Large Dimensional Random Matrices: Some Models and Applications

A tour through some pioneering breakthroughs

Victor Pérez-Abreu
Department of Probability and Statistics
CIMAT, Guanajuato, Mexico

60th ISI World Statistics Congress

Rio de Janeiro, Brazil
July 27, 2015

Goal of the Lecture

Some pioneering work and further developments: A personal taste.

1928-1930s: Wishart, Fisher. Multivariate data analysis.

1950s: Wigner. RMT in physics.

1962: Dyson.Time evolution of energy levels.

1967: Marchenko-Pastur. Large dimensional covariance.

1972: Dyson-Montgomery. Unexpected connections.

1991: Voiculescu. Asymptotically free random matrices.

1998: Tracy- Widom. Maximum eigenvalue.

1999: Telatar. RMT and wireless communication.

2001: Johnstone. RMT & non standard PCA

Plan of the Lecture

1. Notation and classical large sample theory.
2. Wigner law.
3. Random covariance matrices.
   - 3.1 Data dimension less than sample size.
   - 3.2 Large dimensional covariance matrices.
   - 3.3 Data dimension equal or bigger than sample size.
   - 3.4 Distribution of the maximum eigenvalue.
4. Random Matrices and wireless communication.
5. Asymptotically free random matrices.
6. Time-varying random matrices.
7. Unexpected connections.
8. Conclusions.

I. Some classical large sample theory

Notation and elementary facts

- $X_1, \ldots, X_n$ observations from an unkonwn distribution F.
- Empirical distribution function:

$$\widehat{F}_n(x) = \frac{1}{n} \sum_{j=1}^{n} 1_{\{X_j \le x\}} = \frac{1}{n} \# \{i : X_i \le x\}.$$

- Unbiased estimator of $F(x)$: for each x

$$\mathbb{E}(\widehat{F}_n(x)) = F_n(x).$$

- Many sample statistics like sample mean, variance, moments, medians, quantiles, etc., are functionals of $\widehat{F}_n$, for example

$$p\text{th-sample moment} = \int x^p F_n(\mathrm{d}x) = \frac{1}{n} \sum_{j=1}^{n} X_j^p.$$

I. Large sample theory

The so-called functional results

- If observations $X_1, \ldots, X_n$ are independent or weakly dependent, one has
  - Glivenko-Cantelli theorem:

$$\mathbb{P}\left(\sup_{x \in \mathbb{R}} \left| \widehat{F}_n(x) - F(x) \right| \underset{n \to \infty}{\longrightarrow} 0\right) = 1.$$

  - Donsker´s theorem: (functional convergence) The sequence $G_n(x) = \sqrt{n}(\widehat{F}_n(x) - F(x))$ converges to a zero-mean Gaussian process G with covariance

$$\mathrm{Cov}(G(s)G(t)) = \min\{F(s), F(t)\} - F(s)F(t).$$

- Distributions of several goodness of fit tests are found using these results: Kolmogorov-Smirnov, Cramer-von Mises, Anderson-Darling, etc.

I. Notation random matrices and spectral statistics

- A random matrix $A = (A_{ij})$ has random entries A_{ij}.
- For $n \times n$ A, its spectrum is the set of eigenvalues $\{\lambda_1, ..., \lambda_n\}$

  $$\{\lambda : \det(A - \lambda \mathrm{I}_n) = 0\} = \{\lambda : \text{there is a vector } x \neq 0,\ Ax = \lambda x\}.$$

- (spectral statistics) are function of the eigenvalues:
  - Trace $\operatorname{tr}(A) = \sum \lambda_j$, Determinant $\det(A) = \prod \lambda_j$.
  - Gaps: $\lambda_i - \lambda_j$, maximum and minimum eigenvalue.
  - Correlation functions, logarithm of determinant, capacity of a communication channel, and more.

- Empirical Spectral Distribution (ESD)

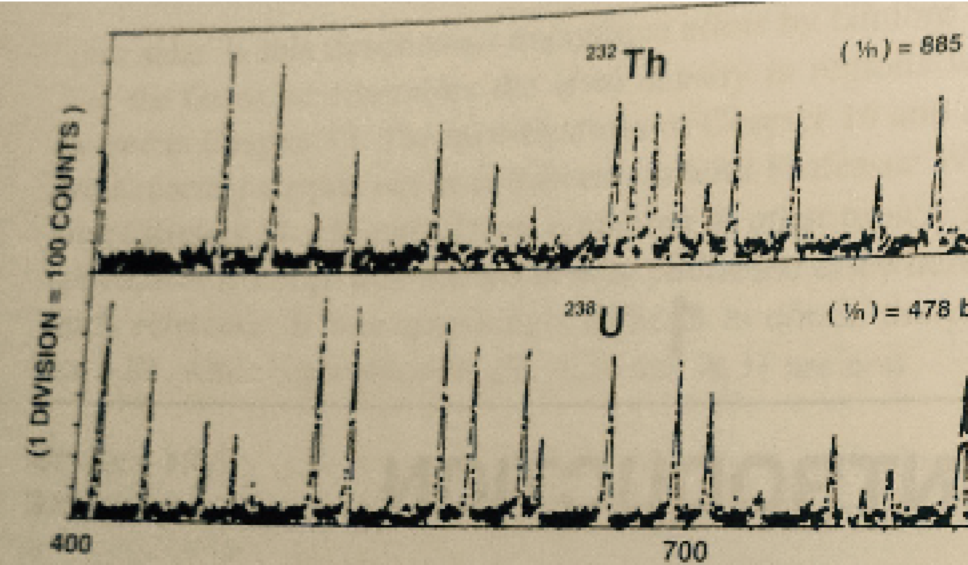
  $$\widehat{F}_n^A(x) = \frac{1}{n} \sum_{j=1}^{n} \mathbf{1}_{\{\lambda_{,j} \leq x\}} = \frac{1}{n} \# \{\text{eigenvalues of } A \leq x\}.$$

II. Random Matrix Theory in Nuclear Physics

Pioneering work of Eugene Wigner

II. Random matrices and nuclear physics

Slow neutron resonance cross-sections on thorium 232 & uranium 238 nuclei. Energy(eV)



II. Gaussian Orthogonal Ensemble (GOE)

- Ensemble: $\mathbf{Z} = (Z_n)$, Z_n is $n \times n$ matrix with random entries.
- A) GOE: $Z_n = (Z_n(j,k))$ is $n \times n$ symmetric matrix with independent Gaussian entries in the upper triangular part:

$$\begin{aligned} Z_n(j,k) &= Z_n(k,j) \sim N(0,1), \quad j \neq k, \\ Z_n(j,j) &\sim N(0,2). \end{aligned}$$

- B) Distribution of Z_n is orthogonal invariant: $OZ_nO^\top$ & Z_n have same distribution for each orthogonal matrix O.
- Characterization GOE: A and B holds.

II. Gaussian Orthogonal Ensemble (GOE)

- Joint density of eigenvalues of $\lambda_1, \ldots, \lambda_n$ of Z_n:

$$f_{\lambda_1,\ldots,\lambda_n}(x_1, \ldots, x_n) = k_n \underbrace{\left[\prod_{j=1}^{n} \exp\left(-\frac{1}{4}x_j^2\right)\right]}_{\text{independence}} \underbrace{\left[\prod_{j<k} |x_j - x_k|\right]}_{\text{strong dependence}}$$

- Nondiagonal RM: eigenvalues are strongly dependent due to Vandermont determinant: $x = (x_1, \ldots, x_n) \in \mathbb{C}^n$

$$\Delta(x) = \det\left(\left\{x_j^{k-1}\right\}_{j,k=1}^{n}\right) = \prod_{j<k}(x_j - x_k)\,.$$

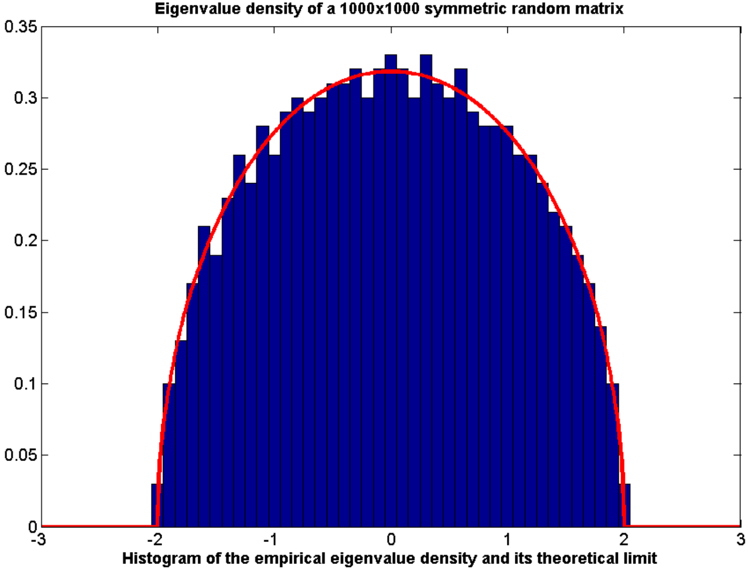
II. Wigner semicircle law

Wigner 1950s: Birth of RMT when both dimensions goes to ∞.

- A heavy nucleus is a liquid drop composed of many particles with unknown strong interactions,
- so a random matrix would be a possible model for the Hamiltonian of a heavy nucleus.
- Which random matrix should be used?
- $\lambda_1 \leq ... \leq \lambda_n$ eigenvalues of scaled GOE: $X_n = Z_n / \sqrt{n}$.
- Sample Spectral Distribution $\widehat{F}_n^{X_n}$
- Limiting Spectral Distribution (LSD): $\widehat{F}_n^{X_n}$ goes, as $n \to \infty$, to Semicircle distribution on $(-2, 2)$

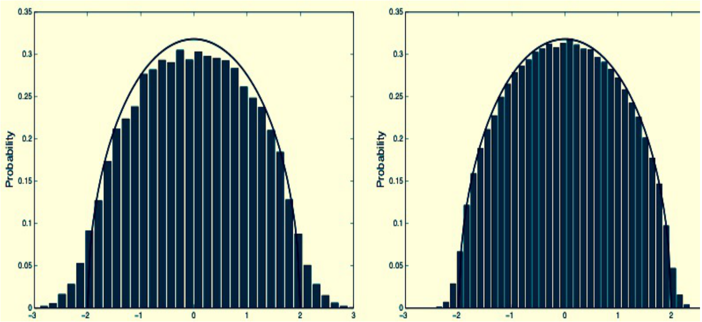
$$w(x) = \frac{1}{2\pi}\sqrt{4 - x^2}, \quad |x| \leq 2.$$

II. Simulation of Wigner semicircle law



II. Good predictions for small n

n=5 and 20



II. Precise statement of Wigner semicircle law

Semicircle distribution approximates the spectral distribution

Theorem: For each continuous bounded function f and $\epsilon > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\left| \int f(x) \mathrm{d}\widehat{F}_n^{X_n}(x) - \int f(x) w(\mathrm{d}x) \right| > \varepsilon \right) = 0.$$

where $w(x)$ is the density of *semicircle distribution on* $(-2, 2)$

$$w(x) = \frac{1}{2\pi}\sqrt{4 - x^2}, \quad |x| \leq 2.$$

- Good predictions for moderate dimension n.
- Breakthrough work by Eugene Wigner: *Ann. Math.*, 1955, 1957, 1958.

II. Gaussian Unitary Ensemble (GUE)

Wigner law also holds

- GUE: $Z_n = (Z_n(j,k))$ is $n \times n$ Hermitian with independent Gaussian entries:

$$Z = \begin{pmatrix} Z_n(1,1) & Z_n(1,2) & \ldots & Z_n(1,n) \\ \overline{Z}_n(1,2) & Z_n(2,2) & & \\ & & & \\ \overline{Z}_n(1,n) & & & Z_n(n,n) \end{pmatrix}$$

$$\mathrm{Re}\,(Z_n(j,k)) \sim \mathrm{Im}\,(Z_n(j,k)) \sim N(0, t(1+\delta_{jk})/2),$$

$$\mathrm{Re}\,(Z_n(j,k))\,, \mathrm{Im}\,(Z_n(j,k))\,, 1 \le j \le k \le n,$$

are independent random variables.

- Distribution of Z_n is unitary invariant: UZ_nU^* & Z_n have same distribution for each unitary non-random matrix U.

II. Universality

- Wigner semicircle law holds for **Wigner ensembles**:

$$X_n(k,j) = X_n(j,k) = \frac{1}{\sqrt{n}} \begin{cases} Z_{j,k}, & \text{if } j < k \\ Y_j, & \text{if } j = k \end{cases}$$

$\{Z_{j,k}\}_{j \le k}, \{Y_j\}_{j \ge 1}$ independent sequences of i.i.d.r.v. with

$$\mathbb{E} Z_{1,2} = \mathbb{E} Y_1 = 0, \mathbb{E} Z_{1,2}^2 = 1, \mathbb{E} Y_1^2 < \infty.$$

- Whatever values the random entries take, the LSD (Semicircle) has bounded support.
- Joint density of eigenvalues of a Wigner matrix is not easy.
- Limit spectral statistics known for Gaussian matrices, also hold for Wigner ensembles (Tao and Vu, 2012).

III. Sample Covariance Matrix

A. The pioneering work of Wishart

Data dimension fixed, less than sample size (varying)

III.A. Pioneering work of J. Wishart.

Wishart (1928), The generalized product moment distribution in samples from multivariate population, *Biometrika*.

- $H = H_{p\times n} = \left(Z_{j,k} : j = 1, .., p, k = 1, ..., n\right)$ is a $p \times n$ rectangular random matrix

$$H = \begin{pmatrix} Z_{1,1} & & Z_{1,n} \\ & & \\ Z_{p,1} & & Z_{p,n} \end{pmatrix} = \left(\underline{Z}_1 \cdots \underline{Z}_n\right),$$

  $\underline{Z}_1, ..., \underline{Z}_n$ is a sample from p-variate normal distribution with zero mean-vector and covariance matrix Σ_p.

- p is the data dimension and n is the *sample size*, $n \geq p$.
- Sample covariance matrix is the $p \times p$ random matrix

$$S_n = \frac{1}{n} HH^\top = \frac{1}{n} \sum_{i=1}^{n} \underline{Z}_i \underline{Z}_i^\top .$$

III.A. Pioneering work of J. Wishart and others

- $W_n = nS_n$ is the Wishart random matrix, ($W_n \sim \mathrm{W}_p(n, \Sigma)$).
- Wishart (1928): Found a formula for the density of $\mathrm{W}_p(n, \Sigma)$, it is the matrix version of the chi-square distribution.
- 1930's: Different aspects of the Wishart random matrices and its eigenvalues: Fisher, Hsu, Girshick, Roy, Lévy.
- Fisher (1939): Joint distribution of the ordered eigenvalues $\lambda_1 < \ldots < \lambda_p$ of W_n with distribution $\mathrm{W}_p(n, \mathrm{I}_p)$

$$c_{n,p} \underbrace{\exp\Big(-\sum_{j=1}^{p} \lambda_j\Big) \prod_{j=1}^{p} \lambda_j^{n-p}} \underbrace{\prod_{i<j}^{p} (\lambda_i - \lambda_j)^2}$$

- Again, strong interaction between eigenvalues.

III.A. Pioneering work of J. Wishart and others

- $W_n = nS_n$ is the *Wishart random matrix*, $(W_n \sim \mathrm{W}_p(n, \Sigma))$.
- Anderson (1957):
  - Asymptotic results for S_n and its eigenvalues when p is fixed and n is large.
  - Sample covariance matrix S_n is a good estimator of Σ:
    - $\mathbb{E}(S_n) = \Sigma$,

$$\mathbb{P}\left(S_n = \frac{1}{n}\sum_{i=1}^{n} \underline{Z}_i \underline{Z}_i^{\top} \to \Sigma\right) = 1.$$

  - Eigenvalues of S_n are good estimators of eigenvalues of Σ.
  - In particular, when $\Sigma = \mathrm{I}_p$ the eigenvalues of $\sqrt{n}(S_n - \mathrm{I}_p)$ converge to the eigenvalues of a GOE random matrix.

III. Sample Covariance Matrix

B. The Marchenko-Pastur Law

Both, data dimension and sample size large

III.B. Marchenko-Pastur law

Marchenko-Pastur (1967), *Mat. Sb.*

- $H = H_{p\times n} = (Z_{j,k} : j = 1, .., p, k = 1, ..., n)$ i.i.d.r.v.

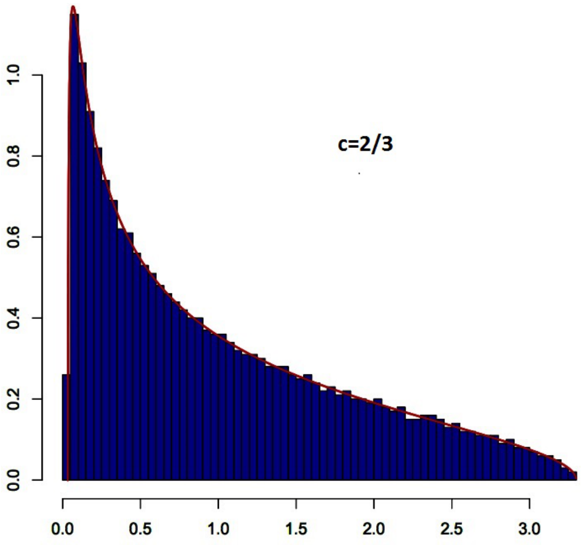
$$\mathbb{E}(Z_{1,1}) = 0,\ \mathbb{E}(\,|Z_{1,1}|^2) = 1,\ \mathbb{E}(\,|Z_{1,1}|^4) < \infty.$$

- Sample covariance matrix $S_n = \frac{1}{n} HH^*$, ESD $\widehat{F}_p^{S_n} = \widehat{F}_p^{\frac{1}{n}HH^*}$.
- If $p/n \to c > 0$, $\widehat{F}_p^{S_n}$ goes to **MP distribution**:

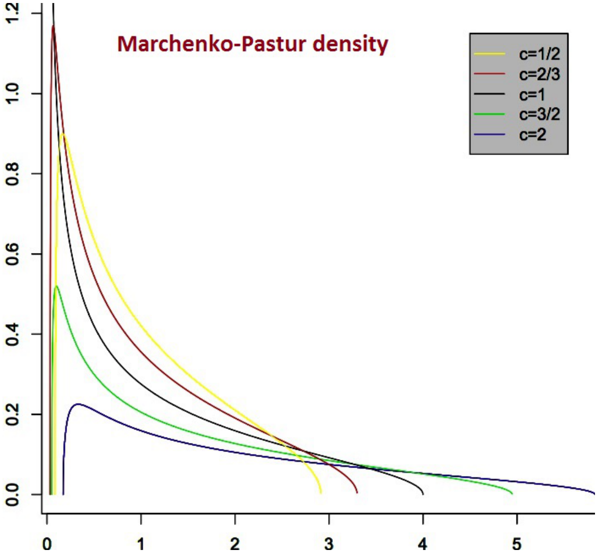
$$\mu_c(\mathrm{d}x) = \begin{cases} f_c(x)\mathrm{d}x, & \text{if} \quad c \geq 1 \\ (1-c)\delta_0(\mathrm{d}x) + f_c(x)\mathrm{d}x, & \text{if} \quad 0 < c < 1, \end{cases}$$

$$f_c(x) = \frac{c}{2\pi x}\sqrt{(x-a)(b-x)}\mathbf{1}_{[a,b]}(x)$$
$$a = (1-\sqrt{c})^2, \ \ b = (1+\sqrt{c})^2.$$

III.B. Simulation Marchenko-Pastur law

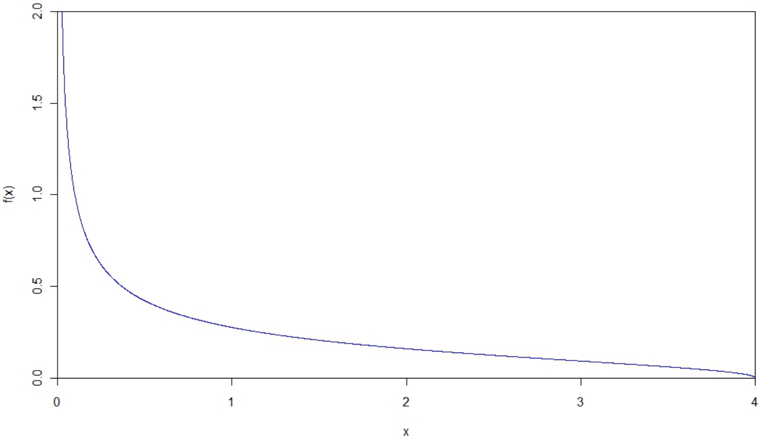


III.B. Marchenko-Pastur distribution



III.B. MP distribution parameter 1

$c = 1$, zero mean and standard deviation 2,
density $f_1(x) = \frac{1}{2\pi x}\sqrt{x(4-x)}\mathbf{1}_{[0,4]}(x)$



III. Sample Covariance Matrix

C. Data dimension equal or bigger than sample size

Learning from the Marchenko-Pastur law

III.C. Random matrices and PCA

- $p/n \to c > 0$, gives the **Marchenko-Pastur support**:

$$\left[a = (1-\sqrt{c})^2,\ b = (1+\sqrt{c})^2\right].$$

- Wishart case: $nS_n = H_n H_n^\top \sim W_p(n, \mathrm{I}_p)$, H_n sample of n i.i.d. Gaussian vectors with mean zero and covariance $\Sigma = \mathrm{I}_p$.

- Problem in Statistics: How well eigenvalues $\lambda_1^n \leq ... \leq \lambda_p^n$ (of S_n) estimate the population eigenvalues (of Σ) when data dimension p and sample size n are of the same order?.

- **Johnstone** (2001). On the distribution of the largest eigenvalue in principal component analysis. *Ann. Statist.*
- **Johnstone** (2007). High dimensional statistical inference and random matrices. *Proc. ICM.* Madrid

III.C. PCA and RMT

Data dimension of same order than sample size

- Take $p = n$, $c = 1$, $n = 10$ independent observations from a multivariate Gaussian distribution with zero mean and covariance $\Sigma = \mathrm{I}_p$. Eigenvalues of S_{10} are

$$(0.003, 0.036, 0.095, 0.16, 0.30, 0.51, 0.78, 1.12, 1.40, 3.07).$$

- Extreme spread in sample eigenvalues and not all close to one.
- This phenomenon is explained by the MP since the ESD with $p = n$ goes to MP law with support

$$a = (1 - \sqrt{1})^2 = 0 \text{ and } b = (1 + \sqrt{1})^2 = 4.$$

- It is not easy to estimate the population eigenvalues of Σ when p, n are of the same order..

III.C. Tracy-Widom distribution

Motivation to consider the asymptotic distribution for the maximum eigenvalue of sample covariance matrix, obtaining Tracy-Widom distribution.

Let $\lambda_{\max} = \lambda_p^n$ be the largest eigenvalue of $S_n \sim \mathrm{W}_p(n, \mathrm{I}_p)$. Define

$$r_n = \sqrt{n-1} + \sqrt{p},\ q_n = r_n\left(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{p}}\right)^{1/3}$$

$$\lim_n \mathbb{P}\left[\frac{\lambda_{\max} - r_n^2}{q_n} \leq t\right] = F_1(t)$$

where F_1 is the **Tracy-Widom distribution** defined by

$$F_1(t) = \exp\left(-\frac{1}{2}\int_t^\infty \left(q(x) + (x-t)q(x)\right)^2 \mathrm{d}x\right)$$

where q is a solution of a Painlevé II differential equation

$$q\prime\prime = tq + 2q^3, \quad q(t) \sim Ai(t.) \text{ as } t \to \infty.$$

III.C. PCA and RMT

Example

Suppose the observed largest sample eigenvalue is equal to 4.25. Is this consistent with $H_0 : \Sigma = I_p$?. By the last result

$$\lim_n \mathbb{P}\left[\frac{\lambda_{\max} - r_n^2}{q_n} \leq t\right] = F_1(t)$$

$$r_n = \sqrt{n-1} + \sqrt{p},\ q_n = r_n\left(\frac{1}{\sqrt{n-1}} + \frac{1}{\sqrt{p}}\right)^{1/3}$$

Then

$$\mathbb{P}\left(\lambda_{\max} > 4.25\right) \approx F_1\left(\frac{10(4.25) - r_n^2}{q_n}\right) = 0.06.$$

Therefore we do not reject H_0 with significance level 5%.

III.C. Tracy-Widom distribution

Tracy, Widom (*Phys. Lett. B* 1993, *Comm. Math. Phys.* 1994).

- Limiting distribution for the largest eigenvalue of GOE, GUE and other matrices.
- Due to strong dependence of the eigenvalues, the limiting distribution cannot be one of the classical distributions for extremes (Gumbel, Fréchet, Weibull).
- Universality: It appears in a variety of increasing contexts:

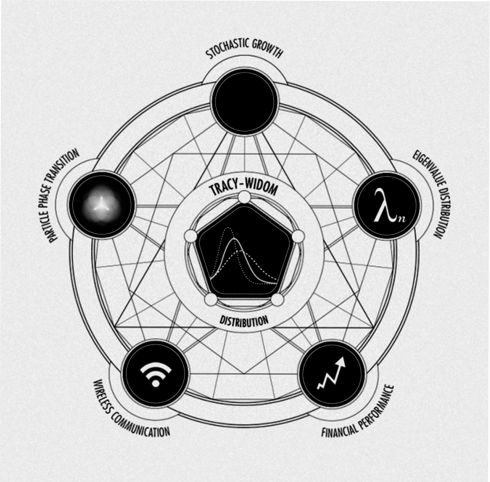
Universality of Tracy-Widom distribution

QUANTA MAGAZINE
illuminating science

STATISTICAL PHYSICS

At the Far Ends of a New Universal Law

A potent theory has emerged explaining a mysterious statistical law that arise throughout physics and mathematics.



IV. RMT and Wireless Communications

Pioneering work of Emre Telatar

Telatar (1999), Capacity of multi-antenna Gaussian channels. *European Transactions on Telecommunications.*

- A $p \times 1$ complex Gaussian random vector $\mathbf{u} = (u_1 \ \cdots \ u_p)^\top$ has a *Q-circularly symmetric complex Gaussian distribution* if

$$\mathbb{E}[(\hat{\mathbf{u}} - \mathbb{E}[\hat{\mathbf{u}}])(\hat{\mathbf{u}} - \mathbb{E}[\hat{\mathbf{u}}])^*] = \frac{1}{2}\begin{bmatrix} \operatorname{Re}[Q] & -\operatorname{Im}[Q] \\ \operatorname{Im}[Q] & \operatorname{Re}[Q] \end{bmatrix},$$

  for some nonnegative definite Hermitian $p \times p$ matrix Q where

$$\hat{\mathbf{u}} = [\operatorname{Re}(u_1), \ldots, \operatorname{Re}(u_p), \operatorname{Im}(u_1), \ldots, \operatorname{Im}(u_p)]^\top .$$

IV. Telatar: RMT and Channel Capacity

- n_T antennas at trasmitter and n_R antennas at receiver.
- Linear channel with Gaussian noise

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}.$$

- $\mathbf{x}$ is the n_T-dimensional input vector. $(n_T = n)$.
- $\mathbf{y}$ is the n_R-dimensional output vector. $(n_R = p)$.
- $\mathbf{n}$ is the receiver 0-mean Gaussian noise, $\mathbb{E}\left(\mathbf{n}\mathbf{n}^*\right) = \mathrm{I}_{n_T}$.
- The $n_R \times n_T$ random matrix $\mathbf{H}$ is the channel matrix.
- $\mathbf{H} = \{h_{jk}\}$ is a random matrix. It models the propagation coefficients between each pair of trasmitter-receiver antennas.
- $\mathbf{x}, \mathbf{H}$ and $\mathbf{n}$ are independent.

- h_{jk} are i.i.d. complex r.v. with 0-mean and variance one ($\mathrm{Re}(Z_{jk}) \sim N(0, \frac{1}{2})$ independent of $\mathrm{Im}(Z_{jk}) \sim N(0, \frac{1}{2})$).
- Total power constraint P: upper bound for variance $\mathbb{E}||\mathbf{x}||^2$ of the input signal amplitude.
- Signal to Noise Ratio (SNR)

$$SNR = \frac{\mathbb{E}||\mathbf{x}||^2 / n_T}{\mathbb{E}||\mathbf{n}||^2 / n_R} = \frac{P}{n_T}.$$

- Channel capacity is the maximum data rate which can be transmitted reliably over a channel (Shannon (1948)).
- The capacity of this MIMO system channel is

$$C(n_R, n_T) = \max_{Q} \mathbb{E}_{\mathbf{H}} \left[\log_2 \det \left(\mathrm{I}_{n_R} + \mathbf{H} Q \mathbf{H}^* \right) \right].$$

- Maximum capacity when $Q = SNR\mathrm{I}_{n_T}$

$$C(n_R, n_T) = \mathbb{E}_{\mathbf{H}}\left[\log_2 \det\left(\mathrm{I}_{n_R} + \frac{P}{n_T}\mathbf{H}\mathbf{H}^*\right)\right]$$

- Maximum capacity when $Q = SNR\mathrm{I}_{n_T}$

$$C(n_R, n_T) = \mathbb{E}_{\mathbf{H}}\left[\log_2 \det\left(\mathrm{I}_{n_R} + \frac{P}{n_T}\mathbf{HH}^*\right)\right]$$

- In terms of ESD $\widehat{F}_{n_T}^{\frac{1}{n_T}\mathbf{HH}^*}$ of sample covariance $\frac{1}{n_T}\mathbf{HH}^*$

$$C(n_R, n_T) = n_R \int_0^\infty \log_2(1 + Px)\,\mathrm{d}\widehat{F}_{n_T}^{\frac{1}{n_T}\mathbf{HH}^*}.$$

- Maximum capacity when $Q = SNR\mathrm{I}_{n_T}$

$$C(n_R, n_T) = \mathbb{E}_{\mathbf{H}}\left[\log_2 \det\left(\mathrm{I}_{n_R} + \frac{P}{n_T}\mathbf{HH}^*\right)\right]$$

- In terms of ESD $\widehat{F}_{n_T}^{\frac{1}{n_T}\mathbf{HH}^*}$ of sample covariance $\frac{1}{n_T}\mathbf{HH}^*$

$$C(n_R, n_T) = n_R \int_0^\infty \log_2(1 + Px)\, \mathrm{d}\widehat{F}_{n_T}^{\frac{1}{n_T}\mathbf{HH}^*}.$$

- By Marchenko-Pastur law, if $n_R/n_T \to c$,

$$\frac{C(n_R, n_T)}{n_R} \to \int_a^b \log_2(1 + Px)\, \mathrm{d}\mu_c(x) = K(c, P).$$

- Maximum capacity when $Q = SNR\mathrm{I}_{n_T}$

$$C(n_R, n_T) = \mathbb{E}_{\mathbf{H}}\left[\log_2 \det\left(\mathrm{I}_{n_R} + \frac{P}{n_T}\mathbf{H}\mathbf{H}^*\right)\right]$$

- In terms of ESD $\widehat{F}_{n_T}^{\frac{1}{n_T}\mathbf{H}\mathbf{H}^*}$ of sample covariance $\frac{1}{n_T}\mathbf{H}\mathbf{H}^*$

$$C(n_R, n_T) = n_R \int_0^\infty \log_2(1 + Px)\,\mathrm{d}\widehat{F}_{n_T}^{\frac{1}{n_T}\mathbf{H}\mathbf{H}^*}.$$

- By Marchenko-Pastur law, if $n_R/n_T \to c$,

$$\frac{C(n_R, n_T)}{n_R} \to \int_a^b \log_2(1 + Px)\,\mathrm{d}\mu_c(x) = K(c, P).$$

- For fixed P

$$C(n_R, n_T) \sim n_R K(c, P).$$

- Maximum capacity when $Q = SNR\mathrm{I}_{n_T}$

$$C(n_R, n_T) = \mathbb{E}_{\mathbf{H}}\left[\log_2 \det\left(\mathrm{I}_{n_R} + \frac{P}{n_T}\mathbf{H}\mathbf{H}^*\right)\right]$$

- In terms of ESD $\widehat{F}_{n_T}^{\frac{1}{n_T}\mathbf{H}\mathbf{H}^*}$ of sample covariance $\frac{1}{n_T}\mathbf{H}\mathbf{H}^*$

$$C(n_R, n_T) = n_R \int_0^\infty \log_2(1 + Px)\,\mathrm{d}\widehat{F}_{n_T}^{\frac{1}{n_T}\mathbf{H}\mathbf{H}^*}.$$

- By Marchenko-Pastur law, if $n_R / n_T \to c$,

$$\frac{C(n_R, n_T)}{n_R} \to \int_a^b \log_2(1 + Px)\,\mathrm{d}\mu_c(x) = K(c, P).$$

- For fixed P

$$C(n_R, n_T) \sim n_R K(c, P).$$

- **Increase capacity with more transmitter and receiver antennas with same total power constraint** P.

IV. RMT and Wireless Communication

Some further developments

- Non Gaussian distribution for i.i.d. entries h_{ij} of the channel matrix $\mathbf{H}$: universality of the Marchenko-Pastor law.
  - Bai & Silverstein (2010). *Spectral Analysis of Large Dimensional Random Matrices.*
- Correlation models for $\mathbf{H}$, Kronecker correlation, etc..
  - Lozano, Tulino & Verdú. (2005). Impact of antenna correlation on the capacity of multiantenna channels. *IEEE Trans. Inform. Theo.*
  - Lozano, Tulino & Verdú (2006). Capacity-achieving input covariance for single-user multi-antenna channels. *IEEE Trans. Wireless Comm.*

IV. RMT and Wireless Communication

Further developments

- Books on RMT and Wireless Communications:
  - Tulino & Verdú (2004). *Random Matrix Theory and Wireless Communications.*
  - Couillet & Debbah (2011). *Random Matrix Methods for Wireless Communications.*
  - Bai, Fang & Ying-Chang (2014). *Spectral Theory of Large Dimensional Random Matrices and Its Applications to Wireless Communications and Finance Statistics*.
- Main problem is the computation of the asymptotic channel capacity, mainly done by a technique introduced by Girko (1990), solving a non-linear system of functional equations.
  - Couillet, R., Debbah, M., and Silverstein, J. (2011). A deterministic equivalent for the analysis of correlated MIMO multiple access channels. *IEEE Trans. Inform.Theo.*

- Recently, tools from Operator-valued free probability theory have been successful used as alternative to approximate the asymptotic capacity of new models:
- Ding (2014), Götze, Kösters & Tikhomirov (2015), Hachem, Loubaton & Najim (2007), Shlyakhtenko (1996), Helton, Far & Speicher (2007), Speicher, Vargas & Mai (2012), Belinschi, Speicher, Treilhard & Vargas (2014), Belinschi, Mai & Speicher, R. (2015), Diaz-Torres & PA (2015).
- Operator-valued free probability (Speicher, 1988) is an extension of **free probability theory** of Voiculescu (1985).
- There is a relation between block based large dimensional random matrices and operator-valued free probability, similar to the relation between:
- *Large dimensional matrices and free probability: Asymptotically free random matrices, Voiculescu (1991).*

V. Random Matrices and Free Probability

The work of Dan Voiculescu

(briefly)

V. RMT and Free Probability

From the Blog of Terence Tao (Free Probability):

- The significance of free probability to random matrix theory lies in the fundamental observation that *random matrices which are independent in the classical sense, also tend to be independent in the free probabilistic sense*, in the large limit.
- Because of this, many tedious computations in random matrix theory, particularly those of an algebraic or enumerative combinatorial nature, can be done more quickly and systematically by using the *framework of free probability.*
- **Voiculescu (1991)**, Limit Laws for random matrices and free products. *Invent. Math.*
- Books on random matrices and wireless communications include free probability.

V. RMT and Free Probability: Why useful?

- Knowing eigenvalues of $n \times n$ random matrices X_n & Y_n, what are the eigenvalues of $X_n + Y_n$? $X_n Y_n$?
- In general if X_n and Y_n do not commute,

$$\lambda(X_n + Y_n) \neq \lambda(X_n) + \lambda(Y_n)$$

- The same for the eigenvalues of the product:

$$\lambda(X_n Y_n) \neq \lambda(X_n)\lambda(Y_n).$$

- However, if X_n & Y_n are *asymptotically free*, LSD of $X_n + Y_n$ can be computed as (*free convolution)*.

$$LSD(X_n + Y_n) = LSD(X_n) \boxplus LSD(Y_n)$$

V. RMT and Free Probability: Why useful?

- The problem is similar to the computation to the distribution of the sum of two independent random variables: product of characteristic functions or moment generating functions (classical convolution).
- This allows for more general non linear channels like $H_1 H_2 + H_3$.
- *Operator valued free probability* arises from considering block-based random matrices

$$\begin{pmatrix} H_1 & H_2 \\ H_3 & H_4 \end{pmatrix}.$$

- It also allows for more models and *alternative computation of capacity channels*

VI. Time-varying random matrices

The pioneering work of Freeman Dyson

VI. Time-varying random matrices: why?

Couillet & Debbah (2011), *Random Matrix Methods for Wireless Communications.* Chapter 19, Perspectives:

- Performance analysis of a typical network with users in motion according to some stochastic behavior, is not accessible to this date in the restrictive framework of random matrix theory.

VI. Time-varying random matrices: why?

Couillet & Debbah (2011), *Random Matrix Methods for Wireless Communications.* Chapter 19, Perspectives:

- Performance analysis of a typical network with users in motion according to some stochastic behavior, is not accessible to this date in the restrictive framework of random matrix theory.

- It is to be believed that random matrix theory for wireless communications may move on a more or less long-term basis towards random matrix process theory for wireless communications. Nonetheless, these random matrix processes are nothing new and have been the interest of several generations of mathematicians.

VI. Time-varying random matrices

Pioneering work of Dyson

Dyson (1962):, A Brownian-motion model for the eigenvalues of a random matrix. *J. Math. Phys.*:

- Description of the motion on time of the energy levels.
- Time dependence of the statistics of the eigenvalues of a Gaussian matrix.

VI. Time-varying GOE: matrix Brownian motion

- $\mathbf{B}(t) = (B_n(t))_{n \geq 1}$.
- $B_n(t)$ is $n \times n$ symmetric Brownian motion

$$B_n(t) = (b_n^{i,j}(t))_{1 \leq i,j \leq n}, \quad b_n^{j,i}(t) = \overline{b}_n^{i,j}(t),$$

$$b_n^{j,i}(t) \sim N(0, t(1 + \delta_{ij})/2),$$

$$b_n^{j,i}(t), \; 1 \leq i \leq j \leq n,$$

are independent one-dimensional Brownian motions.

- $(\lambda_1(t), ..., \lambda_n(t))_{t \geq 0}$ process of eigenvalues of $B_n(t)$.

VI. Dyson-Brownian motion

Time dynamics of the eigenvalues, dimension n fixed

Dyson (1962):

a) If eigenvalues start at different positions, they never collide

$$\mathbb{P}\left(\lambda_1(t) > \lambda_2(t) > \dots > \lambda_n(t) \quad \forall t > 0\right) = 1.$$

b) They satisfy the Stochastic Differential Equation

$$\lambda_i(t) = W_i(t) + \sum_{j \neq i} \int_0^t \frac{\mathrm{d}s}{\lambda_j(s) - \lambda_i(s)}, \quad i = 1, \dots, n, \forall t > 0.$$

where $W_1, .., W_n$ are one-dimensional Brownian motions.

- Brownian part + repulsion part (at any time t).

VI. Dyson-Brownian motion

Time dynamics of the eigenvalues, dimension n fixed

Dyson (1962):

a) If eigenvalues start at different positions, they never collide

$$\mathbb{P}\left(\lambda_1(t) > \lambda_2(t) > \ldots > \lambda_n(t) \quad \forall t > 0\right) = 1.$$

b) They satisfy the Stochastic Differential Equation

$$\lambda_i(t) = W_i(t) + \sum_{j \neq i} \int_0^t \frac{\mathrm{d}s}{\lambda_j(s) - \lambda_i(s)}, \quad i = 1, \ldots, n, \forall t > 0.$$

where $W_1, .., W_n$ are one-dimensional Brownian motions.

- Brownian part + repulsion part (at any time t).
- Anderson, Guionnet, Zeitouni (2010), Tao (2012).

VI. Some extensions of Dyson-Brownian motion

- Matrix processes with semicircle limit (free Brownian motion):Chan (1992), Rogers & Shi (1993), Katori & Tanemura (2004, 2013), Cépa & Lepingle (1997), PA & Tudor (2007).
- Wishart covariance process: **Bru** (1989, J. Multivart. Analys., 1991, J. Theoret. Probab.), Cabanal Duvillard & Guionnet (2001), Konig & Connell (2001), PA & Tudor (2009).

$$W_n(t) = B_n(t)B_n(t)^*, \ t \geq 0.$$

- Matrix Fractional Brownian Motion: Nualart & PA (2014), Pardo, Pérez G. & PA (2015) $\rightarrow$ Free fractional Brownian.
- Hermitian Lévy process with distribution invariant under unitary conjugations: PA & Rocha-Arteaga (2015), Pérez G., PA & Rocha-Arteaga (2015).
- Wishart fractional Brownian: Pardo, Pérez G. & PA (2015).

VI. The matrix Fractional Brownian motion case

- Consider $n(n+1)/2$ independent fractional Brownian motions of parameter $H \in (1/2, 1)$,

$$b^H = \{\{b^H_{i,j}(t), t \geq 0\}, 1 \leq i, j \leq n\}.$$

  - $b^H_{i,j}(t)$ is a zero-mean Gaussian process with covariance

$$\mathbb{E} b^H_{i,j}(t) b^H_{i,j}(s)) = \frac{1}{2}\left(t^{2H} + s^{2H} - |t-s|^{2H}\right)$$

  - It has stationary increments:

$$\mathbb{E}\left|b^H_{i,j}(t) - b^H_{i,j}(s))\right| = |t-s|^{2H}$$

  - $H = 1/2$ is Brownian motion (independent increments).

- Define the symmetric matrix Fractional Brownian motion B^H by $B^H_{ij}(t) = b^H_{i,j}$ if $i < j$ and $B^H_{ii}(t) = \sqrt{2} b^H_{i,i}(t)$.

VI. The Fractional Brownian motion case

Theorem (Nualart & PA 2014, SPA)

a) If the eigenvalues start at different positions, they never collide at any time,

$$\mathbb{P}\left(\lambda_1(t) > \lambda_2(t) > \ldots > \lambda_n(t) \quad \forall t > 0\right) = 1.$$

b) For any $t > 0$ and $i = 1, \ldots, n\lambda$

$$_i(t) = \lambda_i(0) + Y^i(t) + 2H \sum_{j \neq i} \int_0^t \frac{1}{\lambda_i(s) - \lambda_j(s)} \mathrm{d}s$$

$$Y^i(t) = \sum_{k \leq h} \int_0^t \frac{\partial \lambda_i}{\partial b_{kh}^H} \delta b_{kh}^H.$$

VI. Time-varying Wigner theorem

ESD process of re-scaled matrix process $B_n^H(t)/\sqrt{n}$

$$\mu_t^{(n)}(x) = \frac{1}{n}\sum_{j=1}^{n} \mathbf{1}_{\{\lambda_j(t)/\sqrt{n}\le x\}}.$$

Theorem

Fix $T>0$. For all continuous bounded functions f and any $\varepsilon > 0$

$$\lim_{n\to\infty} \mathbb{P}\left(\sup_{0\le t\le T}\left|\int f(x)\mathrm{d}\mu_t^{(n)}(x) - \int f(x)\mathrm{w}_t^H(x)\mathrm{d}x\right| > \varepsilon\right) = 0$$

where w_t^H is the semicircle distribution on $(-2t^H, 2t^H)$

$$\mathrm{w}_t^H(x) = \frac{1}{2\pi}\sqrt{4t^{2H} - x^2}\mathrm{d}x, \quad |x| \le 2t^H$$

where w_t^H is the semicircle distribution on $(-2t^H, 2t^H)$

VI. The Fractional Brownian motion case

- Pardo, Perez-G, PA, 2015, JTP: The family of measure-valued processes $\{(\mu_t^{(n)})_{t\geq 0} : n \geq 1\}$ converges to the family $(\mu_t)_{t\geq 0}$ that corresponds to the law of a free fractional Brownian motion of parameter $H \in (1/2, 1)$.
- $H = 1/2$ is the free Brownian motion:
  - **Biane (1997).** Free Brownian motion, free stochastic calculus and random matrices. *Amer. Math. Soc.*
  - Cabanal Duvillard & Guionnet (2001), *Ann. Probab.*
- $H \neq 1/2$: Free fractional Brownian motion
  - Introduced by Nordin and Taqqu (2012), *J. Theoret. Probab.*

VI. Precise statement and tools

Theorem

a)The family of measure-valued processes $\{(\mu_t^{(n)})_{t\geq 0} : n \geq 1\}$ *converges weakly in* $C(\mathbb{R}_+, \mathrm{Pr}(\mathbb{R}))$ *to the unique continuous probability-measure valued function satisfying, for each* $t \geq 0$ $f \in C_b^2(\mathbb{R})$,

$$\langle \mu_t, f\rangle = \langle \mu_0, f\rangle + H \int_0^t ds \int_{\mathbb{R}^2} \frac{f'(x) - f'(y)}{x - y} s^{2H-1} \mu_s(dx)\mu_s(dy).$$

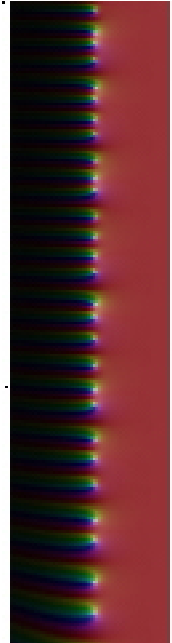
b) The Cauchy transform $G_t(z) = \int_{\mathbb{R}} \frac{\mu_t(dx)}{x-z}$ *of* μ_t *is the unique solution to the initial value problem*

$$\begin{cases} \frac{\partial}{\partial t} G_t(z) = H t^{2H-1} G_t(z) \frac{\partial}{\partial z} G_t(z), & t > 0, \\ G_0(z) = -\int_{\mathbb{R}} \frac{\mu_0(dx)}{x - z}, & z \in \mathbb{C}^+. \end{cases}$$

VI. Unexpected connections

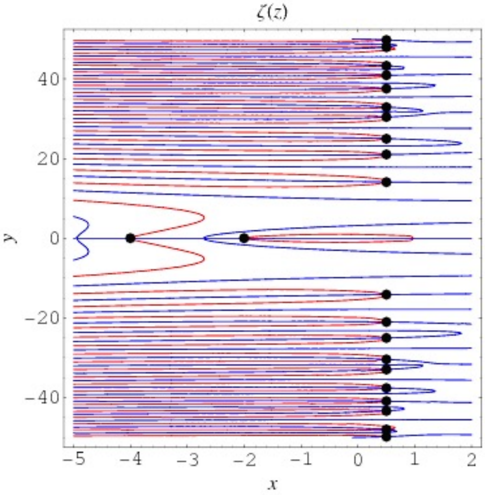
The unsolved Dyson-Montgomery conjecture

VI. Strong interactions between?



VI. Zeta Riemann function

Conjecture: Dyson and Montgomery 1973:



VI. Random matrices and the zeta Riemann function

- Riemann function: For $\mathrm{Re}(s) > 1$

$$\zeta(s) = \sum_{n\geq 1} \frac{1}{n^s} = \prod_{p \text{ prime}} \frac{1}{1 - p^{-s}}.$$

- For complex number $s \neq 1$

$$\zeta(s) = 2^s \pi^{s-1} \mathrm{sen}(\frac{\pi s}{2}) \Gamma(1-s) \zeta(1-s).$$

- Trivial zeros in $-2, -4, \ldots$& nontrivial with real part in $(0, 1)$.
- Riemann hypothesis: Nontrivial zeros in $\mathrm{Re}(s) = 1/2$.
- Dyson-Montgomery conjecture (1972), still open: Same laws of distribution seem to govern the zeros of the Riemann zeta function and the eigenvalues of random matrices.
- Opened a tantalizing unexpected connection between prime numbers, nuclear physics and random matrices.
- Mezzadri & Snaith (2005). *Recent Perspectives in Random Matrix Theory and Number Theory.* Cambridge.

VI. Dyson-Montgomery conjecture

Zeros of Riemann zeta function and eigenvalues of random matrices have same law

- Dyson knew the pair correlation function of eigenvalues of GUE:

$$R_2(x) = 1 - \left(\frac{\sin(\pi x)}{tx}\right)^2.$$

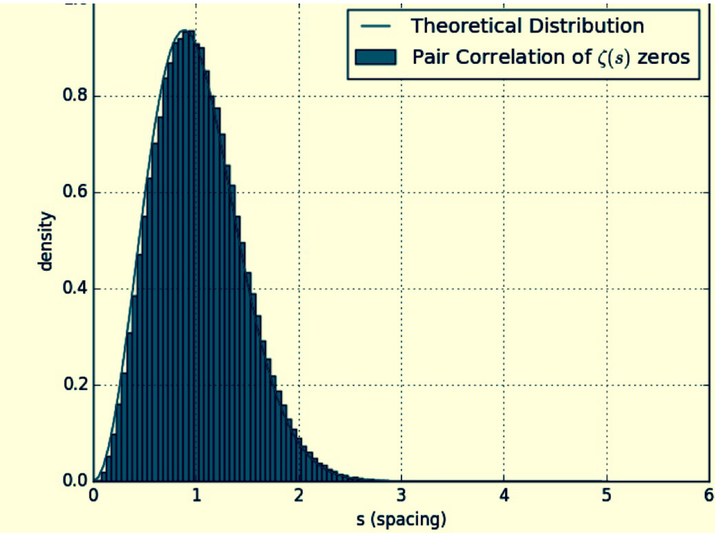
- Montgomery calculations for pair correlations between zeros of Riemann function:

$$R_2(x) = 1 - \left(\frac{\sin(\pi x)}{tx}\right)^2 + O(\frac{1}{|x|})$$

- There are over 70 millions of zeros computed by Andrew Odlyzko (1987).
- *BigData, Complex data?*
- *A lot of dependence.*

VI. Dyson-Montgomery conjecture

Odlyzko´s numerical comparisons



VI. Statistical approach to Dyson-Montgomery conjecture

Zeros of Riemann zeta function and eigenvalues of random matrices have same law

- Coram and Diaconis (2003): New tests of the correspondence between unitary eigenvalues and the zeros of Riemann's zeta. J. Physics A: Math. and General.
  - Formulation of the problem as tests of statistical hypothesis.
  - Need of simulation of large dimensional random matrices.
  - Use exponential families, consistent tests, Anderson-Darling goodness of fit test.

Conclusion: No evidence to reject the hypotheses.

Conclusions

- Random matrices appear in several fields like Multivariate statistics, Physics, Engineering, Probability, Stochastic processes and Number theory.
- Not mentioned: Finances, knot theory, random permutations, complex networks, random graphs, RNA studies, among many other fields.
- Distribution of eigenvalues are useful to model phenomena with strong interactions.
- There was a need for a completely new treatment and scope of asymptotic results.

Conclusions

- Classical ensembles of random matrices: GOE, GUE, Wigner, Wishart, sample covariance.
- Not mentioned: Unitary, Orthogonal, Circular, etc.
- Increasing number of applications and needs in wireless communications.
- Need for time varying random matrices.
- Unexpected and useful connections with free probability, (graphs and combinatorics.)

Books

Other than those on Random Matrices and Wireless Communications

- Anderson (1957), *An Introduction to Multivariate Statistical Analysis*, 1984, 2003, Wiley.
- Mehta M.L. (1967) *Random Matrices.* (1990, 2004). Elsevier.Anderson, Guionnet & Zeitouni (2010). *An Introduction to Random Matrices*. Cambridge.
- Voiculescu, Dykema & Nica (1992). *Free Random Variables.* American Mathematical Society.
- Hiai, & Petz (2000). *The Semicircle Law, Free Random Variables and Entropy*. Mathematical Surveys and Monographs **77,** American Mathematical Society. Providence.
- Nica & Speicher (2006). *Lectures on the Combinatorics of Free Probability.* London Mathematical Society Lecture Notes Series **335,** Cambridge University Press, Cambridge.
- Bai & Silverstein (2010). *Spectral Analysis of Large Dimensional Random Matrices*. Springer.
- Anderson, Guionnet & Zeitouni (2010). *An Introduction to Random Matrices.* Cambridge.

Random Matrices:

A useful, beautiful, deep and fertile field.

Thanks, Obrigado.