Free Probability Theory and Random Matrices

Roland Speicher
Queen's University
Kingston, Canada

We are interested in the limiting eigenvalue distribution of

$$\boldsymbol{N \times N} \text{ random matrices}$$

for

$$\boldsymbol{N \to \infty.}$$

Usually, large N distributions are close to the $N \to \infty$ limit, and asymptotic results give good predictions for finite N.

We can consider the convergence for $N \to \infty$ of

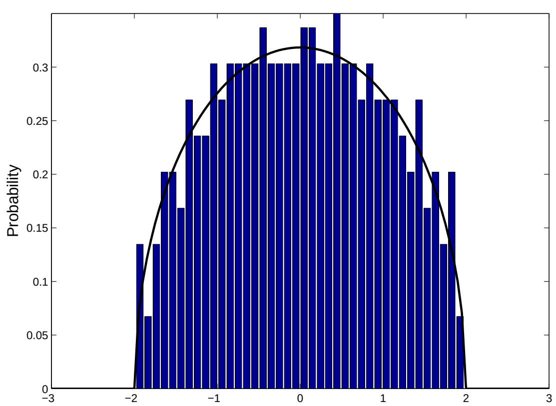
- the eigenvalue distribution of one "typical" realization of the $N \times N$ random matrix

- the averaged eigenvalue distribution over many realizations of the $N \times N$ random matrices

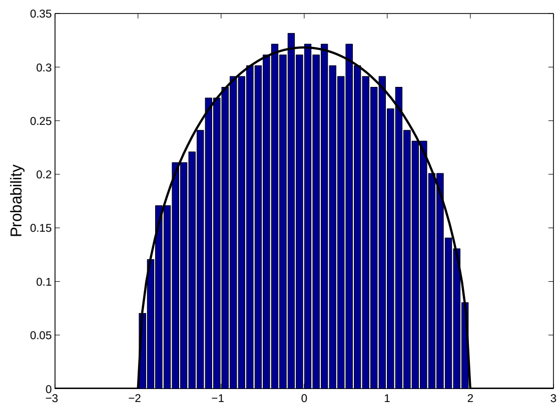
Consider (selfadjoint!) **Gaussian $N \times N$ random matrix**.

We have almost sure convergence (convergence of ”typical” realization) of its eigenvalue distribution towards

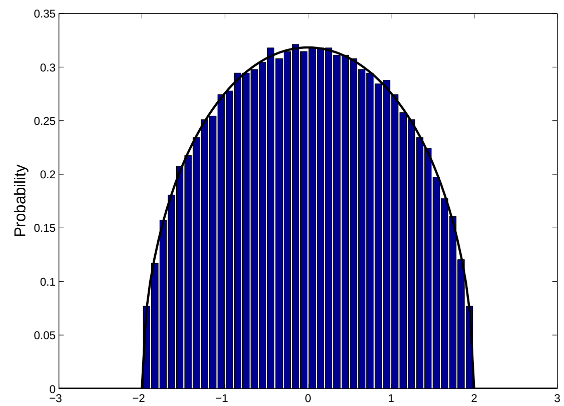
Wigner’s semicircle.



N=300

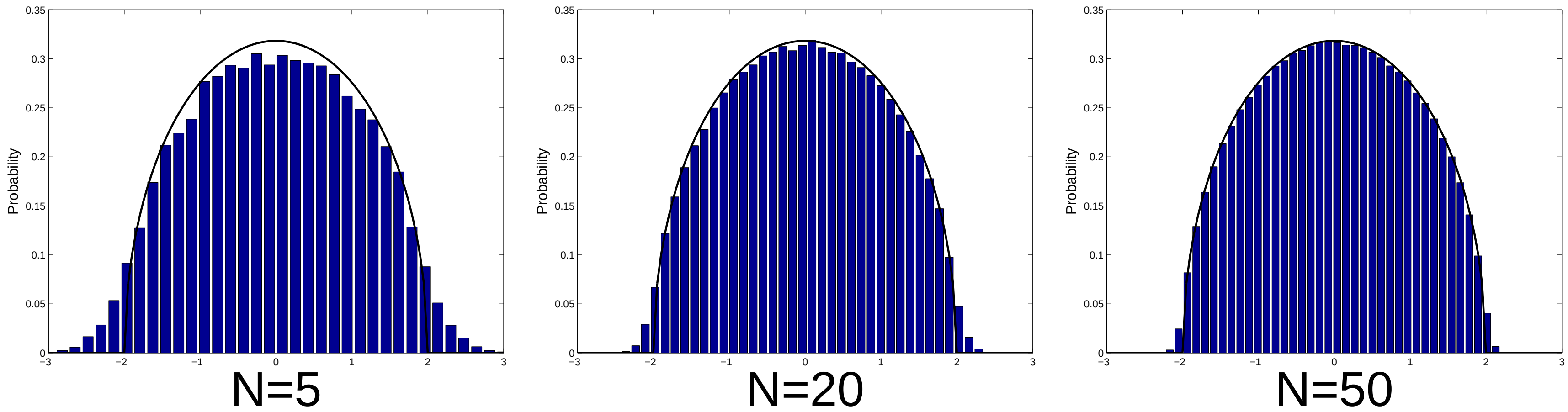


N=1000



N=3000

Convergence of the averaged eigenvalue distribution happens usually much faster, very good agreement with asymptotic limit for moderate N.

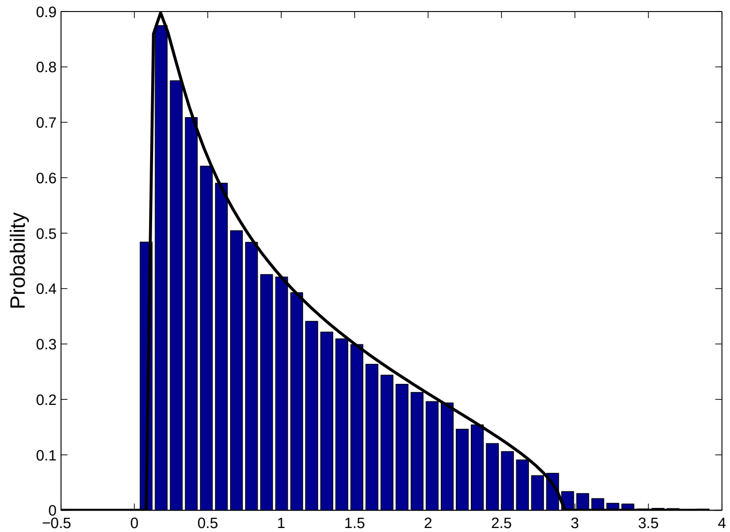


trials=5000

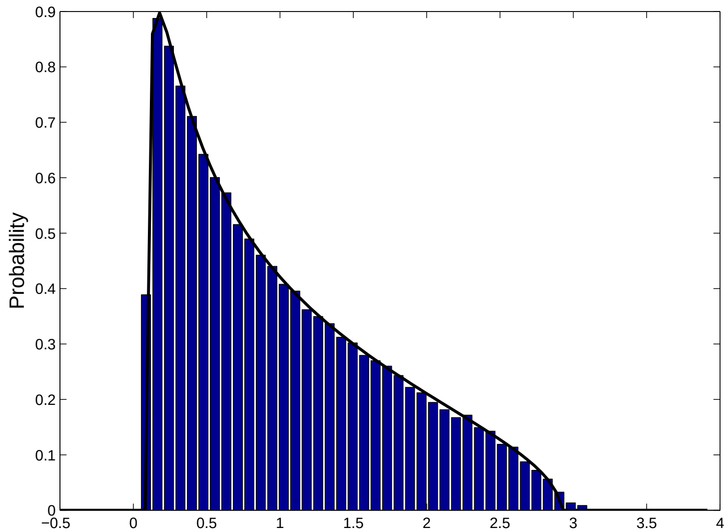
Consider **Wishart random matrix** $A = XX^*$, where X is $N \times M$ random matrix with independent Gaussian entries

Its eigenvalue distribution converges (averaged and almost surely) towards **Marchenko-Pastur distribution.**

Example: $M = 2N$, 2000 trials



N=10



N=50

We want to consider more complicated situations, built out of simple cases (like Gaussian or Wishart) by doing operations like

- taking the sum of two matrices
- taking the product of two matrices
- taking corners of matrices

Note: If different $N \times N$ random matrices A and B are involved then the eigenvalue distribution of non-trivial functions $f(A, B)$ (like $A+B$ or AB) will of course depend on the relation between the eigenspaces of A and of B.

However: It turns out there is a deterministic and treatable result if

- the eigenspaces are in **"generic" position** and
- if $\boldsymbol{N \to \infty}$

This is the realm of **free probability theory**.

Consider $N \times N$ random matrices A and C such that

- A has an asymptotic eigenvalue distribution for $N \to \infty$ and C has an asymptotic eigenvalue distribution for $N \to \infty$

- A and C are independent (i.e., entries of A are independent from entries of C)

Then eigenspaces of A and of C might still be in special relation (e.g., both A and C could be diagonal).

However, consider now

$$A \qquad \text{and} \qquad B := UCU^*,$$

where U is Haar unitary $N \times N$ random matrix.

Then, eigenspaces of A and of B are in "generic" position and the asymptotic eigenvalue distribution of $A + B$ depends only on the asymptotic eigenvalue distribution of A and the asymptotic eigenvalue distribution of B (which is the same as the one of C).

We can expect that the asymptotic eigenvalue distribution of $f(A, B)$ depends only on the asymptotic eigenvalue distribution of A and the asymptotic eigenvalue distribution of B if

- A and B are independent

- one of them is unitarily invariant
  (i.e., the joint distribution of the entries does not change under unitary conjugation)

Note: Gaussian and Wishart random matrices are unitarily invariant

Thus: the asymptotic eigenvalue distribution of

- the sum of random matrices in generic position

$$A + UCU^*$$

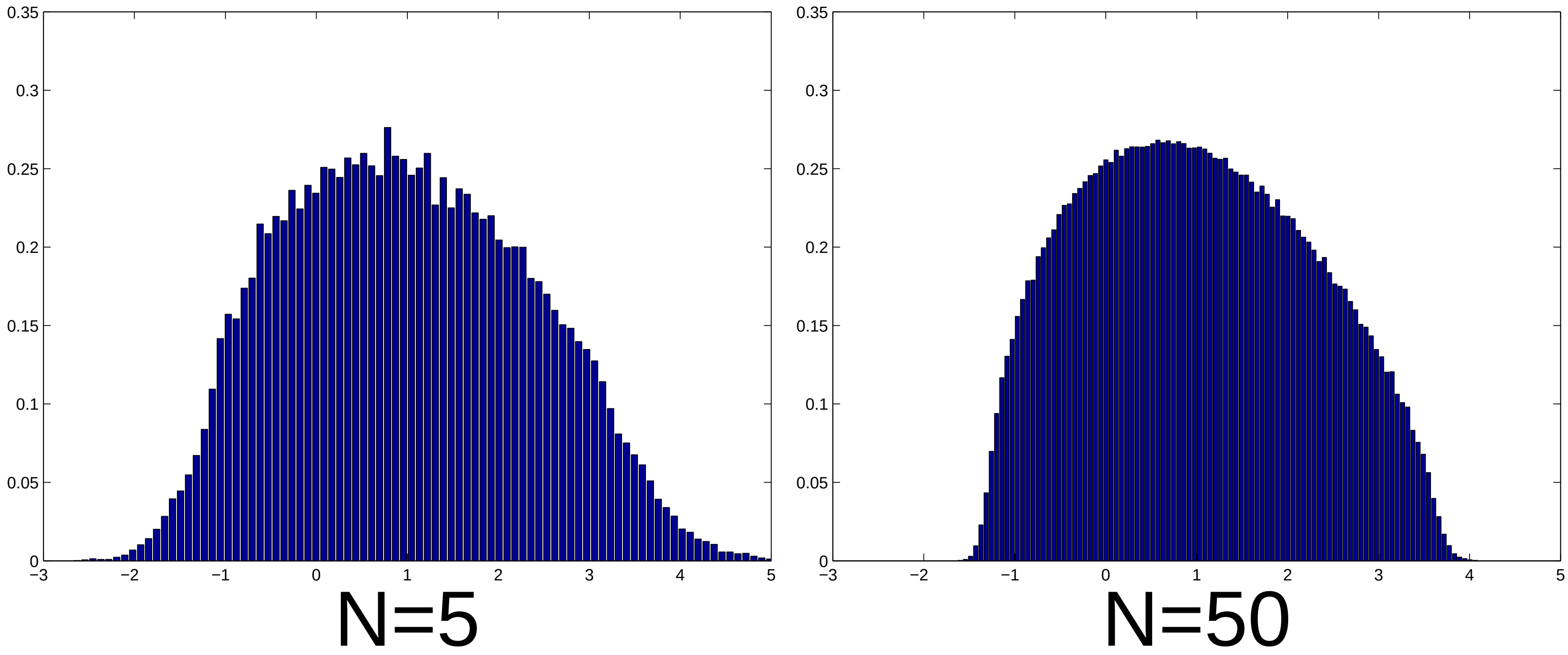
- the product of random matrices in generic position

$$AUCU^*$$

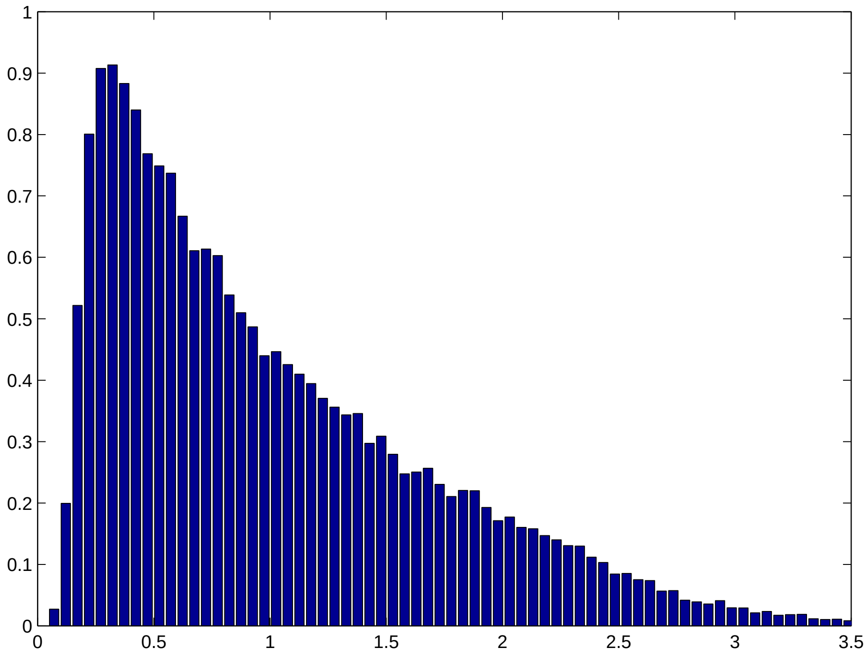
- corners of unitarily invariant matrices UCU^*

should only depend on the asymptotic eigenvalue distribution of A and of C.

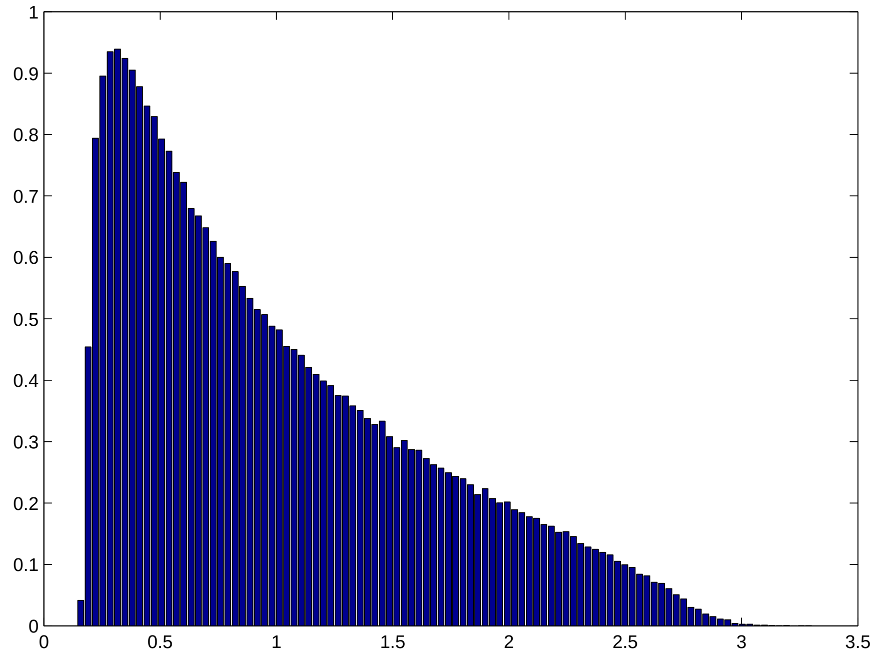
Example: sum of independent Gaussian and Wishart ($M = 2N$) random matrices, averaged over 10000 trials



Example: product of two independent Wishart ($M = 5N$) random matrices, averaged over 10000 trials

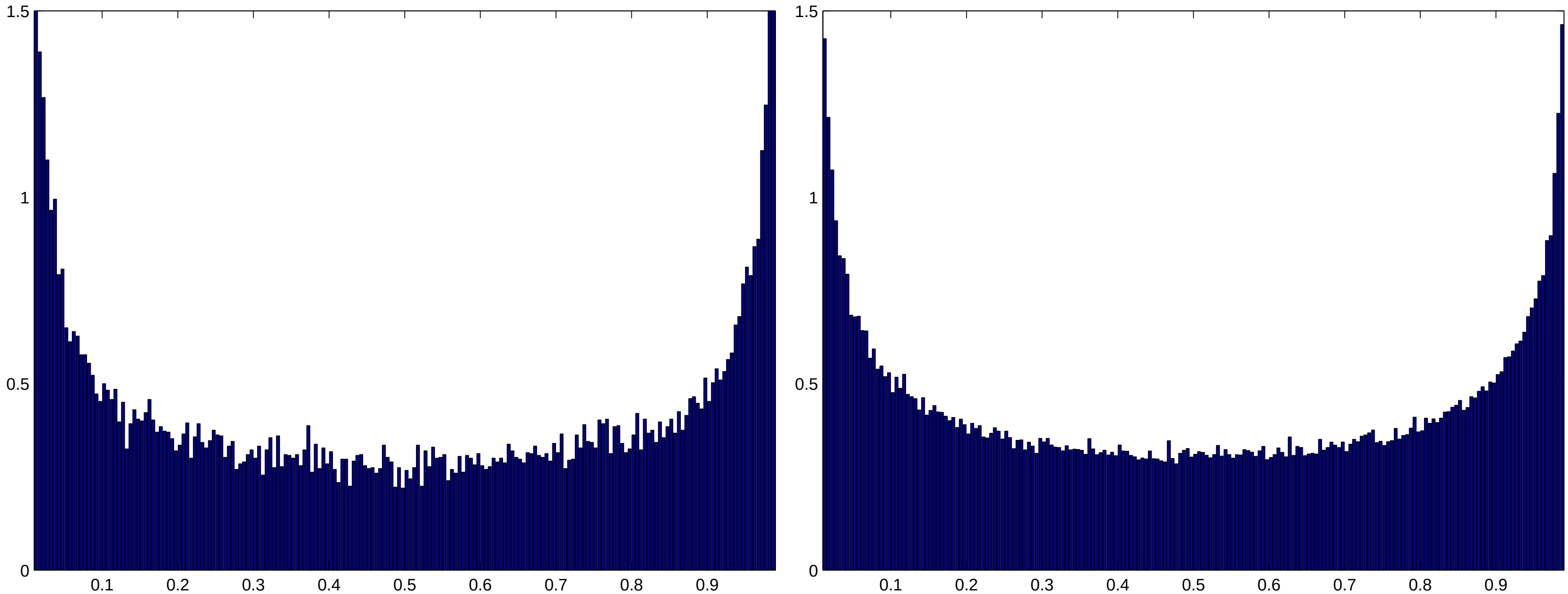


N=5



N=50

Example: upper left corner of size $N/2 \times N/2$ of a randomly rotated $N \times N$ projection matrix,
with half of the eigenvalues 0 and half of the eigenvalues 1, averaged over 10000 trials



N=8 N=32

Problems:

- Do we have a conceptual way of understanding these asymptotic eigenvalue distributions?

- Is there an algorithm for actually calculating these asymptotic eigenvalue distributions?

How do we analyze the eigenvalue distributions?

eigenvalue distribution of matrix A	$\hat{=}$	knowledge of traces of powers, $\text{tr}(A^k)$
$\frac{1}{N}\left(\lambda_1^k + \cdots + \lambda_N^k\right)$	$=$	$\text{tr}(A^k)$
averaged eigenvalue distribution of random matrix A	$\hat{=}$	knowledge of expectations of traces of powers, $E[\text{tr}(A^k)]$

Stieltjes inversion formula. If one knows the asymptotic moments

$$\alpha_k := \lim_{N\to\infty} E[\mathrm{tr}(A^k)]$$

of a random matrix A, then one can get its asymptotic eigenvalue distribution μ as follows:

Form Cauchy (or Stieltjes) transform

$$G(z) := \sum_{k=0}^{\infty} \frac{\alpha_k}{z^{k+1}}$$

Then:

$$d\mu(t) = -\frac{1}{\pi} \lim_{\varepsilon\to 0} \Im G(t + i\varepsilon)$$

Consider random matrices A and B in generic position.

We want to understand $A + B$, i.e., for all $k \in \mathbb{N}$

$$E\left[\mathrm{tr}\Big((A + B)^k\Big)\right].$$

But

$$E[\mathrm{tr}((A+B)^6)] = E[\mathrm{tr}(A^6)]+\cdots+E[\mathrm{tr}(ABAABA)]+\cdots+E[\mathrm{tr}(B^6)],$$

thus we need to understand

mixed moments in A and B

Use following notation:

$$\varphi(A) := \lim_{N\to\infty} E[\mathrm{tr}(A)].$$

Question: If A and B are in generic position, can we understand

$$\varphi\left(A^{n_1}B^{m_1}A^{n_2}B^{m_2}\cdots\right)$$

in terms of

$$\Big(\varphi(A^k)\Big)_{k\in\mathbb{N}} \qquad \text{and} \qquad \Big(\varphi(B^k)\Big)_{k\in\mathbb{N}}$$

Example: Consider two Gaussian random matrices A, B which are independent (and thus in generic position).

Then the asymptotic mixed moments in A and B

$$\varphi\left(A^{n_1} B^{m_1} A^{n_2} B^{m_2} \cdots\right)$$

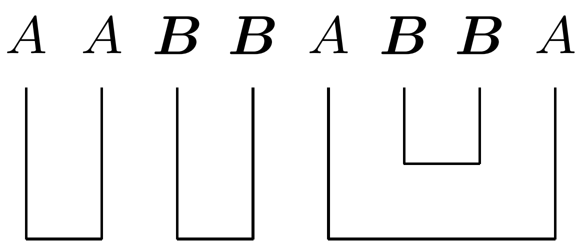
are given by

$$\#\Big\{\text{non-crossing/planar pairings of the pattern}$$

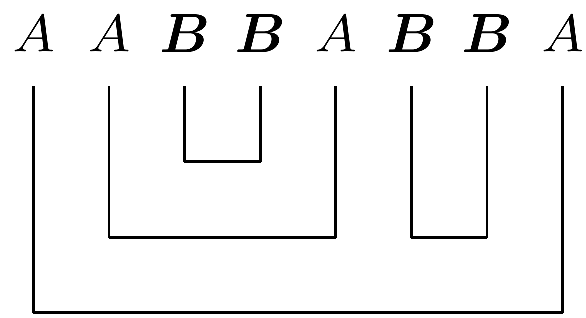
$$\underbrace{A \cdot A \cdots A}_{n_1\text{-times}} \cdot \underbrace{B \cdot B \cdots B}_{m_1\text{-times}} \cdot \underbrace{A \cdot A \cdots A}_{n_2\text{-times}} \cdot \underbrace{B \cdot B \cdots B}_{m_2\text{-times}} \cdots,$$

$$\text{which do not pair } A \text{ with } B\Big\}$$

Example: $\varphi(AA\boldsymbol{BB}A\boldsymbol{BB}A) = 2$ since there are two such non-crossing pairings:



and



Note: each of the pairings connects at least one of the groups A^{n_1}, B^{m_1}, A^{n_2}, ... only among itself!

and thus:

$$\varphi\Big(\big(A^2-\varphi(A^2)1\big)\big(B^2-\varphi(B^2)1\big)\big(A-\varphi(A)1\big)\big(B^2-\varphi(B^2)1\big)\big(A-\varphi(A)1\big)\Big) = 0$$

In general we have

$$\varphi\Big(\big(\boldsymbol{A}^{n_1}-\varphi(\boldsymbol{A}^{n_1})\cdot\mathbf{1}\big)\cdot\big(\boldsymbol{B}^{m_1}-\varphi(\boldsymbol{B}^{m_1})\cdot\mathbf{1}\big)\cdot\big(\boldsymbol{A}^{n_2}-\varphi(\boldsymbol{A}^{n_2})\cdot\mathbf{1}\big)\cdots\Big)$$

$= \#\Big\{$non-crossing pairings which do not pair A with B,

and for which each group is connected with some other group$\Big\}$

$$=\mathbf{0}$$

Actual equation for the calculation of the mixed moments

$$\varphi_1\left(A^{n_1}B^{m_1}A^{n_2}B^{m_2}\cdots\right)$$

is different for different random matrix ensembles.

However, the relation between the mixed moments,

$$\varphi\Big(\left(A^{n_1}-\varphi(A^{n_1})\cdot 1\right)\cdot\left(B^{m_1}-\varphi(B^{m_1})\cdot 1\right)\cdots\Big)=0$$

remains the same for matrix ensembles in generic position and constitutes the **definition of freeness**.

Definition [Voiculescu 1985]: A and B are **free** (with respect to φ) if we have for all $n_1, m_1, n_2, \cdots \geq 1$ that

$$\varphi\Big(\big(A^{n_1} - \varphi(A^{n_1}) \cdot 1\big) \cdot \big(B^{m_1} - \varphi(B^{m_1}) \cdot 1\big) \cdot \big(A^{n_2} - \varphi(A^{n_2}) \cdot 1\big) \cdots \Big) = 0$$

$$\varphi\Big(\big(B^{n_1} - \varphi(B^{n_1}) \cdot 1\big) \cdot \big(A^{m_1} - \varphi(A^{m_1}) \cdot 1\big) \cdot \big(B^{n_2} - \varphi(B^{n_2}) \cdot 1\big) \cdots \Big) = 0$$

$$\varphi\Big(\text{alternating product in centered words in } A \text{ and in } B\Big) = 0$$

Note: freeness is a rule for calculating mixed moments in A and B from the moments of A and the moments of B.

Example:

$$\varphi\Big(\big(A^n - \varphi(A^n)1\big)\big(B^m - \varphi(B^m)1\big)\Big) = 0,$$

thus

$$\varphi(A^nB^m) - \varphi(A^n \cdot 1)\varphi(B^m) - \varphi(A^n)\varphi(1 \cdot B^m) + \varphi(A^n)\varphi(B^m)\varphi(1 \cdot 1) = 0,$$

and hence

$$\boldsymbol{\varphi(A^nB^m) = \varphi(A^n) \cdot \varphi(B^m).}$$

Freeness is a rule for calculating mixed moments, analogous to the concept of independence for random variables.

Thus freeness is also called **free independence**

Note: free independence is a different rule from classical independence; free independence occurs typically for **non-commuting random variables**.

Example:

$$\varphi\Big(\big(A-\varphi(A)1\big)\cdot\big(B-\varphi(B)1\big)\cdot\big(A-\varphi(A)1\big)\cdot\big(B-\varphi(B)1\big)\Big)=0,$$

which results in

$$\begin{aligned}\boldsymbol{\varphi(ABAB)=\varphi(AA)\cdot\varphi(B)\cdot\varphi(B)+\varphi(A)\cdot\varphi(A)\cdot\varphi(BB)}\\ \boldsymbol{-\,\varphi(A)\cdot\varphi(B)\cdot\varphi(A)\cdot\varphi(B)}\end{aligned}$$

Consider A, B free.

Then, by freeness, the moments of $A+B$ are uniquely determined by the moments of A and the moments of B.

Notation: We say the distribution of $A + B$ is the

free convolution

of the distribution of A and the distribution of B,

$$\mu_{A+B} = \mu_A \boxplus \mu_B.$$

In principle, freeness determines this, but the concrete nature of this rule is not clear.

Examples: We have

$$\varphi\Big((A+B)^1\Big) = \varphi(A) + \varphi(B)$$

$$\varphi\Big((A+B)^2\Big) = \varphi(A^2) + 2\varphi(A)\varphi(B) + \varphi(B^2)$$

$$\varphi\Big((A+B)^3\Big) = \varphi(A^3) + 3\varphi(A^2)\varphi(B) + 3\varphi(A)\varphi(B^2) + \varphi(B^3)$$

$$\begin{aligned}\varphi\Big((A+B)^4\Big) = {} & \varphi(A^4) + 4\varphi(A^3)\varphi(B) + 4\varphi(A^2)\varphi(B^2) \\ & + 2\Big(\varphi(A^2)\varphi(B)\varphi(B) + \varphi(A)\varphi(A)\varphi(B^2) \\ & \qquad - \varphi(A)\varphi(B)\varphi(A)\varphi(B)\Big) \\ & + 4\varphi(A)\varphi(B^3) + \varphi(B^4)\end{aligned}$$

To treat these formulas in general, linearize the free convolution by going over from moments $(\varphi(A^m))_{m\in\mathbb{N}}$ to **free cumulants** $(\kappa_m)_{m\in\mathbb{N}}$.

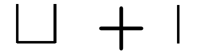
Those are defined by relations like:

$$\begin{aligned}
\varphi(A^1) &= \kappa_1 \\
\varphi(A^2) &= \kappa_2 + \kappa_1^2 \\
\varphi(A^3) &= \kappa_3 + 3\kappa_1\kappa_2 + \kappa_1^3 \\
\varphi(A^4) &= \kappa_4 + 4\kappa_1\kappa_3 + 2\kappa_2^2 + 6\kappa_1^2\kappa_2 + \kappa_1^4 \\
&\vdots
\end{aligned}$$

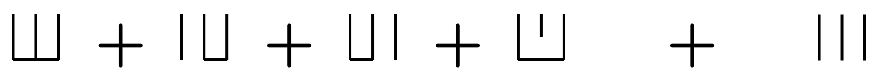
There is a combinatorial structure behind these formulas, the sums are running over **non-crossing partitions**:

$\varphi(A^1) =$

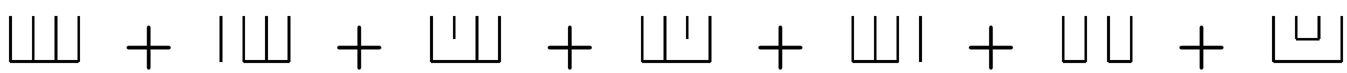
$= \kappa_1$

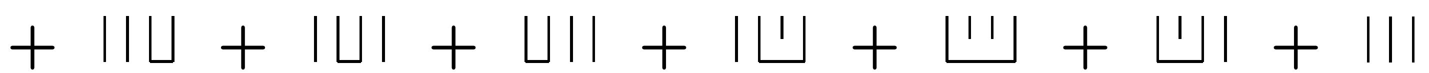
$\varphi(A^2) =$ 

$= \kappa_2 + \kappa_1\kappa_1$

$\varphi(A^3) =$ 

$= \kappa_3 + \kappa_1\kappa_2 + \kappa_2\kappa_1 + \kappa_2\kappa_1 + \kappa_1\kappa_1\kappa_1$

$\varphi(A^4) =$ 



$= \kappa_4 + 4\kappa_1\kappa_3 + 2\kappa_2^2 + 6\kappa_1^2\kappa_2 + \kappa_1^4$

This combinatorial relation between moments $(\varphi(A^m))_{m\in\mathbb{N}}$ and cumulants $(\kappa_m)_{m\in\mathbb{N}}$ can be translated into generating power series.

Put

$$G(z) = \frac{1}{z} + \sum_{m=1}^{\infty} \frac{\varphi(A^m)}{z^{m+1}} \qquad \text{Cauchy transform}$$

and

$$\mathcal{R}(z) = \sum_{m=1}^{\infty} \kappa_m z^{m-1} \qquad \boldsymbol{\mathcal{R}}\textbf{-transform}$$

Then we have the relation

$$\frac{1}{G(z)} + \mathcal{R}(G(z)) = z.$$

Theorem [Voiculescu 1986, Speicher 1994]:
Let A and B be free. Then one has

$$\mathcal{R}^{A+B}(z) = \mathcal{R}^{A}(z) + \mathcal{R}^{B}(z),$$

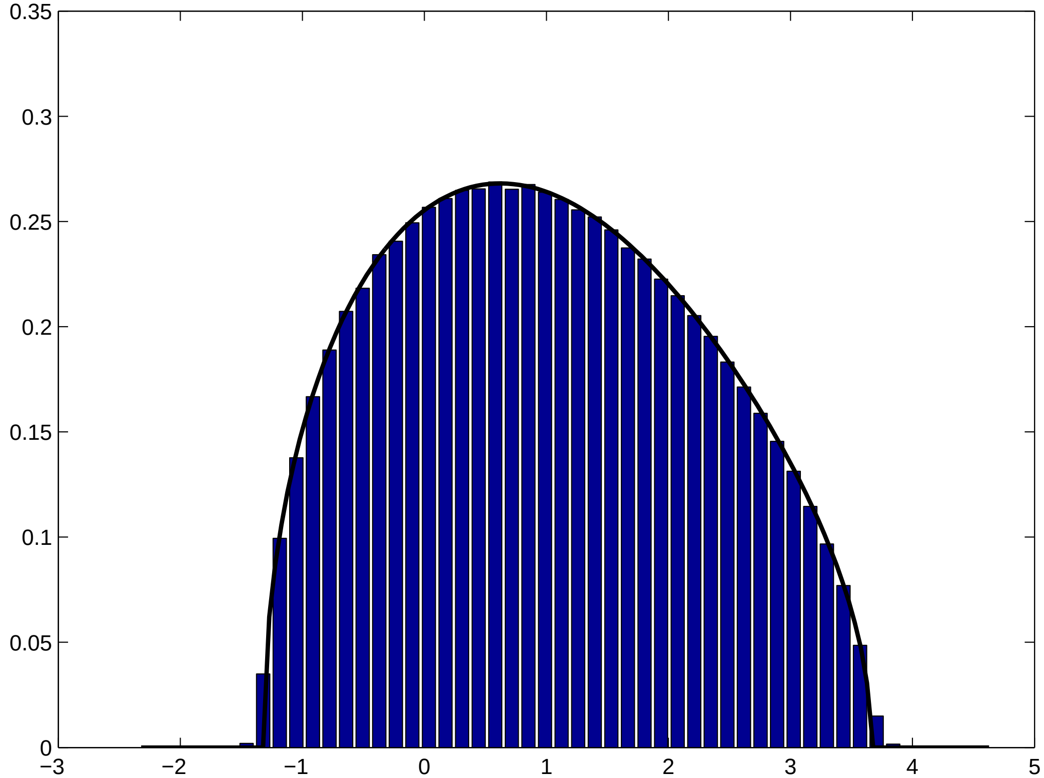
or equivalently

$$\kappa_m^{A+B} = \kappa_m^{A} + \kappa_m^{B} \qquad \forall\, m.$$

This, together with the relation between Cauchy transform and $\mathcal{R}$-transform and with the Stieltjes inversion formula, gives an effective algorithm for calculating free convolutions, i.e., sums of random matrices in generic position.

$$
\begin{array}{ccccccccc}
A & \rightsquigarrow & G^A & \rightsquigarrow & R^A & & & & \\
 & & & & \downarrow & & & & \\
 & & & & R^A + R^B = R^{A+B} & \rightsquigarrow & G^{A+B} & \rightsquigarrow & A+B \\
 & & & & \uparrow & & & & \\
B & \rightsquigarrow & G^B & \rightsquigarrow & R^B & & & &
\end{array}
$$

Example: Wigner + Wishart ($M = 2N$), trials = 4000



N=100

One has similar analytic description for product.

Theorem [Voiculescu 1987, Haagerup 1997, Nica + Speicher 1997]:
Put

$$M_A(z) := \sum_{m=0}^{\infty} \varphi(A^m) z^m$$

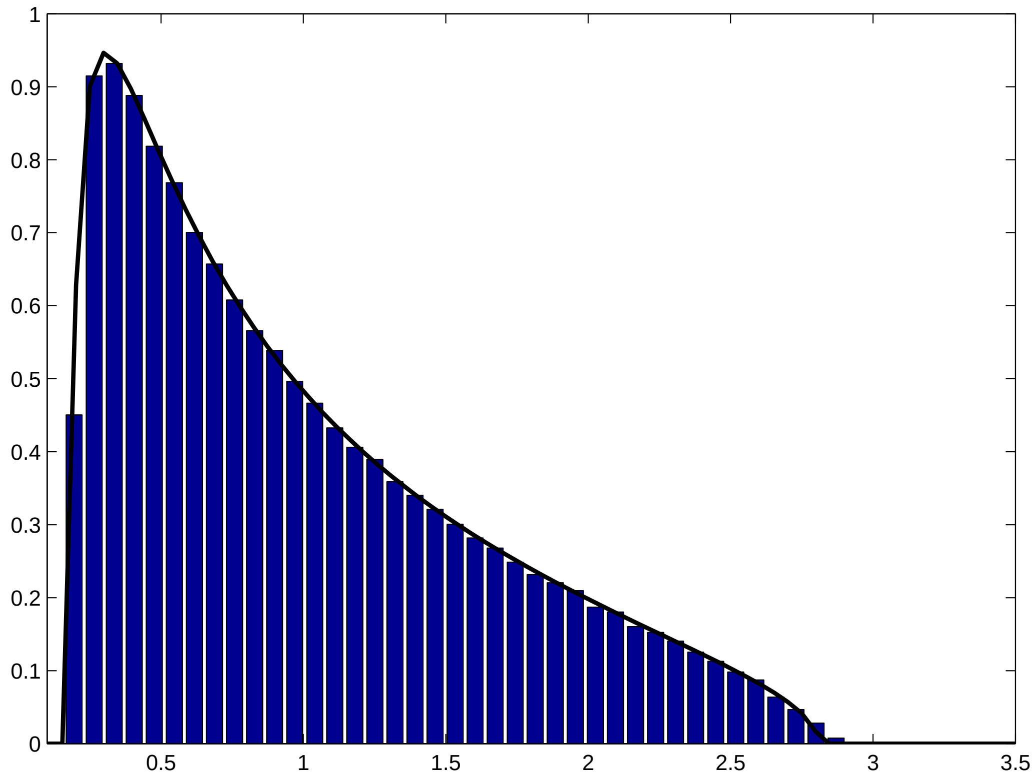
and define

$$S_A(z) := \frac{1+z}{z} M_A^{<-1>}(z) \qquad \textbf{\textit{S}-transform of } \boldsymbol{A}$$

Then: If A and B are free, we have

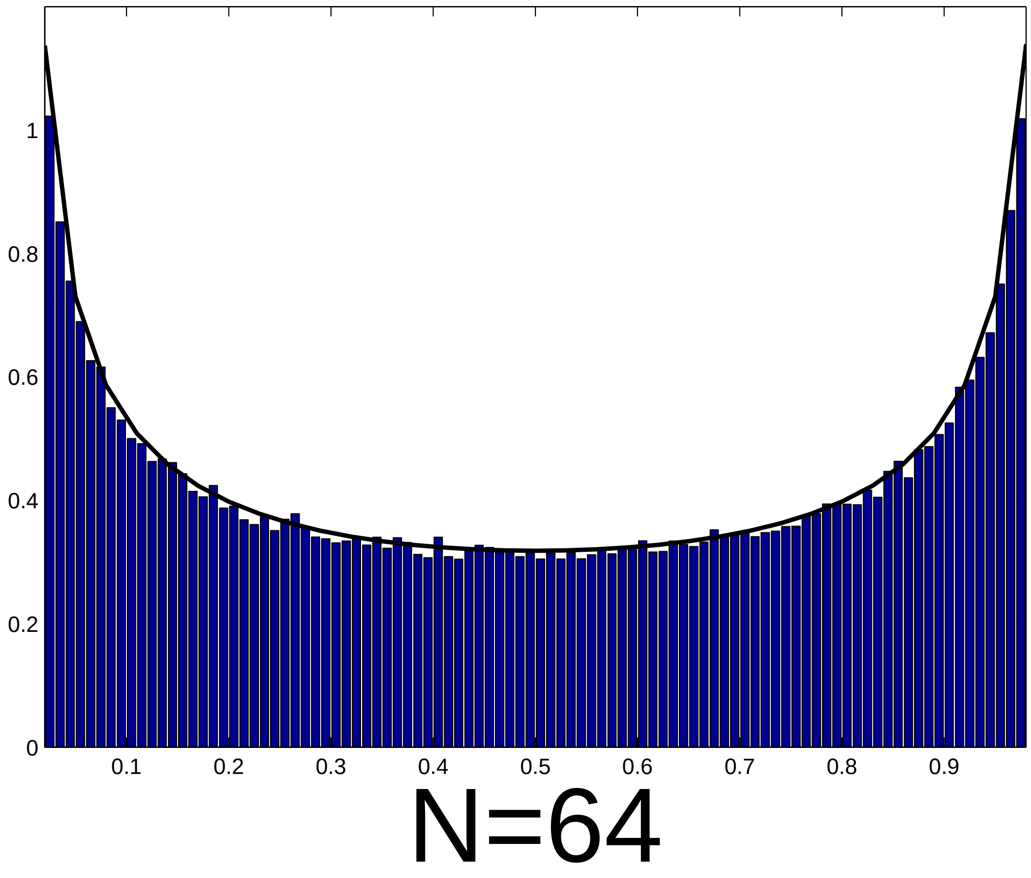
$$S_{AB}(z) = S_A(z) \cdot S_B(z).$$

Example: Wishart x Wishart ($M = 5N$), trials=1000



N=100

upper left corner of size $N/2 \times N/2$ of a projection matrix, with $N/2$ eigenvalues 0 and $N/2$ eigenvalues 1; trials=5000



- Free Calculator by Raj Rao and Alan Edelman

- A. Nica and R. Speicher: Lectures on the Combinatorics of Free Probability.
  To appear soon in the *London Mathematical Society Lecture Note Series*, vol. 335, Cambridge University Press

Outlook on other talks around free probability

- Anshelevich: "free" orthogonal and Meixner polynomials
- Burda: free random Levy matrices
- Chatterjee: concentration of measures and free probability
- Demni: free stochastic processes
- Kargin: large deviations in free probability
- Mingo + Speicher: fluctuations of random matrices
- Rashidi Far: operator-valued free probability theory and block matrices