Random matrix theory: lecture 1

1) General overview

The focus of the theory is on the statistics of eigenvalues of random matrices.

It has applications in a variety of fields:

- multivariate statistics
- nuclear theoretical physics
- wireless communications
- mathematical finance
- number theory (!) and more...

What are the typical questions of the theory?

Let H be a $n \times n$ matrix whose entries are random variables with a given joint distribution.

Reminder: the eigenvalues of H are the roots of the characteristic polynomial

$$C_H(\lambda) = \det(H - \lambda I)$$

Fact: for any H, C_H has n complex roots $\lambda_1 \dots \lambda_n$

i.e. $C_H(\lambda) = c \prod_{j=1}^{n} (\lambda - \lambda_j)$ for some $c \in \mathbb{C}$

NB: since H is a random matrix, the eigenvalues $\lambda_1 .. \lambda_n$ are random variables

Question 1: what is the joint distribution $p(\lambda_1 .. \lambda_n)$ of the eigenvalues of H?

In general, it is very difficult to answer this question, but the answer is known for some specific ensembles of random matrices.

From now on (for this class), let us assume that H is real symmetric, so that the eigenvalues $\lambda_1 .. \lambda_n$ are real.

Question 2: what can be said about the marginal distributions:

main focus
$$\begin{cases} \cdot\ p(\lambda) = \int_{\mathbb{R}^{n-1}} p(\lambda, \lambda_2, .., \lambda_n)\, d\lambda_2 .. d\lambda_n \\ \cdot\ p(\lambda, \mu) = \int_{\mathbb{R}^{n-2}} p(\lambda, \mu, \lambda_3, .., \lambda_n)\, d\lambda_3 .. d\lambda_n \end{cases}$$

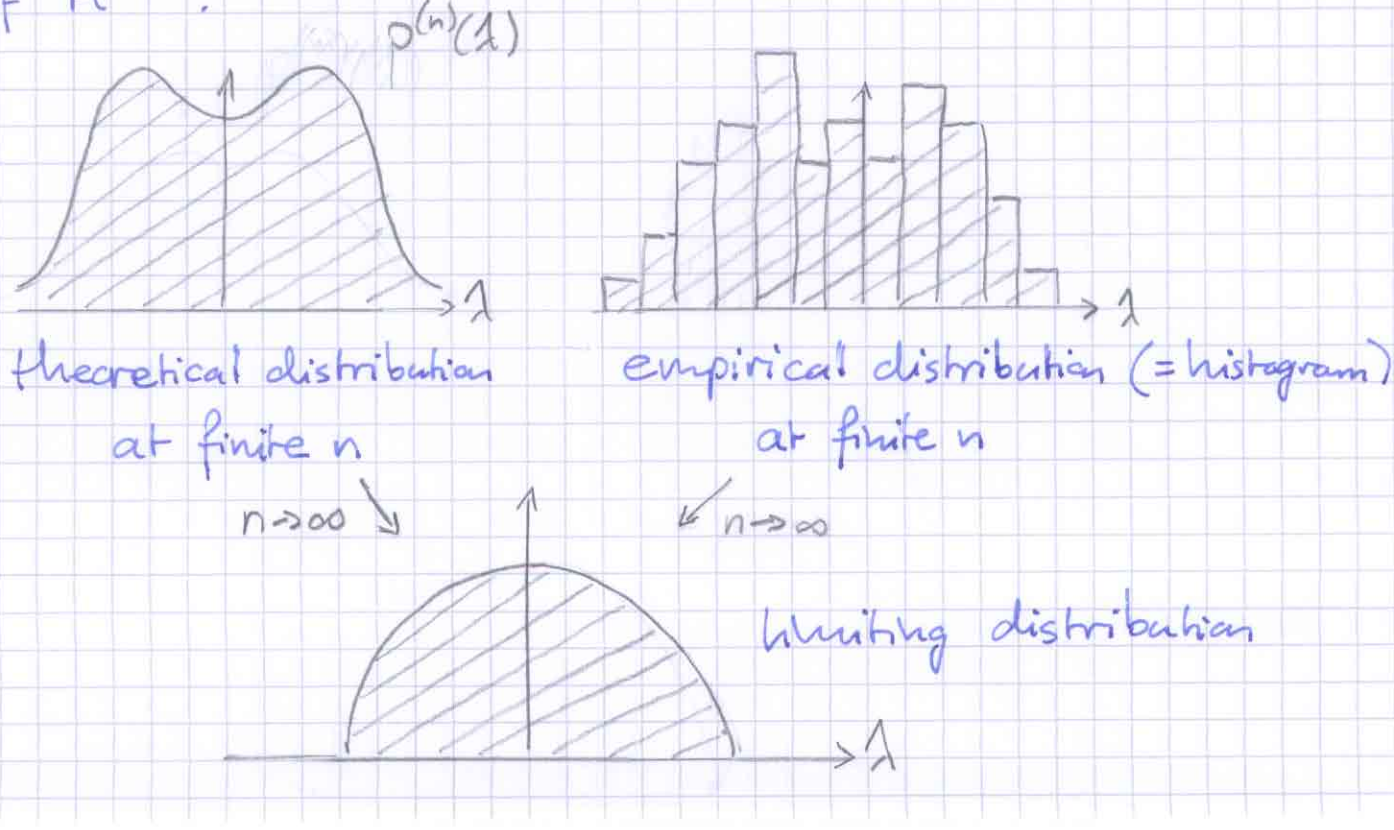
$\cdot\ p(\lambda, \mu, \nu) = \ldots$ etc. ?

Again, the answer is known for specific ensembles only.

Let now $(H^{(n)})_{n\geq 1}$ be a sequence of $n\times n$ random matrices (therefore of increasing size).
Let us assume that the corresponding marginal distributions $\rho^{(n)}(\lambda)$, $\rho^{(n)}(\lambda,\mu)$ are known for each n.

<u>Question 3</u>: what can be said about the limits $\lim_{n\to\infty}\rho^{(n)}(\lambda)$ & $\lim_{n\to\infty}\rho^{(n)}(\lambda,\mu)$?

Surprisingly, an answer to the first part of the question can be obtained for a much larger ensemble of random matrices, even if $\rho^{(n)}(\lambda)$ is not known for each n. This is done by looking at the empirical eigenvalue distribution of $H^{(n)}$:



2) History

1930-1950+ : multivariate statistics

John Wishart, 1928 : let $x^{(1)} \ldots x^{(m)}$ be m independent samples of a n-variate Gaussian random vector

× $x \sim N_{\mathbb{R}}(0, I)$. Wishart computes the joint distribution of the entries of the $n \times n$ sample covariance matrix W defined as

× $$W_{jk} = \frac{1}{m} \sum_{\ell=1}^{m} x_j^{(\ell)} x_k^{(\ell)} \qquad \left[W = \frac{1}{m} X X^T \right]$$

Fisher, Girshik, Hsu, Roy, 1939: compute the joint distribution of the eigenvalues of W.

James, 1964: extension to the complex case

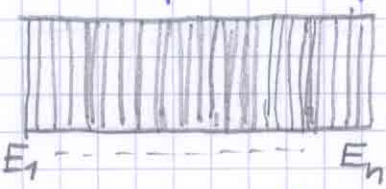
+ various extensions since then:

- joint eigenvalue distribution of $\frac{1}{m} X A X^*$, where A is a deterministic matrix
- distribution of the largest and smallest eigenvalues
- many more ...

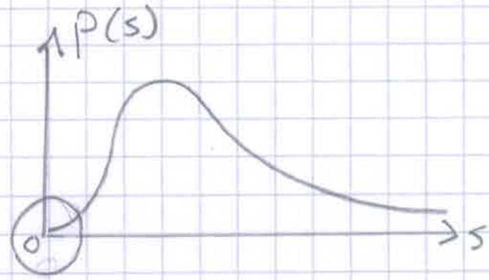
1950–1960+ : nuclear physics

Eugene Wigner, Freeman Dysan, Madan Lal Mehta

Spectra of energy levels in heavy nuclei ($n \sim 235$):



$\Rightarrow$ statistics of spacings :



repulsion of energy levels

In quantum physics, energy levels are eigenvalues of a real symmetric matrix H, called the Hamiltonian.

With 235 particles or more inside a nucleus, it is hopeless to describe H exactly.

Wigner's idea: let us assume that H is completely randam (but symmetric, still)!

Surprisingly, it works! (i.e. reproduces the above feature)

As a "by-product", Wigner obtains the limiting eigenvalue distribution of the studied matrices.

Many refinements and developments of Wigner's result have been obtained since and are still subject of active research.

1960+: random matrix theory (RMT)

The field expands rapidly and becomes a research subject on its own, mainly led by mathematical physicists at first, but also soon by pure mathematicians.

Some important dates:

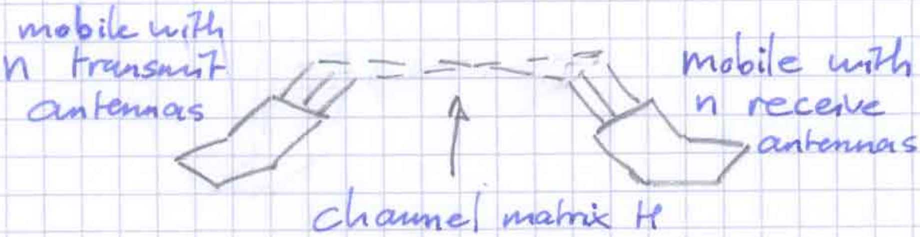
- 1967, Marčenko-Pastur:
  new technique for analyzing the limiting eigenvalue distribution of matrices of the form $A + XTX^*$, where X has i.i.d. entries
- 1991, Voiculescu: free probability and RMT
- 1994, Tracy-Widom: extreme eigenvalues
- 1995, Bogomolny & al: link with number theory

Finally, here are some names of active researchers in the field: Bai, Silverstein, Forrester, Baik, Ben Arous, Guionnet, Zeitouni, Soshnikov, Khorunzhy, Khoruzhenko, Goldsheid, Bohigas, Anderson, Its, Dembo, Diaconis and many others...

1990+: applications to wireless communications

One example: multiple antenna systems

(Foschini-Gans, Telatar, 1995)



H is again too complicated to be described exactly $\Longrightarrow$ modelled as completely random

The capacity of such a system (i.e. the maximum number of bits that can be exchanged per second between the two mobile phones) is shown to be:

power constraint

$$C = \mathbb{E}\left(\log\det\left(I + \frac{P}{n} HH^*\right)\right)$$

$$= \mathbb{E}\left(\sum_{j=1}^{n} \log(1 + P\lambda_j)\right) \qquad \lambda_j = \text{e.v. of } \tfrac{1}{n} HH^*$$

$$= \int_{\mathbb{R}_+^n} \sum_{j=1}^{n} \log(1 + P\lambda_j)\, p(\lambda_1, \dots, \lambda_n)\, d\lambda_1 \dots d\lambda_n$$

$$= n \int_{\mathbb{R}_+} \log(1 + P\lambda)\, p(\lambda)\, d\lambda$$

Many other applications of RMT occur in wireless communications ...

3) Course plan

- Reviews: matrix analysis
  probability
- Finite-size analysis (n fixed)
- Asymptotic analysis ($n \to \infty$)
- Free probability

All along the way: applications to communications:

- Multiple antenna systems
- CDMA systems
- Ad hoc networks
- ISI channels

Projects to be defined on related topics

Random matrix theory: lecture 2

Finite-size analysis (part I)

Problem: let H be a $n\times n$ random matrix with a given distribution; what can be said about the joint distribution of its eigenvalues $p(\lambda_1 .. \lambda_n)$?

Linear algebra reminder (H $n\times n$ complex matrix)

- H is said to be diagonalizable if there exist an invertible matrix S and a diagonal matrix Λ such that $H = S \Lambda S^{-1}$

× in this case, $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$ where $\lambda_1, \dots, \lambda_n$ are the eigenvalues of H

- not every matrix H is diagonalizable, but the following is true in general: there always exist an invertible matrix S and an upper triangular matrix T such that $H = S\,T\,S^{-1}$

moreover, T is block-diagonal, with blocks

× of the form $\begin{pmatrix} \lambda & 1 & 0 \\ & \ddots & 1 \\ 0 & & \lambda \end{pmatrix}$ [Jordan decomposition]

again, the elements of the diagonal of T are the eigenvalues of H

<u>Particular cases:</u> [complex conjugate transpose]

- if H is <u>normal</u>, i.e. $HH^* = H^*H$, then H is <u>unitarily diagonalizable</u>, i.e. there exist a <u>unitary</u> matrix U (i.e. $UU^* = I$) and a <u>diagonal</u> matrix Λ such that $H = U\Lambda U^*$

  <u>NB</u>: this is known as the <u>spectral theorem</u>

- there are three important sub-cases of the above:

  a) if H is <u>Hermitian</u>, i.e. $H = H^*$, then H is normal and $H = U\Lambda U^*$, where $\Lambda = \text{diag}(\lambda_1 .. \lambda_n)$ and the eigenvalues $\lambda_1 .. \lambda_n$ are <u>real</u>

  b) if H is <u>non-negative definite</u>, i.e. $x^* H x \geq 0$ for any vector $x \in \mathbb{C}^n$, then(*) H is normal and $H = U\Lambda U^*$, where $\Lambda = \text{diag}(\lambda_1 .. \lambda_n)$ and the eigenvalues $\lambda_1 .. \lambda_n$ are <u>non-negative</u>

  c) if H is <u>unitary</u>, i.e. $HH^* = I$, then H is normal and $H = U\Lambda U^*$, where $\Lambda = \text{diag}(\lambda_1 .. \lambda_n)$ and the eigenvalues $\lambda_1 .. \lambda_n$ are of <u>modulus 1</u> (i.e. $|\lambda_j| = 1 \quad \forall j$)

(*) H is Hermitian, so ...

For reasons that will become apparent in the class, it is (much) easier to deal with random matrices whose eigenvalues are distributed on a particular curve in the complex plane, and not in the whole plane. We will therefore focus on the last three subcases.

<u>Back to the joint eigenvalue distribution problem</u>

General strategy: given an ensemble of normal $n\times n$ random matrices H, we may interpret the spectral theorem $H = U\Lambda U^*$ as a <u>change of variables</u> $H \longmapsto (\Lambda, U)$.

Provided that H is distributed according to $p(H)$, we therefore have $p(H)\,dH = p(U\Lambda U^*)\cdot|J(\Lambda,U)|\,d\Lambda\,dU$, where $J(\Lambda,U)$ is the Jacobian of the change of variables.

The joint distribution of (Λ, U) is given by

$$\tilde{p}(\underset{\text{eigenvalues}}{\Lambda},\underset{\text{eigenvectors}}{U}) = p(U\Lambda U^*)\cdot|J(\Lambda,U)|$$

And as we will see, this expression simplifies drastically in some particular cases...

Warm-up (case $n=1$!)

Let x, y be iid r.v. $\sim N_{\mathbb{R}}(0,\frac{1}{2})$, i.e. their joint density is given by $p(x,y) = \frac{1}{\sqrt{\pi}}\exp(-x^2)\cdot\frac{1}{\sqrt{\pi}}\exp(-y^2) = \frac{1}{\pi}e^{-x^2-y^2}$

NB: the complex r.v. $z = x+iy$ has therefore a density $p(z) = \frac{1}{\pi}e^{-|z|^2}$; notation: $z \sim N_{\mathbb{C}}(0,1)$ (*)

Let us consider the change of variable $x+iy = re^{i\theta}$:

The Jacobian is given by [i.e. $x = r\cos\theta$, $y = r\sin\theta$]

$$J(r,\theta) = \det\begin{pmatrix} \frac{\partial x}{\partial r} & \frac{\partial x}{\partial \theta} \\ \frac{\partial y}{\partial r} & \frac{\partial y}{\partial \theta}\end{pmatrix}$$

$$= \det\begin{pmatrix} \cos\theta & \sin\theta \\ -r\sin\theta & r\cos\theta\end{pmatrix} = r$$

Therefore, $\tilde{p}(r,\theta) = p(x(r,\theta), y(r,\theta))\cdot r = \frac{1}{\pi}e^{-r^2}r$

$$= \underbrace{2re^{-r^2}}_{\text{(Rayleigh dist.) } \tilde{p}(r)} \cdot \underbrace{\frac{1}{2\pi}}_{\tilde{p}(\theta)}$$

Remark:

$\tilde{p}(r,\theta)$ does actually not depend on θ; this implies:

a) the distribution is uniform in θ

b) r and θ are independent (since factorization)

c) for any given φ (deterministic), z and $ze^{i\varphi}$ have the same distribution

(*) in addition, since $p(z)$ depends only on $|z|$, the r.v. z is said to be "circularly symmetric"

× name

Gaussian Orthogonal Ensemble (GOE)

Let H be a $n\times n$ real symmetric random matrix such that: • $\{h_{jk},\ j\le k\}$ are independent r.v. (& $h_{kj}=h_{jk}$)

• $h_{jj}\sim N_{\mathbb{R}}(0,1)$, $h_{jk}\sim N_{\mathbb{R}}(0,\frac{1}{2})$ $\forall j<k$

• Distribution of H:

$$p(H)=\prod_{j=1}^{n}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{h_{jj}^2}{2}\right)\cdot\prod_{j<k}\frac{1}{\sqrt{\pi}}\exp\left(-h_{jk}^2\right)$$

×
$$=C_n\exp\left(-\frac{1}{2}\sum_{j=1}^{n}h_{jj}^2-\sum_{j<k}h_{jk}^2\right)$$

$$=C_n\exp\left(-\frac{1}{2}\sum_{j=1}^{n}h_{jj}^2-\frac{1}{2}\sum_{j\neq k}h_{jk}^2\right)$$

$$=C_n\exp\left(-\frac{1}{2}\sum_{j,k}h_{jk}^2\right)=C_n\exp\left(-\frac{1}{2}\operatorname{Tr}(HH^T)\right)$$

real case
$$=C_n\exp\left(-\frac{1}{2}\operatorname{Tr}(H^2)\right)\quad\text{since } H=H^T$$

By the spectral theorem, there exists a $n\times n$ orthogonal matrix V (i.e. $VV^T=I$) and $\Lambda=\operatorname{diag}(\lambda_1\ldots\lambda_n)$, with $\lambda_1\ldots\lambda_n$ real, such that $H=V\Lambda V^T$

i.e. $h_{jk}=\sum_{\ell=1}^{n}\lambda_\ell v_{j\ell}v_{k\ell}$ $\quad j,k=1..n$

→ change of variables

• Sanity check: how many free (real) parameters do we have on each side?

• on the left (H): n diag. parameters $+\ \frac{n(n-1)}{2}$ upper diag. parameters

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

- on the right: n diag. parameters in Λ

  (Λ, V) $\quad + \frac{n(n-1)}{2}$ parameters in V (see construction below)

  $= \frac{n^2+n}{2}$ parameters ✓

Aside: how many free parameters are there in an orthogonal matrix V ?

Reminder: $VV^T = I$ means the rows of V are orthonormal vectors $v_1, \dots, v_n$ in $\mathbb{R}^n$

so:
- for the first row v_1, there are $n-1$ free parameters (since $\|v_1\| = 1$)
- for the second row v_2, there are $n-2$ " (since $\|v_2\| = 1$ & $v_2 v_1^T = 0$)
- etc.
- in total, this leads to $(n-1) + (n-2) + \dots + 1 + 0 = \frac{n(n-1)}{2}$ param. ✓

- Jacobian:

The Jacobian of the change of variables $H \mapsto (\Lambda, V)$ is given by:

$$J(\{\lambda_\ell\}, \{v_{\ell m}\}) = \det\left(\left\{ \frac{\partial h_{jk}}{\partial \lambda_\ell} \right\} \middle| \left\{ \frac{\partial h_{jk}}{\partial v_{\ell m}} \right\} \right)$$

$\frac{n(n+1)}{2} \times n$ matrix $\qquad \frac{n(n+1)}{2} \times \frac{n(n-1)}{2}$ matrix

Result of the computation:

$$J(\{\lambda_\ell\}, \{v_{\ell m}\}) = \prod_{j<k} (\lambda_j - \lambda_k)$$

[Homework 1 → explicit simple case $(n=2)$]

Heuristics for the above computation :

- $h_{jk} = \sum_{\ell=1}^{n} \lambda_\ell v_{j\ell} v_{k\ell}$

$$\Longrightarrow \begin{cases} \dfrac{\partial h_{jk}}{\partial \lambda_\ell} = v_{j\ell} v_{k\ell} = \text{cst in } \lambda \\ \dfrac{\partial h_{jk}}{\partial v_{\ell m}} = (\delta_{j\ell} v_{km} + \delta_{k\ell} v_{jm}) \lambda_\ell = \text{linear fn in } \lambda \end{cases}$$

$$\Longrightarrow \begin{cases} J = \det(\;\cdot\;) = \text{polynomial in } \lambda \text{ of max. degree } \frac{n(n-1)}{2} \\ \text{if } \lambda_\ell = \lambda_p, \text{ then } \dfrac{\partial h_{jk}}{\partial v_{\ell m}} = \dfrac{\partial h_{jk}}{\partial v_{pm}} \text{ i.e. } J = 0 \end{cases}$$

so the only polynomial satisfying these two conditions is of the form $\prod_{j<k} (\lambda_j - \lambda_k)$

NB: a remarkable fact is that J does not depend on V (similar to the polar coordinates example)!

[to be proven]

- Conclusion for the GOE:

$$\tilde{p}(\Lambda, V) = p(V\Lambda V^T) \cdot |J(\Lambda, V)|$$

$$= C_n \exp\Big(-\tfrac{1}{2} \underbrace{\mathrm{Tr}((V\Lambda V^T)^2)}\Big) \cdot \prod_{j<k} |\lambda_j - \lambda_k|$$

$$= \mathrm{Tr}(V\Lambda V^T V \Lambda V^T)$$

$$= \mathrm{Tr}(V\Lambda^2 V^T) = \mathrm{Tr}(\Lambda^2 V^T V)$$

$$= \mathrm{Tr}(\Lambda^2)$$

$$= C_n \exp\Big(-\tfrac{1}{2} \sum_{j=1}^{n} \lambda_j^2\Big) \cdot \prod_{j<k} |\lambda_j - \lambda_k|$$

<u>Same remark as before:</u>

$\tilde{p}(\Lambda, V)$ does not depend on V at all

$\Rightarrow$ a) the distribution of V is <u>uniform</u> over the set of orthogonal matrices ("Haar distribution")

b) Λ and V are actually <u>independent</u>, i.e. the eigenvalues of H are independent from its eigenvectors!

c) for any given deterministic orthogonal matrix W, one obtains that H and $W^T H W$ have the same distribution, i.e. the distribution of H is <u>invariant under orthogonal transformations</u>, therefore the name of the ensemble.

<u>NB</u>: the above computation was made possible by the fact that the distribution of H only depends on $Tr(H^2) = Tr(\Lambda^2)$.

Random matrix theory: lecture 3

Real Wishart Ensemble (non-std terminology)

Let H be a $n \times m$ real random matrix such that $\{h_{jk}, 1 \le j \le n, 1 \le k \le m\}$ are i.i.d. r.v. $\sim N_{\mathbb{R}}(0,1)$.
Let $W = H H^T$ ($n \times n$ matrix) and $\lambda_1 .. \lambda_n$ be the eigenvalues of W. (not those of H (!))

Remarks: • since W is symmetric $\underset{(W^T = HH^T = W)}{}$ & non-negative definite, $\underset{(u^T W u = \|H^T u\|^2 \ge 0)}{}$

$\lambda_j \ge 0 \quad \forall 1 \le j \le n$

• if $m < n$, then $n-m$ eigenvalues are zero, since $\text{rank}(W) \le \text{rank}(H) \le \min(n,m) = m$

In the case where $m<n$, there is therefore no joint eigenvalue density $p(\lambda_1,..,\lambda_n)$, but we may as well consider $\tilde{W} = H^T H$ and $\tilde{\lambda}_1 .. \tilde{\lambda}_m$ the eigenvalues of $\tilde{W}$, since the non-zero eigenvalues of W and $\tilde{W}$ coincide.

Proof, together with a linear algebra reminder:

Let A be a $n \times n$ matrix, B a $n \times m$ matrix

C a $m \times n$ matrix, D a $m \times m$ matrix

If A is invertible, then

$$\det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(A) \cdot \det(D - C A^{-1} B)$$

If D is invertible, then

$$\det\begin{pmatrix} A & B \\ C & D \end{pmatrix} = \det(D) \cdot \det(A - B D^{-1} C)$$

In particular, for any B $n \times m$ & C $m \times n$, we have

$$\det(I_n - BC) = \det\begin{pmatrix} I_n & B \\ C & I_m \end{pmatrix} = \det(I_m - CB)$$

Likewise, for any $\lambda \neq 0$, we have

$$\det(\lambda I_n - BC) = \det(\lambda I_m - CB)$$

[⚠ this fails for $\lambda = 0$]

i.e. BC and CB share the same non-zero eigenvalues.

In particular, so do $W = HH^T$ and $\widetilde{W} = H^T H$. #

Without loss of generality, we may therefore assume that the $n \times m$ matrices H are "fat" ($\underline{m \geq n}$).

The case of "tall" matrices H ($m < n$) is handled by replacing H by H^T and "adding" $n - m$ zero eigenvalues for the matrix W.

Preliminary question:

What is the joint distribution of the entries of W?

(= question asked by Wishart in 1928)

a) joint distribution of the entries of H:

$$p(H) = \prod_{j,k=1}^{n,m} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{h_{jk}^2}{2}\right)$$
$$= C_{n,m} \exp\left(-\frac{1}{2}\sum_{j,k=1}^{n,m} h_{jk}^2\right)$$
$$= C_{n,m} \exp\left(-\frac{1}{2}\operatorname{Tr}(HH^T)\right)$$

b) LQ-decomposition of H: $(m \geq n)$

Any $n \times m$ matrix H can be decomposed as

$H = LQ$

where L is a $n \times n$ lower-triangular matrix

and Q is a $n \times m$ matrix such that $QQ^T = I_n$

[generalization of orthogonal matrix]

c) Choleski decomposition of W: $(m \geq n)$

Consequently, $W = HH^T = LQQ^TL^T = LL^T$

General strategy: $p(H) \to p(L) \to p(W)$

(NB: $\neq$ method adopted by Wishart)

d) joint distribution of the <u>entries of L</u> : $(H = LQ)$

The Jacobian of the change of var. $H \mapsto (L, Q)$ is given by

$$J_1(L, Q) = \prod_{j=1}^{n} \ell_{jj}^{m-j} \quad (\geq 0) \qquad \text{(nm variables)}$$

Therefore,

$$p(L, Q) = C_{n,m}\, \exp\left(-\tfrac{1}{2}\, \mathrm{Tr}(LL^T)\right) \cdot \prod_{j=1}^{n} \ell_{jj}^{m-j} = p(L)$$

(this expression again does not depend on Q)

<u>NB</u>: since $\mathrm{Tr}(LL^T) = \sum_{j=1}^{n} \ell_{jj}^2 + \sum_{k<j} \ell_{jk}^2$, the above means that:

- the entries of L are independent
- the off-diagonal entries ℓ_{jk} $(k<j)$ are Gaussian
- the diagonal entries ℓ_{jj} are χ^2_{m+1-j}

e) joint distribution of the <u>entries of W</u>: $(W = LL^T)$

The Jacobian of the change of var. $W \mapsto L$ is given by

$$J_2(L) = 2^n \prod_{j=1}^{n} \ell_{jj}^{n+1-j} \quad (\geq 0) \qquad \left(\tfrac{n(n+1)}{2} \text{ variables}\right)$$

Therefore,

$$p(W) = \frac{p(L)}{J_2(L)} = \tilde{C}_{n,m} \cdot \exp\left(-\tfrac{1}{2}\, \mathrm{Tr}(LL^T)\right) \prod_{j=1}^{n} \ell_{jj}^{m-n-1}$$

note that $\prod_{j=1}^{n} \ell_{jj} = \det L = \sqrt{\det W}$ and $LL^T = W$, so finally

$$p(W) = \tilde{C}_{n,m} \cdot \exp\left(-\tfrac{1}{2}\, \mathrm{Tr}\, W\right) \cdot (\det W)^{\frac{m-n-1}{2}}$$

(≥ 0)

Final step: joint eigenvalue distribution $p(\lambda_1, \dots, \lambda_n)$

There exist a $n \times n$ orthogonal matrix V and $\Lambda = \mathrm{diag}(\lambda_1 .. \lambda_n)$ such that $W = V \Lambda V^T$. The Jacobian of the change of var. $W \mapsto (\Lambda, V)$ is again given by

×
$$J_3(\Lambda, V) = \prod_{j<k} (\lambda_j - \lambda_k)$$

So

$$p(\Lambda, V) = p(V \Lambda V^T) \cdot |J(\Lambda, V)|$$

$$= \tilde{C}_{n,m} \exp(-\tfrac{1}{2} \mathrm{Tr}\, \Lambda) \cdot (\det \Lambda)^{\frac{m-n-1}{2}} \cdot |J(\Lambda, V)|$$

$$= \tilde{C}_{n,m} \exp\Big(-\tfrac{1}{2} \sum_{j=1}^{n} \lambda_j\Big) \cdot \prod_{j=1}^{n} \lambda_j^{\frac{m-n-1}{2}} \cdot \prod_{j<k} |\lambda_k - \lambda_j|$$

$$= p(\lambda_1 .. \lambda_n)$$

(since again the above expression does not depend on V; we therefore have the same conclusions as in the GOE case)

References: (finite-size analysis)

- Madan Lal Mehta, "Random matrices"
- Alan Edelman's PhD thesis (available on the web)
- R. J. Muirhead, "Aspects of multivariate statistical theory"
- A. Gupta - D. Nagar, "Matrix variate distributions"
- V. Girko, "Theory of random determinants"

Corresponding complex ensembles

Gaussian Unitary Ensemble (GUE)

Let H be a $n \times n$ Hermitian random matrix such that

- $\{h_{jk}, j \leq k\}$ are independent random variables ($\& \; h_{kj} = \overline{h_{jk}}$)
- $h_{jj} \sim N_{\mathbb{R}}(0,1)$, $h_{jk} \sim N_{\mathbb{C}}(0,1)$ $j<k$ ("circularly symmetric")

  [i.e. $\mathrm{Re}\, h_{jk}, \mathrm{Im}\, h_{jk} \sim N_{\mathbb{R}}(0,\frac{1}{2})$ indep.]

joint distribution of entries:

× $$p(H) = C_n \exp\Big(-\frac{1}{2}\sum_{j,k=1}^{n} |h_{jk}|^2\Big) = C_n \exp\Big(-\frac{1}{2}\,\mathrm{Tr}(HH^*)\Big)$$

Jacobian of the transformation $H = U \Lambda U^*$:

× $$J(\Lambda, U) = \prod_{j<k} (\lambda_k - \lambda_j)^2 \qquad (\underline{\text{NB}}: \lambda_j \in \mathbb{R})$$

resulting joint eigenvalue distribution:

$$p(\lambda_1 .. \lambda_n) = C_n \exp\Big(-\frac{1}{2}\sum_{j=1}^{n} \lambda_j^2\Big) \cdot \prod_{j<k} (\lambda_k - \lambda_j)^2$$

Complex Wishart Ensemble

Let H be a $n \times m$ random matrix ($m \geq n$) such that

$\{h_{jk}, j=1..n, k=1..m\}$ are i.i.d. r.v. $\sim N_{\mathbb{C}}(0,1)$

Let $W = HH^*$ and $\lambda_1 .. \lambda_n$ be its eigenvalues (≥ 0)

Then $$p(\lambda_1 .. \lambda_n) = C_{n,m} \exp\Big(-\sum_{j=1}^{n} \lambda_j\Big) \Big(\prod_{j=1}^{n} \lambda_j^{m-n}\Big) \prod_{j<k} (\lambda_k - \lambda_j)^2$$

$$(\& \; p(W) = C_{n,m} \exp(-\mathrm{Tr}\, W)\, (\det W)^{m-n})$$

Random matrix theory: lecture 4

Computation of marginals

Square Complex Wishart Ensemble

Let H be a $n\times n$ matrix with iid $\sim N_{\mathbb{C}}(0,1)$ entries

and $\lambda_1 .. \lambda_n$ be the (non-negative) eigenvalues of $W = HH^*$.

Their joint distribution is given by

$$p(\lambda_1 .. \lambda_n) = C_n \exp\left(-\sum_{j=1}^{n} \lambda_j\right) \prod_{j<k} (\lambda_k - \lambda_j)^2$$

NB:

In order for the transformation $H = U \Lambda U^*$ to be unique, we need to order the eigenvalues, e.g. as $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n (\geq 0)$. The above distribution p is therefore defined on this set.

× Note however that p is symmetric under any permutation of $(\lambda_1 .. \lambda_n)$. The distribution of the unordered eigenvalues of W is therefore given by the same above expression (only the constant changes: it is divided by $n!$, but we do not care).

× We will focus on this second object in the following.

× We are interested in computing the first order and second order marginal distributions:

$$\begin{cases} p(\lambda) = \int_{\mathbb{R}_+^{n-1}} p(\lambda, \lambda_2, \dots, \lambda_n)\, d\lambda_2 \dots d\lambda_n \\ p(\lambda, \mu) = \int_{\mathbb{R}_+^{n-2}} p(\lambda, \mu, \lambda_3, \dots, \lambda_n)\, d\lambda_2 \dots d\lambda_n \end{cases}$$

<u>Step 1</u>: Van der Monde determinant and Laguerre polynomials

In order to marginalize $p(\lambda_1 \dots \lambda_n)$, we need to take care of the term $\prod_{j<k} (\lambda_k - \lambda_j)^2$.

× <u>Van der Monde identity</u>:

$$\prod_{j<k} (\lambda_k - \lambda_j) = \det \begin{pmatrix} 1 & \dots & 1 \\ \lambda_1 & \dots & \lambda_n \\ & & \\ \lambda_1^{n-1} & \dots & \lambda_n^{n-1} \end{pmatrix}$$

<u>Laguerre polynomials</u>:

Let us define for $k \geq 0$:

$$L_k(\lambda) = \frac{1}{k!} \cdot e^{\lambda} \cdot \frac{d^k}{d\lambda^k}\left(e^{-\lambda} \lambda^k\right) \quad \lambda \geq 0$$

i.e. $L_0(\lambda) \equiv 1$, $L_1(\lambda) = 1 - \lambda$, $L_2(\lambda) = \frac{1}{2}(\lambda - 2)^2 - 1$ ...

<u>Property A</u>:

L_k is a polynomial of degree k in λ:

$L_k(\lambda) = \gamma_k \lambda^k + \dots$ with $\gamma_k \neq 0$ (actually, $\gamma_k = \frac{(-1)^k}{k!}$)

Therefore,

$$\det\begin{pmatrix}1 & \cdots & 1\\ \vdots & & \vdots\\ \lambda_1^{n-1} & \cdots & \lambda_n^{n-1}\end{pmatrix} = \underbrace{\left(\prod_{j=0}^{n-1}\frac{1}{\gamma_j}\right)}_{\substack{\text{some constant}\\ c_n\neq 0}} \cdot \det\underbrace{\begin{pmatrix}L_0(\lambda_1) & \cdots & L_0(\lambda_n)\\ \vdots & & \vdots\\ L_{n-1}(\lambda_1) & \cdots & L_{n-1}(\lambda_n)\end{pmatrix}}_{:=\{L_{j-1}(\lambda_k)\}}$$

i.e.

$$\prod_{j<k}(\lambda_k-\lambda_j)^2 = c_n^2 \det\left(\{L_{j-1}(\lambda_k)\}\right)^2$$

$$= c_n^2 \det\left(\{L_{j-1}(\lambda_k)\}\cdot\{L_{j-1}(\lambda_k)\}^T\right)$$

$$= c_n^2 \det\left(\{K(\lambda_j,\lambda_k)\}\right)$$

where $K(\lambda,\mu) := \sum_{l=0}^{n-1} L_l(\lambda)L_l(\mu)$

Remark: for the Real Wishart Ensemble, the square is missing in the Jacobian, so the above trick does not work and the analysis is (much) more complicated.

—

So far, our reasoning works if we replace $\{L_k(\lambda)\}$ by any sequence of polynomials. Why then choose the specific Laguerre polynomials? Because of

Property B ("orthogonal polynomials")

$$\int_{\mathbb{R}_+} L_k(\lambda)L_l(\lambda)e^{-\lambda}\,d\lambda = \delta_{kl} = \begin{cases}1 & \text{if } k=l\\ 0 & \text{otherwise}\end{cases}$$

21.3.07

4

Step 2: properties of K and Mehta's lemma

Because of property B, we are able to show the following.

Proposition: (o) $K(\mu,\lambda) = K(\lambda,\mu)$

(i) $\int_{\mathbb{R}_+} K(\lambda,\lambda)\, e^{-\lambda}\, d\lambda = n$

(ii) $\int_{\mathbb{R}_+} K(\lambda,\mu)\, K(\mu,\nu)\, e^{-\mu}\, d\mu = K(\lambda,\nu)$

[i.e. $K(\lambda,\mu)$ is a "self-reproducing kernel"]

Proof: (o) clear

(i) $\sum_{\ell=0}^{n-1} \underbrace{\int_{\mathbb{R}_+} L_\ell(\lambda)\, e^{-\lambda}\, d\lambda}_{=1} = n$ ✓

(ii) $\sum_{\ell,m=0}^{n-1} L_\ell(\lambda)\, L_m(\nu) \underbrace{\int_{\mathbb{R}_+} L_\ell(\mu)\, L_m(\mu)\, e^{-\mu}\, d\lambda}_{=\delta_{\ell m}} = K(\lambda,\nu)$ ✓

#

Recall that

$$p(\lambda_1 .. \lambda_n) = C_n \exp(-\sum_{j=1}^{n} \lambda_j) \cdot \det(\{K(\lambda_j,\lambda_k)\})$$

We have the following lemma from Mehta.

Lemma

a) $p(\lambda) = \int_{\mathbb{R}_+^{n-1}} p(\lambda,\lambda_2 .. \lambda_n)\, d\lambda_2 .. d\lambda_n = C_{n,1}\, K(\lambda,\lambda)\, e^{-\lambda}$

b) $p(\lambda,\mu) = \int_{\mathbb{R}_+^{n-2}} p(\lambda,\mu,\lambda_3 .. \lambda_n)\, d\lambda_3 .. d\lambda_n$

$$= C_{n,2} \left(K(\lambda,\lambda) K(\mu,\mu) - K(\lambda,\mu)^2 \right) e^{-(\lambda+\mu)}$$

Sketch of proof: ($n=2$!)

b) $p(\lambda_1,\lambda_2) = C_2 e^{-(\lambda_1+\lambda_2)} \det \begin{pmatrix} k(\lambda_1,\lambda_1) & k(\lambda_1,\lambda_2) \\ k(\lambda_2,\lambda_1) & k(\lambda_2,\lambda_2) \end{pmatrix}$ ✓

a)
$$p(\lambda_1) = \int_{\mathbb{R}_+} p(\lambda_1,\lambda_2)\, d\lambda_2$$
$$= C_2 e^{-\lambda_1} \Big(k(\lambda_1,\lambda_1) \overbrace{\int_{\mathbb{R}_+} k(\lambda_2,\lambda_2) e^{-\lambda_2} d\lambda_2}^{=2 \text{ by (i)}} - \underbrace{\int_{\mathbb{R}_+} k(\lambda_1,\lambda_2) k(\lambda_2,\lambda_1) e^{-\lambda_2} d\lambda_2}_{=k(\lambda_1,\lambda_1) \text{ by (ii)}} \Big)$$
$$= C_2 e^{-\lambda_1} k(\lambda_1,\lambda_1) \checkmark$$

#

Last step: determination of the constants $C_{n,1}$ & $C_{n,2}$

- $\int_{\mathbb{R}_+} p(\lambda)\, d\lambda = 1 \Rightarrow C_{n,1} \underbrace{\int_{\mathbb{R}_+} k(\lambda,\lambda) e^{-\lambda} d\lambda}_{=n \text{ by (i)}} = 1$

  i.e. $C_{n,1} = \frac{1}{n}$

- $\int_{\mathbb{R}_+} p(\lambda,\mu)\, d\mu = p(\lambda)$

  $\Rightarrow C_{n,2} \big(\underbrace{k(\lambda,\lambda) \cdot n}_{\text{by (i)}} - \underbrace{k(\lambda,\lambda)}_{\text{by (ii)}} \big) e^{-\lambda} = C_{n,1} k(\lambda,\lambda) e^{-\lambda}$

  i.e. $C_{n,2} = \frac{C_{n,1}}{n-1} = \frac{1}{n(n-1)}$

Finally:
$$\begin{cases} p(\lambda) = \frac{1}{n} k(\lambda,\lambda) e^{-\lambda} = \frac{1}{n} \sum_{\ell=0}^{n-1} L_\ell(\lambda)^2 e^{-\lambda} \\ p(\lambda,\mu) = \frac{1}{n(n-1)} \left(k(\lambda,\lambda) k(\mu,\mu) - k(\lambda,\mu)^2 \right) e^{-(\lambda+\mu)} \\ \text{with } k(\lambda,\mu) = \sum_{\ell=0}^{n-1} L_\ell(\lambda) L_\ell(\mu) \end{cases}$$

Marginals for the GUE

joint eigenvalue distribution: ($\lambda_j \in \mathbb{R}$)

$$p(\lambda_1 .. \lambda_n) = C_n \exp\left(-\tfrac{1}{2}\sum_{j=1}^{n}\lambda_j^2\right) \cdot \prod_{j<k}(\lambda_k - \lambda_j)^2$$

$\Rightarrow$ the analysis is totally similar; the only difference comes from the fact that one has to consider Hermite polynomials (instead of Laguerre pol.)

$$H_k(\lambda) = c_k e^{\lambda^2/2} \frac{d^k}{d\lambda^k}\left(e^{-\lambda^2/2}\right) \qquad \lambda \in \mathbb{R}$$

that satisfy:

$$\int_{\mathbb{R}} H_k(\lambda) H_\ell(\lambda) e^{-\lambda^2/2} d\lambda = \delta_{k\ell}$$

We obtain that

$$\begin{cases} p(\lambda) = \frac{1}{n}\sum_{\ell=0}^{n-1} H_\ell(\lambda)^2 e^{-\lambda^2/2} \\ p(\lambda,\mu) = \frac{1}{n(n-1)}\left(K(\lambda,\lambda)K(\mu,\mu) - K(\lambda,\mu)^2\right) e^{-\frac{\lambda^2+\mu^2}{2}} \\ \text{with } K(\lambda,\mu) = \sum_{\ell=0}^{n-1} H_\ell(\lambda) H_\ell(\mu) \end{cases}$$

<u>NB</u>: The method therefore generalizes to

$$p(\lambda_1 .. \lambda_n) = C_n \exp\left(-\sum_{j=1}^{n} V(\lambda_j)\right) \cdot \prod_{j<k}(\lambda_k - \lambda_j)^2$$

as soon as there exist polynomials satisfying

$$\int_{\mathbb{R}} P_k(\lambda) P_\ell(\lambda) \underbrace{e^{-V(\lambda)}}_{\text{weight}} d\lambda = \delta_{k\ell}$$

21.3.07

A

RMT: linear algebra reminder

× **Eigenvalues and singular values**

- Fact 1: Let H be a $n \times n$ complex Hermitian matrix.

There exist U $n \times n$ unitary and $\Lambda = \mathrm{diag}(\lambda_1 .. \lambda_n)$

such that $H = U \Lambda U^*$. This implies: $HU = U\Lambda$

× i.e. λ_k = eigenvalue of H with corr. eigenvector $u^{(k)} = (u_{jk})$
(column of U)

- Fact 2: Let H be a $n \times m$ complex matrix.

There exist U $n \times n$ unitary, V $m \times m$ unitary and

× Σ = pseudodiag $(\sigma_1 .. \sigma_r)$ with $r = \min(n,m)$,
($n \times m$ matrix with off-diagonal entries $= 0$)

such that $H = U \Sigma V^*$.

$\sigma_1 .. \sigma_r$ are called the singular values of H.

This implies:

× a) $HH^*U = U\Sigma\Sigma^*$, i.e. the columns of U are the eigenvectors of HH^*, with non-zero eigenvalues $\sigma_1^2 .. \sigma_r^2$.

× b) $H^*HV = V\Sigma^*\Sigma$, i.e. the columns of V are the eigenvectors of H^*H, again with non-zero eigenvalues $\sigma_1^2 .. \sigma_r^2$.

The squares of the singular values of H are therefore the non-zero eigenvalues of HH^* (or H^*H).

Relation between the eigenvalues and the singular values:

(valid only for square matrices H !)

We know that any $n \times n$ complex matrix H also has n eigenvalues $\lambda_1 \ldots \lambda_n$. What is the relation between these and $\sigma_1 \ldots \sigma_n$, the singular values of H ?

- In general, there is no specific relation, except some inequalities, such as $\max_k |\lambda_k| \leq \max_k \sigma_k$. (*)
- If H is Hermitian, then $HH^* = H^*H = H^2$, so $U = V$ and $\sigma_j^2 = \lambda_j^2$ (provided that they are ordered accordingly), so $\sigma_j = |\lambda_j|$ [recall that $\sigma_j \geq 0$, and $\lambda_j \in \mathbb{R}$ when $H = H^*$].
- Ex: $H = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \longrightarrow \lambda_1 = \lambda_2 = 1$

$$HH^* = \begin{pmatrix} 2 & 1 \\ 1 & 1 \end{pmatrix} \longrightarrow \sigma_1^2 = \frac{3+\sqrt{5}}{2}, \ \sigma_2 = \frac{3-\sqrt{5}}{2}$$

(*) and of course: $\left| \prod_{k=1}^{n} \lambda_k \right| = |\det H| = \sqrt{\det(HH^*)} = \prod_{k=1}^{n} \sigma_k$

Random matrix theory : lecture 5

Preliminary : Haar "uniform" distribution

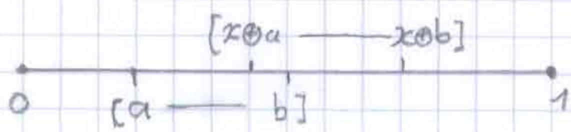
Theorem

Let G be a compact group (not necessarily commutative).
There exists a unique Borel probability distribution μ on G
such that $\mu(B) = \mu(g\cdot B) = \mu(B\cdot g) \quad \forall g \in G,\ B \in \mathcal{B}(G)$,
where
$$\begin{cases} g\cdot B := \{ h \in G : \exists g_0 \in B \text{ s.t. } h = g\, g_0 \} \\ B\cdot g := \{ h \in G : \exists g_0 \in B \text{ s.t. } h = g_0\, g \} \end{cases}$$
μ is called the Haar distribution on G.

Examples

0) Let $G = S(n)$ the group of permutations σ on $\{1..n\}$ equipped with the usual composition law $\sigma \circ \tau$.
Then $\mu(\{\sigma\}) = \frac{1}{n!}\ \forall \sigma$ is the Haar distribution on $S(n)$.

1) Let $G = [0,1[$ equipped with the addition modulo 1 $(x \oplus y)$.
Then $\mu([a,b]) = b - a$ is the Haar distribution on G
i.e. $\mu(x \oplus [a,b]) = \mu([a,b]) \quad \forall x, a, b \in [0,1[$



$$\int_0^1 f(x)\, d\mu(x) = \int_0^1 f(x)\, dx \quad , \quad f : [0,1[\to \mathbb{R}$$

× 2) Let $G = O(n) = \{ V \text{ } n\times n \text{ real matrix s.t. } VV^T = I \}$ (i.e. orthogonal matrix)

equipped with the standard matrix product.

Then the Haar distribution μ on G satisfies

× $\mu(V\cdot B) = \mu(B\cdot V) = \mu(B) \quad \forall V_{\text{orthogonal}},\ B \in \mathcal{B}(G)$

2b) Let $G = SO(n) = \{ V \in O(n) : \det V = +1 \}$ [rotations in $\mathbb{R}^n$]

(equipped again with the standard matrix product)

<u>NB</u>: $SO(2) = \left\{ \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} : 0 \le \theta < 2\pi \right\}$

is isomorphic to example 1.

3) Let $G = \mathcal{U}(n) = \{ U \text{ } n\times n \text{ complex matrix s.t. } UU^* = I \}$ (i.e. unitary matrix)

<u>NB</u>: $\mathcal{U}(1) = \{ e^{i\theta} : 0 \le \theta < 2\pi \}$

is also isomorphic to example 1

3b) Let $G = S\mathcal{U}(n) = \{ U \in \mathcal{U}(n) : \det U = +1 \}$

<u>Concrete example of Haar distribution</u>: $G = SO(3)$ [rotations in $\mathbb{R}^3$]

- Let $Z(\theta) = \begin{pmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}$ and $X(\varphi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\varphi & \sin\varphi \\ 0 & -\sin\varphi & \cos\varphi \end{pmatrix}$
- Any $V \in SO(3)$ may be represented as $Z = X(\varphi_1)\, Z(\theta)\, X(\varphi_2)$

  for some $\varphi_1 \in [0, 2\pi]$, $\theta \in [0, \pi]$, $\varphi_2 \in [0, 2\pi]$ (Euler angles)

× • $\displaystyle\int_{SO(3)} f(V)\, d\mu(V) = \frac{1}{8\pi^2} \int_0^{\pi} d\theta \sin\theta \int_0^{2\pi} d\varphi_1 \int_0^{2\pi} d\varphi_2\, f(X(\varphi_1)\, Z(\theta)\, X(\varphi_2))$

$(f : SO(3) \to \mathbb{R})$

Circular Orthogonal Ensemble (COE)

Let V be a $n \times n$ real orthogonal matrix picked "uniformly at random" i.e. according to the Haar distribution on $O(n)$.

Remarks:

- These matrices arise frequently in wireless commucations (as well as their sisters from the CUE).
- As already seen, there are only $\frac{n(n-1)}{2}$ free parameters in an orthogonal matrix V, so not all the entries of V can be independent. It is actually quite difficult to describe the joint distribution of the entries of V.

Question: how to pick V from the COE in practice?

1) Let $v_1 \dots v_n$ be the columns of V:
   - pick v_1 uniformly on the unit sphere ($\|v_1\| = 1$)
   - pick v_2 uniformly on the set $\{v_2 \in \mathbb{R}^n : \|v_2\| = 1 \ \& \ v_2 \perp v_1\}$
   - and so on ... until v_n, which is fixed. (up to a $\pm$)

× 2) Pick H from the GOE; compute its eigenvectors; the resulting matrix V of eigenvectors is then a matrix from the COE.

3) Let H be a matrix with iid $\sim \mathcal{N}_{\mathbb{R}}(0,1)$ entries; perform the Gram-Schmidt decomposition of H; the resulting orthogonal matrix is again a matrix from the COE. [ref: Eaton, Multivariate Statistics]

<u>Joint eigenvalue distribution</u>

<u>NB</u>: V is orthogonal ($VV^T = V^TV = I$), so:

- its eigenvalues are located on the unit circle in $\mathbb{C}$, i.e. $|\lambda_j| = 1 \; \forall j$ (but the λ_j are <u>not</u> real in general) ×
- V is normal ($VV^* = V^*V$), so it is unitarily diagonalizable (but <u>not</u> necessarily orthogonally diagonalizable) ×

Let us write $\lambda_j = e^{i\theta_j}$ with $\theta_j \in [0, 2\pi[$

Then $p(\theta_1 \ldots \theta_n) = C_n \cdot \underbrace{\prod_{j<k} |e^{i\theta_k} - e^{i\theta_j}|}_{\text{Jacobian (cf. GOE)}}$ (no proof)

interpretation: 1 (unif. dist.)

Circular Unitary Ensemble (CUE)

$\left\{ \begin{array}{l} \text{Let } U \text{ be a } n\times n \text{ complex unitary matrix} \\ \qquad \text{picked according to the Haar distribution on } U(n). \end{array} \right.$

Similar remarks as above apply. Notice one difference: in a unitary matrix, there are $\underset{\text{choice for } u_1}{(2n-1)} + \underset{\text{choice for } u_2}{(2n-3)} + \ldots + 1 = n^2$ free real parameters

Again, since U is unitary ($UU^* = U^*U = I$), it is normal, therefore unitarily diagonalizable and its eigenvalues are of the form $\lambda_j = e^{i\theta_j}$, $\theta_j \in [0, 2\pi[$.

Joint eigenvalue distribution:

$$p(\theta_1, \ldots, \theta_n) = \underset{1,\text{ again}}{C_n} \cdot \underbrace{\prod_{j<k} |e^{i\theta_k - i\theta_j}|^2}_{\text{Jacobian (cf. GUE)}}$$

× Marginals:

$$\left\{ \begin{array}{l} p(\theta) = \frac{1}{2\pi} \quad \text{(uniform distribution on the circle)} \\ p(\theta, \varphi) = \frac{1}{4\pi^2} \cdot \frac{n}{n-1} \left(1 - \left(\frac{\sin\left(\frac{n(\theta-\varphi)}{2}\right)}{n \sin\left(\frac{\theta-\varphi}{2}\right)} \right)^2 \right) \end{array} \right.$$

[→ homework 3 (+ proof of Mehta's lemma in a simple case)]

Generalization and "physical" interpretation

All the random matrix ensembles seen so far fit into the general model:

$$p(\lambda_1 .. \lambda_n) = C_n \exp\left(-\sum_{j=1}^{n} V(\lambda_j)\right) \cdot \prod_{j<k} |\lambda_k - \lambda_j|^\beta$$

where V is a given function [NB: the corresponding distribution of entries is not known in general]

and $\beta = 1$ for real matrices, $\beta = 2$ for complex matrices

Observation: the Jacobian term $\prod_{j<k} |\lambda_k - \lambda_j|^\beta$ is very small whenever two eigenvalues are close to each other; i.e., on average, the eigenvalues tend to repel each other (& more in the complex case than in the real case)

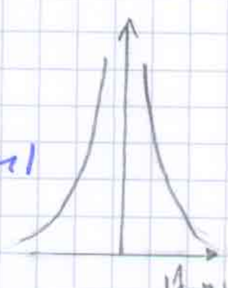
Rewriting:

× $$\underbrace{p(\lambda_1 .. \lambda_n)}_{\text{Gibbs distribution}} = C_n \cdot \exp\Bigg(-\underbrace{\Big(\underbrace{\sum_{j=1}^{n} V(\lambda_j)}_{\text{potential term}} - \underbrace{\beta \sum_{j<k} \log |\lambda_k - \lambda_j|}_{\substack{\text{interaction term}\\ \text{(repulsion)}}}\Big)}_{\text{energy}}\Bigg)$$

$\lambda_1 .. \lambda_n$ = "particles" tending to minimize energy

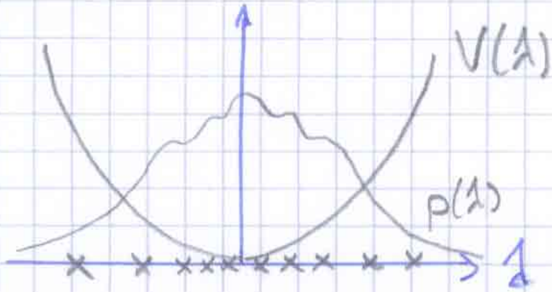
i.e., a) to find the minima of the potential V

b) coping with the repulsion term $-\beta \log |\lambda - \mu|$



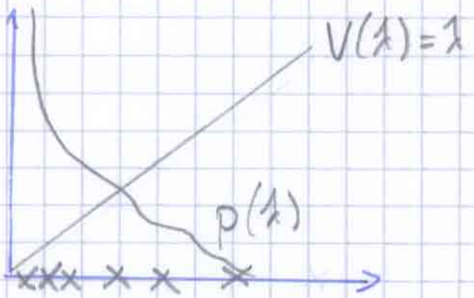
Examples

- **GOE or GUE**: $\lambda_j \in \mathbb{R}$ and $V(\lambda) = \frac{\lambda^2}{2}$

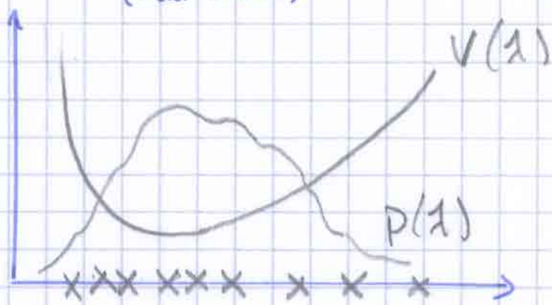


- **Real or complex Wishart ensemble**: $(m \geq n)$

$$\lambda_j \geq 0 \quad \text{and} \quad V(\lambda) = \underbrace{\frac{\lambda}{2} - \left(\frac{m-n-1}{2}\right)\log\lambda}_{\text{(real case)}} \quad \text{or} \quad \underbrace{\lambda - (m-n)\log\lambda}_{\text{(complex case)}}$$

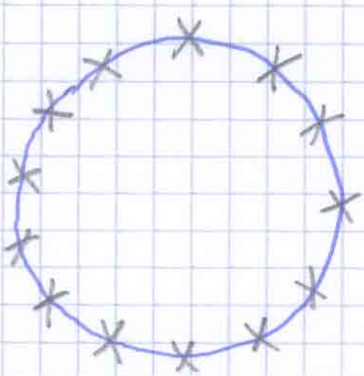


$m = n$ (complex case)



$m > n$ (complex case)

- **COE or CUE**: $|\lambda_j| = 1$, i.e. $\lambda_j = e^{i\theta_j}$ and $V(\lambda) \equiv 1$



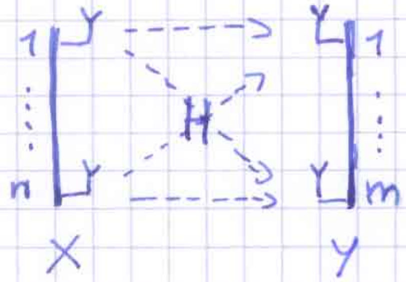
uniform and regular distribution of the eigenvalues on the unit circle ($\neq$ iid points!)

Random matrix theory: lecture 6

Capacity of multi-antenna channels [ref: Telatar 1995]

(also known as MIMO [multiple input - multiple output] channels)

Model:



n transmit antennas

m receive antennas

$$Y = HX + Z, \quad \text{i.e. } y_j = \sum_{k=1}^{n} h_{jk} x_k + z_j, \quad j = 1..m$$

- $X = (x_1 .. x_n)$ input: x_k = signal sent by antenna #k

  ↳ subject to the global power constraint $\sum_{k=1}^{m} \mathbb{E}(|x_k|^2) \leq P$

- $Y = (y_1 .. y_m)$ output: y_j = signal received by antenna #j

- $Z = (z_1 .. z_m)$ background noise: $z_1 .. z_m$ = iid r.v. $\sim N_{\mathbb{C}}(0,1)$

  notation: $Z \sim N_{\mathbb{C}}(0, I)$

- Z has i.i.d. realizations over time

- X and Z are assumed to be independent

- $H = (h_{jk})_{j,k=1}^{m,n}$ $m \times n$ channel matrix;

  h_{jk} = attenuation factor between transmit antenna #k and receive antenna #j

× We want to compute the capacity of the channel $X \to Y$ in three different scenarios.

First scenario: H is deterministic

Preliminary:

- Let X be a complex random vector with density p_X.

  differential entropy: $h(X) = -\int_{\mathbb{C}^n} p_X(x) \log p_X(x)\, dx$

× for a given covariance matrix $Q_X = \mathbb{E}(XX^*)$,
  $h(X)$ is maximized whenever X is Gaussian
  (notation: $X \sim N_{\mathbb{C}}(0, Q_X)$), in which case we have
  $h(X) = \log \det(\pi e Q_X)$

- Let X, Y be two complex random vectors

× mutual information: $I(X;Y) = h(Y) - h(Y|X)$

  in the case where $Y = HX + Z$ with H deterministic,

× we therefore have $I(X;Y) = h(Y) - h(Z)$

× The capacity of the channel $X \to Y$ is given by

$$\boxed{C = \max_{p_X : \sum_{k=1}^{n} \mathbb{E}(|x_k|^2) \le P} I(X,Y)}$$

NB: $\sum_{k=1}^{n} \mathbb{E}(|x_k|^2) = \mathbb{E}(\|X\|^2) = \mathbb{E}(X^*X)$
$= \mathbb{E}(\mathrm{Tr}(XX^*)) = \mathrm{Tr}(\mathbb{E}(XX^*)) = \mathrm{Tr}\, Q_X$

Therefore, $\boxed{C = \max_{p_X : \mathrm{Tr}\, Q_X \leq P} h(Y) - h(Z)}$

Now, $h(Z) = \log\det(\pi e Q_Z) = \log\det(\pi e I)$ (indep. of p_X)

and for a given Q_X, $Q_Y = H Q_X H^* + I$ is fixed,

so $h(Y)$ is maximum when Y is Gaussian,

and $Y = HX + Z$ is Gaussian if X is Gaussian.

In this case, we have $h(Y) = \log\det(\pi e (H Q_X H^* + I))$

so $\boxed{C = \max_{Q_X \geq 0 : \mathrm{Tr}\, Q_X \leq P} \log\det(I + H Q_X H^*)}$

(notation for "Q_X is non-negative definite" [covariance matrix])

Recall that $\det(I + H Q_X H^*) = \det(I + H^* H Q_X)$

and that there exist U $n \times n$ unitary and $\Lambda = \mathrm{diag}(\lambda_1 \ldots \lambda_n)$

such that $H^* H = U \Lambda U^*$. Therefore:

$$\det(I + H Q_X H^*) = \det(I + \Lambda^{1/2} U^* Q_X U \Lambda^{1/2})$$

Observe that $\widetilde{Q}_X := U^* Q_X U$ is non-negative definite

if and only if Q_X is, and that $\mathrm{Tr}\, \widetilde{Q}_X = \mathrm{Tr}\, Q_X$

Therefore, $\boxed{C = \max_{\widetilde{Q}_X \geq 0;\ \mathrm{Tr}\, \widetilde{Q}_X \leq P} \log\det(I + \Lambda^{1/2} \widetilde{Q}_X \Lambda^{1/2})}$

(NB: $\Lambda^{1/2} = \mathrm{diag}(\sqrt{\lambda_1}, \ldots, \sqrt{\lambda_n})$ [recall that $\lambda_j \geq 0$])

$\Big\{$ Hadamard's inequality: (one version)

If A is non-negative definite, then $\det A \le \prod_{j=1}^{n} a_{jj}$.

(NB: recall that $A \ge 0$ implies $A = A^*$ in the complex case)

It is the case that $I + \Lambda^{1/2} \tilde{Q}_x \Lambda^{1/2} \ge 0$, so

$$\det(I + \Lambda^{1/2} \tilde{Q}_x \Lambda^{1/2}) \le \prod_{k=1}^{n} (1 + (\tilde{Q}_x)_{kk} \lambda_k)$$

with equality whenever $\tilde{Q}_x$ is diagonal $(= D)$.

× So $C = \max_{D \text{ diag} \ge 0 : \operatorname{Tr} D \le P} \log\det(I + \Lambda D)$

i.e. $\boxed{C = \max_{d_k \ge 0 : \sum_{k=1}^{n} d_k \le P} \sum_{k=1}^{n} \log(1 + d_k \lambda_k)}$

where $\lambda_1 \ldots \lambda_n$ are the eigenvalues of $H^* H$.

The solution to the above maximization problem is the so-called "water-filling" solution, which can be found through Kuhn-Tucker conditions:

$$d_k = \left(\nu - \frac{1}{\lambda_k}\right)^+ \quad \text{for some parameter } \nu$$

i.e. $$\boxed{\begin{array}{l} C = \sum_{k=1}^{n} (\log(\nu \lambda_k))^+ \\ \text{where } \sum_{k=1}^{n} \left(\nu - \frac{1}{\lambda_k}\right)^+ \le P \end{array}}$$

PS: an alternate derivation of this result may be obtained via the singular value decomposition of H

<u>Second scenario</u>: H is a random matrix that varies in an ergodic manner over time ["fast fading"]

At each time instant, we assume that H has the same stationary distribution $p(H)$, but at this point, we make no particular assumption on what $p(H)$ is.

<u>Remark 1</u>: if $p(H)$ is not known, then the channel transition probability $p(Y|X)$ is not known, so there is no notion of capacity in this case.

Let us therefore assume that $p(H)$ is known (to everybody).

A further question is: Who knows the <u>realizations</u> of the matrix H over time? This is an important question, since the communication strategy described in the preceding scenario (H deterministic) requires both the transmitter and the receiver to know H.

a) nobody knows the realizations of H a priori:

This seems a plausible assumption at first sight; the resulting analysis for computing the capacity in this case is quite difficult; we will skip that.

b) both the transmitter and the receiver know the realizations of H: apart from the "genie aided" interpretation, this assumption is reasonable when the matrix H varies slowly over time (but still fast enough so that it remains an ergodic process...), so that the receiver is able to estimate H through a sequence of pilot symbols sent by the transmitter, and then feed H back to the transmitter. In this case, it is as if the channel matrix H were deterministic, so the channel capacity is given by

$$\boxed{C = \mathbb{E}_H\left(\max_{Q_x \geq 0 \,:\, \mathrm{Tr}\, Q_x \leq P} \log\det\left(I + H Q_x H^*\right)\right)}$$

× where the expectation $\mathbb{E}_H$ comes from the fact that H varies ergodically over time.

(NB: not much more can be said in this case)
[the random water-filling solution is not explicit]

c) another interesting situation is when H varies slowly, so that the receiver can estimate H accurately, but the feedback link is weak, so that the transmitter does not get the estimate.

In this case, the channel becomes (ideally)

$$X \longrightarrow (Y, H)$$

[it is as if the receiver also "receives" the matrix H]

× By the chain rule,

$$I(X; Y, H) = \underbrace{I(X; H)}_{=0 \text{ since we assume } X \,\&\, H \text{ indep.}} + I(X; Y | H)$$

× and

× $$I(X; Y | H) = \int dG \, \overset{\text{dist. of } H}{p(G)} \, I(X; Y | H = G)$$

For a given input covariance matrix Q_x and a given G, we have $I(X; Y | H=G) \leq \log \det (I + G Q_x G^*)$,

so

$$I(X; Y | H) \leq \int dG \, p(G) \log \det (I + G Q_x G^*)$$

$$= \mathbb{E}_H \left(\log \det (I + H Q_x H^*) \right)$$

And the ("ergodic") capacity is given by

$$C = \max_{p_X : \mathrm{Tr}\, Q_X \le P} I(X; Y, H)$$

so
$$\boxed{C = \max_{Q_X \ge 0 : \mathrm{Tr}\, Q_X \le P} \mathbb{E}_H \left(\log \det (I + H Q_X H^*) \right)}$$

<u>Remark 2</u>: at this point, it is <u>not</u> appropriate to say that "since the transmitter does not know H, the best possible input covariance matrix Q_X is proportional to the identity matrix" (i.i.d. signals)

× Depending on the distribution of H, the true maximizing Q_X might be quite different.

On the other hand, it is clear that the solution to the above maximization problem is <u>not</u> the water-filling solution, because of the fact that the expectation is inside now.

(this is the penalty we have to pay for the transmitter not knowing H, actually)

Random matrix theory : lecture 7

Capacity of multi-antenna channels (cont'd)

Reminder: we study the multi-antenna channel $X \to Y = HX + Z$

main assumptions : • H $m \times n$ random matrix

• H varies ergodically over time ("second scenario")

× • at each time instant, H has the same distribution $p(H)$

• the realizations of H are known to the receiver, but not to the transmitter

Under these assumptions, we get the following capacity expression:

$$C = \max_{Q \geq 0 : \mathrm{Tr}\, Q \leq P} \mathbb{E}\left(\underbrace{\log\det\left(I + HQH^*\right)}_{:= \varphi_H(Q)} \right)$$

(NB : in order to lighten the notation, we skip the H in $\mathbb{E}_H$ and also skip the X in Q_X)

We are going to see how to simplify the above maximization problem when the distribution of H is invariant under some transformations.

ref : Abbé - Telatar - Zheng , Allerton 2005

Preliminary proposition:

× The map $A \mapsto \log\det A$ is concave on the set of positive-definite matrices ($A>0$)

Therefore, $Q \mapsto \varphi_H(Q) := \log\det(I + HQH^*)$

× is also concave on the set of non-negative definite matrices (if $Q \geq 0$, then $I + HQH^* > 0$).

Lemma 1

If H has the same distribution as $H\Sigma$ (notation: $H \sim H\Sigma$) for any $n\times n$ matrix Σ of the form $\Sigma = \begin{pmatrix} \pm1 & & 0 \\ & \pm1 & \\ 0 & & \pm1 \ddots \end{pmatrix}$, then

× $$C = \max_{Q \text{ diag} \geq 0 : \operatorname{Tr} Q \leq P} \mathbb{E}(\varphi_H(Q))$$

Proof

- By assumption, $\varphi_H(Q)$ has the same distribution as

$$\begin{aligned} \varphi_{H\Sigma}(Q) &= \log\det(I + (H\Sigma)Q(H\Sigma)^*) \\ &= \log\det(I + H(\Sigma Q \Sigma^*)H^*) \\ &= \varphi_H(\Sigma Q \Sigma^*) \end{aligned}$$

× So $\mathbb{E}(\varphi_H(Q)) = \mathbb{E}(\varphi_H(\Sigma Q \Sigma^*))$

$$= \frac{1}{2}\left(\mathbb{E}(\varphi_H(Q)) + \mathbb{E}(\varphi_H(\Sigma Q \Sigma^*))\right)$$

i.e. $\mathbb{E}(\psi_H(Q)) = \mathbb{E}\left(\frac{1}{2}(\psi_H(Q) + \psi_H(\Sigma Q \Sigma^*))\right)$

- Since $Q \longmapsto \psi_H(Q)$ is concave, we further obtain

$$\mathbb{E}(\psi_H(Q)) \leq \mathbb{E}\left(\psi_H\left(\tfrac{1}{2}(Q + \Sigma Q \Sigma^*)\right)\right) \quad \text{[Jensen inequality]}$$

for any matrix $\Sigma = \mathrm{diag}(\pm 1)$ [NB: $\Sigma^* = \Sigma$]

- Consider e.g. $\Sigma = \begin{pmatrix} -1 & & 0 \\ & +1 & \\ 0 & & \ddots \;\; +1 \end{pmatrix}$:

$$\Sigma Q \Sigma^* = \begin{pmatrix} q_{11} & -q_{12} & \cdots & -q_{1n} \\ -q_{21} & & & \\ \vdots & & Q_1 & \\ -q_{n1} & & & \end{pmatrix}$$

so $\frac{1}{2}(Q + \Sigma Q \Sigma^*) = \begin{pmatrix} q_{11} & 0 & \cdots & 0 \\ 0 & & & \\ \vdots & & Q_1 & \\ 0 & & & \end{pmatrix} := \tilde{Q}$

- That is, for any given input covariance matrix Q, there exists another covariance matrix $\tilde{Q}$ (with same trace) such that $\mathbb{E}(\psi(Q)) \leq \mathbb{E}(\psi(\tilde{Q}))$ and $\tilde{Q}$ has off-diagonal elements of the first row & first column which are all equal to zero.

- Proceeding recursively, we can "erase" all the off-diagonal elements and therefore show that choosing Q diagonal is sufficient to maximize the expectation. #

Lemma 2

× If $H \sim H\Pi$ for any $n \times n$ permutation matrix Π,

then $C = \max\limits_{\rho \in [-\frac{1}{n-1}, 1]} \mathbb{E}(\psi_H(Q_\rho))$

where $Q_\rho = \frac{P}{n}\begin{pmatrix} 1 & & \rho \\ & \ddots & \\ \rho & & 1 \end{pmatrix}$ equal diagonal elements & equal off-diagonal elements

× [NB: $\operatorname{Tr} Q_\rho = \frac{P}{n} \cdot n = P$ and $Q_\rho \geq 0$ iff $\rho \in [-\frac{1}{n}, 1]$]

Proof (same idea as before)

- By assumption, $\psi_H(Q) \sim \psi_{H\Pi}(Q) = \psi_H(\Pi Q \Pi^*)$,

  so $\mathbb{E}(\psi_H(Q)) = \mathbb{E}\left(\frac{1}{2}(\psi_H(Q) + \psi_H(\Pi Q \Pi^*))\right)$

  $\leq \mathbb{E}\left(\psi_H\left(\frac{1}{2}(Q + \Pi Q \Pi^*)\right)\right)$ concavity & Jensen

- Similarly, let $\tilde{Q} := \frac{1}{n!}\sum\limits_{\Pi \in P(n)} \Pi Q \Pi^*$

  $\llcorner$ set of permutation matrices [NB: $\Pi^* \neq \Pi$ in general!]

  Then $\mathbb{E}(\psi_H(Q)) \leq \mathbb{E}(\psi_H(\tilde{Q}))$

- But note that $\tilde{Q}$ has equal diagonal elements & equal off-diagonal elements. Ex. in the case $n=2$:

× $$\frac{1}{2}(Q + \underbrace{\Pi}_{=\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}} Q \Pi^*) = \frac{1}{2}\left(\begin{pmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{pmatrix} + \begin{pmatrix} q_{22} & q_{21} \\ q_{12} & q_{11} \end{pmatrix}\right)$$

$$= \begin{pmatrix} \frac{1}{2}(q_{11}+q_{22}) & \frac{1}{2}(q_{12}+q_{21}) \\ \frac{1}{2}(q_{12}+q_{21}) & \frac{1}{2}(q_{11}+q_{22}) \end{pmatrix}$$

- Because of the trace constraint and the constraint that $\tilde{Q}$ is non-negative definite, $\tilde{Q}$ is necessarily of the form given in the lemma.

#

Remarks:

- The above argument had been already used in [Telatar 95], in the case where the optimal Q was known to be diagonal.
- The same argument could have been used in lemma 1 also, observing that $\tilde{Q} := \frac{1}{2^n} \sum_{\Sigma = \mathrm{diag}(\pm 1)} \Sigma Q \Sigma^*$ is diagonal.

Application:

1) Let H be a $m \times n$ matrix with independent entries such that $h_{jk} \sim -h_{jk}$ for all j, k.
× Then $H\Sigma \sim H \quad \forall \Sigma$ (ex: $H \begin{pmatrix} -1 & & 0 \\ & 1 & \\ 0 & & \ddots \; 1 \end{pmatrix} = \begin{pmatrix} -h_{11} & h_{12} & \cdots & h_{1n} \\ \vdots & & & \\ -h_{m1} & h_{m2} & \cdots & h_{mn} \end{pmatrix} \sim H$),
so lemma 1 applies, i.e. the optimal Q is diagonal.
2) Let H be a $m \times n$ matrix with i.i.d. entries
× Then $H\Pi \sim H \quad \forall \Pi$ (clear), so lemma 2 applies, i.e. the optimal Q has equal diagonal el. & equal off-diag. el.

1+2) Let H be a $m \times n$ matrix with iid entries such that $h_{jk} \sim -h_{jk} \; \forall j k$. Then $H \sim H\Sigma \sim H\Pi$ for any Σ & Π, so the optimal Q is of the form
× $Q = \frac{P}{n} I$.

A particular case: "i.i.d. Rayleigh fading" [Telatar 95]

Let us assume that H is a $m \times n$ matrix with i.i.d. entries distributed as $N_{\mathbb{C}}(0,1)$ r.v.

Remark: this assumption is an assumption of the same flavor as that made by Wigner in physics: rather than trying to model precisely the $n \cdot m$ fading coefficients with a complicated deterministic model, let us consider simply the matrix H as "completely random".

The above distribution falls into the above case (1+2), so the optimal input covariance matrix is $Q = \frac{P}{n} I$ and the capacity is given by $C = \mathbb{E}\left(\log\det\left(I + \frac{P}{n} H H^*\right)\right)$.

Remark: an alternate way to derive the result is to notice that in this very particular case, the distribution of H is unitarily invariant, i.e. that $H \sim HU$ for any $n \times n$ unitary matrix U. This implies in particular that $H \sim H\Sigma$ and $H \sim H\Pi$, since both Σ and Π are unitary matrices.

The capacity may be further written as

$$C = \mathbb{E}\left(\sum_{j=1}^{m} \log\left(1 + \frac{P}{n} \lambda_j\right)\right)$$

with λ_j the eigenvalues of the $m \times m$ matrix HH^*.

For simplicity, let us consider the case where $\underline{m=n}$:

$$\begin{aligned} C &= \int_{\mathbb{R}_+^n} d\lambda_1 \ldots d\lambda_n \cdot p(\lambda_1 \ldots \lambda_n) \cdot \left(\sum_{j=1}^{n} \log\left(1 + \frac{P}{n} \lambda_j\right)\right) \\ &= \sum_{j=1}^{n} \int_{\mathbb{R}_+} d\lambda_j \, p(\lambda_j) \log\left(1 + \frac{P}{n} \lambda_j\right) \\ &= n \int_{\mathbb{R}_+} d\lambda \, p(\lambda) \log\left(1 + \frac{P}{n} \lambda\right) \\ \text{(see lecture 4)} &= \sum_{\ell=0}^{n-1} \int_{\mathbb{R}_+} d\lambda \, e^{-\lambda} \underbrace{L_\ell(\lambda)}_{\text{(= Laguerre polynomials)}}{}^2 \log\left(1 + \frac{P}{n} \lambda\right) \end{aligned}$$

It turns out that this expression is proportional to n as n gets large (see the forthcoming asymptotic analysis).

More generally, C is proportional to $\min(m,n)$ when $m \neq n$.

Third scenario: H is a random matrix

× that is fixed once and for all ("slow fading")

Moreover, we assume again that the receiver knows the realizations of H, but not the transmitter (situation c). We also assume that H is a $m \times n$ matrix with iid $\sim \mathcal{N}_{\mathbb{C}}(0,1)$ entries.

For a given X with input covariance Q and a given H, the mutual information between X and $Y = HX + Z$ is $\log\det(I + HQH^*)$.

There is always a strictly positive probability that this

× expression is arbitrarily small, so the capacity is zero.

We therefore shift our attention to a new quantity; the

× outage probability: for a target rate R, we define

× $$\mathbb{P}_{out}(R) = \min_{Q \geq 0 : \mathrm{Tr}\, Q \leq P} \underbrace{\mathbb{P}\left(\log\det(I + HQH^*) < R \right)}_{= \text{ cumulative distribution function}}$$

This optimization problem is considerably harder than the preceding; actually, the solution is not known!

NB: C = upper bound on achievable rate

$\mathbb{P}_{out}(R)$ = lower bound on error probability

Since $H \sim HU$ for any $n \times n$ unitary matrix U

and the constraint is also unitarily invariant, we have

$$P_{out}(R) = \min_{Q \underline{\underline{\text{diag}}} \geq 0 :\, \mathrm{Tr}\, Q \leq P} \mathbb{P}\left(\log\det\left(I + HQH^*\right) < R\right)$$

<u>Conjecture</u> [Telatar 95] [resolved so far only for $m=1$]

The optimal input covariance matrix Q is of

the form $Q = \mathrm{diag}(\underbrace{\tfrac{P}{k}, \dots, \tfrac{P}{k}}_{k \text{ times}}, 0, \dots, 0)$ [the order is irrelevant here]

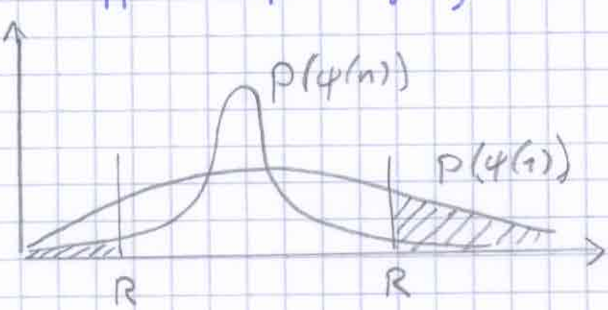
for some $1 \leq k \leq n$.

<u>Remark</u>: the optimal k should depend on the target rate R:

Let $\varphi(k) = \log\det\left(I + H\,\mathrm{diag}\left(\tfrac{P}{k}, \dots, \tfrac{P}{k}, 0, \dots, 0\right) H^*\right)$

= mutual information obtained by using k antennas

- If R is sufficiently small, then $\varphi(n)$ is greater than R with high probability, so $k=n$ is optimal (averaging effect)
- If R is sufficiently large, then $k=1$ is optimal: ("all eggs in one basket")



At high SNR ($P \to \infty$), this problem simplifies

and more can be said [next time].

Random matrix theory: lecture 8

Rate-diversity tradeoff in multi-antenna channels

Back to the third scenario: the quantity of interest is

$$\mathbb{P}_{out}(R) = \min_{Q \geq 0 : \operatorname{Tr} Q \leq P} \mathbb{P}\left(\log\det(I + HQH^*) < R\right)$$

We again assume that h_{jk} are iid $\sim N_{\mathbb{C}}(0,1)$, and would like to perform the analysis of $\mathbb{P}_{out}$ in the high SNR regime ($P \to \infty$). [ref: Zheng-Tse 03]

In the ergodic case, we have seen that

$$C = \max_{Q \geq 0 : \operatorname{Tr} Q \leq P} \mathbb{E}\left(\log\det(I + HQH^*)\right) = \mathbb{E}\left(\log\det(I + \tfrac{P}{n} HH^*)\right)$$

$$= \mathbb{E}\left(\sum_{j=1}^{n} \log(1 + P\lambda_j)\right) \simeq \min(m,n) \cdot \log P \quad \text{as } P \to \infty$$

(as shown by a simple heuristics)

Let us therefore choose a target rate $R = r \log P$, for some $0 \leq r \leq \min(m,n)$. We expect that

$$\mathbb{P}_{out}(r \log P) \approx P^{-d(r)} \quad \text{for some } d(r) > 0.$$

$\Rightarrow$ definition: the diversity order $d(r)$ associated to a rate r is: $d(r) := \lim_{P \to \infty} -\dfrac{\log(\mathbb{P}_{out}(r \log P))}{\log P}$

[= upper bound on the diversity order of the error probability arising from a given coding scheme]

Our aim now is to compute $d(r)$ for

H $m \times n$ matrix with iid $\sim N_{\mathbb{C}}(0,1)$ entries and $\underline{m \geq n}$.

[without loss of generality]

Notation: $f(P) \doteq g(P)$ if $\lim_{P\to\infty} \frac{\log f(P)}{\log P} = \lim_{P\to\infty} \frac{\log g(P)}{\log P}$

[same "diversity order" for f & g]

Lemma

× $\quad P_{out}(r \log P) \doteq \mathbb{P}\left(\log\det(I + P H^* H) < r \log P\right)$

Proof

- $A(P) := \mathbb{P}(\underbrace{\log\det(I + \tfrac{P}{n} H^* H))}_{< r\log P} \geq P_{out}(r \log P)$ (choose $Q = \frac{P}{n} I$)
- × for any $Q \geq 0$ s.t. $\mathrm{Tr}\, Q \leq P$, we have $Q \leq P I$;

  × since $A \mapsto \log\det A$ is increasing on the set of positive definite matrices, this implies

  $\log\det(I + H Q H^*) \leq \log\det(I + P H^* H)$ for any Q,

  so $P_{out}(r\log P) \geq \mathbb{P}\left(\log\det(I + P H^* H) < r \log P\right) := B(P)$
- finally, $A(P) \doteq B(P)$; indeed:

$$B(P) \leq A(P) = \mathbb{P}\left(\log\det(I + \tfrac{P}{n} H^* H) < r \log P\right)$$

$$\doteq \mathbb{P}\left(\log\det(I + P H^* H) < r \log (nP)\right)$$

$$\doteq \mathbb{P}\left(\log\det(I + P H^* H) < r \log P\right) = B(P)$$

#

Bottom line: the hard optimization problem is gone!

Therefore,

$$\mathbb{P}_{out}(r\log P) \doteq \mathbb{P}\Big(\sum_{j=1}^{n} \log(1+P\underbrace{\lambda_j}_{\text{ev. of } H^*H}) < r\log P\Big)$$

$$= \int_{D_\lambda(r)} p(\lambda_1 .. \lambda_n)\, d\lambda_1 .. d\lambda_n$$

where $p(\lambda_1, .., \lambda_n) = C_n \cdot \exp\Big(-\sum_{j=1}^{n} \lambda_j + (m-n)\sum_{j=1}^{n} \log \lambda_j\Big)$

$$\cdot \prod_{j<k} (\lambda_k - \lambda_j)^2 \quad \text{(see lecture 3)}$$

and $D_\lambda(r) = \Big\{ \underbrace{0 \le \lambda_1 \le .. \le \lambda_n}_{\text{eigenvalues ordered in ascending order}} : \sum_{j=1}^{n} \log(1+P\lambda_j) < r\log P \Big\}$

Let us now make the following <u>change of variables</u>:

$$\underset{\ge 0}{\lambda_j} = P^{-\overset{\in \mathbb{R}}{\alpha_j}} = e^{-\alpha_j \log P} \quad \Rightarrow d\lambda_j = -(\log P)\, e^{-\alpha_j \log P} d\alpha_j$$

so

$$\mathbb{P}_{out}(r\log P) \doteq \int_{D_\alpha(r)} q(\alpha_1 .. \alpha_n)\, d\alpha_1 .. d\alpha_n$$

where $q(\alpha_1 .. q_n) = C_n \exp\Big(-\sum_{j=1}^{n} P^{-\alpha_j} - (m-n)\sum_{j=1}^{n} \alpha_j \log P\Big)$

$$\cdot \prod_{j<k} (P^{-\alpha_k} - P^{-\alpha_j})^2 (\log P)^n \exp\Big(-\sum_{j=1}^{n} \alpha_j \log P\Big)$$

and $D_\alpha(r) = \Big\{ \underset{(\alpha_j \in \mathbb{R})}{\alpha_1 \ge ... \ge \alpha_n} : \sum_{j=1}^{n} \log(1+P^{1-\alpha_j}) < r\log P \Big\}$

At this point, let us make a couple of observations:

a) as $P \to \infty$, $\exp(-P^{-\alpha_j}) \begin{cases} \text{decays super-polynomially to } 0 & \text{if } \alpha_j < 0 \\ \text{tends to } 1 & \text{if } \alpha_j \geq 0 \end{cases}$

so we may restrict the integral to $\alpha_j \geq 0 \ \forall j$

b) as $P \to \infty$, $\log(1 + P^{1-\alpha_j}) \approx \begin{cases} (1-\alpha_j) \log P & \text{if } \alpha_j \leq 1 \\ 0 & \text{if } \alpha_j > 1 \end{cases}$

so $\log(1+P^{1-\alpha_j}) \approx \underbrace{(1-\alpha_j)^+}_{\text{positive part}} \log P$

We may therefore replace the domain of integration $D_\alpha(r)$ by $\widetilde{D}_\alpha(r) = \left\{ \alpha_1 \geq \dots \geq \alpha_n \geq 0 : \sum_{j=1}^{n} (1-\alpha_j)^+ < r \right\}$

So $\mathbb{P}_{out}(r \log P) \doteq \int_{\widetilde{D}_\alpha(r)} \widetilde{q}(\alpha_1, \dots, \alpha_n) \, d\alpha_1 \dots d\alpha_n$

where $\widetilde{q}(\alpha_1 \dots \alpha_n) = C_n (\log P)^n \cdot e^{-(m-n+1)\sum_j \alpha_j \log P}$

$$\cdot \prod_{j<k} \left(P^{-\alpha_k} - P^{-\alpha_j} \right)^2$$

Now, since $\alpha_1 \geq \dots \geq \alpha_n$, we have

$$\prod_{j<k} \left(P^{-\alpha_k} - P^{-\alpha_j}\right)^2 \approx \prod_{j<k} P^{-2\alpha_k} = \prod_k P^{-2(k-1)\alpha_k}$$

$$= e^{-2\sum_k (k-1)\alpha_k \log P}$$

Note also that in terms of "diversity order", $(\log P)^n$ is equivalent to a constant, since $\lim_{P \to \infty} \frac{\log((\log P)^n)}{\log P} = 0$
(which is in turn equivalent to 1)

Finally,

$$P_{out}(r \log P) \doteq \int_{\tilde{D}_\alpha(r)} e^{-\overbrace{\sum_j (m-n+2j-1)\,\alpha_j}^{:=f(\alpha)} \log P} d\alpha_1 \ldots d\alpha_n$$

$$= \int_{\tilde{D}_\alpha(r)} P^{-f(\alpha)} d\alpha_1 \ldots d\alpha_n \underset{\uparrow}{\doteq} P^{-\min_{\tilde{D}_\alpha(r)} f(\alpha)}$$

Laplace integration method

The final answer is therefore: $P_{out}(r \log P) \doteq P^{-d(r)}$, where

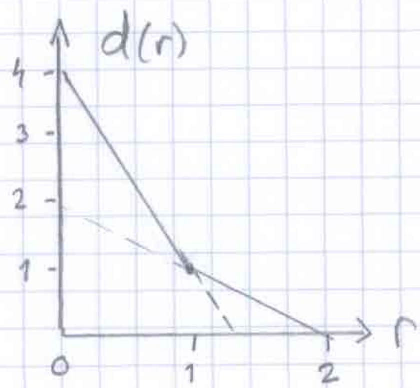
$$\boxed{d(r) = \min_{\tilde{D}_\alpha(r)} f(\alpha) = \min_{\left\{\alpha_1 \geq \ldots \geq \alpha_n \geq 0 : \sum_{j=1}^n (1-\alpha_j)^+ < r\right\}} \sum_{j=1}^n (m-n+2j-1)\,\alpha_j}$$

Ex: $m = n = 2$

$$d(r) = \min_{\{\alpha_1 \geq \alpha_2 \geq 0 : (1-\alpha_1)^+ + (1-\alpha_2)^+ < r\}} \alpha_1 + 3\alpha_2$$

$0 < r < 1$: $\alpha_1 = 1$ & $\alpha_2 = 1 - r$ & $d(r) = 4 - 3r$
i.e. $\lambda_1 \sim P^{-1}$ $\quad\lambda_2 \sim P^{r-1}$

$1 < r < 2$: $\alpha_1 = 2 - r$ & $\alpha_2 = 0$ & $d(r) = 2 - r$
i.e. $\lambda_1 \sim P^{r-2}$ $\quad\lambda_2 \sim 1$



tradeoff:

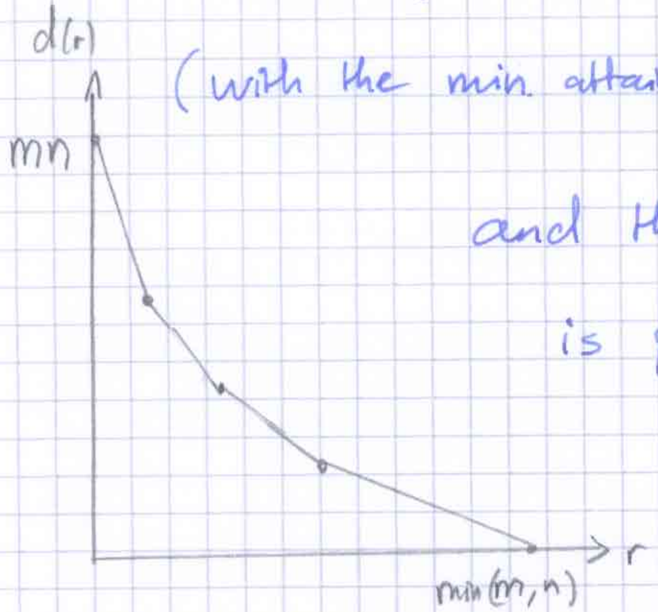
$\begin{cases} r \sim 2 : \text{full rate, low diversity} \\ r \sim 0 : \text{low rate, full diversity} \end{cases}$

Picture in the general case

It turns out that for integer $r = k$, the solution of the above minimization problem is:

$$d(k) = (m-k)(n-k)$$

(with the min attained at $\alpha_1 = \dots = \alpha_{n-k} = 1$, $\alpha_{n-k+1} = \dots = \alpha_n = 0$)



and the function inbetween is piecewise linear (& convex)

Interpretation

With m receive antennas and n transmit antennas, a rate $k \log P$ can be "achieved" if $k \leq \min(m,n)$; k antennas on each side contribute to this rate, while the remaining $m-k$ and $n-k$ antennas provide diversity.

NB: for a target rate $R = k \log P$, the outage event is the event that the channel matrix H gives rise to less than k independent reliable scalar channels.

Random matrix theory: lecture 9

Asymptotic analysis: first approach

Observation: the capacity of a $n \times n$ MIMO channel whose channel matrix H has i.i.d. $N_{\mathbb{C}}(0,1)$ entries is given by

$$C = \mathbb{E}\left(\log\det\left(I + \frac{P}{n} HH^*\right)\right)$$

Notice that, for each j fixed, $\left(\frac{1}{n}HH^*\right)_{jj} = \frac{1}{n}\sum_{\ell=1}^{n} |h_{j\ell}|^2 \xrightarrow[n\to\infty]{} \mathbb{E}(|h_{jj}|^2) = 1$ by the law of large numbers

Similarly $\left(\frac{1}{n}HH^*\right)_{jk} = \frac{1}{n}\sum_{\ell=1}^{n} h_{j\ell}\overline{h_{k\ell}} \xrightarrow[n\to\infty]{} \mathbb{E}(h_{j1}\overline{h_{k1}}) = 0 \quad \forall j \neq k$

Therefore, $\frac{1}{n}HH^* \underset{n\to\infty}{\sim} I$, so

$$C \underset{n\to\infty}{\sim} \mathbb{E}\left(\log\det(I + PI)\right) = n\log(1+P) \quad ???$$

Why would we need random matrix analysis, then?

The problem is that the determinant (and likewise, the eigenvalues) of a matrix is not a continuous function of the entries, as the size of the matrix goes to infinity. Therefore, even though $\frac{1}{n}HH^*$ converges entrywise to I as $n\to\infty$, $\det(I + \frac{P}{n}HH^*)$ diverges from $\det(I + PI)$.

Here is another simple (counter-)example: let $A = (1-\frac{1}{n})I$ of size $n \times n$

Then $A \underset{n\to\infty}{\sim} I$, but $\det A = \left(1-\frac{1}{n}\right)^n \underset{n\to\infty}{\sim} \frac{1}{e} \neq \det I = 1$.

General formulation of the problem

Let $(A^{(n)})_{n\geq 1}$ be a sequence of random matrices of increasing size ($A^{(n)}$ is of size $n \times n$), $\lambda_1^{(n)} \dots \lambda_n^{(n)}$ denote the (ordered or un-ordered) eigenvalues of $A^{(n)}$ and let $\lambda^{(n)}$ be one eigenvalue picked uniformly at random among $\lambda_1^{(n)} \dots \lambda_n^{(n)}$. Let also $p^{(n)}(\lambda)$ denote the distribution of $\lambda^{(n)}$. We are interested in the limiting behaviour of $p^{(n)}(\lambda)$ as $n \to \infty$.

First remark: for most random matrix ensembles, a rescaling of the matrices in n is needed in order to obtain the convergence of $p^{(n)}(\lambda)$.

a) (Square) complex Wishart Ensemble

Let $\underset{(n \times n \text{ matrix})}{H^{(n)}}$ have i.i.d. $N_{\mathbb{C}}(0,1)$ entries and $A^{(n)} = H^{(n)}(H^{(n)})^*$.

What is the first moment of $p^{(n)}(\lambda)$?

$$\int_0^\infty \lambda\, p^{(n)}(\lambda)\, d\lambda = \mathbb{E}(\lambda^{(n)}) = \mathbb{E}\left(\frac{1}{n}(\lambda_1^{(n)} + \dots + \lambda_n^{(n)})\right)$$

$$= \mathbb{E}\left(\frac{1}{n}\operatorname{Tr}(A^{(n)})\right) = \frac{1}{n}\sum_{j,k=1}^{n} \underbrace{\mathbb{E}\left(|h_{jk}^{(n)}|^2\right)}_{=1} = \frac{1}{n}\cdot n^2 = n$$

This heuristics indicates that in order to obtain the convergence of at least the first moment of $p^{(n)}(\lambda)$, we need to scale down the matrix $A^{(n)}$ by a factor $\frac{1}{n}$ (note that this heuristics coincides by the way with the "wrong" analysis of page 1, saying that $\frac{1}{n} H^{(n)} (H^{(n)})^*$ converges entrywise to I).

In the case where H is a rectangular $n \times m$ matrix, where $\frac{m}{n} = c$ is a fixed parameter, the same heuristics shows that we should scale down the matrix $A^{(n)}$ by a factor $\frac{1}{n}$.

b) GUE

Let $A^{(n)} = H^{(n)}$ (n×n matrix) have independent entries $h_{jk}^{(n)}$, $j \le k$, with $h_{jj}^{(n)} \sim N_{\mathbb{R}}(0,1)$, $h_{jk}^{(n)} \sim N_{\mathbb{C}}(0,1)$ and $h_{kj}^{(n)} = \overline{h_{jk}^{(n)}}$

First moment of $p^{(n)}(\lambda)$:

$$\int_{\mathbb{R}} \lambda\, p^{(n)}(\lambda)\, d\lambda = \mathbb{E}(\lambda^{(n)}) = \mathbb{E}\Big(\frac{1}{n}\sum_{j=1}^{n} \lambda_j^{(n)}\Big) = \mathbb{E}\Big(\frac{1}{n} \operatorname{Tr} A^{(n)}\Big)$$

$$= \frac{1}{n}\sum_{j=1}^{n} \underbrace{\mathbb{E}(h_{jj}^{(n)})}_{0} = 0$$

So far, it therefore seems that no rescaling is needed, but...

Second moment of $p^{(n)}(\lambda)$:

$$\int_{\mathbb{R}} \lambda^2 p^{(n)}(\lambda)\,d\lambda = \mathbb{E}\left((\lambda^{(n)})^2\right) = \mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{n}(\lambda_j^{(n)})^2\right)$$

$$= \mathbb{E}\Big(\frac{1}{n}\,\mathrm{Tr}\underbrace{\left((A^{(n)})^2\right)}_{=A^{(n)}(A^{(n)})^*}\Big) = \frac{1}{n}\sum_{j,k=1}^{n}\underbrace{\mathbb{E}\left(|h_{jk}^{(n)}|^2\right)}_{=1} = n$$

According to this heuristics, we therefore need to scale down the matrix $A^{(n)}$ by a factor $\frac{1}{\sqrt{n}}$.

c) CUE

Let $\underset{(n\times n \text{ matrix})}{A^{(n)}}$ be distributed according to the Haar dist. on $U(n)$.

In this case, we have:

$$\int_0^{2\pi} |e^{i\theta}|^2 p^{(n)}(\theta)\,d\theta \overset{(*)}{=} \mathbb{E}\left(|\lambda^{(n)}|^2\right) = \mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{n}|\lambda_j^{(n)}|^2\right)$$

$$= \mathbb{E}\left(\frac{1}{n}\,\mathrm{Tr}\left(A^{(n)}(A^{(n)})^*\right)\right) = \mathbb{E}\left(\frac{1}{n}\,\mathrm{Tr}\, I\right) = 1$$

Therefore, no rescaling is needed. Moreover, $p^{(n)}(\theta) = \frac{1}{2\pi}$ is the uniform distribution on $[0, 2\pi)$ for all n, so the asymptotic analysis is particularly simple!

(*) NB: it is clear that this integral is equal to 1, since $|e^{i\theta}| = 1$ & $p^{(n)}(\theta)$ is a p.d.f.!

Aside: asymptotic analysis of the second order marginal distribution in the CUE

We have: $p^{(n)}(\theta,\varphi) = \frac{1}{4\pi^2}\cdot\frac{n}{n-1}\left(1-\left(\frac{\sin\left(\frac{n(\theta-\varphi)}{2}\right)}{n\sin\left(\frac{\theta-\varphi}{2}\right)}\right)^2\right)$

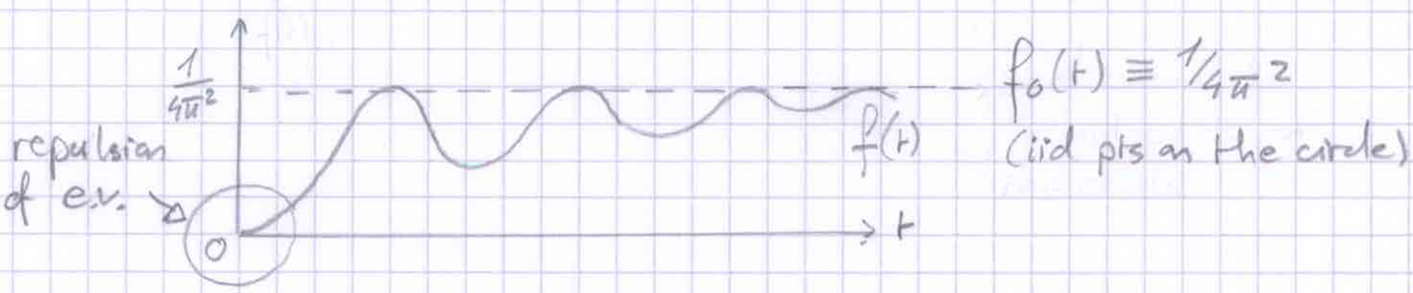
Since there are n eigenvalues on the unit circle of (fixed) length 2π, the natural scaling for the spacings of eigenvalues is of order $\frac{1}{n}$. Let us therefore define $\theta = \varphi + \frac{t}{n}$:

$$p^{(n)}(\varphi+\tfrac{t}{n},\varphi) = \frac{1}{4\pi^2}\cdot\frac{n}{n-1}\left(1-\frac{(\sin(t/2))^2}{(n\sin(t/2n))^2}\right)$$

$$\underset{n\to\infty}{\sim} \frac{1}{4\pi^2}\left(1-\frac{(\sin(t/2))^2}{(t/2)^2}\right) := f(t)$$

"two-point correlation function"

since $n\sin(x/n)\to x$ as $n\to\infty$.



A similar behaviour near zero can be obtained for the probability that two neighbouring ev. are separated by a distance t/n (but the analysis is more difficult).

Back to the first order marginal, in the GUE:

We had computed, for $A^{(n)} = H^{(n)}$:

$$p^{(n)}(\lambda) = \sum_{l=0}^{n-1} H_l(\lambda)^2 \, e^{-\lambda^2/2} \quad \text{with } H_l = \text{Hermite polynomials}$$

Considering now the rescaled matrices $\tilde{A}^{(n)} = \frac{1}{\sqrt{n}} H^{(n)}$, we obtain:

$$\tilde{p}^{(n)}(\lambda) = \sqrt{n}\, p^{(n)}(\sqrt{n}\lambda) = \frac{1}{\sqrt{n}} \sum_{l=0}^{n-1} H_l(\sqrt{n}\lambda)^2 \, e^{-n\lambda^2/2}$$

Analyzing this expression further requires a good knowledge of Hermite polynomials! By the Christoffel-Darboux formula, we obtain:

$$\tilde{p}^{(n)}(\lambda) = \left(\sqrt{n}\, H_n(\sqrt{n}\lambda)^2 - \sqrt{n+1}\, H_{n-1}(\sqrt{n}\lambda) H_{n+1}(\sqrt{n}\lambda)\right) e^{-n\lambda^2/2}$$

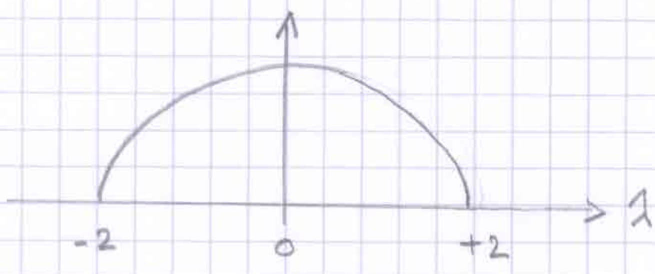
Next, the Plancherel-Rotach formula gives:

$$(2n)^{1/4} H_n(\sqrt{n}\lambda)\, e^{-n\lambda^2/4} = \text{some } f(\lambda) + O\left(\tfrac{1}{n}\right)$$

... we "therefore" have:

$$\lim_{n\to\infty} \tilde{p}^{(n)}(\lambda) = \frac{1}{2\pi}\sqrt{4-\lambda^2} \cdot 1_{|\lambda|\leq 2}$$

Wigner's semi-circle law



Remarks

- The above "proof" is quite technical.
- It generalizes to other random matrix ensembles, but requires different formulas for different orthogonal polynomials
- It relies heavily on the fact that both $p^{(n)}(\lambda_1 \ldots \lambda_n)$ and the marginal $p^{(n)}(\lambda)$ are computable at finite n.

We will see that much more powerful techniques allow to strenghten the result in two main directions:

- the result can be shown for more general random matrix ensembles (for which $p^{(n)}(\lambda)$ is unknown at finite n)
- the convergence of $p^{(n)}(\lambda)$ to a limit as $n \to \infty$ is a convergence "in expectation" (*); it can be shown that an "almost sure" convergence also holds for random matrices.

$$(*) \quad \mathbb{E}\left(\frac{1}{n}\sum_{j=1}^{n} f(\lambda_j^{(n)})\right) = \mathbb{E}\left(f(\lambda^{(n)})\right)$$

$$= \int_{\mathbb{R}} f(\lambda)\, p^{(n)}(\lambda)\, d\lambda \xrightarrow[n\to\infty]{} \int_{\mathbb{R}} f(\lambda)\, p(\lambda)\, d\lambda$$

Random matrix theory: lecture 10

Asymptotic analysis of deterministic (Toeplitz) matrices

[ref: Bob Gray's report]

Circulant matrices

$$C = \begin{pmatrix} c_0 & c_1 & \cdots & c_{n-1} \\ c_{n-1} & c_0 & \cdots & c_{n-2} \\ & & \ddots & \\ c_1 & \cdots & c_{n-1} & c_0 \end{pmatrix}$$ cyclic shifts to the right

$n \times n$ matrix, $c_0 \dots c_{n-1} \in \mathbb{C}$

notation: $C = \mathrm{circ}(c_0, c_1, \dots, c_{n-1})$

Lemma

Let $\alpha \in \mathbb{C}$ be such that $\alpha^n = 1$ (n^{th} root of unity).

Then $u = \begin{pmatrix} \alpha \\ \alpha^2 \\ \vdots \\ \alpha^{n-1} \\ 1 \end{pmatrix}$ is an eigenvector of C

with corresponding eigenvalue $\lambda = \sum_{\ell=0}^{n-1} c_\ell \alpha^\ell$

Proof

One has to check that $Cu = \lambda u$:

$$(Cu)_j = \sum_{k=1}^{n} c_{jk} u_k = \sum_{k=1}^{j-1} c_{n-j+k} \alpha^k + \sum_{k=j}^{n} c_{k-j} \alpha^k$$

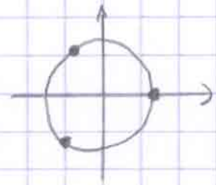
$\ell = n-j+k$ $\qquad$ $\ell = k-j$

$$= \sum_{\ell=n-j+1}^{n-1} c_\ell \underbrace{\alpha^{\ell+j-n}}_{=\alpha^{\ell+j} \text{ since } \alpha^{-n}=1} + \sum_{\ell=0}^{n-j} c_\ell \alpha^{\ell+j}$$

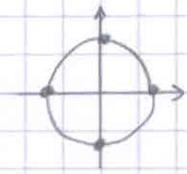
$$= \sum_{\ell=0}^{n-1} c_\ell \alpha^{\ell+j} = \left(\sum_{\ell=0}^{n-1} c_\ell \alpha^\ell\right) \alpha^j = \lambda u_j \quad \#$$

There are n different α's such that $\alpha^n = 1$:

$$\alpha_k = \exp\left(\frac{2\pi i k}{n}\right) \quad k = 1..n$$



$n=3$



$n=4$

and it turns out that the n different eigenvectors $u_1..u_n$ generated by $\alpha_1..\alpha_n$ are orthogonal.

$\Rightarrow$ Proposition: there exist U unitary and $\Lambda = \mathrm{diag}(\lambda_1..\lambda_n)$ such that $C = U \Lambda U^*$, where

$$\begin{cases} \lambda_k = \sum_{\ell=0}^{n-1} c_\ell (\alpha_k)^\ell = \sum_{\ell=0}^{n-1} c_\ell \exp\left(\frac{2\pi i k \ell}{n}\right) \\ U_{jk} = \frac{1}{\sqrt{n}} (\alpha_k)^j = \frac{1}{\sqrt{n}} \exp\left(\frac{2\pi i j k}{n}\right) \qquad \text{[DFT matrix]} \end{cases}$$

Consequences

- All circulant matrices share the same set of eigenvectors! Only the eigenvalues depend (linearly!) on the values of $c_1..c_n$!
- If C is a circulant matrix, then:
  - C^* is circulant; $CC^* = C^*C$, i.e. C is normal
  - if C is invertible, then C^{-1} is circulant
- If C_1, C_2 are circulant with eigenvalues $\lambda_k^{(1)}, \lambda_k^{(2)}$, and $\alpha, \beta \in \mathbb{C}$:
  - $\alpha C_1 + \beta C_2$ is circulant, with eigenvalues $\lambda_k = \alpha \lambda_k^{(1)} + \beta \lambda_k^{(2)}$
  - $C_1 C_2$ is circulant, with eigenvalues $\lambda_k = \lambda_k^{(1)} \lambda_k^{(2)}$

(Finite-order) Toeplitz matrices

$$T^{(n)} = \begin{pmatrix} t_0 & t_1 \cdots t_{\ell_0} & & 0 \\ t_{-1} & \ddots & \ddots & t_{\ell_0} \\ t_{-\ell_0} & \ddots & \ddots & \vdots \\ & & \ddots & t_1 \\ 0 & t_{-\ell_0} & \cdots t_{-1} & t_0 \end{pmatrix} \quad n \times n \text{ matrix}, \quad \left(T^{(n)}\right)_{jk} = t_{k-j} \in \mathbb{C}$$

and $t_\ell = 0$ if $|\ell| > \ell_0$.

Let $\lambda_1^{(n)} \ldots \lambda_n^{(n)}$ be the eigenvalues of $T^{(n)}$; we are interested in the asymptotic behaviour of these eigenvalues as $n \to \infty$.

First remark: Contrary to circulant matrices, there is no general expression at finite n for the eigenvalues of $T^{(n)}$ in terms of the numbers t_ℓ; and the DFT matrix is not the matrix of eigenvectors of $T^{(n)}$.

Let us define $g(x) := \sum_{\ell=-\ell_0}^{\ell_0} t_\ell e^{i\ell x}$, $x \in [0, 2\pi]$

g is complex-valued, bounded and continuous,

t_ℓ are the Fourier coefficients of g: $t_\ell = \frac{1}{2\pi} \int_0^{2\pi} g(x) e^{-i\ell x} dx$

Lemma 1 (connection between the λ's and g)

For any $m \geq 0$, we have

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} \left(\lambda_k^{(n)}\right)^m = \frac{1}{2\pi} \int_0^{2\pi} (g(x))^m dx$$

<u>Proof idea</u> (details are left to homework :))

- consider the case $\ell_0 = 1$ for simplicity; the matrices

$$T^{(n)} = \begin{pmatrix} t_0 & t_1 & & 0 \\ t_{-1} & \ddots & \ddots & \\ & \ddots & \ddots & t_1 \\ 0 & & t_{-1} & t_0 \end{pmatrix} \quad \text{and} \quad C^{(n)} = \begin{pmatrix} t_0 & t_1 & 0 & t_{-1} \\ t_{-1} & \ddots & \ddots & \\ & \ddots & \ddots & t_1 \\ t_1 & 0 & t_{-1} & t_0 \end{pmatrix}$$

with e.v. $\lambda_k^{(n)}$ with e.v. $\mu_k^{(n)}$

can be shown to be "asymptotically equivalent," which implies that $\lim\limits_{n\to\infty} \frac{1}{n}\sum\limits_{k=1}^{n} \left(\lambda_k^{(n)}\right)^m = \lim\limits_{n\to\infty} \frac{1}{n}\sum\limits_{k=1}^{n} \left(\mu_k^{(n)}\right)^m$ $\forall m \geq 0$

- note that $C^{(n)}$ is circulant, so

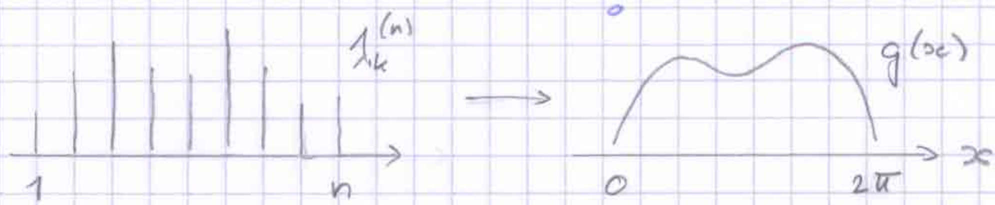
$$\mu_k^{(n)} = \sum_{\ell=-\ell_0}^{\ell_0} t_\ell \exp\left(\frac{2\pi i k \ell}{n}\right) = g\left(\frac{2\pi k}{n}\right)$$

(for $n \geq 2\ell_0 + 1$)

- Therefore, $\lim\limits_{n\to\infty} \frac{1}{n}\sum\limits_{k=1}^{n} \left(\lambda_k^{(n)}\right)^m = \lim\limits_{n\to\infty} \frac{1}{n}\sum\limits_{k=1}^{n} g\left(\frac{2\pi k}{n}\right)$

$$\text{(Riemann sums)} \quad = \frac{1}{2\pi}\int_0^{2\pi} \left(g(x)\right)^m dx$$

illustration for $m=1$:



#

<u>NB</u>: The above result actually says that

$$\lim_{n\to\infty} \frac{1}{n} \operatorname{Tr}\left(\left(T^{(n)}\right)^m\right) = \frac{1}{2\pi}\int_0^{2\pi} \left(g(x)\right)^m dx \quad \forall m \geq 1$$

<u>Assumption H</u>: $t_{-\ell} = \overline{t_\ell}$ for all $|\ell| \le \ell_0$

× Under assumption H: • $T^{(n)}$ is Hermitian, so $\lambda_1^{(n)} .. \lambda_n^{(n)} \in \mathbb{R}$, $\forall n$.

× • g is real-valued

<u>Lemma 2</u> (proof → homework again ☺)

Under assumption H, $a \le \lambda_k^{(n)} \le b \quad \forall 1 \le k \le n$

where $a := \inf_{x\in[0,2\pi]} g(x) \le \sup_{x\in[0,2\pi]} g(x) =: b$

<u>Theorem</u> (Grenander - Szegö 1958, Gray 1972)

Under assumption H, we have for any continuous function $f : [a,b] \longrightarrow \mathbb{R}$:

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(\underbrace{\lambda_k^{(n)}}_{\in [a,b] \text{ by lemma 2}}) = \frac{1}{2\pi} \int_0^{2\pi} f(\underbrace{g(x)}_{\in [a,b] \text{ by def.}})\, dx$$

<u>Proof</u>

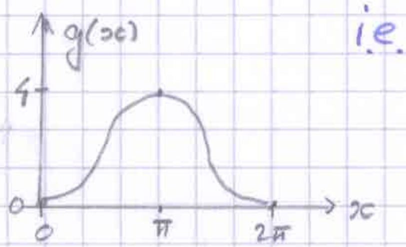
- lemma 1 proves the thm for f of the form $f(y) = y^m$, $y \in [a,b]$.
- by linearity of the sum & integral, the theorem also holds for any f of the form $f(y) = \sum_{m=0}^{m_0} c_m y^m$, i.e. any polynomial.
- × by Weierstrass theorem, any continuous function f on $[a,b]$ may be approximated uniformly by a sequence of polynomials, so the thm extends to continuous functions. #

<u>Example</u>: let $t_0 = 2$, $t_1 = t_{-1} = -1$, $t_\ell = 0 \quad \forall |\ell| > 1$

$$T^{(n)} = \begin{pmatrix} 2 & -1 & & 0 \\ -1 & \ddots & \ddots & \\ & \ddots & \ddots & -1 \\ 0 & & -1 & 2 \end{pmatrix}, \quad g(x) = 2 - e^{ix} - e^{-ix} = 2(1-\cos x)$$

i.e. $a = 0$, $b = 4$

limiting eigenvalue profile



Important remark

Without assumption H, the theorem fails!

(Counter)-example: let $t_0 = 1$, $t_1 = -1$, $t_\ell = 0$ otherwise

$$T^{(n)} = \begin{pmatrix} 1 & -1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & -1 \\ 0 & & & 1 \end{pmatrix}, \quad g(x) = 1 - e^{ix} \in \mathbb{C},$$

but all eigenvalues of $T^{(n)}$ are equal to 1!

it still holds that $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} (\lambda_k)^m = 1 = \frac{1}{2\pi} \int_0^{2\pi} (1-e^{ix})^m \, dx \quad \forall m \geq 0$

but $\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} f(\lambda_k) = f(1) \neq \frac{1}{2\pi} \int_0^{2\pi} f(1-e^{ix}) \, dx$

for any continuous $f: \mathbb{C} \to \mathbb{C}$

Take home message:

When dealing with sequences of non-Hermitian matrices $A^{(n)}$, knowing $\lim_{n\to\infty} \frac{1}{n} \mathrm{Tr}\left((A^{(n)})^m\right) \quad \forall m \geq 0$ is not sufficient to determine the asymptotic behaviour of eigenvalues.

Another perspective on the Grenander-Szegö theorem

- By the change of variable $y = g(x)$, we have

$$\frac{1}{2\pi}\int_0^{2\pi} f(g(x))\,dx = \int_a^b f(y)\,p(y)\,dy \quad \text{for same } p(y)$$

[NB: this works only if g is (piecewise) 1-to-1]

Choosing $f(y) \equiv 1$, we get $\frac{1}{2\pi}\int_0^{2\pi} 1\,dx = 1 = \int_a^b p(y)\,dy$

and $\int_a^b f(y)p(y)\,dy \geq 0$ for any $f(y) \geq 0$, so $p(y) \geq 0$,

i.e. $p(y)$ is the density of a (probability) distribution μ on $\mathbb{R}$.

notation: $\int_a^b f(y)\,p(y)\,dy = \int_a^b f(y)\,d\mu(y)$ $\quad \left(\frac{d\mu}{dy} = p(y)\right)$

- Recall now Dirac's δ-distribution on $\mathbb{R}$:

for $c \in \mathbb{R}$, δ_c is the distribution s.t. $\int_{\mathbb{R}} f(y)\,d\delta_c(y) = f(c)$ $\forall f$

The empirical eigenvalue distribution of $T^{(n)}$ is defined as:

$$\mu_n := \frac{1}{n}\sum_{k=1}^{n} \delta_{\lambda_k^{(n)}} \quad \text{(supported on } [a,b] \text{ by lemma 2)}$$

i.e. $\int_a^b f(y)\,d\mu_n(y) = \frac{1}{n}\sum_{k=1}^{n} f(\lambda_k^{(n)})$ $\quad \forall f : [a,b] \to \mathbb{R}$

- What Grenanger-Szegö's thm says is therefore:

$$\lim_{n\to\infty} \int_a^b f(x)\,d\mu_n(x) = \int_a^b f(y)\,d\mu(y) \quad \forall f : [a,b] \to \mathbb{R} \text{ continuous}$$

i.e. the sequence of distributions $(\mu_n)_{n\geq 1}$ converges weakly to the distribution μ as $n \to \infty$.

<u>Example</u>: $t_0 = 2,\ t_1 = t_{-1} = -1,\ t_\ell = 0 \quad \forall |\ell| > 1$

$\Rightarrow g(x) = 2(1-\cos x) = 4\sin\left(\frac{x}{2}\right)^2$ limiting eigenvalue profile

$$\frac{1}{n}\sum_{k=1}^{n} f(\lambda_k^{(n)}) \xrightarrow[n\to\infty]{} \frac{1}{2\pi}\int_0^{2\pi} f\left(4\sin\left(\tfrac{x}{2}\right)^2\right) dx$$

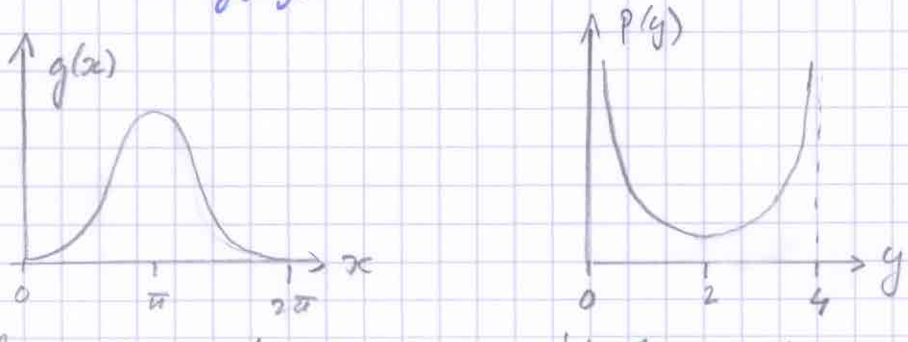
$$= \frac{1}{\pi}\int_0^{\pi} f(\underbrace{4\sin\left(\tfrac{x}{2}\right)^2}_{=y})\, dx$$

$$dy = 4\sin\left(\tfrac{x}{2}\right)\cos\left(\tfrac{x}{2}\right)\tfrac{1}{2}\, dx .$$

$$\Rightarrow dx = \frac{2}{\sqrt{y(4-y)}}\, dy \quad , \quad x=0 \leftrightarrow y=0,\ x=\pi \leftrightarrow y=4$$

$$\frac{1}{n}\sum_{k=1}^{n} f(\lambda_k^{(n)}) \xrightarrow[n\to\infty]{} \int_0^4 f(y)\,\frac{2}{\pi\sqrt{y(4-y)}}\, dy$$

$p(y) = \frac{2}{\pi\sqrt{y(4-y)}} \cdot 1_{[0,4]}(y)$ limiting eigenvalue distribution



limiting eigenvalue profile

limiting eigenvalue distribution

<u>NB</u>: it is therefore possible to talk about "eigenvalue distribution" even for deterministic matrices!

Further generalizations of the theorem

1. If the sequence $(t_\ell, \ell \in \mathbb{Z})$ is not of finite order but satisfies still $\sum_{\ell \in \mathbb{Z}} |t_\ell| < \infty$, then a proof similar to the preceding leads to the same result (but the approximation of $T^{(n)}$ by a circulant matrix is more involved, since every entry of $T^{(n)}$ is possibly non-zero).
   ref: Gray's web report
2. If only the weaker condition $\sum_{\ell \in \mathbb{Z}} |t_\ell|^2 < \infty$ is satisfied, then the proof gets even more involved, but the result still holds true (ref: Grenander - Szegö).
3. The following is not exactly a generalization, but rather a rephrasing of the theorem:
   Let $f(y) = 1_{y \leq t}$ for some fixed $t \in [a,b]$.
   f is not continuous on $[a,b]$, but can be approximated by continuous functions on $[a,b]$, so it can be shown that the theorem still holds true for such f, i.e. that
   $$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} 1_{\lambda_k^{(n)} \leq t} = \frac{1}{2\pi} \int_0^{2\pi} 1_{g(x) \leq t} \, dx \quad \forall t \in [a,b]$$

Let us define $F_n(t) := \frac{1}{n} \sum_{k=1}^{n} 1_{\lambda_k^{(n)} \leq t}$

and $F(t) := \frac{1}{2\pi} \int_0^{2\pi} 1_{g(x) \leq t} \, dx$. $\qquad t \in [a,b]$

Note that

- $F_n(t) = \frac{1}{n} \# \{ k : \lambda_k^{(n)} \leq t \}$ is the proportion of eigenvalues of $T^{(n)}$ less than or equal to t
- by the same change of variable as above ($y = g(x)$)

  $F(t) = \int_a^t p(y) \, dy$

Both F_n and F are therefore cumulative distribution functions$^{(*)}$, and the theorem says that

$$F_n(t) \xrightarrow[n \to \infty]{} F(t) \qquad \forall t \in [a,b]$$

which is a second characterization of the weak convergence of the corresponding sequence of distributions.

(*) i.e. $F_{(n)}(a) = 0$, $F_{(n)}(b) = 1$, $F_{(n)}$ is non-decreasing on $[a,b]$

(and $F_{(n)}$ is a right-continuous function on $[a,b]$)

Random matrix theory : lecture 11

Distributions without random variables !

Def: a (probability) distribution on $\mathbb{R}$ is an application

$\mu : \mathcal{B}(\mathbb{R}) \longrightarrow [0,1]$ such that:

[:= {Borel subsets of $\mathbb{R}$} := smallest σ-field containing all open subsets of $\mathbb{R}$]

$$\begin{cases} \cdot\ \mu(\emptyset) = \emptyset, \quad \mu(\mathbb{R}) = 1 \\ \cdot\ \text{if } B_n \cap B_m = \emptyset \ \forall n \neq m, \text{ then } \mu\left(\dot{\bigcup}_{n=1}^{\infty} B_n\right) = \sum_{n=1}^{\infty} \mu(B_n) \end{cases}$$

Remark: a distribution might therefore exist independently of any underlying random variable !

Def: the cumulative distribution function (cdf) associated to a distribution μ is the application $F_\mu : \mathbb{R} \to [0,1]$ defined as $F_\mu(t) := \mu(]-\infty, t])$, $t \in \mathbb{R}$.

Properties:

- $\lim_{t \to \infty} F_\mu(t) = \mu(\mathbb{R}) = 1$, $\lim_{t \to -\infty} F_\mu(t) = \mu(\emptyset) = 0$
- F_μ is non-decreasing: $t_1 \leq t_2 \Rightarrow F_\mu(t_1) \leq F_\mu(t_2)$
- F_μ is right-continuous: $\lim_{\varepsilon \downarrow 0} F_\mu(t+\varepsilon) = F_\mu(t) \quad \forall t \in \mathbb{R}$
- the knowledge of F_μ characterizes μ entirely !
  (and reciprocally, of course)

Two canonical classes of distributions

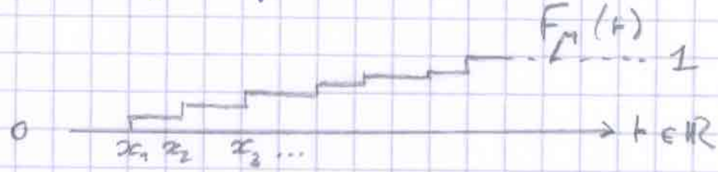
A) discrete distributions:

$\exists\ D = \{x_n\}_{n=1}^{\infty}$ (countable subset) such that $\mu(D) = 1$

(i.e. all the weight of the distribution μ is on D)

Let $p_n = \mu(\{x_n\})$: $\mu(B) = \sum_{x_n \in B} p_n$, $B \in \mathcal{B}(\mathbb{R})$

and $F_\mu(t) = \sum_{x_n \leq t} p_n$ step function



B) continuous distributions:

$\mu(B) = 0$ if $|B| = 0$ (in particular: $\mu(\{x\}) = 0 \quad \forall x \in \mathbb{R}$)

("length" of B)

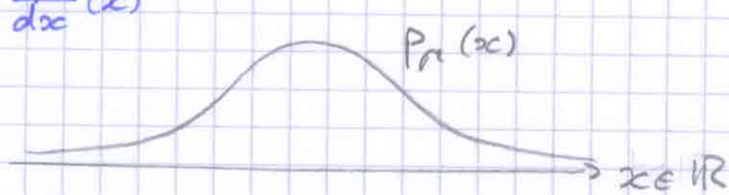
$\Rightarrow$ There exists a function p_μ such that $p_\mu(x) \geq 0$,

$\int_{\mathbb{R}} p_\mu(x)\,dx = 1$ and $\mu(B) = \int_B p_\mu(x)\,dx$, $B \in \mathcal{B}(\mathbb{R})$

p_μ is the probability density function (pdf) of μ

Moreover, $F_\mu(t) = \int_{-\infty}^{t} p_\mu(x)\,dx$ smooth function

× $\underline{\underline{NB}}$: $p_\mu(x) = F_\mu'(x) = \frac{d\mu}{dx}(x)$



<u>Riemann-Stieltjes integral</u> with respect to a distribution μ

<u>Def</u>: a function $f : \mathbb{R} \to \mathbb{R}$ is continuous if $\forall a < b$,

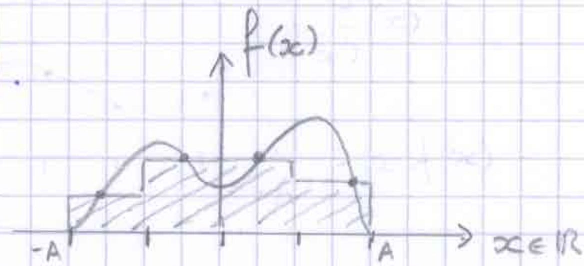
$$\{x \in \mathbb{R} : a < f(x) < b\} \text{ is an open subset of } \mathbb{R}$$

Let f be a continuous function on $\mathbb{R}$ such that $f(x) = 0$ $\forall |x| > A$ (fixed) and μ be a (general) distribution on $\mathbb{R}$.

Let $-A = a_0 < a_1 < \ldots < a_n = A$ be a subdivision of $[-A, A]$.

and $I_n := \sum_{j=1}^{n} f(\xi_j)\, \mu(]a_{j-1}, a_j])$.

where ξ_j is any point in $]a_{j-1}, a_j]$.



<u>Thm</u>:

For any continuous function f vanishing outside $[-A, A]$ and any sequence of subsivisions such that $\max_{1 \le j \le n} |a_j - a_{j-1}| \xrightarrow[n \to \infty]{} 0$ the sequence I_n converges to $I := \int_{\mathbb{R}} f(x)\, d\mu(x)$ as $n \to \infty$.

<u>Alternate notations</u>: • $I = \int_{\mathbb{R}} f(x)\, \mu(dx)$

• since $\mu(]a_{j-1}, a_j]) = F_\mu(a_j) - F_\mu(a_{j-1})$, one still writes

$$I = \int_{\mathbb{R}} f(x)\, dF_\mu(x) \quad \text{or} \quad I = \int_{\mathbb{R}} f(x)\, F_\mu(dx)$$

<u>Remark</u>:

The Riemann-Stieltjes integral can be extended to non-vanishing continuous functions on $\mathbb{R}$ (letting $A \to \infty$).

Lebesgue's integral with respect to a distribution μ

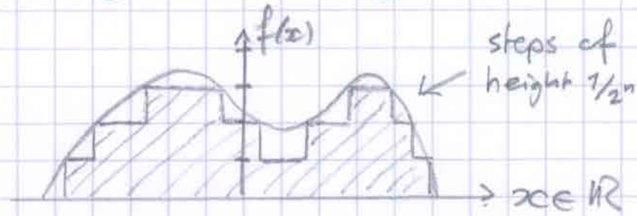
Def: a function $f : \mathbb{R} \to \mathbb{R}$ is Borel-measurable if $\forall a < b$

$$\{x \in \mathbb{R} : a < f(x) < b\} \text{ is a Borel subset of } \mathbb{R}$$

NB: this is a much weaker condition than being continuous!

- Let f be a non-negative Borel-measurable function on $\mathbb{R}$ (*)

and $I_n := \sum_{j=1}^{\infty} \frac{j-1}{2^n} \cdot \mu\left(\left\{x \in \mathbb{R} : \frac{j-1}{2^n} \le f(x) < \frac{j}{2^n}\right\}\right)$



Remarks: • for fixed n, $I_n \in [0, \infty]$

- $I_n \le I_{n+1}$ $\forall n$

(since the height of the steps is divided by 2 from n to $n+1$)

So $\lim_{n\to\infty} I_n = I$ exists and belongs to $[0, \infty]$.

I is the Lebesgue integral of f with respect to μ

and is denoted as $\int_{\mathbb{R}} f(x)\, d\mu(x)$. [same notation as Riemann-Stieltjes integral]

- Let f be a Borel-measurable function on $\mathbb{R}$ such that $\int_{\mathbb{R}} |f(x)|\, d\mu(x) < \infty$. Then

$$\int_{\mathbb{R}} f(x)\, d\mu(x) := \int_{\mathbb{R}} f^+(x)\, d\mu(x) - \int_{\mathbb{R}} f^-(x)\, d\mu(x)$$

where $f^+(x) := \max(0, f(x)) \ge 0$ and $f^-(x) := \max(0, -f(x)) \ge 0$.

(*) and μ be a (general) distribution on $\mathbb{R}$

Remarks:

- Both Riemann-Stieltjes and Lebesgue's integrals are well defined for $f: \mathbb{R} \to \mathbb{R}$ bounded and continuous (and they coincide for such f). Let us check this for Lebesgue: $\int_{\mathbb{R}} |f(x)|\, d\mu(x) \leq \underbrace{\sup_{x \in \mathbb{R}} |f(x)|}_{< \infty} \cdot \underbrace{\int_{\mathbb{R}} 1 \cdot d\mu(x)}_{= \mu(\mathbb{R}) = 1} < \infty$.

- Lebesgue's integral is more general (except for some particular cases), so we will always refer implicitly to this second definition.

Special cases of distributions μ:

A) Integral with respect to a discrete distribution μ:

$$\int_{\mathbb{R}} f(x)\, d\mu(x) = \sum_{n=1}^{\infty} f(x_n)\, p_n \qquad p_n = \mu(\{x_n\})$$

B) Integral with respect to a continuous distribution μ:

$$\int_{\mathbb{R}} f(x)\, d\mu(x) = \int_{\mathbb{R}} f(x)\, p_\mu(x)\, dx \qquad p_\mu = \frac{d\mu}{dx}$$

Special cases of functions f:

1) For a given $t \in \mathbb{R}$, let $f(x) = 1_{]-\infty,t]}(x)$.
   f is Borel-measurable and bounded and
   $\int_{\mathbb{R}} 1_{]-\infty,t]}(x)\, d\mu(x) = \mu(]-\infty,t]) = F_\mu(t)$, cdf of μ.

2) For a given $t \in \mathbb{R}$, let $f(x) = e^{itx}$; f is bounded and continuous and $\int_{\mathbb{R}} e^{itx}\, d\mu(x) = \phi_\mu(t)$,
   Fourier transform or characteristic function of μ.

3) For a given $k \geq 0$, let $f(x) = x^k$; f is continuous and unbounded; $\int_{\mathbb{R}} |f(x)|\, d\mu(x)$ is therefore not necessarily finite. When this is the case, we define
   $\int_{\mathbb{R}} x^k\, d\mu(x) = m_k$ moment of order k of μ.

4) For a given $z \in \mathbb{C} \setminus \mathbb{R}$, let $f(x) = \frac{1}{x-z}$; f is complex-valued, bounded and continuous (since $z \notin \mathbb{R}$)
   $\int_{\mathbb{R}} \frac{1}{x-z}\, d\mu(x) = g_\mu(z)$ Stieltjes transform of μ

Weak convergence of sequences of distributions

Def: A sequence of distributions $(\mu_n)_{n=1}^{\infty}$ converges weakly to a distribution μ if

$$\lim_{n\to\infty} F_{\mu_n}(t) = F_\mu(t)$$

$\forall t \in \mathbb{R}$ continuity point of F_μ

(Unfortunate) notation: $\mu_n \underset{n\to\infty}{\Longrightarrow} \mu$

Two equivalent definitions:

Def': $\mu_n \underset{n\to\infty}{\Longrightarrow} \mu$ iff $\lim_{n\to\infty} \mu_n(]a,b]) = \mu(]a,b])$

$\forall a<b$ such that $\mu(\{a\}) = \mu(\{b\}) = 0$

(Proof: use $\mu(]a,b]) = F_\mu(b) - F_\mu(a)$)

Def'': $\mu_n \underset{n\to\infty}{\Longrightarrow} \mu$ iff $\lim_{n\to\infty} \int_{\mathbb{R}} f(x)\, d\mu_n(x) = \int f(x)\, d\mu(x)$

$\forall f : \mathbb{R} \to \mathbb{R}$ bounded and continuous

(Proof: approximate $1_{]-\infty,t]}$ by a sequence of bdd and continuous functions; reciprocally, approximate f by a sequence of step functions)

Weak convergence and Fourier transform

Proposition (inversion formula)

The knowledge of the function $\phi_\mu(t) = \int_{\mathbb{R}} e^{itx} \, d\mu(x)$, $t \in \mathbb{R}$ characterizes μ entirely. Moreover, $\forall a < b$,

$$\mu(]a,b[) + \frac{1}{2}\mu(\{a,b\}) = \lim_{T \to \infty} \frac{1}{2\pi} \int_{-T}^{T} \frac{e^{-ita} - e^{-itb}}{it} \cdot \phi_\mu(t) \, dt.$$

If μ is a continuous distribution with pdf p_μ, then

$$p_\mu(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \phi_\mu(t) \, dt.$$

Proposition:

$$\mu_n \underset{n\to\infty}{\Rightarrow} \mu \quad \text{iff} \quad \lim_{n \to \infty} \phi_{\mu_n}(t) = \phi_\mu(t) \quad \forall t \in \mathbb{R}$$

Remarks:

- this proposition is of most importance in probability (it is used for proving the central limit theorem, e.g.)
- unfortunately, it is mostly useless for random matrices (explanation coming)

Random matrix theory: lecture 12

Weak convergence and moments

Recall: if for $k \geq 0$, $\int_{\mathbb{R}} |x|^k \, d\mu(x) < \infty$ for a given distribution μ,

$$\text{we set } m_k = \int_{\mathbb{R}} x^k \, d\mu(x)$$

Clearly, not all moments of a distribution are necessarily finite.

Example: let μ be the Cauchy distribution, with pdf

× $$p_\mu(x) = \frac{1}{\pi(1+x^2)} \quad x \in \mathbb{R} \text{ ; then } \int_{\mathbb{R}} |x|^k \, d\mu(x) = \infty \quad \forall k \geq 1$$

But even in the case where all moments are finite, the sequence $(m_k)_{k=0}^{\infty}$ does not necessarily characterize entirely the distribution μ.

Carleman's condition:

If $\sum_{k=0}^{\infty} (m_{2k})^{-\frac{1}{2k}} = \infty$, then μ is the only distribution with the sequence of moments $(m_k)_{k=0}^{\infty}$.

Remark: Carleman's condition is a condition on the growth of the moments m_k; it basically requires that $m_k \lesssim e^{k \log k}$.

Example: if $\mu([-c,c]) = 1$ for some $c > 0$, then

$$|m_k| \leq \int_{\mathbb{R}} |x|^k \, d\mu(x) = \int_{-c}^{c} |x|^k \, d\mu(x) \leq c^k \text{ satisfies the condition.}$$

Counter-example: the log-normal distribution μ with pdf

$$p_\mu(x) = \frac{1}{\sqrt{2\pi}} \frac{1}{x} \exp\left(-\frac{(\log x)^2}{2}\right), \ x > 0, \text{ has moments } m_k = e^{k^2/2}.$$

Proposition

Let $(\mu_n)_{n=1}^{\infty}$ be a sequence of distributions. If there exists a sequence of real numbers $(m_k)_{k=0}^{\infty}$ satisfying Carleman's condition and such that

$$\lim_{n\to\infty} \int_{\mathbb{R}} x^k \, d\mu_n(x) = m_k \quad \forall k \geq 0$$

then there exists a unique distribution μ with the sequence of moments $(m_k)_{k=0}^{\infty}$ such that $\mu_n \underset{n\to\infty}{\Longrightarrow} \mu$.

Remarks

- this proposition provides an easy criterion for checking weak convergence of sequences of distributions.
- drawback: in some cases, weak convergence holds even though not all moments are finite.

Weak convergence and Stieltjes transform

Recall: for $z \in \mathbb{C} \setminus \mathbb{R}$, $g_\mu(z) = \int_{\mathbb{R}} \frac{1}{x-z} \, d\mu(x)$

× Properties:

- g_μ is analytic on $\mathbb{C} \setminus \mathbb{R}$
- $\operatorname{Im} g_\mu(z) > 0$ for z such that $\operatorname{Im} z > 0$
- $\lim_{v\to\infty} v \, |g_\mu(iv)| = 1$

<u>Proposition</u> (inversion formula)

The knowledge of the function $g_\mu(z)$, $z \in \mathbb{C} \setminus \mathbb{R}$ characterizes μ entirely. Moreover, $\forall a < b$,

$$\mu(]a,b[) + \frac{1}{2}\mu(\{a,b\}) = \lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \int_a^b \mathrm{Im}(g_\mu(u+i\varepsilon))\, du$$

If μ is a continuous distribution with pdf p_μ, then

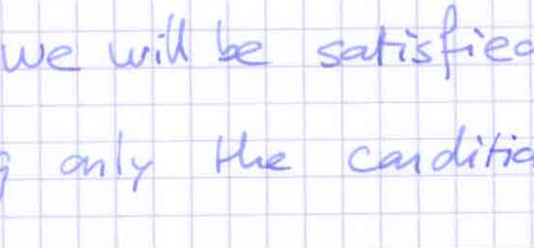
$$p_\mu(x) = \lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \mathrm{Im}(g_\mu(x+i\varepsilon))$$

<u>Proposition</u>

$$\mu_n \underset{n\to\infty}{\Longrightarrow} \mu \quad \text{iff} \quad \lim_{n\to\infty} \sup_{z \in K} |g_{\mu_n}(z) - g_\mu(z)| = 0$$

$$\forall K = [-A, A] \times [b, B] \subset \mathbb{C} \quad \begin{pmatrix} A > 0 \\ B > b > 0 \end{pmatrix}$$

"uniform convergence on compacts"



<u>Remark:</u> in practice, we will be satisfied with checking only the condition

$$\lim_{n\to\infty} g_{\mu_n}(z) = g_\mu(z) \quad \forall z \in \mathbb{C}_+ := \{z \in \mathbb{C} : \mathrm{Im}\, z > 0\}$$

Relation between all this and matrices ?

Let $(A^{(n)})_{n=1}^{\infty}$ be a sequence of Hermitian deterministic matrices of increasing size ($A^{(n)}$ is of size $n \times n$) and let $\lambda_1^{(n)} \dots \lambda_n^{(n)}$ be the (real) eigenvalues of $A^{(n)}$.

We define the following sequence of (discrete) cdf's:

$$F_n(t) := \frac{1}{n} \# \{ j : \lambda_j^{(n)} \leq t \} \quad t \in \mathbb{R}$$

and would like to study the (weak) convergence of the sequence $(F_n)_{n=1}^{\infty}$ as $n \to \infty$.

Corresponding distributions μ_n :

In Dirac's notation, $\mu_n = \frac{1}{n} \sum_{j=1}^{n} \delta_{\lambda_j^{(n)}}$

(ie μ_n = uniform distribution on the $\lambda_j^{(n)}$: $\mu_n(\{\lambda_j^{(n)}\}) = \frac{1}{n}$)

Note that $\mu_n(]a,b]) = \frac{1}{n} \# \{ j : a < \lambda_j^{(n)} \leq b \} = F_n(b) - F_n(a)$

Now, what is $\int_{\mathbb{R}} f(x)\, d\mu_n(x)$?

We already know that $A^{(n)} = U^{(n)} \Lambda^{(n)} (U^{(n)})^*$ where $U^{(n)}$ is unitary and $\Lambda^{(n)} = \mathrm{diag}(\lambda_1^{(n)}, \dots, \lambda_n^{(n)})$.

For $f: \mathbb{R} \to \mathbb{R}$, let us define $f(A^{(n)}) := U^{(n)} f(\Lambda^{(n)}) (U^{(n)})^*$, with $f(\Lambda^{(n)}) := \mathrm{diag}(f(\lambda_1^{(n)}), \dots, f(\lambda_n^{(n)}))$.

We have: $\frac{1}{n} \mathrm{Tr}(f(A^{(n)})) = \frac{1}{n} \mathrm{Tr}(f(\Lambda^{(n)})) = \frac{1}{n} \sum_{j=1}^{n} f(\lambda_j^{(n)})$

$$= \int_{\mathbb{R}} f(x)\, d\mu_n(x) \; !$$

Moments:

$$m_k^{(n)} = \int_{\mathbb{R}} x^k \, d\mu_n(x) = \frac{1}{n}\sum_{j=1}^{n} \left(\lambda_j^{(n)}\right)^k = \frac{1}{n}\,\mathrm{Tr}\left(\left(A^{(n)}\right)^k\right) \quad \forall k \geq 0$$

NB: for a given n, the $m_k^{(n)}$ are finite $\forall k \geq 0$

(but they might not converge in the limit $n \to \infty$)

Stieltjes transform:

$$g_n(z) = \int_{\mathbb{R}} \frac{1}{x-z}\, d\mu_n(x) = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{\lambda_j^{(n)} - z}$$

$$= \frac{1}{n}\,\mathrm{Tr}\Big(\big(\underbrace{A^{(n)} - zI_n}_{\text{invertible since } z \notin \mathbb{R}}\big)^{-1}\Big) \qquad z \in \mathbb{C}\setminus\mathbb{R}$$

Fourier transform:

$$\phi_n(t) = \int_{\mathbb{R}} e^{itx}\, d\mu_n(x) = \frac{1}{n}\sum_{j=1}^{n} e^{i\lambda_j^{(n)} t}$$

$$= \frac{1}{n}\,\mathrm{Tr}\left(e^{itA^{(n)}}\right) \qquad t \in \mathbb{R}$$

problem: In order to compute the exponential of a matrix, we need to compute either its eigenvalues or all the moments (using the formula: $e^{itA^{(n)}} = \sum_{m=0}^{\infty} \frac{(it)^m}{m!}\cdot\left(A^{(n)}\right)^m$). The Fourier transform is therefore not that useful.

Aside:

Note that $\frac{1}{n}\log\det A^{(n)} = \frac{1}{n}\sum_{j=1}^{n}\log\left(\lambda_j^{(n)}\right) = \frac{1}{n}\,\mathrm{Tr}\left(\log A^{(n)}\right)$

(assume $A^{(n)} > 0$) $\qquad = \int_{\mathbb{R}_+} \log(x)\, d\mu_n(x)$

Important remark

For a sequence of random matrices $A^{(n)}$, the eigenvalues $\lambda_1^{(n)} \ldots \lambda_n^{(n)}$ are random variables; the μ_n's are therefore random distributions, which means that $F_n(t)$, $m_k^{(n)}$, $g_n(z)$ and $\phi_n(t)$ are all random variables. We therefore need to specify in what sense the convergence takes place (i.e. writing $\lim_{n\to\infty} F_n(t) = F_\mu(t)$ is not enough).

Let $(X_n)_{n=1}^{\infty}$ be a generic sequence of random variables.

1) Convergence in expectation: $\lim_{n\to\infty} \mathbb{E}(X_n) = \mathbb{E}(X)$

2) Convergence in probability: (denoted as $X_n \xrightarrow[n\to\infty]{\mathbb{P}} X$)

$$\forall \varepsilon > 0, \quad \lim_{n\to\infty} \mathbb{P}(|X_n - X| > \varepsilon) = 0$$

3) Almost sure convergence: (denoted as $X_n \xrightarrow[n\to\infty]{} X$ a.s.)

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1$$

Note that 3) $\Longrightarrow$ 2) $\Longrightarrow$ 1)

(provided that the sequence (X_n) is "uniformly integrable")

Special case: convergence to a deterministic limit $X \equiv c$

1) Convergence in expectation simply means $\lim_{n\to\infty} \mathbb{E}(X_n) = c$

2) Proposition: if $\lim_{n\to\infty} \mathbb{E}(X_n) = c$ and $\lim_{n\to\infty} \mathrm{Var}(X_n) = 0$,

then $X_n \xrightarrow[n\to\infty]{\mathbb{P}} c$.

Proof: $\mathbb{P}(|X_n - c| > \varepsilon) \leq \dfrac{\mathbb{E}((X_n-c)^2)}{\varepsilon^2}$ (Markov inequality)

$$= \frac{1}{\varepsilon^2} \mathbb{E}\left(\left(X_n - \mathbb{E}(X_n) + \mathbb{E}(X_n) - c\right)^2\right)$$

$$\leq \frac{2}{\varepsilon^2} \Big\{ \underbrace{\mathbb{E}\left(\left(X_n - \mathbb{E}(X_n)\right)^2\right)}_{= \mathrm{Var}(X_n) \to 0} + \underbrace{\left(\mathbb{E}(X_n) - c\right)^2}_{\to 0} \Big\} \xrightarrow[n\to\infty]{} 0, \quad \forall \varepsilon > 0$$

#

3) Proposition: if $|\mathbb{E}(X_n) - c| = O(\frac{1}{n})$ and $\mathrm{Var}(X_n) = O(\frac{1}{n^2})$,

then $X_n \xrightarrow[n\to\infty]{} c$ a.s.

Proof: By the same sequence of inequalities as above,

we obtain that $\mathbb{P}(|X_n - c| > \varepsilon) = O(\frac{1}{n^2})$ $\forall \varepsilon > 0$,

so $\sum_{n=1}^{\infty} \mathbb{P}(|X_n - c| > \varepsilon) < \infty$ $\forall \varepsilon > 0$.

The Borel-Cantelli lemma thus implies that

$\mathbb{P}(|X_n - c| > \varepsilon$ infinitely often$) = 0$ $\forall \varepsilon > 0$

which is an equivalent condition for the

almost sure convergence of the sequence (X_n) towards c.

#

A famous example of (weak) convergence of a sequence of random distributions

- Let $(X_n)_{n=1}^{\infty}$ be a sequence of i.i.d. random variables with distribution μ (i.e. $\mathbb{P}(X_n \in B) = \mu(B)$ $\forall B \in \mathcal{B}(\mathbb{R})$, $\forall n \geq 1$)
- Let $F_n(t) := \frac{1}{n} \cdot \#\{j : X_j \leq t\}$, $t \in \mathbb{R}$. That is, F_n is the cdf of the empirical distribution $\mu_n = \frac{1}{n} \sum_{j=1}^{n} \delta_{X_j}$ of the first n random variables.
- Note that $F_n(t)$ is a random variable for a given n & t, so that μ_n is a random distribution for a given n.

Proposition: (convergence of the empirical distribution)

$$\lim_{n \to \infty} F_n(t) = F_\mu(t) \quad \text{a.s.}, \ \forall t \in \mathbb{R}$$

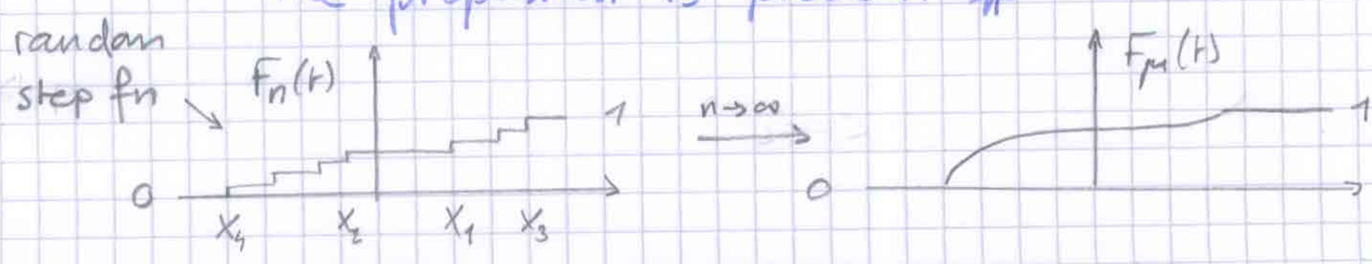
Proof: For a given $t \in \mathbb{R}$, let $Y_j := 1_{\{X_j \leq t\}}$ i.i.d. random var.

$F_n(t) = \frac{1}{n} \sum_{j=1}^{n} Y_j$, so by the law of large numbers,

$$\lim_{n \to \infty} F_n(t) = \mathbb{E}(Y_1) \quad \text{a.s.}$$

Since $\mathbb{E}(Y_1) = \mathbb{P}(X_1 \leq t) = F_\mu(t)$,

the proposition is proved. #



Random matrix theory: lecture 13

Wigner's theorem

Preliminary: the semi-circle distribution and the Catalan numbers

Let μ be the distribution on $\mathbb{R}$ with pdf

$$p_\mu(x) = \frac{1}{2\pi}\sqrt{4-x^2}\cdot 1_{|x|\leq 2} \qquad x \in \mathbb{R}$$

All moments of μ are finite and given by (homework)

$$m_{2k+1} = 0\,, \quad m_{2k} = \frac{(2k)!}{k!\,(k+1)!} = \frac{1}{k+1}\binom{2k}{k} := t_k \quad , k \geq 0 \qquad (*)$$

t_k are called the Catalan numbers; there are several combinatorial characterizations of these numbers. Let us mention two:

- t_k is the number of planar planted rooted trees with k branches

  illustration for $k=3$:

$$t_3 = \frac{6!}{3!\,4!} = 5$$

$(*)$ Note moreover that

$$t_k \leq \frac{(2k)!}{(k!)^2} \leq \frac{\big((2k)(2k-2)\cdots 2\big)^2}{(k!)^2} \leq \left(\frac{2^k k!}{k!}\right)^2 \leq 4^k$$

so the sequence (m_k) satisfies Carleman's condition.

<u>Def</u>: a Dyck path of length $2k$ is a path starting at zero and ending at zero in $2k$ steps, going up or down a each step, and conditioned to stay non-negative over the whole period.

illustration:

a Dyck path

- t_k is the number of Dyck paths of length $2k$

illustration for $k=3$: $t_3 = 5$ again

<u>Identification</u> between the two descriptions:

explore the tree from the root and go up and down along the branches; the record of these ups and downs is the corresponding Dyck path.

<u>Counting Dyck paths</u>: <u>the reflexion principle</u>

The number of Dyck paths of length $2k$ is the number of <u>all</u> paths from $(0,0)$ to $(2k,0)$ minus the number of paths crossing the -1 horizontal line.

Note that to each of these paths crossing the -1 line corresponds a symmetric path reaching $(2k,-2)$ at the end:

reflexion principle

Note moreover that <u>all</u> paths from $(0,0)$ to $(2k,-2)$ necessarily cross the horizontal line -1. Therefore, the number of Dyck paths of length $2k$ is:

$$\binom{2k}{k} - \binom{2k}{k+1} = \frac{(2k)!}{k!\cdot k!} - \frac{(2k)!}{(k+1)!(k-1)!}$$

\# all paths from 0 to 0 (k downs)

\# all paths from 0 to -2 ($k+1$ downs)

$$= \frac{(2k)!}{k!\,k!}\left(1-\frac{k}{k+1}\right) = \binom{2k}{k}\frac{1}{k+1}$$

$$= t_k \quad \checkmark$$

We have therefore shown that the Catalan numbers are indeed the number of Dyck paths of length $2k$, or equivalently, the number of planar planted rooted trees with k branches.

We consider now the following random matrix ensemble: let H be a $n\times n$ real symmetric matrix such that

(i) $\{h_{jk},\ j\leq k\}$ are i.i.d. random variables (and $h_{kj}:=h_{jk}$)

× (ii) all moments of h_{11} are finite

(iii) $\mathbb{E}(h_{11}^{2\ell+1})=0 \quad \forall \ell\geq 0$

(iv) $\mathbb{E}(h_{11}^2)=1$

Let moreover $H^{(n)}=\frac{1}{\sqrt{n}}H$ and $\lambda_1^{(n)}\dots\lambda_n^{(n)}$ be the eigenvalues of $H^{(n)}$.

Theorem (Wigner, 1955)

$$F_n(t):=\frac{1}{n}\#\{j:\lambda_j^{(n)}\leq t\}\xrightarrow[n\to\infty]{}\int_{-\infty}^{t}p_\mu(x)\,dx \quad \text{a.s. } \forall t\in\mathbb{R}$$

where

$$p_\mu(x):=\frac{1}{2\pi}\sqrt{4-x^2}\,1_{|x|\leq 2} \quad \text{semi-circle distribution}$$

Remarks

- examples of distributions of h_{11} satisfying the assumptions are: $h_{11}\sim N_{\mathbb{R}}(0,1)$, $h_{11}=\pm 1$ w.p. $\frac{1}{2}$, $h_{11}\sim \mathcal{U}([-\sqrt{3},\sqrt{3}])$, but the limit does not depend on this distribution; it actually only depends on its variance ($\mathbb{E}(h_{11}^2)$), so the result is universal (c.f. central limit theorem)

- the assumptions do not include the GOE, because of (i), that assumes iid diagonal and off-diagonal entries. It can actually be shown that the distribution of the diagonal entries does not influence the limit (as soon as these entries have bounded variance).
- there is another qualitative difference between the above theorem and the result that one obtains from finite-size analysis. The present theorem says that for any function $f : \mathbb{R} \to \mathbb{R}$ bounded and continuous

$$\frac{1}{n} \sum_{j=1}^{n} f(\lambda_j^{(n)}) \xrightarrow[n \to \infty]{} \underbrace{\int_{\mathbb{R}} f(x)\, p_\mu(x)\, dx}_{\text{NB: deterministic limit!}} \quad \text{a.s.}$$

(c.f. definition of weak convergence), whereas the result obtained from finite-size analysis only says that

$$\mathbb{E}\Big(\frac{1}{n} \sum_{j=1}^{n} f(\lambda_j^{(n)})\Big) = \int_{\mathbb{R}} f(x)\, p^{(n)}(x)\, dx$$
$$\xrightarrow[n \to \infty]{} \int_{\mathbb{R}} f(x)\, p_\mu(x)\, dx.$$

- in general, the iid assumption may be relaxed in various ways; the important fact is that entries are independent with the same variance.

Proof of the theorem

- For convenience, we will prove the theorem under the slightly stronger assumption:

× (ii)' h_{11} is a bounded random variable

(i.e. $\exists C > 0$ such that $|h_{11}| \le C$ a.s.)

- The technique used for proving weak convergence is via moments, i.e., we will show that $\forall \ell \ge 0$,

× $$m_\ell^{(n)} := \frac{1}{n} \sum_{j=1}^{n} \left(\lambda_j^{(n)}\right)^\ell = \frac{1}{n} \operatorname{Tr}\left(\left(H^{(n)}\right)^\ell\right) \xrightarrow[n\to\infty]{} m_\ell \text{ a.s.}$$

where m_ℓ are the moments of the semi-circle distribution.

- Given the criterion of the last lecture for a.s. convergence, it is sufficient to show that $\forall \ell \ge 0$,

1) $\left|\mathbb{E}(m_\ell^{(n)}) - m_\ell\right| = O\left(\frac{1}{n}\right)$
2) $\operatorname{Var}(m_\ell^{(n)}) = O\left(\frac{1}{n^2}\right)$

We will focus on the proof of (1) in the following, as the technique for proving (2) is similar.

- $\mathbb{E}(m_\ell^{(n)}) = \frac{1}{n}\,\mathbb{E}\left(\mathrm{Tr}\left((H^{(n)})^\ell\right)\right) = \frac{1}{n^{1+\ell/2}}\,\mathbb{E}\left(\mathrm{Tr}(H^\ell)\right)$

$$= \frac{1}{n^{1+\ell/2}} \sum_{j_1 \ldots j_\ell = 1}^{n} \mathbb{E}\left(h_{j_1 j_2} h_{j_2 j_3} \ldots h_{j_\ell j_1}\right)$$

- First note that $\mathbb{E}(m_{2\ell+1}^{(n)}) = 0 = m_{2\ell+1}$ $\forall \ell \geq 0$:

Indeed, $m_{2\ell+1} = 0$ since the semi-circle distribution is symmetric around zero.

Moreover, in the expectation of any of the products $\mathbb{E}(h_{j_1 j_2} h_{j_2 j_3} \ldots h_{j_{2\ell+1} j_1})$, at least one of the entries h_{jk} appears an odd number of times (we identify here h_{jk} and h_{kj}), so by the independence assumption and the assumption that $\mathbb{E}(h_{jk}^{2\ell+1}) = 0$ $\forall \ell \geq 0$, we obtain that the overall expectation is zero.

- Let us therefore focus our attention on

$$\mathbb{E}(m_{2\ell}^{(n)}) = \frac{1}{n^{1+\ell}} \sum_{j_1 \cdots j_{2\ell} = 1}^{n} \mathbb{E}\left(h_{j_1 j_2} h_{j_2 j_3} \ldots h_{j_{2\ell} j_1}\right)$$

We will see next time that $|\mathbb{E}(m_{2k}^{(n)}) - t_k| = O(\frac{1}{n})$, where t_k are the Catalan numbers ($= m_{2k}$). This will conclude the proof of (1).

Random matrix theory: lecture 14

Proof of Wigner's theorem (cont'd)

iid random variables ($\& \; h_{kj} = h_{jk}$)

Recall: $\mathbb{E}\left(m_{2\ell}^{(n)}\right) = \frac{1}{n^{1+\ell}} \sum_{j_1 \dots j_{2\ell}=1}^{n} \mathbb{E}\left(h_{j_1 j_2} h_{j_2 j_3} \dots h_{j_{2\ell} j_1}\right)$

We are going to show that for any $\ell \geq 0$,

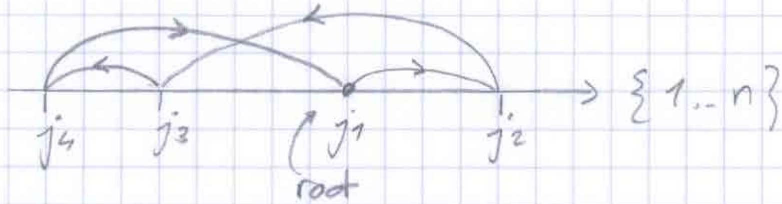
$$\left| \mathbb{E}\left(m_{2\ell}^{(n)}\right) - t_\ell \right| = O\left(\frac{1}{n}\right) \qquad (1)$$

where t_ℓ are the Catalan numbers.

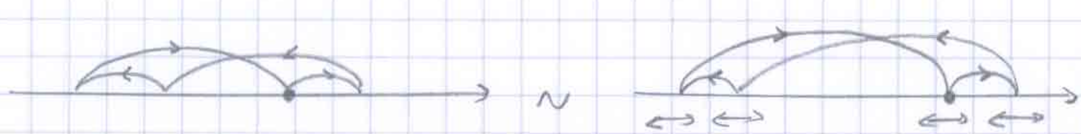
- notation: $J_{2\ell} := (j_1 \dots j_{2\ell})$, $h(J_{2\ell}) := h_{j_1 j_2} h_{j_2 j_3} \dots h_{j_{2\ell} j_1}$

$$\Rightarrow \mathbb{E}\left(m_{2\ell}^{(n)}\right) = \frac{1}{n^{1+\ell}} \sum_{J_{2\ell}} \mathbb{E}\left(h(J_{2\ell})\right)$$

To each sequence $J_{2\ell}$, associate a directed graph (*) $g(J_{2\ell})$:



We say that two sequences $J_{2\ell}$ and $J_{2\ell}'$ are equivalent ($J_{2\ell} \sim J_{2\ell}'$) if their corresponding graphs are the same:



notation: $g_{2\ell} := g(J_{2\ell}) = g(J_{2\ell}')$ or $J_{2\ell} \in g_{2\ell}$

(the graph is an equivalence class for the sequences)

(*) with labelled edges!
(12, 23, 34, etc ...)

- Because the h_{jk} are identically distributed, and the corresponding graphs are the same,

$$\mathbb{E}(h(\mathcal{J}_{2e})) = \mathbb{E}(h(\mathcal{J}_{2e}')) \quad \text{if } \mathcal{J}_{2e} \sim \mathcal{J}_{2e}'$$

So

$$\mathbb{E}(m_{2e}^{(n)}) = \frac{1}{n^{1+e}} \sum_{g_{2e}} \sum_{\mathcal{J}_{2e} \in g_{2e}} \underbrace{\mathbb{E}(h(\mathcal{J}_{2e}))}_{:= Q(g_{2e})}$$

$$= \frac{1}{n^{1+e}} \sum_{g_{2e}} (\#\{\mathcal{J}_{2e} \in g_{2e}\})\, Q(g_{2e})$$

- Let $V(g_{2e})$ be the set of vertices of g_{2e} and $|V(g_{2e})|$ be the number of such vertices.

We have:

$$\#\{\mathcal{J}_{2e} \in g_{2e}\} = \text{number of possibilities of placing } |V(g_{2e})| \text{ ordered points on } \{1..n\}$$

$$= n(n-1)\cdots(n-|V(g_{2e})|+1)$$

$$= n^{|V(g_{2e})|}\left(1+O\left(\tfrac{1}{n}\right)\right)$$

So

$$\mathbb{E}(m_{2e}^{(n)}) = \frac{1}{n^{1+e}} \sum_{g_{2e}} n^{|V(g_{2e})|}\, Q(g_{2e})\left(1+O\left(\tfrac{1}{n}\right)\right)$$

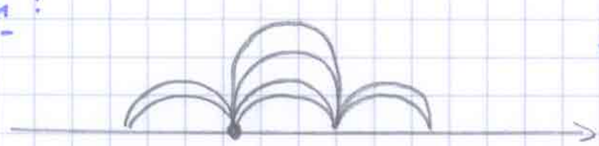
× • Notice that if an edge appears an odd number of times in the graph, then for the same reason as last time(*):

$$Q(g_{2e}) = \mathbb{E}(h(\mathcal{J}_{2e})) = \mathbb{E}(h_{j_1 j_2} \cdots h_{j_{2e} j_1}) = 0$$

(*) assumptions (i) & (iii)

Therefore, $Q(g_{2\ell}) > 0$ iff every edge in the graph appears an even number of times (the direction is indifferent, since $h_{kj} = h_{jk}$).

Illustration:



even graph

Terminology: the graph is said to be even in this case

- In an even graph, each edge appears at least twice, so each "new" vertex costs at least two edges, therefore $|V(g_{2\ell})| \leq 1+\ell$ (since # edges $= 2\ell$) and

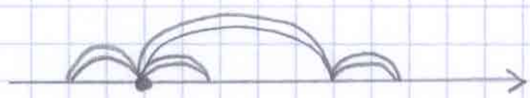
$$\mathbb{E}(m_{2\ell}^{(n)}) = \frac{1}{n^{1+\ell}} \sum_{\substack{g_{2\ell} \text{ even} \\ |V(g_{2\ell})| \leq 1+\ell}} n^{|V(g_{2\ell})|} Q(g_{2\ell})\left(1+O\left(\tfrac{1}{n}\right)\right)$$

- By assumption (ii)', $|Q(g_{2\ell})| \leq |\mathbb{E}(h_{j_1j_2} \cdots h_{j_{2\ell}j_1})| \leq C^{2\ell}$ independently of n, so the only graphs contributing in a non-negligible manner to the above sum are those for which $|V(g_{2\ell})| = 1+\ell$, i.e.

$$\mathbb{E}(m_{2\ell}^{(n)}) = \frac{1}{\cancel{n^{1+\ell}}} \sum_{\substack{g_{2\ell} \text{ even} \\ |V(g_{2\ell})| = 1+\ell}} \cancel{n^{1+\ell}}\, Q(g_{2\ell})\left(1+O\left(\tfrac{1}{n}\right)\right)$$

- Finally, for an even graph $g_{2\ell}$ such that $|V(g_{2\ell})| = 1+\ell$, each edge appears exactly twice.

Illustration:



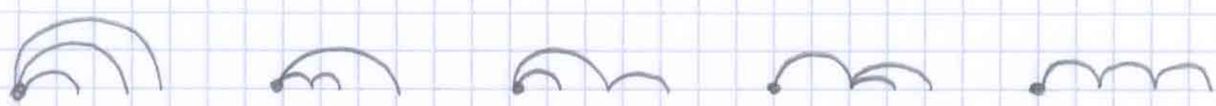
$(1+k=5,\ 2k=8)$

So $Q(g_{2\ell}) = \mathbb{E}(h_{j_1j_2}^2) \cdots \mathbb{E}(h_{j_kj_1}^2) = 1$ (by assumption (iv)).

and $\mathbb{E}(m_{2\ell}^{(n)}) = \#\{g_{2\ell} \text{ even} : |V(g_{2\ell})| = 1+\ell\} + O(\frac{1}{n})$

- How many even (and rooted) graphs are there with 2ℓ branches and $\ell+1$ vertices on the line?

Illustration: $\ell = 3$



unfold:

i.e. $\#\{g_{2\ell} \text{ even} : |V(g_{2\ell})| = 1+\ell\}$

$= \#\{\text{planar planted rooted trees with } \ell \text{ branches}\} = t_\ell$

So $|\mathbb{E}(m_{2\ell}^{(n)}) - t_\ell| = O(\frac{1}{n})$; this concludes the proof of (1). #

What about (2)$^{(*)}$: $Var(m_\ell^{(n)}) = O(\frac{1}{n^2})$?

First remark: this behaviour of the variance is atypical!

- Indeed, $m_\ell^{(n)} = \frac{1}{n} \sum_{j=1}^{n} (\lambda_j^{(n)})^\ell$.

  If the random variables $\lambda_j^{(n)}$ were iid,

  then we would have

  $$Var(m_\ell^{(n)}) = \frac{1}{n^2} \sum_{j=1}^{n} Var((\lambda_j^{(n)})^\ell) = \frac{1}{n} Var((\lambda_1^{(n)})^\ell) = O(\tfrac{1}{n})$$

  But the eigenvalues of a random matrix are everything but iid (as already seen from the joint distribution of the eigenvalues of the GOE at finite n), which explains the different behaviour of the variance.

- A simple heuristics for explaining the $O(\frac{1}{n^2})$ is the following: $m_\ell^{(n)} = \frac{1}{n} Tr((H^{(n)})^\ell)$;

  $m_\ell^{(n)}$ can therefore be seen as a function of the order n^2 iid entries of the matrix $H^{(n)}$, which "explains" the variance of order $\frac{1}{n^2}$, as opposed to the classical case with n iid random variables and variance $\frac{1}{n}$.

(*) a rigorous proof of (2) can be found in $\begin{cases} \text{Jonsson 82} \\ \text{Anderson-Zeitouni 04} \end{cases}$

Concentration

Let us shift our attention from $m_\ell^{(n)} = \frac{1}{n}\sum_{j=1}^{n} (\lambda_j^{(n)})^\ell$ to $\frac{1}{n}\sum_{j=1}^{n} f(\lambda_j^{(n)})$, where $f : \mathbb{R} \to \mathbb{R}$ is a continuous function satisfying moreover:

(v) f is convex

(vi) f is Lipschitz with constant L, i.e. $|f(x)-f(y)| \leq L|x-y|$ $\forall x,y \in \mathbb{R}$

Remember that $\frac{1}{n}\sum_{j=1}^{n} f(\lambda_j^{(n)}) = \frac{1}{n}\mathrm{Tr}\left(f(H^{(n)})\right)$,

so that this object can also be seen as a function of the order n^2 iid entries of the matrix $H^{(n)}$.

Theorem (Guionnet - Zeitouni 2000)

Under assumptions (i), (ii)', (iii) — (vi),

$$\mathbb{P}\left(\left| \frac{1}{n}\sum_{j=1}^{n} f(\lambda_j^{(n)}) - \int_{\mathbb{R}} f(x)\, p_{sc}(x)\, dx \right| > t \right)$$

$$\leq 4 \exp\left(-n^2 \left(t - O(\tfrac{1}{n})\right)^2 / 16\, C^2 L^2\right) \qquad \forall t > 0$$

recall that $|h_{jk}| \leq C$ (assumption (ii)')

x More concretely, this theorem says that

$$\frac{1}{n}\sum_{j=1}^{n} f(\lambda_j^{(n)}) = \int_{\mathbb{R}} f(x)\, p_{sc}(x)\, dx + O\left(\frac{1}{n}\right)$$ (in the probabilistic sense)

i.e. that the variance is of order $\frac{1}{n^2}$ again.

Proof idea

- $\frac{1}{n}\sum_{j=1}^{n} f(\lambda_j^{(n)}) = \frac{1}{n}\operatorname{Tr}\left(f(H^{(n)})\right) := F_n(\{\underbrace{h_{jk}}_{\in[-c,c]},\ j\le k\})$

  It can be shown that

  - $f:\mathbb{R}\to\mathbb{R}$ convex $\Rightarrow F_n:[-c,c]^{\frac{n(n+1)}{2}}\to\mathbb{R}$ is convex
  - $f:\mathbb{R}\to\mathbb{R}$ Lipschitz with constant L

    $\Rightarrow F_n:[-c,c]^{\frac{n(n+1)}{2}}\to\mathbb{R}$ is Lipschitz with cst $\frac{L}{n}$

    i.e. $|F_n(u)-F_n(v)|\le\frac{L}{n}\|u-v\|\quad\forall u,v\in[-c,c]^{\frac{n(n+1)}{2}}$

- Talagrand's concentration inequality (Annals of Prob. 1995):

  If $Y_1..Y_m$ are iid random variables such that $|Y_j|\le c$
  and $F:[-c,c]^m\to\mathbb{R}$ is convex and Lipschitz with cst K,
  then $\mathbb{P}(|F(Y_1..Y_m)-M_F|\ge t)\le 4\exp(-\frac{t^2}{16c^2K^2})$
  where M_F is the median of $F(Y_1..Y_m)$.

- Here, $m=\frac{n(n+1)}{2}$ and $K=\frac{L}{n}$, so

$$\mathbb{P}\left(|F_n(\{h_{jk},\ j\le k\})-M_{F_n}|\ge t\right)\le 4\exp\left(-\frac{n^2t^2}{16c^2L^2}\right)$$

- The last step consists in showing that

$$M_{F_n}=\mathbb{E}(F_n)+O(\tfrac{1}{n})=\int_{\mathbb{R}} f(x)\,p_n(x)\,dx+O(\tfrac{1}{n})$$

#

Random matrix theory: lecture 15

Marčenko-Pastur's theorem

Let H be a $n \times n$ real or complex random matrix such that

(i) $\{h_{jk},\ j,k = 1..n\}$ are iid random variables

(ii) h_{11} is a bounded random variable (i.e. $|h_{11}(\omega)| \leq C\ \forall \omega$)

(iii) $\mathbb{E}(h_{11}) = 0$, $\mathbb{E}(|h_{11}|^2) = 1$

and let $W^{(n)} := \frac{1}{n} H H^*$, $\lambda_1^{(n)} .. \lambda_n^{(n)}$ be the eigenvalues of $W^{(n)}$

($\underline{NB}$: $\lambda_j^{(n)} \geq 0$)

Remarks:

- assumption (ii) may be dropped without affecting the result
- the assumption that $\mathbb{E}(h_{11}) = 0$ may also be dropped (see lecture 17) and one may also assume a variance different than 1; this changes slightly the limiting distribution.

Theorem (Marčenko-Pastur 1967, Bai-Silverstein 1995)

Under assumptions (i)-(iii),

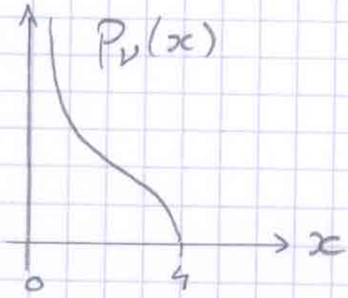
$$F_n(t) := \frac{1}{n} \# \{j : \lambda_j^{(n)} \leq t\} \xrightarrow[n\to\infty]{} \int_0^{\infty} p_\nu(x)\, dx \quad \text{a.s. } \forall t \geq 0$$

where $p_\nu(x) = \frac{1}{\pi} \sqrt{\frac{1}{x} - \frac{1}{4}}\ 1_{0 < x < 4}$

"quarter-circle" distribution

Remarks:

- Marčenko-Pastur's theorem is more general: see lecture 16
- The "quarter-circle" distribution p_ν looks like:

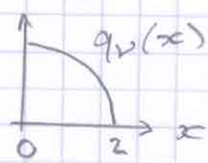


and is therefore not a quater-circle ...

But let us consider instead the limiting distribution of the singular values of $H^{(n)} := \frac{1}{\sqrt{n}} H$: since these are the square roots of the $\lambda_j^{(n)}$, their limiting distribution is:

$$q_\nu(x) = p_\nu(x^2) \cdot \underset{\substack{\uparrow \\ \text{(Jacobian of } x \mapsto x^2)}}{2x} = \frac{1}{\pi} \sqrt{4 - x^2}\, 1_{0<x<2}$$

which is indeed a quater-circle:



- Note that the singular values of $H^{(n)}$, which is non-Hermitian, have a priori no relation with its eigenvalues. It can be shown however that the limiting eigenvalue distribution of $H^{(n)}$ is the uniform distribution on the disc of radius 2 in the complex plane!

<u>Proof</u> (main ideas)

This time, we are going to use the Stieltjes transform:

- $$g_n(z) := \int_0^\infty \frac{1}{x-z}\, dF_n(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{\lambda_j^{(n)} - z}$$
$$= \frac{1}{n} \operatorname{Tr}\left(\left(W^{(n)} - z I \right)^{-1} \right)$$

- $$g_\nu(z) := \int_0^\infty \frac{1}{x-z}\, p_\nu(x)\, dx = \frac{1}{\pi} \int_0^4 \frac{1}{x-z} \sqrt{\frac{1}{x} - \frac{1}{4}}\, dx$$
$$= \ldots = -\frac{1}{2} + \sqrt{\frac{1}{4} - \frac{1}{z}} \quad \text{for } \operatorname{Im} z > 0$$

<u>NB</u>: the reverse formula $p_\nu(x) = \lim_{\varepsilon \downarrow 0} \frac{1}{\pi} \operatorname{Im}(g_\nu(x + i\varepsilon))$

is actually easier to check!

More importantly, note that $g_\nu(z)$ is solution of the second order equation:

$$z \cdot g(z)^2 + z \cdot g(z) + 1 = 0$$

- We are now going to prove that:

$$\lim_{n \to \infty} \sup_{z \in K} |g_n(z) - g_\nu(z)| = 0 \quad \text{a.s.}, \; \forall K = [-A, A] \times [b, B]$$
$$A > 0,\ B > b > 0$$

(we do not show uniform convergence here) $\qquad \forall z \in \mathbb{C}$ s.t. $\operatorname{Im} z > 0$

which implies the result.

First notice that $W^{(n)} = \frac{1}{n} H H^* = \frac{1}{n} \sum_{k=1}^{n} \underbrace{h_k h_k^*}_{n\times n \text{ matrix}}$,

where $h_k = k^{th}$ column of H.

Let $W_k^{(n)} := W^{(n)} - \frac{1}{n} h_k h_k^* = \frac{1}{n} \sum_{\substack{\ell=1 \\ \ell \neq k}}^{n} h_\ell h_\ell^*$.

and $G^{(n)}(z) = \left(W^{(n)} - zI\right)^{-1}$, $G_k^{(n)}(z) = \left(W_k^{(n)} - zI\right)^{-1}$.

Note that $g_n(z) = \frac{1}{n} \operatorname{Tr} G^{(n)}(z)$; we are interested in finding a limiting equation for $g_n(z)$.

Lemma 1

$$\frac{1}{n} h_k^* G^{(n)}(z) h_k = \frac{\frac{1}{n} h_k^* G_k^{(n)}(z) h_k}{1 + \frac{1}{n} h_k^* G_k^{(n)}(z) h_k}$$

Proof

$$h_k^* G_k^{(n)}(z) \overbrace{\left(W^{(n)} - zI\right)}^{= G^{(n)}(z)^{-1}} = h_k^* G_k^{(n)}(z) \left(W_k^{(n)} - zI + \frac{1}{n} h_k h_k^*\right)$$

$$= h_k^* + \frac{1}{n} h_k^* G_k^{(n)}(z) h_k h_k^* = \left(1 + \frac{1}{n} h_k^* G_k^{(n)}(z) h_k\right) h_k^*$$

$$\Rightarrow h_k^* G_k^{(n)}(z) = \Big(1 + \underbrace{\tfrac{1}{n} h_k^* G_k^{(n)}(z) h_k}_{\text{scalar}}\Big) h_k^* G^{(n)}(z) \qquad (*)$$

so $h_k^* G_k^{(n)}(z) h_k = \left(1 + \frac{1}{n} h_k^* G_k^{(n)} h_k\right) h_k^* G^{(n)}(z) h_k$. #

Lemma 2

$$g_n(z) = \frac{1}{n} \operatorname{Tr} G^{(n)}(z) = -\frac{1}{nz} \sum_{k=1}^{n} \frac{1}{1 + \frac{1}{n} h_k^* G_k^{(n)}(z) h_k}$$

Proof

$$1 = \frac{1}{n}\operatorname{Tr}(I) = \frac{1}{n}\operatorname{Tr}\left((W^{(n)} - zI)\, G^{(n)}(z)\right)$$

$$= \frac{1}{n}\operatorname{Tr}\left(\frac{1}{n}\sum_{k=1}^{n} h_k h_k^* G^{(n)}(z) - z\, G^{(n)}(z)\right)$$

$$= \frac{1}{n}\sum_{k=1}^{n}\frac{1}{n} h_k^* G^{(n)}(z)\, h_k - z\, g_n(z)$$

$$= \frac{1}{n}\sum_{k=1}^{n}\frac{\frac{1}{n} h_k^* G_k^{(n)}(z)\, h_k}{1 + \frac{1}{n} h_k^* G_k^{(n)}(z) h_k} - z\, g_n(z) \qquad \text{by Lemma 1}$$

$$\Rightarrow z\, g_n(z) = \frac{1}{n}\sum_{k=1}^{n}\left(\frac{\frac{1}{n} h_k^* G_k^{(n)}(z)\, h_k}{1 + \frac{1}{n} h_k^* G_k^{(n)}(z) h_k} - 1\right)$$

$$= \frac{1}{n}\sum_{k=1}^{n}\frac{1}{1 + \frac{1}{n} h_k^* G_k^{(n)}(z)\, h_k} \qquad \#$$

NB: this formula holds for any matrix of the form $W^{(n)} = \frac{1}{n} HH^*$.

Lemma 3

$$\frac{1}{n} h_k^* G_k^{(n)}(z)\, h_k \cong \frac{1}{n}\operatorname{Tr}\left(G_k^{(n)}(z)\right) \quad \text{as } n \to \infty$$

more precisely:

$$\mathbb{P}\left(\left|\frac{1}{n} h_k^* G_k^{(n)}(z)\, h_k - \frac{1}{n}\operatorname{Tr}\left(G_k^{(n)}(z)\right)\right| > a\right) \leq \frac{C(a)}{n^2}$$

which implies, by the Borel-Cantelli lemma,

that the difference converges to zero a.s. as $n \to \infty$.

Proof idea

We only check here that expectations are equal:

$$\mathbb{E}\left(\frac{1}{n} h_k^* G_k^{(n)}(z)\, h_k\right) = \frac{1}{n}\sum_{j,\ell=1}^{n} \mathbb{E}\left(\overline{h_{jk}}\,(W_k^{(n)} - zI)^{-1}_{j\ell}\, h_{\ell k}\right)$$

Since $W_k^{(n)}$ does not "contain" the k^{th} column of H, it is independent of both $\overline{h_{jk}}$ and $h_{\ell k}$ (assumption (i)), so

$$\ldots = \frac{1}{n}\sum_{j,\ell=1}^{n} \underbrace{\mathbb{E}\left(\overline{h_{jk}}\, h_{\ell k}\right)}_{=\delta_{j\ell} \text{ by assumption (iii)}} \mathbb{E}\left(\left(W_k^{(n)} - zI\right)^{-1}_{j\ell}\right)$$

$$= \frac{1}{n}\sum_{j=1}^{n} \mathbb{E}\left(\left(W_k^{(n)} - zI\right)^{-1}_{jj}\right) = \mathbb{E}\left(\frac{1}{n}\operatorname{Tr} G_k^{(n)}(z)\right) \quad \text{"}\#\text{"}$$

<u>NB</u>: assumption (ii) is needed in order to prove that a.s. convergence holds, but there is also a way to get rid of it.

Lemma 4

$$\left|\frac{1}{n}\cdot\operatorname{Tr} G_k^{(n)}(z) - \frac{1}{n}\cdot\operatorname{Tr} G^{(n)}(z)\right| \leq \frac{1}{n\cdot|\operatorname{Im} z|} \quad \forall z \text{ s.t. } \operatorname{Im} z \neq 0$$

<u>Proof</u>: homework

<u>NB</u>: this lemma also holds for any matrix of the form $W^{(n)} = \frac{1}{n} HH^*$.

Conclusion of the proof of the theorem:

By Lemmas 2, 3 & 4, we have:

$$g_n(z) = -\frac{1}{nz}\sum_{k=1}^{n}\frac{1}{1+\frac{1}{n}h_k^* G_k^{(n)}(z)h_k}$$

$$\underset{n\to\infty}{\simeq} -\frac{1}{nz}\sum_{k=1}^{n}\frac{1}{1+g_n(z)} = -\frac{1}{z}\frac{1}{(1+g_n(z))}$$

So $\quad z\,g_n(z)^2 + z\,g_n(z) + 1 \underset{n\to\infty}{\simeq} 0$

i.e. $g_n(z) \xrightarrow[n\to\infty]{} g_\nu(z)$ a.s., where $g_\nu(z)$ is solution of

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

Note that there are a priori two solutions of this equation, but only one of them is a Stieltjes transform, and we already know that the functions g_n are Stieltjes transforms, so a continuity argument allows to conclude that:

$$g_\nu(z) = -\frac{1}{2} + \sqrt{\frac{1}{4} - \frac{1}{z}} \qquad \text{for } \operatorname{Im} z > 0$$

#

<u>NB</u>: we have skipped quite a lot of technical difficulties in this proof!

Random matrix theory: lecture 16

Generalizations: 1°) non-square matrices

Let H be a $n \times m$ random matrix with i.i.d. entries such that $\mathbb{E}(h_{11}) = 0$, $\mathbb{E}(|h_{11}|^2) = 1$, and $W^{(n)} = \frac{1}{n} H H^*$.

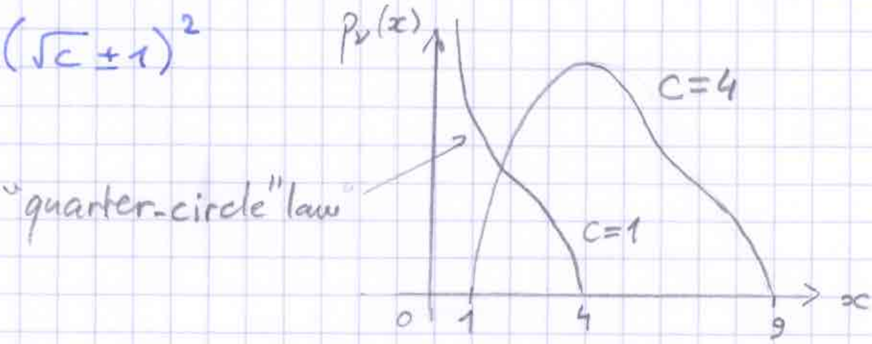
We assume that $m, n \to \infty$ with $\frac{m}{n} = c \geq 1$ fixed.

Let $\lambda_1^{(n)} \ldots \lambda_n^{(n)}$ be the eigenvalues of $W^{(n)}$. Then

$$F_n(t) := \frac{1}{n} \#\{j : \lambda_j^{(n)} \leq t\} \xrightarrow[n \to \infty]{} \int_0^t p_\nu(x)\, dx \quad \text{a.s. } \forall t \geq 0$$

where $p_\nu(x) = \frac{1}{2\pi} \sqrt{\left(\frac{x_+}{x} - 1\right)\left(1 - \frac{x_-}{x}\right)} \cdot 1_{x_- < x < x_+}$

and $x_\pm = (\sqrt{c} \pm 1)^2$



Remarks:

- the Stieltjes transform of p_ν, given by
$$g_\nu(z) = \int_{\mathbb{R}} \frac{1}{x - z} \cdot p_\nu(x)\, dx$$
is solution of the quadratic equation:
$$z\, g_\nu(z)^2 + (z + 1 - c)\, g_\nu(z) + 1 = 0$$
- if $\frac{m}{n} = c < 1$, then p_ν is the limiting distribution of the non-zero eigenvalues of $W^{(n)}$ (there are m of them: the remaining $n-m$ eigenvalues are equal to zero).

2°) insertion of a diagonal matrix Q

Let H be a $n\times n$ matrix with i.i.d. entries such that $\mathbb{E}(h_{11})=0$ and $\mathbb{E}(|h_{11}|^2)=1$; let $Q^{(n)}=\mathrm{diag}(q_1\cdots q_n)$ (*) be a deterministic diagonal matrix and $W^{(n)}=\frac{1}{n}HQ^{(n)}H^*$, with corresponding eigenvalues $\lambda_1^{(n)}\cdots\lambda_n^{(n)}$. (*) with $q_j\geq 0$

If $\frac{1}{n}\#\{1\leq j\leq n : q_j\leq t\}\xrightarrow[n\to\infty]{} F_Q(t)$ $\forall t\geq 0$, with corresponding

× Stieltjes transform $g_Q(z)=\int_0^\infty \frac{1}{x-z}\,dF_Q(x)$, then

$$F_n(t):=\frac{1}{n}\#\{j : \lambda_j^{(n)}\leq t\}\xrightarrow[n\to\infty]{} F_W(t)\quad \text{a.s.}\ \forall t\geq 0$$

whose Stieltjes transform $g_W(z)$ satisfies the equation:

$$z\,g_W(z)^2+g_Q\left(-\frac{1}{g_W(z)}\right)=0$$

<u>Example</u>: if $Q^{(n)}=I_n=\mathrm{diag}(1,..,1)$, then $g_Q(z)=\frac{1}{1-z}$

so $g_Q\left(-\frac{1}{g_W(z)}\right)=\dfrac{1}{1+\frac{1}{g_W(z)}}=\dfrac{g_W(z)}{1+g_W(z)}$

$$\Longrightarrow\quad z\,g_W(z)^2+\frac{g_W(z)}{1+g_W(z)}=0$$

$$z\,g_W(z)\,(1+g_W(z))+1=0$$

$$z\,g_W(z)^2+z\,g_W(z)+1=0$$

again, the "quarter-circle" law.

3°) application: multiplication of (non-negative definite) random matrices

- It has been shown in (Silverstein, 1995) that if we replace the diagonal matrix $Q^{(n)}$ with any (deterministic) non-negative definite matrix $Q^{(n)}$ whose eigenvalues $q_1^{(n)} \dots q_n^{(n)}$ satisfy the same hypothesis as above, then the same conclusion holds: $z\, g_W(z)^2 + g_Q\left(-\frac{1}{g_W(z)}\right) = 0 \quad (*)$

- A further generalization of this result is that if $Q^{(n)}$ is a non-negative definite random matrix independent of H with limiting eigenvalue distribution F_Q and corresponding Stieltjes transform g_Q, then the result continues to hold!

Example: let $Q^{(n)} = \frac{1}{n} \tilde{H}\tilde{H}^*$, where $\tilde{H}$ and H are iid

Then we know that in the limit $n \to \infty$,

$$z\, g_Q(z)^2 + z\, g_Q(z) + 1 = 0$$

$$\Rightarrow -\frac{1}{g_W(z)}\, g_Q\left(-\frac{1}{g_W(z)}\right)^2 - \frac{1}{g_W(z)}\, g_Q\left(-\frac{1}{g_W(z)}\right) + 1 = 0 \qquad z \mapsto -\frac{1}{g_W(z)}$$

use $(*)$

$$\Rightarrow -z^2 g_W(z)^3 + z\, g_W(z) + 1 = 0$$

cubic equation for g_W (NB: $W^{(n)} = \frac{1}{n^2} H\tilde{H}\tilde{H}^* H^*$)

16.5.07

4

4°) addition of random matrices

Let H be a $n \times n$ random matrix with i.i.d. entries such that $\mathbb{E}(h_{11}) = 0$ and $\mathbb{E}(|h_{11}|^2) = 1$; let $A^{(n)}$ be a deterministic $n \times n$ Hermitian matrix with eigenvalues $a_1^{(n)} .. a_n^{(n)}$ satisfying $\frac{1}{n} \# \{j : a_j^{(n)} \leq t\} \underset{n \to \infty}{\to} F_A(t)$ with corresponding Stieltjes transform $g_A(z)$; and let $W^{(n)} = A^{(n)} + \frac{1}{n} H H^*$ with eigenvalues $\lambda_1^{(n)} .. \lambda_n^{(n)}$.

Then

$$F_n(t) := \frac{1}{n} \# \{j : \lambda_j^{(n)} \leq t\} \xrightarrow[n \to \infty]{} F_W(t) \quad \text{a.s.} \quad \forall t \in \mathbb{R}$$

whose Stieltjes transform $g_W(z)$ satisfies the equation:

$$g_W(z) = g_A\left(z - \frac{1}{1 + g_W(z)}\right)$$

Remarks:

- The result again naturally generalizes to the case where $A^{(n)}$ is a random matrix independent of H.
- When $A^{(n)} = 0$, $g_A(z) = -\frac{1}{z}$ and we again recover the "quarter-circle" law.

5°) <u>all combinations of these are possible</u>!

The most general result is actually already contained in Marčenko-Pastur (1967) and Bai-Silverstein (1995). For reference, it says that:

- if $A^{(n)}$ is deterministic Hermitian $n \times n$ with limiting S.T. $g_A(z)$
- if $Q^{(n)}$ is diagonal $m \times m$ with limiting S.T. $g_Q(z)$
- if H is $n \times m$ with iid entries such that $\mathbb{E}(h_{11}) = 0$ (*), $\mathbb{E}(|h_{11}|^2) = 1$ and $\frac{m}{n} = c$
- then $W^{(n)} = A^{(n)} + \frac{1}{n} H Q^{(n)} H^*$ has a limiting eigenvalue distribution whose Stieltjes transform $g_W(z)$ is solution of the equation:

$$(!) \quad g_W(z) = g_A\left(z - \frac{c}{g_W(z)}\left(1 - \frac{1}{g_W(z)} \cdot g_Q\left(-\frac{1}{g_W(z)}\right)\right)\right)$$

The core of the proof is the one developed in the last lecture; we will therefore not redo it.

(*) this assumption can actually be dropped; we will come back to this later in the class.

Back to the addition of random matrices and an alternate proof of Wigner's theorem

(using Stieltjes transform)

- Let H be a $n \times n$ real symmetric matrix such that $\{h_{jk}, j \leq k\}$ are iid random variables with $\mathbb{E}(h_{11}) = 0$ and $\mathbb{E}(h_{11}^2) = 1$.
- let $A^{(n)} = \operatorname{diag}(a_1 \dots a_n)$ be deterministic ($a_j \in \mathbb{R}$) and such that $\frac{1}{n} \#\{1 \leq j \leq n : a_j \leq t\} \xrightarrow[n\to\infty]{} F_A(t)$ $\forall t \in \mathbb{R}$, with corresponding Stieltjes transform $g_A(z) = \int_{\mathbb{R}} \frac{1}{x-z} \, dF_A(x)$.
- Let $B^{(n)} = A^{(n)} + \frac{1}{\sqrt{n}} H$, with eigenvalues $\lambda_1^{(n)} \dots \lambda_n^{(n)}$.

Theorem (Pastur et al.)

$$F_n(t) := \frac{1}{n} \#\{j : \lambda_j^{(n)} \leq t\} \xrightarrow[n\to\infty]{} F_B(t) \quad \text{a.s.} \quad \forall t \geq 0$$

where the Stieltjes transform of F_B satisfies the equation

$$g_B(z) = g_A(z + g_B(z)), \quad z \in \mathbb{C} \setminus \mathbb{R}.$$

Remarks:

- When $A^{(n)} = 0$, $g_A(z) = -\frac{1}{z}$, so $-g_B(z)(z + g_B(z)) = 1$,

$$g_B(z)^2 + z g_B(z) + 1 = 0, \quad g_B(z) = -\frac{z}{2} \pm \sqrt{\frac{z^2}{4} - 1}$$

and $p_B(x) = \frac{1}{\pi} \lim_{\varepsilon \downarrow 0} \operatorname{Im} g_B(x + i\varepsilon) = \frac{1}{2\pi} \sqrt{4 - x^2} \cdot 1_{|x| < 2}$

i.e. we recover Wigner's semi-circle law.

- the result generalizes in various directions:

1°) complex Hermitian case

2°) $A^{(n)}$ Hermitian with eigenvalues $a_1^{(n)} \dots a_n^{(n)}$

3°) $A^{(n)}$ random & independent of H

<u>Proof of the result</u> (main ideas)

- Let M be a $n \times n$ real symmetric matrix and $z \in \mathbb{C} \setminus \mathbb{R}$. Then (matrix inversion lemma)

$$(M - zI_n)^{-1}_{11} = 1/\left(m_{11} - z - m_1^T (M_1 - zI_{n-1})^{-1} m_1\right)$$

where
$$M = \begin{array}{cc} \begin{pmatrix} m_{11} & m_1^T \\ m_1 & M_1 \end{pmatrix} & \leftarrow \text{one row} \\ \uparrow \\ \text{one column} & \end{array}.$$

Similarly,

k^{th} diag. coeff. — matrix with column & row k removed — k^{th} column without diag. coeff.

$$\mathrm{Tr}\left((M - zI_n)^{-1}\right) = \sum_{k=1}^{n} 1/\left(m_{kk} - z - m_k^T (M_k - zI_{n-1})^{-1} m_k\right)$$

- $G^{(n)}(z) := (B^{(n)} - zI_n)^{-1}$, $G_k^{(n)}(z) := (B_k^{(n)} - zI_{n-1})^{-1}$

$g_n(z) = \frac{1}{n} \mathrm{Tr}\, G^{(n)}(z)$

Using the matrix inversion lemma, we obtain

$$g_n(z) = \frac{1}{n} \sum_{k=1}^{n} 1/\left(b_{kk}^{(n)} - z - (b_k^{(n)})^T G_k^{(n)}(z)\, b_k^{(n)}\right)$$

(cf. formula last time)

- Now: $\begin{cases} b_{kk}^{(n)} = a_k + \frac{1}{\sqrt{n}} h_{kk} \simeq a_k & \text{as } n\to\infty \\ b_k^{(n)} = \frac{1}{\sqrt{n}} h_k & \text{(since this vector contains no diag. element)} \end{cases}$

so $g_n(z) \underset{n\to\infty}{\simeq} \frac{1}{n}\sum_{k=1}^{n} 1 \Big/ \left(a_k - z - \frac{1}{n} h_k^T G_k^{(n)}(z) h_k\right)$

- Same reasoning as last time:

1°) $\frac{1}{n} h_k^T G_k^{(n)}(z) h_k \simeq \frac{1}{n} \mathrm{Tr}\, G_k^{(n)}(z)$ (w.h.p.)

2°) $\frac{1}{n}\mathrm{Tr}\, G_k^{(n)}(z) \simeq \frac{1}{n}\mathrm{Tr}\, G^{(n)}(z) = g_n(z)$

- Therefore: $g_n(z) \simeq \frac{1}{n}\sum_{k=1}^{n} 1 \big/ (a_k - z - g_n(z))$

(assumption on $A^{(n)}\to$) $\simeq \int_{\mathbb{R}} \frac{1}{x - z - g_n(z)}\, dF_A(x)$

(definition $\to$) $= g_A(z + g_n(z))$

So as $n\to\infty$, $g_n(z) \to g_B(z)$ solution of $g_B(z) = g_A(z + g_B(z))$ #

<u>Example</u>: let $A^{(n)} = \begin{pmatrix} 0 & 1 & & 0 \\ 1 & \ddots & \ddots & \\ & \ddots & \ddots & 1 \\ 0 & & 1 & 0 \end{pmatrix}$; $A^{(n)}$ has the following limiting

pdf for its eigenvalues: $p_A(x) = \frac{1}{\pi\sqrt{4-x^2}}\, 1_{|x|<2}$

(cf. class on Toeplitz matrices)

The corresponding Stieltjes transform is $g_A(z) = \frac{1}{\sqrt{z^2-4}}$

$\Rightarrow g_B(z) = 1 \Big/ \sqrt{(z + g_B(z))^2 - 4}$

i.e. $g_B(z)^4 + 2z\, g_B(z)^3 + (z^2-4)\, g_B(z)^2 - 1 = 0$

(quartic equation)

Random matrix theory: lecture 17

Largest eigenvalue of Wigner's matrices

Preliminary: matrix norms

Def.: a norm on the space of $n \times n$ (complex) matrices is an application

$$\|\cdot\| : M_n \to \mathbb{R} \quad \text{such that}$$

- $\|A\| \geq 0$, $\|A\| = 0$ iff $A = 0$, $\forall A$
- $\|cA\| = |c| \cdot \|A\|$, $\forall c \in \mathbb{C}$, $\forall A$
- $\|A+B\| \leq \|A\| + \|B\|$, $\forall A, B$

Def: $\|\cdot\|$ is called a matrix norm if moreover

$$\|AB\| \leq \|A\| \cdot \|B\|, \quad \forall A, B$$

Properties: if $\|\cdot\|$ is a matrix norm, then

- $\|A^2\| \leq \|A\|^2$; likewise, $\|A^k\| \leq \|A\|^k$
- $A^2 = A \Rightarrow \|A\| = \|A^2\| \leq \|A\|^2 \Rightarrow \|A\| \geq 1$
- in particular, $\|I\| \geq 1$
- A invertible $\Rightarrow \|I\| = \|AA^{-1}\| \leq \|A\| \cdot \|A^{-1}\|$

$$\text{so } \|A^{-1}\| \geq \frac{\|I\|}{\|A\|} \geq \frac{1}{\|A\|}$$

Examples and counter-examples (⚠ notations ⚠)

1) The ℓ^1 norm $\|A\|_1 := \sum_{j,k=1}^{n} |a_{jk}|$ is a matrix norm.

2) The ℓ^2 norm (or Euclidean norm or Frobenius norm)

defined as $\|A\|_2^2 := \sum_{j,k=1}^{n} |a_{jk}|^2$ is a matrix norm.

But note that the modified version of this norm:

$$\|A\|_{\tilde{2}}^2 := \frac{1}{n} \sum_{j,k=1}^{n} |a_{jk}|^2 \quad (= \|A\|_2^2 \text{ in homework 4})$$

is not a matrix norm ($\|AB\|_{\tilde{2}} \not\leq \|A\|_{\tilde{2}} \cdot \|B\|_{\tilde{2}} \ \forall A, B$).

NB: both $\|\cdot\|_2$ and $\|\cdot\|_{\tilde{2}}$ are unitarily invariant, i.e.

$$\|UAV\|_2 = \|UA\|_2 = \|AV\|_2 = \|A\|_2 \quad \forall U, V \text{ unitary}$$

3) The ℓ^∞ norm $\|A\|_\infty := \max_{1 \leq j,k \leq n} |a_{jk}|$ is not a matrix norm

but its modified version $\|A\|_{\tilde{\infty}} := n \max_{1 \leq j,k \leq n} |a_{jk}|$ is.

Induced norms

Def: let $\|\cdot\|$ be a (vector) norm on $\mathbb{C}^n$. We define the following induced norm on M_n:

$$|||A||| := \max_{\|x\|=1} \|Ax\| \quad \left(= \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} \right)$$

Proposition: Induced norms are matrix norms.

In addition, $\|Ax\| \leq |||A||| \cdot \|x\|$ and $|||I||| = 1$.
$\forall A, x$

Examples:

1) The maximum column sum matrix norm $|||\cdot|||_1$ defined as

$$|||A|||_1 := \max_{1\le k\le n} \sum_{j=1}^{n} |a_{jk}|$$

is induced by the ℓ^1 norm on $\mathbb{C}^n$, i.e.

$$|||A|||_1 = \max_{\|x\|_1=1} \|Ax\|_1 \quad \text{where } \|x\|_1 = \sum_{j=1}^{n} |x_j| .$$

2) The spectral norm $|||\cdot|||_2$ defined as

$$|||A|||_2 := \max \{ \sqrt{\lambda} : \lambda \text{ is an eigenvalue of } A^*A \}$$

is induced by the ℓ^2 norm on $\mathbb{C}^n$, i.e.

$$|||A|||_2 = \max_{\|x\|_2=1} \|Ax\|_2 \quad (= \|A\|_1 \text{ in homework 4})$$

NB: $|||\cdot|||_2$ is unitarily invariant.

3) The maximum raw sum norm $|||\cdot|||_\infty$ defined as

$$|||A|||_\infty := \max_{1\le j\le n} \sum_{k=1}^{n} |a_{jk}|$$

is induced by the ℓ^∞ norm on $\mathbb{C}^n$, i.e.

$$|||A|||_\infty = \max_{\|x\|_\infty=1} \|Ax\|_\infty \quad \text{where } \|x\|_\infty = \max_{1\le j\le n} |x_j|$$

(Pfs: first show that $\|Ax\| \le |||A||| \cdot \|x\| \quad \forall x \in \mathbb{C}^n$
then find a x such that there is equality)

Spectral radius

$$\rho(A) := \max\{|\lambda| : \lambda \text{ is an eigenvalue of } A\}$$

Proposition

If A is normal, then $\rho(A) = \max_{\|x\|_2=1} |x^* A x|$

(proof goes along the same lines as the proof of $|||A|||_2 = \max_{\|x\|_2=1} \|Ax\|$)

Note that $\rho(\cdot)$ is not a norm, but the following holds:

Proposition

For any matrix norm $\|\cdot\|$, $\rho(A) \leq \|A\|$.

Proof

$\exists \lambda, x$ st $Ax = \lambda x$ and $|\lambda| = \rho(A)$.

Let X be the matrix whose columns are all equal to x.

Then $AX = \lambda X$ and for any matrix norm $\|\cdot\|$, we have

$$|\lambda| \cdot \|X\| = \|\lambda X\| = \|AX\| \leq \|A\| \cdot \|X\|$$

so $\rho(A) = |\lambda| \leq \|A\|$, since $\|X\| \neq 0$. #

Proposition (whose proof is more involved)

For any matrix norm $\|\cdot\|$, $\rho(A) = \lim_{\ell\to\infty} \|A^\ell\|^{1/\ell}$.

[ref: Horn-Johnson p.299]

Back to Wigner's matrices

Let H^0 be a $n\times n$ real symmetric random matrix s.t.

(i) $\{h^0_{jk},\ j\leq k\}$ are iid random variables ($\&$ $h_{kj}=h_{jk}$)

(ii) $|h^0_{11}(\omega)|\leq C$ $\forall\omega$ (bdd random variable)

(iii) $\mathbb{E}(h^0_{11})=0$, $\mathbb{E}((h^0_{11})^2)=1$

and $H^{(0,n)}:=\frac{1}{\sqrt{n}}H^0$, $\lambda_1^{(0,n)}\ldots\lambda_n^{(0,n)}:=$ eigenvalues of $H^{(0,n)}$

We already know that

$$F_n^0(t):=\frac{1}{n}\#\{j:\lambda_j^{(0,n)}\leq t\}\xrightarrow[n\to\infty]{}\int_{-\infty}^{t}p_\mu(x)\,dx\quad \text{a.s.}$$

where p_μ is the semi-circle distribution:

$$p_\mu(x)=\frac{1}{2\pi}\sqrt{4-x^2}\cdot 1_{|x|\leq 2}$$

First remark

The same result holds if we replace assumption (iii)

by the weaker assumption (iii)' $\mathbb{E}(h_{11})=a\in\mathbb{R}$, $\mathrm{Var}(h_{11})=1$

Proof

× • Notice that $H=a\,\mathbb{1}+H^0$, where $\mathbb{1}$ is the all-ones matrix and H^0 satisfies assumptions (i)-(iii).

Therefore, $\mathrm{rank}(H-H^0)=1$.

(NB: $a\mathbb{1}$ has 1 eigenvalue equal to na and $n-1$ zero eigenvalues)

- We will need Weyl's inequalities: if A, B are two real symmetric matrices with respective eigenvalues $\lambda_1^A \geq \dots \geq \lambda_n^A$ and $\lambda_1^B \geq \dots \geq \lambda_n^B$, then

$$\lambda_{j+k-1}^{A+B} \leq \lambda_j^A + \lambda_k^B \qquad \forall j, k \geq 1 \quad \text{s.t. } j+k-1 \leq n$$

- Let $F_n(t) := \frac{1}{n} \# \{ j : \lambda_j^{(n)} \leq t \}$ with $\lambda_j^{(n)}$ = e.v. of $H^{(n)} = \frac{1}{\sqrt{n}} H$.

Then $\sup\limits_{t \in \mathbb{R}} | F_n(t) - F_n^0(t) | \leq \frac{\text{rank}(H - H^0)}{n} = \frac{1}{n}$ (∗)

Therefore, F_n and F_n^0 converge to the same limit as $n \to \infty$. #

- Proof of (∗) using Weyl's inequalities:

let $A = \frac{1}{\sqrt{n}} a \mathbb{1}$: since A is rank one, $\lambda_2^A = \dots = \lambda_n^A = 0$

Let $B = H^{(0,n)}$; i.e. $A + B = \frac{1}{\sqrt{n}} (a \mathbb{1} + H^0) = H^{(n)}$

Weyl's inequalities therefore imply that

$$\lambda_{k+1}^{(n)} \leq \lambda_k^{(0,n)} \qquad \forall k \leq n-1 \quad (\text{take } j = 2)$$

Likewise, one can show that $\lambda_{k+1}^{(0,n)} \leq \lambda_k^{(n)}$.

So $|F_n(t) - F_n^0(t)| = \frac{1}{n} \left| \# \{ j : \lambda_j^{(n)} \leq t \} - \# \{ j : \lambda_j^{(0,n)} \leq t \} \right| \leq \frac{1}{n}$ #

Conclusion: The global regime (that of the semi-circle distribution) does not "see" the non-zero mean of the entries. The situation is quite different for local properties (such as the position of the largest eigenvalue).

Second remark

Wigner's result implies a lower bound on the largest eigenvalue of Wigner's matrices: for any $\varepsilon > 0$,

$$\frac{1}{n}\mathbb{E}\left(\#\{j : \lambda_j^{(n)} \in [2-\varepsilon, 2]\}\right) \xrightarrow[n\to\infty]{} \int_{2-\varepsilon}^{2} p_p(x)\,dx = c_\varepsilon > 0$$

i.e. the number of eigenvalues lying between $2-\varepsilon$ & 2 is equal to nc_ε in expectation. Therefore, there is in expectation at least one eigenvalue larger than $2-\varepsilon$,

i.e. $\lim_{n\to\infty} \mathbb{E}(\lambda_{max}(H^{(n)})) \geq 2-\varepsilon \quad \forall \varepsilon > 0$, i.e. ≥ 2.

Third remark

On the other hand, the largest eigenvalue of $H^{(n)}$ might be much greater than 2!

Example:

Consider $\mathbb{P}(h_{11} = 2) = \mathbb{P}(h_{11} = 0) = \frac{1}{2}$. Then $\mathbb{E}(h_{11}) = 1$, $Var(h_{11}) = 1$.

(satisfying hyp. (iii)')

It can easily be shown that

$$\sqrt{n} \leq \mathbb{E}\left(\lambda_{max}(H^{(n)})\right) \leq \sqrt{2n}$$

($\to$ homework)

21.5.07

8

Finally: let us show that in the case where

$$\mathbb{P}(h_{11}=+1)=\mathbb{P}(h_{11}=-1)=\frac{1}{2},$$

we have $\mathbb{E}(\lambda_{max}(H^{(n)}))\leq 2, \ \forall n$.

Proof:

- We will actually show that $\mathbb{E}(\rho(H^{(n)}))\leq 2, \ \forall n$.

  implying that both the largest and lowest eigenvalues of $H^{(n)}$ lie in the interval $[-2,2]$ in expectation.

- $\mathbb{E}(\rho(H^{(n)}))=\mathbb{E}\left(\lim_{\ell\to\infty}\|(H^{(n)})^\ell\|^{1/\ell}\right)$ for any matrix norm $\|\cdot\|$ (cf. preceding proposition page 4)

  Let us choose $\|A\|^2:=\sum_{j,k=1}^{n}|a_{jk}|^2=\mathrm{Tr}(A^*A)$:

  $$\mathbb{E}(\rho(H^{(n)}))=\mathbb{E}\left(\lim_{\ell\to\infty}\mathrm{Tr}((H^{(n)})^{2\ell})^{1/2\ell}\right)\qquad ((H^{(n)})^*=H^{(n)})$$

  $$\underset{\text{DCT}}{=}\lim_{\ell\to\infty}\mathbb{E}\left(\mathrm{Tr}((H^{(n)})^{2\ell})^{1/2\ell}\right)\underset{\text{Jensen}}{\leq}\lim_{\ell\to\infty}\mathbb{E}\left(\mathrm{Tr}((H^{(n)})^{2\ell})\right)^{1/2\ell}$$

- We have already computed $\mathbb{E}(\mathrm{Tr}((H^{(n)})^{2\ell}))$: (see lect. 13)

  $$=\frac{1}{n^\ell}\sum_{j_1\cdots j_{2\ell}=1}^{n}\mathbb{E}(h_{j_1j_2}\cdots h_{j_{2\ell}j_1})$$

  $$=\frac{1}{n^\ell}\sum_{g_{2\ell}}\sum_{J_{2\ell}\in g_{2\ell}}\mathbb{E}(h(J_{2\ell}))$$

  $$=\frac{1}{n^\ell}\sum_{\substack{g_{2\ell}\text{ even}\\ |v(g_{2\ell})|\leq 1+\ell}} n(n-1)\cdots(n-|v(g_{2\ell})|+1)\cdot\underbrace{Q(g_{2\ell})}_{\equiv 1\text{ here, since } h_{jk}^{2m}\equiv 1\ \forall m}$$

Note that allowing for "repetitions" of vertices, each graph with strictly less than $1+\ell$ vertices may be seen as a graph with $1+\ell$ vertices, allowing some of the vertices to be identical; and there are less than(*) $n^{1+\ell}$ such graphs, so

(*) since we are overcounting some graphs here

$$\mathbb{E}\left(\mathrm{Tr}\left(\left(H^{(n)}\right)^{2\ell}\right)\right) \le \frac{1}{n^\ell} \sum_{\substack{g_{2\ell} \text{ even} \\ |v(g_{2\ell})| = 1+\ell}} n^{1+\ell}$$

$$= n \cdot \#\{g_{2\ell} \text{ even} : |v(g_{2\ell})| = 1+\ell\} = n \underbrace{t_\ell}_{\text{Catalan numbers}}$$

Therefore,

$$\mathbb{E}\left(\rho(H^{(n)})\right) \le \lim_{\ell\to\infty} (n t_\ell)^{1/2\ell} = \lim_{\ell\to\infty} t_\ell^{\,1/2\ell} = 2, \quad \forall n \qquad \#$$

computed before (lect. 13)

$\left(\underline{\text{NB}}: t_\ell = \int_{-2}^{2} x^{2\ell} \rho_{R}(x)\,dx\right)$

<u>Remarks</u>:

- The inequality $\rho(A) \le \|A\|$ with the same choice of matrix norm $\|A\|^2 = \mathrm{Tr}(A^*A)$ does not suffice here:

$$\mathbb{E}\left(\rho(H^{(n)})\right) \le \mathbb{E}\left(\sqrt{\mathrm{Tr}\left(\left(H^{(n)}\right)^2\right)}\right) \le \sqrt{\frac{1}{n}\sum_{j,k=1}^{n} \underbrace{\mathbb{E}\left(h_{jk}^2\right)}_{=1}} = \sqrt{n}$$

and recall that $\|A\|^2 = \frac{1}{n}\mathrm{Tr}(A^*A)$ is <u>not</u> a matrix norm.

- Similarly, one can get bounds on $\mathbb{P}\left(\rho(H^{(n)}) \ge n^\varepsilon\right)$ using Chebyshev's inequality, but a more careful analysis is required for tighter bounds. [<u>ref</u>: Soshnikov].

Random matrix theory: lecture 18

Capacity scaling of multi-antenna channels (MIMO)

× • n transmit antennas and n receive antennas (simplifying assumption)

• channel model: $Y = HX + Z$

• Z = additive white Gaussian noise (of variance 1)

• H = channel matrix with iid entries

such that $\mathbb{E}(|h_{jk}|^2) = 1$ & $h_{jk} \sim -h_{jk}$

(NB: implies that $\mathbb{E}(h_{jk}) = 0$)

• global power constraint at the transmitter: $\sum_{j=1}^{n} \mathbb{E}(|x_j|^2) \leq P$

• assumption: only the receiver knows the channel realizations

$$\Rightarrow \quad C_n = \max_{Q \geq 0 : \operatorname{Tr} Q \leq P} \mathbb{E}\left(\log\det\left(I + HQH^*\right)\right)$$

$$= \mathbb{E}\left(\log\det\left(I + \frac{P}{n} HH^*\right)\right)$$

(see lecture 6 and lemmas 1 & 2 in lecture 7)

• we are now interested in the scaling of the capacity as $n \to \infty$; let $\lambda_1^{(n)} .. \lambda_n^{(n)}$ be the eigenvalues of $W^{(n)} = \frac{1}{n} HH^*$. Then

$$C_n = \mathbb{E}\left(\sum_{j=1}^{n} \log\left(1 + P\lambda_j^{(n)}\right)\right)$$

- By the Marčenko-Pastur theorem, we know that $F_n(t) := \frac{1}{n} \# \{ j : \lambda_j^{(n)} \leq t \} \xrightarrow[n \to \infty]{} \int_0^t p_\nu(x)\, dx$ a.s. $\forall t \geq 0$

× where $p_\nu(x) = \frac{1}{\pi} \sqrt{\frac{1}{x} - \frac{1}{4}}\ 1_{0 < x < 4}$ "quater-circle" distribution

- This implies that for any $f : \mathbb{R} \to \mathbb{R}$ bounded and continuous,

$$\frac{1}{n} \sum_{j=1}^{n} f(\lambda_j^{(n)}) \xrightarrow[n \to \infty]{} \int_0^4 f(x)\, p_\nu(x)\, dx \quad \text{a.s.}$$

- Moreover, convergence in expectation also holds:

$$\frac{1}{n} \mathbb{E} \Big(\sum_{j=1}^{n} f(\lambda_j^{(n)}) \Big) \xrightarrow[n \to \infty]{} \int_0^4 f(x)\, p_\nu(x)\, dx$$

× • A little extra work is required to show that the theorem holds for $f(x) = \log(1 + Px)$, which is continuous on $[0, \infty)$ (note that $\lambda_j^{(n)} \geq 0 \ \forall n$), but unbounded.

Therefore, $\frac{1}{n} C_n \xrightarrow[n \to \infty]{} \int_0^4 \log(1 + Px)\, p_\nu(x)\, dx$.

× Main conclusion:

The capacity scales linearly with the number of antennas, for a fixed global power budget P.

Note however that the model of i.i.d. channel gains eventually breaks down as $n \to \infty$.

Capacity scaling of ad hoc wireless networks

The previous result relies on the study of the global regime of random matrices. Various extensions of the above result exist for more general channel matrix models. Most of them again rely on the study of the global regime. In the following, we are going to see an example where the largest eigenvalue of random matrices plays a major role.

Model: • $2n$ nodes, distributed independently and uniformly in a square domain of area n.
($\Rightarrow$ constant density of nodes as n increases)

- nodes are paired up at random so as to form n source-destination pairs (logical links)
- individual power constraint P at each node
- attenuation of signals over distance r: $\frac{e^{i\phi}}{r^{\alpha/2}}$ where ϕ is a random phase and $\alpha \geq 2$.
- additive white Gaussian noise at each receiver

Assume now that the n S-D pairs wish to establish communication at a common rate R_n, without the help of any fixed infrastructure, but using other nodes as relays for their communication.

Question: how does the maximum achievable rate R_n scale with n? Likewise, how does the overall capacity of the network $C_n = nR_n$ scale with n?

Previous answers:

- $C_n \overset{\sim}{=} \sqrt{n}$ is achievable with multi-hop strategy (Gupta-Kumar 2000)
- for $\alpha > 4$, $\sqrt{n}$ is the best we can do (Kumar-Xie 2006)

× New result: (Ozgur-Leveque-Tse 07)

- for $2 \leq \alpha \leq 3$, $C_n = O(n^{2-\alpha/2+\varepsilon})$ $\forall \varepsilon > 0$
- for $\alpha > 3$, $C_n = O(n^{\frac{1}{2}+\varepsilon})$ $\forall \varepsilon > 0$

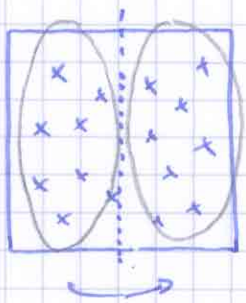
and the first bound is achievable (up to a n^{ε})

× via MIMO and hierarchical cooperation;

the second is achievable via multihop.

Proof of the two upper bounds (simultaneously)

- Cut-set argument (*) : let us divide the network into two equal parts:



we assume here full cooperation on both sides (turning the network into a MIMO channel)

What is the maximum information flow going from left to right? Since there are statistically order n S-D pairs willing to establish communication from the left-hand side to the right-hand side, this information flow is an upper bound on nR_n, i.e. is an upper bound on C_n (up to some constant).

From left to right, we have the following MIMO channel:

$$y_j = \sum_{k=1}^{n} h_{jk}\, x_k + z_j$$

(NB: questionable assumptions)

where $h_{jk} = \dfrac{e^{i\phi_{jk}}}{r_{jk}^{\alpha/2}}$, with ϕ_{jk} = i.i.d. random phases varying ergodically over time and r_{jk} = distances (fixed).

(*) = the only tool at hand for dealing with information theoretic capacity of large networks.

At this point we have to specify who knows the channel realizations h_{jk} (i.e. the phases ϕ_{jk}, since we already assume that the positions of the nodes are fixed and known to everybody).

We will study two cases:

1°) only the receivers know the phases ϕ_{jk} (not that realistic in a wireless network, since receivers can always feed these phases back to transmitters)

2°) the phases ϕ_{jk} are known to everybody.

NB: the case where the phases are not known to anybody remains an interesting open problem!

1°) By the above cut-set argument, we have

$$C_n \leq \max_{Q \geq 0 \,:\, Q_{kk} \leq P,\ \forall k} \mathbb{E}\left(\log\det\left(I + HQH^*\right)\right)$$

Since the h_{jk} are independent and $h_{jk} \sim -h_{jk}$, we have by Lemma 1 of lecture 7 that the optimal covariance matrix Q is diagonal.

But because of the individual power constraint, it is then clear that the optimal $Q = P\,I$, so that

$$C_n \le \mathbb{E}\left(\log\det(I + P H H^*)\right)$$
$$\le \mathbb{E}\left(\mathrm{Tr}(P H H^*)\right)$$

[since $\log\det(I+A) = \sum_{j=1}^n \log(1+\lambda_j^A) \underset{\uparrow}{\le} \sum_{j=1}^n \lambda_j^A = \mathrm{Tr}\,A$]
$(A \ge 0)$ $\quad$ $(\log(1+x) \le x)$

× Note moreover that the last inequality is reasonably tight, since the eigenvalues of HH^* are relatively small.

Therefore,

$$C_n \le P\,\mathbb{E}\left(\sum_{j,k=1}^n |h_{jk}|^2\right)$$
$$= P \sum_{j,k=1}^n \frac{1}{r_{jk}^\alpha} \qquad \left(\text{since } |e^{i\phi_{jk}}| = 1\right)$$

and estimating the above sum leads to

× $$\sum_{j,k=1}^n \frac{1}{r_{jk}^\alpha} \sim \begin{cases} n^{2-\alpha/2} & \text{if } 2 \le \alpha \le 3 \\ \sqrt{n} & \text{if } \alpha > 3 \end{cases} \qquad \#$$

(⚠ technical detail skipped)

<u>Interpretation</u>:

- if $\alpha \le 3$, long range MIMO communications are worth it, but still limited by the power transfer.
- if $\alpha > 3$, then multi-hop communication is optimal
  (long range communications are too onerous)

2°) If phases are known to everybody, then we only have the a priori looser upper bound:

$$C_n \le \mathbb{E}\left(\max_{Q \geq 0 :\, Q_{kk} \le P,\, \forall k} \log\det(I + HQH^*) \right)$$

since the transmitters can theoretically tune their covariance matrix to the channel realization H.
But optimizing Q for a given H is a difficult optimization problem (because of the individual power constraint, which is not unitarily invariant).
Nevertheless, we are only interested in scaling laws here; we are therefore going to show that up to some $\log n$'s, the same upper bound as before applies:

$$C_n \le \mathbb{E}\left(\max_{Q \geq 0 :\, Q_{kk} \le P,\, \forall k} \operatorname{Tr}(HQH^*) \right)$$

Notice that $\operatorname{Tr}(HQH^*) \le |||H|||_2^2 \cdot \operatorname{Tr} Q$,
where $|||H|||_2$ is the spectral norm of H.

$$\left[\begin{aligned} &\text{since } \operatorname{Tr}(HQH^*) = \mathbb{E}_x(\operatorname{Tr}(Hxx^*H^*)) = \mathbb{E}_x(x^*H^*Hx) \\ &\quad = \mathbb{E}_x(\|Hx\|_2^2) \le |||H|||_2^2 \cdot \mathbb{E}_x(\|x\|_2^2) = |||H|||_2^2 \cdot \operatorname{Tr} Q \end{aligned}\right]$$

However since $\mathrm{Tr}\, Q \leq nP$, this upper bound only gives $C_n \leq \mathbb{E}(|||H|||_2^2) \cdot nP$
and therefore fails to give the correct order (since it can easily be shown that $\mathbb{E}(|||H|||_2^2) \geq cst$).

Instead, let us consider the rescaled matrices

$$\begin{cases} \tilde{h}_{jk} := h_{jk}/\sqrt{d_k} \\ \tilde{Q}_{jk} := \sqrt{d_j}\, Q_{jk} \sqrt{d_k} \end{cases} \quad \text{where } d_k := \sum_{j=1}^n \frac{1}{r_{jk}^\alpha}$$

The upper bound now becomes:

$$C_n \leq \mathbb{E}\Big(\max_{\tilde{Q} \geq 0 :\ \tilde{Q}_{kk} \leq P d_k} \mathrm{Tr}(\tilde{H} \tilde{Q} \tilde{H}^*) \Big)$$

Using again the inequality $\mathrm{Tr}(\tilde{H}\tilde{Q}\tilde{H}^*) \leq |||\tilde{H}|||_2^2\, \mathrm{Tr}\, \tilde{Q}$, we obtain

$$C_n \leq \underbrace{\mathbb{E}(|||\tilde{H}|||_2^2)}_{\swarrow} \cdot \underbrace{P \underbrace{\sum_{k=1}^n d_k}_{= \sum_{j,k=1}^n \frac{1}{r_{jk}^\alpha} = \begin{cases} O(n^{2-\alpha/2}) & 2 \leq \alpha \leq 3 \\ O(\sqrt{n}) & \alpha > 3 \end{cases}}}_{= \text{correct order!}}$$

There remains to prove that this term does not scale faster than some $\log n$'s.

Note that $|||\tilde{H}|||_2^2 = \rho(\tilde{H}\tilde{H}^*) = \lim_{\ell\to\infty} \|(\tilde{H}\tilde{H}^*)^\ell\|^{1/\ell}$

for any matrix norm $\|\cdot\|$. Choosing $\|A\| = \sqrt{Tr(A^*A)}$,

we obtain $|||\tilde{H}|||_2^2 = \lim_{\ell\to\infty} Tr\left((\tilde{H}\tilde{H}^*)^{2\ell}\right)^{1/2\ell}$

$= \lim_{\ell\to\infty} Tr\left((\tilde{H}\tilde{H}^*)^{\ell}\right)^{1/\ell}$. Therefore,

$$\mathbb{E}\left(|||\tilde{H}|||_2^2\right) \underset{DCT}{=} \lim_{\ell\to\infty} \mathbb{E}\left(Tr\left((\tilde{H}\tilde{H}^*)^\ell\right)^{1/\ell}\right) \underset{Jensen}{\le} \lim_{\ell\to\infty} \left(\mathbb{E}\left(Tr\left((\tilde{H}\tilde{H}^*)^\ell\right)\right)\right)^{1/\ell}$$

The moment computation gives

$$\mathbb{E}\left(Tr\left((\tilde{H}\tilde{H}^*)^\ell\right)\right) \le t_\ell\, n (\log n)^{3\ell} \quad (*)$$

so

$$\mathbb{E}\left(|||\tilde{H}|||_2^2\right) \le \lim_{\ell\to\infty} \left(t_\ell^{1/\ell}\right)(\log n)^3 = 4(\log n)^3 \qquad \#$$

(*) illustration for $\ell = 2$ and a 1D regular network:

$$r_{jk} = j+k \implies d_k = \sum_{j=1}^n \frac{1}{(j+k)^\alpha} \sim k^{1-\alpha}$$

$$\implies \tilde{h}_{jk} \cong \frac{e^{i\phi_{jk}}}{(j+k)^{\alpha/2}}\sqrt{k^{\alpha-1}} \ : \ |\tilde{h}_{jk}|^2 = \frac{k^{\alpha-1}}{(j+k)^\alpha} \le \frac{1}{j+k}$$

$$\mathbb{E}\left(Tr\left((\tilde{H}\tilde{H}^*)^2\right)\right) = \sum_{j_1,j_2,k_1,k_2=1}^n \mathbb{E}\left(\tilde{h}_{j_1k_1}\overline{\tilde{h}_{j_2k_1}}\tilde{h}_{j_2k_2}\overline{\tilde{h}_{j_1k_2}}\right)$$

$$= \sum_{j,k_1\neq k_2} \mathbb{E}(|\tilde{h}_{jk_1}|^2)\,\mathbb{E}(|\tilde{h}_{jk_2}|^2) + \sum_{j_1\neq j_2,k} \mathbb{E}(|\tilde{h}_{j_1k}|^2)\,\mathbb{E}(|\tilde{h}_{j_2k}|^2)$$

$$+ \sum_{j,k} \mathbb{E}(|\tilde{h}_{jk}|^4) \le \sum_{j,k_1\neq k_2} \frac{1}{j+k_1}\frac{1}{j+k_2} + \sum_{j_1\neq j_2,k} \frac{1}{j_1+k}\frac{1}{j_2+k} + \sum_{j,k}\frac{1}{(j+k)^2}$$

$$\le 2\sum_j \Big(\underbrace{\sum_k \frac{1}{j+k}}_{\le \log n}\Big)^2 \le \underbrace{2}_{=t_2} n \underbrace{(\log n)^2}_{\text{or } (\log n)^6 \text{ for random placement of nodes (binning argument)}}$$

Randam matrix theory: lecture 19

Non-negative and positive definite matrices [ref: Horn-Johnson chap. 7]

× <u>Def</u>: • an $n\times n$ matrix A is non-negative definite

× if $x^*Ax \geq 0 \quad \forall x \in \mathbb{C}^n$; notation: $A \geq 0$

• A is moreover positive definite if $x^*Ax > 0 \quad \forall x \neq 0$;

notation: $A > 0$

× <u>Properties</u>:

• $A \geq 0 \implies A = A^*$

• $A \geq 0 \overset{(*)}{\implies} A = U\Lambda U^*$ with U $n\times n$ unitary

and $\Lambda = \mathrm{diag}(\lambda_1 \dots \lambda_n)$ with $\lambda_j \geq 0$

• $A, B \geq 0$, $\underset{(\in\mathbb{R})}{\alpha,\beta} \geq 0 \implies \alpha A + \beta B \geq 0$

• $A \geq 0$ and $\mathrm{rank}\, A = m \leq n \overset{(*)}{\implies} \exists\, V$ $n\times m$ matrix s.t. $A = VV^*$

<u>Pf</u>: $A = U\Lambda U^* = \sum_{j=1}^{n} \lambda_j u_j u_j^*$ with $\lambda_1 \dots \lambda_m > 0$, $\lambda_{m+1} = \dots = \lambda_n = 0$

define $v_j = \frac{1}{\sqrt{\lambda_j}} u_j \leftarrow$ (j^{th} column of U) for $1 \leq j \leq m$

then $A = \sum_{j=1}^{m} \underbrace{v_j v_j^*}_{\text{rank one } n\times n \text{ matrices}} = VV^*$ where $V = (v_1 | \dots | v_m)$

#

• $A > 0$ iff $A \geq 0$ and A is invertible

iff $A \geq 0$ and $\mathrm{rank}\, A = n$

iff $A = VV^*$ with V $n\times n$ invertible

(*) <u>NB</u>: the reverse implications are clearly also true!

<u>Def:</u> Schur or Hadamard product:

$$(A \circ B)_{jk} := a_{jk}\, b_{jk} \qquad A, B \ n\times n \text{ matrices}$$

<u>Property:</u> $A, B \geq 0 \implies A \circ B \geq 0$

<u>Pf:</u> $A = VV^*$ for some $n\times m$ V, $B = WW^*$ for some $n \times p$ W

i.e. $a_{jk} = \sum_{\ell=1}^{m} V_{j\ell} \overline{V_{k\ell}}$, $b_{jk} = \sum_{\ell'=1}^{p} W_{j\ell'} \overline{W_{k\ell'}}$

so $(A\circ B)_{jk} = \sum_{\ell,\ell'=1}^{m,p} \underbrace{(V_{j\ell} W_{j\ell'})}_{= y_j^{(\ell,\ell')}} \underbrace{\overline{(V_{k\ell} W_{k\ell'})}}_{= \overline{y_k^{(\ell,\ell')}}}$

i.e. $A \circ B$ is a sum of rank one non-negative def. matrices #

However, it is <u>not</u> true that $A, B \geq 0 \implies A \cdot B \geq 0$ (usual matrix product)

This is because:

$$A = A^*, B = B^* \not\Longrightarrow AB = (AB)^* \quad (\text{true only if } AB = BA)$$

<u>Examples:</u>

- $a_{jk} = e^{i(\phi_j - \phi_k)}$ $\phi_j \in \mathbb{R}$ $\implies A = vv^* \geq 0$, with $v_j = e^{i\phi_j}$
- $b_{jk} = \frac{1}{x_j + x_k} = \int_0^\infty e^{-t(x_j + x_k)}\, dt \implies B \geq 0$ (Riemann sum of non-neg. def. matrices)
- $(A\circ B)_{jk} = \frac{e^{i(\phi_j - \phi_k)}}{x_j + x_k} \implies (A \circ B) \geq 0$
- $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $B = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \implies A\cdot B = \begin{pmatrix} 2 & 1 \\ 2 & 1 \end{pmatrix} \neq (AB)^*$

$\implies A \cdot B \not\geq 0$ (in the "complex" sense)

Simultaneous diagonalization (and its applications)

Thm: if $A > 0$ and $B = B^*$ (both $n \times n$ matrices),

then $\exists C$ $n\times n$ invertible such that $CAC^* = I$, $CBC^* = D$ diagonal

Pf: we know that $\exists V$ $n \times n$ invertible such that $A = VV^*$, i.e. $V^{-1}A(V^{-1})^* = I$

since $V^{-1}B(V^{-1})^*$ is Hermitian, $\exists U$ $n\times n$ unitary such that

$UV^{-1}B(V^{-1})^*U^* = D$ diagonal. Let therefore $C = UV^{-1}$:

we have both $CBC^* = D$ and $CAC^* = UIU^* = I$ #

Thm: the map $A \mapsto \log\det A$ is concave on the set

of $n\times n$ positive definite matrices.

Pf: We need to check that for any $\alpha \in (0,1)$, $A, B > 0$:

$$\log\det(\alpha A + (1-\alpha)B) \geq \alpha \log\det A + (1-\alpha)\log\det B \quad (1)$$

By simultaneous diagonalization, $\exists C$ $n\times n$ invertible such

× that $A = CC^*$ and $B = CDC^*$. Thus, (1) is equivalent to:

$$\log\det(\alpha I + (1-\alpha)D) \geq \alpha \log\det I + (1-\alpha)\log\det D$$

i.e. $\sum_{j=1}^{n} \log(\alpha + (1-\alpha)d_j) \geq 0 + (1-\alpha)\sum_{j=1}^{n}\log d_j$

× which holds because of the concavity of the log itself. #

Corollary: $A \mapsto \log\det A^{-1}$ is convex on $\{A > 0\}$

NB: more generally, $A \mapsto \mathrm{Tr}\, f(A)$ is convex on $\{A > 0\}$

if $f: \mathbb{R}_+ \to \mathbb{R}$ is convex

Partial order on $\{A \geq 0\}$

<u>Def:</u> $A \geq B$ if $A - B \geq 0$

Note that $A \leq cI$ iff all e.v. of A are less than or equal to c

<u>Thm:</u> a) $A \geq B > 0 \implies \operatorname{Tr} A \geq \operatorname{Tr} B$

b) $A \geq B > 0 \implies \det A \geq \det B$ (& same is true for log det)

<u>Pf:</u> By simultaneous diagonalization, $\exists\, C$ $n \times n$ invertible s.t.

$A = CC^*$ and $B = CDC^*$, so

exercise!

× $A \geq B$ iff $C(I - D)C^* \geq 0$ $\boxed{\text{iff}}$ $(I - D) \geq 0$

iff all $d_j \leq 1$

a) $\operatorname{Tr} A = \operatorname{Tr} CC^* = \sum_{j,k=1}^{n} |c_{jk}|^2$

$\operatorname{Tr} B = \operatorname{Tr}(CDC^*) = \sum_{j,k=1}^{n} d_k |c_{jk}|^2 \leq \sum_{j,k=1}^{n} |c_{jk}|^2 = \operatorname{Tr} A$ ✓

b)$^{(*)}$ $BA^{-1} = CDC^*(CC^*)^{-1} = CDC^{-1}$

BA^{-1} and D have therefore the same e.v., namely $d_j \leq 1$ (and $d_j \geq 0$)

so $\det(BA^{-1}) \leq 1$ i.e. $\det B \leq \det A$ ✓

\#

Note that "reciprocally", if $A \geq 0$ and $\operatorname{Tr} A \leq c$,

then all e.v. of A are less than or equal to c (since their sum is), i.e. $A \leq cI$

[but note also that $\operatorname{Tr}(cI) = nc$, not c]

(*) <u>or</u>: $\det B = \det(CDC^*) = \det(C^*C)\underbrace{\det D}_{\leq 1} \leq \det(C^*C) = \det A$ ✓

Further matrix inequalities

Hadamard's inequality (version I)

If $A \geq 0$, then $\det A \leq \prod_{j=1}^{n} a_{jj}$. (and equality holds if A is diagonal)

Proof

- If $\det A = 0$, then there is nothing to prove.
- If $\det A \neq 0$, then A is invertible, so all $a_{jj} > 0$

(recall that for $A \geq 0$, $a_{jj} = 0 \Rightarrow a_{jk} = 0 \ \forall k \neq j$)

Let $D = \mathrm{diag}(a_{11}^{-1/2}, \ldots, a_{nn}^{-1/2})$: $\det A \leq \prod_{j=1}^{n} a_{jj}$ iff $\det(DAD) \leq 1$, so we may as well assume that all diagonal entries of A are equal to 1, in which case:

$$\det A = \prod_{j=1}^{n} \lambda_j \underset{\substack{\uparrow \\ \text{arithmetic-geometric} \\ \text{mean inequality}}}{\leq} \left(\frac{1}{n}\sum_{j=1}^{n} \lambda_j\right)^{1/n} = \left(\frac{1}{n}\operatorname{Tr} A\right)^{1/n} = 1^{1/n} = 1$$

(or concavity of the log) #

Note on the other hand that it is not true that

$$A \leq \mathrm{diag}(a_{11}, \ldots, a_{nn})$$

even in the case where all a_{jj} are equal.

Counter-ex: A = all ones matrix: $I - A \not\geq 0$

(but $\det A = 0 \leq 1 = \det I$)

Hadamard's inequality (version II)

For any $n \times n$ matrix B,

$$|\det B| \le \prod_{j=1}^{n} \sqrt{\sum_{k=1}^{n} |b_{jk}|^2} \quad \left(\begin{array}{l}\text{and equality holds if the}\\ \text{rows of } B \text{ are orthogonal}\end{array}\right)$$

Proof

Let $A = BB^*$: $|\det B| = \sqrt{\det A} \le \prod_{j=1}^{n} \sqrt{a_{jj}} = \prod_{j=1}^{n} \sqrt{\sum_{k=1}^{n} |b_{jk}|^2}$ #

Block Hadamard or Fischer's inequality

Let A be a $n \times n$ matrix, B be a $n \times m$ matrix, C be a $m \times m$ matrix such that $M = \begin{pmatrix} A & B \\ B^* & C \end{pmatrix}$ is positive definite

(this is true iff $A > 0$ and $C - B^*A^{-1}B > 0$)

Then $\det M \le \det A \cdot \det C$

Proof

Let $X = -A^{-1}B$. We have

$$\begin{pmatrix} I & 0 \\ X^* & I \end{pmatrix} \begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \begin{pmatrix} I & X \\ 0 & I \end{pmatrix} = \begin{pmatrix} A & 0 \\ 0 & C - B^*A^{-1}B \end{pmatrix}$$

so $\det \begin{pmatrix} A & B \\ B^* & C \end{pmatrix} = \underbrace{\det A \cdot \det (C - B^*A^{-1}B)}$

Since $C - \underbrace{B^*A^{-1}B}_{\ge 0} \le C$, we have $\le \det A \cdot \det C$ #

This inequality generalizes in an obvious manner to the situation with more than 2 blocks.

Further inequalities (without proofs)

- Oppenheim's inequality:

$$A, B \geq 0 \Rightarrow \det(A \circ B) \geq (\det A) \cdot \prod_{j=1}^{n} b_{jj}$$

(with $B = I$, we recover Hadamard's inequality)

corollary: $\det(A \circ B) \geq \det A \cdot \det B$

- By concavity of $A \mapsto \log \det A$, we have:

$$\begin{cases} A, B > 0 \\ \alpha \in (0,1) \end{cases} \Rightarrow (\det A)^{\alpha} (\det B)^{1-\alpha} \leq \det(\alpha A + (1-\alpha) B)$$

- Minkowski's inequality:

$$\begin{cases} A, B > 0 \\ n \times n \text{ matrices} \end{cases} \Rightarrow (\det(A+B))^{1/n} \geq (\det A)^{1/n} + (\det B)^{1/n}$$

and equality holds iff $B = cA$ for some $c > 0$ (scalar)

- Lieb's inequality:

$$A, B \geq 0 \Rightarrow \log \det(I + A + B) \leq \log \det(I + A) + \log \det(I + B)$$

- Finally, note this additional property:

for any $m \times n$ matrix B, the map

$A \mapsto \log \det(I + B A^{-1} B^*)$ is convex

on the set of $n \times n$ positive definite matrices

Random matrix theory: lecture 20

Matrix inequalities: proofs from information theory

[ref: Dembo-Cover-Thomas, Diggavi-Cover]

Hadamard's inequality

Let $X=(X_1..X_n)$ be a complex random vector $\left(\begin{matrix}\text{assuming a}\\ \text{joint pdf exists}\end{matrix}\right)$.

Then $h(X_1..X_n) \leq \sum_{j=1}^{n} h(X_j)$

(obtained by the chain rule and the fact that conditioning reduces entropy)

Take $X \sim N_{\mathbb{C}}(0,A)$, with $A>0$. The above inequality reads: $\log\det(\pi e A) \leq \sum_{j=1}^{n} \log(\pi e\, a_{jj})$

i.e. $\det A \leq \prod_{j=1}^{n} a_{jj}$ #

Fischer's inequality

Same idea: $h(X_1..X_m, X_{m+1}..X_{m+n}) \leq h(X_1..X_m) + h(X_{m+1}..X_{m+n})$

Take $X \sim N_{\mathbb{C}}\left(0, \begin{pmatrix} A & B \\ B^* & C \end{pmatrix}\right)$ with $A>0$ $m\times m$, $C>0$ $n\times n$ and $\underset{(m\times n)}{B}$ such that $\begin{pmatrix} A & B \\ B^* & C \end{pmatrix} > 0$. The above reads:

$$\log\det\left(\pi e \begin{pmatrix} A & B \\ B^* & C \end{pmatrix}\right) \leq \log\det(\pi e A) + \log\det(\pi e C)$$

i.e. $\det\begin{pmatrix} A & B \\ B^* & C \end{pmatrix} \leq \det A \cdot \det C$ #

The application $A \mapsto \log\det A$ is concave on $\{A > 0\}$:

Let $X \sim N_{\mathbb{C}}(0, A)$, $Y \sim N_{\mathbb{C}}(0, B)$ be independent

Let Θ be indep. of both X and Y and be such that

$$\mathbb{P}(\Theta = 1) = \alpha = 1 - \mathbb{P}(\Theta = 0)$$

Let also $Z = \begin{cases} X & \text{if } \Theta = 1, \\ Y & \text{if } \Theta = 0. \end{cases}$ (NB: Z is not Gaussian)

(1) Then $h(Z) \geq h(Z \mid \Theta) = \alpha\, h(Z \mid \Theta = 1) + (1-\alpha)\, h(Z \mid \Theta = 0)$

$$= \alpha\, h(X) + (1-\alpha)\, h(Y)$$

(2) Now, $h(Z) \leq h(Z_G)$, where Z_G is Gaussian with the same covariance matrix Q as Z:

$$Q = \mathbb{E}(ZZ^*) = \alpha\, \mathbb{E}(XX^*) + (1-\alpha)\, \mathbb{E}(YY^*)$$

$$= \alpha A + (1-\alpha) B$$

(1+2) $\Rightarrow$ $\alpha \log\det(\pi e A) + (1-\alpha) \log\det(\pi e B)$

$$\leq \log\det(\pi e(\alpha A + (1-\alpha) B))$$

i.e. $\alpha \log\det(A) + (1-\alpha)\log\det B \leq \log\det(\alpha A + (1-\alpha) B)$

#

The application $A \mapsto \log\det(I + BA^{-1}B^*)$ is convex

Lemma: "The Gaussian noise is the worst noise" (without proof)

Let $X \sim N_{\mathbb{C}}(0, Q)$ and Z, Z_G be random vectors independent of X, with the same covariance matrix A. If $Z_G \sim N_{\mathbb{C}}(0, A)$, then $I(X; X+Z) \geq I(X; X+Z_G)$.

Let now $Z_1 \sim N_{\mathbb{C}}(0, A_1)$ and $Z_2 \sim N_{\mathbb{C}}(0, A_2)$ be independent and Θ be independent of both Z_1 & Z_2 and such that $\mathbb{P}(\Theta = 1) = \alpha = 1 - \mathbb{P}(\Theta = 0)$. Let also $Y = \begin{cases} X + Z_1 & \text{if } \Theta = 1 \\ X + Z_2 & \text{if } \Theta = 0 \end{cases}$

(1) Then $I(X; Y) \leq I(X; Y, \Theta) = \overbrace{I(X; \Theta)}^{=0} + I(X; Y \mid \Theta)$

$$= \alpha I(X; Y \mid \Theta = 1) + (1-\alpha) I(X; Y \mid \Theta = 0)$$

$$= \alpha I(X; X + Z_1) + (1-\alpha) I(X; X + Z_2)$$

$$= \alpha \log\det(I + QA_1^{-1}) + (1-\alpha) \log\det(I + QA_2^{-1})$$

(2) Also, $I(X; Y) \geq I(X; Y_G)$ where $Y_G = X + Z_G$ and $Z_G \sim N_{\mathbb{C}}(0, A)$ with $A = \alpha A_1 + (1-\alpha) A_2$

$$\Rightarrow I(X; Y) \geq \log\det\left(I + Q(\alpha A_1 + (1-\alpha) A_2)^{-1}\right)$$

(1+2) $\Rightarrow$ convexity of $A \mapsto \log\det(I + QA^{-1})$

$Q = B^* B \Rightarrow$ convexity of $A \mapsto \log\det(I + BA^{-1}B^*)$ #

Minkowski's inequality

Def: for X a continuous complex random $\overbrace{\text{vector}}^{\text{of dimension } n}$, we set

$$N(X) := \frac{1}{\pi e} \exp\left(\frac{1}{n} h(X)\right).$$

Entropy power inequality: if X and Y are independent,

then $N(X+Y) \geq N(X) + N(Y)$.

So if $X \sim N_{\mathbb{C}}(0, A)$ and $Y \sim N_{\mathbb{C}}(0, B)$, then

$$N(X) = \frac{1}{\pi e} \exp\left(\frac{1}{n} \log \det(\pi e A)\right) = (\det A)^{1/n}$$

$$\Rightarrow (\det(\underbrace{A+B}_{=\mathrm{Cov}(X+Y)}))^{1/n} \geq (\det A)^{1/n} + (\det B)^{1/n}$$

#

Lieb's inequality

Let X, Y, Z be three independent random vectors

$Y \to Y + X \to Y + X + Z$ forms a Markov chain,

so by the data processing inequality:

$$I(Y; X+Y+Z) \leq I(Y; X+Y)$$

i.e. $h(X+Y+Z) - \underbrace{h(X+Y+Z \mid Y)}_{= h(X+Z)} \leq h(X+Y) - \underbrace{h(X+Y \mid Y)}_{= h(X)}$

i.e. $h(X+Y+Z) + h(X) \leq h(X+Y) + h(X+Z)$

Consider $X \sim N_{\mathbb{C}}(0, I)$, $Y \sim N_{\mathbb{C}}(0, A)$, $Z \sim N_{\mathbb{C}}(0, B)$

$$\Rightarrow \log\det(I + A + B) + \underset{(= \log\det(I))}{0} \leq \log\det(I + A) + \log\det(I + B)$$

#

Randan matrix theory: Lecture 21

Gaussian randan matrices and free probability

We have already seen that the following result holds (lecture 16):

- Let $A^{(n)} := \mathrm{diag}(a_1 \dots a_n)$, with $a_j \in \mathbb{R}$ (deterministic), be such that $F_n^A(t) := \frac{1}{n} \# \{1 \le j \le n : a_j \le t\} \xrightarrow[n\to\infty]{} F^A(t)$, with corresponding Stieltjes transform $g_A(z)$.
- Let H be a $n\times n$ real symmetric matrix with iid $\sim N_{\mathbb{R}}(0,1)$ entries in the upper triangular part, and $H^{(n)} = \frac{1}{\sqrt{n}} H$. (GOE model)
- Let $B^{(n)} := A^{(n)} + H^{(n)}$ and $\lambda_j^{(n)}$ be the e.v. of $B^{(n)}$. Then

$$F_n^B(t) := \frac{1}{n} \# \{ j : \lambda_j^{(n)} \le t \} \xrightarrow[n\to\infty]{} F^B(t) \quad \text{a.s.}$$

whose Stieltjes transform $g_B(z)$ satisfies $\boxed{g_B(z) = g_A(z + g_B(z))}$

Generalizations of this result:

- The result still holds if $A^{(n)}$ is deterministic and real symmetric with eigenvalues $a_1^{(n)} \dots a_n^{(n)}$ and $F_n^A(t) = \frac{1}{n} \# \{ j : a_j^{(n)} \le t \} \xrightarrow[n\to\infty]{} F^A(t)$

Proof:

$\exists\, O^{(n)}$ orthogonal and $D^{(n)} = \mathrm{diag}(a_1^{(n)} \dots a_n^{(n)})$ s.t. $A^{(n)} = O^{(n)} D^{(n)} (O^{(n)})^T$

$$\Rightarrow B^{(n)} = O^{(n)} \underbrace{\left(D^{(n)} + \underbrace{(O^{(n)})^T H^{(n)} O^{(n)}}_{\text{same dist. as } H^{(n)}} \right)}_{\text{same e.v. as } B^{(n)}} (O^{(n)})^T$$

\#

see lecture 2

- The result still holds if $A^{(n)}$ is <u>random</u> and <u>independent</u> of $H^{(n)}$ with $F_n^A(t) \to F^A(t)$ a.s.

<u>Proof</u>:

× Conditioned on $A^{(n)}$, the result holds since $F_n^A(t) \to F^A(t)$ a.s. and F^A is deterministic #

<u>Remark</u>:

The result still holds for H with non-Gaussian entries, (i.e. non-orthogonally invariant) but this requires further work.

In lecture 16, we have also seen a result of the same flavor:

- Let $A^{(n)}$ be real symmetric & independent of H such that $F_n^A(t) \xrightarrow[n\to\infty]{} F^A(t)$ a.s. with Stieltjes transform $g_A(z)$.
- Let H be $n \times n$ with iid $\sim N_{\mathbb{R}}(0,1)$ entries and $W^{(n)} = \frac{1}{n} HH^T$.
- Let $B^{(n)} = A^{(n)} + W^{(n)}$. Then $F_n^B(t) \xrightarrow[n\to\infty]{} F^B(t)$ a.s., whose Stieltjes transform $g_B(z)$ satisfies

$$\boxed{g_B(z) = g_A\left(z - \frac{1}{1+g_B(z)}\right)}$$

Question

Is there a general rule for computing the limiting eigenvalue distribution of the sum of two independent random matrices $A^{(n)} + B^{(n)}$?

Answer 1

× • A particular case of independent random matrices are deterministic matrices; and in this case, we know that there is no simple rule for computing the eigenvalues of $A^{(n)} + B^{(n)}$ from the eigenvalues of $A^{(n)}$ and $B^{(n)}$ separately, mainly because of the fact that they do not share the same eigenvectors in general.

• Moreover, even in the case where $A^{(n)}$ and $B^{(n)}$ share the same eigenvectors (when they are both diagonal, or both circulant, e.g.), everything is possible regarding the limiting eigenvalue distribution of $A^{(n)} + B^{(n)}$.

Example:

- Let $A^{(n)} = \mathrm{diag}\left(\frac{1}{n}, \frac{2}{n}, \dots, \frac{n-1}{n}, \frac{n}{n}\right) = B^{(n)}$.
  Then the limiting eigenvalue distribution of both $A^{(n)}$ and $B^{(n)}$ is the uniform distribution on $[0,1]$.
  Also, $A^{(n)} + B^{(n)} = \mathrm{diag}\left(\frac{2}{n}, \frac{4}{n} \dots, \frac{2n}{n}\right)$ has for limiting eigenvalue distribution the uniform distribution on $[0,2]$.
- Let now $\widetilde{B}^{(n)} = \mathrm{diag}\left(\frac{n}{n}, \frac{n-1}{n}, \dots \frac{2}{n}, \frac{1}{n}\right)$
  Then the limiting eigenvalue distribution of $\widetilde{B}^{(n)}$ is also the uniform distribution on $[0,1]$, but $A^{(n)} + \widetilde{B}^{(n)} = \mathrm{diag}\left(\frac{n+1}{n}, \frac{n+1}{n}, \dots, \frac{n+1}{n}\right)$ has for limiting eigenvalue distribution the Dirac distribution at point $x=1$.

In order to find a general rule for the limiting eigenvalue distribution of sums of random matrices, we need therefore to find a more restrictive condition than the independence of $A^{(n)}$ and $B^{(n)}$.

Important observation

When dealing with distributions of (eigenvalues of) random matrices, the framework of classical probability is not the best one, since any two classical random variables X and Y commute: $XY = YX$, but the same is not true for random matrices.

$\Rightarrow$ Non-commutative probability

Let A be the set of $n\times n$ $\overbrace{\text{matrices}}^{\text{(real)}}$; A is a non-commutative algebra, with addition $A+B$, multiplication $A\cdot B$ and unit element $A=I$. (*)

Def: an expectation on A is an application $\varphi : A \to \mathbb{R}$ s.t.

- φ is linear: $\varphi(A + cB) = \varphi(A) + c\,\varphi(B)$
- $\varphi(I) = 1$
- $\varphi(A) \geq 0$ if $A \geq 0$

Examples:

- $\varphi(A) := \frac{1}{n} \operatorname{Tr} A$
- $\varphi(A) := a_{11}$

(*) and matrices are called non-commutative random variables.

Remarks:

- The set of classical random variables also forms an algebra, which is moreover commutative.
- So far, non-commutative random variables are $n \times n$ deterministic matrices (random matrices will come later).

What is the distribution of a non-commutative r.v.?

- The "distribution" of a matrix A is defined through its moments: $m_k = \varphi(A^k)$, $k \geq 0$, but there is in general no corresponding classical distribution μ_A on $\mathbb{R}$.
- For $\varphi(A) = \frac{1}{n} \cdot \operatorname{Tr} A$ and A real symmetric, there is:

$$\mu_A = \frac{1}{n} \sum_{j=1}^{n} \delta_{\lambda_j^A}, \quad \text{where } \lambda_j^A = \text{e.v. of } A$$

$$\Rightarrow \begin{cases} m_k^A = \int_{\mathbb{R}} x^k \, d\mu_A(x) = \frac{1}{n} \sum_{j=1}^{n} (\lambda_j^A)^k = \frac{1}{n} \operatorname{Tr}(A^k) = \varphi(A^k). \\ g_A(z) = \int_{\mathbb{R}} \frac{1}{x-z} \, d\mu_A(z) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{\lambda_j^A - z} \end{cases}$$

$$= \frac{1}{n} \operatorname{Tr}\left((A - zI)^{-1}\right) = \varphi\left((A - zI)^{-1}\right), \quad z \in \mathbb{C} \setminus \mathbb{R}.$$

- Note moreover that in this case, we have

$$\varphi(AB) = \frac{1}{n} \operatorname{Tr}(AB) = \frac{1}{n} \operatorname{Tr}(BA) = \varphi(BA).$$

Random matrix theory: lecture 22

Free probability (cont'd) [ref: Voiculescu]

Reminder: non-commutative probability

- A algebra with unit element I and involution $A \mapsto A^T$.
- $\varphi : A \to \mathbb{R}$ linear and positive application s.t. $\varphi(I) = 1$

Example:

$A = \{ n \times n$ real (deterministic) matrices $\}$

Let $\rho \in A$ be s.t. $\rho \geq 0$ and $Tr(\rho) = 1$ ("density matrix")

$\varphi(A) := Tr(\rho A)$

sub-examples: $\rho = \begin{pmatrix} 1 & & 0 \\ & 0 & \\ 0 & & \ddots \, 0 \end{pmatrix} \Rightarrow \varphi(A) = a_{11}$

$\rho = \frac{1}{n} I \Rightarrow \varphi(A) = \frac{1}{n} Tr A$

Terminology:

- elements of A are non-commutative random variables
- $\varphi(A)$ is the expectation of A
- the distribution of A is specified through its moments $M_k^A = \varphi(A^k)$, $k \geq 0$
  (NB: in the present framework, these always exist)

Note that in general, $\varphi(ABAB) \neq \varphi(A^2 B^2)$.

Def: classical independence

Two non-commutative random variables A and B are classically independent if:

- they commute, i.e. $AB = BA$
- $\varphi(A^k B^\ell) = \varphi(A^k)\varphi(B^\ell) \quad \forall k, \ell \geq 0$

Def: free independence

Two non-commutative random variables A and B are freely independent if for any $k \geq 0$ and any polynomials $P_1 .. P_m$, $Q_1 .. Q_m$ such that $\varphi(P_j(A)) = \varphi(Q_j(A)) = 0 \quad \forall 1 \leq j \leq m$, we have $\varphi(P_1(A) Q_1(B) P_2(A) Q_2(B) \ldots P_m(A) Q_m(B)) = 0$

[NB: $P_1(A)$ or $Q_m(B)$ might be replaced by I also]

Example: if A, B are freely independent, then

$$\varphi((A - \varphi(A)I)(B - \varphi(B)I)(A - \varphi(A)I)(B - \varphi(B)I)) = 0$$

since $\varphi(A - \varphi(A)I) = \varphi(A) - \varphi(A)\underbrace{\varphi(I)}_{=1} = 0$, and

similarly: $\varphi(B - \varphi(B)I) = 0$.

We will see below that the notion of free independence is too restrictive in the context of classical probability (i.e. commuting random variables).

Lemma 1

If A and B commute and are freely independent, then they are classically independent.

Proof:

Consider $m=1$ & $P_1(A)=A^k-\varphi(A^k)I$, $Q_1(B)=B^\ell-\varphi(B^\ell)I$

A,B free, $\varphi(P_1(A))=\varphi(Q_1(B))=0 \Rightarrow \varphi(P_1(A)Q_1(B))=0$

i.e. $\varphi((A^k-\varphi(A^k)I)(B^\ell-\varphi(B^\ell)I))=0$

i.e. $\varphi(A^kB^\ell)=\varphi(A^k)\varphi(B^\ell)\quad \forall k,\ell\geq 0$ #

Lemma 2

If A and B are freely independent, then

$$\varphi(A^2B^2)-\varphi(ABAB)=(\varphi(A^2)-\varphi(A)^2)(\varphi(B^2)-\varphi(B)^2)$$

Proof:

- A,B free $\Rightarrow \varphi((A^2-\varphi(A^2)I)(B^2-\varphi(B^2)I))=0$

  i.e. $\varphi(A^2B^2)=\varphi(A^2)\varphi(B^2)$

- define now $A_0=A-\varphi(A)I$, $B_0=B-\varphi(B)I$: $\varphi(A_0)=\varphi(B_0)=0$

  so $\underbrace{\varphi(A_0B_0A_0B_0)}=0$

$$=\varphi(AB_0A_0B)-\varphi(A)\underbrace{\varphi(B_0A_0B_0)}_{=0}-\underbrace{\varphi(A_0B_0A_0)}_{=0}\varphi(B)$$

$$-\varphi(A)\underbrace{\varphi(B_0A_0)}_{=0}\varphi(B)$$

6.6.07

5

In turn,

$$0 = \varphi(A B_0 A_0 B) = \varphi(ABAB) - \overbrace{\varphi(A^2B)}^{=\varphi(A^2)\varphi(B)}\varphi(B) - \overbrace{\varphi(AB^2)}^{=\varphi(A)\varphi(B^2)}\varphi(A)$$

$$+ \varphi(B)\underbrace{\varphi(AB)}_{=\varphi(B)\varphi(A)}\varphi(A)$$

So $\varphi(ABAB) = \varphi(A^2)\varphi(B)^2 + \varphi(A)^2\varphi(B^2) - \varphi(A)^2\varphi(B)^2$

and $\varphi(ABAB) - \varphi(A^2B^2) = (\varphi(A^2) - \varphi(A)^2)(\varphi(B^2) - \varphi(B)^2)$ #

Corollary 1

If A and B commute and are freely independent, then $0 = \varphi(A^2B^2) - \varphi(ABAB) = (\varphi(A^2) - \varphi(A)^2)(\varphi(B^2) - \varphi(B)^2)$ so either $\varphi(A^2) - \varphi(A)^2 = 0$ or $\varphi(B^2) - \varphi(B)^2 = 0$.

Corollary 2

× If A and B are classical random variables and are freely independant, then either $Var(A) = \varphi(A^2) - \varphi(A)^2 = 0$ or $Var(B) = \varphi(B^2) - \varphi(B)^2 = 0$ i.e. one of the two random variables is a constant. In classical probability, there are therefore only trivial examples of free random variables.

Lemma 3

If A and B are freely independant, then the sequence $(\varphi((A+B)^k), k \geq 0)$ is entirely determined by the two sequences $(\varphi(A^k), k \geq 0)$ and $(\varphi(B^k), k \geq 0)$, i.e. the distribution of $A+B$ is determined by the distribution of A and the distribution of B separately.

Remark: This is to relate to what we know in classical probability: if two random variables X and Y are independent, then the distribution of $X+Y$ is determined by the distributions of X and Y; more precisely, the distribution of $X+Y$ is the (classical) convolution:

$$\mu_{X+Y} = \mu_X * \mu_Y$$

Proof of the above lemma: (main idea)

The lemma can be shown by induction on k. The main idea is that
- $\varphi((A+B)^k)$ can be expanded into terms of the form $\varphi(A^{k_1} B^{k_2} A^{k_3} B^{k_4} \dots)$
- such terms can in turn be expanded into lower order terms, using free independence. (see e.g. lemma 2)

#

Terminology:

The distribution of $A+B$ is called the additive free convolution of the distributions of A and B.

Question: In classical probability, the Fourier transform possesses the nice property that if μ and ν are two distributions, then $\phi_{\mu * \nu}(t) = \phi_\mu(t) \cdot \phi_\nu(t)$. Is there such a function for the free additive convolution?

Answer: yes; the so-called R-transform.

Defs:

- For $z \in \mathbb{C} \setminus \mathbb{R}$, we define the Stieltjes transform of A: (satisfying $A = A^T$)

$$g_A(z) := \varphi\left((A - zI)^{-1} \right)$$

- The R-transform of A is then defined as:

$$R_A(z) := g_A^{-1}(-z) - \frac{1}{z}$$ (well defined $\forall z \in \mathbb{C}$, actually)

(= inverse function, $\neq \frac{1}{g_A(z)}$)

Proposition (to be proven next time)

If A and B are freely independent, then $R_{A+B}(z) = R_A(z) + R_B(z)$.

(so to be precise, $R_A(z)$ plays the role of $\log \phi_\mu(t)$ here)

Example

Catalan numbers

- If B is distributed according to the semi-circle distribution (i.e. $m_k = \varphi(B^k) = \begin{cases} r_\ell & \text{if } k=2\ell \\ 0 & \text{if } k=2\ell+1 \end{cases}$)
then we know that its Stieltjes transform satisfies the equation:

$$g_B(z)^2 + z\, g_B(z) + 1 = 0$$

i.e. $z = -g_B(z) - \frac{1}{g_A(z)}$

i.e. $g_B^{-1}(z) = -z - \frac{1}{z}$

× i.e. $R_B(z) = g_B^{-1}(-z) - \frac{1}{z} = z + \frac{1}{z} - \frac{1}{z} = z$

- So in this case, $R_{A+B}(z) = R_A(z) + z$

i.e. $g_{A+B}^{-1}(-z) - \frac{1}{z} = z + g_A^{-1}(-z) - \frac{1}{z}$

$-z \to g_{A+B}(z)$

i.e. $z = -g_{A+B}(z) + g_A^{-1}(g_{A+B}(z))$

i.e. $g_A(z + g_{A+B}(z)) = g_{A+B}(z)$

which was the composition rule found (in the limit $n \to \infty$) for the addition of a random matrix A and a GOE matrix $\mathbf{B}$, assumed to be independent.

Connection with random matrices

The connection follows from the following important proposition of Voiculescu:

Proposition

- Let $A^{(n)}$ and $B^{(n)}$ be two $\underbrace{\text{independent random}}_{n\times n \text{ real symmetric}}$ matrices such that $F_n^A(t) \xrightarrow[n\to\infty]{} F^A(t)$ a.s., $F_n^B(t) \xrightarrow[n\to\infty]{} F^B(t)$ a.s.,

× with corresponding R-transforms $R_A(z)$ and $R_B(z)$.

- Let moreover $V^{(n)}$ be a $n\times n$ orthogonal matrix distributed according to the Haar distribution and independent of both $A^{(n)}$ and $B^{(n)}$.
- Then $A^{(n)}$ and $\tilde{B}^{(n)} = V^{(n)} B^{(n)} (V^{(n)})^T$ are asymptotically freely independent, i.e. $\forall m$,

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E}\left(\mathrm{Tr}\left(P_1(A^{(n)}) Q_1(\tilde{B}^{(n)}) \dots P_m(A^{(n)}) Q_m(\tilde{B}^{(n)})\right)\right) = 0$$

whenever $P_1 .. P_m$, $Q_1 .. Q_m$ are polynomials such that (*)

$$\lim_{n\to\infty} \frac{1}{n} \mathbb{E}\left(\mathrm{Tr}\, P_j(A^{(n)})\right) = \lim_{n\to\infty} \frac{1}{n} \mathbb{E}\left(\mathrm{Tr}\, Q_j(\tilde{B}^{(n)})\right) = 0 \quad \forall 1\le j\le m$$

× • As a consequence, the limiting eigenvalue distribution of $A^{(n)} + \tilde{B}^{(n)} = A^{(n)} + V^{(n)} B^{(n)} (V^{(n)})^T$ is the distribution whose R-transform is given by $R_A(z) + R_B(z)$.

(*) here also, $P_1(A^{(n)}) = I$ or $Q_m(\tilde{B}^{(n)}) = I$ are also valid

Remarks

- The above proposition therefore also applies to $A^{(n)} + H^{(n)}$, where $A^{(n)}$ and $H^{(n)}$ are independent and $H^{(n)}$ is from the GOE (since $H^{(n)} = V^{(n)} \Lambda^{(n)} (V^{(n)})^T$, with $V^{(n)}$ distributed according to the Haar distribution, and $A^{(n)}$ is independent of both $\Lambda^{(n)}$ and $V^{(n)}$, which are in turn mutually independent (see Lecture 2)).

- We have found here a composition rule for random matrices whose eigenvectors have no relation whatsoever between them. This is coherent with the fact seen last time that whenever two matrices share the same set of eigenvectors, then no general composition rule exists.

Randan matrix theory: lecture 23

Reminder

- A, B free non-commutative randan variables, with corresponding distributions $\mu_A := (m_k^A, k \geq 0)$, $\mu_B := (m_k^B, k \geq 0)$

$\Rightarrow$ the distribution of $A+B$ given by $\mu_{A+B} := (m_k^{A+B}, k \geq 0)$ is the free additive convolution of μ_A and μ_B (denoted as $\mu_{A+B} = \mu_A \boxplus \mu_B$), which can be computed using the R-transform: $R_{A+B}(z) = R_A(z) + R_B(z)$.

- $A^{(n)}, B^{(n)}$ independent random matrices with limiting eigenvalue distributions μ_A, μ_B (nxn real symmetric)

$\Rightarrow$ let $V^{(n)}$ be orthogonal (Haar distributed) & independent of both $A^{(n)}$ and $B^{(n)}$; then $A^{(n)}$ & $V^{(n)} B^{(n)} (V^{(n)})^T$ are asymptotically free so the limiting eigenvalue distribution of $A^{(n)} + V^{(n)} B^{(n)} (V^{(n)})^T$ is given by $\mu_{A+B} = \mu_A \boxplus \mu_B$ and can be computed via the R-transform.

Today's program: [ref: Haagerup, Thorbjörnsen]

- construction of free random variables
- proof of the R-transform additivity rule.
- (free multiplicative convolution)

Construction of free random variables

Preliminary:

x • An $n \times n$ matrix A "is" a linear application $\mathbb{C}^n \to \mathbb{C}^n$.
It is entirely determined by its action on the elements of any basis of $\mathbb{C}^n$ (e.g. $e_1, .., e_n$):

• More generally, let $\mathcal{T}$ be a Hilbert space with a countable basis (possibly infinite) and a be a linear and bounded operator $\mathcal{T} \to \mathcal{T}$.
Then a is entirely determined by its action on the basis elements of $\mathcal{T}$.

Example:

Let $\mathcal{H} := \mathbb{C}^2$, with basis (e_1, e_2)

$\mathcal{H}^{\otimes n}$:= tensor product, with basis $(e_{i_1} \otimes .. \otimes e_{i_n},\ i_1 .. i_n \in \{1,2\})$

(with the convention $\mathcal{H}^{\otimes 0} := \mathbb{C} \cdot 1$ with basis (1))

$\mathcal{T} := \bigoplus_{n \geq 0} \mathcal{H}^{\otimes n}$; $\mathcal{T}$ has the following countable basis:

$(1, e_1, e_2, e_1 \otimes e_1, e_1 \otimes e_2, e_2 \otimes e_1, e_2 \otimes e_2, e_1 \otimes e_1 \otimes e_1, \dots)$

Interpretation: elements of $\mathcal{T}$ represent a physical system in a given state: n represents the number of particles, each of which is either in state e_1 or state e_2; 1 (corresponding to $n=0$) represents the empty state.

11.6.07
3

In the following, we will consider non-commutative random variables as linear and bounded operators $a : \mathcal{T} \to \mathcal{T}$.

We also define the expectation: $\varphi(a) := \langle 1, a1 \rangle$
(cf. $\varphi(A) = a_{11}$ in the matrix case)

NB: • the scalar product $\langle \cdot , \cdot \rangle$ on $\mathcal{T}$ is "defined" by

$$\langle e_{i_1} \otimes .. \otimes e_{i_n}, e_{j_1} \otimes .. \otimes e_{j_m} \rangle = \begin{cases} 1 & \text{if } m=n,\ i_1=j_1 \ldots i_m=j_m \\ 0 & \text{otherwise} \end{cases}$$

• the corresponding norm on $\mathcal{T}$ is defined by

$$\| h \| := \sqrt{\langle h, h \rangle} \qquad h \in \mathcal{T}$$

• and $a : \mathcal{T} \to \mathcal{T}$ is bounded if $\exists k > 0$ such that

$$\| a h \| \leq k \| h \| \qquad \forall h \in \mathcal{T}$$

(note that an $n \times n$ matrix A always satisfies this condition)

Examples of non-commutative random variables on $\mathcal{T}$:

• "creation operator":

$$a_i (e_{i_1} \otimes .. \otimes e_{i_n}) := \underset{\substack{\uparrow \\ \text{one more particle in state } i}}{e_i} \otimes e_{i_1} \otimes .. \otimes e_{i_n} \qquad i = 1, 2$$

• "annihilation operator":

$$a_i^* (e_{i_1} \otimes .. \otimes e_{i_n}) := \begin{cases} \overset{\text{one less particle in state } i}{e_{i_2}} \otimes \ldots \otimes e_{i_n} & \text{if } i_1 = i \\ \underset{\substack{\uparrow \\ \ldots \text{or even nothing!}}}{0} & \text{otherwise} \end{cases}$$

convention: $a_i^* 1 := 0$

Basic properties:

- $\forall h_1, h_2 \in \mathcal{T}, \quad \langle a_i^* h_1, h_2 \rangle = \langle h_1, a_i h_2 \rangle$; a_i^* = dual of a_i

  Proof: check this for basis elements:

  $$\langle a_i^* e_{i_1} \otimes .. \otimes e_{i_n}, e_{j_1} \otimes .. \otimes e_{j_m} \rangle =$$
  $$\stackrel{?}{=} \langle e_{i_1} \otimes .. \otimes e_{i_n}, a_i e_{j_1} \otimes .. \otimes e_{j_m} \rangle$$

  ie. $\delta_{i,i_1} \cdot \langle e_{i_2} \otimes .. \otimes e_{i_n}, e_{j_1} \otimes .. \otimes e_{j_m} \rangle$
  $$\stackrel{?}{=} \langle e_{i_1} \otimes .. \otimes e_{i_n}, e_i \otimes e_{j_1} \otimes .. \otimes e_{j_m} \rangle$$

  Both sides are equal to 1 iff $n = m+1$, $i = i_1$, $i_2 = j_1 \dots, i_n = j_m$ #

- $a_i^* a_i = I$: $a_i^* a_i e_{i_1} \otimes .. \otimes e_{i_n} = a_i^* e_i \otimes e_{i_1} \otimes \dots \otimes e_{i_n}$
  $$= e_{i_1} \otimes .. \otimes e_{i_n} \quad \#$$

- $a_i^* a_j = 0$ if $i \neq j$: clear, since a_j creates a particle in state j; so a_i^* annihilates the whole system #

Proposition

Let $p(x,y)$, $q(x,y)$ be any two polynomials.

Then $p(a_1, a_1^*)$ and $q(a_2, a_2^*)$ are freely independent.

Proof:

- Since $a_1^* a_1 = I$, $p(a_1, a_1^*)$ may always be written as a sum of terms of the form $a_1^k (a_1^*)^l$ with $k+l>0$ (and the identity term). Same is true for $q(a_2, a_2^*)$.

- Let us therefore check the free independence of terms of the form $a_1^k(a_1^*)^\ell$ and $b_1^k(b_1^*)^\ell$ with $k+\ell>0$:

  - first note that: $\varphi(a_i^k(a_i^*)^\ell) = \langle 1, a_1^k(a_i^*)^\ell 1\rangle$

    $= \langle (a_i^*)^k 1, (a_i^*)^\ell 1\rangle = 0$ since at least $k>0$ or $\ell>0$

    (recall that a_i^* = annihilation operator)

  - there remains to show therefore that:

$$\varphi\left(a_1^{k_1}(a_1^*)^{\ell_1} a_2^{k_2}(a_2^*)^{\ell_2} \cdots a_1^{k_{2m-1}}(a_1^*)^{\ell_{2m-1}} a_2^{k_{2m}}(a_2^*)^{\ell_{2m}}\right) = 0$$

suppose it is not the case; then $k_1=0$, otherwise

$$\varphi(a_1^{k_1}\ldots) = \langle 1, a_1^{k_1}\ldots 1\rangle = \langle (a_1^*)^{k_1}1, \ldots 1\rangle = 0$$

now, $k_1+\ell_1>0 \Rightarrow \ell_1>0$, so $k_2=0$, otherwise

$$(a_1^*)^{\ell_1} a_2^{k_2} h = 0 \quad \forall h \quad \Rightarrow \varphi(\ldots) = 0$$

now $k_2+\ell_2>0 \Rightarrow \ell_2>0$, so $k_3=0$ etc..

Finally, (by induction) $k_1=k_2=\ldots=k_{2m}=0$ and

$$\varphi(\ldots) = \varphi\left((a_1^*)^{\ell_1}(a_2^*)^{\ell_2}\cdots(a_1^*)^{\ell_{2m-1}}(a_2^*)^{\ell_{2m}}\right) = 0:$$

contradiction.

so $\varphi(\ldots)$ must be equal to zero, and therefore $p(a_1,a_1^*)$ and $q(a_2,a_2^*)$ are indeed freely independent. #

Now that we know that free non-commutative random variables indeed exist, let us derive the additivity rule for the R-transform.

Proposition

Let $p(x)$ be a polynomial and $A_1 = a_1 + p(a_1^*)$.

Then $R_{A_1}(z) = p(z)$. [Example: $A_1 = a_1 + a_1^* \rightarrow R(z) = z$ semi-circle dist!]

Recall: • the distribution of A_1 is given by $m_k^{A_1} = \varphi(A_1^k)$

• its Stieltjes transform is given by $g_{A_1}(z) = \varphi((A_1 - zI)^{-1})$.

• its R-transform is given by $R_{A_1}(z) = g_{A_1}^{-1}(-z) - \frac{1}{z}$.

Proof:

× Let $\omega_z := \sum_{n\geq 0} z^n e_1^{\otimes n} = 1 + z e_1 + z^2 e_1 \otimes e_1 + \ldots \in \mathcal{T}$ $(|z|<1)$

Then $a_1^* \omega_z = \sum_{n \geq 1} z^n e_1^{\otimes(n-1)} = z \sum_{n\geq 0} z^n e_1^{\otimes n} = z\,\omega_z$

scalar multiplication!

so $p(a_1^*)\,\omega_z = p(z)\,\omega_z$

Also, $a_1 \omega_z = \sum_{n\geq 0} z^n e_1^{\otimes(n+1)} = \frac{1}{z} \sum_{n \geq 1} z^n e_1^{\otimes n} = \frac{1}{z}(\omega_z - 1)$

and $A_1 \omega_z = \frac{1}{z}(\omega_z - 1) + p(z)\,\omega_z = \left(\frac{1}{z} + p(z)\right)\omega_z - \frac{1}{z} 1$

i.e. $\left(A_1 - \frac{1}{z} - p(z)\right)\omega_z = -\frac{1}{z} 1$ or $\left(A_1 - \frac{1}{z} - p(z)\right)^{-1} 1 = -z\,\omega_z$

$\Rightarrow \underbrace{\varphi\left(\left(A_1 - \frac{1}{z} - p(z)\right)^{-1}\right)} = -z \langle 1, \omega_z \rangle = -z$

$= g_{A_1}\left(\frac{1}{z} + p(z)\right)$ so $g_{A_1}^{-1}(-z) = \frac{1}{z} + p(z)$ and $R(z) = p(z)$ #

Theorem

Let $p(x)$, $q(x)$ be any two polynomials $\left(\begin{array}{l}\text{may be extended to}\\ p,q \text{ analytic functions}\end{array}\right)$

and $A_1 = a_1 + p(a_1^*)$, $A_2 = a_2 + q(a_2^*)$

× Then A_1 and A_2 are freely independent

and $R_{A_1+A_2}(z) = R_{A_1}(z) + R_{A_2}(z)$.

Proof:

× One has to prove that $R_{A_1+A_2}(z) = p(z) + q(z)$.

Let $\rho_z := \sum_{n\geq 0} z^n (e_1+e_2)^{\otimes n} \qquad |z| < \frac{1}{2}$

Then $(a_1+a_2)\rho_z = \sum_{n\geq 0} z^n (e_1+e_2)^{\otimes(n+1)} = \frac{1}{z}(\rho_z - 1)$

$$a_1^* \rho_z = \sum_{n\geq 0} z^n a_1^*(a_1+a_2)^{\otimes n} 1$$

$$= \sum_{n\geq 1} z^n (a_1+a_2)^{\otimes(n-1)} 1 \quad \text{since } a_1^* a_1 = I,\ a_1^* a_2 = 0$$

$$= \sum_{n\geq 1} z^n (e_1+e_2)^{\otimes(n-1)} = z\rho_z$$

Similarly, $p(a_1^*)\rho_z = p(z)\rho_z$; $a_2^* = z\rho_z$, $q(a_2^*)\rho_z = q(z)\rho_z$

So $(A_1+A_2)\rho_z = \frac{1}{z}(\rho_z - 1) + (p(z)+q(z))\rho_z$

$\left(A_1+A_2 - \frac{1}{z} - p(z) - q(z)\right)\rho_z = -\frac{1}{z} 1$

$\left(A_1+A_2 - \frac{1}{z} - p(z) - q(z)\right)^{-1} 1 = -z\rho_z$

$g_{A_1+A_2}\left(\frac{1}{z} + p(z) + q(z)\right) = -z \quad$ since $\langle 1, \rho_z\rangle = 1$

i.e. $g_{A_1+A_2}^{-1}(-z) = \frac{1}{z} + p(z) + q(z)$

and $R_{A_1+A_2}(z) = g_{A_1+A_2}^{-1}(-z) - \frac{1}{z} = p(z) + q(z)$ #

Random matrix theory: Lecture 24

Recall:

- $\mathcal{T} = \bigoplus_{n \geq 0} \mathcal{H}^{\otimes n}$, with $\mathcal{H} = \mathbb{C}^2$, basis (e_1, e_2)
- $\mathcal{A} = \{ a : \mathcal{T} \to \mathcal{T}$ linear bounded operator $\}$
- $\varphi(a) = \langle 1, a1 \rangle$
- a_i, a_i^* creation and annihilation operators
- $A_i = a_i + a_i^*$ distributed according to the semi-circle distribution: $R_{A_i}(z) = z$
- A_1 and A_2 freely independent

Comment 1

$$R_{A_1+A_2}(z) = R_{A_1}(z) + R_{A_2}(z) = z + z = 2z$$

ie. $A_1 + A_2$ is again distributed according to the semi-circle distribution (with a different variance).

ie. the semi-circle distribution plays the same role in free probability as the Gaussian distribution in classical probability. The analogy goes on with the following theorem (turn the page)..

13.6.04

2

Free central limit theorem

Let $a_1, \dots, a_n, \dots$ be a sequence of freely independent random variables, identically distributed and such that $\varphi(a_1) = 0$ and $\varphi(a_1^2) = 1$.

Let μ_n be the distribution of $\frac{1}{\sqrt{n}}(a_1 + \dots + a_n)$.

Then μ_n converges to the semi-circle distribution as $n \to \infty$.

(i.e. $R_{\mu_n}(z) \xrightarrow[n\to\infty]{} z \quad \forall z \in \mathbb{C}$)

Proof idea:

1) For c a constant and A a random variable, $R_{cA}(z) = c R_A(cz)$:

$$\begin{cases} g_{cA}(z) = \varphi\left((cA - zI)^{-1}\right) = \frac{1}{c}\varphi\left(\left(A - \frac{z}{c}I\right)^{-1}\right) = \frac{1}{c} g_A\left(\frac{z}{c}\right) \\ \text{so } g_{cA}^{-1}(z) = c\, g_A^{-1}(cz) \quad \left[\frac{1}{c} g_A\left(\frac{z}{c}\right) = \zeta \text{ iff } c\, g_A^{-1}(c\zeta) = z\right] \\ \text{and } R_{cA}(z) = g_{cA}^{-1}(-z) - \frac{1}{z} = c R_A(cz). \end{cases}$$

2) $R_{\frac{1}{\sqrt{n}}(a_1+\dots+a_n)}(z) = \frac{1}{\sqrt{n}} R_{a_1+\dots+a_n}\left(\frac{z}{\sqrt{n}}\right)$ by (1)

(by free indep.) $= \frac{1}{\sqrt{n}} \sum_{j=1}^{n} R_{a_j}\left(\frac{z}{\sqrt{n}}\right) = \sqrt{n}\, R_{a_1}\left(\frac{z}{\sqrt{n}}\right)$

× 3) expansion: $R_{a_1}(z) = \sum_{k\geq 0} c_k z^k \quad \begin{cases} c_0 = \varphi(a_1) = 0 \\ c_1 = \varphi(a_1^2) - \varphi(a_1)^2 = 1 \end{cases}$

$\Rightarrow R_{a_1}(z) = z + o(z)$

i.e. $R_{\frac{1}{\sqrt{n}}(a_1+\dots+a_n)}(z) = \sqrt{n}\left(\frac{z}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right)\right) = z + o(1)$ #

<u>Comment 2</u>

The fact that $A_1 = a_1 + a_1^*$ is distributed according to the semi-circle distribution can be put in relation with the following:

- Let $A^{(n)} := \{ n \times n \text{ (deterministic) matrices} \}$ (with $n \to \infty$)

$$\varphi(A) := a_{11}^{(n)}$$

$$A^{(n)} := \begin{pmatrix} 0 & 1 & & 0 \\ 1 & \ddots & \ddots & \\ & \ddots & \ddots & 1 \\ 0 & & 1 & 0 \end{pmatrix} \quad \text{deterministic Toeplitz matrix}$$

Catalan numbers!

Then $\lim_{n\to\infty} \varphi\left(\left(A^{(n)}\right)^k \right) = \begin{cases} \frac{1}{\ell+1}\binom{2\ell}{\ell} & \text{if } k = 2\ell \\ 0 & \text{if } k = 2\ell+1 \end{cases}$

i.e as $n \to \infty$, $A^{(n)}$ is distributed according to the semi-circle distribution (with respect to expectation φ)!

"<u>Proof</u>"

× $\varphi\left(\left(A^{(n)}\right)^k\right) = \left(A^{(n)}\right)^k_{11} \xrightarrow[n\to\infty]{} \#$ Dyck paths of length k

$$= \begin{cases} c_\ell & \text{if } k = 2\ell \\ 0 & \text{otherwise} \end{cases} \quad \#$$

- Note that with such an expectation φ, changing only two entries in the Toeplitz matrix changes drastically the limit: Let $B^{(n)} := \begin{pmatrix} 0 & 1 & & 0 & 1 \\ 1 & \ddots & \ddots & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & 1 \\ 1 & 0 & & 1 & 0 \end{pmatrix}$ [circulant matrix!]

× Then $\lim_{n\to\infty} \varphi\left(\left(B^{(n)}\right)^k\right) = \begin{cases} \binom{2\ell}{\ell} & \text{if } k = 2\ell \\ 0 & \text{otherwise} \end{cases}$ arcsine distribution

<u>Proof</u>: use $b_{11}^{(n)} = \frac{1}{n} \mathrm{Tr}\left(B^{(n)}\right)$

13.6.04

5

Free multiplicative convolution and S-transform

Let A be a non-commutative random variable such that $\varphi(A) \neq 0$.

Let $\begin{cases} \psi_A(z) := \sum_{k\geq 1} \varphi(A^k) z^k & \left[= \varphi((I - zA)^{-1}) = -\frac{1}{z} \underbrace{g_A(\tfrac{1}{z})}_{\text{Stieltjes transform}} - 1 \right] \\ S_A(z) := \frac{z+1}{z} \underbrace{\psi_A^{-1}(z)}_{\text{inverse function}} \end{cases}$

(*)

Proposition:

If A_1, A_2 are freely independent and s.t. $\varphi(A_1) \neq 0$, $\varphi(A_2) \neq 0$,

Then $S_{A_1A_2}(z) = S_{A_1}(z) S_{A_2}(z)$

and the distribution of $A_1 A_2$ is called the free multiplicative convolution and is denoted as $\mu_{A_1A_2} = \mu_{A_1} \boxtimes \mu_{A_2}$.

Proof idea:

Same technique as for the R-transform:

× Let $A_1 = (I + a_1) \cdot p(a_1^*)$ with $p(z)$ some polynomial s.t. $p(0) \neq 0$.

× Then $S_{A_1}(z) = \frac{1}{p(z)}$; similarly $S_{A_2}(z) = \frac{1}{q(z)}$

and $S_{A_1A_2}(z) = \frac{1}{p(z)q(z)}$ --- "#"

× (*) Example: $A \sim$ "quarter circle", i.e. $z g_A(z)^2 + z g_A(z) + 1 = 0$

$$\Rightarrow z(\psi_A(z)+1)^2 = \psi_A(z) \Rightarrow S_A(z) = \frac{1}{z+1}$$

Application to random matrices

- Let $A^{(n)}, B^{(n)}$ be $n \times n$ real symmetric independent random matrices with limiting eigenvalue distributions μ_A, μ_B.$^{(*)}$
  Let $V^{(n)}$ be orthogonal (Haar dist.) & indep. of both $A^{(n)}$ and $B^{(n)}$.
- Then $A^{(n)}$ and $V^{(n)} B^{(n)} (V^{(n)})^T$ are asymptotically free, so the limiting eigenvalue distribution of $A^{(n)} V^{(n)} B^{(n)} (V^{(n)})^T$ is given by $\mu_{AB} = \mu_A \boxtimes \mu_B$ and can be computed via the S-transform.

(*) such that $\int_{\mathbb{R}} x \, d\mu_A(x) \neq 0$ and $\int_{\mathbb{R}} x \, d\mu_B(x) \neq 0$