Random Matrices

1. Introduction

1.1. 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 atomic nucleus
  studied in particular asymptotic for $N \to \infty$
  "large N limit"

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

- since 1960's random matrices are important tool in physics ⇝ quantum chaos
  ⇝ universality

  Mehta, Dyson

  ↳ influential (first) book
  "Random Matrices" 1967

- ~ 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
  ⇝ Tracy-Widom distribution of largest eigenvalue
  ⇝ free probability theory
  ⇝ universality of fluctuations
  ⇝ "circular law"

  etc...

1.2. 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)

simple example: Choose $a_{ij} \in \{-1, +1\}$ with $a_{ij} = a_{ji} \quad \forall i,j$

Consider all such matrices and ask for typical or generic behaviour!

[in a more probabilistic language: all allowed matrices have the same probability]

1.3. Quantity of interest:

We are mainly interested in the eigenvalues of the matrices.

Consider situation from above with $a_{ij} \in \{\pm 1\}$ for different size N of matrices

$N = 1$:

matrix	eigenvalues	probability
$A_1 = (1)$	$+1$	$1/2$
$A_2 = (-1)$	-1	$1/2$

$N=2$:

matrix	eigenvalues	probability
$A_1 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$	$0, 2$	$1/8$
$A_2 = \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix}$	$-\sqrt{2}, \sqrt{2}$	$1/8$
$A_3 = \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$	$0, 2$	
$A_4 = \begin{pmatrix} -1 & 1 \\ 1 & 1 \end{pmatrix}$	$-\sqrt{2}, \sqrt{2}$	
$A_5 = \begin{pmatrix} 1 & -1 \\ -1 & -1 \end{pmatrix}$	$-\sqrt{2}, \sqrt{2}$	
$A_6 = \begin{pmatrix} -1 & 1 \\ 1 & -1 \end{pmatrix}$	$-2, 0$	
$A_7 = \begin{pmatrix} -1 & -1 \\ -1 & 1 \end{pmatrix}$	$-\sqrt{2}, \sqrt{2}$	
$A_8 = \begin{pmatrix} -1 & -1 \\ -1 & -1 \end{pmatrix}$	$-2, 0$	$1/8$

general N: for N there are

$2^{N(N+1)/2}$ such matrices

(each with probability $2^{-N(N+1)/2}$)

there are always very special ones, like

$A = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ 1 & 1 & & \\ 1 & & \ddots & \\ 1 & & & 1 \end{pmatrix}$ eigenvalues: $0, N$

Σ multiplicity $N-1$

or

$$A = \begin{pmatrix} -1 & -1 & \cdots & -1 \\ -1 & -1 & & \\ \vdots & & & \\ -1 & -1 & \cdots & -1 \end{pmatrix}$$ eigenvalues: $0, -N$

↑ multiplicity $N-1$

they have small probability and are "untypical".

Question: What is the "typical" behaviour of the eigenvalues?

→ computer pictures

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

This is a statement on global behaviour of eigenvalues, not of single eigenvalues.

1.5. Universality: This statement is valid much more general: Choose the a_{ij} not just from $\{\pm 1\}$, but, for example,

- $a_{ij} \in \{1, 2, 3, 4, 5, 6\}$

or

- a_{ij} normally ($G_{\alpha\beta}$) distributed

(hence continuous)

- a_{ij} distributed according to your favorite distribution

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

1.6 Concentration phenomena: The (quite amazing) fact that the apriori random eigenval distribution is, for $N \to \infty$, not random any more, but concentrates on one deterministic distribution (namely the semicircle) is an example of the general high-dimensional phenomena of "measure concentration".

1.7. Example of such "strange" concentration in high dimensions: in high dimensions the volume of a ball is essentially sitting in the surface

$$\text{vol}(B_r(0)) = r^n \cdot \frac{\pi^{n/2}}{(\frac{n}{2}-1)!}$$

↑ ball of radius r in $\mathbb{R}^n$

$$\text{vol}(B_1) = \frac{\pi^{n/2}}{(\frac{n}{2}-1)!}$$

$$\mathrm{vol}\left(\{x \in \mathbb{R}^n \mid 1-\varepsilon \le \|x\| \le 1\}\right) =$$

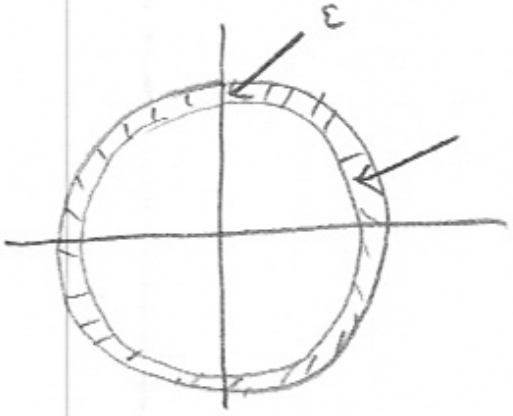
$$= \mathrm{vol}(B_1) - \mathrm{vol}(B_{1-\varepsilon})$$

$$= \frac{\pi^{n/2}}{\left(\frac{n}{2}-1\right)!}\left(1-(1-\varepsilon)^n\right)$$

$$\Rightarrow \frac{\mathrm{vol}\left(\{x \mid 1-\varepsilon \le \|x\| \le 1\}\right)}{\mathrm{vol}(B_1)} = 1-(1-\varepsilon)^n$$

$\to 1$ for $n \to \infty$

$\forall\, \varepsilon > 0$

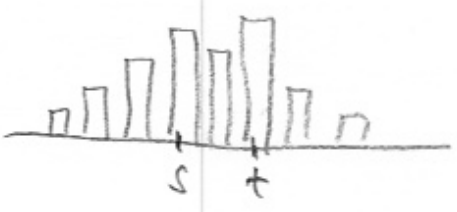


almost all volume is sitting here in high dimensions

1.8. From histograms to moments:

Let $A = A_N = (a_{ij})_{i,j=1}^{N}$ be our selfadjoint matrix with $a_{ij} = \pm 1$ randomly chosen

Then we see typically eigenvalues of A



$N \to \infty$ ⟶

$\mu_W = p_W \cdot \lambda$

This convergence means

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

this is difficult to calculate directly,
but note that it is the same as

$$(*)\quad \frac{1}{N}\sum_{i=1}^{N} 1_{[s,t]}(\lambda_i) \longrightarrow \int 1_{[s,t]}(x)\,d\mu$$

where

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

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

Hence in (*) we are claiming that

$$\frac{1}{N}\sum_{i=1}^{N} f(\lambda_i) \xrightarrow{N\to\infty} \int f(x)\,d\mu_w(x)$$

for all $f = 1_{[s,t]}$

It is easier to calculate this for other functions f, in particular for those of the form $f(x) = x^n$, i.e.

$$(**) \quad \frac{1}{N} \sum_{i=1}^{N} \lambda_i^n \xrightarrow{N \to \infty} \int x^n \, d\mu_w(x)$$

"moments of μ_w"

We will see later that in our case the validity of (*) for all $f = 1_{[s,t]}$ is equivalent to (**) for all n.

Hence we want to show (**) for all n!

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

note: $A = A^*$ selfadjoint

$\Rightarrow$ A is diagonizable, i.e.

$A = U D U^*$, where

- U is unitary
- D is diagonal with $D = \begin{pmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_N \end{pmatrix}$

But then we have

$$A^n = U D^n U^* \quad \text{with} \quad D^n = \begin{pmatrix} \lambda_1^n & & 0 \\ & \ddots & \\ 0 & & \lambda_N^n \end{pmatrix}$$ (0-1

hence

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

$$= \mathrm{Tr}(U D^n U^*)$$

$$= \mathrm{Tr}(A^n)$$

and thus

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

1.10 Notation: We denote by $\mathrm{tr} := \frac{1}{N}\mathrm{Tr}$ the normalized trace of our 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 we have typically for our matrices that

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

1.11 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 $a_{ij} \in \{\pm 1\}$, $A_N = A_N^*$, we have

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

$$\underbrace{\qquad\qquad}_{(\pm 1)^2 = 1}$$

$$= \frac{1}{N} \sum_{i,j=1}^{N} 1$$

$$= \frac{1}{N} \cdot N^2$$

$$= N$$

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

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

i.e. we consider the matrices

$$A_N = \frac{1}{\sqrt{N}} (a_{ij})_{i,j=1}^{N} \quad \text{where } a_{ij} = \pm 1$$

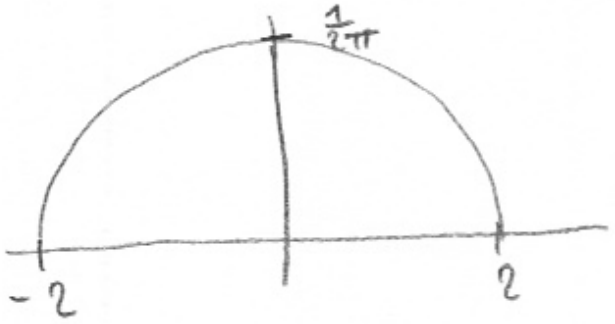
For this scaling we claim that we have typically that

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

for a deterministic probability measure μ_{sc}

1.12. Definition: 1) The (standard) semicircle distribution μ_w is the measure on $[-2,+2]$ with density

$$d\mu_w(x) = \frac{1}{2\pi}\sqrt{4-x^2}\,dx$$



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

$$c_k = \frac{1}{k+1}\binom{2k}{k}$$

$$1, 1, 2, 5, 14, 42, 132, \ldots$$

1.13. Theorem: 1) i) The Catalan numbers satisfy the 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 semicircle distribution μ_{sc} is a probability measure, i.e.

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

and its moments are given by

$$\frac{1}{2\pi}\int_{-2}^{+2} x^n\sqrt{4-x}\,dx = \begin{cases} 0 & n \text{ odd} \\ c_k & n=2k \text{ even} \end{cases}$$

1.14. Type of convergence: So we are claiming that typically

$$tr(A_N^2) \to 1 \qquad tr(A^8) \to 14$$

$$tr(A_N^4) \to 2 \qquad tr(A^{10}) \to 42$$

$$tr(A_N^6) \to 5 \qquad \text{etc}$$

But what do we mean with "typically". The mathematical word for this is "almost surely", but let us for now look on the more intuitive "convergence in probability" for

$$tr(A_N^{2k}) \to c_k$$

Denote by Ω_N the set of our considered matrices,

$$\Omega_N := \{ A_N = \frac{1}{\sqrt{N}} (a_{ij})_{i,j=1}^{N} \mid A_N = A_N^* \quad a_{ij} \in \{\pm 1\} \}$$

then convergence in probability means $\forall \varepsilon > 0$:

$$(*) \quad P(A_N \mid |tr(A_N^{2k}) - c_k| > \varepsilon) \xrightarrow{N \to \infty} 0$$

$$\| $$

$$\frac{\#\{A_N \in \Omega_N \mid |tr(A_N^{2k}) - c_k| > \varepsilon\}}{\#\Omega_N}$$

How can we show $(*)$?

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

$$E[tr(A_N^{2k})] \xrightarrow{N \to \infty} c_k$$

$$\| $$

$$\frac{\sum_{A_N \in \Omega_N} tr(A_N^{2k})}{\#\Omega_N}$$

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

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

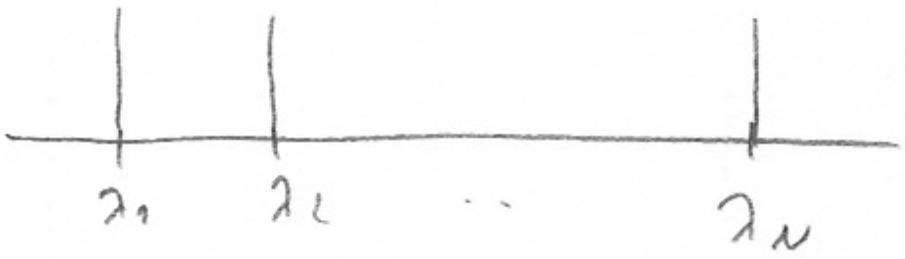
1.15. Remark: Note that

$$(*)\quad \frac{1}{N}\sum_{i=1}^{N} f(\lambda_i) \xrightarrow{N\to\infty} \int f(x)\, d\mu_w(x)$$

is actually a statement on convergence of measures, since

$$\frac{1}{N}\sum_{i=1}^{N} f(\lambda_i) = \int f(x)\, d\mu_N(x)$$

for $\mu_N = \frac{1}{N}(\delta_{\lambda_1} + \dots + \delta_{\lambda_N})$ eigenvalue distribution of A_N



δ_λ is Dirac measure $\delta_\lambda(E) = \begin{cases} 0 & \lambda \notin E \\ 1 & \lambda \in E \end{cases}$

Hence (*) says that

$$\int f(x)\, d\mu_N \xrightarrow{N\to\infty} \int f(x)\, d\mu$$

If we require this for sufficiently many f this is a kind of convergence

$$\mu_N \longrightarrow \mu$$

of measures.

We will need to understand such convergence better and develops tools (Cauchy- or Stieltjes transform) to deal with them