mathrm{d}\gamma_{X_j}^{\sigma^2}(x_j) \int_{\mathbb{R}} x_i h(x_1, \ldots, x_k) \, \mathrm{d}\gamma_{X_i}^{\sigma^2}(x_i) \\
&\overset{(\star)}{=} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \prod_{\substack{j=1 \\ j \neq i}}^{k} \mathrm{d}\gamma_{X_j}^{\sigma^2}(x_j) \int_{\mathbb{R}} \partial_i h(x_1, \ldots, x_k) \, \mathrm{d}\gamma_{X_i}^{\sigma^2}(x_i) \\
&= \sigma^2 \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \partial_i h(x_1, \ldots, x_k) \, \mathrm{d}\gamma_{X_1}^{\sigma^2}(x_1) \cdots \mathrm{d}\gamma_{X_1}^{\sigma^2}(x_1) \\
&= \sigma^2 \mathbb{E}[\partial_i h(X_1, \ldots, X_k)].
\end{aligned}$$

□

Theorem 4.6. *Let A_N be a GOE matrix with $a_{ij} \sim \mathcal{N}(0, \sigma^2)$. Then for all $z \in \mathbb{C}_+$,*

$$\mathbb{E}[g_{A_N}(z)] \to g(z),$$

where g is the Stieltjes transform of the semicircle distribution (of variance σ^2) given by

$$g(z) = \frac{-z + \sqrt{z^2 - 4\sigma^2}}{2}$$

and satisfying the quadratic equation

$$g(z) = -\frac{1}{z} - \frac{\sigma^2}{z} g(z)^2.$$

Proof. Just consider the case $\sigma^2 = 1$. Then

$$\mathbb{E}[g_{A_N}(z)] = \mathbb{E}\left[\frac{1}{N} \operatorname{Tr}(A_N - zI)^{-1}\right] = \frac{1}{N} \mathbb{E}[\operatorname{Tr} G_{A_n}(z)].$$

Now, $(A_N - zI)G_{A_n}(z) = I$ such that

$$G_{A_n}(z) = -\frac{1}{z}I + \frac{1}{z}A_n G_{A_N}(z)$$

and hence, by Proposition 4.5 and Lemma 4.1,

$$\begin{aligned}
\frac{1}{N}\mathbb{E}\left[\operatorname{Tr} G_{A_n}(z)\right] &= -\frac{1}{z} + \frac{1}{zN}\mathbb{E}\left[\operatorname{Tr}(A_n G_{A_N}(z))\right] \\
&= -\frac{1}{z} + \frac{1}{zN^{\frac{3}{2}}}\sum_{i,j=1}^{N}\mathbb{E}\left[a_{ij}G_{ji}\right] \\
&= -\frac{1}{z} + \frac{1}{zN^{\frac{3}{2}}}\sum_{i,j=1}^{N}\mathbb{E}\left[\partial_{a_{ij}}G_{ji}\right] \\
&= -\frac{1}{z} - \frac{1}{zN^2}\sum_{i,j=1}^{N}\mathbb{E}\left[-G_{ji}G_{ij} - G_{jj}G_{ii}\right] \\
&= -\frac{1}{z} - \frac{1}{zN^2}\mathbb{E}\left[\operatorname{Tr} G^2\right] - \frac{1}{z}\mathbb{E}\left[\left(\frac{1}{N}\operatorname{Tr} G\right)^2\right].
\end{aligned}$$

Since A_N is selfadjoint, $A_N = U \operatorname{diag}(\lambda_1, \ldots, \lambda_N)U^*$ such that

$$G_{A_N}^2 = U \operatorname{diag}\left(\left(\frac{1}{\lambda_1 - z}\right)^2, \ldots, \left(\frac{1}{\lambda_N - z}\right)^2\right)U^*.$$

Furthermore, $\frac{1}{|\lambda_i - z|^2} \leq \frac{1}{|\operatorname{Im} z|^2}$, for all i. Thus,

$$\left|\frac{1}{zN^2}\mathbb{E}\left[\operatorname{Tr} G^2\right]\right| = \left|\frac{1}{zN^2}\sum_{i=1}^{N}\frac{1}{(\lambda_i - z)^2}\right| \leq \frac{1}{N^2\,|z|\,|\operatorname{Im} z|^2} \xrightarrow{N\to\infty} 0.$$

We shall admit for this lecture that the contribution of $\mathbb{E}\left[\left(\frac{1}{N}\operatorname{Tr} G\right)^2\right]$ is the same as that of $\mathbb{E}\left[\frac{1}{N}\operatorname{Tr} G\right]^2$ and thus we shall replace it. So,

$$\frac{1}{N}\mathbb{E}\left[\operatorname{Tr} G_{A_N}(z)\right] = -\frac{1}{z} + \frac{1}{z}\mathbb{E}\left[\frac{1}{N}\operatorname{Tr} G\right]^2 + E_N,$$

where E_N is an error term that converges to 0 as $N \to \infty$. Thus,

$$1 + z\mathbb{E}\left[g_{A_N}(z)\right] + \mathbb{E}\left[g_{A_N}(z)\right]^2 \xrightarrow{N\to\infty} 0$$

for fixed $z \in \mathbb{C}_+$ and $g_N(z) = g_{A_N}(z)$ is contained in the closed ball of radius $\frac{1}{|\operatorname{Im} z|}$, which is compact. Then there exists a subsequence $g_{N_k}(z)$ that converges to a limit $g(z)$ satisfying

$$1 + zg(z) + g(z)^2 = 0.$$

This has the two solutions

$$g(z) = \frac{-z \pm \sqrt{z^2 - 4}}{2},$$

but as the limit is a Stieltjes transform we find

$$g(z) = \frac{-z + \sqrt{z^2 - 4}}{2}$$

and thus

$$\mathbb{E}\left[g_{A_{N_k}}(z)\right] \xrightarrow{N\to\infty} g(z).$$

Since by the same argument any subsequence of $g_N(z)$ must contain a further subsequence which converges to $g(z)$, the whole sequence must converge to $g(z)$. □

4.1 Universality of Wigner's semicircle law

Consider the selfadjoint $N \times N$-matrix

$$X_N = \frac{1}{\sqrt{N}} \begin{cases} X_{ij}, & j \leq i, \\ X_{ji}, & j > i \end{cases}$$

Theorem 4.7. *Let $(X_{ij})_{i\leq j\leq i\leq N}$ be a family of independently identically distributed random variables such that $\mathbb{E}[X_{11}] = 0$ and $\mathbb{E}[X_{11}^2] = \sigma^2$. Then $\mu_{X_n} \xrightarrow{w} \mu_\sigma$ almost surely for $N \to \infty$, where μ_σ is the semicircle distribution. In terms of the Stieltjes transform this means that for any $z \in \mathbb{C}_+$,*

$$g_{X_N}(z) \xrightarrow{N\to\infty} g_{\sigma^2}(u) \qquad \textit{almost surely.}$$

Remark 4.8. (i) Theorem 4.7 is a universality result in the sense that the limiting distribution does not depend on the distribution of the entries, since only the variance shows up in the limit. What really gives the semicircle distribution is the structure of the matrix and the independence of the entries (up to symmetry).

(ii) Recall that $\mathbb{E}\left[g_{A_N}(z)\right] \xrightarrow{N\to\infty} g_{\sigma^2}(z)$ for GOE. Then proving the weak version of Theorem 4.7 is equivalent to proving that for any $z \in \mathbb{C}_+$,

$$\left|\mathbb{E}\left[g_{X_N}(z)\right] - \mathbb{E}\left[g_{A_N}(z)\right]\right| \xrightarrow{N\to\infty} 0. \tag{$\star$}$$

Definition 4.9. Let A be an $N \times N$-matrix.

(i) The **spectral norm** of A is defined by

$$\|A\| = \max\left\{\sqrt{\lambda};\ \lambda \text{ is an eigenvalue of } A^*A\right\}.$$

If A is symmetric then

$$\|A\| = \max\left\{|\lambda|\,;\ \lambda \text{ is an eigenvalue of } A\right\}.$$

(ii) The **Hilbert-Schmidt norm** of A is given by

$$\|A\|_2 = \left(\sum_{i,j=1}^{N} |A_{ij}|^2\right)^{\frac{1}{2}}.$$

In addition to the usual norm properties, the Hilbert-Schmidt norm satisfies the following:

Proposition 4.10. *Let $A, B \in M_N(\mathbb{C})$.*

(i) $|\mathrm{Tr}(AB)| \le \|A\|_2 \|B\|_2$

(ii) *If U is a **unitary** matrix, i.e., $UU^* = U^*U = I$, then*

$$\|UA\|_2 = \|AU\|_2 = \|A\|_2.$$

(iii) *If B is a **normal** matrix, i.e., B admits a spectral decomposition, then*

$$\max\left\{\|AB\|_2, \|BA\|_2\right\} \le \|B\| \|A\|_2.$$

We will also use the following easy estimate.

Lemma 4.11. *Let A be a selfadjoint matrix and consider its resolvent given by $G_A(z) = (A - zI)^{-1}$ for $z \in \mathbb{C}\backslash\mathbb{R}$. Then*

$$\|G_A(z)\| \le \frac{1}{|\mathrm{Im}\, z|}.$$

In order to prove ($\star$) we shall use an approximation technique known as the Lindeberg method (replacement trick).

4.2 What is the Lindeberg method?

Let $(X_k)_{k\geq 1}$ and $(Y_k)_{k\geq 1}$ be two independent families of independently identically distributed random variables such that

$$\mathbb{E}[X_1] = \mathbb{E}[Y_1] = 0 \qquad \text{and} \qquad \mathbb{E}\left[X_1^2\right] = \mathbb{E}\left[Y_1^2\right] = \sigma^2 < \infty.$$

Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a three times differentiable function. Under what conditions can we approximate $\mathbb{E}[f(X_1,\dots,X_n)]$ by $\mathbb{E}[f(Y_1,\dots,Y_n)]$?
In other words: What is the order of

$$|\mathbb{E}[f(X_1,\dots,X_n)] - \mathbb{E}[f(Y_1,\dots,Y_n)]|?$$

To see that, we first need to introduce some notation. For any $i = 1,\dots,n$ and $c \in [0,1]$ let

$$Z_i = (X_1,\dots,X_i,Y_{i+1},\dots Y_n) \qquad \text{and} \qquad Z_i^c = (X_1,\dots,X_{i-1},cX_i,Y_{i+1},\dots Y_n).$$

Note that $Z_i^0 = (X_1,\dots,X_{i-1},0,Y_{i+1},\dots Y_n)$, $Z_n = (X_1,\dots,X_n)$ and $Z_0 = (Y_1,\dots,Y_n)$.
Now, calculate

$$\begin{aligned}\mathbb{E}[f(X_1,\dots,X_n)] - \mathbb{E}[f(Y_1,\dots,Y_n)] &= \mathbb{E}[f(Z_n)] - \mathbb{E}[f(Z_0)] \\ &= \sum_{i=1}^{n} \left(\mathbb{E}[f(Z_i)] - \mathbb{E}[f(Z_{i-1})]\right).\end{aligned}$$

We have written the difference as a telescopic sum of n terms and we shall control each one of them. Notice that the only difference between Z_i and Z_{i-1} is in the i-th component. We fix i and consider

$$\mathbb{E}[f(Z_i)] - \mathbb{E}[f(Z_{i-1})] = \mathbb{E}[f(Z_i)] - \mathbb{E}\left[f(Z_i^0)\right] - \left(\mathbb{E}[f(Z_{i-1})] - \mathbb{E}\left[f(Z_i^0)\right]\right).$$

By using the Taylor expansion with Lagrange's remainder for functions of several variables we get

$$f(Z_i) = f(Z_i^0) + X_i\partial_i f(Z_i^0) + \frac{X_i^2}{2!}\partial_i^2 f(Z_i^0) + \frac{X_i^3}{3!}\partial_i^3 f(Z_i^{c_1})$$

and

$$f(Z_{i-1}) = f(Z_i^0) + Y_i\partial_i f(Z_i^0) + \frac{Y_i^2}{2!}\partial_i^2 f(Z_i^0) + \frac{Y_i^3}{3!}\partial_i^3 f(Z_i^{c_2})$$

for some $c_1, c_2 \in (0, c)$. Hence

$$\begin{aligned}\mathbb{E}\left[f(Z_i)\right] - \mathbb{E}\left[f(Z_{i-1})\right] = \mathbb{E}\left[(X_i - Y_i)\partial_i f(Z_i^0)\right] + \frac{1}{2}\mathbb{E}\left[(X_i^2 - Y_i^2)\partial_i^2 f(Z_i^0)\right] \\ + \frac{1}{6}\mathbb{E}\left[X_i^3 \partial_i^3 f(Z_i^{c_1})\right] - \frac{1}{6}\mathbb{E}\left[Y_i^3 \partial_i^3 f(Z_i^{c_2})\right].\end{aligned}$$

First, by independence,

$$\begin{aligned}\mathbb{E}\left[(X_i - Y_i)\partial_i f(Z_i^0)\right] &= \mathbb{E}\left[(X_i - Y_i)\right]\mathbb{E}\left[\partial_i f(Z_i^0)\right] \\ &= \left(\mathbb{E}\left[X_i\right] - \mathbb{E}\left[Y_i\right]\right)\mathbb{E}\left[\partial_i f(Z_i^0)\right] \\ &= 0.\end{aligned}$$

For the term of second order we also have, by independence and the assumption about our variables,

$$\mathbb{E}\left[(X_i^2 - Y_i^2)\partial_i^2 f(Z_i^0)\right] = \mathbb{E}\left[(X_i^2 - Y_i^2)\right]\mathbb{E}\left[\partial_i^2 f(Z_i^0)\right] = 0.$$

Thus,

$$\left|\mathbb{E}\left[f(Z_i)\right] - \mathbb{E}\left[f(Z_{i-1})\right]\right| \leq \frac{1}{6}\mathbb{E}\left[\left|X_i^3 \partial_i^3 f(Z_i^{c_1})\right|\right] + \frac{1}{6}\mathbb{E}\left[\left|Y_i^3 \partial_i^3 f(Z_i^{c_2})\right|\right]$$

Assume that

$$\sup_{i=1,\ldots,n} \sup_{x\in\mathbb{R}^n} \left|\partial_i^3 f(x)\right| \leq L_3(f)$$

and

$$\max_{i=1,\ldots,n} \left\{\mathbb{E}\left[\left|X_i^3\right|\right], \mathbb{E}\left[\left|Y_i^3\right|\right]\right\} \leq K < \infty.$$

Then

$$\left|\mathbb{E}\left[f(Z_i)\right] - \mathbb{E}\left[f(Z_{i-1})\right]\right| \leq \frac{1}{3}K L_3(f)$$

such that

$$\left|\mathbb{E}\left[f(Z_n)\right] - \mathbb{E}\left[f(Z_0)\right]\right| \leq \frac{1}{3}K L_3(f) n.$$

Theorem 4.12. *Let $f\colon \mathbb{R}^n \to \mathbb{R}$ be a three times differentiable function such that*

$$\sup_{i=1,\ldots,n} \sup_{x\in\mathbb{R}^n} \left|\partial_i^3 f(x)\right| \leq L_3(f).$$

Let $(X_k)_{k\geq 1}$ and $(Y_k)_{k\geq 1}$ be two independent families of independently identically distributed random variables satisfying

$$\mathbb{E}\left[X_1\right] = \mathbb{E}\left[Y_1\right] = 0 \qquad \text{and} \qquad \mathbb{E}\left[X_1^2\right] = \mathbb{E}\left[Y_1^2\right] = \sigma^2 < \infty$$

and such that

$$\max_{i=1,\ldots,n}\left\{\mathbb{E}\left[\left|X_i^3\right|\right],\mathbb{E}\left[\left|Y_i^3\right|\right]\right\}\leq K<\infty. \tag{C1}$$

Then

$$\left|\mathbb{E}\left[f(X_1,\ldots,X_n)\right]-\mathbb{E}\left[f(Y_1,\ldots,Y_n)\right]\right|\leq\frac{1}{3}KL_3(f)n.$$

We now want to apply this theorem to the Stieltjes transform. For any $z\in\mathbb{C}_+$, let

$$h_z\colon\mathbb{R}^{\frac{N(N+1)}{2}}\to\mathbb{C}_+,\ x=(x_{ij})_{1\leq j\leq i\leq N}\mapsto\frac{1}{N}\operatorname{Tr}G_z(x),$$

where $G_z(x)=(A(x)-zI)^{-1}$ with

$$[A(x)]_{ij}=\frac{1}{N}\begin{cases}x_{ij}, & j\leq i,\\ x_{ji}, & j>i.\end{cases}$$

We will see in the exercises that for some constant C_z depending on z,

$$L_3(h_z)\leq C_z\frac{1}{N^{\frac{5}{2}}}.$$

Applying Theorem 4.12 to the real and imaginary parts of h_z gives for all $z\in\mathbb{C}_+$,

$$\begin{aligned}\left|\mathbb{E}\left[g_{X_N}(z)\right]-\mathbb{E}\left[g_{A_n}(z)\right]\right|&=\left|\mathbb{E}\left[h_z\left((x_{ij})_{1\leq j\leq i\leq N}\right)\right]-\mathbb{E}\left[h_z\left((a_{ij})_{1\leq j\leq i\leq N}\right)\right]\right|\\&\leq\frac{1}{3}\frac{C_z}{N^{\frac{5}{2}}}L_3(f)\frac{N(N+1)}{2}\\&\leq\frac{1}{3}C_z\frac{1}{\sqrt{N}}\xrightarrow{N\to\infty}0,\end{aligned}$$

thus completing the proof.

Remark 4.13. We shall see how one can truncate the matrix entries and consider bounded random variables such that (C1) is satisfied.

5 Concentration inequalities

5.1 Preliminaries

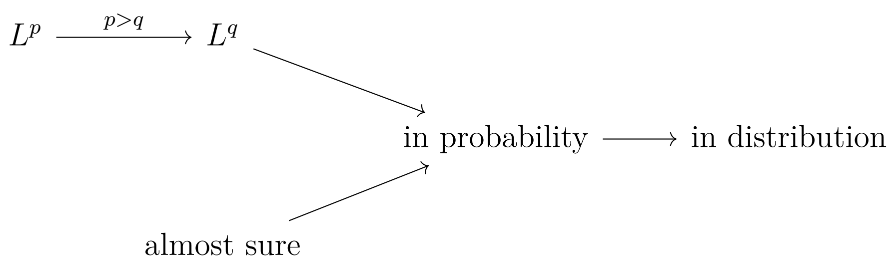
Recall the following definitions:

- A **probability space** $(\Omega, \mathcal{F}, \mathbb{P})$ is a measure space with $\mathbb{P}$ being a finite measure of mass 1. We interpret $A \in \mathcal{F}$ as events and $\mathbb{P}[A]$ as the probability that A happens.
- Measurable functions $X\colon (\Omega, \mathcal{F}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ are called **random variables**.
- We say that $A \in \mathcal{F}$ holds **with probability 1** or **almost surely** if $\mathbb{P}[A] = 1$.

Let $(X_n)_{n\geq 1}$ be a sequence of random variables. Then we consider the following modes of convergence:

- **Pointwise convergence**: $X_n(w) \to X(w)$ for any $w \in \Omega$.
- **Almost sure convergence**: There is $A \in \mathcal{F}$ with $\mathbb{P}[A] = 1$ such that $X_n(w) \to X(w)$ for all $w \in A$.
- **Convergence in probability**: For all $\varepsilon > 0$, $\mathbb{P}[|X_n - X| > \varepsilon] \to 0$.
- **L^p-convergence**: $X_n, X \in L^p$ and $\mathbb{E}[|X_n - X|^p] \to 0$.
- **Convergence in distribution**: $\mathbb{P}[X_n \leq x] \to \mathbb{P}[X \leq x]$ for all points x where $x \mapsto \mathbb{P}[X \leq x]$ is continuous.

  (This is equivalent to weak convergence.)



Proposition 5.1 (Criterion for almost sure convergence)**.** *$X_n \to X$ almost surely if and only if for every $\varepsilon > 0$*

$$\mathbb{P}\left[\sup_{k\geq n} |X_k - X| \geq \varepsilon\right] \xrightarrow{n\to\infty} 0.$$

Note that the event $\{\sup_{k\geq n} |X_k - X| \geq \varepsilon\}$ is the $\limsup$ of the events $E_n = \{|X_n - X| \geq \varepsilon\}$; recall the general definition of the $\limsup$ for sets as those elements which are in infinitely many of the sets, i.e.,

$$\limsup_n E_n = \bigcap_n \bigcup_{k\geq n} E_k.$$

Theorem 5.2 (Borel-Cantelli lemma)**.** *Let $(E_n)_n$ be a sequence of events in Ω.*

(i) *(The first Borel-Cantelli lemma) If $\sum_{n=1}^{\infty} \mathbb{P}[E_n] < \infty$ then $\mathbb{P}[\limsup_n E_n] = 0$.*

(ii) *(The second Borel-Cantelli lemma) If $\sum_{n=1}^{\infty} \mathbb{P}[E_n] = \infty$ and the $(E_n)_n$ are independent then $\mathbb{P}[\limsup_n E_n] = 1$.*

We would like to prove that for any $z \in \mathbb{C}_+$ and $\varepsilon > 0$,

$$\sum_{n=1}^{\infty} \mathbb{P}\left[\left|g_{X_n}(z) - \mathbb{E}\left[g_{X_n(z)}\right]\right| \geq \varepsilon\right] < \infty$$

and then, using the first Borel-Cantelli lemma and Proposition 5.1, we prove that for all $z \in \mathbb{C}_+$,

$$\left|g_{X_n}(z) - \mathbb{E}\left[g_{X_n(z)}\right]\right| \to 0 \qquad \text{almost surely.}$$

To do this, we need to find for each $n \geq 1$ upper bounds on $\mathbb{P}\left[\left|g_{X_n}(z) - \mathbb{E}\left[g_{X_n(z)}\right]\right| \geq \varepsilon\right]$.

5.2 Concentration phenomena

The basic example for mass concentration is the Gaussian measure. As we can see, the mass is concentrated around the mean 0 and decreases rapidly away from 0. However, this is not true for all distributions.

Example 5.3. Let $X \sim \text{Bernoulli}(\frac{1}{2})$, that is, $X \in \{0, 1\}$ and $\mathbb{P}[X = 0] = \mathbb{P}[X = 1] = \frac{1}{2}$. Is this case,

$$\mathbb{E}[X] = 0 \cdot \mathbb{P}[X = 0] + 1 \cdot \mathbb{P}[X = 1] = \frac{1}{2}.$$

Theorem 5.4 (Strong law of large numbers). *If* $(X_n)_{n\geq 1}$ *is a sequence of independently identically distributed random variables such that* $\mathbb{E}[|X_1|] < \infty$, *then*

$$\frac{1}{n}\sum_{i=1}^{n} X_i \to \mathbb{E}[X_1] \qquad \textit{almost surely.}$$

Remark 5.5. Let $\mu_N = \frac{1}{N}\sum_{i=1}^{N}\delta_{\lambda_i}$, where $\lambda_1, \ldots, \lambda_N$ are the eigenvalues of a GOE matrix. Then Wigner's semicircle law tells us that

$$\mu_N \xrightarrow{w} \mu \qquad \text{almost surely.}$$

μ_N is a random empirical measure converging almost surely to a deterministic measure. It has the same flavor as the law of large numbers.

Proposition 5.6 (Markov's inequality). *Let* X *be a random variable taking non-negative values. Then, for any* $t > 0$,

$$\mathbb{P}[X \geq t] \leq \frac{\mathbb{E}[X]}{t}.$$

Proof. As $X \geq 0$,

$$\begin{aligned}\mathbb{E}[X] &= \int_0^\infty x\,\mathrm{d}X(x) = \int_0^t x\,\mathrm{d}X(x) + \int_t^\infty x\,\mathrm{d}X(x)\\ &\geq \int_t^\infty x\,\mathrm{d}X(x) \geq t\int_t^\infty \mathrm{d}X(x) = t\int_0^\infty 1_{\{X\geq t\}}\,\mathrm{d}X(x) = t\mathbb{P}[X \geq t].\end{aligned}$$

□

Note that for any $p \geq 1$,

$$\mathbb{P}[|X| \geq t] = \mathbb{P}\left[|X|^P \geq t^p\right] \leq \frac{\mathbb{E}[|X|^p]}{t^p},$$

provided that $\mathbb{E}[|X|^p] < \infty$.

Proposition 5.7 (Chebyshev's inequality). *Let* X *be a random variable with finite mean* μ *and variance* σ^2. *Then, for any* $t > 0$,

$$\mathbb{P}[|X - \mu| \geq t] \leq \frac{\sigma^2}{t^2}.$$

Proof. We have

$$\mathbb{P}\left[|X-\mu| \geq t\right] = \mathbb{P}\left[|X-\mu|^2 \geq t^2\right] \leq \frac{\mathbb{E}\left[|X-\mu|^2\right]}{t^2} = \frac{\mathbb{V}\mathrm{ar}\left[X\right]}{t^2} = \frac{\sigma^2}{t^2}.$$

□

Example 5.8. Suppose that a post office handles on average 10000 letters a day. The probability that it will handle at least 15000 letters the next day is

$$\mathbb{P}\left[X \geq 15000\right] \leq \frac{\mathbb{E}\left[X\right]}{15000} = \frac{10000}{15000} = \frac{2}{3}.$$

However, if we know that the variance is 2000 letters, then

$$\begin{aligned}\mathbb{P}\left[X \geq 15000\right] = \mathbb{P}\left[X - 10000 \geq 5000\right] &\leq \mathbb{P}\left[|X - 10000| \geq 5000\right] \\ &\leq \frac{\mathbb{V}\mathrm{ar}\left[X\right]}{(5000)^2} = \frac{1}{12500} << \frac{2}{3}.\end{aligned}$$

We can see in the above example that if the random variable possesses more moments, then the tail inequalities can be improved to get better bounds for the tail probabilities.

Definition 5.9. Let X be a random variable. Its moment generating function M_X is given by

$$M_X \colon \mathbb{R} \to \mathbb{R},\ M_X(\lambda) = \mathbb{E}\left[e^{\lambda X}\right].$$

Remark 5.10. We have the following:

- $M_X(0) = \mathbb{E}\left[e^0\right] = 1$.
- $M'_X(\lambda) = \mathbb{E}\left[Xe^{\lambda X}\right]$, thus $M'_X(0) = \mathbb{E}\left[X\right]$.
- $M^{(k)}_X(\lambda) = \mathbb{E}\left[X^k e^{\lambda X}\right]$, thus $M^{(k)}_X(0) = \mathbb{E}\left[X^k\right]$.

Example 5.11. Let $X \sim \mathcal{N}(0,1)$. Then

$$\begin{aligned} M_X(\lambda) = \mathbb{E}\left[e^{\lambda X}\right] &= \int_{-\infty}^{\infty} e^{\lambda x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, \mathrm{d}x \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x^2 - 2\lambda x + \lambda^2)} e^{-\frac{\lambda^2}{2}} \, \mathrm{d}x \\ &= e^{\frac{\lambda^2}{2}} \underbrace{\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-\frac{1}{2}(x-\lambda)^2} \, \mathrm{d}x}_{=1} \\ &= e^{\frac{\lambda^2}{2}}. \end{aligned}$$

Lemma 5.12 (Chernoff bound)**.** *Let X be a real random variable. Then, for all $t \in \mathbb{R}$,*

$$\mathbb{P}\left[X \geq t\right] \leq \inf_{\lambda > 0} \left\{ e^{-\lambda t} M_X(\lambda) \right\}.$$

Proof. Let $\lambda > 0$, then

$$\mathbb{P}\left[X \geq t\right] = \mathbb{P}\left[\lambda X \geq \lambda t\right] = \mathbb{P}\left[e^{\lambda X} \geq e^{\lambda t}\right] \leq \frac{\mathbb{E}\left[e^{\lambda X}\right]}{e^{\lambda t}} = e^{-\lambda t} M_X(\lambda),$$

such that taking the infimum concludes the proof. □

Remark 5.13. If $X \sim \mathcal{N}(0,1)$ then

$$\mathbb{P}\left[X \geq t\right] \leq \inf_{\lambda > 0} \left\{ e^{-\lambda t} e^{\frac{\lambda^2}{2}} \right\} = e^{-\frac{t^2}{2}}.$$

The Chernoff bound captures the correct behaviour of Gaussian tails.

Lemma 5.14 (Hoeffding's lemma)**.** *Let X be a real centered random variable with values in $[a,b]$ almost surely. Then for any $\lambda > 0$,*

$$M_X(\lambda) \leq e^{\frac{\lambda^2 (b-a)^2}{8}}.$$

Proof. Consider the convex function

$$f : x \mapsto e^{\lambda x}.$$

Recall that convex means that for any $\alpha \in (0,1)$ and all y, z,

$$f(\alpha y + (1-\alpha)z) \leq \alpha f(y) + (1-\alpha) f(z).$$

Fix $x \in [a,b]$ and set $\alpha = \frac{b-x}{b-a}$. Then

$$\begin{aligned}\lambda x &= \frac{b-x}{b-a}\lambda a + \frac{x-a}{b-a}\lambda b \\ &= \alpha(\lambda a) + (1-\alpha)(\lambda b)\end{aligned}$$

such that

$$e^{\lambda X} \leq \frac{b-X}{b-a} e^{\lambda a} + \frac{X-a}{b-a} e^{\lambda b}.$$

Notice that

$$0 \leq \mathbb{E}\left[e^{\lambda X}\right] \leq \frac{b}{b-a} e^{\lambda a} - \frac{a}{b-a} e^{\lambda b} = g(\lambda) = e^{\log g(\lambda)},$$

since $\mathbb{E}[X] = 0$. Now put $p = \frac{-a}{b-a}$ and $u = (b-a)\lambda$. Consider the function

$$\begin{aligned}\varphi(u) &= \log g(\lambda) \\ &= \log\left[(1-p)e^{\lambda a} + p e^{\lambda b}\right] \\ &= \log\left[\left(1 - p + p e^{\lambda(b-a)}\right) e^{\lambda a}\right] \\ &= \lambda a + \log\left[1 - p + p e^{u}\right] \\ &= -pu + \log\left[1 - p + p e^{u}\right].\end{aligned}$$

This function is smooth and thus, by Taylor's theorem, for all $u \in \mathbb{R}$ there exists $v \in \mathbb{R}$ such that

$$\varphi(u) = \varphi(0) + u\varphi'(0) + \frac{u^2}{2}\varphi''(v).$$

Now calculate $\varphi(0) = 0$,

$$\varphi'(u) = -p + \frac{p e^u}{1 - p + p e^u}$$

such that $\varphi'(0) = -p + p = 0$ and

$$\varphi''(u) = \frac{(1-p) p e^u}{(1 - p + p e^u)^2} \leq \frac{1}{4}.$$

Thus, $\varphi(u) \leq \frac{1}{8}u^2$, such that

$$M_X(\lambda) \leq e^{\varphi(u)} \leq e^{\frac{1}{8}u^2} = e^{\frac{\lambda^2 (b-a)^2}{8}}.$$

□

Theorem 5.15 (Hoeffding's inequality)**.** *Let $X_1, \ldots, X_n$ be independent real random variables such that $X_i \in [a_i, b_i]$ almost surely for all $i = 1, \ldots, n$. Then, for any $t > 0$:*

(i)
$$\mathbb{P}\left[\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \geq t\right] \leq \exp\left(\frac{-2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

(ii)
$$\mathbb{P}\left[\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i]) \leq -t\right] \leq \exp\left(\frac{-2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

(iii)
$$\mathbb{P}\left[\left|\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])\right| \geq t\right] \leq 2\exp\left(\frac{-2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

Proof. Set $S_n = \sum_{i=1}^{n} X_i$. Then, by the Chernoff bound applied to $S_n - \mathbb{E}[S_n]$ we get

$$\mathbb{P}[S_n - \mathbb{E}[S_n] \geq t] \leq \inf_{\lambda > 0}\left\{e^{-\lambda t}\mathbb{E}\left[e^{\lambda(S_n - \mathbb{E}[S_n])}\right]\right\}.$$

Calculate, by Hoeffding's lemma,

$$\begin{aligned}
\mathbb{E}\left[e^{\lambda(S_n - \mathbb{E}[S_n])}\right] &= \mathbb{E}\left[e^{\lambda\sum_{i=1}^{n}(X_i - \mathbb{E}[X_i])}\right] \\
&= \mathbb{E}\left[\prod_{i=1}^{n} e^{\lambda(X_i - \mathbb{E}[X_i])}\right] \\
&= \prod_{i=1}^{n}\mathbb{E}\left[e^{\lambda(X_i - \mathbb{E}[X_i])}\right] \\
&= \prod_{i=1}^{n} M_{X_i - \mathbb{E}[X_i]}(\lambda) \\
&\leq \prod_{i=1}^{n} e^{\frac{\lambda^2(b_i - a_i)^2}{8}} \\
&= e^{\frac{\lambda^2}{8}\sum_{i=1}^{n}(b_i - a_i)^2}.
\end{aligned}$$

Plugging this in the Chernoff bound yields

$$\mathbb{P}[S_n - \mathbb{E}[S_n] \geq t] \leq \inf_{\lambda > 0} g(\lambda),$$

where

$$g(\lambda) = \exp\left(-\lambda t + \frac{\lambda^2}{8}\sum_i (b_i - a_i)^2\right).$$

Now, $g'(\lambda) = 0$ implies

$$\lambda = \frac{4t}{\sum_{i=1}^n (b_i - a_i)^2}.$$

Therefore

$$\mathbb{P}\left[S_n - \mathbb{E}\left[S_n\right] \geq t\right] \leq \exp\left(\frac{-2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).$$

To obtain the second inequality, it suffices to apply the first bound to $-X_1, \ldots, -X_n$. Finally, the third inequality follows from

$$\begin{aligned}\mathbb{P}\left[S_n - \mathbb{E}\left[S_n\right] \geq t\right] &= \mathbb{P}\left[\{S_n - \mathbb{E}\left[S_n\right] \geq t\} \cup \{S_n - \mathbb{E}\left[S_n\right] \leq -t\}\right] \\ &\leq \mathbb{P}\left[S_n - \mathbb{E}\left[S_n\right] \geq t\right] + \mathbb{P}\left[S_n - \mathbb{E}\left[S_n\right] \leq -t\right] \\ &\leq 2\exp\left(\frac{-2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right).\end{aligned}$$

□

Example 5.16. Let $X \sim \text{Bernoulli}(p)$, that is, $X \in \{0, 1\}$ and $\mathbb{P}\left[X = 1\right] = p$ and $\mathbb{P}\left[X = 0\right] = 1 - p$. Is this case, $\mathbb{E}\left[X\right] = p$ and $\mathbb{V}\text{ar}\left[X\right] = p(1 - p)$. Let $\overline{X}_n = \frac{1}{n}\sum_{i=1}^n X_i$, where the X_i are independent copies of X. Then $\mathbb{E}\left[\overline{X}_n\right] = p$ and

$$\mathbb{V}\text{ar}\left[\overline{X}_n\right] = \mathbb{V}\text{ar}\left[\frac{1}{n}\sum_{i=1}^n X_i\right] = \frac{1}{n^2}\sum_{i=1}^n \mathbb{V}\text{ar}\left[X_i\right] = \frac{1}{n}p(1 - p).$$

For $t > 0$ Chebyshev's inequality gives

$$\mathbb{P}\left[\left|\overline{X}_n - p\right| \geq t\right] \leq \frac{p(1 - p)}{nt^2} \leq \frac{1}{4nt^2}$$

for any $p \in (0, 1)$. In particular, for $t = 0.2$ and $n = 100$,

$$\mathbb{P}\left[\left|\overline{X}_n - p\right| \geq t\right] \leq 0.0625.$$

On the other hand, Hoeffding's inequality gives

$$\begin{aligned}\mathbb{P}\left[\left|\overline{X}_n - p\right| \geq t\right] &\leq 2\mathbb{P}\left[\frac{1}{n}\left|\sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right| \geq t\right] \\ &= 2\mathbb{P}\left[\left|\sum_{i=1}^n (X_i - \mathbb{E}[X_i])\right| \geq nt\right] \\ &\leq 2\exp\left(-\frac{2n^2t^2}{\sum_{i=1}^n (b_i - a_i)^2}\right) \\ &= 2\exp\left(-2nt^2\right).\end{aligned}$$

In particular, for $t = 0.2$ and $n = 100$,

$$\mathbb{P}\left[\left|\overline{X}_n - p\right| \geq t\right] \leq 0.000067,$$

which is almost 100 times smaller than the bound obtained by Chebyshev.

Remark 5.17. We know, by the law of large numbers, that

$$\frac{1}{n}\sum_{i=1}^n X_i \to \mathbb{E}[X_i] = p \qquad \text{almost surely.}$$

By Chebyshev,

$$\sum_{n=1}^\infty \mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^n X_i - p\right| \geq t\right] \leq \sum_{n=1}^\infty \frac{1}{4nt^2} = \infty,$$

which does not give bounds good enough to prove that the above series is finite. However, by Hoeffding,

$$\sum_{n=1}^\infty \mathbb{P}\left[\left|\frac{1}{n}\sum_{i=1}^n X_i - p\right| \geq t\right] \leq 2\sum_{n=1}^\infty e^{-2nt^2} = 2\left(\frac{1}{1 - e^{-2t^2}} - 1\right) < \infty.$$

The first Borel-Cantelli lemma then ascertains the almost sure convergence in the law of large numbers.

Theorem 5.18 (Azuma-Hoeffding inequality, McDiarmid's inequality)**.** *Let $(X_n)_{n\geq 1}$ be a family of independent random variables and let $H\colon \mathbb{R}^n \to \mathbb{R}$ be a measurable function such that the random variable $H(X_1, \ldots, X_n)$ is integrable. Then*

$$\mathbb{P}\left[\,|H(X_1, \ldots, X_n) - \mathbb{E}[H(X_1, \ldots, X_n)]| \geq t\right] \leq 2\exp\left(\frac{-2t^2}{\sum_{k=1}^n c_k^2}\right),$$

where

$$c_k = \sup_{x, x', x_i \in \mathbb{R}} \left|H(x_1, \ldots, x_{k-1}, x, x_{k+1}, \ldots, x_n) - H(x_1, \ldots, x_{k-1}, x', x_{k+1}, \ldots, x_n)\right|.$$

We want to apply this to the Stieltjes transform of the Wigner matrix $X_N = \left(\frac{1}{\sqrt{N}} X_{ij}\right)$, where the X_{ij} are independently identically distributed with $\mathbb{E}\left[X_{ij}\right] = 0$.

Theorem 5.19 (Rank theorem)**.** *Let $A, B \in M_N(\mathbb{C})$ be Hermitian. Then*

$$\|F_A - F_B\|_\infty \leq \frac{1}{N} \operatorname{rank}(A - B),$$

where, for any matrix M with eigenvalues $\lambda_1, \ldots, \lambda_N$,

$$F_M(x) = \frac{1}{N} \sum_{i=1}^{N} 1_{\{\lambda_i \leq x\}}$$

Let $f \colon \mathbb{R} \to \mathbb{R}$ be integrable. By integration of parts and the rank theorem,

$$\begin{aligned} \left| \int f \, \mathrm{d}\mu_A - \int f \, \mathrm{d}\mu_B \right| &= \left| \int f'(t) \left(F_A(t) - F_B(t)\right) \, \mathrm{d}t \right| \\ &\leq \frac{\operatorname{rank}(A - B)}{N} \int |f'(t)| \, \mathrm{d}t. \end{aligned}$$

As before, write X_N as a function of its entries $X_N = A\left(x_1, \ldots, x_N\right)$, where $x_1, x_2, \ldots, x_N$ are the rows of the lower triangular matrix, i.e., they are vectors in $\mathbb{C}, \mathbb{C}^2, \ldots, \mathbb{C}^N$. Then

$$\mu_{X_N} = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i(A(x_1, \ldots, x_N))} = \mu_{A(x_1, \ldots, x_N)}$$

and for the Stieltjes transform we have

$$g_z(x_1, \ldots, x_N) = \int \frac{1}{\lambda - z} \, \mathrm{d}\mu_{A(x_1, \ldots, x_N)}(\lambda).$$

We let $y = (x_1, \ldots, x_N)$ and $y' = (x_1, \ldots, x_{k-1}, x', x_{k+1}, \ldots, x_N)$, where the only difference between y and y' is their k-th component, i.e., the k-th row of the lower triangular matrix. Then

$$|g_z(y) - g_z(y')| \leq \frac{\operatorname{rank}\left(A(y) - A(y')\right)}{N} \|f'\|_{L_1} \leq \frac{2}{N} \|f'\|_{L_1}.$$

We apply the Azuma-Hoeffding inequality to the real and to the imaginary part of g_{X_N} and can easily check that

$$\left\| \left(\operatorname{Re} \frac{1}{\lambda - z} \right)' \right\|_{L_1} = \left\| \left(\operatorname{Im} \frac{1}{\lambda - z} \right)' \right\|_{L_1} \leq \frac{2}{\operatorname{Im} z},$$

thus, for all $k = 1, \dots, N$,

$$c_k \leq \frac{4}{N \operatorname{Im} z}.$$

Hence we obtain

$$\mathbb{P}\left[|g_{X_N}(z) - \mathbb{E}\left[g_{X_N}(z)\right]| \geq t\right] \leq 4 \exp\left(\frac{-2t^2}{\sum\limits_{k=1}^{N} \frac{16}{(\operatorname{Im} z)^2 N^2}}\right) \leq 4 \exp\left(-\frac{Nt^2}{8}(\operatorname{Im} z)^2\right).$$

Set $t = N^{-\frac{1}{2}+\varepsilon}$ for some $\varepsilon > 0$ such that

$$\sum_{N \geq 1} \mathbb{P}\left[|g_{X_N}(z) - \mathbb{E}\left[g_{X_N}(z)\right]| \geq t\right] < \infty.$$

Therefore, by Borel-Cantelli,

$$|g_{X_N}(z) - \mathbb{E}\left[g_{X_N}(z)\right]| \to 0 \qquad \text{almost surely.}$$

6 Analytic description of the eigenvalue distribution of Gaussian random matrices

In Problem 8 we showed that the joint distribution of the entries $a_{ij} = x_{ij} + y_{ij}$ of a GUE $A = (a_{ij})_{i,j=1}^N$ has density

$$c \exp\left(-\frac{N}{2} \operatorname{Tr} A^2\right) \mathrm{d}A.$$

This clearly shows the invariance of the distribution under unitary transformations: Let U be a unitary $N \times N$-matrix and let $B = U^* A U = (b_{ij})_{i,j=1}^N$. Then we have $\operatorname{Tr} B^2 = \operatorname{Tr} A^2$ and the volume element is invariant under unitary transformations, $\mathrm{d}B = \mathrm{d}A$. Therefore, for the joint distributions of eigenvalues of A and B respectively,

$$c \exp\left(-\frac{N}{2} \operatorname{Tr} B^2\right) \mathrm{d}B = c \exp\left(-\frac{N}{2} \operatorname{Tr} A^2\right) \mathrm{d}A.$$

Thus the joint distribution of entries of a GUE is invariant under unitary transformations, which explains the name **G**aussian **U**nitary **E**nsemble. We would like to transform this density from entries to eigenvalues. Instead of GUE, we will consider the real case, i.e., GOE.

Definition 6.1 (GOE matrices)**.** A **Gaussian orthogonal random matrix (GOE)** $A = (x_{ij})_{i,j=1}^N$ is given by real-valued entries x_{ij} with $x_{ij} = x_{ji}$ for all $i, j = 1, \dots, N$ and joint distribution

$$c_N \exp\left(-\frac{N}{4} \operatorname{Tr} A^2\right) \mathrm{d}A,$$

where

$$\mathrm{d}A = \prod_{i \geq j} \mathrm{d}x_{ij}.$$

Remark 6.2. (i) This is clearly invariant under orthogonal transformation of the entries.

(ii) This is equivalent to independent Gaussian random variables. But the variance for the diagonal entries has to be chosen differently from the off-diagonals. Consider the example $N = 2$ with

$$A = \begin{pmatrix} x_{11} & x_{12} \\ x_{12} & x_{22}. \end{pmatrix}$$

Then

$$\begin{aligned} \exp\left(-\frac{N}{4}\operatorname{Tr}\begin{pmatrix} x_{11} & x_{12} \\ x_{12} & x_{22}. \end{pmatrix}^2\right) &= \exp\left(-\frac{N}{4}\left(x_{11}^2 + 2x_{12}^2 + x_{22}^2\right)\right) \\ &= \exp\left(-\frac{N}{4}x_{11}^2\right)\exp\left(-\frac{N}{2}x_{12}^2\right)\exp\left(-\frac{N}{4}x_{22}^2\right). \end{aligned}$$

(iii) From this one can easily determine the normalization constant c_N (as a function of N).

Since we usually are interested in functions of the eigenvalues, we will now transform this density to eigenvalues.

Example 6.3. As a warmup, let us consider the GOE for $N = 2$

$$A = \begin{pmatrix} x_{11} & x_{12} \\ x_{12} & x_{22} \end{pmatrix}$$

with density

$$p(A) = c_2 \exp\left(-\frac{N}{4}\operatorname{Tr} A^2\right).$$

We parametrize A by its eigenvalues λ_1 and λ_2 and an angle θ by diagonalization $A = O^T D O$, where

$$D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} \qquad \text{and} \qquad O = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

I.e.,

$$\begin{aligned} x_{11} &= \lambda_1 \cos^2\theta + \lambda_2 \sin^2\theta, \\ x_{12} &= (\lambda_1 - \lambda_2)\cos\theta\sin\theta, \\ x_{22} &= \lambda_1 \sin^2\theta + \lambda_2 \cos^2\theta. \end{aligned}$$

Note that O and D are not uniquely determined by A. In particular, if $\lambda_1 = \lambda_2$ then any orthogonal O works. However, this case has probability zero and thus can be ignored (comes later). If $\lambda_1 < \lambda_2$ then O contains the normalized eigenvectors for λ_1 and λ_2. Those are unique up to a sign, which can be fixed by requiring that $\cos\theta \geq 0$. Hence θ is not running from $-\pi$ to π, but instead it can be restricted to $\left[-\frac{\pi}{2}, \frac{\pi}{2}\right]$. We will now transform

$$p(x_{11}, x_{22}, x_{12})\,\mathrm{d}x_{11}\,\mathrm{d}x_{22}\,\mathrm{d}x_{12} \to q(\lambda_1, \lambda_2, \theta)\,\mathrm{d}\lambda_1\,\mathrm{d}\lambda_2\,\mathrm{d}\theta$$

by the change of variable formula

$$q = p\,|\det \mathrm{D}F|\,,$$

where $\mathrm{D}F$ is the Jacobian of

$$F\colon (x_{11}, x_{22}, x_{12}) \mapsto (\lambda_1, \lambda_2, \theta).$$

We calculate

$$\det \mathrm{D}F = \det \begin{pmatrix} \cos^2\theta & \sin^2\theta & -2(\lambda_1 - \lambda_2)\sin\theta\cos\theta \\ \cos\theta\sin\theta & -\cos\theta\sin\theta & (\lambda_1 - \lambda_2)\left(-\sin^2\theta\cos^2\theta\right) \\ \sin^2\theta & \cos^2\theta & 2(\lambda_1 - \lambda_2)\sin\theta\cos\theta \end{pmatrix} = -(\lambda_1 - \lambda_2)$$

such that $|\det \mathrm{D}F| = |\lambda_1 - \lambda_2|$. Thus,

$$q(\lambda_1, \lambda_2, \theta) = c_2 e^{-\frac{N}{4}(\lambda_1^2 + \lambda_2^2)}\,|\lambda_1 - \lambda_2|\,.$$

Note that q is independent of θ, i.e., we have a uniform distribution in θ. Consider a function $f = f(\lambda_1, \lambda_2)$ of the eigenvalues. Then

$$\begin{aligned} \mathbb{E}\left[f(\lambda_1, \lambda_2)\right] &= \int\int\int q(\lambda_1, \lambda_2, \theta) f(\lambda_1, \lambda_2)\,\mathrm{d}\lambda_1\,\mathrm{d}\lambda_2\,\mathrm{d}\theta \\ &= \int_{-\frac{\pi}{2}}^{\frac{\pi}{2}} \iint_{\lambda_1 < \lambda_2} f(\lambda_1, \lambda_2) c_2 e^{-\frac{N}{4}(\lambda_1^2 + \lambda_2^2)}\,|\lambda_1 - \lambda_2|\,\mathrm{d}\lambda_1\,\mathrm{d}\lambda_2\,\mathrm{d}\theta \\ &= \pi c_2 \iint_{\lambda_1 < \lambda_2} f(\lambda_1, \lambda_2) e^{-\frac{N}{4}(\lambda_1^2 + \lambda_2^2)}\,|\lambda_1 - \lambda_2|\,\mathrm{d}\lambda_1\,\mathrm{d}\lambda_2. \end{aligned}$$

Thus, the density for the joint distribution of the eigenvalues on $\{(\lambda_1, \lambda_2);\, \lambda_1 < \lambda_2\}$ is given by

$$\tilde{c}_2 e^{-\frac{N}{4}(\lambda_1^2 + \lambda_2^2)}\,|\lambda_1 - \lambda_2|$$

with $\tilde{c}_2 = \pi c_2$.

Remark 6.4. Let us check that the probability of $\lambda_1 = \lambda_2$ is zero.
λ_1, λ_2 are the solutions of the characteristic equation

$$\begin{aligned} 0 = \det(\lambda I - A) &= (\lambda - x_{11})(\lambda - x_{22}) - x_{12}^2 \\ &= \lambda^2 - (x_{11} + x_{22})\lambda + (x_{11}x_{22} - x_{12}^2) \\ &= \lambda^2 - b\lambda + c. \end{aligned}$$

Then there is only one solution if and only if the discriminant $d = b^2 - 4ac$ is zero. However,

$$\{(x_{11}, x_{22}, x_{12});\ d(x_{11}, x_{22}, x_{12}) = 0\}$$

is a two-dimensional surface in $\mathbb{R}^3$, i.e., its Lebesgue measure is zero.

Now we consider general GOE(N).

Theorem 6.5. *The joint distribution of the eigenvalues of a GOE(N) is given by a density*

$$\tilde{c}_N e^{-\frac{N}{4}(\lambda_1^2 + \cdots + \lambda_N^2)} \prod_{k<l} (\lambda_l - \lambda_k)$$

restricted on $\lambda_1 < \cdots < \lambda_N$.

Proof. In terms of the entries of the GOE matrix A we have density

$$p\,(x_{kl} \mid k \geq l) = c_N e^{-\frac{N}{4}\operatorname{Tr} A^2},$$

where $A = (x_{kl})_{k,l=1}^N$ with x_{kl} real independent Gaussians and $x_{kl} = x_{lk}$ for all l, k. Again we diagonalize $A = O^T D O$ with O diagonal and $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_N)$ with $\lambda_1 \leq \cdots \leq \lambda_N$. As before, degenerated eigenvalues have probability zero, hence this case can be neglected and we assume $\lambda_1 < \cdots < \lambda_N$. We parametrize O via $O = e^{-H}$ by a **skew-symmetric** matrix H, that is, $H^T = -H$, i.e., $H = (h_{ij})_{i,j=1}^N$ with $h_{ij} \in \mathbb{R}$ and $h_{ij} = -h_{ji}$ for all i, j. In particular, $h_{ii} = 0$ for all i. We have

$$O^T = \left(e^{-H}\right)^T = e^{-H^T} = e^H$$

and thus

$$O^T O = e^H e^{-H} = e^{H-H} = e^0 = I = O O^T.$$

Note that $A = e^H D e^{-H}$. For A we have the variables

$$\{x_{ij};\ j \leq i\}$$

and for $e^H D e^{-H}$ we have

$$\{\lambda_1, \dots, \lambda_N\} \cup \{h_{ij};\, i > j\}.$$

We thus see that the parametrization has the right number of parameters. This parametrization is locally bijective ($O = e^{-H}$ is a parametrization of the Lie group $\mathrm{SO}(N)$ by the Lie algebra $\mathrm{so}(N)$ of skew-symmetric matrices). So we need to compute the Jacobian of the map

$$S\colon A \mapsto e^H D e^{-H}.$$

We have

$$\begin{aligned} \mathrm{d}A &= (\mathrm{d}e^H) D e^{-H} + e^H (\mathrm{d}D) e^{-H} + e^H D (\mathrm{d}e^{-H}) \\ &= e^H \left[e^{-H} (\mathrm{d}e^H) D + \mathrm{d}D + D(\mathrm{d}e^{-H}) e^H \right] e^{-H}. \end{aligned}$$

This transports the calculation of the derivative at any arbitrary point e^H to the identity element $I = e^0$. It suffices to calculate the Jacobian at $H = 0$, i.e., for $e^H = I$ and $\mathrm{d}e^H = \mathrm{d}H$. Then

$$\mathrm{d}A = \mathrm{d}H D - D\, \mathrm{d}H + \mathrm{d}D,$$

i.e.,

$$\mathrm{d}x_{ij} = \mathrm{d}h_{ij} \lambda_j - \lambda_i\, \mathrm{d}h_{ij} + \mathrm{d}\lambda_i \delta_{ij}$$

This means that we have

$$\frac{\partial x_{ij}}{\partial \lambda_k} = \delta_{ij} \delta_{ik}$$

and

$$\frac{\partial x_{ij}}{\partial h_{kl}} = \delta_{ik} \delta_{jl} (\lambda_l - \lambda_k).$$

Hence the Jacobian is given by

$$J = \det DS = \prod_{k<l} (\lambda_l - \lambda_k).$$

Thus,

$$q(\lambda_1, \dots, \lambda_N, h_{kl}) = p(x_{ij} \,|\, i \geq j) J = c_N e^{-\frac{N}{4} \operatorname{Tr} A^2} \prod_{k<l} (\lambda_l - \lambda_k).$$

This is independent of the "angles" h_{kl}, so integrating over those variables just changes the constant c_N into another constant $\tilde{c}_N$. Rewriting $\operatorname{Tr} A^2 = \lambda_1^2 + \cdots + \lambda_N^2$ in terms of the eigenvalues gives the statement. □

In a similar way, the complex case (GUE) can be treated and we get the following:

Theorem 6.6. *The joint distribution of the eigenvalues of a GUE(N) is given by a density*

$$\hat{c}_N e^{-\frac{N}{2}(\lambda_1^2+\cdots+\lambda_N^2)} \prod_{k<l}(\lambda_l - \lambda_k)^2$$

restricted on $\lambda_1 < \cdots < \lambda_N$.

Definition 6.7. The function

$$\Delta(\lambda_1, \ldots, \lambda_N) = \prod_{\substack{k,l=1 \\ k<l}}^{N} (\lambda_l - \lambda_k)$$

is called the **Vandermonde determinant**.

Proposition 6.8. *For* $\lambda_1, \ldots, \lambda_N \in \mathbb{R}$ *we have that*

$$\Delta(\lambda_1, \ldots, \lambda_N) = \det\left(\lambda_j^{i-1}\right)_{i,j=1}^{N} = \det\begin{pmatrix} 1 & 1 & \cdots & 1 \\ \lambda_1 & \lambda_2 & \ldots & \lambda_N \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_1^{N-1} & \lambda_2^{N-1} & \ldots & \lambda_N^{N-1} \end{pmatrix}.$$

Proof. $\det\left(\lambda_j^{i-1}\right)_{i,j=1}^{N}$ is a polynomial in $\lambda_1, \ldots, \lambda_N$ and if $\lambda_l = \lambda_k$ for some $l, k \in \{1, \ldots, N\}$ then

$$\det\left(\lambda_j^{i-1}\right)_{i,j=1}^{N} = 0 = \Delta(\lambda_1, \ldots, \lambda_N).$$

For $N = 1$ the statement is obviously true. Suppose that the statement is true up to $N - 1$. Consider the determinant

$$p(x) = \det\begin{pmatrix} 1 & 1 & \cdots & 1 \\ x & \lambda_2 & \ldots & \lambda_N \\ \vdots & \vdots & \ddots & \vdots \\ x^{N-1} & \lambda_2^{N-1} & \ldots & \lambda_N^{N-1} \end{pmatrix}.$$

Expanding with respect to the first column, one can see that p has degree at most $N - 1$. But

$$p(\lambda_2) = \cdots = p(\lambda_N) = 0,$$

so $(x - \lambda_k)$ divides p for all $k = 2, \ldots, N$ i.e.,

$$p(x) = c(x - \lambda_2) \cdots (x - \lambda_N)$$

with some constant c not depending on x. Expanding the determinant with respect to x^{N-1} we see that

$$\begin{aligned} c &= (-1)^{N-1} \det \begin{pmatrix} 1 & \cdots & 1 \\ \lambda_2 & \dots & \lambda_N \\ \vdots & \ddots & \vdots \\ \lambda_2^{N-2} & \dots & \lambda_N^{N-2} \end{pmatrix} \\ &= (-1)^{N-1} \Delta^{(N-1)}(\lambda_2, \dots, \lambda_N) \\ &= (-1)^{N-1} \prod_{\substack{k,l=2 \\ k<l}}^{n} (\lambda_l - \lambda_k). \end{aligned}$$

Therefore,

$$\det \left(\lambda_j^{i-1} \right)_{i,j=1}^N = p(\lambda_1) = \prod_{\substack{k,l=1 \\ k<l}}^{n} (\lambda_l - \lambda_k) = \Delta(\lambda_1, \dots, \lambda_N).$$

□

Note: In $\det \left(\lambda_j^{i-1} \right)_{i,j=1}^N$ we can add arbitrary linear combinations of smaller rows to the k-th row without changing the value of the determinant, i.e., we can replace λ^k by any arbitrary monic polynomial $p_k(\lambda) = \lambda^k + \alpha_{k-1}\lambda^{k-1} + \dots + \alpha_1 \lambda + \alpha_0$ of degree k. Hence we have the following statement:

Proposition 6.9. *Let $p_0, \dots, p_{N-1}$ be monic polynomials with $\deg p_k = k$. Then we have*

$$\det \left(p_{i-1} \left(\lambda_j \right) \right)_{i,j=1}^N = \Delta(\lambda_1, \dots, \lambda_N) = \prod_{\substack{k,l=1 \\ k<l}}^{N} (\lambda_l - \lambda_k).$$

In the following, we will make a special choice for the p_k. We will choose them as the Hermite polynomials, which are orthogonal with respect to the Gaussian distribution $\frac{1}{c} e^{-\frac{1}{2}\lambda^2}$.

Definition 6.10. The **Hermite polynomials** H_n are defined by the following rules.

(i) H_n is a monic polynomial of degree n.

(ii) For all $n, m \geq 0$:

$$\int_{\mathbb{R}} H_n(x) \overline{H_m(x)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \, \mathrm{d}x = \delta_{nm} n!$$

Remark 6.11. (i) One can get the $H_n(x)$ from the monomials $1, x, x^2, \ldots$ via Gram-Schmidt orthogonalization as follows.

(1) We define an inner product on the polynomials by

$$\langle f, g \rangle = \int_{\mathbb{R}} f(x)\overline{g(x)} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \,\mathrm{d}x.$$

(2) We put $H_0(x) = 1$. This is monic of degree 0 with

$$\langle H_0, H_0 \rangle = \int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \,\mathrm{d}x = 1 = 0!.$$

(3) We put $H_1(x) = x$. This is monic of degree 1 with

$$\langle H_1, H_0 \rangle = \int_{\mathbb{R}} x \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \,\mathrm{d}x = 0$$

and

$$\langle H_1, H_1 \rangle = \int_{\mathbb{R}} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \,\mathrm{d}x = 1 = 1!.$$

(4) For H_2, note that

$$\langle x^2, H_1 \rangle = \int_{\mathbb{R}} x^3 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \,\mathrm{d}x = 0$$

and

$$\langle x^2, H_0 \rangle = \int_{\mathbb{R}} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \,\mathrm{d}x = 1.$$

Hence we set $H_2(x) = x^2 - H_0(x) = x^2 - 1$. Then we have

$$\langle H_2, H_0 \rangle = 0 = \langle H_2, H_1 \rangle$$

and

$$\langle H_2, H_2 \rangle = \int_{\mathbb{R}} (x^2 - 1)^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \,\mathrm{d}x = 3 - 2 + 1 = 2!$$

(5) Continue in this way.

Note that the H_n are uniquely determined by the requirements that H_n is monic and that $\langle H_m, H_n\rangle = 0$ for all $m \neq n$. That we have $\langle H_n, H_n\rangle = n!$, is then a statement which has to be proved.

(ii) The Hermite polynomials satisfy many explicit relations; important is the three-term recurrence relation

$$xH_n(x) = H_{n+1}(x) + nH_{n-1}(x)$$

for all $n \geq 1$.

(iii) The first few H_n are

$$\begin{aligned} H_0(x) &= 1, \\ H_1(x) &= x, \\ H_2(x) &= x^2 - 1, \\ H_3(x) &= x^3 - 3x, \\ H_4(x) &= x^4 - 6x^2 + 3. \end{aligned}$$

(iv) By Proposition 6.9 we can now use the H_n for writing our Vandermonde determinant as

$$\Delta(\lambda_1, \dots, \lambda_N) = \det\left(H_{i-1}\left(\lambda_j\right)\right)_{i,j=1}^{N}.$$

We want to use this for our GUE(N) density

$$\begin{aligned} q(\lambda_1, \dots, \lambda_N) &= \hat{c}_N e^{-\frac{N}{2}(\lambda_1^2+\cdots+\lambda_N^2)} \Delta(\lambda_1, \dots, \lambda_N)^2 \\ &= \hat{c}_N e^{-\frac{1}{2}(\mu_1^2+\cdots+\mu_N^2)} \Delta\left(\frac{\mu_1}{\sqrt{N}}, \dots, \frac{\mu_N}{\sqrt{N}}\right)^2. \\ &= \hat{c}_N e^{-\frac{1}{2}(\mu_1^2+\cdots+\mu_N^2)} \Delta(\mu_1, \dots, \mu_N)^2 \left(\frac{1}{\sqrt{N}}\right)^{N(N-1)}, \end{aligned}$$

where the $\mu_i = \sqrt{N}\lambda_i$ are the eigenvalues of the "unnormalized" GUE matrix $\sqrt{N}A_N$. It will be easier to deal with those. We now will also go over from ordered eigenvalues $\lambda_1 < \lambda_2 < \cdots < \lambda_N$ to unordered eigenvalues $(\mu_1, \dots, \mu_N) \in \mathbb{R}^N$. Since in the latter case each ordered tuple shows up $N!$ times, this gives an additional factor $N!$ in our density. We collect all these

N-dependent factors in our constant $\tilde{c}_N$. So we now have the density

$$\begin{aligned} p(\mu_1, \dots, \mu_N) &= \tilde{c}_N e^{-\frac{1}{2}(\mu_1^2 + \dots + \mu_N^2)} \Delta(\mu_1, \dots, \mu_N)^2 \\ &= \tilde{c}_N e^{-\frac{1}{2}(\mu_1^2 + \dots + \mu_N^2)} \left[\det\left(H_{i-1}\left(\mu_j\right)\right)_{i,j=1}^N\right]^2 \\ &= \tilde{c}_N \left[\det\left(e^{-\frac{1}{4}\mu_j^2} H_{i-1}\left(\mu_j\right)\right)_{i,j=1}^N\right]^2. \end{aligned}$$

Definition 6.12. The **Hermite functions** Ψ_n are defined by

$$\Psi_n(x) = (2\pi)^{-\frac{1}{4}}(n!)^{-\frac{1}{2}} e^{-\frac{1}{4}x^2} H_n(x).$$

Remark 6.13. (i) We have

$$\int_{\mathbb{R}} \Psi_n(x)\Psi_m(x)\,\mathrm{d}x = \frac{1}{\sqrt{2\pi}}\frac{1}{\sqrt{n!m!}} \int_{\mathbb{R}} e^{-\frac{1}{4}x^2} H_n(x) H_m(x)\,\mathrm{d}x = \delta_{nm},$$

i.e., the Ψ_n are orthonormal with respect to the Lebesgue measure. Actually, they form an orthonormal Hilbert space basis of $L^2(\mathbb{R})$.

(ii) Now we can continue the calculation

$$p(\mu_1, \dots, \mu_N) = c_N \left[\det\left(\Psi_{i-1}\left(\mu_j\right)\right)_{i,j=1}^N\right]^2$$

with a new constant c_N. Denote $V_{ij} = \Psi_{i-1}(\mu_j)$. Then we have

$$(\det V)^2 = \det V \det V^t = \det(V^t V)$$

such that

$$(V^t V)_{ij} = \sum_{k=1}^N V_{ki} V_{kj} = \sum_{k=1}^N \Psi_{k-1}(\mu_i)\Psi_{k-1}(\mu_j).$$

Definition 6.14. The N-th **Hermite kernel** K_N is defined by

$$K_N(x, y) = \sum_{k=0}^{N-1} \Psi_k(x)\Psi_k(y).$$

Theorem 6.15. *The unordered joint eigenvalue distribution of an unnormalized GUE(N) is given by the density*

$$p(\mu_1, \dots, \mu_N) = c_N \det\left(K_N(\mu_i, \mu_j)\right)_{i,j=1}^N.$$

Proposition 6.16. *K_N is a* ***reproducing kernel****, i.e.,*

$$\int_{\mathbb{R}} K_N(x,u)K_N(u,y)\,\mathrm{d}u = K_N(x,y).$$

Proof. We calculate

$$\begin{aligned}\int_{\mathbb{R}} K_N(x,u)K_N(u,y)\,\mathrm{d}u &= \int_{\mathbb{R}} \left(\sum_{k=0}^{N-1}\Psi_k(x)\Psi_k(u)\right)\left(\sum_{l=0}^{N-1}\Psi_l(u)\Psi_l(y)\right)\mathrm{d}u \\ &= \sum_{k,l=0}^{N-1}\Psi_k(x)\Psi_l(y)\int_{\mathbb{R}}\Psi_k(u)\Psi_l(u)\,\mathrm{d}u \\ &= \sum_{k=0}^{N-1}\Psi_k(x)\Psi_k(y) \\ &= K_N(x,y).\end{aligned}$$

□

Lemma 6.17. *Let $K\colon \mathbb{R}^2 \to \mathbb{R}$ be a reproducing kernel, i.e.,*

$$\int_{\mathbb{R}} K(x,u)K(u,y)\,\mathrm{d}u = K(x,y).$$

Put $d = \int_{\mathbb{R}} K(x,x)\,\mathrm{d}x$. Then, for all $n \geq 2$,

$$\int_{\mathbb{R}} \det\left(K(\mu_i,\mu_j)\right)_{i,j=1}^{n}\,\mathrm{d}\mu_n = (d-n+1)\det\left(K(\mu_i,\mu_j)\right)_{i,j=1}^{n-1}.$$

Proof. Consider the case $n = 2$. Then

$$\begin{aligned}\int_{\mathbb{R}} \det\begin{pmatrix}K(\mu_1,\mu_1) & K(\mu_1,\mu_2)\\ K(\mu_2,\mu_1) & K(\mu_2,\mu_2)\end{pmatrix}\mathrm{d}\mu_2 &= K(\mu_1,\mu_1)\int_{\mathbb{R}} K(\mu_2,\mu_2)\,\mathrm{d}\mu_2 \\ &\qquad - \int_{\mathbb{R}} K(\mu_1,\mu_2)K(\mu_2,\mu_1)\,\mathrm{d}\mu_2 \\ &= (d-1)K(\mu_1,\mu_1) \\ &= (d-1)K(\mu_1,\mu_1)\det\left(K(\mu_1,\mu_1)\right).\end{aligned}$$

For $n = 3$,

$$\det \begin{pmatrix} K(\mu_1,\mu_1) & K(\mu_1,\mu_2) & K(\mu_1,\mu_3) \\ K(\mu_2,\mu_1) & K(\mu_2,\mu_2) & K(\mu_2,\mu_3) \\ K(\mu_3,\mu_1) & K(\mu_3,\mu_2) & K(\mu_3,\mu_3) \end{pmatrix} = \det \begin{pmatrix} K(\mu_2,\mu_1) & K(\mu_2,\mu_2) \\ K(\mu_3,\mu_1) & K(\mu_3,\mu_2) \end{pmatrix} K(\mu_1,\mu_3)$$
$$- \det \begin{pmatrix} K(\mu_1,\mu_1) & K(\mu_1,\mu_2) \\ K(\mu_3,\mu_1) & K(\mu_3,\mu_2) \end{pmatrix} K(\mu_2,\mu_3)$$
$$+ \det \begin{pmatrix} K(\mu_1,\mu_1) & K(\mu_1,\mu_2) \\ K(\mu_2,\mu_1) & K(\mu_2,\mu_2) \end{pmatrix} K(\mu_3,\mu_3),$$

with

$$\int_{\mathbb{R}} \det \begin{pmatrix} K(\mu_1,\mu_1) & K(\mu_1,\mu_2) \\ K(\mu_2,\mu_1) & K(\mu_2,\mu_2) \end{pmatrix} K(\mu_3,\mu_3)\,\mathrm{d}\mu_3 = \det \begin{pmatrix} K(\mu_1,\mu_1) & K(\mu_1,\mu_2) \\ K(\mu_2,\mu_1) & K(\mu_2,\mu_2) \end{pmatrix} d,$$

and

$$-\int_{\mathbb{R}} \det \begin{pmatrix} K(\mu_1,\mu_1) & K(\mu_1,\mu_2) \\ K(\mu_3,\mu_1) & K(\mu_3,\mu_2) \end{pmatrix} K(\mu_2,\mu_3)\,\mathrm{d}\mu_3$$
$$= -\int_{\mathbb{R}} \det \begin{pmatrix} K(\mu_1,\mu_1) & K(\mu_1,\mu_2) \\ K(\mu_2,\mu_3)K(\mu_3,\mu_1) & K(\mu_2,\mu_3)K(\mu_3,\mu_2) \end{pmatrix} \mathrm{d}\mu_3$$
$$= -\det \begin{pmatrix} K(\mu_1,\mu_1) & K(\mu_1,\mu_2) \\ K(\mu_2,\mu_1) & K(\mu_2,\mu_2) \end{pmatrix},$$

and

$$\int_{\mathbb{R}} \det \begin{pmatrix} K(\mu_2,\mu_1) & K(\mu_2,\mu_2) \\ K(\mu_3,\mu_1) & K(\mu_3,\mu_2) \end{pmatrix} K(\mu_1,\mu_3)\,\mathrm{d}\mu_3$$
$$= \int_{\mathbb{R}} \det \begin{pmatrix} K(\mu_2,\mu_1) & K(\mu_2,\mu_2) \\ K(\mu_1,\mu_3)K(\mu_3,\mu_1) & K(\mu_1,\mu_3)K(\mu_3,\mu_2) \end{pmatrix} \mathrm{d}\mu_3$$
$$= \det \begin{pmatrix} K(\mu_2,\mu_1) & K(\mu_2,\mu_2) \\ K(\mu_1,\mu_1) & K(\mu_1,\mu_2) \end{pmatrix}$$
$$= -\det \begin{pmatrix} K(\mu_1,\mu_1) & K(\mu_1,\mu_2) \\ K(\mu_2,\mu_1) & K(\mu_2,\mu_2) \end{pmatrix}.$$

The general case works in the same way. □

Corollary 6.18. *Under the assumptions of Lemma 6.17 we have*

$$\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \det\left(K(\mu_i, \mu_j)\right)_{i,j=1}^{n} \,\mathrm{d}\mu_1 \cdots \mathrm{d}\mu_n = (d-n+1)(d-n+2)\cdots(d-1)d.$$

Remark 6.19. We want to apply this to the Hermite kernel $K = K_N$. In this case we have

$$\begin{aligned} d &= \int_{\mathbb{R}} K_N(x,x)\,\mathrm{d}x \\ &= \int_{\mathbb{R}} \sum_{k=0}^{N-1} \Psi_k(x)\Psi_k(x)\,\mathrm{d}x \\ &= \sum_{k=0}^{N-1} \int_{\mathbb{R}} \Psi_k(x)\Psi_k(x)\,\mathrm{d}x \\ &= N, \end{aligned}$$

and thus

$$\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \det\left(K_N(\mu_i, \mu_j)\right)_{i,j=1}^{N} \,\mathrm{d}\mu_1 \cdots \mathrm{d}\mu_n = N!.$$

This now allows us to determine the constant c_N in the density $p(\mu_1, \ldots, \mu_n)$. Since p is a probability density on $\mathbb{R}^N$, we have

$$\begin{aligned} 1 &= \int_{\mathbb{R}^N} p(\mu_1, \ldots, \mu_n)\,\mathrm{d}\mu_1 \cdots \mathrm{d}\mu_N \\ &= c_N \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \det\left(K_N(\mu_i, \mu_j)\right)_{i,j=1}^{N} \,\mathrm{d}\mu_1 \cdots \mathrm{d}\mu_N \\ &= c_N N!, \end{aligned}$$

such that $c_N = \frac{1}{N!}$.

Theorem 6.20. *The unordered joint eigenvalue distribution of an unnormalized GUE(N) is given by a density*

$$p(\mu_1, \ldots, \mu_N) = \frac{1}{N!} \det\left(K_N(\mu_i, \mu_j)\right)_{i,j=1}^{N},$$

where K_N is the Hermite kernel

$$K_N(x,y) = \sum_{k=0}^{N-1} \Psi_k(x)\Psi_k(x).$$

Theorem 6.21. *The averaged eigenvalue density of an unnormalized GUE(N) is given by*

$$p_N(\mu) = \frac{1}{N} K_N(\mu, \mu) = \frac{1}{\sqrt{2\pi}} \frac{1}{N} \sum_{k=0}^{N-1} \frac{1}{k!} H_k(\mu)^2 e^{-\frac{\mu^2}{2}}.$$

Proof. With the notation $\mu_N = \mu$ we get

$$\begin{aligned} p_N(\mu) &= \int_{\mathbb{R}^N} p(\mu_1, \dots, \mu_{N-1}, \mu) \, \mathrm{d}\mu_1 \cdots \mathrm{d}\mu_{N-1} \\ &= \frac{1}{N!} \int_{\mathbb{R}^N} \det \left(K_N(\mu_i, \mu_j) \right)_{i,j=1}^{N} \, \mathrm{d}\mu_1 \cdots \mathrm{d}\mu_{N-1} \\ &= \frac{1}{N!} (N-1)! \det(K_N(\mu, \mu)) \\ &= \frac{1}{N} K_N(\mu, \mu). \end{aligned}$$

□

7 Determinantal processes and non-crossing paths: Karlin-McGregor and Gessel-Viennot

Remark 7.1. Our probability distributions for the eigenvalues of GUE have a determinantal structure, i.e., are of the form

$$p(\mu_1, \dots, \mu_n) = \frac{1}{N!} \det \left(K_N(\mu_i, \mu_j)\right)_{i,j=1}^{N}.$$

They describe N eigenvalues which repel each other (via the factor $(\mu_i - \mu_j)^2$). If we consider corresponding processes, then the paths of the eigenvalues should not cross. There is a quite general relation between determinants as above and non-crossing paths. This appeared independently in different contexts:

(i) Karlin-McGregor, 1958, Markov chains and Brownian motion

(ii) Lindström, 1973, matroids

(iii) Gessel-Viennot, 1985, combinatorics

7.2 Stochastic version à la Karlin-McGregor

Consider a random walk on the integers $\mathbb{Z}$:

- Y_k: position at time k
- $\mathbb{Z}$: possible positions
- Transition probability (to the two neighbors) might depend on position:

$$i-1 \xleftarrow{q_i} i \xrightarrow{p_i} i+1, \qquad q_i + p_1 = 1$$

We now consider n copies of such a random walk, which at time $k = 0$ start at different positions x_i. We are interested in the probability that the paths don't cross. Let x_i be such that all distances are even, i.e., if two paths cross they have to meet.

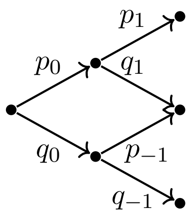
Theorem 7.3 (Karlin-McGregor). *Consider n copies of Y_k, i.e., $(Y_k^{(1)}, \ldots, Y_k^{(n)})$ at the same time with $Y_0^{(i)} = x_i$, where $x_1 > x_2 > \cdots > x_n$. Consider now $t \in \mathbb{N}$ and $y_1 > y_2 > \cdots > y_n$. Denote by*

$$P_t(x_i, y_j) = \mathbb{P}\left[Y_t = y_j \,|\, Y_0 = x_i\right]$$

the probability of one random walk to get from x_i to y_j in t steps. Then we have

$$\mathbb{P}\left[Y_t^{(i)} = y_i \text{ for all } i, Y_s^{(1)} > Y_s^{(2)} > \cdots > Y_s^{(n)} \text{ for all } 0 \leq s \leq t\right] = \det\left(P_t(x_i, y_j)\right)_{i,j=1}^n.$$

Example 7.4. For one symmetric random walk Y_t we have the following probabilities to go in two steps from 0 to -2,0,2:



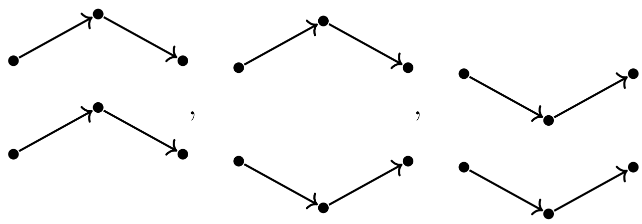
$p_0 p_1 = \frac{1}{4}$

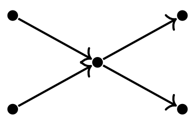
$p_0 q_1 + q_0 p_{-1} = \frac{1}{2}$

$q_0 q_{-1} = \frac{1}{4}$

Now consider two such symmetric random walks and set $x_1 = 2 = y_1$, $x_2 = 0 = y_2$. Then

$$\mathbb{P}\left[Y_2^{(1)} = 2 = Y_0^{(1)}, Y_2^{(2)} = 0 = Y_0^{(2)}, Y_1^{(1)} > Y_1^{(2)}\right]$$

$= \mathbb{P}\left[\left\{\right.\right.$  $\left.\left.\right\}\right] = \frac{3}{16}.$

Note that  is not allowed.

Theorem 7.3 says that we also obtain this probability from the transition probabilities of one random walk as

$$\det\begin{pmatrix} 1/2 & 1/4 \\ 1/4 & 1/2 \end{pmatrix} = \frac{1}{4} - \frac{1}{16} = \frac{3}{16}.$$

Proof of Theorem 7.3. Let Ω_{ij} be the set of all possible paths in t steps from x_i to y_j. Denote by $\mathbb{P}[\pi]$ the probability for such a path $\pi \in \Omega_{ij}$. Then we have

$$P_t(x_i, y_j) = \sum_{\pi \in \Omega_{ij}} \mathbb{P}[\pi]$$

and we have to consider the determinant

$$\det \left(P_t(x_i, y_j)\right)_{i,j=1}^n = \det \left(\sum_{\pi \in \Omega_{ij}} \mathbb{P}[\pi] \right)_{i,j=1}^n .$$

Let us consider the case $n = 2$:

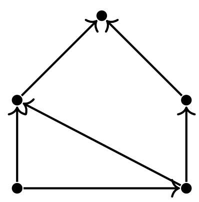
$$\det \begin{pmatrix} \sum_{\pi\in\Omega_{11}} \mathbb{P}[\pi] & \sum_{\pi\in\Omega_{12}} \mathbb{P}[\pi] \\ \sum_{\pi\in\Omega_{21}} \mathbb{P}[\pi] & \sum_{\pi\in\Omega_{22}} \mathbb{P}[\pi] \end{pmatrix} = \sum_{\pi\in\Omega_{11}} \mathbb{P}[\pi] \sum_{\sigma\in\Omega_{22}} \mathbb{P}[\sigma] - \sum_{\pi\in\Omega_{12}} \mathbb{P}[\pi] \sum_{\sigma\in\Omega_{21}} \mathbb{P}[\sigma]$$

Here, the first term counts all pairs of paths $x_1 \to y_1$ and $x_2 \to y_2$. Hence non-crossing ones, but also crossing ones. However, such a crossing pair of paths is, via the "reflection principle" (where we exchange the parts of the two paths after their first crossing), in bijection with a pair of paths $x_1 \to y_2$ and $x_2 \to y_1$; this bijection also preserves the probabilities.
Those paths, $x_1 \to y_2$ and $x_2 \to y_1$, are counted by the second term in the determinant. Hence they cancel each other out, leaving only the non-crossing paths.

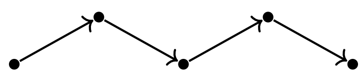
For general n it works similarly. □

7.5 Combinatorial version à la Gessel-Viennot

Let G be a weighted directed graph without directed cycles, e.g.



where we have weights $m_{ij} = m_e$ on each edge $i \xrightarrow{e} j$. This gives weights for directed paths

$P =$ 

via

$$m(P) = \sum_{e \in P} m_e,$$

and then also a weight for connecting vertices a, b,

$$m(a, b) = \sum_{P\colon a \to b} m(P),$$

where we sum over all directed paths from a to b. Note that this is a finite sum, because we do not have directed cycles in our graph.
Consider now two n-tuples of vertices $A = (a_1, \ldots, a_n)$ and $B = (b_1, \ldots, b_n)$. A **path system** $P\colon A \to B$ is given by $\sigma \in S_n$ and paths $P_i\colon a_i \to b_{\sigma(i)}$ for $i = 1, \ldots, n$. We also put $\sigma(P) = \sigma$ and $\operatorname{sgn} P = \operatorname{sgn} \sigma$. A **vertex-disjoint** path system is a path system $(P_1, \ldots, P_n)$, where the paths $P_1, \ldots, P_n$ do not have a common vertex.

Lemma 7.6 (Gessel-Viennot)**.** *Let G be a finite acyclic weighted directed graph and let $A = (a_1, \ldots, a_n)$ and $B = (b_1, \ldots, b_n)$ be two n-sets of vertices. Then we have*

$$\det\left(m(a_i, b_j)\right)_{i,j=1}^{n} = \sum_{\substack{P\colon A \to B \\ \text{vertex-disjoint}}} \operatorname{sgn} \sigma(P) \prod_{i=1}^{n} m(P_i).$$

Proof. Similarly to the proof of Theorem 7.3 the crossing paths cancel each other out in the determinant. □

Example 7.7. Let C_n be the Catalan numbers

$$C_0 = 1, C_1 = 1, C_2 = 2, C_3 = 5, C_4 = 14, \ldots$$

and consider

$$M_n = \begin{pmatrix} C_0 & C_1 & \cdots & C_n \\ C_1 & C_2 & \cdots & C_{n+1} \\ \vdots & & \ddots & \vdots \\ C_n & C_{n+1} & \cdots & C_{2n} \end{pmatrix}.$$

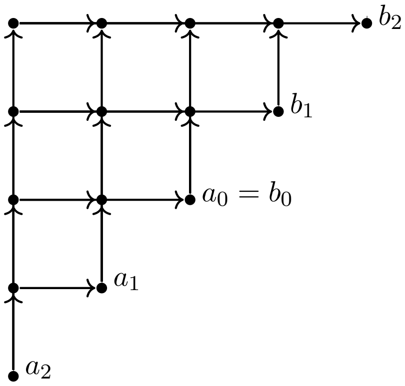
Then we have

$$\begin{aligned} \det M_0 &= 1, \\ \det M_1 &= 2 - 1 = 1, \\ \det M_2 &= 28 + 10 + 10 - 8 - 14 - 25 = 1. \end{aligned}$$

This is actually true for all n:

$$\det M_n = 1$$

This follows from Gessel-Viennot. Let us show it for M_2. For this, consider the graph



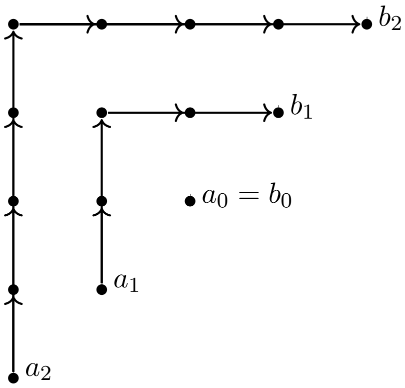
All weights are chosen as 1. Paths in this graph correspond to Dyck graphs, and thus the weights for connecting the a's with the b's are counted by Catalan numbers; e.g.,

$$\begin{aligned} m(a_0, b_0) &= C_0, \\ m(a_0, b_1) &= C_1, \\ m(a_0, b_2) &= C_2, \\ m(a_2, b_2) &= C_4. \end{aligned}$$

Hence, by Gessel-Viennot,

$$\det M_2 = \det \left(m(a_i, b_j)\right)_{i,j=0}^{2} = \sum_{\substack{P\colon (a_0,a_1,a_2)\to(b_0,b_1,b_2) \\ \text{vertex-disjoint}}} 1 = 1,$$

since there is only one such vertex-disjoint system of three paths, corresponding to $\sigma = \text{id}$. This is given as follows; note that the path from a_0 to b_0 is actually a path with 0 steps.


b_2
b_1
$a_0 = b_0$
a_1
a_2

8 Statistics of the largest eigenvalue and Tracy-Widom distribution

Consider GUE(N) or GOE(N). For large N, the eigenvalue distribution is close to a semircircle with density

$$p(x) = \frac{1}{2\pi}\sqrt{4 - x^2}.$$

We will now zoom to a microscopic level and try to understand the behaviour of a single eigenvalue. The behaviour in the *bulk* and at the *edge* is different. We are particularly interested in the largest eigenvalue. Note that at the moment we do not know whether the largest eigenvalue sticks close to 2 with high probability. Wigner's semicircle law implies that it cannot go far below 2, but it does not prevent it from being very large. We will in particular see that this cannot happen.

8.1 Some heuristics on single eigenvalues

Let us first check heuristically what we expect as typical order of fluctuations of the eigenvalues. For this we assume (without justification) that the semicircle predicts the behaviour of eigenvalues down to the microscopic level.

Behaviour in the bulk: In $[\lambda, \lambda + t]$ there should be $\sim tp(\lambda)N$ eigenvalues. This is of order 1 if we choose $t \sim \frac{1}{N}$. This means that eigenvalues in the bulk have for their position an interval of size $\sim \frac{1}{N}$, so this is a good guess for the order of fluctuations for an eigenvalue in the bulk.

Behaviour at the edge: In $[2 - t, 2]$ there should be roughly

$$N \int_{2-t}^{2} p(x)\,\mathrm{d}x = \frac{N}{2\pi} \int_{2-t}^{2} \sqrt{(2-x)(2+x)}\,\mathrm{d}x$$

many eigenvalues. To have this of order 1, we should choose t as follows:

$$\begin{aligned}1 &\approx \frac{N}{2\pi}\int_{2-t}^{2}\sqrt{(2-x)(2+x)}\,\mathrm{d}x\\ &\approx \frac{N}{\pi}\int_{2-t}^{2}\sqrt{2-x}\,\mathrm{d}x\\ &= \frac{N}{\pi}\frac{2}{3}t^{\frac{3}{2}}\end{aligned}$$

Thus $1 \sim t^{\frac{3}{2}}N$, i.e., $t \sim N^{-\frac{2}{3}}$. Hence we expect for the largest eigenvalue an interval or fluctuation of size $N^{-\frac{2}{3}}$. Very optimistically, we might expect

$$\lambda_{\max} \approx 2 + N^{-\frac{2}{3}}x,$$

where x has N-independent distribution.

8.2 The miracle

This heuristic (at least its implication) is indeed true and one has that the limit

$$F_\beta(x) = \lim_{N\to\infty}\mathbb{P}\left[N^{\frac{2}{3}}(\lambda_{\max}-2) \le x\right]$$

exists. It is called the **Tracy-Widom** distribution.

Remark 8.3. (i) Note the parameter β! This corresponds to:

GOE	$\beta = 1$	$(\lambda_i - \lambda_j)^1$
GUE	$\beta = 2$	$(\lambda_i - \lambda_j)^2$
GSE	$\beta = 4$	$(\lambda_i - \lambda_j)^4$

It turns out that the statistics of the largest eigenvalue is different for real, complex, quaternionic Gaussian random matrices. The behaviour on the microscopic level is more sensitive to the underlying symmetry than the macroscopic behaviour.

(ii) On the other hand, when β is fixed, there is a large universality class for the corresponding Tracy-Widom distribution. F_β shows up as limiting fluctuations for

(a) GUE (Tracy-Widom, 1993),

(b) more general Wigner matrices (Soshnikov, 1999),

(c) general unitarily invariant matrix ensembles (Deift + Co, 1994-2000),

(d) length of the longest increasing subsequence of random permutations (Baik, Deift, Johansson, 1999; Okounkov, 2000),

(e) fluctuations of arctic cicle for Aztec diamond (Johansson, 2005),

(f) various growth processes like ASEP, TASEP.

(iii) There is still no uniform explanation for this universality. The feeling is that Tracy-Widom is somehow the analogue of the normal distribution for a kind of central limit theorem, where independence is replaced by some kind of dependence. But no one can make this precise at the moment.

(iv) Proving Tracy-Widom for GUE is out of reach for us, but we will give some ideas. In particular, we try to derive rigorous estimates which show that our $N^{-\frac{2}{3}}$-heuristic is of the right order and that the largest eigenvalue converges to 2.

8.4 How to estimate $\mathbb{P}\left[\lambda_{\max} \geq 2+\varepsilon\right]$?

We want to derive an estimate, in the GUE case, for the probability $\mathbb{P}\left[\lambda_{\max} \geq 2+\varepsilon\right]$, which is compatible with our heuristic that $\varepsilon = N^{-\frac{2}{3}}x$. We will refine our moment method for this. A_N is our normalized GUE(N). We have for all $k \in \mathbb{N}$:

$$
\begin{aligned}
\mathbb{P}\left[\lambda_{\max} \geq 2+\varepsilon\right] &= \mathbb{P}\left[(\lambda_{\max})^{2k} \geq (2+\varepsilon)^{2k}\right] \\
&\leq \mathbb{P}\left[\sum_{j=1}^{N} \lambda_j^{2k} \geq (2+\varepsilon)^{2k}\right] \\
&= \mathbb{P}\left[\operatorname{tr} A_N^{2k} \geq \frac{(2+\varepsilon)^{2k}}{N}\right] \\
&\leq \frac{N}{(2+\varepsilon)^{2k}} \mathbb{E}\left[\operatorname{tr} A_N^{2k}\right]
\end{aligned}
$$

In the last step we used the Markov inequality; note that we have even powers, and hence the random variable $\operatorname{tr}(A_N^{2k})$ is positive.
In Theorem 2.15 we calculated the expectation in terms of a genus expansion as

$$
\begin{aligned}
\mathbb{E}\left[\operatorname{tr}(A_N^{2k})\right] &= \sum_{\pi \in \mathcal{P}_2(2k)} N^{\#(\gamma\pi)-k-1} \\
&= \sum_{g \geq 0} \varepsilon_g(k) N^{-2g},
\end{aligned}
$$

where

$$\varepsilon_g(k) = \#\left\{\pi \in \mathcal{P}_2(2k);\ \pi \text{ has genus } g\right\}.$$

The inequality

$$\mathbb{P}\left[\lambda_{\max} \geq 2+\varepsilon\right] \leq \frac{N}{(2+\varepsilon)^{2k}}\mathbb{E}\left[\operatorname{tr} A_N^{2k}\right]$$

is useless if k is fixed for $N \to \infty$, because then the right hand side goes to ∞. Hence we also have to scale k with N (we will use $k \sim N^{\frac{2}{3}}$), but then the subleading terms in the genus expansion become important. Up to now we only know that $\varepsilon_0(k) = C_k$, but now we need some information on the other $\varepsilon_g(k)$. This is provided by a theorem of Harer and Zagier.

Theorem 8.5 (Harer-Zagier, 1986)**.** *Let us define b_k by*

$$\sum_{g \geq 0} \varepsilon_g(k) N^{-2g} = C_k b_k,$$

where C_k are the Catalan numbers. (Note that the b_k depend also on N, but we suppress this dependency in the notation.) Then we have the recursion formula

$$b_{k+1} = b_k + \frac{k(k+1)}{4N^2} b_{k-1}$$

for all $k \geq 2$.

Example 8.6. We will prove this later. For now, let us check it for small examples. From Remark 2.14 we know

$$\begin{aligned}
C_1 b_1 &= \mathbb{E}\left[\operatorname{tr} A_N^2\right] = 1, \\
C_2 b_2 &= \mathbb{E}\left[\operatorname{tr} A_N^4\right] = 2 + \frac{1}{N^2}, \\
C_3 b_3 &= \mathbb{E}\left[\operatorname{tr} A_N^6\right] = 5 + \frac{10}{N^2}, \\
C_4 b_4 &= \mathbb{E}\left[\operatorname{tr} A_N^8\right] = 14 + \frac{70}{N^2} + \frac{21}{N^4},
\end{aligned}$$

such that

$$
\begin{aligned}
b_1 &= 1, \\
b_2 &= 1 + \frac{1}{2N^2}, \\
b_3 &= 1 + \frac{2}{N^2}, \\
b_4 &= 1 + \frac{5}{N^2} + \frac{3}{2N^4}.
\end{aligned}
$$

We now check the recursion from Theorem 8.5 for $k = 3$:

$$
\begin{aligned}
b_3 + \frac{k(k+1)}{4N^2} b_2 &= 1 + \frac{2}{N^2} + \frac{12}{4N^2}\left(1 + \frac{1}{2N^2}\right) \\
&= 1 + \frac{5}{N^2} + \frac{3}{2N^4} \\
&= b_4
\end{aligned}
$$

Corollary 8.7. *For all $N, k \in \mathbb{N}$ we have for a GUE(N) matrix A_N that*

$$
\mathbb{E}\left[\operatorname{tr} A_N^{2k}\right] \leq C_k \exp\left(\frac{k^3}{2N^2}\right).
$$

Proof. Note that by definition, $b_k > 0$ for all $k \in \mathbb{N}$ and hence $b_{k+1} > b_k$. Thus,

$$
\begin{aligned}
b_{k+1} &= b_k + \frac{k(k+1)}{4N^2} b_{k-1} \\
&\leq b_k \left(1 + \frac{k(k+1)}{4N^2}\right) \\
&\leq b_k \left(1 + \frac{k^2}{2N^2}\right),
\end{aligned}
$$

such that, since $1 + x \leq e^x$,

$$
\begin{aligned}
b_k &\leq \left(1 + \frac{(k-1)^2}{2N^2}\right)\left(1 + \frac{(k-2)^2}{2N^2}\right) \cdots \left(1 + \frac{1^2}{2N^2}\right) \\
&\leq \left(1 + \frac{k^2}{2N^2}\right)^k \\
&\leq \exp\left(\frac{k^2}{2N^2}\right)^k \\
&= \exp\left(\frac{k^3}{2N^2}\right).
\end{aligned}
$$

□

Almost sure convergence of the largest eigenvalue

We can now continue our estimate from Remark 8.4:

$$\begin{aligned}\mathbb{P}[\lambda_{\max} \geq 2+\varepsilon] &\leq \frac{N}{(2+\varepsilon)^{2k}} \mathbb{E}\left[\operatorname{tr} A_N^{2k}\right] \\ &\leq \frac{N}{(2+\varepsilon)^{2k}} C_k \exp\left(\frac{k^3}{2N^2}\right) \\ &\leq \frac{N}{(2+\varepsilon)^{2k}} \frac{4^k}{k^{\frac{3}{2}}} \exp\left(\frac{k^3}{2N^2}\right).\end{aligned}$$

For the last estimate, we used

$$C_k \leq \frac{4^k}{\sqrt{\pi} k^{\frac{3}{2}}} \leq \frac{4^k}{k^{\frac{3}{2}}}.$$

Let us first fix $\varepsilon > 0$ and choose

$$k_N = \left\lfloor N^{\frac{2}{3}} \right\rfloor.$$

Then

$$\frac{N}{k_N^{3/2}} \xrightarrow{N\to\infty} 1, \qquad \frac{k_N^3}{2N^2} \xrightarrow{N\to\infty} \frac{1}{2}.$$

Hence

$$\limsup_{N\to\infty} \mathbb{P}[\lambda_{\max} \geq 2+\varepsilon] \leq \lim_{N\to\infty} \left(\frac{2}{2+\varepsilon}\right)^{2k_N} e^{\frac{1}{2}} = 0,$$

such that for all $\varepsilon > 0$,

$$\lim_{N\to\infty} \mathbb{P}[\lambda_{\max} \geq 2+\varepsilon] = 0.$$

Thus, $\lambda_{\max}$ converges in probability to 2.

Corollary 8.8. *For a GUE(N) matrix A_N we have that its largest eigenvalue converges in probability, and almost surely, to* 2*, i.e.,*

$$\lambda_{\max}(A_N) \xrightarrow{N\to\infty} 2 \qquad \textit{almost surely.}$$

Proof. For the almost sure version one has to use Borel-Cantelli and the fact that

$$\sum_N \left(\frac{2}{2+\varepsilon}\right)^{2k_N} < \infty.$$

□

Estimate for fluctuations

Our estimate from Corollary 8.7 now also gives some information about the fluctuations of $\lambda_{\max}$ about 2, if we choose ε also depending on N. Let us use there now

$$k_N = \left\lfloor N^{\frac{2}{3}} r \right\rfloor$$

and

$$\varepsilon_N = N^{-\frac{2}{3}} t.$$

Then

$$\frac{N}{k_N^{3/2}} \xrightarrow{N\to\infty} \frac{1}{r^{\frac{3}{2}}}, \qquad \frac{k_N^3}{2N^2} \xrightarrow{N\to\infty} \frac{r^3}{2},$$

and

$$\frac{4^{k_N}}{(2+\varepsilon_N)^{2k_N}} = \left(\frac{1}{1+\frac{1}{2N^{2/3}}t} \right)^{2N^{2/3}r} \xrightarrow{N\to\infty} e^{-rt},$$

and thus

$$\limsup_{N\to\infty} \mathbb{P}\left[\lambda_{\max} \geq 2 + tN^{-\frac{2}{3}}\right] \leq \frac{1}{r^{3/2}} e^{r^3/2} e^{-rt}$$

for arbitrary $r > 0$. We optimize this now by choosing $r = \sqrt{t}$ for $t > 0$ and get

$$\limsup_{N\to\infty} \mathbb{P}\left[\lambda_{\max} \geq 2 + tN^{-\frac{2}{3}}\right] \leq t^{-\frac{3}{4}} \exp\left(-\frac{1}{2} t^{\frac{3}{2}}\right).$$

Although this estimate does not prove the existence of the limit on the left hand side, it turns out that the right hand side is quite sharp and captures the tail behaviour of the Tracy-Widom distribution quite well.

Proof of the Harer-Zagier recursion

Proof of Theorem 8.5. Let us denote

$$T(k,N) = \mathbb{E}\left[\operatorname{tr} A_N^{2k}\right] = \sum_{g\geq 0} \varepsilon_g(k) N^{-2g}.$$

The genus expansion shows that $T(k,N)$ is, for fixed k, a polynomial in N^{-1}. Expressing it in terms of integrating over eigenvalues reveals the surprising fact that, up to a Gaussian factor, it is also a polynomial in k for fixed N, as will be shown in the following Lemma 8.9. We now have that

$$\begin{aligned} N^k \frac{1}{(2k-1)!!} T(k,N) &= N^k \frac{1}{(2k-1)!!} \sum_{\pi \in \mathcal{P}_2(2k)} N^{\#(\gamma\pi)-k-1} \\ &= \frac{1}{N} \frac{1}{(2k-1)!!} \sum_{\pi \in \mathcal{P}_2(2k)} N^{\#(\gamma\pi)} \\ &= \frac{1}{N} \frac{1}{(2k-1)!!} t(k,N), \end{aligned}$$

where the last equality defines $t(k,N)$.
By Lemma 8.9, $\frac{t(k,N)}{(2k-1)!!}$ is a polynomial of degree $N-1$ in k. We interpret it as follows:

$$t(k,N) = \sum_{\pi \in \mathcal{P}_2(2k)} \# \{\text{coloring cycles of } \gamma\pi \text{ with at most } N \text{ different colors}\}$$

Let us introduce

$$\tilde{t}(k,L) = \sum_{\pi \in \mathcal{P}_2(2k)} \# \{\text{coloring cycles of } \gamma\pi \text{ with exactly } L \text{ different colors}\},$$

then we have

$$t(k,N) = \sum_{L=1}^{N} \tilde{t}(k,L) \binom{N}{L},$$

because if we want to use at most N different colors, then we can do this by using exactly L different colors (for any L between 1 and N), and after fixing L we have $\binom{N}{L}$ many possibilities to choose the L colors among the N colors.
This relation can be inverted by

$$\tilde{t}(k,N) = \sum_{L=1}^{N} (-1)^{N-L} \binom{N}{L} t(k,L)$$

and hence $\frac{\tilde{t}(k,N)}{(2k-1)!!}$ is also a polynomial in k of degree $N-1$. But now we have

$$0 = \tilde{t}(0,N) = \tilde{t}(1,N) = \cdots = \tilde{t}(N-2,N),$$

since $\gamma\pi$ has at most $k+1$ cycles (see 2.18) for $\pi \in \mathcal{P}_2(2k)$ and thus $\tilde{t}(k+1,N) = 0$ if $k+1 < N$, as we need at least N cycles if we want to use N different colors. So,

$\frac{\tilde{t}(k,N)}{(2k-1)!!}$ is a polynomial in k of degree $N-1$ and we know $N-1$ zeros; hence it must be of the form

$$\frac{\tilde{t}(k,N)}{(2k-1)!!} = \alpha_N k(k-1)\cdots(k-N+2) = \alpha_N \binom{k}{N-1}(N-1)!.$$

Hence,

$$t(k,N) = \sum_{L=1}^{N} \binom{N}{L}\binom{k}{L-1}(L-1)!\alpha_L(2k-1)!!.$$

To identify α_N we look at

$$\alpha_{N+1}\binom{N}{N}N!(2N-1)!! = \tilde{t}(N,N+1) = C_N(N+1)!.$$

Note that only the NC pairings can be colored with exactly $N+1$ colors, and for each such π there are $(N+1)!$ ways of doing so. We conclude

$$\begin{aligned}
\alpha_{N+1} &= \frac{C_N(N+1)!}{N!(2N-1)!!} \\
&= \frac{C_N(N+1)}{(2N-1)!!} \\
&= \frac{1}{N+1}\binom{2N}{N}\frac{N+1}{(2N-1)!!} \\
&= \frac{(2N)!}{N!N!(2N-1)!!} \\
&= \frac{2^N}{N!}.
\end{aligned}$$

Thus we have

$$\begin{aligned}
T(k,N) &= \frac{1}{N^{k+1}}t(k,N) \\
&= \frac{1}{N^{k+1}}\sum_{L=1}^{N}\binom{N}{L}\binom{k}{L-1}(L-1)!\frac{2^{L-1}}{(L-1)!}(2k-1)!! \\
&= (2k-1)!!\frac{1}{N^{k+1}}\sum_{L=1}^{N}\binom{N}{L}\binom{k}{L-1}2^{L-1}.
\end{aligned}$$

To get information from this on how this changes in k we consider a generating function in k,

$$\begin{aligned}
\mathcal{T}(s,N) &= 1 + 2\sum_{k=0}^{\infty} \frac{T(k,N)}{(2k-1)!!}(Ns)^{k+1} \\
&= 1 + 2\sum_{k=0}^{\infty}\sum_{L=1}^{N} \binom{N}{L}\binom{k}{L-1} 2^{L-1} s^{k+1} \\
&= \sum_{L=0}^{N} \binom{N}{L} 2^L \sum_{k=L-1}^{\infty} \binom{k}{L-1} s^{k+1} \\
&= \sum_{L=0}^{N} \binom{N}{L} 2^L \left(\frac{s}{1-s}\right)^L \\
&= \sum_{L=0}^{N} \binom{N}{L} \left(\frac{2s}{1-s}\right)^L \\
&= \left(1 + \frac{2s}{1-s}\right)^N \\
&= \left(\frac{1+s}{1-s}\right)^N .
\end{aligned}$$

Note that

$$\frac{1}{(2k-1)!!} = \frac{2^k}{k!C_k(k+1)}$$

and

$$b_k^{(N)} = \frac{T(k,N)}{C_k}.$$

(We write now $b_k^{(N)}$ instead of b_k to make the dependence on N explicit.) Therefore,

$$\begin{aligned}
\left(\frac{1+s}{1-s}\right)^N = \mathcal{T}(s,N) &= 1 + 2\sum_{k=0}^{\infty} \frac{T(k,N)}{(k+1)!C_k} 2^k (Ns)^{k+1} \\
&= 1 + \sum_{k=0}^{\infty} \frac{b_k^{(N)}}{(k+1)!} (2Ns)^{k+1}.
\end{aligned}$$

Also,

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{T}(s,N) &= N\left(\frac{1+s}{1-s}\right)^{N-1}\frac{(1-s)+(1+s)}{(1-s)^2} \\
&= 2N\left(\frac{1+s}{1-s}\right)^{N-1}\frac{1}{(1-s)^2} \\
&= N\left(\frac{1+s}{1-s}\right)^{N-1}\frac{(1-s)+(1+s)}{(1-s)^2} \\
&= 2N\left(\frac{1+s}{1-s}\right)^{N}\frac{1}{(1-s)(1+s)},
\end{aligned}
$$

such that

$$
(1-s^2)\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{T}(s,N) = 2N\mathcal{T}(s,N).
$$

Note that we have

$$
\frac{\mathrm{d}}{\mathrm{d}s}\mathcal{T}(s,N) = \sum_{k=0}^{\infty}\frac{b_k^{(N)}}{k!}(2Ns)^k 2N.
$$

Thus, by comparing coefficients in our equation from above, we conclude

$$
\frac{b_{k+1}^{(N)}}{(k+1)!}(2N)^{k+2} - \frac{b_{k-1}^{(N)}}{(k-1)!}(2N)^{k} = 2N\frac{b_{k+1}^{(N)}}{(k+1)!}(2N)^{k+1},
$$

such that, finally,

$$
b_{k+1}^{(N)} = b_k^{(N)} + b_{k-1}^{(N)}\frac{(k+1)k}{(2N)^2}.
$$

□

To finish the proof of the Harer-Zagier theorem it only remains to prove the following lemma. Note that this is the only place where we need the random matrix interpretation of our quantities.

Lemma 8.9. *The expression*

$$
N^k\frac{1}{(2k-1)!!}T(k,N)
$$

is a polynomial of degree $N-1$ in k.

Proof. First check the easy case $N = 1$: $T(k, 1) = (2k-1)!!$ is the $2k$-th moment of a normal variable and

$$\frac{T(k,1)}{(2k-1)!!} = 1$$

is a polynomial of degree 0 in k.

For general N we have

$$\begin{aligned}
T(k,N) &= \mathbb{E}\left[\operatorname{tr} A_N^{2k}\right] \\
&= c_N \int\limits_{\mathbb{R}^N} \left(\lambda_1^{2k} + \cdots + \lambda_N^{2k}\right) e^{-\frac{N}{2}\left(\lambda_1^2+\cdots+\lambda_N^2\right)} \prod_{i<j} (\lambda_i - \lambda_j)^2 \,\mathrm{d}\lambda_1 \cdots \mathrm{d}\lambda_N \\
&= N c_N \int\limits_{\mathbb{R}^N} \lambda_1^{2k} e^{-\frac{N}{2}\left(\lambda_1^2+\cdots+\lambda_N^2\right)} \prod_{i \neq j} |\lambda_i - \lambda_j| \,\mathrm{d}\lambda_1 \cdots \mathrm{d}\lambda_N \\
&= N c_N \int\limits_{\mathbb{R}} \lambda_1^{2k} e^{-\frac{N}{2}\lambda_1^2} p_N(\lambda_1) \,\mathrm{d}\lambda_1,
\end{aligned}$$

where p_N is the result of integrating the Vandermonde over $\lambda_2, \ldots, \lambda_N$. It is an even polynomial in λ_1 of degree $2(N-1)$, whose coefficients depend only on N and not on k. So

$$p_N(\lambda_1) = \sum_{l=0}^{N-1} \alpha_l \lambda_1^{2l}$$

with α_l possibly depending on N. Thus,

$$\begin{aligned}
T(k,N) &= N c_N \sum_{l=0}^{N-1} \alpha_l \int\limits_{\mathbb{R}} \lambda_1^{2k+2l} e^{-\frac{N}{2}\lambda_1^2} \,\mathrm{d}\lambda_1 \\
&= N c_N \sum_{l=0}^{N-1} \alpha_l k_N (2k+2l-1)!! N^{-k},
\end{aligned}$$

where k_N contains the N-dependent normalization constants of the Gaussian measure; hence

$$\frac{N^k T(k,N)}{(2k-1)!!}$$

is a linear combination (with N-dependent coefficients) of terms of the form

$$\frac{(2k+2l-1)!!}{(2k-1)!!}.$$

These terms are polynomials in k of degree l. □

8.10 How does one get to Tracy-Widom?

For determining the Tracy-Widom fluctuations in the limit $N \to \infty$ one has to use the analytic description of the GUE's joint density. Recall from Theorem 6.20 that the joint density of the unordered eigenvalues of an unnormalized GUE(N) is given by

$$p(\mu_1, \dots, \mu_N) = \frac{1}{N!} \det \left(K_N(\mu_i, \mu_j)\right)_{i,j=1}^{N},$$

where K_N is the Hermite kernel

$$K_N(x, y) = \sum_{k=0}^{N-1} \Psi_k(x)\Psi_k(x)$$

with the Hermite functions Ψ_k. Because K_N is a reproducing kernel, we can integrate out some of the eigenvalues and get a density of the same form. If we are integrate out all but r eigenvalues we get, by Corollary 6.18,

$$\begin{aligned} \int_{\mathbb{R}} \cdots \int_{\mathbb{R}} p(\mu_1, \dots, \mu_N)\, \mathrm{d}\mu_{r+1} \cdots \mathrm{d}\mu_N &= \frac{1}{N!} \cdot 1 \cdot 2 \cdots (N-r) \cdot \det \left(K(\mu_i, \mu_j)\right)_{i,j=1}^{r} \\ &= \frac{(N-r)}{N!} \det \left(K(\mu_i, \mu_j)\right)_{i,j=1}^{r} \\ &=: p_N(\mu_1, \dots, \mu_r). \end{aligned}$$

Now consider

$$\begin{aligned} \mathbb{P}\left[\mu_{\max}^{(N)} \le t\right] &= \mathbb{P}\left[\text{there is no eigenvalue in } (t, \infty)\right] \\ &= 1 - \mathbb{P}\left[\text{there is an eigenvalue in } (t, \infty)\right] \\ &= 1 - \Bigg[N\mathbb{P}\left[\mu_1 \in (t, \infty)\right] - \binom{N}{2}\mathbb{P}\left[\mu_1, \mu_2 \in (t, \infty)\right] \\ &\qquad + \binom{N}{3}\mathbb{P}\left[\mu_1, \mu_2, \mu_3 \in (t, \infty)\right] - \cdots \Bigg] \\ &= 1 + \sum_{r=1}^{N} (-1)^r \binom{N}{R} \int_t^{\infty} \cdots \int_t^{\infty} p_N(\mu_1, \dots, \mu_r)\, \mathrm{d}\mu_1 \cdots \mathrm{d}\mu_r \\ &= 1 + \sum_{r=1}^{N} (-1)^r \frac{1}{r!} \int_t^{\infty} \cdots \int_t^{\infty} \det \left(K(\mu_i, \mu_j)\right)_{i,j=1}^{r} \mathrm{d}\mu_1 \cdots \mathrm{d}\mu_r. \end{aligned}$$

Does this have a limit for $N \to \infty$?

Note that p is the distribution for a GUE(N) without normalization, i.e.,

$$\mu_{\max}^{(N)} \approx 2\sqrt{N}.$$

More precisely, we expect flucttutations

$$\mu_{\max}^{(N)} \approx \sqrt{N}\left(2 + tN^{-\frac{2}{3}}\right) = 2\sqrt{N} + tN^{-\frac{1}{6}}.$$

We put

$$\tilde{K}_N(x,y) = N^{-\frac{1}{6}} K_N\left(2\sqrt{N} + xN^{-\frac{1}{6}}, 2\sqrt{N} + yN^{-\frac{1}{6}}\right)$$

so that we have

$$\mathbb{P}\left[N^{\frac{2}{3}}\left(\frac{\mu_{\max}^{(N)} - 2}{\sqrt{N}} - 2\right) \leq t\right] = \sum_{r=0}^{N} \frac{(-1)^r}{r!} \int_t^\infty \cdots \int_t^\infty \det\left(\tilde{K}(x_i, x_j)\right)_{i,j=1}^{r} \mathrm{d}x_1 \cdots \mathrm{d}x_r.$$

We expect that the limit

$$F_2(t) = \lim_{N\to\infty} \mathbb{P}\left[N^{\frac{2}{3}}\left(\frac{\mu_{\max}^{(N)} - 2}{\sqrt{N}} - 2\right) \leq t\right]$$

exists. For this, we need the limit

$$\lim_{N\to\infty} \tilde{K}_N(x,y).$$

Recall that

$$K_N(x,y) = \sum_{k=0}^{N-1} \Psi_k(x)\Psi_k(y).$$

For the Hermite functions we have the Christoffel-Darboux identity (Problem 31)

$$\sum_{k=0}^{n-1} \frac{H_k(x)H_k(y)}{k!} = \frac{H_n(x)H_{n-1}(y) - H_{n-1}(x)H_n(y)}{(x-y)\ (n-1)!}$$

and with

$$\Psi_k(x) = (2\pi)^{-\frac{1}{4}}(k!)^{-\frac{1}{2}} e^{-\frac{1}{4}x^2} H_k(x)$$

as defined in Definition 6.12, we can rewrite

$$\begin{aligned}
K_N(x,y) &= \frac{1}{\sqrt{2\pi}} \sum_{k=0}^{N-1} \frac{1}{k!} e^{-\frac{1}{4}\left(x^2+y^2\right)} H_k(x)H_k(y) \\
&= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{4}\left(x^2+y^2\right)} \frac{H_N(x)H_{N-1}(y) - H_{N-1}(x)H_N(y)}{(x-y)\ (N-1)!} \\
&= \sqrt{N} \cdot \frac{\Psi_N(x)\Psi_{N-1}(y) - \Psi_{N-1}(x)\Psi_N(y)}{x-y}.
\end{aligned}$$

Note that the Ψ_N satisfy the differential equation

$$\psi_N'(x) = -\frac{x}{2}\psi_N(x) + \sqrt{N}\psi_{N-1}(x)$$

such that

$$\begin{aligned} K_N(x,y) &= \frac{\Psi_N(x)\left[\Psi_N'(y) + \frac{y}{2}\Psi_N(y)\right] - \left[\Psi_N'(x) + \frac{x}{2}\Psi_N(x)\right]\Psi_N(y)}{x-y} \\ &= \frac{\Psi_N(x)\Psi_N'(y) - \Psi_N'(x)\Psi_N(y)}{x-y} - \frac{1}{2}\Psi_N(x)\Psi_N(y). \end{aligned}$$

Now put

$$\widetilde{\Psi}_N(x) = N^{\frac{1}{12}}\Psi_N\left(2\sqrt{N} + xN^{-\frac{1}{6}}\right),$$

thus

$$\widetilde{\Psi}_N'(x) = N^{\frac{1}{12}}\Psi_N'\left(2\sqrt{N} + xN^{-\frac{1}{6}}\right)N^{-\frac{1}{6}}.$$

Then

$$\tilde{K}(x,y) = \frac{\widetilde{\Psi}_N(x)\widetilde{\Psi}_N'(y) - \widetilde{\Psi}_N'(x)\widetilde{\Psi}_N(y)}{x-y} - \frac{1}{2N^{\frac{1}{3}}}\widetilde{\Psi}_N'(x)\widetilde{\Psi}_N'(y).$$

One can show, by a quite non-trivial steepest descent method, that $\widetilde{\Psi}_N(x)$ converges to a limit. Let us call this limit the **Airy function**

$$\operatorname{Ai}(x) = \lim_{N\to\infty}\widetilde{\Psi}_N(x).$$

The convergence is actually so strong that also

$$\operatorname{Ai}'(x) = \lim_{N\to\infty}\widetilde{\Psi}_N'(x),$$

and hence

$$\begin{aligned} \lim_{N\to\infty}\tilde{K}(x,y) &= \frac{\operatorname{Ai}(x)\operatorname{Ai}'(y) - \operatorname{Ai}'(x)\operatorname{Ai}(y)}{x-y} \\ &=: \mathrm{A}(x,y). \end{aligned}$$

A is called the **Airy kernel**. For the Hermite functions we have

$$\psi_N''(x) + \left(N + \frac{1}{2} - \frac{x^2}{4}\right)\psi_N(x) = 0.$$

For the $\widetilde{\Psi}_N$ we have

$$\widetilde{\Psi}'_N(x) = N^{-\frac{1}{12}}\Psi'_N\left(2\sqrt{N} + xN^{-\frac{1}{6}}\right)$$

and

$$\widetilde{\Psi}''_N(x) = N^{-\frac{1}{4}}\Psi''_N\left(2\sqrt{N} + xN^{-\frac{1}{6}}\right).$$

Thus,

$$\begin{aligned}
\widetilde{\Psi}''_N(x) &= -N^{-\frac{1}{4}}\left[N + \frac{1}{2} - \frac{\left(2\sqrt{N} + xN^{-\frac{1}{6}}\right)^2}{4}\right]\Psi_N\left(2\sqrt{N} + xN^{-\frac{1}{6}}\right) \\
&= -N^{-\frac{1}{3}}\left[N + \frac{1}{2} - \frac{4N + 4\sqrt{N}xN^{-\frac{1}{6}} + x^2N^{-\frac{1}{3}}}{4}\right]\widetilde{\Psi}_N(x) \\
&= -N^{-\frac{1}{3}}\left[\frac{1}{2} - \frac{4xN^{\frac{1}{3}} + x^2N^{-\frac{1}{3}}}{4}\right]\widetilde{\Psi}_N(x) \\
&\approx -x\widetilde{\Psi}_N(x).
\end{aligned}$$

Hence we expect that Ai should satisfy the differential equation

$$\mathrm{Ai}''(x) - x\,\mathrm{Ai}(x) = 0.$$

This is indeed the case, but the proof is again beyond our tools. Let us just give the formal definiton of the Airy function and formulate the final result.

Definition 8.11. The **Airy function** $\mathrm{Ai}\colon \mathbb{R} \to \mathbb{R}$ is a solution of the **Airy ODE**

$$u''(x) = xu(x)$$

determined by the following asymptotic as $x \to \infty$:

$$\mathrm{Ai}(x) \sim \frac{1}{2}\pi^{-\frac{1}{2}}x^{-\frac{1}{4}}e^{-\frac{2}{3}x^{\frac{3}{2}}}$$

The **Airy kernel** is defined by

$$\mathrm{A}(x, y) = \frac{\mathrm{Ai}(x)\,\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\,\mathrm{Ai}(y)}{x - y}.$$

Theorem 8.12. *The random variable $N^{\frac{2}{3}}(\lambda_{\max} - 2)$ of the normalized GUE has a limiting distribution as $N \to \infty$. Its limiting distribution function is*

$$\begin{aligned} F_2(t) &= \lim_{N\to\infty} \mathbb{P}\left[N^{\frac{2}{3}}(\lambda_{\max} - 2) \leq t\right] \\ &= \sum_{r=0}^{N} \frac{(-1)^r}{r!} \int_t^\infty \cdots \int_t^\infty \det\left(A(x_i, x_j)\right)_{i,j=1}^r \mathrm{d}x_1 \cdots \mathrm{d}x_r. \end{aligned}$$

The main contribution of Tracy-Widom in this context was that they were able to derive another, quite astonishing, representation of the limiting distribution.

Theorem 8.13 (Tracy-Widom, 1994)**.** *The distribution function F_2 satisfies*

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

where q is a solution of the Painlevé II equation

$$q''(x) - xq(x) + 2q(x)^3 = 0$$

with $q(x) \sim \mathrm{Ai}(x)$ as $x \to \infty$.

9 Statistics of the longest increasing subsequence

Definition 9.1. A permutation $\sigma \in S_n$ is said to have an **increasing subsequence of length** k if there exist indices $1 \leq i_1 < \cdots < i_k \leq n$ such that

$$\sigma(i_1) < \cdots < \sigma(i_k).$$

For a **decreasing subsequence of length k** the above holds with the second set of inequalities reversed. For a given $\sigma \in S_n$ we denote the length of an increasing subsequence of maximal length by $L_n(\sigma)$.

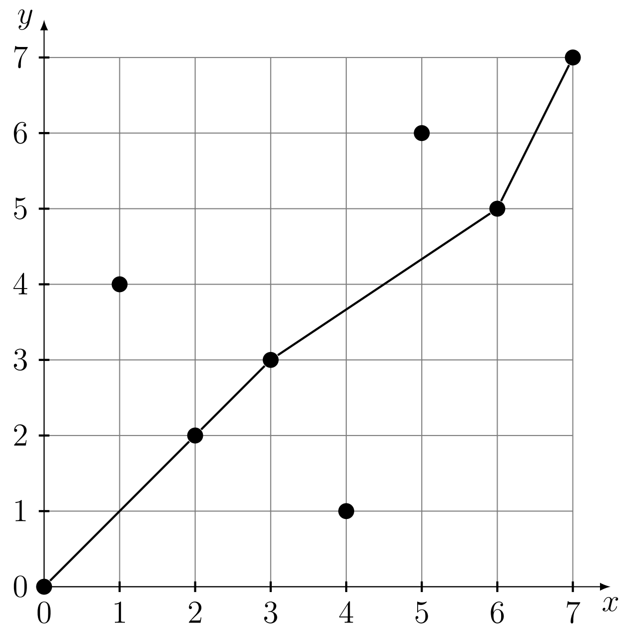
Example 9.2. (i) $\sigma = \mathrm{id}$ has an increasing subsequence of length n, hence $L_n(\mathrm{id}) = n$. All decreasing subsequences have length 1.

(ii) $\sigma = \begin{pmatrix} 1 & 2 & \cdots & n-1 & n \\ n & n-1 & \cdots & 2 & 1 \end{pmatrix}$ has $L_n(\sigma) = 1$, but there is a decreasing subsequence of length n.

(iii) Consider

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 2 & 3 & 1 & 6 & 5 & 7 \end{pmatrix};$$

this has $(2, 3, 5, 7)$ and $(2, 3, 6, 7)$ as longest increasing subsequences, thus $L_7(\sigma) = 4$. Its longest decreasing subsequences are $(4, 2, 1)$ and $(4, 3, 1)$ with length 3. In the graphical representation



an increasing subsequence corresponds to a path that always goes up.

(iv) Longest increasing subsequences are relevant for sorting algorithms. Consider a library of n books, labeled bijectively with numbers $1, \dots, n$, arranged somehow on a single long bookshelf. The configuration of the books correspond to a permutation $\sigma \in S_n$. How many operations does one need to sort the books in a canonical ascending order $1, 2, \dots, n$? It turns out that the minimum number is $n - L_n(\sigma)$. One can sort around an increasing subsequence.
Example. Around the longest increasing subsequence $(1, 2, 6, 8)$ we sort

$$\begin{array}{rccccccccc} & 4 & 1 & 9 & 3 & 2 & 7 & 6 & 8 & 5 \\ \to & 4 & 1 & 9 & 2 & 3 & 7 & 6 & 8 & 5 \\ \to & 1 & 9 & 2 & 3 & 4 & 7 & 6 & 8 & 5 \\ \to & 1 & 9 & 2 & 3 & 4 & 5 & 7 & 6 & 8 \\ \to & 1 & 9 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ \to & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \end{array}$$

in $9 - 4 = 5$ operations.

Remark 9.3. One has situations with only small increasing subsequences, like $\sigma = \begin{pmatrix} 1 & \cdots & n \\ n & \cdots & 1 \end{pmatrix}$, but then one has long decreasing subsequences. This is true in general;

one cannot avoid both long decreasing and long increasing subsequences at the same time. According to the slogan

"Complete order is impossible." (Motzkin)

Theorem 9.4 (Erdős, Szekeres, 1935)**.** *Every permutation $\sigma \in S_{n^2+1}$ has a monotone subsequence of length more than n.*

Proof. Write $\sigma = a_1 a_2 \cdots a_{n^2+1}$. Assign labels (x_k, y_k), where x_k is the length of a longest increasing subsequence ending at a_k and y_k is the length of a longest decreasing subsequence ending at a_k. Assume now that there is no monotone subsequence of length $n+1$. Hence for all k,

$$1 \leq x_k, y_k \leq n,$$

i.e., there are only n^2 possible labels. By the pigeonhole principle there are $i < j$ with $(x_i, y_i) = (x_j, y_j)$. If $a_i < a_j$ we can append a_j to a longest increasing subsequence ending at a_i, but then $x_j > x_i$. If $a_i > a_j$ we can append a_j to a longest decreasing subsequence ending at a_i, but then $y_j > y_i$. Hence we have a contradiction. □

9.5 A bit of history and relation to Tracy-Widom

We are now interested in the distribution of $L_n(\sigma)$ for $n \to \infty$. This means, we put the uniform distribution on permutations, i.e.,

$$\mathbb{P}[\sigma] = \frac{1}{n!}$$

for all $\sigma \in S_n$, and consider $L_n \colon S_n \to \mathbb{R}$ as a random variable. What is the asymptotic distribution of L_n? This question is called Ulan's problem and was raised in the 1960's. In 1972, Hammersley showed that the limit

$$\Lambda = \lim_{n\to\infty} \frac{\mathbb{E}[L_n]}{\sqrt{n}}$$

exists and that $L_n/\sqrt{n}$ converges to Λ in probability. In 1977, both Vershik / Kerov and Logan / Shepp showed independently that $\Lambda = 2$. Then in 1998, Baik, Deift and Johansson proved the asymptotic behaviour of the fluctuations of L_n; quite surprisingly, this is also captured by the Tracy-Widom distribution:

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{L_n - 2\sqrt{n}}{n^{\frac{1}{6}}} \leq t\right] = F_2(t).$$

A very rough sketch of the proof of the Baik, Deift, Johansson theorem

(i) RSK correspondence relates permutations to Young diagrams. L_n goes under this mapping to the length of the first row of the diagram.

(ii) These Young diagrams correspond to non-intersecting paths.

(iii) Via Gessel-Viennot the relevant quantities in terms of NC paths have a determinantal form.

(iv) Show that the involved kernel, suitably rescaled, converges to the Airy kernel.

In the following we want to give some idea of the first two items in the above list; the main (and very hard part of the proof) is to show the convergence to the Airy kernel.

9.6 RSK correspondence

RSK stands for Robinson-Schensted-Knuth after papers from 1938, 1961 and 1973. It gives a bijection

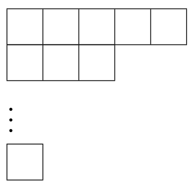
$$S_n \longleftrightarrow \bigcup_{\substack{\lambda \\ \text{Young diagram} \\ \text{of size } n}} (\operatorname{Tab}\lambda \times \operatorname{Tab}\lambda),$$

where $\operatorname{Tab}\lambda$ is the set of Young tableaux of shape λ.

Definition 9.7. (i) Let $n \geq 1$. A **partition** of n is a sequence of natural numbers $\lambda = (\lambda_1, \dots, \lambda_r)$ such that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r \qquad \text{and} \qquad \sum_{i=1}^{r} \lambda_i = n.$$

We denote this by $\lambda \vdash n$. Graphically, a partition $\lambda \vdash n$ is represented by a **Young diagram** with n boxes.



(ii) A **Young tableau** of shape λ is the Young diagram λ filled with numbers $1, \dots, n$ such that in any row the numbers are increasing from left to right and in any column the numbers are increasing from top to bottom. We denote the set of all Young tableaux of shape λ by $\operatorname{Tab}\lambda$.

Example 9.8. (i) For $n = 1$ there is only one Young diagram and one corresponding Young tableau:

$$\lambda = \ydiagram{1}, \qquad \mathrm{Tab}\,\ydiagram{1} = \ytableaushort{1}$$

For $n = 2$, there are two Young diagrams, each of them having one corresponding Young tableau

$$\lambda = \ydiagram{2}, \quad \ydiagram{1,1} \qquad \mathrm{Tab}\,\ydiagram{2} = \ytableaushort{12}, \quad \mathrm{Tab}\,\ydiagram{1,1} = \ytableaushort{1,2}$$

For $n = 3$, there are three Young diagrams

$$\lambda = \ydiagram{3}, \quad \ydiagram{2,1}, \quad \ydiagram{1,1,1},$$

one of them has two corresponding tableaux:

$$\mathrm{Tab}\,\ydiagram{2,1} = \ytableaushort{12,3}, \quad \ytableaushort{13,2}$$

(ii) Note that a tableau of shape λ corresponds to a walk from $\emptyset$ to λ by adding one box in each step and only visiting Young diagrams, e.g.,

$$\ytableaushort{1248,37,5,6}$$

corresponds to

$$\ydiagram{1} \to \ydiagram{2} \to \ydiagram{2,1} \to \ydiagram{3,1} \to \ydiagram{3,1,1} \to \ydiagram{3,1,1,1} \to \ydiagram{3,2,1,1} \to \ydiagram{4,2,1,1}$$

Those objects are extremely important since they parametrize the irreducible representations of S_n:

$$\lambda \vdash n \longleftrightarrow \text{irreducible representation } \pi_\lambda \text{ of } S_n.$$

Furthermore, the dimension of such a representation π_λ is given by the number of tableaux of shape λ. If one recalls that for any finite group one has the general statement that the sum of the squares of the dimensions over all irreducible representations of the group gives the number of elements in the group, then one has for the symmetric group the statement that

$$\sum_{\lambda \vdash n} (\# Tab \lambda)^2 = n!.$$

This shows that there is a bijection between elements in S_n and pairs of tableaux of the same shape $\lambda \vdash n$. The RSK correspondence is such a concrete bijection, given by an explicit algorithm. It has the property, that L_n goes under this bijection over to the length of the first row of the corresponding Young diagram λ.

For example, under the RSK correspondence, the permutation

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 4 & 2 & 3 & 6 & 5 & 1 & 7 \end{pmatrix}$$

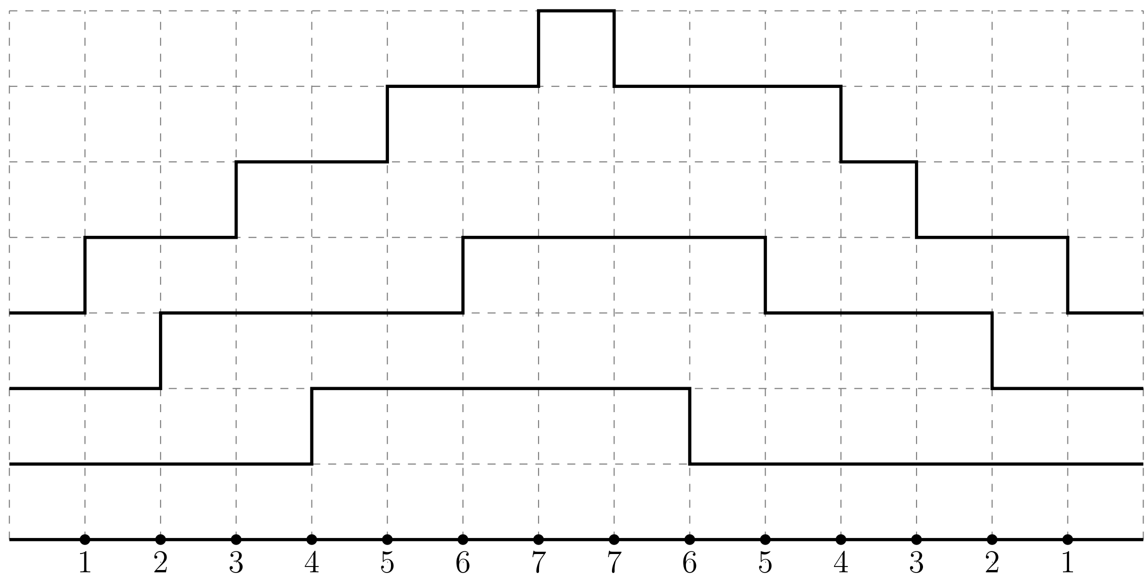
corresponds to the pair of Young tableaux

$$\begin{array}{|c|c|c|c|} \hline 1 & 3 & 5 & 7 \\ \hline 2 & 6 \\ \cline{1-2} 4 \\ \cline{1-1} \end{array}\,, \qquad \begin{array}{|c|c|c|c|} \hline 1 & 3 & 4 & 7 \\ \hline 2 & 5 \\ \cline{1-2} 6 \\ \cline{1-1} \end{array}\,.$$

Note that $L_7(\sigma) = 4$ is the length of the first row.

Relation to non-intersecting paths

Pairs $(Q, P) \in \mathrm{Tab}\,\lambda \times \mathrm{Tab}\,\lambda$ can be identified with $r = \#\mathrm{rows}(\lambda)$ paths. Q gives the positions of where to go up and P of where to go down; the conditions on the Young tableau guarantee that the paths will be non-intersecting; e.g. the pair corresponding to the σ from above gives the following non-intersecting paths:



10 The circular law

The non-selfadjoint analogue of GUE is given by the Ginibre ensemble, where all entries are independent and complex Gaussians. A standard complex Gaussian is of the form

$$z = \frac{x + iy}{\sqrt{2}},$$

where x and y are independent standard real Gaussians, i.e., with joint distribution

$$p(x, y)\, \mathrm{d}x\, \mathrm{d}y = \frac{1}{2\pi} e^{-\frac{x^2}{2}} e^{-\frac{y^2}{2}}\, \mathrm{d}x\, \mathrm{d}y.$$

If we rewrite this in terms of a density with respect to the Lebesgue measure for real and imaginary part

$$z = \frac{x + iy}{\sqrt{2}} = t_1 + it_2, \qquad \overline{z} = \frac{x - iy}{\sqrt{2}} = t_1 - it_2,$$

we get

$$p(t_1, t_2)\, \mathrm{d}t_1\, \mathrm{d}t_2 = \frac{1}{\pi} e^{-(t_1^2 + t_2^2)}\, \mathrm{d}t_1\, \mathrm{d}t_2 = \frac{1}{\pi} e^{-|z|^2}\, \mathrm{d}^2 z,$$

where $\mathrm{d}^2 z = \mathrm{d}t_1\, \mathrm{d}t_2$.

Definition 10.1. A **(complex) unnormalized Ginibre ensemble** $A_N = (a_{ij})_{i,j=1}^N$ is given by complex-valued entries with joint distribution

$$\frac{1}{\pi^{N^2}} \exp\left(- \sum_{i,j=1}^N |a_{ij}|^2 \right) \mathrm{d}A = \frac{1}{\pi^{N^2}} \exp\left(- \operatorname{Tr}(AA^*) \right) \mathrm{d}A,$$

where

$$\mathrm{d}A = \prod_{i,j=1}^N \mathrm{d}^2 a_{ij}.$$

Theorem 10.2. *The joint distribution of the complex eigenvalues of an $N \times N$ Ginibre ensemble is given by a density*

$$p(z_1, \ldots, z_N) = c_N \exp\left(- \sum_{k=1}^N |z_k|^2 \right) \prod_{1 \le i < j \le N} |z_i - z_j|^2.$$

Remark 10.3. (i) Note that typically Ginibre matrices are not normal, i.e.,

$$AA^* \neq A^*A.$$

This means that one loses the relation between functions in eigenvalues and traces of functions in the matrix. The latter is what we can control, the former is what we want to understand.

(ii) As in the selfadjoint case the eigenvalues repel, hence there will almost surely be no multiple eigenvalues. Thus we can also in the Ginibre case diagonalize our matrix, i.e., $A = VDV^{-1}$, where $D = \operatorname{diag}(z_1, \dots, z_N)$ contains the eigenvalues. However, V is now not unitary anymore, i.e., eigenvectors for different eigenvalues are in general not orthogonal. We can also diagonalize A^* via

$$A^* = \left(V^{-1}\right)^* D^* V^*,$$

but since $V^{-1} \neq V^*$ (if A is not normal) we cannot diagonalize A and A^* simultaneously. This means that in general, for example $\operatorname{Tr}(AA^*A^*A)$ has no clear relation to

$$\sum_{i=1}^{N} z_i \bar{z}_i \bar{z}_i z_i.$$

Note that $\operatorname{Tr}(AA^*A^*A) \neq \operatorname{Tr}(AA^*AA^*)$ if $AA^* \neq A^*A$, but of course

$$\sum_{i=1}^{N} z_i \bar{z}_i \bar{z}_i z_i = \sum_{i=1}^{N} z_i \bar{z}_i z_1 \bar{z}_i.$$

(iii) Hence there is a subtle point in the proof of Theorem 10.2, as we apparently have rewritten the density $\exp(-\operatorname{Tr}(AA^*))$ as

$$\exp\left(-\sum_{k=1}^{N} |z_k|^2\right).$$

This is okay though, as we alway have

$$\operatorname{Tr}(AA^*) = \sum_{k=1}^{N} |z_k|^2 = \operatorname{Tr}(A^*A).$$

This relies on the fact that we can always bring a matrix via a unitary conjugation in a triangular form $A = UNU^*$, where U is unitary and

$$N = \begin{pmatrix} z_1 & \star & \cdots & \star \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & \star \\ 0 & \cdots & 0 & z_n \end{pmatrix}.$$

Then $A^* = UN^*U^*$ with

$$N^* = \begin{pmatrix} \bar{z_1} & 0 & \cdots & 0 \\ \star & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \star & \cdots & \star & \bar{z_n} \end{pmatrix}$$

and

$$\mathrm{Tr}(AA^*) = \mathrm{Tr}(UNU^*UN^*U^*) = \mathrm{Tr}(NN^*) = \sum_{k=1}^{N} |z_k|^2 .$$

(iv) As for GUE (Theorem 6.15) we can write the Vandermonde density in a determinantal form. The only difference is that we have to replace the Hermite polynomials $H_k(x)$, which orthogonalize the real Gauß distribution, by monomials z^k, which orthogonalize the complex Gauß distribution.

Theorem 10.4. *The joint eigenvalue distribution of the Ginibre ensemble is of the determinantal form*

$$p(z_1, \dots, z_n) = \frac{1}{N!} \det \left(K_N(z_i, z_j)\right)_{i,j=1}^{N}$$

with the kernel

$$K_N(z, w) = \sum_{k=0}^{N-1} f_k(z)\overline{f_k(w)},$$

where

$$f_k(z) = \frac{1}{\sqrt{\pi}} e^{-\frac{1}{2}|z|^2} \frac{1}{\sqrt{k!}} z^k.$$

In particular, for the averaged eigenvalue density of an unnormalized Ginibre eigenvalue matrix we have the density

$$p_N(z) = \frac{1}{N} K_N(z, z) = \frac{1}{N\pi} e^{-|z|^2} \sum_{k=0}^{N-1} \frac{|z|^{2k}}{k!}.$$

Theorem 10.5 (Circular law for the Ginibre ensemble)**.** *The averaged eigenvalue distribution for a normalized Ginibre random matrix $\frac{1}{\sqrt{N}}A_N$ converges for $N \to \infty$ weakly to the uniform distribution on the unit disc of $\mathbb{C}$ with density*

$$z \mapsto \frac{1}{\pi} 1_{\{z \in \mathbb{C};\, |z| \le 1\}}.$$

Proof. The density q of the normalized Ginibre is given by

$$\begin{aligned} q_N(z) &= N p_N\left(\sqrt{N} z\right) \\ &= \frac{1}{\pi} e^{-N|z|^2} \sum_{k=0}^{N-1} \frac{\left(N|z|^2\right)^k}{k!} \end{aligned}$$

For $|z| < 1$ we have

$$\begin{aligned} e^{N|z|^2} - \sum_{k=0}^{N-1} \frac{\left(N|z|^2\right)^k}{k!} &= \sum_{k=N}^{\infty} \frac{\left(N|z|^2\right)^k}{k!} \\ &\leq \frac{\left(N|z|^2\right)^N}{N!} \sum_{l=0}^{\infty} \frac{\left(N|z|^2\right)^l}{(N+1)^l} \\ &\leq \frac{\left(N|z|^2\right)^N}{N!} \frac{1}{1 - \frac{N|z|^2}{N+1}}, \end{aligned}$$

Furthermore, using the lower bound

$$N! \geq \sqrt{2\pi} N^{N+\frac{1}{2}} e^{-N}$$

on $N!$, we calculate

$$\begin{aligned} e^{-N|z|^2} \frac{\left(N|z|^2\right)^N}{N!} &\leq e^{-N|z|^2} N^N |z|^{2N} \frac{1}{\sqrt{2\pi}} \frac{1}{N^{N+\frac{1}{2}}} e^N \\ &= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{N}} e^{-N|z|^2} e^{N \ln|z|^2} e^N \\ &= \frac{1}{\sqrt{2\pi}} \frac{\exp\left[N\left(-|z|^2 + \ln|z|^2 + 1\right)\right]}{\sqrt{N}} \\ &\xrightarrow{N\to\infty} 0. \end{aligned}$$

Here, we used that

$$-|z|^2 + \ln|z|^2 + 1 < 0$$

for $|z| < 1$. Hence we conclude

$$1 - e^{-N|z|^2} \sum_{k=0}^{N-1} \frac{\left(N|z|^2\right)^k}{k!} \leq e^{-N|z|^2} \frac{\left(N|z|^2\right)^N}{N!} \frac{1}{1 - \frac{N|z|^2}{N+1}} \xrightarrow{N\to\infty} 0.$$

Similarly, for $|z| > 1$,

$$\sum_{k=0}^{N-1} \frac{\left(N|z|^2\right)^k}{k!} \leq \frac{(N|z|)^{N-1}}{(N-1)!} \sum_{l=0}^{N-1} \frac{(N-1)^l}{\left(N|z|^2\right)^l}$$
$$\leq \frac{(N|z|)^{N-1}}{(N-1)!} \frac{1}{1 - \frac{N-1}{N|z|^2}}.$$

□

Remark 10.6. (i) The convergence also holds almost surely.

(ii) The circular law also holds for non-Gaussian entries, but proving this is much harder than the extension for the semicircle law from the Gaussian case to Wigner matrices.

Theorem 10.7 (General circular law)**.** *Consider a complex random matrix*

$$A_N = \frac{1}{\sqrt{N}} (a_{ij})_{i,j=1}^N ,$$

where the a_{ij} are independently and identically distributed random variables with variance 1, i.e., $\mathbb{E}\left[|a_{ij}|^2\right] - \mathbb{E}\left[a_{ij}\right]^2 = 1$. (Note that only the existence of the second moment is required, higher moments don't need to be finite.) Then the eigenvalue distribution of A_N converges weakly almost surely for $N \to \infty$ to the uniform distribution on the unit circle.

10.8 History of the proof

- 60', Mehta: in expectation for Ginibre ensemble
- 80', Silverstein: almost sure convergence for Ginibre
- 80', 90's, Girko: ideas for a proof in the general case
- 1997, Bai: first rigorous proof under additional assumptions on the distribution
- papers by Tao-Vu, Götze-Tikhomirov, Pan-Zhou and others improving more and more on the optimal assumptions
- 2010, Tao-Vu: final version under the assumptions of the existence of the second moment

Remark 10.9. (i) For measures on $\mathbb{C}$ one can use $*$-moments and the Stieltjes transform to describe them, but controlling the convergence properties is the main problem.

(ii) For a matrix A its $*$-moments are all expressions of the form $\operatorname{tr}\left(A^{\varepsilon(1)} \cdots A^{\varepsilon(m)}\right)$, where $m \in \mathbb{N}$ and $\varepsilon(1), \ldots, \varepsilon(m) \in \{1, *\}$. The eigenvalue distribution

$$\mu_A = \frac{1}{N} \left(\delta_{\lambda_1} + \cdots + \delta_{\lambda_N}\right)$$

of A is uniquely determined by the knowledge of all $*$-moments of A, but convergence of $*$-moments does not necessarily imply convergence of the eigenvalue distribution.
Example. Consider

$$A_N = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ \vdots & & & \ddots & 1 \\ 0 & \cdots & \cdots & \cdots & 0 \end{pmatrix} \qquad \text{and} \qquad B_N = \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & \ddots & 0 \\ 0 & & & \ddots & 1 \\ 1 & 0 & \cdots & \cdots & 0 \end{pmatrix}.$$

Then $\mu_{A_N} = \delta_0$, but μ_{B_N} is the uniform distribution on the N-th roots of unity. Hence $\mu_{A_N} \to \delta_0$, whereas μ_{B_N} converges to the uniform distribution on the unit circle. However, the limits of the $*$-moments are the same for A_N and B_N.

(iii) For each measure μ on $\mathbb{C}$ one has the Cauchy-Stieltjes transform

$$m_\mu(u) = \int_{\mathbb{C}} \frac{1}{\lambda - z} \,\mathrm{d}\mu(\lambda).$$

This is almost surely defined. However, it is analytic in z only outside the support of μ. In order to recover μ from m_μ one also needs the information about m_μ inside the support. In order to determine and deal with μ_A one reduces it via Girko's "hermitization method"

$$\int_{\mathbb{C}} \log|\lambda - z| \,\mathrm{d}\mu_A(z) = \int_0^t \log t \,\mathrm{d}\mu_{|A-\lambda 1|}(t)$$

to selfadjoint matrices. There, the left hand side for all λ determines μ_A and the right hand side is about selfadjoint matrices

$$|A - \lambda 1| = \sqrt{(A - \lambda 1)(A - \lambda 1)^*}.$$

Note that the eigenvalues of $|B|$ are related to those of

$$\begin{pmatrix} 0 & B \\ B^* & 0 \end{pmatrix}.$$

In this analytic approach one still needs to control convergence properties. For this, estimates of probabilities of small singular values are crucial.

$\rightarrow$ Survey of Bordenave / Chafai, *Around the circular law*

11 The Marchenko-Pastur law

Let $X_1, \dots, X_p \in \mathbb{R}^N$ with $X_i = (X_{1,i}, \dots, X_{N,i})^T$ be independently identically distributed such that

$$\mathbb{E}[X_i] = \begin{pmatrix} \mathbb{E}[X_{1,i}] \\ \vdots \\ \mathbb{E}[X_{N,i}] \end{pmatrix} = 0$$

and $\mathbb{E}\left[X_i X_i^T\right] = \Sigma_N$. Σ_N is called the **covariance matrix**. It is positive semi-definite and of rank 1. Consider the matrix

$$B_N = B_{N,p} = \frac{1}{p} \sum_{k=1}^{p} X_k X_k^T \in M_N(\mathbb{R}).$$

B_N is a random positive semi-definite matrix and $\mathbb{E}[B_N] = \Sigma_N$. Assume N is fixed, then by the law of large numbers applied entry-wise we get that

$$\lim_{p \to \infty} B_N = \lim_{p \to \infty} \frac{1}{p} \sum_{k=1}^{p} X_k X_k^T = \Sigma_N$$

almost surely. In this case, B_N is a good estimate of the covariance matrix Σ_N. Then a natural question to ask is what would be the behavior of B_N when both N and p go to infinity? In fact, we are in an era where an increasingly large volume of complex data is generated and tools from classical multivariate statistics are not enough to analyze high-dimensional data.

Definition 11.1. A **Wishart matrix** $B_{N,p}$ is an $N \times N$ matrix of the form

$$B_{N,p} = \frac{1}{p} \mathcal{X}_{N,p} \mathcal{X}_{N,p}^T,$$

where $\mathcal{X}_{N,p}$ is an $N \times p$ matrix with independently identically distributed centered entries of variance 1.

To simplify the notation we will drop the index p and just write B_N. Note that B_N can also be written as

$$B_N = \frac{1}{p} \sum_{k=1}^{p} X_k X_k^T,$$

where $X_1, \dots, X_N$ are the columns of $\mathcal{X}_N$. Recall that the empirical spectral measure is given by

$$\mu_{B_N} = \frac{1}{N} \sum_{k=1}^{N} \delta_{\lambda_k}$$

with $\lambda_1, \dots, \lambda_N$ the eigenvalues of B_N. We are interested in the case that $N, p \to \infty$ at the same pace, i.e.,

$$\frac{N}{p} \xrightarrow{N,p\to\infty} c \in (0, \infty).$$

Theorem 11.2 (Marchenko-Pastur, 1967)**.** *Let $(X_{ij})_{i,j}$ be a family of independently identically distributed random variables such that $\mathbb{E}[X_{11}] = 0$ and $\mathbb{E}[X_{11}^2] = \sigma^2 < \infty$. Provided that $N, p \to \infty$ such that $\frac{N}{p} \to c \in (0, \infty)$ then*

$$\mu_{B_N} \xrightarrow{w} \mu_{MP} \qquad \textit{almost surely,}$$

whose density is given by

$$\left(1 - \frac{1}{c}\right)_+ \delta_0 + \frac{1}{2\pi c \sigma^2 x} \sqrt{(\lambda^+ - x)(x - \lambda^-)} 1_{[\lambda^-, \lambda^+]}(x) \, \mathrm{d}x,$$

where $(\cdot)_+ = \max(0, \cdot)$ and $\lambda^\pm = \sigma^2(1 \pm \sqrt{c})^2$.

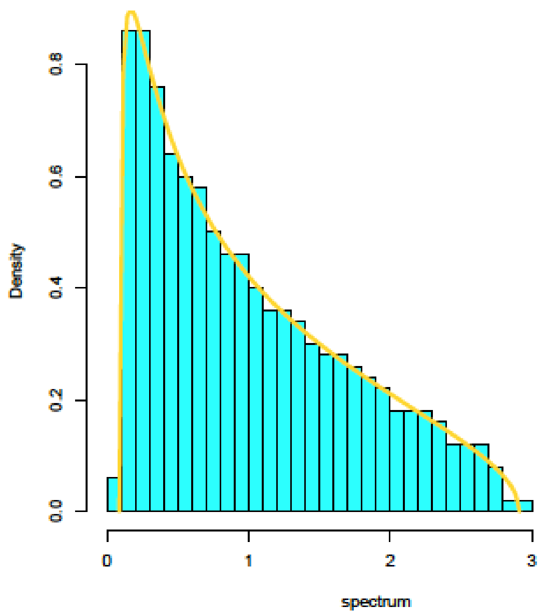


Figure 11.1: Histogram of the eigenvalues of a Wishart matrix with $N = 500$ and $p = 1000$. In yellow, the density of the Marchenko-Pastur distribution for $c = 0.5$.

Remark 11.3. (i) If $c > 1$ then $1 - \frac{1}{c} > 0$ and we have an atom at 0 with mass $1 - \frac{1}{c}$. Since $\operatorname{rank} B_N = \min(N, p)$, B_N has $N - p$ zero eigenvalues such that

$$\mu_{B_N} = \frac{N-p}{N}\delta_0 + \frac{1}{N}\sum_{k=1}^{p}\delta_{\lambda_k}$$

with

$$\frac{N-p}{N} = 1 - \frac{1}{\frac{N}{p}} \to 1 - \frac{1}{c}.$$

If $c < 1$ and $N < p$ then B_N is invertible and does not have zero eigenvalues.

(ii) Observe that the non-zero eigenvalues of XX^T and X^TX are the same so that we have

$$\mu = \left(1 - \frac{1}{c}\right)\delta_0 + \tilde{\mu},$$

where $\tilde{\mu}$ is the limiting distribution of X^TX.

(iii) Apart from the Dirac measure at 0, the support of μ_{MP} is compact and is spread on an interval of length $4\sigma^2\sqrt{c}$ around the variance σ^2.

(iv) The Marchenko-Pastur theorem is a universality result in the sense that the limiting distribution depends on the distribution of the entries only through the variance σ^2.

(v) The mean and variance of the Marchenko-Pastur distribution are

$$\int_{\mathbb{R}} x \,\mathrm{d}\mu_{\mathrm{MP}}(x) = \sigma^2,$$

$$\int_{\mathbb{R}} x^2 \,\mathrm{d}\mu_{\mathrm{MP}}(x) - \left(\int_{\mathbb{R}} x \,\mathrm{d}\mu_{\mathrm{MP}}(x)\right)^2 = \frac{\sigma^4}{c}.$$

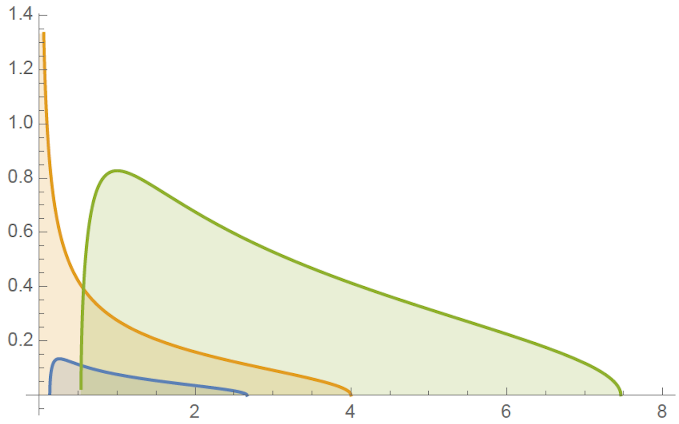


Figure 11.2: The Marchenko-Pastur distribution density function for $\sigma^2 = 1$. In green $c = 3$, in orange $c = 1$ and in blue $c = 0.4$. The mass of the distribution in the last case is 0.4, to which we should add a Dirac mass at 0 with mass 0.6.

(vi) The Stieltjes transform of μ_{MP} is given by

$$g_{\mu_{\mathrm{MP}}}(z) = \frac{\sigma^2(1-c) - z + \sqrt{(z+\lambda^+)(z-\lambda^-)}}{2cz\sigma^2}.$$

It is a solution of the quadratic equation

$$zc\sigma^2 g(z)^2 + \left[z + \sigma^2(c-1)\right] g(z) + 1 = 0.$$

(vii) As for the Wigner semicircle law, the Marchenko-Pastur theorem can be proven via the moment method or analytically via the resolvent method.

Moment method: Since the support of μ_{MP} is compact, μ_{MP} is uniquely determined by its moments and it is enough to show hat

$$\int x^k \mu_{B_N}(x) \xrightarrow{N\to\infty} \int x^k \mu_{\mathrm{MP}}(x) \qquad \text{almost surely}$$

for any $k \geq 1$. For this, it suffices to show that

- $\mathbb{E}\left[\int x^k \mu_{B_N}(x)\right] \xrightarrow{N\to\infty} \int x^k \mu_{\mathrm{MP}}(x)$,
- $\mathbb{V}\mathrm{ar}\left[\int x^k \mu_{B_N}(x)\right] \leq \frac{C_k}{N^2}$,

and then invoke the Borel-Cantelli lemma.

Resolvent method: Show that for all $z \in \mathbb{C}_+$,

$$g_{B_N}(z) \xrightarrow{N\to\infty} g_{\mathrm{MP}}(z) \qquad \text{almost surely.}$$

To do this, we proceed as follows:

(1) $g_{B_N}(z) - \mathbb{E}\left[g_{B_N}(z)\right] \xrightarrow{N\to\infty} 0$ almost surely (concentration)

(2) We prove that $\mathbb{E}\left[g_{B_N}(z)\right]$ satisfies the equation

$$zc\sigma^2 g(z)^2 + \left[z + \sigma^2(c-1)\right] g(z) + 1 = \varepsilon_N \tag{$\star$}$$

with $\varepsilon_N \xrightarrow{N\to\infty} 0$.

(3) We prove the stability of ($\star$).

"Close equation implies close solution."

Theorem 11.4 (Bai-Yin)**.** *Let $B_N = \frac{1}{p}\mathcal{X}_N\mathcal{X}_N^T$ with $\mathcal{X}_N = (X_{ij})_{i,j}$, where the X_{ij} are independently identically distributed, centered, have variance σ^2 and $\mathbb{E}\left[X_{11}^4\right] < \infty$. Then provided that $\frac{N}{p} \to c \in (0,\infty)$,*

$$\lambda_{\max}(B_N) \xrightarrow{N\to\infty} \sigma^2(1+\sqrt{c})^2 \qquad \textit{almost surely,}$$
$$\lambda_{\min}(B_N) \xrightarrow{N\to\infty} \sigma^2(1-\sqrt{c})^2 \qquad \textit{almost surely.}$$

Remark 11.5. (i) The Marchenko-Pastur theorem gives a lower bound for $\lambda_{\max}$:

$$\lambda_{\max}(B_N) \geq \sigma^2(1+\sqrt{c})^2$$

(ii) The condition $\mathbb{E}\left[X_{11}^4\right] < \infty$ is necessary for the convergence of $\lambda_{\max}$ to the edge of the bulk.

11.1 Large covariance matrix

Let $\mathcal{X}_N = (X_1 \ldots, X_N)$ be a $N \times p$ matrix with independently identically distributed centered entries of variance 1. Let R_N be a deterministic hermitian positive semidefinite $N \times N$ matrix. Then

$$\mathcal{Y}_N = R_N^{\frac{1}{2}}\mathcal{X}_N = \left(R_N^{\frac{1}{2}}X_1, \ldots, R_N^{\frac{1}{2}}X_p\right)$$

represents p independently identically distributed sample observations. Let us denote $Y_i = R_N^{\frac{1}{2}} X_i \in \mathbb{R}^N$ and note that $\mathbb{E}[Y_i] = 0$ and $\mathbb{E}\left[Y_i Y_i^T\right] = R_N$. If N is fixed then by the law of large numbers,

$$B_N = \frac{1}{p} R_N^{\frac{1}{2}} \mathcal{X}_N \mathcal{X}_N^T R_N^{\frac{1}{2}} \xrightarrow{p\to\infty} R_N \qquad \text{almost surely.}$$

This situation is sensibly different in the asymptotic setting where $N, p \to \infty$ and $\frac{N}{p} \to c \in (0, \infty)$.

Theorem 11.6. *Assume that the empirical spectral measure μ_{R_N} converges weakly to ν for $N \to \infty$. Then, provided that $N, p \to \infty$ such that $\frac{N}{p} \to c \in (0, \infty)$:*

(i) *The equation*

$$t(z) = \int \frac{1}{-z[1 - \lambda c t(z)] + (1-c)\lambda} \,\mathrm{d}\nu(\lambda), \qquad z \in \mathbb{C}_+,$$

has a unique solution $z \mapsto t(z)$, which is the Stieltjes transform of a probability measure μ, i.e.,

$$t(z) = \int \frac{1}{\lambda - z} \,\mathrm{d}\mu(z).$$

(ii) *For all $z \in \mathbb{C}_+$, $g_{B_N(z)} \to t(z)$ almost surely if and only if $\mu_{B_N} \to \mu$ almost surely.*

Remark 11.7. (i) The entries of the matrix $\mathcal{Y}_N = R_N^{\frac{1}{2}} \mathcal{X}_N$ are no longer independent. In fact, the columns Y_i of $\mathcal{Y}_N$ are independently identically distributed, but we allow correlations between their components.

(ii) The covariance structure of R_N also appears in the limiting distribution through the limiting measure ν.

(iii) In the independently identically distributed case $R_N = \sigma^2 I_N$, the covariance of the entries is zero and only the variance σ^2 shows up in the limiting distribution, the Marchenko-Pastur distribution.

(iv) Contrary to the Marchenko-Pastur distribution, in this case we do not have an explicit equation for the Stieltjes transform $t(z)$. Numerical methods are applied to approximate $t(z)$ and then the inversion formula yields an approximation of the limiting measure μ.

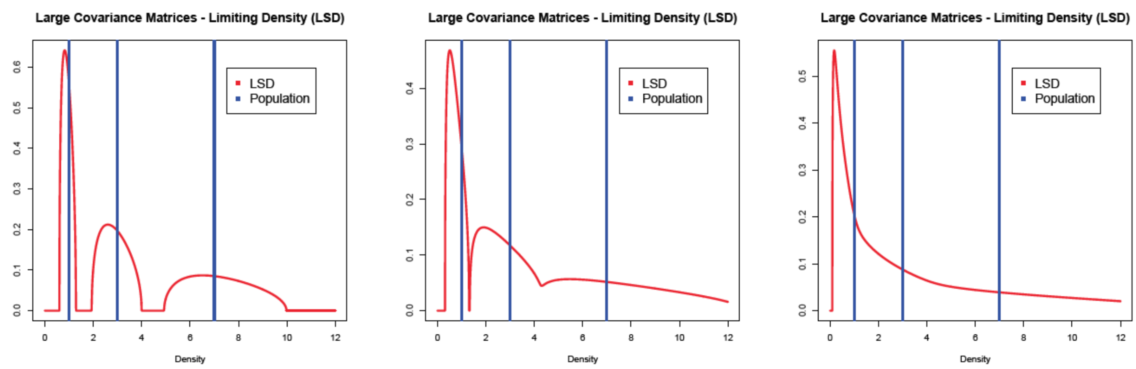


Figure 11.3: In red, the density of the limiting measure for $c = 0.1$, $c = 0.3$ and $c = 0.6$. In blue, the 3 eigenvalues of the covariance matrix of the population R_N, each of equivalent multiplicity.

Remark 11.8. It seems clear that if the ratio c is small, we can somehow guess some information on the covariance matrix R_N (eigenvalues in blue). However, if c is not so small ($c = 0.6$) it is somehow impossible to guess the eigenvalues of R_N from the figure. In other words, the information B_N gives is not as direct as in the conventional setting.

11.2 Small perturbations

Consider the case

$$R_N = \sigma^2 \left(I_N + \sum_{l=1}^{K} \theta_l u_l u_l^T \right),$$

where K is a fixed number, the $\theta_l > 0$ and the u_l are deterministic orthonormal vectors. We have seen that when $R_N = \sigma^2 I_N$ then $\mu_{\frac{\sigma^2}{p} \mathcal{X}_N \mathcal{X}_N^T}$ converges weakly to μ_{MP}. We shall see how the convergence of the largest eigenvalue is affected by the parameters θ_l and u_l. We will consider the case $K = 1$ (rank 1 perturbation),

$$R_N = \sigma^2 \left(I_N + \theta u u^T \right),$$

with $\sigma > 0$ and $\|u\| = 1$.

Theorem 11.9. *Provided that $N, p \to \infty$ such that $\frac{N}{p} \to c \in (0, \infty)$ then:*

(i) *If $\theta \leq \sqrt{c}$ then*

$$\lambda_{\max}(B_N) \xrightarrow{N \to \infty} \sigma^2 \left(1 + \sqrt{c}\right)^2 \qquad \textit{almost surely.}$$

(ii) *If $\theta > \sqrt{c}$ then*

$$\lambda_{\max}(B_N) \xrightarrow{N\to\infty} \sigma^2(1+\theta)\left(1+\frac{c}{\theta}\right) \qquad \textit{almost surely.}$$

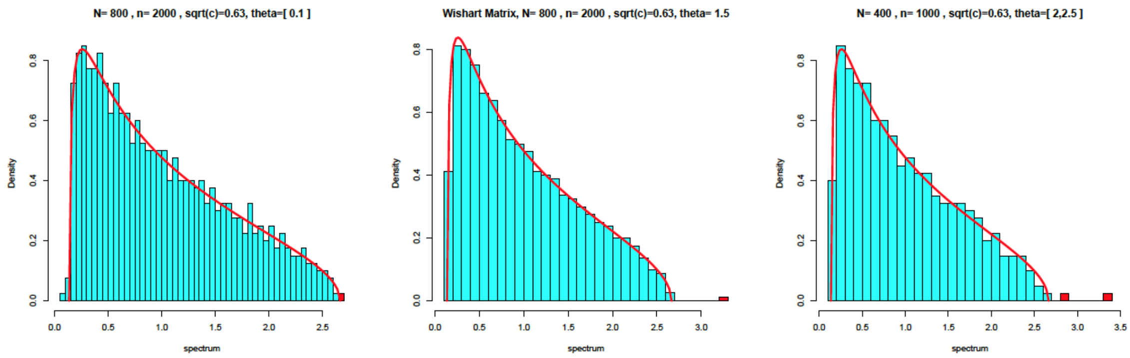


Figure 11.4: In red, the density of the Marchenko-Pastur distribution. In blue, the histogram of the eigenvalues. The red parts represent the largest eigenvalue. The first two simulations correspond to a simple perturbation, whereas the last one corresponds to a double perturbation.

Remark 11.10. (i) We can easily verify that if $\theta > \sqrt{c}$ then

$$\sigma^2(1+\theta)\left(1+\frac{c}{\theta}\right) > \sigma^2\left(1+\sqrt{c}\right)^2.$$

(ii) The intensity of the perturbation θ has an influence on the behavior of the largest eigenvalue. It converges to the edge of the support of the Marchenko-Pastur distribution if θ is sufficiently small and separates otherwise.

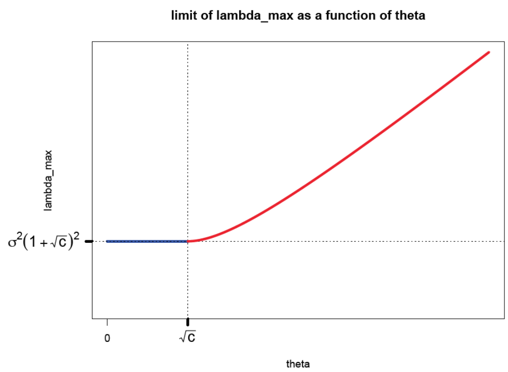


Figure 11.5: The limit of the largest eigenvalue $\lambda_{\max}$ as a function of the perturbation θ.

12 Several independent GUE and asymptotic freeness

Up to now, we have only considered limits $N \to \infty$ of one random matrix A_N, but often one has several matrix ensembles and would like to understand the "joint" distribution, e.g., in order to use them as building blocks for more complicated random matrix models. As an example, $R^{\frac{1}{2}}\mathcal{X}_N\mathcal{X}_N^T R^{\frac{1}{2}}$ is built out of the two random matrices $\mathcal{X}_N$ and R_N. Note that a deterministic matrix like R_N is also a special case of random matrices.

Remark 12.1. (i) Consider two random matrices $A_N = (a_{ij})_{i,j=1}^N$, $B_N = (b_{ij})_{i,j=1}^N$, where the entries a_{ij}, b_{ij} are defined on the same probability space. What do we mean by the joint distribution of the matrices in which we are interested as $N \to \infty$? Note that in general our analytical approach breaks down if A_N and B_N do not commute, since then we cannot diagonalize them simultaneously, hence it makes no sense to talk about a joint distribution of A_N and B_N. The notion μ_{A_N,B_N} has no clear analytic meaning. What still makes sense in the multivariate case is the combinatorial approach via mixed moments with respect to the normalized trace tr. Hence we consider the collection of all **mixed moments**

$$\operatorname{tr}\left(A_{i_1}^{(N)} \cdots A_{i_m}^{(N)}\right)$$

in A_N and B_N, with $m \in \mathbb{N}$, $i_1, \ldots, i_m \in \{1, 2\}$, $A_1^{(N)} = A_N$ and $A_2^{(N)} = B_N$, as the joint distribution of A_N and B_N and denote this by μ_{A_N,B_N}. We want to understand, in interesting cases, the behavior of μ_{A_N,B_N} as $N \to \infty$.

(ii) In the case of one (selfadjoint) matrix A, μ_A has two meanings:

Analytic: $\mu_A = \frac{1}{N}(\delta_{\lambda_1} + \cdots + \delta_{\lambda_N})$ is a probability measure on $\mathbb{R}$.

Combinatorial: μ_A is given by all moments $\operatorname{tr}(A^k)$ for all $k \geq 1$.

These two points of view are the same via

$$\operatorname{tr}(A^k) = \int t^k \,\mathrm{d}\mu_A(t).$$

In the case of two matrices A_1, A_2 the notion μ_{A_1,A_2} has only one meaning, namely the collection of all mixed moments

$$\mathrm{tr}\,(A_{i_1}\cdots A_{i_m})$$

with $m\in\mathbb{N}$ and $i_1,\dots,i_m\in\{1,2\}$. If A_1 and A_2 do not commute then there exists no probability measure μ on $\mathbb{R}^2$ with

$$\mathrm{tr}\,(A_{i_1}\cdots A_{i_m}) = \int t_{i_1}\cdots t_{i_m}\,\mathrm{d}\mu(t_1,t_2)$$

for all $m\in\mathbb{N}$ and $i_1,\dots,i_m\in\{1,2\}$.

12.1 Joint moments of independent GUEs

We will now consider the simplest case of several random matrices, namely r GUEs $A_1^{(N)},\dots,A_r^{(N)}$, which we assume to be independent of each other, i.e., we have

$$A_i^{(N)} = \frac{1}{\sqrt{N}}\left(a_{kl}^{(i)}\right)_{k,l=1}^{N},$$

where $i=1,\dots,r$, each $A_i^{(N)}$ is a GUE and

$$\left\{a_{kl}^{(1)};\,k,l=1,\dots,N\right\},\dots,\left\{a_{kl}^{(r)};\,k,l=1,\dots,N\right\}$$

are independent sets of Gaussian random variables. Equivalently, this can be characterized by the requirement that all entries of all matrices together form a collection of independent standard Gaussian variables (real on the diagonal, complex otherwise). Hence we can express this again in terms of the Wick formula as

$$\mathbb{E}\left[a_{k_1l_1}^{(i_1)}\cdots a_{k_ml_m}^{(i_m)}\right] = \sum_{\pi\in\mathcal{P}_2(m)}\mathbb{E}_\pi\left[a_{k_1l_1}^{(i_1)},\dots,a_{k_ml_m}^{(i_m)}\right]$$

for all $m\in\mathbb{N}$, $1\le k_1,l_1,\dots,k_m,l_m\le N$ and $1\le i_1,\dots,i_m\le r$ and where the second moments are given by

$$\mathbb{E}\left[a_{pq}^{(i)}a_{kl}^{(j)}\right] = \delta_{pl}\delta_{qk}\delta_{ij}.$$

Now we can essentially repeat the calculations from Remark 2.14 for our mixed moments:

$$\begin{aligned}
\mathbb{E}\left[\operatorname{tr}(A_{i_1}\cdots A_{i_m})\right] &= \frac{1}{N^{1+\frac{m}{2}}}\sum_{k_1,\dots,k_m=1}^{N}\mathbb{E}\left[a^{(i_1)}_{k_1k_2}a^{(i_2)}_{k_2k_3}\cdots a^{(i_m)}_{k_mk_1}\right]\\
&= \frac{1}{N^{1+\frac{m}{2}}}\sum_{k_1,\dots,k_m=1}^{N}\sum_{\pi\in\mathcal{P}_2(m)}\mathbb{E}_\pi\left[a^{(i_1)}_{k_1k_2},a^{(i_2)}_{k_2k_3},\dots,a^{(i_m)}_{k_mk_1}\right]\\
&= \frac{1}{N^{1+\frac{m}{2}}}\sum_{k_1,\dots,k_m=1}^{N}\sum_{\pi\in\mathcal{P}_2(m)}\prod_{(p,q)\in\pi}\mathbb{E}\left[a^{(i_p)}_{k_pk_{p+1}}a^{(i_q)}_{k_qk_{q+1}}\right]\\
&= \frac{1}{N^{1+\frac{m}{2}}}\sum_{k_1,\dots,k_m=1}^{N}\sum_{\pi\in\mathcal{P}_2(m)}\prod_{(p,q)\in\pi}\left[k_p=k_{q+1}\right]\left[k_q=k_{p+1}\right]\left[i_p=i_q\right]\\
&= \frac{1}{N^{1+\frac{m}{2}}}\sum_{\substack{\pi\in\mathcal{P}_2(m)\\(p,q)\in\pi\\i_p=i_q}}\sum_{k_1,\dots,k_m=1}^{N}\prod_{p}\left[k_p=k_{\gamma\pi(p)}\right]\\
&= \frac{1}{N^{1+\frac{m}{2}}}\sum_{\substack{\pi\in\mathcal{P}_2(m)\\(p,q)\in\pi\\i_p=i_q}}N^{\#(\gamma\pi)},
\end{aligned}$$

where $\gamma=\begin{pmatrix}1 & 2 & \cdots & m\end{pmatrix}$ is the shift by 1 modulo m. Hence we get the same kind of genus expansion for several GUEs as for one GUE. The only difference is, that in our pairing we only allow to connect the same matrices.

Notation 12.2. For a given $i=(i_1,\dots,i_m)$ with $1\le i_1,\dots,i_m\le r$ we say that $\pi\in\mathcal{P}_2(m)$ **respects** i if we have $i_p=i_q$ for all $(p,q)\in\pi$. We put

$$\mathcal{P}_2^{[i]}(m)=\{\pi\in\mathcal{P}_2(m);\ \pi\text{ respects }i\}$$

and also

$$\mathcal{NC}_2^{[i]}(m)=\{\pi\in\mathcal{NC}_2(m);\ \pi\text{ respects }i\}.$$

Theorem 12.3 (Genus expansion of independent GUEs)**.** *Let $A_1,\dots,A_r$ be r independent $N\times N$ GUEs. Then we have for all $m\in\mathbb{N}$ and all $i_1,\dots,i_m\in[r]$ that*

$$\mathbb{E}\left[\operatorname{tr}(A_{i_1}\cdots A_{i_m})\right]=\sum_{\pi\in\mathcal{P}_2^{[i]}(m)}N^{\#(\gamma\pi)-\frac{m}{2}-1}$$

and thus

$$\lim_{N\to\infty}\mathbb{E}\left[\operatorname{tr}(A_{i_1}\cdots A_{i_m})\right]=\#\mathcal{NC}_2^{[i]}(m).$$

Proof. The genus expansion follows from our computation before. The limit for $N \to \infty$ follows as in Theorem 2.12 from the fact that

$$\lim_{N\to\infty} N^{\#(\gamma\pi)-\frac{m}{2}-1} = \begin{cases} 1, & \pi \in \mathcal{NC}_2(m), \\ 0, & \pi \notin \mathcal{NC}_2(m). \end{cases}$$

The index tuple $(i_1, \ldots, i_m)$ has no say in this limit. □

Remark 12.4. We would like to find some structure in the limiting moments. We prefer to talk directly about the limit instead of making asymptotic statements. In the case of one GUE, we had the semicircle μ_W as a limiting analytic object. Now we do not have an analytic object in the limit, but we can organize our distribution as the limit of moments in a more algebraic way.

Definition 12.5. (i) Let $\mathcal{A} = \mathbb{C}\langle s_1, \ldots, s_r \rangle$ be the algebra of polynomials in the non-commuting variables $s_1, \ldots, s_r$. I.e., there are no non-trivial relations between $s_1, \ldots, s_r$ and it is the linear span of the monomials $s_{i_1} \cdots s_{i_m}$ for $m \geq 0$ and $i_1, \ldots, i_m \in [r]$. Multiplication for monomials is given by concatenation.

(ii) On this algebra $\mathcal{A}$ we define a unital linear functional $\varphi\colon \mathcal{A} \to \mathbb{C}$ by $\varphi(1) = 1$ and

$$\varphi(s_{i_1} \cdots s_{i_m}) = \lim_{N\to\infty} \mathbb{E}\left[\operatorname{tr}(A_{i_1} \cdots A_{i_m})\right] = \#\mathcal{NC}_2^{[i]}(m).$$

(iii) We also address $(\mathcal{A}, \varphi)$ as a **non-commutative probability space** and $s_1, \ldots, s_r \in \mathcal{A}$ as **non-commutative random variables**. The **moments** of $s_1, \ldots, s_r$ are the $\varphi(s_{i_1} \cdots s_{i_m})$ and the collection of those moments is the **(joint) distribution** of $s_1, \ldots, s_r$.

Remark 12.6. (i) Note that if we consider only one of the s_i, then its distribution is just the collection of Catalan numbers, which we understand quite well

(ii) If we consider all $s_1, \ldots, s_r$, then their joint distribution is a large collection of numbers. We claim that the following theorem discovers some important structure in those.

Theorem 12.7. *Let $\mathcal{A} = \mathbb{C}\langle s_1, \ldots, s_r \rangle$ and let $\varphi\colon \mathcal{A} \to \mathbb{C}$ be defined by*

$$\varphi(s_{i_1} \cdots s_{i_m}) = \#\mathcal{NC}_2^{[i]}(m)$$

as before. Then for all $m \geq 1$, $i_1, \ldots, i_m \in [r]$ with

$$i_1 \neq i_2,\ i_2 \neq i_3,\ \ldots,\ i_{m-1} \neq i_m$$

and all polynomials $p_1, \dots, p_m$ in one variable such that

$$\varphi\left(p_k(s_{i_k})\right) = 0$$

we have that

$$\varphi\left(p_1(s_{i_1}) \cdots p_m(s_{i_m})\right) = 0.$$

In other words: The alternating product of centered variables is centered.

We say that $s_1, \dots, s_r$ are *free (or freely independent)*; in terms of the random matrices, we say that $A_1^{(N)}, \dots, A_r^{(N)}$ are *asymtotially free.*

Proof. It suffices to prove the statement for polynomials of the form

$$p_k(s_{i_k}) = s_{i_k}^{p_k} - \varphi\left(s_{i_k}^{p_k}\right)$$

for any power p_k, since general polynomials can be written as linear combinations of those. The general statement then follows by linearity. So we have to prove that

$$\varphi\left[\left(s_{i_1}^{p_1} - \varphi\left(s_{i_1}^{p_1}\right)\right) \cdots \left(s_{i_m}^{p_m} - \varphi\left(s_{i_m}^{p_m}\right)\right)\right] = 0.$$

We have

$$\varphi\left[\left(s_{i_1}^{p_1} - \varphi\left(s_{i_1}^{p_1}\right)\right) \cdots \left(s_{i_m}^{p_m} - \varphi\left(s_{i_m}^{p_m}\right)\right)\right] = \sum_{M \subset [m]} (-1)^{|M|} \prod_{j \in M} \varphi\left(s_{i_j}^{p_j}\right) \varphi\left(\prod_{j \notin M} s_{i_j}^{p_j}\right)$$

with

$$\varphi\left(s_{i_j}^{p_j}\right) = \varphi\left(s_{i_j} \cdots s_{i_j}\right) = \#\mathcal{NC}_2(p_j)$$

and

$$\varphi\left(\prod_{j \notin M} s_{i_j}^{p_j}\right) = \#\mathcal{NC}_2^{\text{[respects indices]}}\left(\sum_{j \notin M} p_j\right).$$

Let us put

$$\begin{aligned} I_1 &= \{1, \dots, p_1\} \\ I_2 &= \{p_1 + 1, \dots, p_1 + p_2\} \\ &\vdots \\ I_m &= \{p_1 + p_2 + \dots + p_{m-1}, \dots, p_1 + p_2 + \dots + p_m\} \end{aligned}$$

and $I = I_1 \cup I_2 \cup \cdots \cup I_m$. Denote

$$[\ldots] = [i_1, \ldots, i_1, i_2, \ldots, i_2, \ldots, i_m, \ldots, i_m].$$

Then

$$\prod_{j \in M} \varphi\left(s_{i_j}^{p_j}\right) \varphi\left(\prod_{j \notin M} s_{i_j}^{p_j}\right) = \#\{\pi \in \mathcal{NC}_2^{[\ldots]}(I); \text{ for all } j \in M \text{ all elements in } I_j \text{ are only paired amongst each other}\}$$

Let us denote

$\mathcal{NC}_2^{[\ldots]}(I : j) = \{\pi \in \mathcal{NC}_2^{[\ldots]}(I);$ elements in I_j are only paried amongst each other$\}$.

Then, by the inclusion-exclusion formula,

$$\begin{aligned}\varphi\left[\left(s_{i_1}^{p_1} - \varphi\left(s_{i_1}^{p_1}\right)\right) \cdots \left(s_{i_m}^{p_m} - \varphi\left(s_{i_m}^{p_m}\right)\right)\right] &= \sum_{M \subset [m]} (-1)^{|M|} \cdot \#\left(\bigcap_{j \in M} \mathcal{NC}_2^{[\ldots]}(I : j)\right) \\ &= \#\left(\mathcal{NC}_2^{[\ldots]}(I) \setminus \bigcup_j \mathcal{NC}_2^{[\ldots]}(I : j)\right).\end{aligned}$$

These are $\pi \in \mathcal{NC}_2^{[\ldots]}(I)$ such that at least one element of each interval I_j is paired with an element from another interval I_k. Since

$$i_1 \neq i_2,\ i_2 \neq i_3,\ \ldots,\ i_{m-1} \neq i_m$$

we cannot connect neighboring intervals and each interval must be connected to another interval in a non-crossing way. But there is no such π, hence

$$\varphi\left[\left(s_{i_1}^{p_1} - \varphi\left(s_{i_1}^{p_1}\right)\right) \cdots \left(s_{i_m}^{p_m} - \varphi\left(s_{i_m}^{p_m}\right)\right)\right] = \#\left(\mathcal{NC}_2^{[\ldots]}(I) \setminus \bigcup_j \mathcal{NC}_2^{[\ldots]}(I : j)\right) = 0$$

as claimed. □

Remark 12.8. (i) Note that in Theorem 12.7 we have traded the explicit description of our moments for implicit relations between the moments.

(ii) For example, the simplest relations from Theorem 12.7 are

$$\varphi\left([s_i^p - \varphi(s_i^p)1][s_j^q - \varphi(s_j^q)1]\right) = 0,$$

for $i \neq j$, which can be reformulated to

$$\varphi(s_i^p s_j^q) - \varphi(s_i^p 1)\varphi(s_j^q) - \varphi(s_i^p)\varphi(s_j^q 1) + \varphi(s_i^p)\varphi(s_j^q)\varphi(1) = 0,$$

i.e.,

$$\varphi(s_i^p s_j^q) = \varphi(s_i^p)\varphi(s_j^q).$$

Those relations are quickly getting more complicated. For example,

$$\varphi\left[(s_1^{p_1} - \varphi(s_1^{p_1})1)(s_2^{q_1} - \varphi(s_2^{q_1})1)(s_1^{p_2} - \varphi(s_1^{p_2})1)(s_2^{q_2} - \varphi(s_2^{q_2})1)\right] = 0$$

leads to

$$\begin{aligned}\varphi\left(s_1^{p_1} s_2^{q_1} s_1^{p_2} s_2^{q_2}\right) = {} & \varphi\left(s_1^{p_1+p_2}\right)\varphi\left(s_2^{q_1}\right)\varphi\left(s_2^{q_2}\right) \\ & + \varphi\left(s_1^{p_1}\right)\varphi\left(s_1^{p_2}\right)\varphi\left(s_2^{q_1+q_2}\right) \\ & - \varphi\left(s_1^{p_1}\right)\varphi\left(s_2^{q_1}\right)\varphi\left(s_1^{p_2}\right)\varphi\left(s_2^{q_2}\right).\end{aligned}$$

(iii) So one might ask: What is it good for to find those relations between the moments, if we know the moments in a more explicit form anyhow?

Answer: Those relations occur in many more situations. For example, independent Wishart matrices satisfy the same relations, even though the explicit form of their mixed moments is quite different from the GUE case. Furthermore, we can control what happens with these relations much better than with the explicit moments if we deform our setting or construct new random matrices out of other ones. Not to mention that those relations also show up in very different corners of mathematics (like operator algebras).

To make a long story short: Those relations from Theorem 12.7 are really worth being investigated further, not just in a random matrix context, but also for its own sake. This will be done in the lecture *Free Probability Theory* next term!