Random Matrices: Universality of Local Eigenvalues Statistics

Van H. Vu

Department of Mathematics
Yale University

(joint works with Terence Tao, UCLA)

Overview

(1) History/models of random matrices.

Overview

(1) History/models of random matrices.

(2) A hierarchy of classical problems concerning limiting distribution of the eigenvalues.

Overview

(1) History/models of random matrices.

(2) A hierarchy of classical problems concerning limiting distribution of the eigenvalues.

(3) Notion of universality .

Overview

(1) History/models of random matrices.

(2) A hierarchy of classical problems concerning limiting distribution of the eigenvalues.

(3) Notion of universality .

(4) Brief survey of recent results.

Overview

(5) Four Moment Theorem, which (roughly) states that the (limiting) joint distribution of any set of k (ordered) eigenvalues depend only on the first four moments of the entry distributions.

Overview

(5) Four Moment Theorem, which (roughly) states that the (limiting) joint distribution of any set of k (ordered) eigenvalues depend only on the first four moments of the entry distributions.

(6) Applications.

(5) Four Moment Theorem, which (roughly) states that the (limiting) joint distribution of any set of k (ordered) eigenvalues depend only on the first four moments of the entry distributions.

(6) Applications.

(7) Non-hermitian model. Universality of correlation functions; Number of real eigenvalues.

Models of random matrices

Random covariance matrix (Wishart matrix). $H_{n,p}$ is an $n \times p$ matrix with independent entries having mean zero and variance one. Set

$$G_n = H_{n,p} H_{n,p}^*.$$

Random (hermitian) matrix with independent entries. M_n is an $n \times n$ hermitian matrix with independent (upper triangular) entries having mean zero and variance one. (The diagonal entries may have a different variance.)

Random (non-hermitian) matrix with independent entries. M_n is an $n \times n$ matrix with n^2 independent entries having mean zero and variance one.

Motivations:

- Statistics (Wishart 1928) Sample covariance matrices; Hypothesis testing:
- Numerical analysis (von Neuman-Goldstine 1947) Complexity of linear algebraic algorithms on random inputs.
- Physics (Wigner 1950)

More recent motivations/connections

- Probability theory (limiting distributions)
- Spectral graph theory (relations between eigenvalues and graph properties)
- Number theory (distribution of roots of zeta functions)
- Combinatorics (longest increasing subsequences)
- Computer science/data analysis (effect of random noise; analysis of random data)

Hermitian case: Wigner matrices

A *Wigner Hermitian matrix* (of size n) is a random Hermitian $n \times n$ matrix M_n with upper triangular complex entries $\zeta_{ij} := \xi_{ij} + \sqrt{-1}\tau_{ij}$ $(1 \leq i < j \leq n)$ and diagonal real entries ξ_{ii} $(1 \leq i \leq n)$ where

- For $1 \leq i < j \leq n$, ξ_{ij}, τ_{ij} are iid copies of a real random variable ξ with mean zero and variance $1/2$.
- For $1 \leq i \leq n$, ξ_{ii} are iid copies of a real random variable $\tilde{\xi}$ with mean zero and variance σ^2.

Generalizations

We will focus on the Wigner model for the sake of convenience.

Generalizations

We will focus on the Wigner model for the sake of convenience.

The results hold for more general models, both real and complex.

Similar results hold for $G_n = H_{n,p} H_{n,p}^T$ (Wishart model; hermitian but not independent) .

Scaling

We refer to

- $\xi, \tilde{\xi}$ as the *atom distributions* of M_n, and ξ_{ij}, τ_{ij} as the *atom variables*;
- $W_n := \frac{1}{\sqrt{n}} M_n$ as the *coarse-scale normalized Wigner Hermitian matrix*;
- $A_n := \sqrt{n} M_n$ as the *fine-scale normalized Wigner Hermitian matrix*.

It is well known that $\|M_n\| = \Theta(\sqrt{n})$ with high probability.

- The coarse-scale normalization W_n places all the eigenvalues in a bounded interval $[-2, 2]$.
- The fine-scale normalization A_n keeps the spacing between adjacent eigenvalues to be roughly of unit size.

Gaussian matrices

Example

An important special case of a Wigner Hermitian matrix is the *gaussian unitary ensemble* (GUE), in which $\xi, \tilde{\xi}$ are gaussian random variables with mean zero and variance $1/2$, 1 respectively.

If one consider matrices with real Gaussian entries, then the corresponding example is the *gaussian orthogonal ensemble* (GOE).

The goal of the theory

The main goal of the theory of random matrices is to understand the behavior of the eigenvalues.

We consider the eigenvalues in increasing order

$$\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n.$$

The Global Distribution: Wigner Semi-Circle Law

The global distribution of the eigenvalues is well understood. Denote by ρ_{sc} the semi-circle density function with support on $[-2, 2]$,

$$\rho_{sc}(x) := \begin{cases} \frac{1}{2\pi}\sqrt{4 - x^2}, & |x| \leq 2 \\ 0, & |x| > 2. \end{cases} \tag{1}$$

Theorem (Wigner Semi-circular law, 1950s; Full generality: Pastur 60s)

Let M_n be a Wigner Hermitian matrix. Then for any real number x,

$$\lim_{n\to\infty} \frac{1}{n}|\{1 \leq i \leq n : \lambda_i(W_n) \leq x\}| = \int_{-2}^{x} \rho_{sc}(y)\, dy$$

in the sense of probability (and also in the almost sure sense, if the M_n are all minors of the same infinite Wigner Hermitian matrix).

The Global Distribution: Wigner Semi-Circle Law

The global distribution of the eigenvalues is well understood. Denote by ρ_{sc} the semi-circle density function with support on $[-2, 2]$,

$$\rho_{sc}(x) := \begin{cases} \frac{1}{2\pi}\sqrt{4-x^2}, & |x| \le 2 \\ 0, & |x| > 2. \end{cases} \tag{1}$$

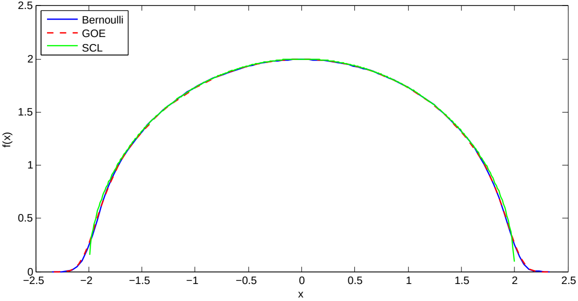
Theorem (Wigner Semi-circular law, 1950s; Full generality: Pastur 60s)

Let M_n be a Wigner Hermitian matrix. Then for any real number x,

$$\lim_{n\to\infty} \frac{1}{n}|\{1 \le i \le n : \lambda_i(W_n) \le x\}| = \int_{-2}^{x} \rho_{sc}(y)\ dy$$

in the sense of probability (and also in the almost sure sense, if the M_n are all minors of the same infinite Wigner Hermitian matrix).

The Global Distribution: Wigner Semi-Circle Law



Local Statistics

The main open problem is to understand the Local Statistics.

- Distribution of the gaps between consecutive eigenvalues. (How many $1 \leq i \leq n$ are there such that $\lambda_{i+1} - \lambda_i \leq s$?)

Local Statistics

The main open problem is to understand the Local Statistics.

- Distribution of the gaps between consecutive eigenvalues. (How many $1 \leq i \leq n$ are there such that $\lambda_{i+1} - \lambda_i \leq s$?)
- k-point correlation functions.

Local Statistics

The main open problem is to understand the Local Statistics.

- Distribution of the gaps between consecutive eigenvalues. (How many $1 \leq i \leq n$ are there such that $\lambda_{i+1} - \lambda_i \leq s$?)
- k-point correlation functions.
- Distribution of individual λ_i, for any $1 \leq i \leq n$.

Local Statistics

The main open problem is to understand the Local Statistics.

- Distribution of the gaps between consecutive eigenvalues. (How many $1 \leq i \leq n$ are there such that $\lambda_{i+1} - \lambda_i \leq s$?)
- k-point correlation functions.
- Distribution of individual λ_i, for any $1 \leq i \leq n$.
- More generally, one wants to know the limiting joint distribution of $(\lambda_{i_1}, \ldots, \lambda_{i_k})$, for any given $1 \leq i_1 < ... < i_k \leq n$.

The GUE case: Ginibre's formula

Ginibre's formula of the joint distribution of the eigenvalues (non-ordered)

$$\rho(x_1, \ldots, x_n) = c(n) \prod_{1\le i<j\le n} |x_j - x_i|^2 \exp(-\frac{1}{2}\sum_{i=1}^{n} x_i^2).$$

This is due to the fact that a matrix from GUE enjoys the decomposition

$$M = UDU^*$$

where U is a random unitary matrix and D is an independent diagonal matrix. (The measure of GUE is unitary invariance.)

Many local statistics can be computed directly from the above joint distribution by integration or via the fact that the corresponding point process is determinatal.

Many local statistics can be computed directly from the above joint distribution by integration or via the fact that the corresponding point process is determinatal.

Example. To compute the distribution of the smallest singular value (von Neumann-Goldstine problem)

$$\mathbf{P}(\textit{no eigenvalue in } [-\theta, \theta]) = \int_{(\mathbf{R} \setminus [-\theta,\theta])^n} \rho(x_1, \ldots, x_n) dx_1 \ldots dx_n.$$

(explicit formula given by Jimbo et. al.)

Distribution of the gaps

For a vector $x = (x_1, \ldots, x_n)$ where $x_1 < x_2 \cdots < x_n$, define the normalized gap distribution $S_n(s; x)$ as

$$S_n(s; x) := \frac{1}{n} |\{1 \leq i \leq n : x_{i+1} - x_i \leq s\}|.$$

One is interested in the distribution of $S_n(s; \lambda)$, with $\lambda = (\lambda_1, \ldots, \lambda_n)$. In particular

$$F_1(s) := \mathbf{E} S_n(s, \lambda); F_2(s) = \mathbf{E} S_n^2(s, \lambda); etc$$

k-correlation functions

k-correlation functions

$$\int f(x_1, \ldots, x_k)\rho_n^k(x_1, \ldots, x_k)dx_1 \ldots dx_k = \mathbf{E} \sum_{1 \le i_1 < \cdots < i_k \le n} f(\lambda_{i_1}, \ldots, \lambda_{i_k}).$$

k-correlation functions

$$\int f(x_1, \dots, x_k)\rho_n^k(x_1, \dots, x_k)dx_1 \dots dx_k = \mathbf{E} \sum_{1 \le i_1 < \dots < i_k \le n} f(\lambda_{i_1}, \dots, \lambda_{i_k}).$$

One can also define $\rho_n^{(k)}(x_1, \dots, x_k)$ to be the quantity such that the probability that there is an eigenvalue of M_n in each of the intervals $\{x : |x - x_i| \le \epsilon\}$ for $i = 1, \dots, k$ is asymptotically $(\rho_n^{(k)}(x_1, \dots, x_k) + o(1))(\pi\epsilon^2)^k$ in the limit $\epsilon \to 0^+$.

From the correlation functions, one can compute several interesting statistics (using simple inclusion-exclusion arguments) such as: the gap distribution, the probability that an interval (of length $\Theta(1/n)$) is empty, the probability that an interval has exactly k eigenvalues.

k-point correlation functions

$\int_I \rho_n^1(x)dx$ computes the expectation of the number of eigenvalues in I:

$$\int_I \rho_n^1(x)dx = \sum_{i=1}^{n} \mathbf{P}(\lambda_i \in I).$$

$\int_{I\times J} \rho_n^2(x,y)dxdy$ computes the expectation of the number of pairs of eigenvalues λ_i, λ_j where $\lambda_i \in I, \lambda_j \in J$.

$$\int_{I\times J} \rho_n^2(x,y)dxdy = \sum_{i,j=1}^{n} \mathbf{P}(\lambda_i \in I \wedge \lambda_j \in J).$$

Consider a small neighborhood around a point u in the spectrum (nu in the fine-scale model).

$$\lim_{n\to\infty} \frac{1}{\rho_{sc}(u)^k} \int_{\mathbf{R}^k} f(t_1, \ldots, t_k)\rho_n^{(k)}(nu+\frac{t_1}{\rho_{sc}(u)}, \ldots, nu+\frac{t_k}{\rho_{sc}(u)})\, dt_1 \ldots dt$$

Correlation in a small neighborhood around u.

Joint distribution of few eigenvalues

For any $1 \leq i \leq n$, consider λ_i in the ordered sequence

$$\lambda_1 \leq \cdots \leq \lambda_n.$$

Does $\frac{\lambda_i - \mu(i,n)}{\sigma(i,n)} \to$ a limiting distribution ? What is the limiting distribution ?

Joint distribution of few eigenvalues

For any $1 \leq i \leq n$, consider λ_i in the ordered sequence

$$\lambda_1 \leq \cdots \leq \lambda_n.$$

Does $\frac{\lambda_i - \mu(i,n)}{\sigma(i,n)} \rightarrow$ a limiting distribution ? What is the limiting distribution ?

In general one wants to know the joint distribution of $(\lambda_{i_1}, \ldots, \lambda_{i_k})$ (after a proper normalization), for any $1 \leq i_1 < \cdots < i_k \leq n$.

A hierarchy of problems

From the k-point correlation one can compute the gap distribution (inclusion-exclusion).

From the joint distribution of any (ordered) k eigenvalues one can (in principle) deduce the k-point correlation functions.

$$\int_{I \times J} \rho_n^2(x, y) dxdy = \sum_{i,j=1}^{n} \mathbf{P}(\lambda_i \in I \wedge \lambda_j \in J).$$

A hierarchy of problems

From the k-point correlation one can compute the gap distribution (inclusion-exclusion).

From the joint distribution of any (ordered) k eigenvalues one can (in principle) deduce the k-point correlation functions.

$$\int_{I \times J} \rho_n^2(x, y)dxdy = \sum_{i,j=1}^{n} \mathbf{P}(\lambda_i \in I \wedge \lambda_j \in J).$$

Gap problem $< - - - - -$ k-correlation problem $< - - - - -$ Joint distribution of (ordered) k eigenvalues.

We will first focus on the last problem.

GUE statistics: Gap distribution

It is known (due to Dyson-Mehta) that

$$F(s) := \lim_{n\to\infty} \mathbf{E} S_n(s, \lambda(A_n)) = \int_0^s p(\sigma)d\sigma, \qquad (2)$$

where $A_n := \sqrt{n}M_n$ is the fine-scale normalization of M_n, and $p(\sigma)$ is the *Gaudin distribution*, given by the formula

$$p(s) := \frac{d^2}{ds^2}\det(I-K)_{L^2(0,s)},$$

where K is the integral operator on $L^2((0,s))$ with the *Dyson sine kernel*

$$K(x,y) := \frac{\sin \pi(x-y)}{\pi(x-y)}. \qquad (3)$$

GUE statistics: k-correlation functions

For GUE, it was established by Gaudin and Mehta that

$$\rho_n^{(k)}(x_1, \ldots, x_k) = \det(K_n(x_i, x_j))_{1 \leq i,j \leq k}$$

where the kernel $K_n(x, y)$ is given by the formula

$$K_n(x, y) := \frac{1}{\sqrt{2n}} e^{-\frac{1}{4n}(x^2+y^2)} \sum_{j=0}^{n-1} h_j(\frac{x}{\sqrt{2n}}) h_j(\frac{y}{\sqrt{2n}})$$

and $h_0, \ldots, h_{n-1}$ are the first n Hermite polynomials, normalized to be orthonormal with respect to $e^{-x^2}\, dx$.

GUE statistics: Dyson kernel

From this and the asymptotics of Hermite polynomials, it was shown by Dyson that

$$\lim_{n\to\infty} \frac{1}{\rho_{sc}(u)^k} \rho_n^{(k)}(nu + \frac{t_1}{\rho_{sc}(u)}, \ldots, nu + \frac{t_k}{\rho_{sc}(u)}) = \det(K(t_i, t_j))_{1 \le i,j \le k}, \quad (4)$$

for any fixed $-2 < u < 2$ and real numbers $t_1, \ldots, t_k$, where K is the Dyson sine kerne. (Universal in u.)

GUE statistics: Gaps in a small neighborhood

Let l_n be any sequence of numbers tending to infinity such that l_n/n tends to zero. Define

$$\tilde{S}_n(s;x,u) := \frac{1}{l_n}|\{1 \le i \le n : x_{i+1} - x_i \le \frac{s}{\rho_{sc}(u)}, |x_i - nu| \le \frac{l_n}{\rho_{sc}(u)}\}|. \quad (5)$$

It is proved (Deift et. al.) that for any fixed $-2 < u < 2$, we have

$$\lim_{n\to\infty} \mathbf{E}\tilde{S}_n(s;\lambda(A_n),u) = \int_0^s p(\sigma)d\sigma. \quad (6)$$

GUE statistics: Fluctuation at the edge

The large (edge) eigenvalues (such as λ_1, λ_2 or λ_n) fluctuate according to the Tracy-Widom law. Consider λ_1 of $W_n = \frac{1}{\sqrt{n}} M_n$, so $\lambda_1 \approx -2$. One has

$$(\lambda_1 + 2) n^{2/3} \to TW.$$

An interesting point here is that the fluctuation is of order $n^{-2/3}$, not n^{-1}. (The semi-circular function decays sharply at the edge.)

Tracy and Widom computed the limiting joint distribution of $(\lambda_1, \ldots, \lambda_k)$ (after a proper normalization).

GUE statistics: Fluctuation in the bulk

Gustavsson (2005) (based on earlier works of Soshnyikov and Costin-Lebowitz) proved that a bulk eigenvalue has gaussian fluctuation.

To be precise choose an index $i = i(n)$ such that $i/n \to c$ as $n \to \infty$ for some $0 < c < 1$, let M_n be drawn from the GUE and $A_n := \sqrt{n} M_n$. Then

$$\sqrt{\frac{4 - t(c)^2}{2}} \frac{\lambda_i(A_n) - t(c)n}{\sqrt{\log n}} \to N(0, 1)$$

in the sense of distributions, where $t(c)$ can be computed from the semi-circular function. More informally,

$$\lambda_i(M_n) \approx t(c)\sqrt{n} + N(0, \frac{2 \log n}{(4 - t(c)^2)n}).$$

The result extends to the joint distribution of k eigenvalues.

The Universality Phenomenon

It is generally believed (with strong numerical evidence) that the local distributions are Universal, namely that results such as the above should hold for Wigner matrices, or even more general classes of random matrices.

The Universality Phenomenon

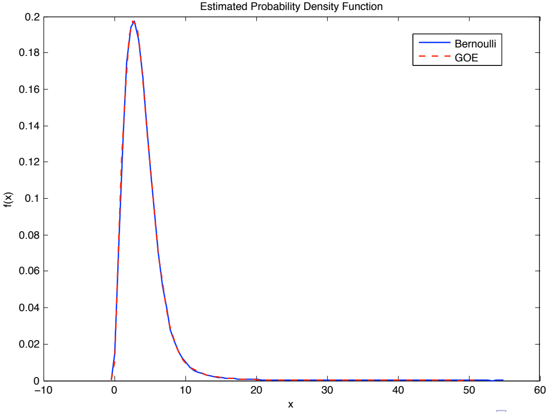
It is generally believed (with strong numerical evidence) that the local distributions are Universal, namely that results such as the above should hold for Wigner matrices, or even more general classes of random matrices.

The Universality Phenomenon

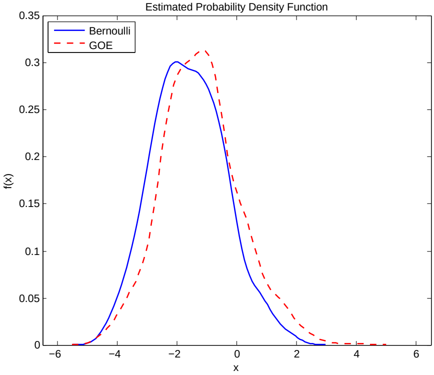
It is generally believed (with strong numerical evidence) that the local distributions are Universal, namely that results such as the above should hold for Wigner matrices, or even more general classes of random matrices.

We will focus on the bulk of the spectrum (interval $[-2+\epsilon, 2-\epsilon]$).

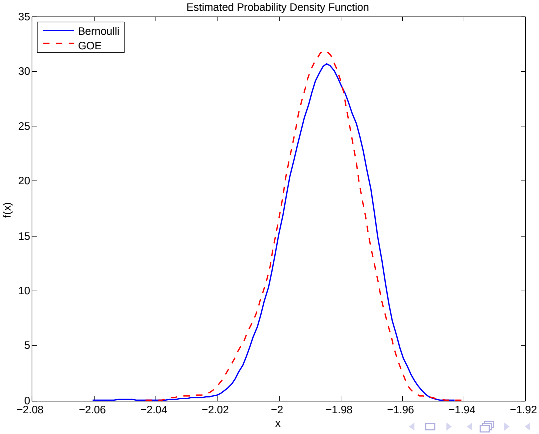
Numerical evidence: Gap distributions



Numerical evidence: Distribution at the edge



Numerical evidence: Distribution in the bulk



Universality Results: Gauss divisible matrices

Consider a "mixed" model

$$M_n = c_1 H_n + c_2 G_n$$

where H_n is a Wigner matrix, and G_n is from GUE, c_1, c_2 are positive constants such that $c_1^2 + c_2^2 = 1$.

Theorem (Johansson 2001)

The gap distribution of a Johansson matrix is the same as for GUE as $n \to \infty$. The k-point correlation is also universal in the weak sense

$$\lim_{n\to} \int f(x_1, \dots, x_k) \rho_n^{(k)}(x_1, \dots, x_k) dx_1 \dots dx_k \to$$

$$\int f(x_1, \dots, x_k) \det(K_n(x_i, x_j))_{1 \le i,j \le k} dx_1 \dots dx_k.$$

Method. Johansson matrices still admit an explicit joint distribution

Universality Results: Gauss divisible matrices

Consider a "mixed" model

$$M_n = c_1 H_n + c_2 G_n$$

where H_n is a Wigner matrix, and G_n is from GUE, c_1, c_2 are positive constants such that $c_1^2 + c_2^2 = 1$.

Theorem (Johansson 2001)

The gap distribution of a Johansson matrix is the same as for GUE as $n \to \infty$. The k-point correlation is also universal in the weak sense

$$\lim_{n\to} \int f(x_1, \ldots, x_k)\rho_n^{(k)}(x_1, \ldots, x_k)dx_1 \ldots dx_k \to$$

$$\int f(x_1, \ldots, x_k) \det(K_n(x_i, x_j))_{1\le i,j\le k} dx_1 \ldots dx_k.$$

Method. Johansson matrices still admit an explicit joint distribution

Recently (2009-2011) Erdős et. al. proved average universality of k correlation functions under various assumptions

$$\lim_{b\to 0}\frac{1}{2b}\int_{u-b}^{u+b}\lim_{n\to\infty}\int f(x_1,\dots,x_k)\rho_n^{(k)}(x_1,\dots,x_k)dx_1\dots dx_k \to$$

$$\int f(x_1,\dots,x_k)\det(K_n(x_i,x_j))_{1\le i,j\le k}dx_1\dots dx_k.$$

The original assumption was that the atom variables are very smooth (6 times differentiable with polynomially bounded derivatives) with some Iob-Sobolev property.
The newest improvement (Erdős-Jin-Yau, 2010-2011) removed the smooth assumption completely. The method works for real and quaternion matrices as well.

Recently (2009-2011) Erdős et. al. proved average universality of k correlation functions under various assumptions

$$\lim_{b\to 0}\frac{1}{2b}\int_{u-b}^{u+b}\lim_{n\to\infty}\int f(x_1,\dots,x_k)\rho_n^{(k)}(x_1,\dots,x_k)dx_1\dots dx_k \to$$

$$\int f(x_1,\dots,x_k)\det(K_n(x_i,x_j))_{1\le i,j\le k}dx_1\dots dx_k.$$

The original assumption was that the atom variables are very smooth (6 times differentiable with polynomially bounded derivatives) with some log-Sobolev property.
The newest improvement (Erdős-Jin-Yau, 2010-2011) removed the smooth assumption completely. The method works for real and quaternion matrices as well.

Problem. The extra averaging (double integral).

Rate of Convergence

In 2002, Bai, Miao, Tsay showed that the spectrum of M_n converges to Wigner law at rate $n^{-1/2}$, using the Stieljes transform method.

Rate of Convergence

In 2002, Bai, Miao, Tsay showed that the spectrum of M_n converges to Wigner law at rate $n^{-1/2}$, using the Stieljes transform method.

In 2008, Erdős-Schlein-Yau refined this method to show strong rate $n^{-1+\epsilon}$.

Their techniques play an important role in both our and Erdős et. al. approach to the universality conjectures, in different ways.

Works on invariant models:

$$\rho(x_1, \ldots, x_n) = c(n) \prod_{1 \le i < j \le n} |x_j - x_i|^\beta \exp(-V(x_1, \ldots, x_n)).$$

Deift, Kriecherbauer, McLaughlin, Venakides and Zhou, Pastur and Shcherbina, Bleher and Its (universality in u and V).

The four moment theorem

Informal: If the first four moments of two atom variables ξ and ξ' match, then the joint distribution of $(\lambda_{i_1}, \ldots, \lambda_{i_k})$ and $(\lambda'_{i_1}, \ldots, \lambda'_{i_k})$ are asymptotically the same, for any $1 \leq i_1 < \cdots < i_k \leq n$.

The four moment theorem

Informal: If the first four moments of two atom variables ξ and ξ' match, then the joint distribution of $(\lambda_{i_1}, \dots, \lambda_{i_k})$ and $(\lambda'_{i_1}, \dots, \lambda'_{i_k})$ are asymptotically the same, for any $1 \leq i_1 < \cdots < i_k \leq n$.

One can deduce

- Universality of gaussian fluctuation, under Four Moments assumption.

The four moment theorem

Informal: If the first four moments of two atom variables ξ and ξ' match, then the joint distribution of $(\lambda_{i_1}, \ldots, \lambda_{i_k})$ and $(\lambda'_{i_1}, \ldots, \lambda'_{i_k})$ are asymptotically the same, for any $1 \leq i_1 < \cdots < i_k \leq n$.

One can deduce

- Universality of gaussian fluctuation, under Four Moments assumption.
- Universality of Tracy-Widom law (at the edge), under Two Moments assumption (and no symmetry).

The four moment theorem

Informal: If the first four moments of two atom variables ξ and ξ' match, then the joint distribution of $(\lambda_{i_1}, \dots, \lambda_{i_k})$ and $(\lambda'_{i_1}, \dots, \lambda'_{i_k})$ are asymptotically the same, for any $1 \leq i_1 < \cdots < i_k \leq n$.

One can deduce

- Universality of gaussian fluctuation, under Four Moments assumption.
- Universality of Tracy-Widom law (at the edge), under Two Moments assumption (and no symmetry).
- Universality of k-point correlation and Gap distribution, under Two moments assumption.

Definition (Moment matching)

We say that two complex random variables ξ and ξ' *match to order* k if

$$\mathbf{E}\Re(\zeta)^m \Im(\zeta)^l = \mathbf{E}\Re(\zeta')^m \Im(\zeta')^l$$

for all $m, l \geq 0$ such that $m + l \leq k$.

Four moment theorem, informal version

Theorem (Tao-V. 2009)

There is a positive constant c such that for every fixed $k \geq 1$ the following holds. Let $M_n = (\zeta_{ij})_{1 \leq i,j \leq n}$ and $M'_n = (\zeta'_{ij})_{1 \leq i,j \leq n}$ be two random matrices satisfying

- $\mathbf{E}\zeta_{ij} = 0, \mathbf{E}|\zeta_{ij}|^2 = 1$.
- $\mathbf{E}|\zeta_{ij}|^C, \mathbf{E}|\zeta'_{ij}|^C < \infty$ *for a sufficiently large C ($C = 10^4$).*
- *Any $1 \leq i < j \leq n$, ζ_{ij} and ζ'_{ij} match to order 4 and for any $1 \leq i \leq n$, ζ_{ii} and ζ'_{ii} match to order 2.*

Then for any k tuples $1 \leq i_1 < \cdots < i_k \leq n$ and a "nice" domain $D \in \mathbf{R}^k$,

$$|\mathbf{P}((\lambda_{i_1}, \ldots, \lambda_{i_k}) \in D) - \mathbf{P}((\lambda'_{i_1}, \ldots, \lambda'_{i_k}) \in D)| \leq n^{-c}.$$

Remark. We always assume the first moment is zero and the second is 1, so the matching assumption is actually about the third and forth moments.

Four moment theorem, formal version

Theorem (Four Moment Theorem)

There is a small positive constant c_0 such that for integer $k \geq 1$ the following holds. Let $M_n = (\zeta_{ij})_{1 \leq i,j \leq n}$ and $M'_n = (\zeta'_{ij})_{1 \leq i,j \leq n}$ be two random matrices satisfying **C1** *(with some sufficiently large parameter C_0). Assume furthermore that for any $1 \leq i < j \leq n$, ζ_{ij} and ζ'_{ij} match to order 4 and for any $1 \leq i \leq n$, ζ_{ii} and ζ'_{ii} match to order 2. Set $A_n := \sqrt{n} M_n$ and $A'_n := \sqrt{n} M'_n$, and let $G : \mathbf{R}^k \to \mathbf{R}$ be a smooth function obeying the derivative bounds*

$$|\nabla^j G(x)| \leq n^{c_0} \qquad (7)$$

for all $0 \leq j \leq 5$ and $x \in \mathbf{R}^k$. Then for any $1 \leq i_1 < i_2 \cdots < i_k \leq n$, and for n sufficiently large we have

$$|\mathbf{E}(G(\lambda_{i_1}(A_n), \dots, \lambda_{i_k}(A_n))) - \mathbf{E}(G(\lambda_{i_1}(A'_n), \dots, \lambda_{i_k}(A'_n)))| \leq n^{-c_0}. \qquad (8)$$

Consequences: Universality of Gaussian Fluctuation

Corollary (Universality of gaussian fluctuation)

The conclusion of Gustavsson theorem holds for any other Wigner Hermitian matrix M_n whose atom distribution ξ satisfies $\mathbf{E}\xi^3 = 0$ and $\mathbf{E}\xi^4 = \frac{3}{4}$. In words, the bulk eigenvalues of such a matrix has gaussian fluctuation.

The same statement holds for the universality of the asymptotic joint distribution law for any k eigenvalues $\lambda_{i_1}(M_n), \ldots, \lambda_{i_k}(M_n)$ in the bulk of the spectrum of a Wigner Hermitian matrix for any fixed k (the GUE case is treated by Gustavsson). The real case is treated recently by O' Rourke (2010)

Consequences: Universality of the gap distribution

Corollary (Universality of the gap distribution)

The limiting gap distribution of GUE holds for random Hermitian matrices whose whose entries have mean zero, variance one and bounded Cth moment, for a sufficiently large constant C and has support on at least 3 points.

Note that in contrast to previous applications, we are making NO assumptions on the third and fourth moments of the atom distribution ξ.

The extra observation here is that we do not always need to compare M_n with GUE . We can compare M_n with any model where the desired statistics have been computed.

The extra observation here is that we do not always need to compare M_n with GUE . We can compare M_n with any model where the desired statistics have been computed.

In this case, we are going to compare M_n with a Johansson matrix. The definition of Johansson matrices provides etra degrees of freedom, and we can use them to remove the condition of the third and fourth moments.

Recall that the atom variable in a Johansson matrix is of the form $c_1\xi' + c_2 N(0,1)$, where $c_1^2 + c_2^2 = 1$. Given a Wigner matrix M_n with atom variable ξ, we want to show that there is a Johansson variable that match ξ in the first four moments. This is a special case of the classical truncated moment problem (Hamburger).

Lemma (Truncated moment matching problem)

For any variable ξ with mean 0 and variance 1 and support on at least 3 points, there is a random variable ξ' with mean 0 and variance 1, and two numbers $0 < c_1, c_2 < 1$ such that $c_1^2 + c_2^2 = 1$ and the first four moments of $c_1\xi' + c_2 N(0,1)$ and ξ match.

One can remove the 3 points assumption (ERSTVY) by an additional trick.

Consequences: Universality of the k-correlation function

One can use an asymptotic version of the Four moment theorem, combined with the above argument and a recent localization result of Erdos-Yau-Yin to prove the (local) Unviversality of the k-point correlation function.

Theorem (Universality of correlation function)

Fix $\epsilon > 0$ and u such that $-2 < u - \epsilon < u + \epsilon < 2$. Let $k \geq 1$ and let $f : \mathbf{R}^k \to \mathbf{R}$ be a continuous, compactly supported function, and let $M = M_n$ be a Wigner random matrix whose entries have mean zero, variance one and bounded Cth moment, for a sufficiently large C. Then

$$\lim_{n\to\infty} \frac{1}{\rho_{sc}(u)^k} \int_{\mathbf{R}^k} f(t_1, \dots, t_k)\rho_n^{(k)}(nu + \frac{t_1}{\rho_{sc}(u)}, \dots, nu + \frac{t_k}{\rho_{sc}(u)})\, dt_1 \dots dt_k$$

$$\to \int f(x_1, \dots, x_k) \det(K_n(x_i, x_j))_{1\leq i,j\leq k} dx_1 \dots dx_k.$$

Mehta conjecture (1967) for Hermitian complex matrices.

Mehta conjecture (1967) for Hermitian complex matrices.

Extension to generalized Wigner matrices: Erdős-Yau (2012).

Mehta conjecture (1967) for Hermitian complex matrices.

Extension to generalized Wigner matrices: Erdős-Yau (2012).

Open question. Mehta conjecture for random real symmetric matrices ?

The above theorem holds with 4 moment assumption (as one does not have Johansson type results).

Extensions of Four Moment theorem

Beyond the Wigner model. The entries do not need to be iid, and can be either real or complex. In the complex case, the real part and the imaginary part are not necessarily independent.

Random covariance matrices. Four moment theorem holds for Random covariance matrices, as far as $\lim \frac{p}{n}$ tends to a positive constant.

Necessity of Four moments

Theorem

Let W_n, W_n' be Wigner matrices whose atom variables η, η' satisfy $\mathbf{E}\eta^3 = \mathbf{E}(\eta')^3 = 0$ but their fourth moments are different $\mathbf{E}\eta^4 \neq \mathbf{E}(\eta')^4$. As before, write $\lambda_i := \lambda_i(W_n)$ and $\lambda_i' := \lambda_i(W_n')$. Then for all sufficiently large n, one has

$$\sum_{i=1}^{n} |\mathbf{E}\lambda_i - \mathbf{E}\lambda_i'| \geq \kappa$$

for some κ depending only on the atom distributions.

Necessity of Four moments

Theorem

Let W_n, W_n' be Wigner matrices whose atom variables η, η' satisfy $\mathbf{E}\eta^3 = \mathbf{E}(\eta')^3 = 0$ but their fourth moments are different $\mathbf{E}\eta^4 \neq \mathbf{E}(\eta')^4$. As before, write $\lambda_i := \lambda_i(W_n)$ and $\lambda_i' := \lambda_i(W_n')$. Then for all sufficiently large n, one has

$$\sum_{i=1}^{n} |\mathbf{E}\lambda_i - \mathbf{E}\lambda_i'| \geq \kappa$$

for some κ depending only on the atom distributions.

On average, eigenvalues move by a considerable amount once the forth moment changes.

Eigenvalues' dance

For $\delta n \leq i \leq (1-\delta)n$ with $\delta > 0$ fixed, one has

$$\mathbf{E}\lambda_i = n^{1/2}\gamma_i + n^{-1/2}C_{i,n} + \frac{1}{4\sqrt{n}}(\gamma_i^3 - 2\gamma_i)\mathbf{E}\eta^4 + O_\delta(n^{-1/2-c})$$

for some absolute constant $c > 0$, where $C_{i,n}$ is some bounded quantity depending only on i, n (in particular independent of η).

This conjecture would imply that if one increases the forth moment, then (in expectation) those λ_i with $\gamma_i \leq -\sqrt{2}$ or $0 \leq \gamma_i \leq \sqrt{2}$ are shift to the left (decreasing), while those λ_i with $\gamma_i \geq \sqrt{2}$ or $0 \geq \gamma_i \geq -\sqrt{2}$ are shifted to the right (increasing). In other words, the eigenvalues in the middle move toward the center of the spectrum, while those closer to the edge move outward.

Non-hermitian matrices

M_n has n^2 iid entries with mean 0 and variance 1. The eigenvalues are now typically complex .

Non-hermitian matrices

M_n has n^2 iid entries with mean 0 and variance 1. The eigenvalues are now typically complex .

Global law: Circular Law. The limiting distribution of the spectrum of $\frac{1}{\sqrt{n}} M_n$ is uniform on the unit circle.

Non-hermitian matrices

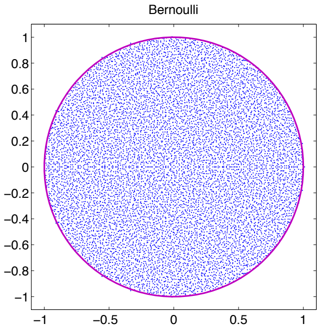
M_n has n^2 iid entries with mean 0 and variance 1. The eigenvalues are now typically complex .

Global law: Circular Law. The limiting distribution of the spectrum of $\frac{1}{\sqrt{n}} M_n$ is uniform on the unit circle.
(Conjectured in the 1950s.)

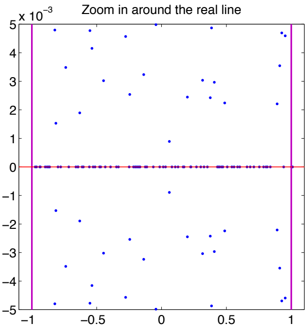
Partial results: Ginibre (1960s), Edelman, Girko (80s), Bai (90s), Götze-Tikhomirov, Pan-Zhou, Tao-V. (around 2005).

Full generality: Tao-V. (2006). Important ingredient. Inverse Littlewood-Offord theorems from additive combinatorics.

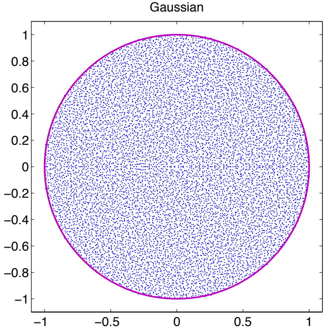
Extension. Krisnapur-Tao-V., Soshnikov et. al., Götze-Tikhomirov, Chafai, Hoi Nguyen, etc. (2006-2012)

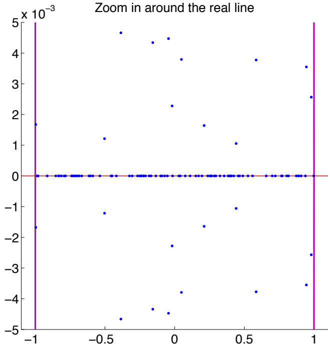


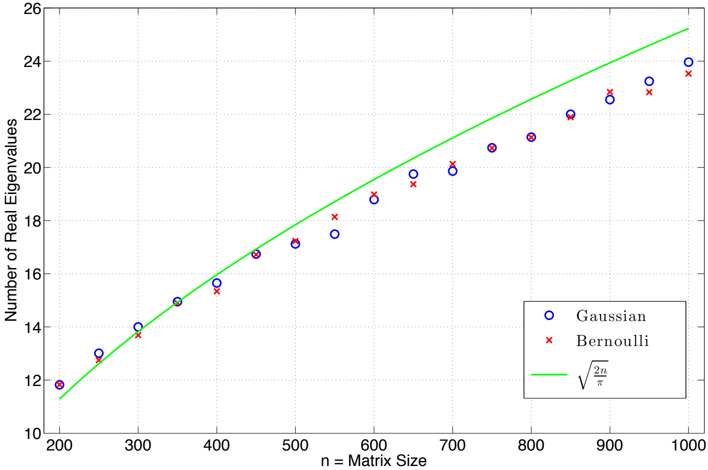
Bernoulli
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
−1
−0.5
0
0.5
1



Zoom in around the real line
x 10⁻³
5
4
3
2
1
0
−1
−2
−3
−4
−5
−1
−0.5
0
0.5
1


Gaussian
1
0.8
0.6
0.4
0.2
0
−0.2
−0.4
−0.6
−0.8
−1
−1
−0.5
0
0.5
1


Zoom in around the real line
x 10^−3
5
4
3
2
1
0
−1
−2
−3
−4
−5
−1
−0.5
0
0.5
1



26
24
22
20
18
16
14
12
10
200
300
400
500
600
700
800
900
1000
Number of Real Eigenvalues
n = Matrix Size
Gaussian
Bernoulli
$\sqrt{\frac{2n}{\pi}}$

Matrix with complex entries: Correlation functions

Definition.

$$\int_{\mathbf{C}^k} F(z_1,\ldots,z_k)\rho_n^{(k)}(z_1,\ldots,z_k)\ dz_1\ldots dz_k$$
$$= \mathbf{E} \sum_{1\le i_1<\cdots<i_k\le n} F(\lambda_{i_1}(M_n),\ldots,\lambda_{i_k}(M_n)) \qquad (9)$$

for all continuous, compactly supported test functions F.

Definition.

$$\int_{\mathbf{C}^k} F(z_1, \dots, z_k)\rho_n^{(k)}(z_1, \dots, z_k)\ dz_1 \dots dz_k = \mathbf{E} \sum_{1 \le i_1 < \dots < i_k \le n} F(\lambda_{i_1}(M_n), \dots, \lambda_{i_k}(M_n)) \tag{9}$$

for all continuous, compactly supported test functions F.

Equivalent: $\rho_n^{(k)}(z_1, \dots, z_k)$ for distinct $z_1, \dots, z_k$ is the quantity such that the probability that there is an eigenvalue of M_n in each of the disks $\{z : |z - z_i| \le \epsilon\}$ for $i = 1, \dots, k$ is asymptotically $(\rho_n^{(k)}(z_1, \dots, z_k) + o(1))(\pi\epsilon^2)^k$ in the limit $\epsilon \to 0^+$.

The Gaussian case

The entries ξ_{ij} are iid with the distribution of a complex gaussian $N(0,1)_{\mathbf{C}}$ with mean zero and variance one. The correlation functions of a complex gaussian matrix are given by the explicit formula

$$\rho_n^{(k)}(z_1,\dots,z_k) = \det(K_n(z_i,z_j))_{1\le i,j\le k} \tag{10}$$

where $K_n : \mathbf{C} \times \mathbf{C} \to \mathbf{C}$ is the kernel

$$K_n(z,w) := \frac{1}{\pi} e^{-(|z|^2+|w|^2)/2} \sum_{j=0}^{n-1} \frac{(z\overline{w})^j}{j!}. \tag{11}$$

In particular, one has

$$\rho_n^{(1)}(z) = K_n(z,z) = \frac{1}{\pi} e^{-|z|^2} \sum_{j=0}^{n-1} \frac{|z|^{2j}}{j!}. \qquad (12)$$

Thus (by Taylor expansion of $e^{-|z|^2}$) one has the asymptotic

$$\rho_n^{(1)}(\sqrt{n}z) \to \frac{1}{\pi} 1_{|z| \leq 1}$$

as $n \to \infty$ for almost every $z \in \mathbf{C}$. This gives the *circular law* for complex gaussian matrices.

Four Moment theorem for the correlation functions

Recently (2012), we succeeded to prove a Four moment theorem for non-hermitian matrices:

Four Moment theorem for the correlation functions

Recently (2012), we succeeded to prove a Four moment theorem for non-hermitian matrices:

Theorem (Four Moment Theorem for complex matrices)

Let $M_n, \tilde{M}_n$ be independent-entry matrix ensembles with independent real and imaginary parts, having vanishing third moment, and the fourth moments match. Then for any compactly support continuous test function F

$$\int_{\mathbf{C}^k} F(w_1, \dots, w_k)\rho_n^{(k)}(\sqrt{n}z_1 + w_1, \dots, \sqrt{n}z_k + w_k)\, dw_1 \dots dw_k =$$
$$\int_{\mathbf{C}^k} F(w_1, \dots, w_k)\tilde{\rho}_n^{(k)}(\sqrt{n}z_1 + w_1, \dots, \sqrt{n}z_k + w_k)\, dw_1 \dots dw_k + o(1).$$

Four Moment theorem for the correlation functions

Recently (2012), we succeeded to prove a Four moment theorem for non-hermitian matrices:

Theorem (Four Moment Theorem for complex matrices)

Let $M_n, \tilde{M}_n$ be independent-entry matrix ensembles with independent real and imaginary parts, having vanishing third moment, and the fourth moments match. Then for any compactly support continuous test function F

$$\int_{\mathbf{C}^k} F(w_1, \dots, w_k)\rho_n^{(k)}(\sqrt{n}z_1 + w_1, \dots, \sqrt{n}z_k + w_k)\, dw_1 \dots dw_k =$$
$$\int_{\mathbf{C}^k} F(w_1, \dots, w_k)\tilde{\rho}_n^{(k)}(\sqrt{n}z_1 + w_1, \dots, \sqrt{n}z_k + w_k)\, dw_1 \dots dw_k + o(1).$$

Better (polynomial) error term if F is smooth.

Matrices with real entries: Correlation functions

The real entries have the effect of forcing the spectrum $\lambda_1(M_n), \ldots, \lambda_n(M_n)$ to split into a union of real eigenvalues and pairs of conjugate complex eigenvalues.

Because of this additional structure of the eigenvalues, we need to work with the correlation functions $\rho_n^{(k,l)} : \mathbf{R}^k \times \mathbf{C}_+^l \to \mathbf{R}^+$, defined for fixed $k, l \geq 0$ by the formula

$$\int_{\mathbf{R}^k} \int_{\mathbf{C}_+^l} F(x_1, \ldots, x_k, z_1, \ldots, z_l) \rho_n^{(k,l)}(x_1, \ldots, x_k, z_1, \ldots, z_l)\ dx_1 \ldots dx_k dz$$

$$= \mathbf{E} \sum_{1 \leq i_1 < \cdots < i_k \leq N_{\mathbf{R}}[M_n]} \sum_{1 \leq j_1 < \cdots < j_l \leq N_{\mathbf{C}_+}[M_n]} F(\lambda_{i_1,\mathbf{R}}(M_n), \ldots, \lambda_{i_k,\mathbf{R}}(M_n), \lambda_{j_1,\mathbf{C}_+}(M_n), \ldots, \lambda_{j_l,\mathbf{C}_+}(M_n)). \quad (13)$$

Again, one can interpret $\rho_n^{(k,l)}(x_1, \ldots, x_k, z_1, \ldots, z_l)$ as the unique real number such that, as $\epsilon \to 0$, the probability of simultaneously having an eigenvalue of M_n in each of the intervals $(x_i - \epsilon, x_i + \epsilon)$ for $i = 1, \ldots, k$ and in each of the disks $B(z_j, \epsilon)$ for $j = 1, \ldots, l$ is equal to

$$(1 + o(1))\rho_n^{(k,l)}(x_1, \ldots, x_k, z_1, \ldots, z_l)(2\epsilon)^k(\pi\epsilon^2)^l$$

in the limit as $\epsilon \to 0$.

The Gaussian case

When M_n is given by the real gaussian ensemble, the correlation functions $\rho_n^{(k,l)}$ were computed by a variety of methods, for both odd and even n, (Edelman, Sommers-Wieczorec, Borodin-Sinclair etc) building in turn on the foundational work of Ginibre.

The Gaussian case

When M_n is given by the real gaussian ensemble, the correlation functions $\rho_n^{(k,l)}$ were computed by a variety of methods, for both odd and even n, (Edelman, Sommers-Wieczorec, Borodin-Sinclair etc) building in turn on the foundational work of Ginibre.

The precise formulae for these correlation functions are somewhat complicated and involve Pfaffians of a certain 2×2 matrix kernel.

The Gaussian case

When M_n is given by the real gaussian ensemble, the correlation functions $\rho_n^{(k,l)}$ were computed by a variety of methods, for both odd and even n, (Edelman, Sommers-Wieczorec, Borodin-Sinclair etc) building in turn on the foundational work of Ginibre.

The precise formulae for these correlation functions are somewhat complicated and involve Pfaffians of a certain 2×2 matrix kernel.

The parity of n also matters. To avoid technicality we assume n is even.

Four moment theorem for real matrices

Theorem (Four Moment Theorem for real matrices)

Let $M_n, \tilde{M}_n$ be independent-entry matrix ensembles with real coefficients, Let $k, l \geq 0$ be fixed integers, and let let $x_1, \ldots, x_k$ and $z_1, \ldots, z_l \in \mathbf{C}$ be bounded. Assume that n is even and F is a compactly supported continuous test function. Then

$$\int_{\mathbf{R}^k} \int_{\mathbf{C}^l} F(y_1, \ldots, y_k, w_1, \ldots, w_l) \rho_n^{(k,l)}(\sqrt{n}x_1 + y_1, \ldots, \sqrt{n}x_k + y_k, \sqrt{n}z_1 + w_1, \ldots, \sqrt{n}z_l + w_l)\ dw_1 \ldots dw_l dy_1 \ldots dy_k$$
$$= \int_{\mathbf{R}^k} \int_{\mathbf{C}^l} F(y_1, \ldots, y_k, w_1, \ldots, w_l) \tilde{\rho}_n^{(k,l)}(\sqrt{n}x_1 + y_1, \ldots, \sqrt{n}x_k + y_k, \sqrt{n}z_1 + w_1, \ldots, \sqrt{n}z_l + w_l)\ dw_1 \ldots dw_l dy_1 \ldots dy_k + o(1).$$

Application: Number of real eigenvalues

Let $N_{\mathbf{R}}(M_n)$ denote the number of real zeroes of a random matrix M_n. We have the following asymptotics:

Theorem (Real eigenvalues of a real gaussian matrix)

Let M_n be drawn from the real gaussian ensemble. Then

$$\mathbf{E} N_{\mathbf{R}}(M_n) = \sqrt{\frac{2n}{\pi}} + O(1) \ \ \textit{Edelman}$$

and

$$N_{\mathbf{R}}(M_n) = (2 - \sqrt{2})\sqrt{\frac{2n}{\pi}} + o(\sqrt{n}) \ \ \textit{Forrester} - \textit{Nagao}$$

By using the above universality results, we may partially extend this result to more general ensembles:

Theorem (Real eigenvalues of a real matrix)

Let M_n be an independent-entry matrix ensemble with real coefficients which matches moments with the real gaussian matrix ensemble to fourth order. Assume n is even. Then

$$\mathbf{E} N_{\mathbf{R}}(M_n) = \sqrt{\frac{2n}{\pi}} + O(n^{1/2-c})$$

and

$$N_{\mathbf{R}}(M_n) = O(n^{1-c})$$

for some fixed $c > 0$.

In particular, from Chebyshev's inequality, we have

$$N_{\mathbf{R}}(M_n) = \sqrt{\frac{2n}{\pi}} + O(n^{1/2-c'})$$

Ideas: Distribution via bump functions

Assume that we want to estimate $\mathbf{P}(X \in I)$ for some random variable X and an interval I. We have

$$\mathbf{P}(X \in I) = \mathbf{E}\chi_I(X)$$

where $\chi_I(u) = 1$ if $u \in I$ and 0 otherwise.
Replace χ_I be a smooth bump function G, we have

$$\mathbf{P}(X \in I) \approx \mathbf{E}(G(X)).$$

Ideas: Distribution via bump functions

Assume that we want to estimate $\mathbf{P}(X \in I)$ for some random variable X and an interval I. We have

$$\mathbf{P}(X \in I) = \mathbf{E}\chi_I(X)$$

where $\chi_I(u) = 1$ if $u \in I$ and 0 otherwise.
Replace χ_I be a smooth bump function G, we have

$$\mathbf{P}(X \in I) \approx \mathbf{E}(G(X)).$$

As G is smooth, we can make use of Taylor expansion

$$G(u) = G(0) + G'(0)u + G''(0)u^2/2 + \dots.$$

Lindeberg method: Central limit theorem

Lindeberg replacement method

$$S_n := \frac{\xi_1 + \cdots + \xi_n}{\sqrt{n}} \to N(0,1).$$

Trivial for ξ_i being gaussian. Replace ξ_i with gaussian one at a time.

Recent applications: Chatterjee, Krishnapur.

Lindeberg method: Eigenvalues

Fix all but one entry of A_n (say ζ_{pq}). Switch ζ_{pq} to ζ'_{pq}. Show that the impact on λ_i is negligible.

$$\lambda := \lambda_i; z := \zeta_{pq}; \lambda' := \lambda'_i; z' := \zeta'_{pq}.$$

We want to estimate

$$\mathbf{E}G(\lambda) - \mathbf{E}G(\lambda').$$

$$G(\lambda) = G(\lambda(z)) := F(z).$$

$$F(z) = F(0) + F'(0)z + F''(0)z^2/2 + ...$$

$$\mathbf{E}G(\lambda) - \mathbf{E}G(\lambda') = \mathbf{E}F(z) - \mathbf{E}F(z') = F(0)(1-1) + F'(0)(z-z') + \frac{1}{2}F''(0)(z$$

If the first four moments of z and z' match, then the first 5 terms vanish. The remaining term is at most

$$\sup_x |F^v(x)| \mathbf{E}(|z|^5 + |z'|^5).$$

Computing of the derivatives

We use *Hadamard variation formulae* for the derivatives of $\lambda_i(A(t))$, which can be derived for instance by repeatedly differentiating the eigenvector equation $A(t)u_i(A(t)) = \lambda_i(A(t))u_i(A(t))$. The formula for the first derivative is

$$\frac{d}{dt}\lambda_i(A(t)) = u_i(A(t))^* A'(0) u_i(A(t)).$$

The second derivative formula reads

$$\frac{d^2}{dt^2}\lambda_i(A(t)) = -2 \sum_{j \neq i} \frac{|u_i(A(t))^* A'(0) u_j(A(t))|^2}{\lambda_j(A(t)) - \lambda_i(A(t))}.$$

Further derivatives can be computed with recursive formulae.

Lemma (Main Lemma 1)

$|F^k(x)| \leq n^{-k+o(1)}$ *for all fixed* k.

We are talking about $A_n\sqrt{n}M_n$, so $\mathbf{E}(|z|^5 + |z'|^5) = \Theta(n^{5/2})$. Thus,

$$|\mathbf{E}G(\lambda) - \mathbf{E}G(\lambda')| = O(n^{-5/2+o(1)}).$$

Since we have to swap roughly $n^2/2$ times, the total change is

$$O(n^{-1/2+o(1)}) = o(1).$$

There is, however, a big problem with conditioning. Main Lemma 1 only holds if the rest of the matrix (entries we do not swap) is nice.

Bad event. There are two eigenvalues very close to each other.

Notice that the typical gap between two consecutive eigenvalues of $A_n = \sqrt{n} M_n$ is of order $\Theta(1)$

Lemma (Main Lemma 2)

For any constant $c > 0$

For any fixed $1 \leq i \leq n$, $\mathbf{P}(\lambda_{i+1}(A_n) - \lambda_i(A_n) < n^{-c}) \leq n^{-.01}$.

This means *Bad event* happens rarely.

However, not rarely enough, as we have to proceed in roughly $n^2/2$ steps. (The bound $n^{-0.1}$ can be improved somewhat, but it cannot be better than $n^{-1/2}$ by a theoretical reason.

Lemma (Main Lemma 3)

One can put the Bad event into the function F.

This enables us to bound the *Bad event* just once.

Cauchy Interlacing Law. The eigenvalues η_j of a principal $(n-1) \times (n-1)$ interlace λ_i.

$$\lambda_i \leq \eta_i \leq \lambda_{i+1}.$$

Recall that the typical gap $\lambda_{i+1} - \lambda$ between two consecutive eigenvalues of $A_n = \sqrt{n} M_n$ is of order $O(1)$ So $\eta_i - \lambda_i = O(1)$. However, at the edge of the spectrum, the $\lambda_{i+1} - \lambda$ is much bigger. For example $\lambda_2 - \lambda_1 = \Theta(n^{1/3})$, with high probability.

Lemma (Main Lemma 4: Bias Interlacing law)

With high probability $\eta_i - \lambda_i = O(1)$, for all $i \leq n/2$. (A mirrow inequality holds for the other end.)

For example, η_1 is not in the middle of λ_1 and λ_2, it is glued to λ_1.

Tools:

- Linear algebra (Perturbation theory, Matrix analysis: Hadamard'sprinciple.)
- Probability (Sharp concentration: Talagrand inequality, Stieljes transform technique: Erdős-Schlein-Yau (strong rate of convergence to SCL), Bai et. al.)
- Combinatorics (Boosting).
- High dimensional geometry (Berry-Essen bound in high dimension) .

Bound the derivatives.
Bound the "bad" event (gap property).
Put the bad event into the test function.
Bias Interlacing law.