Random Matrices: Universality of Local Eigenvalues Statistics

Van H. Vu

Department of Mathematics
Rutgers University

(joint works with Terence Tao, UCLA)

Overview

(1) History/models of random matrices.

Overview

(1) History/models of random matrices.

(2) Classical problems concerning limiting distribution of the eigenvalues.

Overview

(1) History/models of random matrices.

(2) Classical problems concerning limiting distribution of the eigenvalues.

These problems form a hierarchy. The top problem is to determine the (limiting) joint distribution of a set of k (ordered) eigenvalues, for any fixed k.

Overview

(1) History/models of random matrices.

(2) Classical problems concerning limiting distribution of the eigenvalues.

These problems form a hierarchy. The top problem is to determine the (limiting) joint distribution of a set of k (ordered) eigenvalues, for any fixed k.

(3) Notion of universality . (This will be different from the notion of universality from mathematical physics literature, where it typically refers to the universality of k-correlation functions.)

Overview

(4) Brief survey of recent results.

Overview

(4) Brief survey of recent results.

(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

(4) Brief survey of recent results.

(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. A key point here is that for problems which place lower in our hierarchy , one does not need to consider 4 moments.

Overview

(4) Brief survey of recent results.

(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. A key point here is that for problems which place lower in our hierarchy , one does not need to consider 4 moments.

For instance, the universality of k-correlation functions or the gap distribution will require only a 2 moment assumption.

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 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.)

Statistics (Wishart 1928)

Sample covariance matrices; Hypothesis testing:
$H_{n,p}$ is an $n \times p$ matrix with independent entries having mean zero and variance one. Set

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

G_n is positive definite; It eigenvalues are the square of the singular values of $H_{n,p}$.

Numerical analysis (von Neumann-Goldstine 1947)

Inverting a large $n \times n$ matrix M. Time and accuracy.

Critical parameter: The condition number

$$\kappa(M) = \sigma_1(M)/\sigma_n(M) = \|M\|\|M^{-1}\|.$$

$\sigma_1 \geq \cdots \geq \sigma_n(M)$ are the singular values.

Numerical analysis (von Neumann-Goldstine 1947)

Inverting a large $n \times n$ matrix M. Time and accuracy.

Critical parameter: The condition number

$$\kappa(M) = \sigma_1(M)/\sigma_n(M) = \|M\|\|M^{-1}\|.$$

$\sigma_1 \geq \cdots \geq \sigma_n(M)$ are the singular values.

Question. Distribution of $\kappa(M)$ if M is random ? (A very early example of average case analysis.)

Physics (Wigner 1950)

M_n is a hermitian (symmetric) random matrices with upper diagonal entries being iid random variables with mean 0 and variance 1.

Eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ model energy levels.

More recent motivations/connections

- Probability theory (limiting distributions: Wigner, Mehta, Dyson, Tracy, Widom, etc)
- Spectral graph theory (relations between eigenvalues and graph properties: Hoffman, Lovász, Juha'sz, Fűredi-Komlos etc)
- Number theory (distribution of roots of zeta functions:Montgomery-Dyson, Odlyzko, Katz-Sarnak etc.)
- Combinatorics (longest increasing subsequences, singularity problems: Baik-Deift-Johansson, Okounkov, Kahn-Komlos-Szemerédi etc)
- Computer science/data analysis (effect of random noise: Smale, Demmel, Edelman, Spielman-Teng etc)

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.

Generalization

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

Generalization

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) .

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 $[-2, 2]$,

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

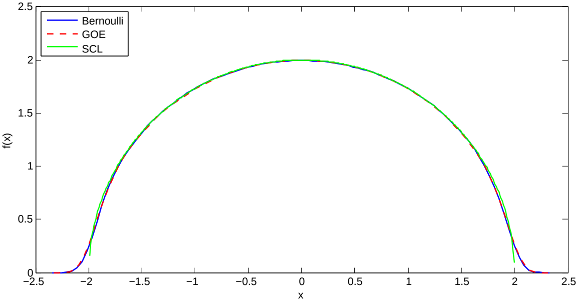
Theorem (Wigner Semi-circular law, 1950s)

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



Proof. Moment method.

One needs to show the moments of the eigenvalues tends to the moments of the semi-circle density function, which are:

$$\beta_k = \frac{1}{k+1}\binom{2k}{k}, k = 0, 2, 4...; \;\; \beta_k = 0, k = 1, 3, 5...$$

One uses the trace identity (in the coarse-scale)

$$\lambda_1 + \cdots + \lambda_n = Trace(W_n).$$

$$\lambda_1^k + \cdots + \lambda_n^k = Trace(W_n^k).$$

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 the limiting joint distribution of $(\lambda_{i_1}, \dots, \lambda_{i_k})$, 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)

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

(explicit formula given by Jimbo et. al.)

Distribution of the gaps

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

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

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

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

k-point correlation functions

$$\rho_n^k(x_1, \dots, x_k) := \int \rho(x_1, \dots, x_n) dx_{k+1} \dots dx_n.$$

$\int_I \rho_n^1(x)dx$ 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 $\lambda_i \in I, \lambda_j \in J$.

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

Typically, one wants to know

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

for a fixed test function f.

Local version.

Let $l_n << n$ tends to infinity with n. Consider a small neighborhood around a point u in the spectrum (nu in the fine-scale model).

$$\tilde{S}_n(s;u) := \frac{1}{l_n}|\{1 \le i \le n : \lambda_{i+1} - \lambda_i \le \frac{s}{\rho(u)}, |\lambda_i - nu| \le \frac{l_n}{\rho(u)}\}|.$$

Local version.
Let $l_n << n$ tends to infinity with n. Consider a small neighborhood around a point u in the spectrum (nu in the fine-scale model).

$$\tilde{S}_n(s;u) := \frac{1}{l_n}|\{1 \leq i \leq n : \lambda_{i+1} - \lambda_i \leq \frac{s}{\rho(u)}, |\lambda_i - nu| \leq \frac{l_n}{\rho(u)}\}|.$$

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

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 ?

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 ?

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 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 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 mainly focus on the last problem.

GUE statistics: Gap distribution

It is known (due to Dyson-Mehta) that

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

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*

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

GUE statistics: k-correlation functions

For GUE, it was established by Gaudin and Mehta that

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

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

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

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

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

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

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

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.

GUE statistics: Distribution of the least singular value

Recall that $\sigma_n(M)$ (the least singular value of M_n) is the minimum absolute value of an eigenvalue: $\sigma_n = \min_i |\lambda_i|$.
Jimbo-Miwa-Tetsuji-Mori-Sato (1980) showed that for fixed any $t > 0$, and M_n drawn from GUE, one has

$$\mathbf{P}(\sigma_n(M_n) \leq \frac{t}{2\sqrt{n}}) \to \exp(\int_0^t \frac{f(x)}{x} dx)$$

as $n \to \infty$, where $f : \mathbf{R} \to \mathbf{R}$ is the solution of the differential equation

$$(tf'')^2 + 4(tf' - f)(tf' - f + (f')^2) = 0$$

with the asymptotics $f(t) = \frac{-t}{\pi} - \frac{t^2}{\pi^2} - \frac{t^3}{\pi^3} + O(t^4)$ as $t \to 0$.

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.

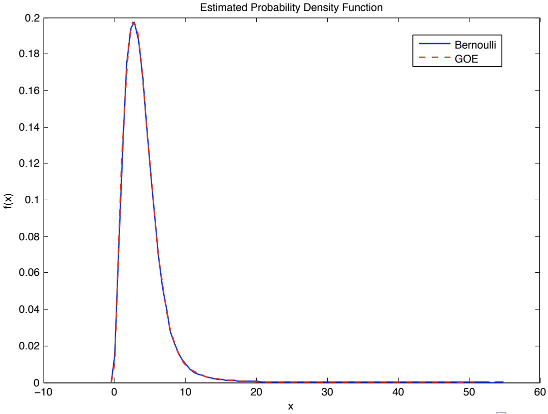
The Universality Phenomenon

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.

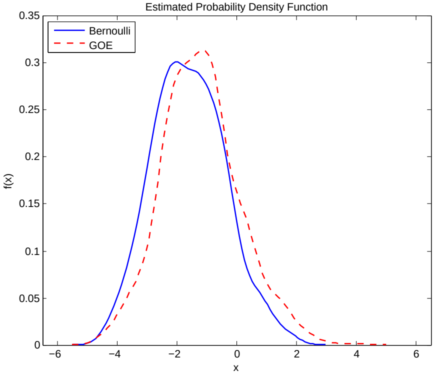
The word "Universality" in mathematical physics refers to the specific problem of the universality of k correlation functions.

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

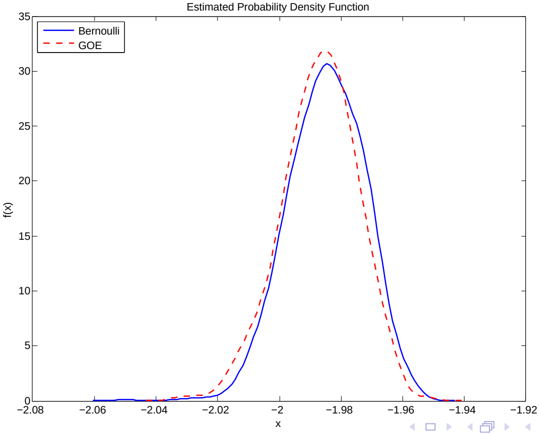
Numerical evidence: Gap distributions



Numerical evidence: Distribution at the edge



Numerical evidence: Distribution in the bulk



Universality Results: Universality at the Edge

Theorem (Soshnikov 1999)

Let k be a fixed integer and M_n a Wigner matrix. Assume furthermore that the distribution of the atom variables are symmetric and have exponential tail. Set $W_n := \frac{1}{\sqrt{n}} M_n$. Then the joint distribution of the k dimensional random vector

$$\left((\lambda_1(W_n) + 2) n^{2/3}, \ldots, (\lambda_k(W_n) + 2) n^{2/3} \right)$$

has a weak limit as $n \to \infty$, which coincides with that in the GUE case. In particular, the limit for $k = 1$ is the Tracy-Widom distribution. The result also holds for $\lambda_{n-k+1}, \ldots, \lambda_n$

Universality Results: Universality at the Edge

Theorem (Soshnikov 1999)

Let k be a fixed integer and M_n a Wigner matrix. Assume furthermore that the distribution of the atom variables are symmetric and have exponential tail. Set $W_n := \frac{1}{\sqrt{n}} M_n$. Then the joint distribution of the k dimensional random vector

$$\left((\lambda_1(W_n) + 2) n^{2/3}, \ldots, (\lambda_k(W_n) + 2) n^{2/3} \right)$$

has a weak limit as $n \to \infty$, which coincides with that in the GUE case. In particular, the limit for $k = 1$ is the Tracy-Widom distribution. The result also holds for $\lambda_{n-k+1}, \ldots, \lambda_n$

Method. Wigner trace method. Computing the number of special walks in a complete graph (Sinai-Soshnikov). Recent improvement: Ruzmiakina, Khorunzhiy replaced the exponential tail by high moment.

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.

Recently (2009-2010), Erdős- Peche-Schlein-Ramirez-Yau considered *average* k correlation functions, and prove universality under more general 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 lob-Sobolev property.
The newest improvement (Erdős-Jin-Yau) removed the smooth assumption completely.

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 and results play an important 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).

New result: 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 $1 \leq i_1 < \cdots < i_k \leq n$.

New result: 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.

New result: 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 $1 \leq i_1 < \dots < i_k \leq n$.

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).

New result: 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).
- 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

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 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.

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). In fact one can even take k growing slowly in n.

The mean of a typical λ_i do shift (by a significant amount) if we change the fourth moment.
For any constant $C > 0$ there is a constant C'
¿0*suchthatif*—$\mathbf{E}\zeta^4 - E\zeta'^4| \geq C$ then for a *typical i*

$$|\lambda_i(A_n) - \lambda_i(A'_n)| \geq C'.$$

Intuition:

$$\lambda_1^2 + \cdots + \lambda_n^2 = \sum_{ij} \zeta_{ij}^2.$$

$$E(\lambda_1 + \cdots + \lambda_n^2)^2 = \mathbf{E}(\sum_{ij} \zeta_{ij}^2)^2.$$

Universality of Tracy-Widom law

One can use a slightly more technical version of the Four Moment theorem to treat eigenvalues at the edge, such as λ_1. Notice that the fluctuation at the edge is of order $n^{-2/3}$, not n^{-1} as in the bulk. It saves one moment. One can save another moment by using Johansson model instead of GUE. There is no need for the entry distributions be symmetric.

Universality of Tracy-Widom law

One can use a slightly more technical version of the Four Moment theorem to treat eigenvalues at the edge, such as λ_1. Notice that the fluctuation at the edge is of order $n^{-2/3}$, not n^{-1} as in the bulk. It saves one moment. One can save another moment by using Johansson model instead of GUE. There is no need for the entry distributions be symmetric.

Corollary (Universality of Tracy-Widom)

Tracy-Widom distribution at the edge of the spectrum is universal with respect to all random Hermitian matrices whose entries have mean zero, variance one and bounded Cth moment, for a sufficiently large constant C.

Corollary (Universality of the least singular value)

Jimbo et. al. for the distribution of the least singular value is universal with respect to all random Hermitian matrices whose entries have mean zero, variance one and bounded Cth moment, for a sufficiently large constant C.

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.

Consequences: Universality of the k-correlation function

One can use the same argument to prove the Unviversality of the k-point correlation function.

Consequences: Universality of the k-correlation function

One can use the same argument to prove the Unviversality of the k-point correlation function.

One can even avoid the assumption that the support has at least 3 points by using the recent result of Erdős et. al. which shows that Johansson theorem extends to the case with c_1 being a negative power of n, combining with a variant of the Four Moment Theorem (Erdős-Ramirez-Schlein-Tao-Vu-Yau 09).

Local version: Universality of the Dyson kernel

Corollary (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

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

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

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. (Tao-V., V.-Wang).

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: Krishnapur, Chatterjee, O' Donnell et. al.

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

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

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) .