The Distribution Functions of Random Matrix Theory

Craig A. Tracy
UC Davis

UC Irvine

November 30, 2006

The Central Limit Theorem

Carl F. Gauss was the first to use the normal law (or Gaussian)

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-t^2/2)\, dt$$

as a bona fide distribution function. Earlier, A. De Moivre in his "Doctrine of Chances" (1733) had shown the normal law is a good *approximation* to the binomial distribution. However, it was Pierre S. Laplace (1820) who gave the first CLT.

In modern notation: $X_1, \ldots, X_n$ are discrete-valued independent and identically distributed random variables. Set $\mu = E(X_i)$ and $\sigma^2 = \mathrm{Var}(X_i)$, then

$$\Pr\left(\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right) \to \Phi(x), \;\; n \to \infty$$

where $S_n = X_1 + \cdots + X_n$.

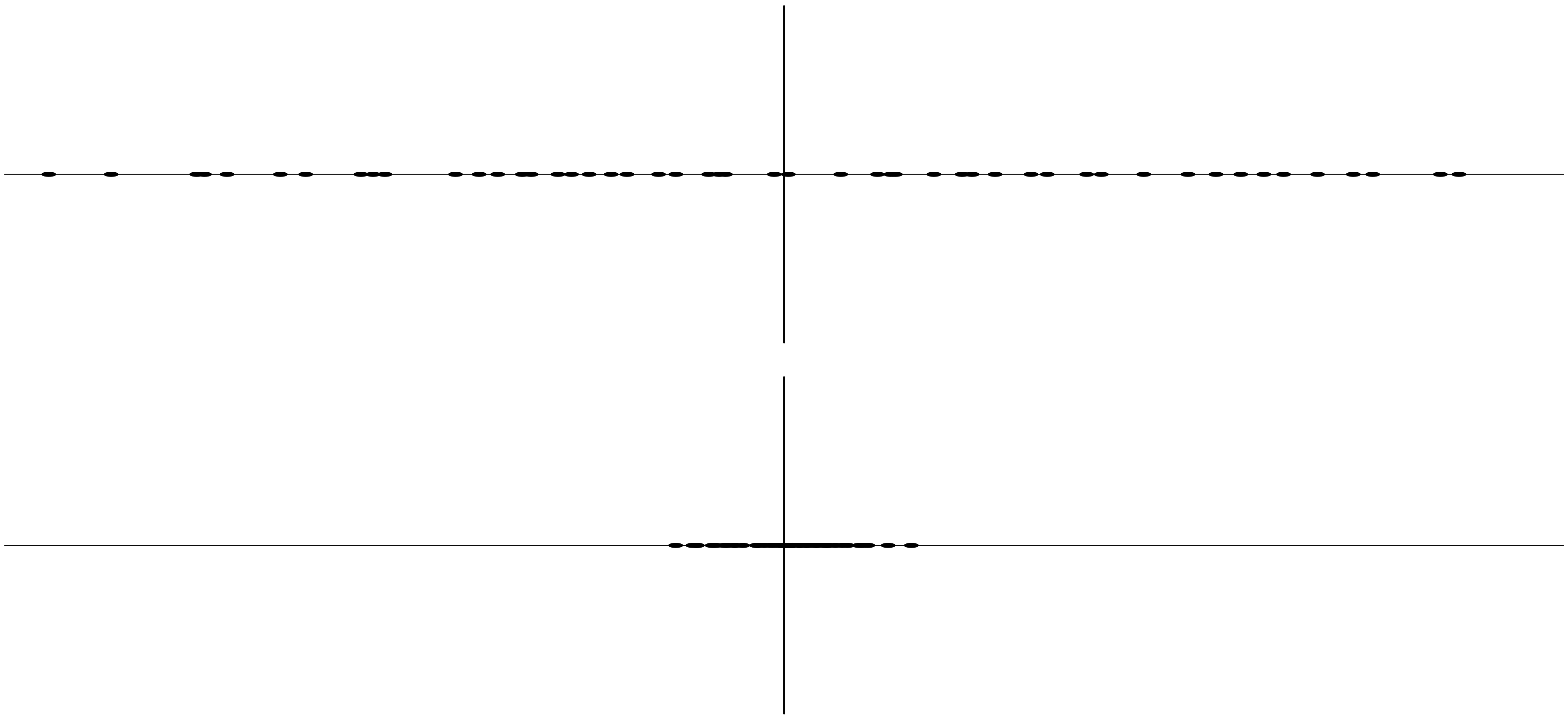
Important features of Laplace's proof

- Introduced the characteristic function $E(\mathrm{e}^{itS_n})$ and used Laplace's method for approximating integrals
- He made the important observation that the limit law depended only upon μ and σ of the underlying distribution. (Universality).

Laplace's ideas were further developed by Poisson, Dirichlet, Cauchy and others. The St. Petersburg School of probability (Chebyshev, Markov, Lyapunov, ...) relaxed the conditions of identically distributed as well as the important assumption of independence. This includes the important "method of moments"

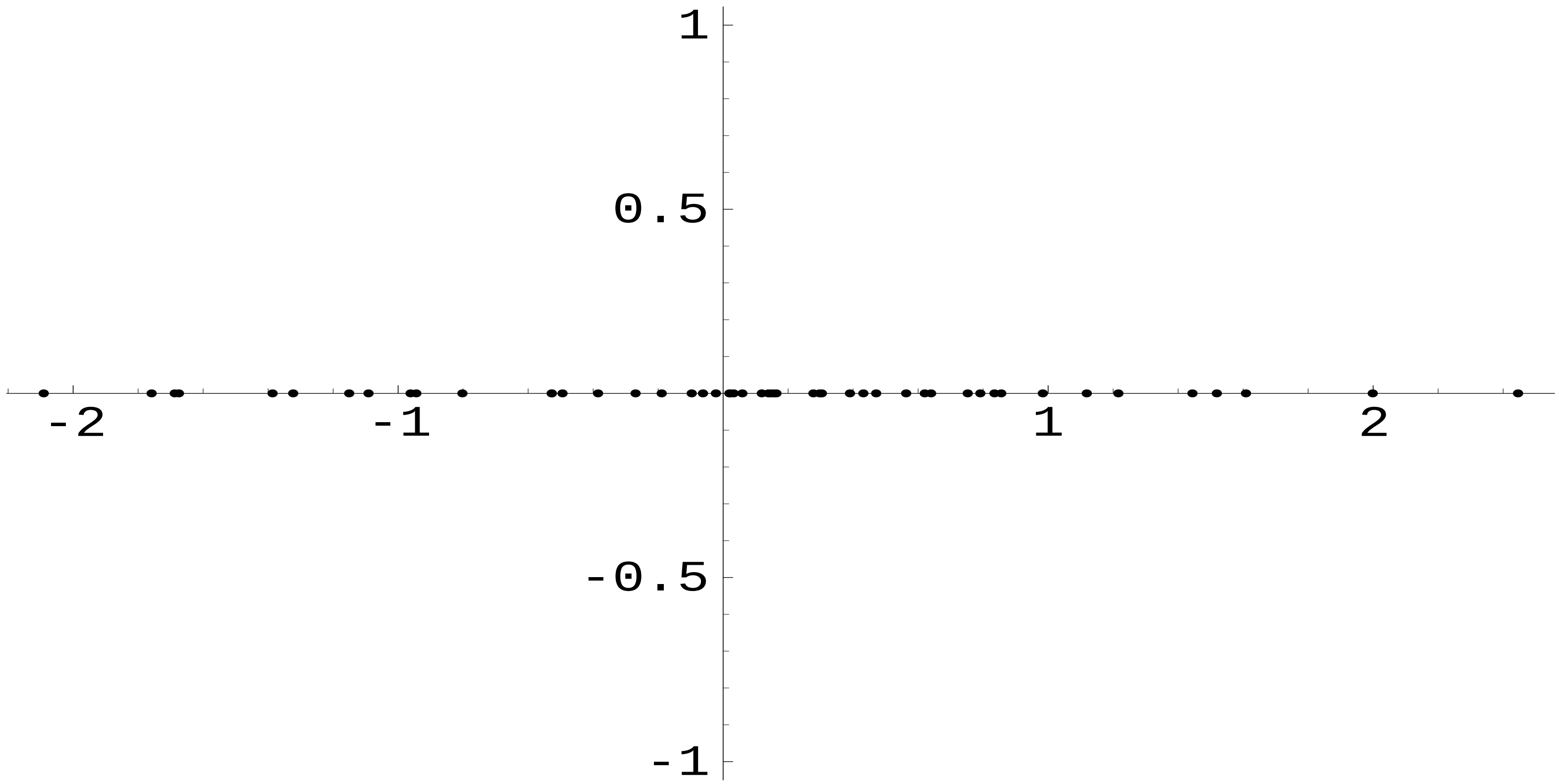
The final form is attributed to Lindeberg, Feller and Levy where necessary and sufficient conditions are given for convergence to the normal law.

Eigenvalues vs. Independent Normals



Top: 50×50 real symmetric matrix with indep. $N(0,1)$ entries.

Bottom: 50 $N(0,1)$ independent random variables.



Same 50×50 real symmetric matrix with indep. $N(0,1)$ entries.
Note spacing is unlike eigenvalue spacing.

Random Matrix Models

Probability Space: $(\Omega, \Pr, \mathcal{F})$:

- Gaussian Orthogonal Ensemble (GOE, $\beta = 1$):
  - $\Omega = N \times N$ real symmetric matrices
  - $\Pr =$ "unique" measure that is invariant under orthogonal transformations and matrix elements are iid random variables. Explicitly,

$$\Pr(A \in \mathcal{B}) = \int_{\mathcal{B}} \mathrm{e}^{-\mathrm{tr}(A^2)}\, dA$$

- Gaussian Unitary Ensemble (GUE, $\beta = 2$)
  - $\Omega = N \times N$ (complex) hermitian matrices
  - $\Pr =$ "unique" measure that is invariant under unitary transformations and the independent real and imaginary matrix elements are iid random variables
- Gaussian Symplectic Ensemble (GSE, $\beta = 4$)

Write matrix $n \times n$

$$A = UDU^*$$

U orthogonal for GOE, unitary for GUE, etc.

D is diagonal matrix with eigenvalues λ_j as entries

The induced density on eigenvalues is

$$P(\lambda_1, \ldots, \lambda_n) = c_{n,\beta} \prod_{1 \le j < k \le n} |\lambda_j - \lambda_k|^\beta \, \mathrm{e}^{-\sum_j \lambda_j^2}$$

$$\beta = \begin{cases} 1 & \text{GOE} \\ 2 & \text{GUE} \\ 4 & \text{GSE} \end{cases}$$

Vandermonde to the β power $\Longrightarrow$ **level repulsion**

Limit Laws: $N \to \infty$

Eigenvalues, which are random variables, are real and with probability one they are distinct.

If $\lambda_{\mathbf{max}}(\mathbf{A})$ denotes the **largest eigenvalue** of the random matrix A, then for each of the **three Gaussian ensembles** we introduce the corresponding distribution function

$$F_{N,\beta}(t) := \mathrm{Pr}_\beta\left(\lambda_{\mathrm{max}} < t\right), \beta = 1, 2, 4.$$

The basic limit laws (Tracy-Widom) state that[a]

$$F_\beta(s) := \lim_{N\to\infty} F_{N,\beta}\left(2\sigma\sqrt{N} + \frac{\sigma s}{N^{1/6}}\right), \, \beta = 1, 2, 4,$$

exist and are given explicitly by

[a] Here σ is the standard deviation of the Gaussian distribution on the off-diagonal matrix elements.

$$
\begin{aligned}
F_2(s) &= \det\left(I - K_{\mathrm{Airy}}\right) \\
&= \exp\left(-\int_s^\infty (x-s)q^2(x)\,dx\right)
\end{aligned}
$$

where

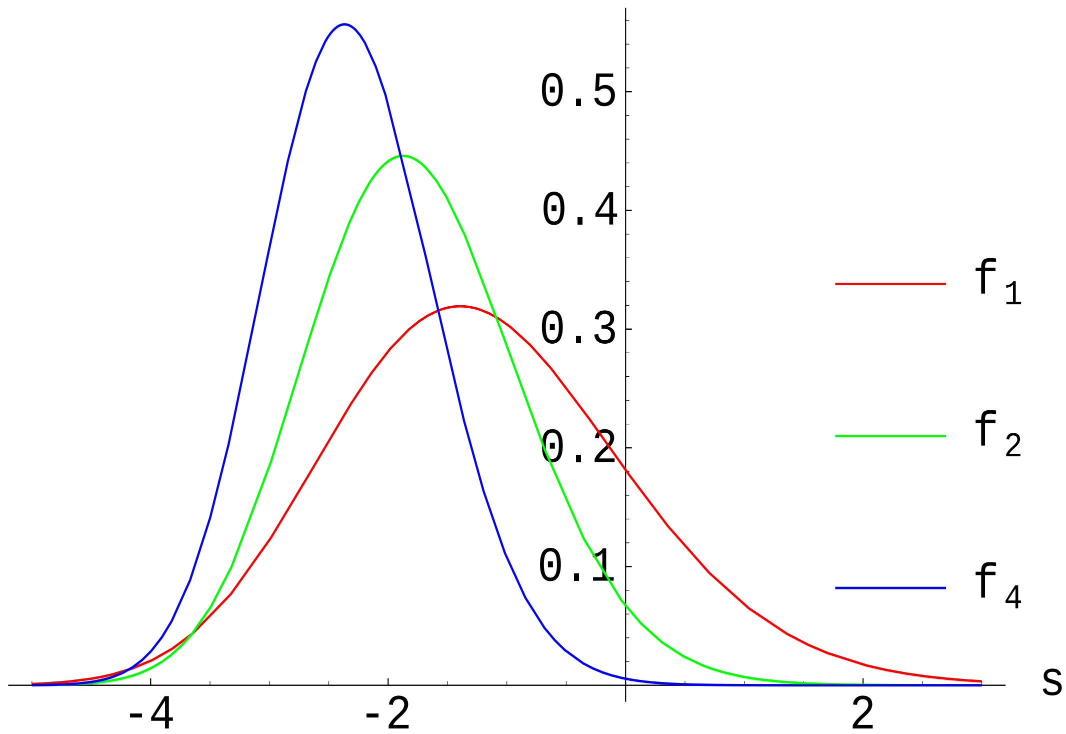
$$
K_{\mathrm{Airy}} \doteq \frac{\mathrm{Ai}(x)\mathrm{Ai}'(y) - \mathrm{Ai}'(x)\mathrm{Ai}(y)}{x-y}
$$

acting on $L^2(s,\infty)$ (**Airy kernel**)

and q is the unique solution to the **Painlevé II equation**

$$
q'' = sq + 2q^3, \quad q(s) \sim \mathrm{Ai}(s) \text{ as } s \to \infty.
$$

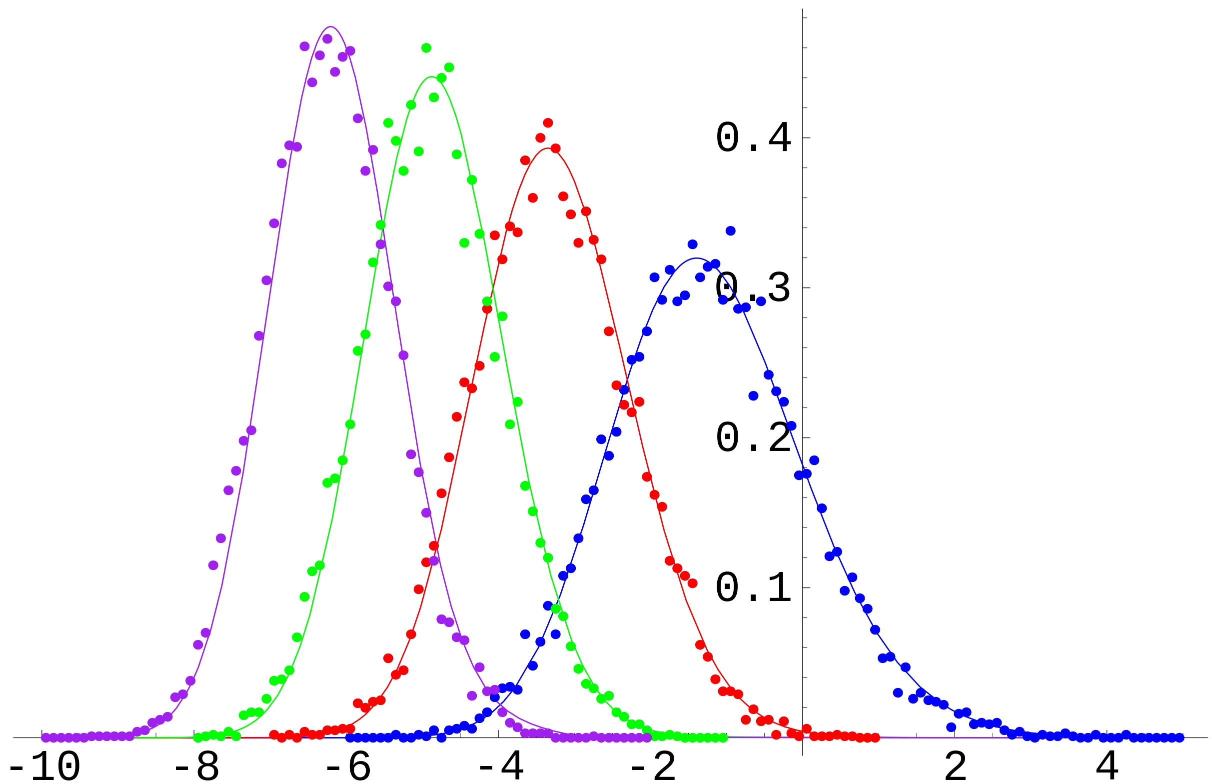
(Called the Hastings-McLeod solution.)



$$F_1(s) = \exp\left(-\frac{1}{2}\int_s^\infty q(x)\,dx\right)\,(F_2(s))^{1/2}\,,$$

$$F_4(s/\sqrt{2}) = \cosh\left(\frac{1}{2}\int_s^\infty q(x)\,dx\right)\,(F_2(s))^{1/2}\;.$$

10^4 realizations of $10^3 \times 10^3$ GOE matrices



Density of largest, next-largest, next-next largest, etc. eigenvalues (Dieng). Eigenvalues have been normalized. Distribution $F(s,k)$ given in terms of Painlevé functions and derivatives.

RMT Universality Theorems

Do limit laws depend upon the underlying Gaussian assumption on the probability measure?

To investigate this for unitarily invariant measures ($\beta = 2$):

$$\exp\left(-\mathrm{tr}(A^2)\right) \to \exp\left(-\mathrm{tr}(V(A))\right).$$

Bleher & Its chose

$$V(A) = gA^4 - A^2, g > 0,$$

and subsequently a large class of potentials V was analyzed by Deift/Kriecherbauer/McLaughlin/Venakides/Zhou.

Requires proving new **Plancherel-Rotac**h type formulas for **nonclassical** orthogonal polynomials. The proofs use **Riemann-Hilbert methods**. Generic behavior is GUE. However, by tuning V new universality classes will emerge.

Universality theorems for **orthogonal & symplectic** invariant measures:

- Stojanovic analyzed the quartic potential.
- Deift & Gioev considered a class of polynomial potentials whose equilibrium measure is supported on a single interval. Their starting point is Widom's representation of the correlation kernels for the $\beta = 1, 4$ cases in terms of the unitary ($\beta = 2$) correlation kernel plus a correction.

All these results can be summarized by

Generic edge behavior is described by law F_β

Noninvariant RMT Measures

Soshnikov proved that for real symmetric Wigner matrices[a] (complex hermitian Wigner matrices) the limiting distribution of the largest eigenvalue is F_1 (respectively, F_2). The **significance** of this result is that nongaussian Wigner measures lie **outside** the **"integrable class"** (e.g. there are no Fredholm determinant representations for the distribution functions) yet the limit laws are the same as in the integrable cases.

[a] A symmetric Wigner matrix is a random matrix whose entries on and above the main diagonal are independent and identically distributed random variables with distribution function F. Soshnikov assumes F is even and all moments are finite.

Applications to Wishart Distribution

Let X denote an $n \times p$ data matrix whose rows are independent $N_p(0, \Sigma)$ random variables. The matrix

$$\frac{1}{n} X^t X,$$

called the SAMPLE COVARIANCE MATRIX, is said to have WISHART DISTRIBUTION $W_p(n, \Sigma)$. The **null case**: $\Sigma = \mathrm{id}$.

Let $\hat{\ell}_1 > \cdots > \hat{\ell}_n$ denote the eigenvalues of $X^t X$—**sample covariance eigenvalues**. Johnstone for $k = 1$ and Soshnikov for $k > 1$ show, for the null case, as $n, p \to \infty$, $n/p \to \gamma$, $0 \le \gamma < \infty$

$$\frac{\hat{\ell}_k - \mu_{np}}{\sigma_{np}} \xrightarrow{\mathcal{D}} F_1(s, k)$$

with explicit expressions for the centering and norming constants.

Further Developments for Wishart Distribution

1. Johnstone, El Karoui, Choup: Second-order accuracy

$$\left| P\left(n\hat{\ell}_1 \leq \mu_{np} + \sigma_{np} s | H_0\right) - F_\beta(s) \right| \leq C p^{-2/3}$$

$$\begin{aligned} \mu_{np} &= \left(\sqrt{n - \frac{1}{2}} + \sqrt{p - \frac{1}{2}}\right)^2 \\ \sigma_{np} &= \left(\sqrt{n - \frac{1}{2}} + \sqrt{p - \frac{1}{2}}\right)\left(\frac{1}{\sqrt{n - \frac{1}{2}}} + \frac{1}{\sqrt{p - \frac{1}{2}}}\right)^{1/3} \end{aligned}$$

2. El Karoui in null case for the largest eigenvalue, proves the limit law for $0 \leq \gamma \leq \infty$. This requires additional estimates to allow $\gamma = \infty$. Soshnikov's theorem for $k > 1$ has not been extended to the $\gamma = \infty$ case.

3. Soshnikov removes Gaussian assumption on the distribution of the matrix elements of X and only requires odd moments are zero and even moments satisfy a Gaussian type bound. Then for the null case and under the restriction that as $n, p \to \infty$ that

$$n - p = \mathrm{O}(p^{1/3})$$

we get the same limit law described by $F_1(s, k)$.

4. **Beyond the Null Hypothesis**: If $A \sim W_p(n, \Sigma)$, then joint eigenvalue density

$$c_{p,n,\Sigma} \prod_{j=1}^{p} l_j^{(n-p-1)/2} \prod_{j<k} |l_j - l_k| \times$$

$$\int_{\mathcal{O}(p)} \mathrm{e}^{-\frac{1}{2}\,\mathrm{tr}(\Sigma^{-1} Q L Q^T)}\, dQ,$$

where $L = \mathrm{diag}(l_1, \ldots, l_p)$ and dQ is normalized Haar measure.

Difficulty is the integral

$$\int_{\mathcal{O}(p)} \mathrm{e}^{-\frac{1}{2}\,\mathrm{tr}(\Sigma^{-1} Q L Q^T)}\, dQ$$

For **complex data** above integral is replaced by

$$\int_{\mathcal{U}(p)} \mathrm{e}^{-\frac{1}{2}\,\mathrm{tr}(\Sigma^{-1} U L U^*)}\, dU$$

This integral can be evaluated in terms of determinants: Harish Chandra–Itzykson–Zuber integral.

Baik, Ben Arous, Péché describe a phase transition for population covariance matrices of the form ($k = 1$ case of their theorem)

$$\Sigma_p = \mathrm{diag}\,(\ell_1, 1, \ldots, 1)\,,$$

the **spiked population model**.

$$\frac{p}{n} \to \gamma, \ \ \mu := (1+\sqrt{\gamma})^2, \ \ \sigma := (1+\sqrt{\gamma})\,(1+\sqrt{\gamma^{-1}})^{1/3}$$

$$n^{2/3}\left(\hat{\ell}_1 - \mu\right)/\sigma \Longrightarrow \begin{cases} F_2 & \ell_1 < 1+\sqrt{\gamma} \\ \tilde{F}_2 & \ell_1 = 1+\sqrt{\gamma} \end{cases}$$

where $\tilde{F}_2$ is related to F_2.

If $\ell_1 > 1+\sqrt{\gamma}$,

$$n^{1/2}\left(\hat{\ell}_1 - \mu(\ell_1)\right)/\sigma(\ell_1) \Longrightarrow N(0,1)$$

with

$$\mu(\ell_1) = \ell_1\left(1+\frac{\gamma}{\ell_1 - 1}\right), \ \ \sigma^2(\ell_1) = \ell_1^2\left(1-\frac{\gamma}{(\ell_1-1)^2}\right)$$

Below the BBP phase transition the distribution of $\hat{\ell}_1$ is unchanged regardless of the value of ℓ_1.

Remarks concerning the BBP PHASE TRANSITION:

1. For **real data**, Baik & Silverstein and Paul show the existence of a phase transition at the same value
$$1 + \sqrt{\gamma}$$
and give a distributional normal law for $\ell_1 > 1 + \sqrt{\gamma}$. One expects an F_1 law below, but no theorem!

2. Patterson, Price & Reich use these results to construct statistical tests to determine if the samples from genetic data are from a population that has structure; that is, can the samples be regarded as randomly chosen from a homogeneous population (null hypothesis), or does the data imply that the population is not genetically homogeneous.

3. See Johnstone's 2006 ICM lecture for further details and references.

Appearance of Limit Laws Outside of RMT

Major breakthrough when Baik, Deift, Johansson proved that the limiting distribution of the **length of the longest increasing subsequence** in a random permutation is F_2.

Random permutation of $\{1, 2, \ldots, 10\}$:

$$\sigma = \{\mathbf{3}, 7, 10, \mathbf{5}, 9, \mathbf{6}, \mathbf{8}, 1, 4, 2\}, \;\; \ell_{10}(\sigma) = 4$$

Patience Sorting Algorithm (Aldous, Diaconis)

$$\begin{array}{cccc} & 2 & & \\ & 4 & 6 & \\ 1 & 5 & 9 & \\ 3 & 7 & 10 & 8 \end{array}$$

BDJ Theorem:

$$\lim_{n\to\infty} \Pr\left(\frac{\ell_n - 2\sqrt{n}}{n^{1/6}} \le x\right) = F_2(x)$$

and with convergence of moments, e.g.

$$\begin{aligned} E(\ell_n) &= 2\sqrt{n} + \int_{-\infty}^{\infty} x f_2(x)\,dx\; n^{1/6} + \mathrm{o}(n^{1/6}) \\ &= 2\sqrt{n} - 1.7710868074\, n^{1/6} + \mathrm{o}(n^{1/6}) \end{aligned}$$

A simulation with 100 trials for $n = 10^5$ gives an average number of piles per trial

621.96

which should be compared with the asymptotic expected value

620.389

The $2\sqrt{n}$ term alone gives **632.456**.

Key Points in the Proof of the BDJ Theorem

- Gessel proved (uses RSK and Jacobi-Trudi identity for s_λ)

$$\sum_{n\geq 0} \Pr(\ell_n \leq k)\,\frac{t^{2n}}{n!} = \det(T_k(\varphi))$$

  where $T_k(\varphi)$ is a $k \times k$ Toeplitz matrix with symbol $\varphi(z) = e^{t(z+1/z)}$.

- Use Case/Geronimo-Borodin/Okounkov identity that relates a Toeplitz determinant to a Fredholm determinant of an operator on $\ell^2(\{0,1,\ldots\})$

$$\sum_{n\geq 0} \Pr(\ell_n \leq k)\,\frac{t^{2n}}{n!} = \det(I - K_k)$$

Specifically, $\varphi = \varphi_+ \varphi_-$, then

$$K_k(i,j) = \sum_{\ell \geq 0} (\varphi_-/\varphi_+)_{k+i+\ell+1} \, (\varphi_+/\varphi_-)_{-k-j-\ell-1}$$

- Show $K_k \to K_{\mathrm{Airy}}$ in trace class norm: Use saddle point method on Fourier coefficients appearing in CGBO identity.

Nontrivial limit only when two saddle points coalesce

Airy function generic behavior

$$K_{\mathrm{Airy}}(x,y) = \int_0^\infty \mathrm{Ai}(x+z)\mathrm{Ai}(y+z)\, dz$$

The one-third scaling is a direct result of this coalescence—viz. the cubic power in the Airy function integral:

$$\mathrm{Ai}(z) = \frac{1}{2\pi i} \int_C e^{\xi^3/3 - z\xi} \, d\xi$$

- A dePoissonization lemma to get limit theorem.

The BDJ Theorem resulted in a burst of activity relating the distribution functions of RMT to problems in combinatorics, representation theory of the symmetric group, growth processes and determinantal random fields

Cast of Players

M. Adler, D. Aldous, J. Baik, P. Bleher, T. Bodineau, A. Borodin, P. Deift, P. Diaconis, P. Ferrari, P. Forrester, J. Gravner, T. Imamura, A. Its, K. Johannson, J. Martin, K. McLaughlin, N. O'Connell, A. Okounkov, G. Olshanski, M. Prähoffer, E. Rains, N. Reshetikhin, T. Sasamoto, A.Soshnikov, H. Spohn, C. Tracy, P. van Moerbeke, H. Widom, ...

But that is another story ...
THANK YOU FOR YOUR ATTENTION