Universality results for the largest eigenvalue of sample covariance matrices

S. Péché

Workshop on free probability, random matrices and planar algebras,

Fields Institute, Toronto, CA, September, 17-21 2007

Plan

- I. Introduction
  Motivations for the study of the largest eigenvalues of sample covariance matrices.

- II. A review of known results.

- III. Universality results:
  - Entries with less than four moments.
  - Entries with enough moments: Tracy-Widom's universal limit.
    * a review of Soshnikov's moment approach
    * idea of the proof.
  - Non white ensembles and possible extensions.

- IV. Concluding remarks.

Sample covariance matrices : basic model

Let μ (resp.μ') be a probability distribution on $\mathbb{C}$ (resp. on $\mathbb{R}$).

- Let X be a $N \times p$ $(p = p(N) \geq N)$ random matrix with i.i.d. entries of distribution μ

$$M_N = \frac{1}{N} X X^*,$$

$$M_p = \frac{1}{p} X^* X.$$

  Archetypical ensemble : Wishart ensemble $\mu = \mathcal{N}(0,1)$ (complex or real).

- Σ a $N \times N$ deterministic (diagonal) matrix, $Y := \Sigma^{1/2} X$

$$M_N = \frac{1}{N} Y Y^*.$$

Question: behavior of extreme eigenvalues of such random matrices as $N, p \to \infty$?

Motivations

Large scale principal component analysis used in:

-image analysis, signal processing, functionnal data analysis (data for each is a curve),

-quantitative finance: Markowitz's portfolio analysis. Minimize the risk of a large portfolio (risk measured via the variance of the returns).

-biology (very large samples of random variables with compact support)

3 regimes:
-N fixed, p large: CLT
-N and p large, $p/N \to \gamma$: finance, climate studies
-N and p large, $p/N \to \infty$: very high dimensional data sets, genetics.

Example of problems: Homogeneity test, $H_o : \Sigma = Id$ vs. $H_a = \Sigma \neq Id$.

Sample covariance matrices: global behavior

$X_{ij}, i \leq j$, i.i.d. with distribution μ such that

$$\int x d\mu = 0, \int |x|^2 d\mu = \sigma^2.$$

$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ the eigenvalues of $M_N = \frac{1}{N} X X^*$, $\mu_N = \frac{1}{N} \sum_{i=1}^{N} \delta_{\lambda_i}$.

Theorem Marchenko-Pastur (67):
A.s. if $\lim p/N = \gamma \geq 1$,

$$\lim_{N \to \infty} \mu_N = \rho_{MP}, \text{ with } \frac{d\rho_{MP}}{du} = \frac{1}{2\pi u \sigma^2} \sqrt{(u_+ - u)(u - u_-)},$$

$$u_\pm = \sigma^2 (1 \pm \sqrt{\gamma})^2.$$

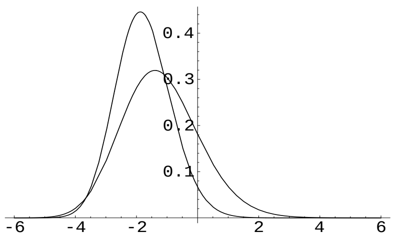
Holds for both real symmetric and complex Hermitian matrices.
Method: resolvent approach.
Behavior of extreme eigenvalues?

The largest eigenvalue

A.s. limit : Bai-Yin (87), Geman (80) Bai-Yin-Krishnaiah (88)
If $\int |x|^4 d\mu(x) < \infty$, then $\lim_{N\to\infty} \lambda_1 = u_+$ a.s.

Fluctuations : Tracy-Widom (94) Johnstone (2001) Johansson (2000)
λ_1 largest eigenvalue of the complex (resp. real) Wishart ensemble, i.e. $\mu = \mathcal{N}(0, \sigma^2)$,

$$\lim_{N\to\infty} P\Big(N^{2/3}(\sqrt{\gamma}+1)^{4/3}\gamma^{-1/6}(\lambda_1 - u_+) \leq x\Big) = F^{TW}_{2(1)}(x),$$ Tracy Widom distribution.



$\beta = 2$, $F_2^{TW}(x) = \det(I - A_x)$, A_x operator on $L^2(x,\infty)$ with the so-called Airy kernel

$$Ai(u,v) = \frac{Ai(u)Ai'(v) - Ai'(u)Ai(v)}{u - v}.$$

The case where $\lim p/N = \infty$

Theorem El Karoui (2005)
Both p and N go to infinity, $p/N \to \infty$. Complex and real Wishart ensembles. Set

$$\mu_{Np} = \left(\sqrt{N} + \sqrt{p}\right)^2, \quad \sigma_{NP} = \left(\sqrt{N} + \sqrt{p}\right)\left(\frac{1}{\sqrt{N}} + \frac{1}{\sqrt{p}}\right)^{1/3}.$$

Then,

$$\frac{N\lambda_1 - \mu_{NP}}{\sigma_{NP}} \xrightarrow{d} F_{2(1)}^{TW}.$$

Method: correlation functions and uniform asymptotic expansions of Laguerre orthogonal polynomials.
Relies on the fact that the j.e.d. can be explicitly computed.

Conjecture

As soon as μ admits moments up to order 4,

the distribution of the largest eigenvalue of M_N

should be Tracy-Widom (real or complex depending on the symmetry of M_N).

Wigner case: A. Soshnikov (2006)

Assume that $1 - F(x) = P(|X_{ij}| \geq x) = L(x)x^{-\alpha}$ where $0 < \alpha < 2$ and L is a slowly varying function, i.e., for all $t > 0$

$$\lim_{x\to\infty} \frac{L(tx)}{L(x)} = 1.$$

Set $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N$ to be the eigenvalues of a Hermitian Wigner random matrix $X = (X_{ij})_{i,j=1}^N$, and $b_N = \inf\{x : 1 - F(x) \leq \frac{2}{N(N+1)}\}$.
Define $\mathcal{P}_N = \sum \delta_{b_N^{-1}\lambda_i} 1_{\lambda_i > 0}$.

Theorem Soshnikov (2006)
The random point process $\mathcal{P}_N$ converges in distribution to the Poisson Point Process $\mathcal{P}$ defined on $(0, \infty)$ with intensity $\rho(x) = \frac{\alpha}{x^{1+\alpha}}$. In particular,

$$\lim_{N\to\infty} \mathbb{P}(\frac{1}{b_N}\lambda_1 \leq x) = \exp(-x^{-\alpha}).$$

Sample covariance matrices.

$$M = XX^* \quad p/N \to 1 \le \gamma < \infty.$$

Set $\lambda_1 \ge \ldots \ge \lambda_N$ to be the ordered eigenvalues of M and

$$\mathcal{P}_N = \sum_i \delta_{b_{Np}^{-2}\lambda_i}.$$

Theorem Joint work with A. Auffinger and G. Ben Arous (forthcoming paper).

Assume that $0 < \alpha < 4$ and $\mathbb{E}(X_{ij}) = 0$ if $2 \le \alpha < 4$.
The random point process $\mathcal{P}_N$ converges in distribution, as N goes to infinity, to the Poisson Point Process $\mathcal{P}$ defined on $(0, \infty)$ with intensity $\rho(x) = \frac{\alpha}{2x^{1+\alpha/2}}$.

Thus,

$$\lim_{N\to\infty} \mathbb{P}(\frac{1}{b_{Np}^2}\lambda_1 \le x) = \exp(-x^{-\frac{\alpha}{2}}).$$

Idea of the proof.

The largest eigenvalues are determined by the largest entries of the random matrix.

In the case where $0 < \alpha < 2$ the largest entries are large enough.
In the case where $2 \leq \alpha < 4$, use a trick due to Biroli, Bouchaud, and Potters (2006):

$$X = \left(X_{ij} 1_{|X_{ij}| \leq N^\beta}\right) + \left(X_{ij} 1_{|X_{ij}| > N^\beta}\right),$$

for some well-chosen β.

One can then bound the spectral norm of $\left(X_{ij} 1_{|X_{ij}| \leq N^\beta}\right)$ using standard tools of random matrix theory and then study the spectral radius of the other matrix (sparse enough).

Tracy-Widom universality results

Non Gaussian samples. Soshnikov (2001):
μ non Gaussian symmetric distribution with sub-Gaussian tails

$$\exists C > 0,\ \forall k > 0,\ \int |x|^{2k} d\mu(x) \leq (Ck)^k \text{ and } \int |x|^2 d\mu(x) = \sigma^2\ (\star).$$

If $p - N = O(N^{1/3})$, then

$$\frac{N\lambda_1 - \mu_{NP}}{\sigma_{NP}} \xrightarrow{d} F_{2(1)}^{TW}, \text{ Tracy Widom distribution.}$$

First universality result which does not make use of the j.e.d: no invariance assumption.
A universality result is proved for any fixed number of largest eigenvalues.
Idea of the proof: the eigenvalues of M_N behave as the squares of the eigenvalues of a Wigner random matrix if $\gamma = 1$.

Universality results for any $\gamma \in [0, \infty]$

Assume that μ is a symmetric distribution with sub-Gaussian tails and assume that $\lim p/N \in [0, \infty]$.

Then

$$\frac{N\lambda_1 - \mu_{NP}}{\sigma_{NP}} \xrightarrow{d} F_{2(1)}^{TW}.$$

Universality holds for any fixed number of largest eigenvalues.

Remark: The moment condition can be relaxed to the assumption

$$P(|X_{ij}| \geq x) \leq C(1+x)^p, \text{ for some } p > 36.$$

(Truncation techniques initially developped by Ruzmaikina).

Soshnikov's method: Tracy-Widom universality results

- Complex Wishart ensemble (or LUE): $\lambda_1 = u_+ + C\xi N^{-2/3}$ with $\xi \sim F_2^{TW}$

- If one computes

$$m_k^N(t_1, \ldots, t_k) = \mathbb{E} \prod_{i=1}^{k} Tr \left(\frac{M_N}{u_+} \right)^{[t_i N^{2/3}]},$$

  for any k, one should find something like the Laplace transform of the joint distribution of largest eigenvalues.

- Instead of computing the asymptotics of m_k^N, show that

$$|m_k^N(t_1, \ldots, t_k) - m_k^N(LUE)(t_1, \ldots, t_k)| = o(1).$$

- One can then deduce that the joint distribution of the largest eigenvalues of L_N exhibit Tracy-Widom fluctuations.

A review of the moment approach: Wigner random matrices

Let $H_N = \frac{1}{\sqrt{N}}(H_{ij})$ be a $N \times N$ Hermitian random matrix with i.i.d. entries H_{ij} (modulo the symmetry condition) with distribution μ.

$$\mathbb{E}\left[N^{s_N/2}\mathrm{Tr}H_N^{2s_N}\right] = \sum_{i_o,\dots i_{2s_N-1}} \mathbb{E}H_{i_o i_1}\cdots H_{i_{2s_N-1}i_o}(\star\star).$$

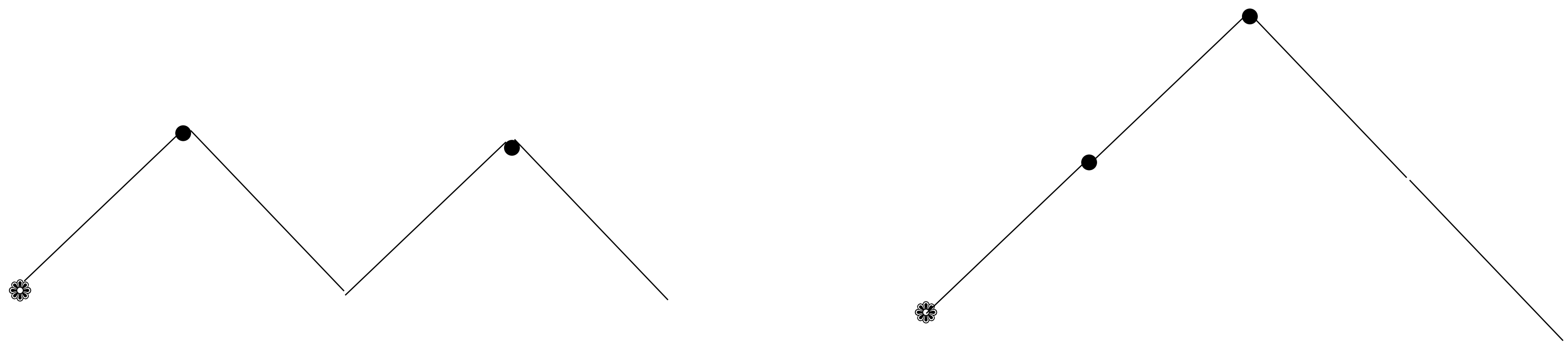
Consider the sequence of edges $(i_o i_1)\cdots(i_j i_{j+1})\cdots(i_{s_N-1}i_o)$.
Due to symmetry, independence and zero mean assumption, each non oriented "edge" (vv') is seen an even number of times.
To each term in $(\star\star)$, we associate:
- a path $i_o i_1 \cdots i_{2s_N-1} i_o$
- a trajectory $x(t), 0 \leq t \leq 2s_N$ starting at the origin and making $\pm$ steps. If at the instant t, the edge we see has been read for an odd number of times, then $+$ step $(1,1)$ and $-$ step $(1,-1)$ otherwise. This defines a Dyck path i.e. a trajectory in the positive quadrant of length $2s_N$ and ending at level 0.

Example

Consider

$$\mathbb{E}(\mathrm{Tr} H_N^4) = \frac{1}{N^2}\mathbb{E}\left(\sum_{i_o,i_1,i_2} H_{i_oi_1}H_{i_1i_o}H_{i_oi_2}H_{i_2i_o} + \sum_{i_o,i_1,i_2\neq i_o} H_{i_oi_1}H_{i_1i_2}H_{i_2i_1}H_{i_1i_o}\right).$$

Two possible trajectories:



The number of trajectories of length $2s_N$ is $\frac{(2s_N)!}{s_N!(s_N+1)!}$.

Marked instants: right endpoint of an up edge.

Same as the classical proof of Wigner's theorem.

Paths

Given a trajectory $x(t)$, assign labels chosen amongst $\{1, \ldots, N\}$

- choose the origin i_o and vertices at marked instants $\sim N^{s_N+1}$ choices,
- then "close" the edges by assigning vertices at non-marked instants.

Wigner's regime: choose the marked vertices and origin pairwise distinct. No choice to close the edges.
Largest eigenvalue: $s_N \sim N^{2/3}$: repeat some marked vertices. This decreases the number of labels of a factor N but s_N^2 moments where a label occurs twice e.g.

Self intersection of type i: v occurs i times as a marked vertex.

Problem: which trajectories are typical? which paths are typical?

More insight

Assume that the trajectory is known as well as $N_i, i = 0, \ldots, s_N$ the number of vertices of type i. Then, one has $\sum_{i=0}^{s_N} N_i = N, \quad \sum_{i=0}^{s_N} iN_i = s_N$.

The number of ways to choose the vertices occuring in the path is $\frac{N!}{\prod_{i=0}^{s_N} N_i!}$.

The number of ways to distribute the marked vertices along the path is $\frac{s_N!}{\prod_{i\geq 2}(i!)^{N_i}}$.

The number of ways to close the edges is at most

$$\prod_{i=2}^{s_N} (2i)^{iN_i}.$$

The expectation of the path is at most of order

$$\sigma^{2s_N} \prod_{i\geq 2} (Ci)^{iN_i},$$

due to the sub-Gaussian tail assumption.

More insight 2

Thus summing over the N_i's

$$\mathbb{E}\mathrm{Tr}\left(H_N^{2s_N}\right) \leq N\sigma^{2s_N}\frac{(2s_N)!}{s_N!(s_N+1)!}\sum_{N_i, i\geq 2}\prod_{i\geq 3}\frac{1}{N_i!}\left(\frac{Cs_N^i}{N^{i-1}}\right)^{N_i}.$$

In the scale $s_N \sim N^{2/3}$, and improving the above upper bound, vertices of type 2 may give a non-trivial contribution.

The actual number of pairwise distinct vertices in the path is $N^{s_N+1}\exp\{-\frac{s_N^2}{N}\}$.

Vertices of type 2:

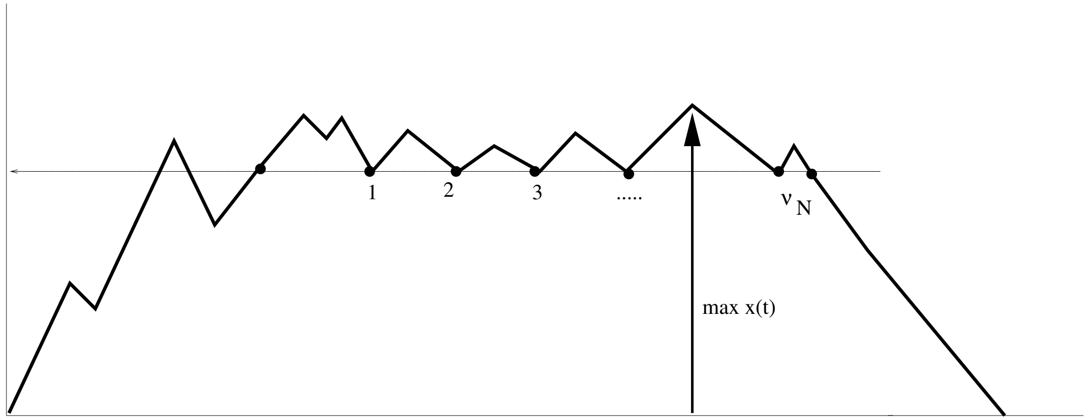
- more than one possible choice to close an edge starting from a vertex of type 2,
- may have edges seen four times.

Their contribution is bounded via typical trajectories.

Show that the contribution of paths with edges seen at least four times is $o(1)$ and that the expectation is bounded.

Typical paths and typical trajectories

Typical trajectories:



$$\max x(t) \sim \sqrt{s_N}, \quad \nu_N << \sqrt{s_N}.$$

Typical paths: The typical number of vertices of type i is $\left(\frac{s_N}{N}\right)^i N$.

In the scale $s_N \sim N^{2/3}$, there are self-intersections of type 3 at most.

Each edge is read at most twice in typical paths

This implies universality.

In the scale $N^{2/3}$ there are multiple choices to close the edges (GOE or GUE TW).

Combinatorics for sample covariance matrices

Developping the trace

$$\mathbb{E}\mathrm{Tr} M_N^{s_N} = \frac{1}{N^{s_N}} \sum_{i_o,\ldots,i_{s_N-1}} \sum_{j_o,\ldots,j_{s_N-1}} \mathbb{E}\left(X_{i_o j_o} \overline{X_{i_1 j_o}} \cdots X_{i_{s_N-1} j_{s_N-1}} \overline{X_{i_o j_{s_N-1}}} \right).$$

What matters:

-number of odd and even marked instants (up steps) in Dyck paths. Indeed, p choices for labels instead of N.

-need to consider oriented edges.

We associate the sequence of oriented edges:

$$\binom{j_o}{i_o}\binom{j_o}{i_1}\binom{j_1}{i_1}\cdots\binom{j_{s_N-1}}{i_{s_N-1}}\binom{j_{s_N-1}}{i_o}$$

Define marked instants as before except that edges are oriented and read from bottom to top. We still get a Dyck trajectory.

Narayana numbers

Let $1 \leq k \leq s_N$.

$$\mathbf{N}(s_N, k) = \frac{1}{s_N} C_{s_N}^k C_{s_N}^{k-1}$$

Narayana number counts the number of Dyck trajectories of length $2s_N$ with k odd up steps.

Jonsson (1982) Bai (1999). Connection with Marchenko-Pastur distribution

$$\sigma^{2l} \sum_{k=1}^{l} N(l,k)\gamma^k = \lim_{N\to\infty} \frac{1}{N} \mathbb{E}\mathrm{Tr} \left(\frac{XX^*}{N}\right)^l, \text{ if } \lim_{N\to\infty} \frac{p}{N} = \gamma.$$

Consider self-intersections on the top and bottom line separately. Define

$$n_i := \sharp\{ \text{ vertices occuring } i \text{ times as a marked instant on the bottom line}\}, i \leq s_N - k,$$

$$p_i := \sharp\{ \text{ vertices occuring } i \text{ times as a marked instant on the top line}\}, i \leq k.$$

Same statistics as for Wigner case.

Mutatis mutandis

$$\text{choice of vertices: } \frac{N!}{\prod_{i=0}^{s_N} N_i!} \to \frac{N!}{\prod_{i=0}^{s_N-k} n_i!} \frac{p!}{\prod_{i=0}^{k} p_i!}.$$

$$\text{self-intersections: } \frac{s_N!}{\prod_{i\geq 2}(i!)^{N_i}} \to \frac{(s_N-k)!}{\prod_{i\geq 2}(i!)^{n_i}} \frac{k!}{\prod_{i\geq 2}(i!)^{p_i}}.$$

$$\text{closing the path: } \prod_{i=2}^{s_N}(2i)^{iN_i} \to \prod_{i=2}^{s_N-k}(2i)^{in_i} \prod_{i=2}^{k}(2i)^{ip_i}.$$

$$\text{expectation: } \sigma^{2s_N} \prod_{i\geq 2}(Ci)^{iN_i} \to \sigma^{2s_N} \prod_{i\geq 2}(Ci)^{in_i} \prod_{i\geq 2}(Ci)^{ip_i}.$$

Typical trajectories have $k \sim \frac{\sqrt{\gamma_N}}{1+\sqrt{\gamma_N}} s_N$ ($\gamma_N = p/N$) odd marked instants.

Same statistics as for Wigner case: $\nu_N^{odd(even)}$ $\sharp$ returns to some level at odd or even instants. Same typical behavior as for Hermitian ensembles (modulo $s_N \mapsto k, s_N - k$).

More complex covariance structure

Extensions Spiked models: $\Sigma = \text{diag}(\pi_1, \pi_2, \ldots, \pi_r, 1, \ldots, 1)$ be a fixed rank perturbation of the identity matrix.
Baik-Ben Arous-P (2004), D. Paul (2005), Baik-Silverstein (2005)

Complex Wishart ensembles: $\pi_1 > \pi_2 \geq \cdots \geq \pi_r$.

If $\pi_1 < 1 + \frac{1}{\sqrt{\gamma}}$ Tracy-Widom fluctuations.

If $\pi_1 > 1 + \frac{1}{\sqrt{\gamma}}$, $\lambda_1 = \gamma\pi_1 + \frac{\pi_1}{\pi_1 - 1} + \frac{1}{\sqrt{N}}G$ where G is a Gaussian random variable.

Test: $H_o : \Sigma = Id$ vs. $H_a : \Sigma \neq Id$
Problem: the test based on the largest eigenvalue will not detect π_1 if it is “too smal”.

Extensions to non white ensembles

Let $\Sigma = \text{diag}(\pi_1, \pi_2, \ldots, \pi_r, 1, \ldots, 1)$ be a fixed rank perturbation of the identity matrix.

$$\mathbb{E}\text{Tr}(\frac{1}{N}YY^*)^{s_N} = \frac{1}{N^{s_N}} \sum_{i_o,\ldots,i_{s_N-1}} \sum_{j_1,\ldots,j_{s_N}} \mathbb{E}X_{i_o j_1}\overline{X_{i_1 j_1}} \cdots X_{i_{s_N-1} j_{s_N}} \overline{X_{i_o j_{s_N}}} \prod_{i=1}^{r} \pi_i^{r_i},$$

where r_i is the number of times the vertex i appears on the bottom line.

Should be possible to count the typical number of such occurences (forthcoming paper with D. Féral).

If Σ is more general, then the combinatorial idea is unclear.
Another combinatorial approach: Anderson-Zeitouni (2006).

Conclusion:

- Statistical implications.
  - The symmetry assumption is expected to be unnecessary to obtain TW.
  - Usual sample covariance matrices: $(X - \overline{X})(X - \overline{X})^*$: harder.
  - One issue:

$$\mu_{Np} \to \left(\sqrt{N+a} + \sqrt{p+b}\right)^2 ,$$

$$\sigma_{NP} \to \left(\sqrt{N+a} + \sqrt{p+b}\right)\left(\frac{1}{\sqrt{N+a}} + \frac{1}{\sqrt{p+b}}\right)^{1/3} .$$

  Result still true, but the choice of a and b for simulations is crucial.

- Tracy-Widom universality when entries have less than $m_o = 36$ moments. Transition at $m_o = 4$? Limiting distribution?