Discrete random matries and universality

Terence Tao

University of California, Los Angeles

Mahler Lecture Series

Universality

- The world is a complex place, and so one would think that complex mathematical models, with many independent variables, are needed to describe it.
- However, it is a curious phenomenon that the cumulative effect of many independent variables in a system becomes more predictable as the number of variables increases, rather than less.

Universality cont.

- In fact, in many cases, the exact behaviour of each individual variable becomes essentially irrelevant (except perhaps for one or two key parameters); one gets the same observed behaviour for the system as a whole regardless of what the individual components are doing.
- This phenomenon is sometimes referred to universality, or the invarance principle.

Examples from the physical world

- **Statistical mechanics:** The behaviour of a system of N particles with respect to changes in energy, volume, etc., would seem almost impossible to compute precisely when N is huge, requiring precise knowledge of the system and its interactions. But, in fact, in the limit $N \to \infty$, the behaviour can be controlled by just a handful of key parameters, such as temperature and entropy.
- **Benford's law:** 30% (or more precisely, $\log_{10} 2 \approx 30.1\%$) of all statistics start with the digit 1! For instance, 30% of all cities have populations beginning with 1, 30% of all word frequencies in a language begin with 1, 30% of all stock prices begin with 1, etc. This is despite the fact that different statistics are governed by completely different laws of nature.

An example from probability theory

- **The law of large numbers** If $X_1, \ldots, X_n$ are n random variables that are independent and identically distributed (iid), then the empirical average $\frac{X_1+\ldots+X_n}{n}$ converges to the mean μ of any one of the variables X_i in the limit $n \to \infty$. (Let me gloss over the technical issue of what "converges" means here.)
- **The central limit theorem** Continuing the above example, the distribution of the normalised deviation $\sqrt{n} \times (\frac{X_1+\ldots+X_n}{n} - \mu)$ converges as $n \to \infty$ to the normal distribution $N(0, \sigma^2)$, where σ^2 is the variance of any of the X_i.

The invariance principle

In the above example, we saw that the only two features of the random variables that are relevant in the limit $n \to \infty$ are the mean and variance; it does not matter, for instance, whether these variables are continuous or discrete.
This is a model case of a more general

- **Invariance principle** In many cases, the behaviour of a combination $F(X_1, \ldots, X_n)$ of iid random variables $X_1, \ldots, X_n$ for n large does not depend very much on the actual distribution of the X_i, but only on some key parameters of that distribution, such as mean and variance.

- In the law of large numbers and central limit theorem, F was a linear combination of the X_i. But the principle also extends to some important nonlinear combinations as well, in particular to spectral statistics of random matrices.
- This is the focus of my talk today.

Random matrix models

We will consider a number of random matrix models, which can be either discrete or continuous:

- **Iid random matrices** These are $n \times n$ matrices $A = (x_{ij})_{1 \leq i,j \leq n}$, where the x_{ij} are iid random variables, normalised to have mean zero and variance 1. Key examples include the Bernoulli matrix ensemble (random sign matrices), in which $x_{ij} = \pm 1$ with an equal probability of each, and the real and complex gaussian matrix ensembles, in which the x_{ij} have either the real or complex gaussian distribution with the indicated mean and variance.

More random matrix models

- **Wigner symmetric matrices** These are similar to iid random matrices, but the coefficients x_{ij} are now assumed to be real and symmetric ($x_{ij} = x_{ji}$). The x_{ij} are now iid just for $1 \leq i \leq j \leq n$. Examples include the symmetric Bernoulli matrix ensemble ($x_{ij} = \pm 1$) and the gaussian orthonormal ensemble (GOE) (x_{ij} are real gaussians).

Even more random matrix models

- **Wigner Hermitian matrices** These are similar to Wigner symmetric matrices, but now the coefficients x_{ij} are complex and Hermitian ($x_{ij} = \overline{x_{ji}}$). The x_{ij} are iid on the upper triangular region $1 \leq i < j \leq n$ and on the diagonal $1 \leq i = j \leq n$, but can have different distributions in the two regions. For instance, in the gaussian unitary ensemble (GUE), the x_{ij} are complex gaussian for $1 \leq i < j \leq n$ but real gaussian for $1 \leq i = j \leq n$. (In all cases, the coefficients have mean zero and variance 1.)

- Many other random matrix models are of interest, but to focus the talk we shall only discuss these particular ones.
- Discrete random matrices arise naturally in numerical linear algebra (as a model for rounding errors), while continuous random models arise naturally in various physical settings (e.g. spectra of atoms).
- The gaussian models are particularly tractable due to their group invariance properties. For instance, the GUE ensemble is invariant under conjugations by the unitary group $U(n)$, and GOE is similarly invariant under the orthogonal group $O(n)$.

Eigenvalues

- Given an $n \times n$ random matrix A_n, we let $\lambda_1(A_n), \ldots, \lambda_n(A_n)$ be the n (generalised) eigenvalues of A_n. In symmetric or Hermitian models, these eigenvalues will be real, and we can order them: $\lambda_1(A_n) \leq \ldots \leq \lambda_n(A_n)$. In the iid random matrix model, the eigenvalues will be complex and unordered; but one can then define the singular values $0 \leq \sigma_n(A_n) \leq \ldots \leq \sigma_1(A_n)$ (the eigenvalues of $(A_n A_n^*)^{1/2}$).

Eigenvalues cont.

- Eigenvalues and singular values are related to many other important matrix quantities, such as the determinant

$$|\det(A_n)| = \prod_{i=1}^{n} |\lambda_i(A_n)| = \prod_{i=1}^{n} \sigma_i(A_n)$$

or the trace

$$\mathrm{tr}(A_n) = \sum_{i=1}^{n} \lambda_i(A_n).$$

For this and other reasons, it is of interest to understand the distribution of these numbers.

Universality

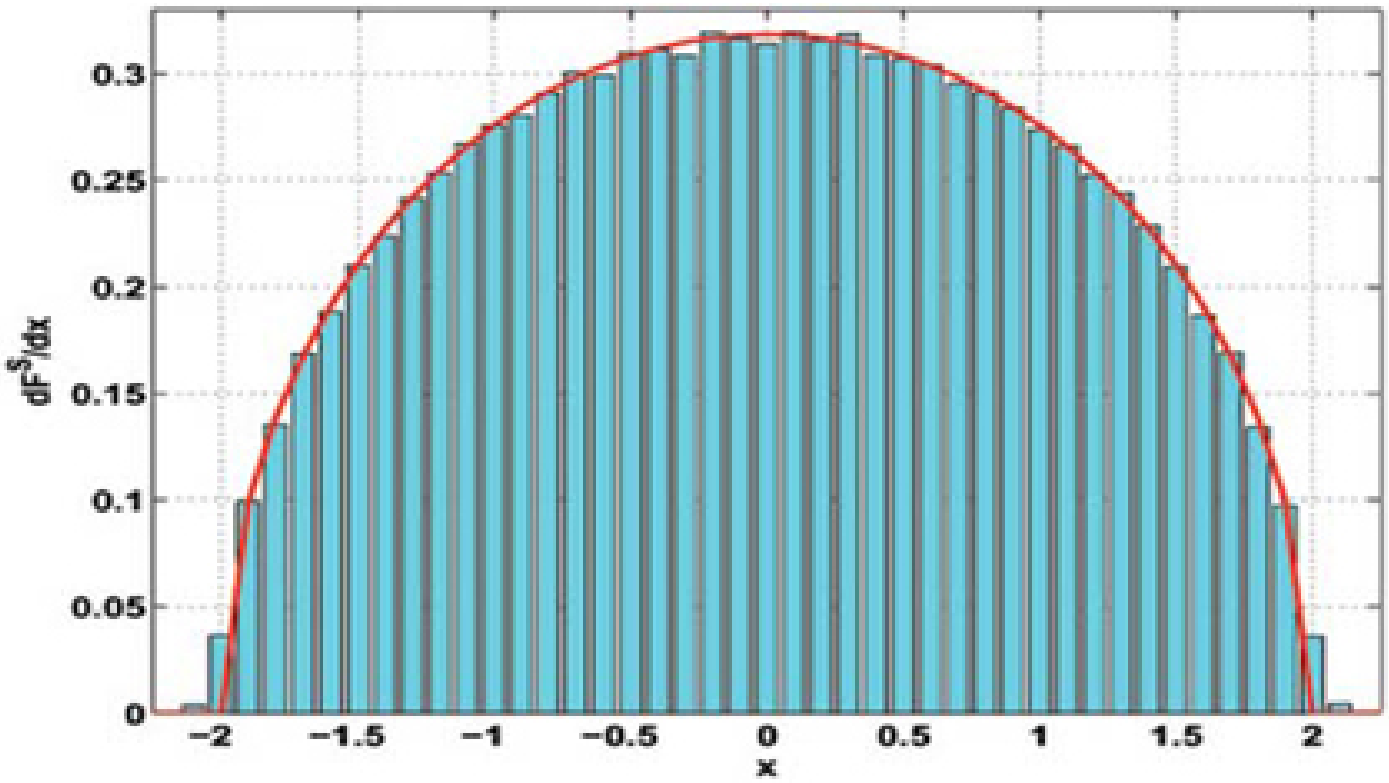
- In accordance with the invariance principle, many facts about the distribution of eigenvalues and singular values of random matrices seem to be universal in the limit $n \to \infty$ - they do not depend on the precise matrix model used. Thus, for instance, continuous and discrete random matrices often have the same statistics in the high-dimensional limit.
- This phenomenon has been observed numerically for many decades. More recently, rigorous explanations of this phenomenon have been found (and there is still work to be done in some cases).

- Many distributions of empirically observed eigenvalues (e.g. atomic spectra) obey the same statistics as random matrix models. The universality phenomenon provides a partial explanation of this fact.

Wigner's semicircular law

The most well-known example of universality is for the bulk distribution of eigenvalues of Wigner matrices:

- **Wigner's semicircular law** For a Wigner symmetric or Hermitian random matrix A_n, the normalised eigenvalues $\frac{1}{\sqrt{n}}\lambda_1(A_n), \ldots, \frac{1}{\sqrt{n}}\lambda_n(A_n)$ are asymptotically distributed according to the semicircular distribution $\frac{1}{2\pi}(4-x^2)_+^{1/2}\,dx$.
- Established by Wigner for GOE in 1955, and then repeatedly generalised (the version above was established by Pastur in 1977). Many, many further refinements and proofs.

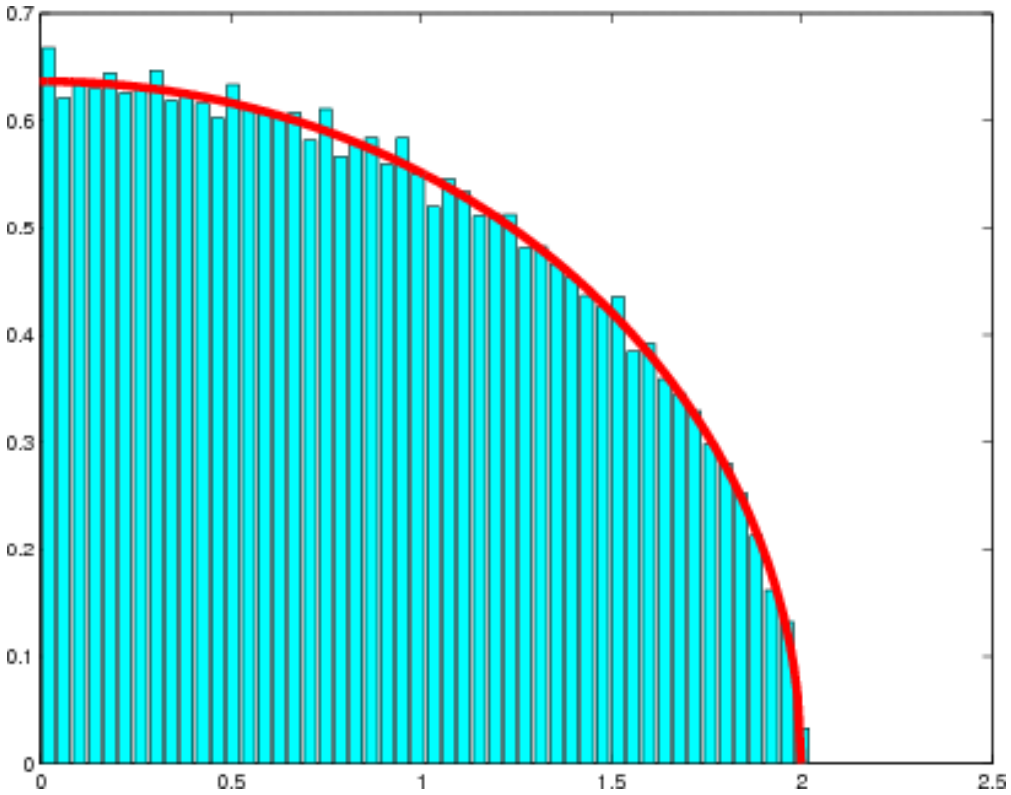


Normalised eigenvalue distribution of a random 100×100 GUE matrix. (Image by Alan Edelman.)

Marchenko-Pastur quarter-circle law

There is an analogous law for bulk distribution of singular values of iid matrices:

- **Quarter-circle law** For an iid random matrix A_n, the normalised singular values $\frac{1}{\sqrt{n}}\sigma_1(A_n), \ldots, \frac{1}{\sqrt{n}}\sigma_n(A_n)$ are asymptotically distributed according to the quarter-circle distribution $\frac{1}{\pi}(4-x^2)^{1/2}1_{[0,2]}(x)\,dx$.
- Established by Marchenko and Pastur in 1967. Again, many further refinements and proofs. (This law has an equivalent formulation in terms of eigenvalues of covariance matrices, known as (a special case of) the Marchenko-Pastur law.)

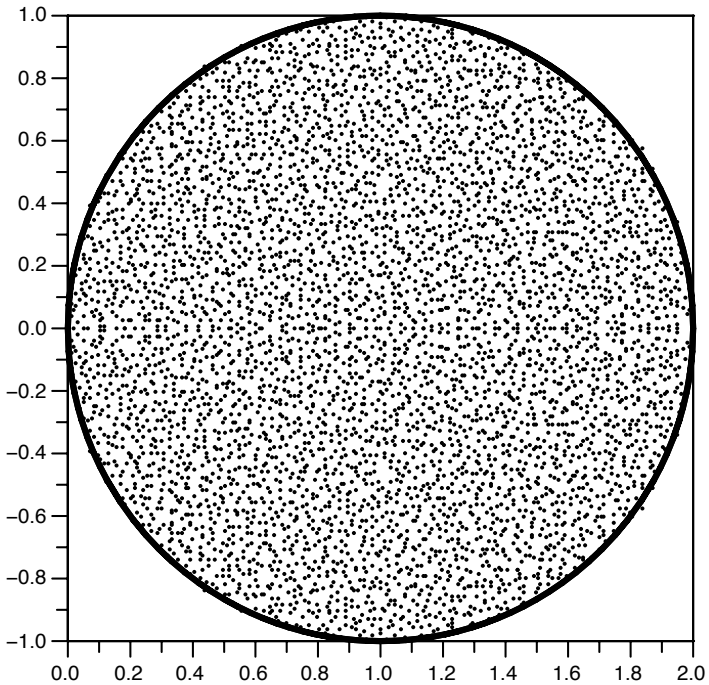


Normalised singular distribution of a 100×100 iid gaussian matrix. (Image by Antonio Tulino)

Circular law

As for the bulk distribution of eigenvalues of iid matrices, we have

- **Circular law** For an iid random matrix A_n, the normalised eigenvalues $\frac{1}{\sqrt{n}}\lambda_1(A_n), \ldots, \frac{1}{\sqrt{n}}\lambda_n(A_n)$ are asymptotically distributed according to the circular law $\frac{1}{\pi}1_{x^2+y^2\leq 1}\, dxdy$.
- Established for gaussian matrices by Mehta in 1967. Generalised by many authors (Girko, Bai, Bai-Silverstein, Götze-Tikhomirov, Pan-Zhou, Tao-Vu); the result above is due to Tao-Vu-Krishnapur, 2008.

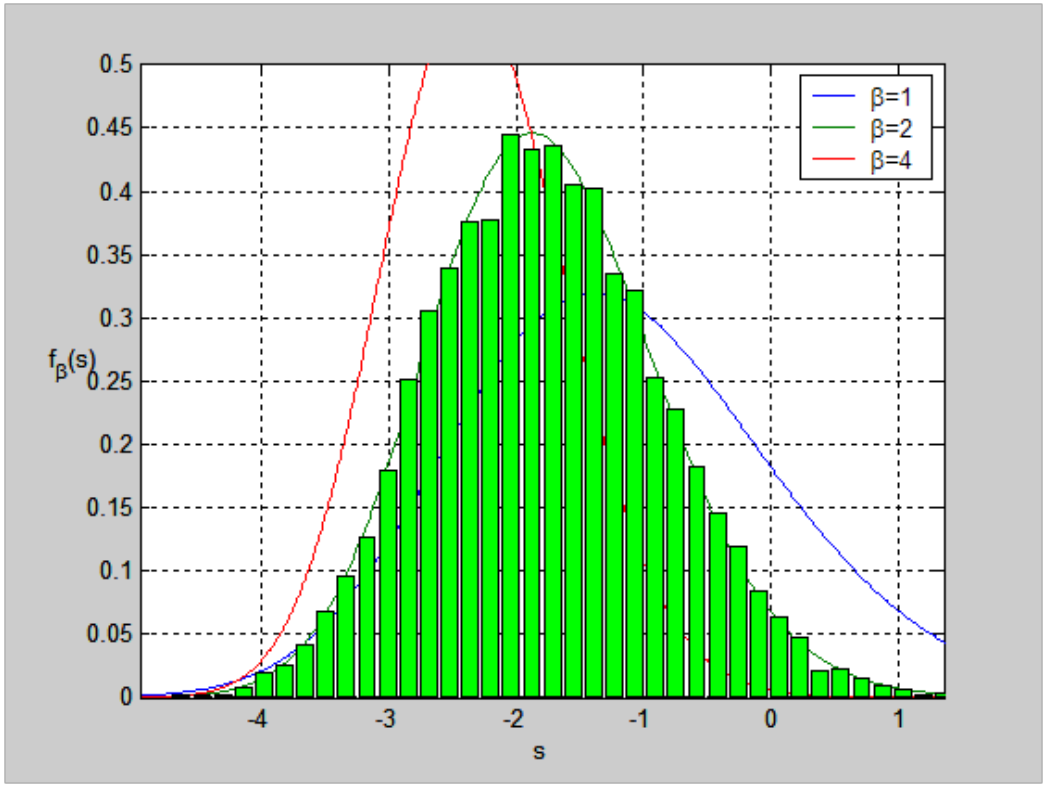


Normalised eigenvalue distribution of a 5000×5000 iid Bernoulli matrix. (Image by Phillip Wood)

Tracy-Widom law

Instead of the bulk distribution, one can ask for finer information about individual eigenvalues, which is harder. A typical result is

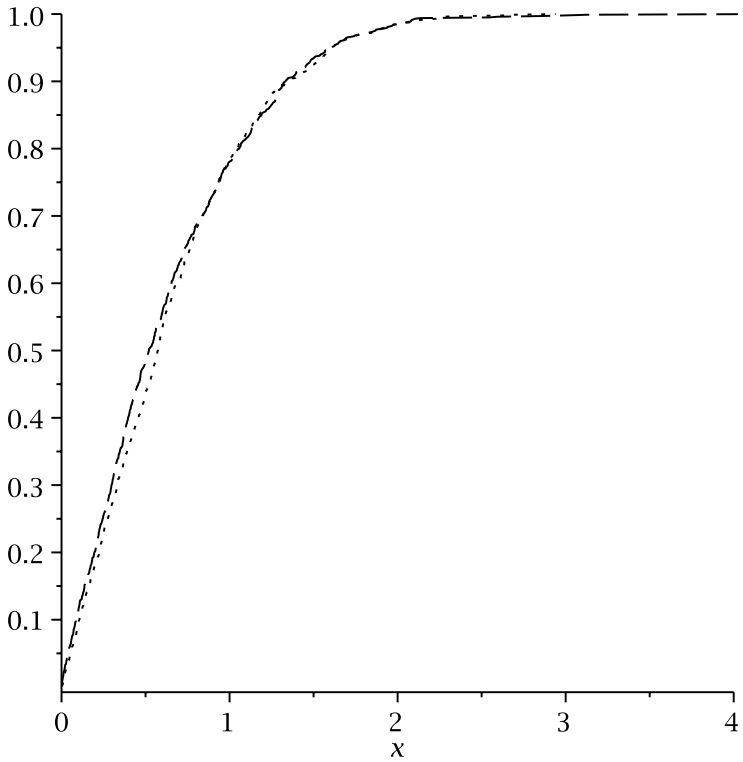
- **Tracy-Widom law** For a Wigner Hermitian matrix A_n, the normalised largest eigenvalue $(\lambda_n(A_n) - 2\sqrt{n})/n^{1/6}$ is asymptotically distributed according to the Tracy-Widom law $F_2(x)\,dx = \det(1-K)\,dx$, where K is the integral operator with Airy kernel $\frac{\mathrm{Ai}(x)\,\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\,\mathrm{Ai}(y)}{x-y}$.
- Established for GUE by Tracy and Widom in 1994. Not fully resolved in general, but known for symmetric decaying distributions (Sinai-Soshnikov 1998, Soshnikov 1999, Ruzmaikina 2006, Khorunzhiy 2009) and decaying distributions with vanishing third moment (Tao-Vu 2009).



Normalised largest eigenvalue of GUE matrices. (Image by Alan Edelman et al..)

Least singular value

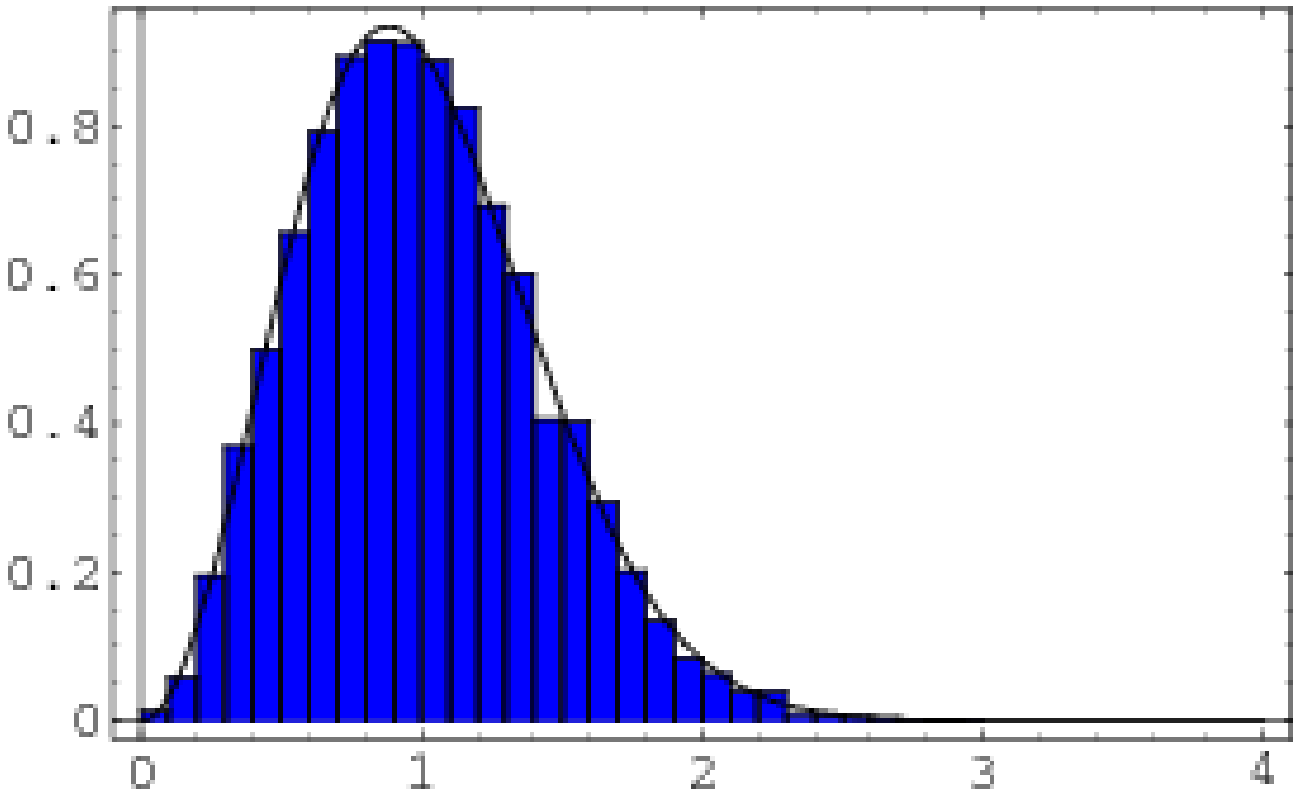
- **Least singular value law** For a real iid random matrix A_n, the normalised least singular value $\sqrt{n}\sigma_n(A_n)$ is asymptotically distributed according to the law $(1+x)e^{-x-x^2}\,dx$.
- Established for GOE by Edelman in 1991, and in general by Tao-Vu in 2009.



Cumulative distribution of normalised singular values of one thousand 100×100 Bernoulli and gaussian iid matrices. (Image by Phillip Wood.)

Eigenvalue gaps

- **GUE spacing** For a Wigner Hermitian matrix A_n, the normalised gap $\sqrt{n}(\lambda_{i+1}(A_n) - \lambda_i(A_n))$ for a randomly chosen $1 \leq i \leq n$ is asymptotically distributed according to the **Gaudin distribution** $\frac{d^2}{dx^2} \det(1-K)_{L^2([0,x])}$, where K has the sine kernel $\frac{\sin \pi(x-y)}{\pi(x-y)}$.
- Established for GOE by Edelman in 1991. Almost proven in full generality; known for all rapidly decreasing distributions by Erdős-Ramirez-Schlein-Tao-Vu-Yau (2009).



Cumulative distribution of eigenvalue spacings of five hundred 100×100 GUE matrices. (Image by Peter Kostelec.)

Many, many techniques go into the proofs of these facts. We will discuss just two key tools:

- **Linear algebra identities** that relate eigenvalues and singular values to more computable quantities, such as moments, resolvents, determinants, and distances.
- The **Lindeberg exchange strategy**, based on exchanging an arbitrary distribution with a gaussian one of the same mean and variance.

But many other tools are used too...

- Symmetry reductions and explicit formulae, Lie groups
- Asymptotics of orthogonal polynomials, Riemann-Hilbert problems
- Free probability
- Dyck paths, combinatorics
- Stieltjes transform, complex analysis
- Concentration of measure, high-dimensional geometry
- Inverse Littlewood-Offord theorems, additive combinatorics
- Estimation of eigenvalues by random sampling
- Dyson Brownian motion
- Ornstein-Uhlenbeck process
- Cauchy interlacing law
- . . .

Linear algebra identities

A surprising amount of mileage can be gained from basic linear algebra identities such as

$$\mathrm{tr}(A^k) = \sum_{i=1}^n \lambda_i(A)^k$$

$$\mathrm{tr}((A^* A)^k) = \sum_{i=1}^n \sigma_i(A)^{2k}$$

$$\log|\det(A - zI)| = \sum_{i=1}^n \log|\lambda_i(A) - z| = \sum_{i=1}^n \log|\sigma_i(A - zI)|$$

$$\log|\det(A)| = \sum_{i=1}^n \log|\mathrm{dist}(X_i, \mathrm{span}(X_1, \ldots, X_{i-1}))|$$

A simple example

Let $A = (x_{ij})_{1 \leq i,j \leq n}$ be an $n \times n$ symmetric Bernoulli matrix. Then

$$\mathrm{tr}(A^2) = \sum_{i=1}^{n} \sum_{j=1}^{n} |x_{ij}|^2 = n^2$$

and thus

$$\sum_{i=1}^{n} \lambda_i(A)^2 = n^2.$$

For non-Bernoulli matrices, one has to use the law of large numbers (thus linear universality is used to deduce nonlinear univerality!)

The Lindeberg strategy

The Lindeberg strategy splits the task of proving a universal law into two distinct parts:

- **The gaussian case** Show that the law holds when all the underlying random variables are gaussian. This is usually achieved by algebraic means, using all the special properties of gaussians (e.g. the group symmetries of the gaussian ensembles).
- **Invariance** Show that the limiting distribution is unchanged when non-gaussian random variables are replaced by gaussian random variables. This is usually achieved by analytic means, showing that the error terms caused by this replacement are asymptotically negligible compared to the main terms.

Example: Lindeberg's proof of the CLT

- To give a simple example of the method, suppose one wants to prove the central limit theorem for the normalised average $S = \frac{X_1+\ldots+X_n}{\sqrt{n}}$ of iid variables $X_1, \ldots, X_n$ of mean zero and variance 1. We replace the X_i by gaussians Y_i of the same mean and variance, and consider the normalised average $T = \frac{Y_1+\ldots+Y_n}{\sqrt{n}}$.
- Under mild decay assumptions on X, it will suffice to show that the moments $\mathbb{E}S^k$, $\mathbb{E}T^k$ asymptotically match as $n \to \infty$ for each $k = 0, 1, 2, \ldots$. (The decay assumptions can be removed by a truncation argument.)

- When one expands out the moments $\mathbb{E}S^k$, one gets a linear combination of terms of the form $\mathbb{E}X_{i_1}^{a_1} \dots X_{i_j}^{a_j}$. The corresponding moment $\mathbb{E}T^k$ has a similar expansion with the X_i replaced by Y_i.
- When all the exponents $a_1, \dots, a_j$ are at most 2, then the term for X and the term for Y are identical (because the X_i, Y_i are iid and have matching moments to second order).
- The terms when one or more of the a_i exceed 2 can be shown to be asymptotically negligible. Putting all this together, one establishes the central limit theorem.