DOI 10.4171/8ECM/34

Bernoulli random matrices

Alice Guionnet

Abstract. Random matrix theory has become a field on its own with a breadth of new results, techniques, and ideas in the last thirty years. In these proceedings, I illustrate some of these advances by describing what we now know about the spectrum and the eigenvectors of Bernoulli matrices.

1. Introduction

Jacques (or Jakob) Bernoulli (1654–1705) was a renowned Swiss mathematician who made important contributions to probability theory and partial differential equations. He was the first to discover the number e. But his most famous result is, at least for probabilists, the first proof of the law of large numbers. To this end, he analyzed the concept of the Bernoulli law, which is the simplest non-trivial distribution you can think of, being the sum of two Dirac masses. It is the distribution of a random variable b which can only take two values 0 and 1. We denote

$$p = \mathbb{P}(b = 1) = 1 - \mathbb{P}(b = 0).$$

A very common example is a coin that, once thrown, falls either on head (modeled by the state 1) or tail (modeled by 0). Even if one would expect in general the probability of each event to be equal to $1/2$, it may well be rather $p \in (0, 1)$ if the coin is rigged. In Ars Conjectandi, Bernoulli showed that if one throws such a coin independently a number n of times, then, with large probability, one should see approximately pn heads if n is large enough. To state this law of large numbers more precisely, he showed that if $b_1, \ldots, b_n$ denotes the outcome of n-independent Bernoulli trials, then for any $a < p < b$

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} b_i \in [a, b]\right) = 1.$$

2020 Mathematics Subject Classification. Primary 60B20; Secondary 05C80.
Keywords. Random matrices, random graphs.

But how close can we choose a, b to p so that this result remains true? Few years later, A. de Moivre (1667–1754) quantified the size of the error and proved the first central limit theorem, namely that a, b can be at a distance of about $1/\sqrt{n}$ of p in the sense that

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{\sqrt{np(1-p)}}\sum_{i=1}^{n}(b_i - p) \in [a,b]\right) = \frac{1}{\sqrt{2\pi}}\int_a^b e^{-\frac{x^2}{2}}\,dx.$$

This was the first occurrence of the central limit theorem and the start of modern probability theory and statistics. Implicitly, we so far assumed that p does not depend on n and belongs to (0, 1). Later on, we shall also be interested in the case where p depends on n. Then, it can be checked that the central limit theorem still holds as long as pn goes to infinity. If pn goes to a finite constant c, then it cannot hold since $\sum_{i=1}^{n} b_i$ is an integer so that the above random variable is discrete. In fact, it converges towards the Poisson distribution

$$\lim_{n\to\infty} \mathbb{P}\left(\frac{1}{\sqrt{np(1-p)}}\sum_{i=1}^{n}(b_i - p) \in [a,b]\right) = \sum_{k\in c+\sqrt{c}[a,b]} \frac{1}{k!}c^k e^{-c}.$$

We will see later that this transition between such continuous and discrete limits is also key to describing the spectrum of Bernoulli random matrices. The last concept which is central in probability theory and important in these notes is entropy. It was introduced by Ludwig Boltzmann (1844–1906) and Claude Shannon (1916–2001) in physics and information theory, respectively, as a way to measure disorder. For again n-independent Bernoulli trials with parameter p, it is defined for any $q \in [0, 1]$ by

$$\lim_{\varepsilon\downarrow 0}\lim_{n\to\infty}\frac{1}{n}\ln \mathbb{P}\left(\frac{1}{n}\sum_{i=1}^{n} b_i \in [q-\varepsilon, q+\varepsilon]\right) = -S_p(q),$$

where $S_p(q) = \frac{q}{p}\ln\frac{q}{p} + \frac{1-q}{1-p}\ln\frac{1-q}{1-p}$ is the entropy or rate function.

In this survey, I will discuss Bernoulli random matrices. A Bernoulli random matrix is an $n \times n$ symmetric matrix with independent Bernoulli entries (modulo the symmetry constraint) whose size n is going to infinity. I will discuss the law of large numbers, the fluctuations, and the entropy for their spectrum and eigenvectors. There are many motivations to study random matrices. The first goes back to Wishart who considered random matrices to study correlations in large data sets. Such questions are very modern, with the need to analyze larger and larger data sets and machine learning. The second comes from physics and works of Wigner and Dyson. They proposed to model the Hamiltonian of excited nuclei by random matrices, an idea which turned out to be quite successful as indeed real nuclei turned out to have energy levels distributed like the eigenvalues of random matrices. But Bernoulli matrices

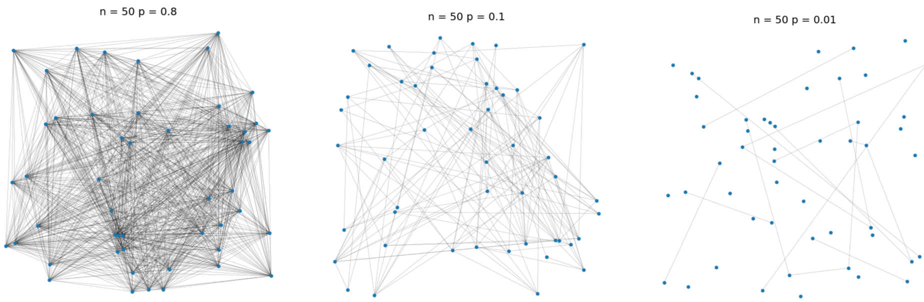

Figure 1. Courtesy of D. Coulette.

are special among all other random matrices because they describe the adjacency matrix of an Erdős–Rényi graph $G(n, p)$. Indeed, the latter is just a graph built on n (labeled) vertices, with an edge drawn independently between each couple of vertices with probability p. Studying the eigenvalues of the adjacency matrix of a graph gives valuable geometric information, such as the size of its boundary (expanders) or the number of specific configurations, such as triangles, that it contains. One can also be interested in the combinatorial properties of such matrices and for instance focus on the probability that the matrix is singular; see e.g. [68]. My viewpoint will be to investigate the properties of the eigenvalues and eigenvectors of Bernoulli random matrices, as a particularly nice and well-documented example of random matrices.

To simplify, I will restrict myself to symmetric Bernoulli matrices $\mathbf{B}_n$ throughout these notes:

$$\mathbf{B}_n(i, j) = \mathbf{B}_n(j, i),$$

and assume that $(B_n(i, j),\ i \le j)$ follows a Bernoulli law with parameter p. Also, I will take $\mathbf{B}_n(i, i)$ random, but could take it equal to zero without changing much the statements of most of the results.

My goal is to understand the spectrum of $\mathbf{B}_n$ as well as the properties of its eigenvectors as n goes to infinity. One can easily guess that these properties should depend on the parameter p. Indeed, thinking about the Erdős–Rényi graph, one sees that the average degree of a vertex is pn. The graph will be very dense if pn goes to infinity fast enough but sparse if it is finite.

Indeed, it is well known since the breakthrough paper of Erdős and Rényi (see Figure 1) that if $np < 1$, $G(n, p)$ will almost surely have no connected component of size greater than $O(\ln n)$; if $np = 1$, there is a giant connected component but it is of size of order $n^{2/3}$; if np goes to a constant $c > 1$, it will have a unique giant component but lots of small components, and isolated vertices will continue to exist until $np < (1 - \varepsilon) \ln n$; whereas if $np > (1 + \varepsilon) \ln n$ the graph will almost surely be connected. Here ε is some positive real number as small as wished. In the case where

np is of order c, the finite size connected components will create small diagonal blocks in the Bernoulli matrix, with entries equal either to zero or one and therefore finitely many possible eigenvalues. Hence, we expect the spectrum to accumulate at these possible values. But should there be other possible eigenvalues? Similarly, we see that the eigenvectors related with these eigenvalues are localized on a few vertices. But should we also have delocalized eigenvectors? On the contrary, in the case where $np > (1+\varepsilon)\ln n$, we may expect eigenvectors to be delocalized and the spectrum to be nicely continuous. In this case, a whole theory has been developed to show that the spectrum and the eigenvalues of Bernoulli matrices have the same properties as those of a random matrix with Gaussian entries. The latter is well known to be much easier to study, for instance, because the joint law of its eigenvalues is rather simple and independent of the eigenvectors. Conversely, Bernoulli matrices resemble more heavy-tailed matrices when pn is of order one, in the sense that it has mostly very tiny entries but a few large entries. Understanding the transition between these two behaviors is at the heart of random matrix theory.

In this survey, I will start discussing the asymptotic behavior of the spectrum in both sparse and dense cases. Then, I will consider its fluctuations, both local and global, as well as the properties of its eigenvectors. Finally, I will discuss the large deviations of the spectrum, for instance how to estimate the probability that the second eigenvalue of Bernoulli matrices takes an unexpected value.

2. Law of large numbers

In this section, we shall see that the limiting distribution of the spectrum differs a lot according to whether pn goes to infinity or not.

A first remark should be made about the matrix $\mathbf{B}_n$: its entries are not centered. It will be more convenient to center them and renormalize the matrix properly. To this end, we make the decomposition

$$\mathbf{B}_n = \sqrt{np(1-p)}\mathbf{X}_n + p\mathbb{1},$$

where $\mathbb{1}$ is a matrix whose entries are all equal to one, whereas the entries of $\mathbf{X}_n$ are centered and renormalized to have covariance $1/n$:

$$\mathbf{X}_n(i,j) = \frac{\mathbf{B}_n(i,j) - p}{\sqrt{np(1-p)}}.$$

The matrix $\mathbb{1}$ has one non-trivial eigenvalue which equals n, and flat eigenvector

$$\mathbf{1} = (1/\sqrt{n}, 1/\sqrt{n}, \ldots, 1/\sqrt{n}).$$

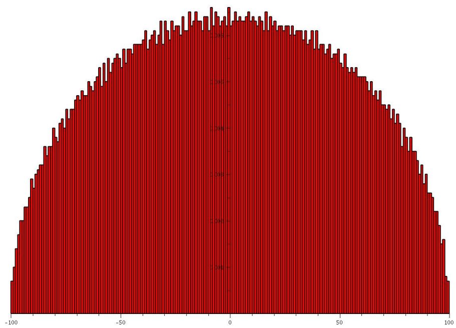

Figure 2.

Conversely, the spectrum of $\mathbf{X}_n$ has eigenvalues mostly of order one in the sense that $\mathbb{E}[\mathrm{Tr}(\mathbf{X}_n^2)] = \mathbb{E}[\sum \lambda_i^2] = n$. Therefore, the above decomposition shows that $\mathbf{B}_n$ has a very large eigenvalue of order n, and the rest is roughly given by the eigenvalues of $\mathbf{X}_n$ taken on $\mathbf{1}^\perp$. Moreover, by Weyl's interlacing properties, the eigenvalues $(\lambda_i^B)_{1\le i\le n}$ of $\mathbf{B}_n/\sqrt{np(1-p)}$ and $(\lambda_i^X)_{1\le i\le n}$ of $\mathbf{X}_n$ are interlaced:

$$\lambda_n^X \le \lambda_n^B \le \lambda_{n-1}^X \cdots \le \lambda_1^X \le \lambda_1^B.$$

Therefore, it is in general not difficult to retrieve the properties of the eigenvalues of $\mathbf{B}_n/\sqrt{np(1-p)}$ from those of $\mathbf{X}_n$. Hereafter, we will therefore concentrate mostly on $\mathbf{X}_n$.

2.1. Dense case

The first result describes the asymptotic distribution of the spectrum in the dense case and shows that the limit is described by the famous semi-circle law; see Figure 2.

Theorem 2.1. *Assume that pn goes to infinity as n goes to infinity. Then, almost surely, for any $a < b$*

$$\lim_{n\to\infty} \frac{1}{n}\#\big\{i : \lambda_i^B \in \sqrt{np(1-p)}[a,b]\big\} = \lim_{n\to\infty} \frac{1}{n}\#\big\{i : \lambda_i^X \in [a,b]\big\} = \sigma\big([a,b]\big),$$

where σ is the semi-circle law given by

$$\sigma(dx) = \frac{1}{2\pi}\sqrt{4-x^2}\,1_{|x|\le 2}\,dx. \tag{2.1}$$

The semi-circle law is ubiquitous to random matrix theory as it describes the asymptotic behavior of random matrices with Gaussian entries, but in fact any random matrix with independent centered entries $(a_{ij})_{i,j}$ such that $\mathbb{E}[|\sqrt{n}a_{ij}|^{2+\varepsilon}]$ is

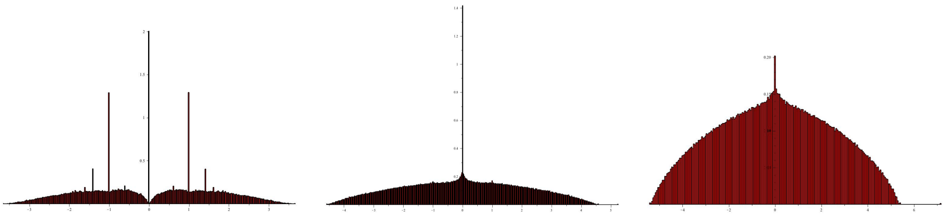

Figure 3. Simulation for $c = 1, 2, 3$ (courtesy of J. Salez).

uniformly bounded for some $\varepsilon > 0$. Such a convergence was proved first by Wigner in the case where p is independent of n based on the computation of the moments $\mathbb{E}[\operatorname{Tr} \mathbf{X}_n^k]$. Indeed, one can expand the trace of moments of matrices in terms of the entries, and observe that the indices which contribute to the first order of this expansion can be described by rooted trees, whereas $\sigma(x^k)$ is equal to the Catalan numbers which enumerate them.

2.2. Sparse case

On the other hand, the limiting distribution of the spectrum is very different when pn is of order one. Namely, we have the following theorem; see [52, 70].

Theorem 2.2. *Assume that pn goes to $c \in (0, +\infty)$ as n goes to infinity. Then, almost surely, for any $a < b$*

$$\lim_{n\to\infty} \frac{1}{n} \#\{i : \lambda_i^B \in \sqrt{np(1-p)}[a,b]\} = \lim_{n\to\infty} \frac{1}{n} \#\{i : \lambda_i^X \in [a,b]\} = \mu_c([a,b]).$$

The limit law μ_c depends on c; some plots are shown in Figure 3.

The simulations indicate the presence of atoms. They were shown to be exactly given by totally real algebraic integers in [58] for all $c > 0$; these are the roots of monic polynomials with integer coefficients. It is easy to understand that the atoms should be totally algebraic integers as finite connected components are diagonal blocks with 0 or 1 entries whose characteristic polynomials have such roots. It is a much stronger statement to show that all such roots are atoms, in particular since totally algebraic integers are dense in the real numbers. μ_c has also a continuous spectrum: it was indeed proved in [30] that μ_c has a non-trivial continuous part if and only if $c > 1$. This result is in fact hard to prove as the limit laws μ_c's are described as the solution of complicated equations [28]; see also [17, 20]. However, such description could be used in [8] to prove the existence of an absolutely continuous part for sufficiently large c. Moreover, the first-order expansion of μ_c in c going to infinity was derived in [38]. The spectrum at the origin seems to have a Dirac mass whose weight could be computed [29].

2.3. Idea of the proof

The first proof of Theorem 2.1 estimated the moments $\frac{1}{n}\operatorname{Tr}(\mathbf{X}_n)^k$ for all integer numbers k; see [69] for the first theorem and [17, 52, 70] for the sparse case. However, in order to go into more local results like the behavior of the eigenvectors or the local fluctuations, and as well to have more explicit formulas for the limit law, it is more convenient to study the resolvent. This path can be used to study the asymptotics of the spectral measure of any self-adjoint matrix $\mathbf{X}_n$ with independent entries modulo the symmetry constraint, and was generalized to study heavy-tailed matrices in [17, 20, 52] based on the ideas from [35]. The idea is to derive the asymptotics of the Stieltjes transform

$$G_n(z) = \frac{1}{n}\operatorname{Tr}(z - \mathbf{X}_n)^{-1} = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{z - \lambda_i^X}$$

for a complex number z away from the real line. To this end, we use the Schur complement formula which reads

$$(z - \mathbf{X}_n)^{-1}_{ii} = \frac{1}{z - X_{ii} - \langle X_i, (z - \mathbf{X}^{(i)})^{-1} X_i\rangle}, \tag{2.2}$$

where $X_i = (X_{ij})_{j\neq i}$ and $\mathbf{X}^{(i)}$ is the associated principal minor, namely the $(N-1)\times(N-1)$ matrix obtained from $\mathbf{X}_n$ by removing the ith row and column. X_{ii} goes to zero with N and we can check (e.g. by estimating the L^2 norm of the difference) that with probability going to one

$$\langle X_i, (z - \mathbf{X}^{(i)})^{-1} X_i\rangle = \sum_{j:j\neq i} X_{ij}^2 (z - \mathbf{X}^{(i)})^{-1}_{jj} + o(1). \tag{2.3}$$

This is where the "light tail" hypothesis pn going to infinity starts to matter. Then, the entries X_{ij}^2 go to zero and have variance $1/n$ so that, since the X_{ij} are independent of $\mathbf{X}^{(i)}$, the law of large numbers (or a second moment computation) asserts that with probability going to one

$$\sum_{j:j\neq i} X_{ij}^2 (z-\mathbf{X}^{(i)})^{-1}_{jj} = \sum_{j:j\neq i}\mathbb{E}[X_{ij}^2](z-\mathbf{X}^{(i)})^{-1}_{jj} + o(1) = \frac{1}{n}\sum_{j:j\neq i}(z - \mathbf{X}^{(i)})^{-1}_{jj} + o(1).$$

But again $\mathbf{X}^{(i)}$ and $\mathbf{X}_n$ vary only by a rank two matrix (if we complete $\mathbf{X}^{(i)}$ by zero entries at the ith row and column), so that their spectrum is interlaced by Weyl's interlacing property. As a consequence

$$\frac{1}{n}\sum_{j\neq i}(z - \mathbf{X}^{(i)})^{-1}_{jj} = \frac{1}{n}\sum_{i}(z - \mathbf{X}_n)^{-1}_{jj} + O\left(\frac{1}{\Im(z)n}\right).$$

This approximation, together with (2.2) and (2.3), implies that with high probability

$$G_n(z) = \frac{1}{n}\sum_i (z-\mathbf{X}_n)^{-1}_{jj} = \frac{1}{z-G_n(z)} + o(1). \tag{2.4}$$

After recalling that $G_n(z)$ goes to zero as N goes to infinity, we conclude that since $G_n(z)$ goes to zero as the imaginary part of z goes to infinity,

$$G_n(z) = \frac{1}{2}(z-\sqrt{z^2-4}) + o(1)$$

is approximately the Stieltjes transform of the semicircle law $G_\sigma(z)=\frac{1}{2}(z-\sqrt{z^2-4})$. Since G_n is analytic and uniformly bounded for $\Im z > \varepsilon$, Montel's theorem implies that G_n converges to this limit away from the real line, which yields the vague convergence of the empirical measure of the eigenvalues. Because $\frac{1}{n}\operatorname{Tr}(\mathbf{X}_n^2)$ is in L^1, the weak convergence follows.

On the contrary, in the heavy-tailed case where pn is of order one, the entries of X_{ij} are often very small but of order one with a positive probability. Hence, the previous law of large numbers does not hold true any more and we cannot expect such a simple equation as (2.4). In fact, $\sum_{j\neq i} X_{ij}^2 (z-\mathbf{X}^{(i)})^{-1}_{jj}$, if it converges, will a priori converge to a random variable. To study this convergence, we make the following assumption on the law μ_n of X_{ij}:

$$\lim_{n\to\infty} n\left(\int (e^{-iux^2}-1)\,d\mu_n(x)\right) = \Phi(u) \tag{2.5}$$

with Φ such that there exists g on $\mathbb{R}^+$, with $g(y)$ bounded by Cy^κ for some $\kappa > -1$, such that for $u \in \mathbb{C}^-$,

$$\Phi(u) = \int_0^\infty g(y)e^{\frac{iy}{u}}\,dy. \tag{2.6}$$

This is satisfied by the adjacency matrix of Erdős–Rényi graph with $\Phi(u)=c(e^{iu}-1)$ if pn goes to c and g is a Bessel function [20], but also for other cases, for instance for α stable laws with $\Phi(u) = c(iu)^{\alpha/2}$ and $g(y) = Cy^{\alpha/2-1}$ for some constants c, C. Then, it was shown in [17, 20] that $G_n(z)=\frac{1}{n}\operatorname{Tr}(z-\mathbf{X}^n)^{-1}$ converges almost surely towards G given by

$$G(z) = i\int e^{itz}e^{\rho_z(t)}\,dt, \quad z\in\mathbb{C}^+, \tag{2.7}$$

where $\rho_z : \mathbb{R}^+ \to \{x+iy; x\le 0\}$ is the unique solution, analytic in $z \in \mathbb{C}^+$, of the non-linear equation

$$\rho_z(t) = \int_0^\infty g(y)e^{\frac{iy}{t}z+\rho_z(\frac{y}{t})}\,dy. \tag{2.8}$$

This entails the convergence of the spectral measure of $\mathbf{X}^n$, with σ replaced by a probability measure with Stieltjes transform given by (2.7). The argument to prove (2.7) and (2.8) is as follows. We first remark that G_n concentrates in the sense that it is close to its average; see Theorem 3.2. We let ρ^n be the order parameter $\rho_z^n(x) := \mathbb{E}[\frac{1}{n}\sum \Phi(x(z-\mathbf{X}^{(i)})_{jj}^{-1})]$. By (2.2) and (2.3), we find that, if $\Im z > 0$,

$$\begin{aligned}
G_n(z) \simeq \mathbb{E}\big[G_n(z)\big] &= -i\mathbb{E}\bigg[\int_0^\infty e^{itz-it\sum_{j\neq i} X_{ij}^2 (z-\mathbf{X}^{(i)})_{jj}^{-1}}\, dt\bigg] + o(1) \\
&= i\int_0^\infty e^{itz}\mathbb{E}\bigg[\prod_{j\neq i}\mathbb{E}[e^{-itX_{ij}^2(z-\mathbf{X}^{(i)})_{jj}^{-1}}]\, dt\bigg] + o(1) \\
&= -i\int_0^\infty e^{itz}\mathbb{E}\bigg[\prod_{j\neq i}\bigg(1+\frac{1}{n}\Phi\big(t(z-\mathbf{X}^{(i)})_{jj}^{-1}\big)\bigg)\, dt\bigg] + o(1) \\
&= i\int_0^\infty e^{itz+\rho_z^n(t)}\, dt + o(1).
\end{aligned}$$

To conclude, we need to show the convergence of ρ^n. But ρ^n can be seen to be analytic away from the real axis, and uniformly bounded under our hypothesis. This is enough to see that it is tight and any limit point will be analytic by Montel theorem. Hence, it is enough to show that it has a unique limit point for z with large imaginary part. To this end, we get an equation for ρ^n which follows from (2.6) by

$$\begin{aligned}
\rho_z^n(t) &= \int_0^\infty g(y)\mathbb{E}\Big[e^{\frac{iy}{x(z-\mathbf{X}^{(i)})_{11}^{-1}}}\Big]\, dy \\
&\simeq \int_0^\infty g(y)\mathbb{E}\Big[e^{\frac{iy}{x}(z-\sum_{j\geq 2} X_{ij}^2 (z-\mathbf{X}^{(1)})_{jj}^{-1})}\Big]\, dy + o(1) \\
&\simeq \int_0^\infty g(y) e^{\frac{iy}{x}z} e^{\rho_z^n(\frac{y}{x})}\, dy + o(1),
\end{aligned}$$

where in the second line we used (2.2) and (2.3). One can conclude by proving the uniqueness of the solutions to this equation when z is far from the real line by showing that the non-linear equation is then a contraction. The above arguments were made complete in [17, 19, 20]. Another approach to heavy-tailed matrices and sparse Bernoulli matrices based on Aldous' Poisson-weighted infinite tree was proposed in [25].

2.4. Extreme eigenvalues

The asymptotic behavior of the extreme eigenvalues also depend on c: they stick to the bulk when $pn \gg \ln n$ and then go away at distance of order $\sqrt{\ln n}$. We, more precisely, have the following result, putting together the article of Benaych-Georges, Bordenave, and Knowles [18] and that of Alt, Ducatez, and Knowles [4]; see also [65].

Theorem 2.3. • *Assume that $pn/\ln n \to +\infty$. Then the largest eigenvalue of $\mathbf{X}_n$ sticks to the bulk: $\lambda_1^X \to 2$.*

- *Assume that $pn/\ln n \to 0$. Then $\lambda_1^X \simeq \sqrt{\ln n/\ln(\ln n/pn)}$.*
- *Assume that $pn \simeq C \ln n$. Then for $C > 1/(\ln 4 - 1) := C^*$ the eigenvalues stick to the bulk, whereas for $C < 1/(\ln 4 - 1)$*

$$\lambda_1^X = \frac{\alpha}{\sqrt{\alpha - 1}}, \quad \alpha = \max \frac{1}{pn} \sum_j B_{ij}.$$

Observe that $\sum_j B_{ij}$ is the degree of vertex i: the largest eigenvalue is hence created by the largest degree in the graph. In fact, in the work of Alt, Ducatez, and Knowles [4], it is shown that all eigenvalues outside the bulk are created by vertices with large degrees when $pn \leq C^* \ln n$.

3. Fluctuations

3.1. Concentration of measure

Concentration of measure has become a central tool in probability and, in particular, in random matrix theory. It allows us to prove that some quantities, such as smooth function of independent variables, are not much random. It was crucial in the previous proof of the convergence of the spectral measure. However, it generally depends on the tails of the random variables. Herbst's argument allows considering random variables with sub-Gaussian tails and more precisely random variables whose distribution satisfies log-Sobolev inequalities, which is the case for instance when their density is strictly log-concave as for Gaussian's variables. To deal with bounded variables such as the entries of Bernoulli matrices, one should rather use the theory developed by Talagrand [61]. This was done in [44], where the spectrum of random matrices was observed to be a smooth function of its entries and the associated Lipschitz norm was computed. It resulted in the following theorem [44, Theorem 1.1]. We hereafter consider a symmetric matrix $\mathbf{A}$ with independent entries above the diagonal with distribution $a_{ij}/\sqrt{n}$, where a_{ij} is distributed according to P_{ij} supported in a compact set K with width $|K|$.

Theorem 3.1. (1) *Take f convex and Lipschitz with Lipschitz norm $\|f\|_L$. Then, for any $\delta > \delta_0(n) = 8|K|\|f\|_L/n$,*

$$\mathbb{P}\left(\left|\frac{1}{n}\operatorname{Tr}\big(f(\mathbf{A})\big) - \mathbb{E}\left[\frac{1}{n}\operatorname{Tr}\big(f(\mathbf{A})\big)\right]\right| > \delta\|f\|_L\right) \leq 4\exp\left\{-n^2\frac{\big(\delta - \delta_0(n)\big)^2}{16|K|^2}\right\}.$$

(2) *There exists a finite constant* $c > 0$ *such that for any* $\delta > \delta_1(n) \simeq \sqrt{\delta_0(n)}$

$$\mathbb{P}\Bigg(\sup_{f \in \mathrm{Lip}_{\mathcal{K}}} \bigg| \frac{1}{n} \operatorname{Tr}\big(f(\mathbf{A})\big) - \mathbb{E}\bigg[\frac{1}{n} \operatorname{Tr}\big(f(\mathbf{A})\big)\bigg]\bigg| > \delta \|f\|_L \Bigg) \leq \exp\bigg\{ - n^2 \frac{\big(\delta - \delta_1(n)\big)^2}{c|K|^2} \bigg\}.$$

(3) *Let* λ_1^A *be the largest eigenvalue of* $\mathbf{A}$. *Then*

$$\mathbb{P}\big(\big|\lambda_1^A - \mathbb{E}[\lambda_1^A]\big| \geq \delta|K|\big) \leq \exp\bigg\{ - \frac{\big(\delta - 8|K|/\sqrt{n}\big)^2 n}{16} \bigg\}.$$

This result is a direct application of Talagrand's beautiful theory and the computation of Lipschitz constants of functions of the spectral measure in terms of the entries; see [6, 45]. The original statement proves concentration around the median rather than the mean, but it is easy to go from one result to the other up to some error $\delta_0(n)$, $\delta_1(n)$. The second point is deducted from the first by approximating a general function by convex functions. It applies to Bernoulli matrices straightforwardly by taking $|K| = 1/\sqrt{p(1-p)}$.

Theorem 3.2. *Take* f *convex and Lipschitz with Lipschitz norm* $\|f\|_L$. *Then, for any* $\delta > \delta_0(n) = 8\sqrt{\pi}|f|_L/np(1-p)$,

$$\mathbb{P}\Bigg(\bigg| \frac{1}{n} \operatorname{Tr}\big(f(\mathbf{X}_n)\big) - \mathbb{E}\bigg[\frac{1}{n} \operatorname{Tr}\big(f(\mathbf{X}_n)\big)\bigg]\bigg| > \delta + \delta_0(n) \Bigg) \leq \exp\bigg\{ - p(1-p)n^2 \frac{(\delta)^2}{16|f|_L^2} \bigg\}.$$

Moreover, for any $\delta > \delta_0'(n) = O(1/\sqrt{p(1-p)n})$

$$\mathbb{P}\big(\big|\lambda_1 - \mathbb{E}[\lambda_1]\big| > \delta + \delta_0'(n)\big) \leq \exp\big\{ - p(1-p)n\delta^2\big\}.$$

As we can see, the speed of the concentration deteriorates with p going to zero to be of order n when np is of order one. In fact, it can be shown that the worse concentration estimates for the empirical measure are of the order of exponential in n. Indeed, we have the following result due to Bordenave, Caputo, and Chafai [24] which is based on the Azuma–Hoeffding inequality and requires only the independence of the vectors of the random matrix.

Lemma 3.3. *Let* $\|f\|_{TV}$ *be the total variation norm:*

$$\|f\|_{TV} = \sup_{x_1 < \cdots < x_p} \sum_{i=2}^{p} \big|f(x_i) - f(x_{i-1})\big|.$$

Then, for any self-adjoint matrix $\mathbf{X}_n$ *with independent vectors* $((X_{ij},\ i \le j),\ 1 \le j \le n)$ *and eigenvalues* $(\lambda_i)_{1\le i\le n}$ *for any function* f *with finite total variation norm so that* $E[|\frac{1}{n}\sum_{i=1}^n f(\lambda_i)|] < \infty$, *and any* $\delta > 0$

$$P\left(\left|\frac{1}{n}\sum_{i=1}^n f(\lambda_i) - \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n f(\lambda_i)\right]\right| \ge \delta \|f\|_{TV}\right) \le 2e^{-\frac{n\delta^2}{8}}.$$

In the general case, however, the extreme eigenvalues do not concentrate and can be very large for heavy-tailed entries [4, 9].

3.2. Global fluctuations

It is a natural question to wonder how the empirical measure of the eigenvalues fluctuates and, in particular, whether the concentration result of Theorem 3.2 is on the optimal scale. In the case where p is of order one, this question was first answered by Jonsson [51] by estimating moments, and in the context of Gaussian matrices by Johansson [50] by using loop equations. The main point is that the central limit theorem does not require a renormalization by the famous $\sqrt{n}$ as for the classical central limit theorems.

Theorem 3.4. *Assume that* $p \in (0,1)$ *independent of* n. *Let* f *be a continuously differentiable function. Let* λ_i *be the eigenvalues of* $\mathbf{X}_n$. *Then*

$$\sum_{i=1}^n f(\lambda_i) - \mathbb{E}\left[\sum_{i=1}^n f(\lambda_i)\right]$$

converges in distribution towards a centered Gaussian variable with variance

$$V(f) = \frac{1}{2\pi^2}\int_{-2}^{2}\int_{-2}^{2}\left(\frac{f(x)-f(y)}{x-y}\right)^2 \frac{(4-xy)}{\sqrt{4-x^2}\sqrt{4-y^2}}\,dx\,dy.$$

The central limit theorem also holds if one recenters with respect to the limit rather than the expectation; see e.g. [56].

On the contrary, if pn goes to a constant c, we see that Theorem 3.3 gives the optimal speed and we have a "more" classical central limit theorem [7, 20, 59]:

Theorem 3.5. *Assume that* pn *goes to* $c \in (0, +\infty)$. *Let* f *be a* C_b^1 *function. Then*

$$\frac{1}{\sqrt{n}}\left(\sum_{i=1}^n f(\lambda_i) - \mathbb{E}\left[\sum_{i=1}^n f(\lambda_i)\right]\right)$$

converges in law towards a centered Gaussian variable with non-trivial variance.

Together with [46], we claim that at least for pn of order one, or in $[n^{\varepsilon}, n^{1-\varepsilon}]$, or p of order one, we have the following theorem.

Theorem 3.6. *Let f be a C_b^1 function. Then*

$$\sqrt{p}\left(\sum_{i=1}^{n} f(\lambda_i) - \mathbb{E}\left[\sum_{i=1}^{n} f(\lambda_i)\right]\right)$$

converges in law towards a centered Gaussian variable with non-trivial covariance.

This result should hold for any $p > 1/n$.

3.3. Local laws

An important breakthrough towards the understanding of local fluctuations and eigenvectors is to analyze the so-called local laws as foreseen in [41]. Namely, to estimate $\sum f(\lambda_i)$ for less smooth functions, in fact for functions on a mesoscopic scale $f(x) = g(N^{\alpha}(x-E))$ for some $\alpha \in (0,1)$. Equivalently, one can look at $f(x) = (z-x)^{-1}$ with $z = E + i\eta$ with η of order $N^{-\alpha}$ (indeed the latter can serve to approximate conveniently the first). In this scale, it was proved that if pn goes to infinity, the mesoscopic distribution of the eigenvalues is still very close from the semi-circle distribution. Indeed, let us define the Stieltjes transform to be given by

$$G_n(z) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{z-\lambda_i}, \quad G_\mu(z) = \int \frac{1}{z-\lambda}\, d\mu(\lambda).$$

In [40, Theorems 2.8 and 2.10], the following result was proved, where ζ-high probability means a probability greater than or equal to $1 - e^{-v(\ln n)^{\zeta}}$ for some $v > 0$.

Theorem 3.7. *There are universal constants $C_1, C_2 > 0$ such that the following holds. Assume that*

$$pn \geq (\ln n)^{C_1 \xi}, \quad \xi = C_2 \ln\ln n.$$

Then, for $E \in [-3,3]$ and $D = \{z = E + i\eta, 0 < \eta < 3\}$,

$$\bigcap_{z\in D}\left\{\left|G_n(z) - m_\sigma(z)\right| \leq (\ln n)^{C_2\xi}\left(\min\left\{\frac{1}{pn\sqrt{\kappa_E+\eta}}, \frac{1}{\sqrt{pn}}\right\} + \frac{1}{n\eta}\right)\right\}$$

holds with ζ-high probability. Moreover, for $\eta > (\ln n)^{C\zeta} n^{-1}$

$$\#\{i : \lambda_i \in [E-\eta, E+\eta]\} = n\sigma\big([E-\eta, E+\eta]\big)\left(1 + O(\ln n)^{C\zeta}\left(\frac{1}{n\eta^{\frac{3}{2}}} + \frac{1}{pn\eta}\right)\right)$$

with ζ-high probability.

The above theorem applies for any p such that pn goes to infinity much faster than any $\ln n$; see e.g. [4]. Below $\ln n$, the extreme eigenvalues were shown to be dictated by the largest degree in the graph [18].

A similar statement in the sparse case where pn goes to a finite constant is still open. Indeed, the fact that μ_c has a dense set of atoms and a continuous part makes the analysis a priori much more involved and the local law more difficult to conjecture. An easier heavy tail matrix model was studied in [17, 26, 35], namely the random matrices with alpha-stable independent entries. In this case, the entries follow the alpha-stable law $\mathbb{P}(|A_{ij}| \geq t) \simeq t^{-\alpha}/n$. When $\alpha < 2$, it was shown in [17, 35] that the empirical measure converges towards a limiting law μ_α which is different from the semi-circle law. One of the advantages of this model is that μ_α is absolutely continuous except possibly for a discrete set of atoms. Of course, one cannot expect the eigenvalues to be as rigid in the heavy-tailed case since this would contradict the central limit theorem (which holds as in Theorem 3.6; see [20]). Hence, in this case, large eigenvalues should be less rigid, creating large fluctuations. The following result was proved if the A_{ij} are α-stable variables in [26, 27]: for all $t \in \mathbb{R}$,

$$\mathbb{E}\big[\exp(itA_{11})\big] = \exp\left(-\frac{1}{n}w_\alpha |t|^\alpha\right), \tag{3.1}$$

for some $0 < \alpha < 2$ and $w_\alpha = \pi/(\sin(\pi\alpha/2)\Gamma(\alpha))$. We put

$$\rho = \begin{cases} \frac{1}{2} & \text{if } \frac{8}{5} \leq \alpha < 2, \\ \frac{\alpha}{8-3\alpha} & \text{if } 1 < \alpha < \frac{8}{5}, \\ \frac{\alpha}{2+3\alpha} & \text{if } 0 < \alpha \leq 1. \end{cases} \tag{3.2}$$

Then, there exists a finite set $\mathcal{E}_\alpha \subset \mathbb{R}$ such that if $K \subset \mathbb{R}\backslash\mathcal{E}_\alpha$ is a compact set and $\delta > 0$, the following holds. There are constants $c_0, c_1 > 0$ such that for all integers $n \geq 1$, if $I \subset K$ is an interval of length $|I| \geq c_1 n^{-\rho}(\ln n)^2$, then

$$\big|N_I - n\mu_\alpha(I)\big| \leq \delta n|I|, \tag{3.3}$$

with probability at least $1 - 2\exp(-c_0 n\delta^2|I|^2)$. The fact that our result might not be true on a finite set of values should only be technical. This result was improved in [2, Theorems 3.4 and 3.5] in order to tackle I of size $n^{-\omega(\alpha)}$ with $\omega(\alpha) > 1/2$ (and $\Re(z)$ small enough when $\alpha < 1$). Such an optimal scale is important in the study of the local fluctuations of the spectrum.

In both light and heavy tails, the main point is to estimate the Stieltjes transform $G_n(z) = \frac{1}{n}\sum_{i=1}^n (z-\lambda_i)^{-1}$ for z going to the real axis: $z = E + i\eta$ with η of order nearly as good as n^{-1} for light tails, $n^{-\rho}$ for heavy tails. This is done by showing that G_n is characterized approximately by a closed set of equations. In the case of lights tails, one has simply a quadratic equation for G_n and needs to show that the error terms remain small as z approaches the real line. In the heavy-tailed case, the equations are much more complicated, see (2.7) and (2.8), and therefore more difficult to handle. Similar questions are completely open for other heavy-tailed matrices, including Bernoulli matrices with pn of order one.

3.4. Local fluctuations

When the average degree pn is large, one expects the eigenvalues to behave exactly as the eigenvalues of a symmetric matrix with independent Gaussian entries (so-called GOE matrices). The advantage of Gaussian matrices is that they are an integrable model of random matrices in the sense that many of their properties can be exactly computed. To start with, the joint distribution of its eigenvalues $(\lambda_i^G)_{1\le i\le n}$ is explicit:

$$d\mathbb{P}(\lambda^G) = \frac{1}{Z}\Delta(\lambda)e^{-\frac{n}{4}\sum(\lambda_i^G)^2}\prod d\lambda_i^G, \tag{3.4}$$

where

$$\Delta(\lambda) = \prod_{i<j}|\lambda_i - \lambda_j|$$

is the Vandermonde determinant. In particular, this formula does not depend on the eigenvectors. Based on this formula, Tracy and Widom could study the local fluctuations of the spectrum $(\lambda_i^G)_{1\le i\le n}$ [66, 67] and they proved that

$$\lim_{n\to\infty}\mathbb{P}\big(n^{2/3}(\lambda_1^G - 2) \le s\big) = F_1(s),$$

where F_1 is the distribution function of the Tracy–Widom law. For the eigenvalues in the bulk, it was proved [55] that, for all smooth compactly supported function,

$$\mathcal{E}_{\mathbf{G}_n}(O, E) = \mathbb{E}\big[O\big(n(\lambda_i^G - E), \dots, n(\lambda_{i+p}^G - E)\big)\big]$$

converges as n goes to infinity and the limit is described in terms of Pfaffian distributions.

The universality in the bulk was obtained after a series of works including notably [41, 42, 62] and [39, Theorem 2.5] (for $\phi \ge 2/3$) and improved in [48] (for $\phi > 0$) to finally get the following theorem.

Theorem 3.8 (Bulk universality). *Suppose that $pn > n^\phi$ with $\phi > 0$. There exists b_n going to zero so that for all smooth compactly supported function O, any $E \in (-2, 2)$,*

$$\lim_{n\to\infty}\int_{E-b_n}^{E+b_n}\frac{dE'}{2b_n}\big(\mathcal{E}_{\mathbf{G}_n}(O, E') - \mathcal{E}_{\mathbf{B}_n}(O, E')\big) = 0.$$

Moreover, the universality at the edge was obtained in [39, Theorem 2.7]; see also [60].

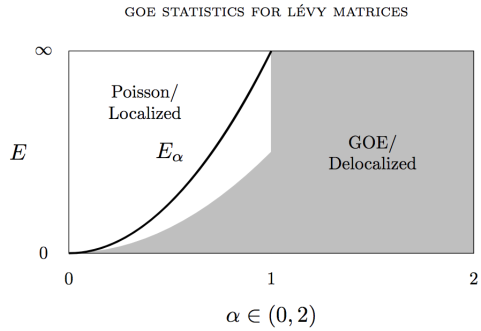

Figure 4.

Theorem 3.9 (Edge universality). *Suppose that $pn > n^\phi$, $\phi > 2/3$. Then there exists $\delta > 0$ such that*

$$\mathbb{P}\big(n^{2/3}(\lambda_2^B - 2) \le s\big) = \mathbb{P}\big(n^{2/3}(\lambda_2^G - 2) \le s + O(n^{-\delta})\big) + O(n^{-\delta}).$$

This statement was generalized to $pn > n^{1/3}$ but the largest eigenvalue then needs to be shifted by a deterministic drift of order $1/pn$ [53]. Beyond this threshold, the fluctuations of the second largest eigenvalue starts to be Gaussian.

When pn decreases below $1/3$, it was proved that universality stops to hold and fluctuations of the largest eigenvalue start to be Gaussian. The precise transition between Tracy–Widom law and Gaussian fluctuations when p is of order $n^{-2/3}$ was described in [49]. When $n^{o(1)} \ll pn \ll n^{1/3}$, the papers [47, 49] show that the fluctuations of the extreme eigenvalues are Gaussian, even if they stick to the bulk. In the case where $pn \ll \ln n$, Theorem 2.3 asserts that the eigenvalues go away from the bulk, at distance of order $\sqrt{\ln n}$. The corresponding eigenvectors are localized close to the vertices with a high degree. In an even more recent preprint [5], the same authors show that these eigenvalues follow a Poisson point process. Such questions are open for Bernoulli random matrices with pn of order $c \in (0, +\infty)$ and eigenvalues in the bulk. Indeed, as we have seen, the limiting density is a mixture of atoms and continuous density and it is not yet clear how to zoom in the spectrum in such a situation. However, such questions could be analyzed for Lévy matrices with α-stable entries in the regime where local law can be obtained on the optimal scale $n^{-1/2}$ [2]. Figure 4 depicts the expected regimes. In fact, one expects the following transition to occur (see [63]).

- If $\alpha \in [1, 2]$, all eigenvectors corresponding to finite eigenvalues are completely delocalized. Further, for any $E \in \mathbb{R}$, the local statistics of the eigenvalues near E converge to those of the GOE as N goes to infinity.
- If $\alpha \in (0, 1)$, there exists a mobility edge E_α such that for $|E| < E_\alpha$ the local statistics of the eigenvalues near E converge to those of the GOE as N goes to

infinity. But if $|E| > E_\alpha$, the local statistics of the eigenvalues near E converge to those of a Poisson point process and all eigenvectors in this region are localized. The fact that local statistics are given by those of Gaussian matrices for $\alpha \in (1,2)$ or $\alpha \in (0,1)$ and E small enough, except for E in some finite set, was proved in [2, Theorems 2.4 and 2.5].

3.5. Properties of the eigenvectors

The properties of the eigenvectors are intimately related with local laws. Indeed, by definition of the eigenvectors, if v is an eigenvector of the symmetric matrix $\mathbf{X}_n$ for the eigenvalue E and we set $\langle v, e_i\rangle = v_i$, then X_1 is the first column vector of $\mathbf{X}_n$ while $\mathbf{X}_n^{(1)}$ is the $(n-1)\times(n-1)$ principal minor of $\mathbf{X}_n$ obtained by removing the column and row vector given by X_1 and X_1^T:

$$v_1^2 = \big(1 + \big\langle X_1, (E - \mathbf{X}_n^{(1)})^{-2} X_1\big\rangle\big)^{-1},$$

where, at least in the dense cases $\langle X_1, (E - \mathbf{X}_n^{(1)})^{-2} X_1\rangle$ is close to $\frac{1}{n}\operatorname{Tr}(E - \mathbf{X}_n)^{-2}$, and so is governed by the local law. In [40, Theorem 2.16], the following theorem was proved.

Theorem 3.10 (Complete delocalization of eigenvectors). *Assume the hypotheses of Theorem 3.7 with $pn > n^\phi$ with $\phi > 0$. Let v_i be the eigenvectors of $\mathbf{B}_n$ for the eigenvalues $\lambda_n \le \lambda_{n-1} \cdots \le \lambda_1$. Then*

$$\max_{i\le n} \|v_i\|_\infty \le \frac{(\ln n)^{4\xi}}{\sqrt{n}}$$

with ζ-high probability.

This result was extended to q going to infinity logarithmically only more recently [3]. We roughly state their result:

- (Semilocalized phase) Assume that $C\sqrt{\ln n}\ln\ln n \le \sqrt{pn} \le 3\ln n$ and let w be a normalized eigenvector of $\mathbf{B}_n$ with non-trivial eigenvalue $E \ge 2 + C\zeta^{1/2}$. We let $\Lambda(\alpha) = \alpha/\sqrt{\alpha - 1}$ and $\alpha_x = \sum_y \mathbf{B}_{xy}/pn$. We let $W_{E,\delta}$ be the set of vertices such that $\Lambda(\alpha_x) \in [E-\delta, E+\delta]$. Then for each $x \in W_{E,\delta}$, there exists a normalized vector $v(x)$ supported in a ball around x and radius $c\sqrt{\ln n}$, such that the support of $v(cx)$ and $v(y)$ is distinct if $x \ne y$ and

$$\sum_{x\in W_{E,\delta}} \big\langle v(x), w\big\rangle^2 \ge 1 - C\left(\sqrt{\ln n}\, pn \ln pn + \sqrt{\ln n}\, pn \frac{1}{E-2}\right)^2 \delta^{-2}.$$

  Moreover,

$$\sum_{y\in B_r(x)} \big(v(x)\big)_y^2 \le \frac{1}{(\alpha_x - 1)^{r+1}}.$$

- (Delocalized phase) For any $\nu > 0$ and $\kappa > 0$, there exists a constant $C > 0$ such that for $pn \in [C\sqrt{\ln n}, (\ln n)^{3/2}]$, if w is a normalized eigenvector for $\mathbf{B}_n$ with eigenvalue $E \in [-2+\kappa, -\kappa] \cup [\kappa, 2-\kappa]$,

$$\|w\|_\infty^2 \le n^{-1+\kappa}$$

with probability greater than $1 - n^{-\nu}$.

This question is completely open for Bernoulli random matrices with pn of order one but the understanding of Lévy matrices is again more complete. Based on [2, 26, 27], we can assert that Tarquini, Biroli, and Tarzia's conjecture [63] is partly proved. Indeed the complete delocalization is proved for $\alpha \in (1, 2)$ and $\alpha \in (0, 1)$ and small enough eigenvalues. A sort of localization for $\alpha \in (0, 1)$ for large enough eigenvalue was derived in [26], and was shown to be not true for small enough eigenvalues in [27]: the transition and the value of the mobility edge is still an open question. In fact, even in the case where the eigenvalue statistics belong to the universality class of Gaussian matrices, the fine properties of the eigenvectors of Lévy matrices differ [1]. Let us also mention [57] which shows under quite general assumptions that eigenvectors are somehow uniformly delocalized in the sense that any subset of at least eight coordinates carries a non-negligible part of the mass of an eigenvector.

4. Rare events

It is sometimes important to estimate the probability of rare events, such as the probability that the extreme eigenvalues take unlikely values or the empirical measure of the eigenvalues shows an unlikely profile, and what kind of optimal strategy can lead to such deviations from the expected behavior. In the case of Gaussian symmetric matrices, the joint density of the eigenvalues is known (3.4). One finds by sort of Laplace's principle [15, 16] the large deviations for the empirical measure and the largest eigenvalue.

Theorem 4.1. *Let $\lambda_n^G \le \lambda_2^G \cdots \le \lambda_1^G$ be the eigenvalues of a GOE matrix. Then, the following holds.*

- *Let $E(\mu) = \frac{1}{2}\iint (\frac{x^2}{4} + \frac{y^2}{4} - \ln|x-y|)\, d\mu(x)\, d\mu(y)$ and set $\mathcal{E}(\mu) = E - \inf E$. Then $\mathcal{E}$ is a good rate function and the distribution of the empirical measure of the eigenvalues $\hat{\mu}_n = \frac{1}{n}\sum \delta_{\lambda_i^G}$ satisfies a large deviations principle (LDP) with speed n^2 with rate function $\mathcal{I}$, that is for every closed set F*

$$\limsup_{n\to\infty} \frac{1}{n^2} \ln \mathbb{P}(\hat{\mu}_n \in F) \le -\inf_F \mathcal{E},$$

whereas for any open set O

$$\limsup_{n\to\infty} \frac{1}{n^2} \ln \mathbb{P}(\hat{\mu}_n \in O) \ge -\inf_O \mathcal{E}.$$

- *Let* $I_G(x) = \frac{1}{2}\int_2^x \sqrt{4-y^2}dy$ *for* $x \geq 2$ *and* $I_G(x) = +\infty$ *for* $x < 2$. *Then* I *is a good rate function and the distribution of* λ_1^G *satisfies an LDP with speed* n *and good rate function* I_G.

In this case, deviations of the spectrum can be created independently from the eigenvectors which stay uniformly distributed. On the other hand, if the entries have sharp exponential decay, large deviations can be created by large entries. Assume that for some $\alpha \in (0, 2)$, there exists $a > 0$ so that for all i, j

$$\lim_{t\to\infty} 2^{-1_{i=j}} t^{-\alpha} \ln \mathbb{P}\big(|\sqrt{n}X_{ij}| \geq t\big) = -a.$$

Theorem 4.2. • *The law of the empirical measure satisfies an LDP in the speed* $n^{1+\frac{\alpha}{2}}$ *and good rate function which is infinite unless* $\mu = \sigma \boxplus \nu$ *and then equals* $a \int |x|^\alpha d\nu(x)$ [22].

- *The law of the largest eigenvalue satisfies an LDP with rate* $n^{\frac{\alpha}{2}}$ *and GRF proportional to* $(\int (x-y)^{-1} d\sigma(y))^{-\alpha}$ [10].

However, the situation is much less understood for Bernoulli matrices and again the sparse and the dense regimes lead to very different results and techniques. We discuss these questions hereafter.

4.1. Large deviations for the extreme eigenvalues

Let us first consider the dense case. In [12, 43], we considered the large deviations for the largest eigenvalue of Wigner matrices and showed that if the entries are Rademacher, then the same large deviation principle holds, whereas in general there is a transition between deviations close to two where the rate function is the Gaussian one whereas for large deviations towards large enough values the rate function is more of a heavy tail type. In a work in progress with F. Augeri, R. Ducatez and J. Husson, we prove the following theorem.

Theorem 4.3. • *Assume that* $p = 1/2$. *Then the law of* λ_1^X *satisfies an LDP in the scale* n *and with the same rate function* I_G *as for the GOE matrix.*

- *Assume that* $p \in (0, 1/2)$. *Then for* x *close enough to* 2, *the probability that* λ_1^X *is close to* x *is the same as in the Gaussian case. But for* x *large enough,*

$$\limsup_{\delta\downarrow 0} \limsup_{n\to\infty} \frac{1}{n} \ln \mathbb{P}\big(|\lambda_1^X - x| < \delta\big) = -I_p(x),$$

where $I_p(x) < I_G(x)$.

The case $p \in (1/2, 1)$ is under investigation. In fact, analyzing the large deviation requires to understand good strategies to create the deviations. For $p = 1/2$, it is shown that an optimal strategy is to tilt the law of the entries in order to change their expectation so that the matrix looks like a rank one deformation of Bernoulli matrix with a delocalized deformation. The eigenvectors also stay delocalized through this deformation. When $p < 1/2$ and x is large, it turns out that the optimal strategy is to create fully connected components of size $\sqrt{n}$. For $p > 1/2$, the picture is less clear and we suspect that vertices with high degree are optimal ways to create large eigenvalues.

Let us now consider the sparse case following [21]: in this case we already saw that large eigenvalues are created by vertices with large degree, namely with row or column vectors with many entries equal to one.

Theorem 4.4. *Let $L_p = \frac{\ln n}{\ln\ln n - \ln(np)}$ and assume that*

$$\ln(1/np) \ll \ln n \quad \text{and} \quad np \ll \sqrt{\ln n / \ln\ln n}.$$

Let λ_2 be the second largest eigenvalue of $\mathbf{B}_n$. Then for any $\delta \geq 0$,

$$\lim_{n\to\infty} \frac{-\ln P\left(\lambda_2 \geq (1+\delta)\sqrt{L_p}\right)}{\ln n} = 2\delta + \delta^2,$$

whereas

$$\lim_{n\to\infty} \frac{-\ln P\left(\lambda_2 \leq (1-\delta)\sqrt{L_p}\right)}{\ln n} = 2\delta - \delta^2.$$

4.2. Large deviations for the empirical measure

In [23, Theorem 1.6], a large deviation for the empirical measure of the eigenvalue in the sparse case was derived: we do not make precise the rate function as it is obtained by contraction from the large deviation for the empirical neighborhood distribution.

Theorem 4.5. *Assume that pn is fixed. Then the law of $\widehat{\mu}_n$ satisfies a large deviation principle with speed n.*

This question is still open when $pn \gg 1$. When p is of order one, we should expect to have a large deviation with speed n^2 according to the concentration of measure, but the rate function should not be equal to the Gaussian one even when $p = 1/2$ because the Dirac at the origin should have rate function bounded above by $\ln p$ (whereas it is infinite in the Gaussian case).

4.3. Large deviations for triangle counts

The traces of Bernoulli matrices have a combinatorial interpretation. For instance, $\mathrm{Tr}(\mathbf{B}_n^3)$ is the number $T_{n,p}$ of triangles in the Erdős–Rényi graph. Observe that its expectation is of order $p^3 n^3$. In the well-known paper [34, Theorem 4.1], the following theorem was proved.

Theorem 4.6. *Let*

$$I_p(f) = \sup_{\phi}\left\{\int_0^1\int_0^1 f(x,y)\phi(x,y)\,dx\,dy - \frac{1}{2}\iint \ln\left(pe^{2\phi(x,y)} + (1-p)\right)dx\,dy\right\}$$

and set $\varphi(p,t) = \in \{I_p(f), \int f(x,y)f(y,v)f(v,x)dxdydv \geq 6t\}$. *Then for each* $p \in (0,1)$,

$$\lim_{n\to\infty}\frac{1}{n^2}\ln \mathbb{P}(T_{n,p} \geq tn^3) = -\varphi(p,t).$$

This result extends to any moment $\mathrm{Tr}(\mathbf{B}_n^k)$. However, observe that it does not tell us about deviations of the empirical measure since $x \to x^k$ is unbounded so that deviations of the extreme eigenvalues matter. It is natural to wonder what happens as well when p goes to zero. This question was attacked in [33,36,37], but we state here [11, Proposition 1.19]

Theorem 4.7. *Let p go to zero with n so that* $(\ln n)^4 \ll np^2$. *Set* $v_n = n^2p^2\ln(1/p)$. *Then for* $t \geq 1$,

$$\lim_{n\to\infty}\frac{1}{v_n}\ln \mathbb{P}\left(\mathrm{Tr}(\mathbf{B}_n^d) \geq tn^d p^d\right) = -\Phi(t),$$

where $\Phi(t) = \frac{1}{2}(t-1)^{2/d}$ *if* $n^{-1} \ll p \ll n^{-1/2}$, *but* $\Phi(t) = \min\{\theta_t, \frac{1}{2}(t-1)^{2/d}\}$ *if* $p \gg n^{-1/2}$ *and θ_t is the solution of* $P_{C_d}(\theta_t) = t$, *where P_{C_d} is the independence polynomial of the d-cycle.*

4.4. The singularity probability

A well-known problem has been to estimate the probability that a matrix $\widetilde{\mathbf{B}}_n$ with all independent Bernoulli entries (hence not self-adjoint) is singular. In a breakthrough paper, Tikhomirov [64] (see also [54]) could exactly estimate it, by showing that the best strategy to achieve singularity is to have a zero column or row vector.

Theorem 4.8. *There exists a finite constant C such that if* $C\ln n/n \leq p \leq \frac{1}{2}$,

$$\mathbb{P}(\widetilde{\mathbf{B}}_n \text{ is singular}) = \left(2 + o_n(1)\right)(1-p)^n n.$$

Such an optimal estimate is not yet known for the symmetric Bernoulli matrix $\mathbf{B}_n$ (even though it is conjectured) but the paper [32] proves that the probability that it is singular is bounded above by $e^{-O(\sqrt{n})}$. This was improved in an exponential upper bound in [31].

5. Open problems

(1) Local law for Bernoulli matrices when pn is of order one. This could be at best on the scale $\sqrt{n}$ but is tricky even to state because of the atoms of the limit law.

(2) Localization/delocalization of the eigenvectors of Bernoulli matrices for pn of order one (one would conjecture that Dirac masses yield localization but the continuous part yields delocalization, however the right criteria to express this remains to be given). Find a critical c^* such that for $np > c^*$ there exists delocalized vectors with connected support with high probability.

(3) Large deviations for the empirical measure of the eigenvalues of Bernoulli matrices (all p so that $pn \gg 1$). Even when $p = 1/2$, one does not expect to retrieve the Gaussian rate function since the entropy should be finite at δ_0 (as can be seen by requiring all entries to be equal).

(4) Precise estimate on the singularity probability in the symmetric case.

(5) In comparison, d-regular graphs which are picked uniformly at random are conjectured to be in the universality class of Gaussian random matrices for all $d \geq 3$. This was proved for d going to infinity fast enough [13, 14], and recently Huang and Yau could get the local law and the delocalization of the eigenvectors up to $d = 3$.

Acknowledgments. I thank Charles Bordenave, Christophe Garban, Jiaoyang Huang, Antti Knowles, Justin Salez, and Ofer Zeitouni for pointing out recent developments of the subject and for their very helpful comments on preliminary versions of this survey.

Funding. This work was supported in part by ERC-2019-ADG Project 884584-LDRAM.

References

[1] A. Aggarwal, P. Lopatto, and J. Marcinek, Eigenvector statistics of Lévy matrices. *Ann. Probab.* **49** (2021), no. 4, 1778–1846 Zbl 1467.60003 MR 4260468

[2] A. Aggarwal, P. Lopatto, and H.-T. Yau, GOE statistics for Lévy matrices. *J. Eur. Math. Soc. (JEMS)* **23** (2021), no. 11, 3707–3800 Zbl 07445595 MR 4310816

[3] J. Alt, R. Ducatez, and A. Knowles, Delocalization transition for critical Erdős–Rényi graphs. *Comm. Math. Phys.* **388** (2021), no. 1, 507–579 Zbl 1477.15029 MR 4328063

[4] J. Alt, R. Ducatez, and A. Knowles, Extremal eigenvalues of critical Erdős–Rényi graphs. *Ann. Probab.* **49** (2021), no. 3, 1347–1401 Zbl 1467.05236 MR 4255147

[5] J. Alt, R. Ducatez, and A. Knowles, Poisson statistics and localization at the spectral edge of sparse Erdős–Rényi graphs. 2021, arXiv:2106.12519

[6] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices*. Cambridge Stud. Adv. Math. 118, Cambridge University Press, Cambridge, 2010 Zbl 1184.15023 MR 2760897

[7] G. W. Anderson and O. Zeitouni, A CLT for regularized sample covariance matrices. *Ann. Statist.* **36** (2008), no. 6, 2553–2576 Zbl 1360.60047 MR 2485007

[8] A. Arras and C. Bordenave, Existence of absolutely continuous spectrum for Galton–Watson random trees. 2021, arXiv:2105.10177

[9] A. Auffinger, G. Ben Arous, and S. Péché, Poisson convergence for the largest eigenvalues of heavy tailed random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.* **45** (2009), no. 3, 589–610 Zbl 1177.15037 MR 2548495

[10] F. Augeri, Large deviations principle for the largest eigenvalue of Wigner matrices without Gaussian tails. *Electron. J. Probab.* **21** (2016), Paper No. 32 Zbl 1338.60010 MR 3492936

[11] F. Augeri, Nonlinear large deviation bounds with applications to Wigner matrices and sparse Erdős–Rényi graphs. *Ann. Probab.* **48** (2020), no. 5, 2404–2448 Zbl 1456.60063 MR 4152647

[12] F. Augeri, A. Guionnet, and J. Husson, Large deviations for the largest eigenvalue of sub-Gaussian matrices. *Comm. Math. Phys.* **383** (2021), no. 2, 997–1050 Zbl 1479.60011 MR 4239836

[13] R. Bauerschmidt, J. Huang, A. Knowles, and H.-T. Yau, Bulk eigenvalue statistics for random regular graphs. *Ann. Probab.* **45** (2017), no. 6A, 3626–3663 Zbl 1379.05098 MR 3729611

[14] R. Bauerschmidt, J. Huang, A. Knowles, and H.-T. Yau, Edge rigidity and universality of random regular graphs of intermediate degree. *Geom. Funct. Anal.* **30** (2020), no. 3, 693–769 Zbl 1453.05117 MR 4135670

[15] G. Ben Arous, A. Dembo, and A. Guionnet, Aging of spherical spin glasses. *Probab. Theory Related Fields* **120** (2001), no. 1, 1–67 Zbl 0993.60055 MR 1856194

[16] G. Ben Arous and A. Guionnet, Large deviations for Wigner's law and Voiculescu's non-commutative entropy. *Probab. Theory Related Fields* **108** (1997), no. 4, 517–542 Zbl 0954.60029 MR 1465640

[17] G. Ben Arous and A. Guionnet, The spectrum of heavy tailed random matrices. *Comm. Math. Phys.* **278** (2008), no. 3, 715–751 Zbl 1157.60005 MR 2373441

[18] F. Benaych-Georges, C. Bordenave, and A. Knowles, Largest eigenvalues of sparse inhomogeneous Erdős–Rényi graphs. *Ann. Probab.* **47** (2019), no. 3, 1653–1676 Zbl 1447.60017 MR 3945756

[19] F. Benaych-Georges and A. Guionnet, Central limit theorem for eigenvectors of heavy tailed matrices. *Electron. J. Probab.* **19** (2014), Paper No. 54 Zbl 1293.15021 MR 3227063

[20] F. Benaych-Georges, A. Guionnet, and C. Male, Central limit theorems for linear statistics of heavy tailed random matrices. *Comm. Math. Phys.* **329** (2014), no. 2, 641–686 Zbl 1294.60039 MR 3210147

[21] B. B. Bhattacharya, S. Bhattacharya, and S. Ganguly, Spectral edge in sparse random graphs: Upper and lower tail large deviations. *Ann. Probab.* **49** (2021), no. 4, 1847–1885 Zbl 1467.05237 MR 4260469

[22] C. Bordenave and P. Caputo, A large deviation principle for Wigner matrices without Gaussian tails. *Ann. Probab.* **42** (2014), no. 6, 2454–2496 Zbl 1330.60012 MR 3265172

[23] C. Bordenave and P. Caputo, Large deviations of empirical neighborhood distribution in sparse random graphs. *Probab. Theory Related Fields* **163** (2015), no. 1-2, 149–222 Zbl 1327.60067 MR 3405616

[24] C. Bordenave, P. Caputo, and D. Chafaï, Spectrum of large random reversible Markov chains: Heavy-tailed weights on the complete graph. *Ann. Probab.* **39** (2011), no. 4, 1544–1590 Zbl 1245.60008 MR 2857250

[25] C. Bordenave, P. Caputo, and D. Chafaï, Spectrum of non-Hermitian heavy tailed random matrices. *Comm. Math. Phys.* **307** (2011), no. 2, 513–560 Zbl 1235.60008 MR 2837123

[26] C. Bordenave and A. Guionnet, Localization and delocalization of eigenvectors for heavy-tailed random matrices. *Probab. Theory Related Fields* **157** (2013), no. 3-4, 885–953 Zbl 1296.15019 MR 3129806

[27] C. Bordenave and A. Guionnet, Delocalization at small energy for heavy-tailed random matrices. *Comm. Math. Phys.* **354** (2017), no. 1, 115–159 Zbl 1388.60022 MR 3656514

[28] C. Bordenave and M. Lelarge, Resolvent of large random graphs. *Random Structures Algorithms* **37** (2010), no. 3, 332–352 Zbl 1209.05222 MR 2724665

[29] C. Bordenave, M. Lelarge, and J. Salez, The rank of diluted random graphs. *Ann. Probab.* **39** (2011), no. 3, 1097–1121 Zbl 1298.05283 MR 2789584

[30] C. Bordenave, A. Sen, and B. Virág, Mean quantum percolation. *J. Eur. Math. Soc. (JEMS)* **19** (2017), no. 12, 3679–3707 Zbl 1385.60057 MR 3730511

[31] M. Campos, M. Jenssen, M. Michelen, and J. Sahasrabudhe, The singularity probability of a random symmetric matrix is exponentially small. 2021, arXiv:2105.11384

[32] M. Campos, L. Mattos, R. Morris, and N. Morrison, On the singularity of random symmetric matrices. *Duke Math. J.* **170** (2021), no. 5, 881–907 Zbl 1467.60005 MR 4255046

[33] S. Chatterjee and A. Dembo, Nonlinear large deviations. *Adv. Math.* **299** (2016), 396–450 Zbl 1356.60045 MR 3519474

[34] S. Chatterjee and S. R. S. Varadhan, The large deviation principle for the Erdős–Rényi random graph. *European J. Combin.* **32** (2011), no. 7, 1000–1017 Zbl 1230.05259 MR 2825532

[35] P. Cizeau and J. P. Bouchaud, Theory of Lévy matrices. *Phys. Rev. E* **50** (1994), no. 3, 1810–1822

[36] N. Cook and A. Dembo, Large deviations of subgraph counts for sparse Erdős–Rényi graphs. *Adv. Math.* **373** (2020), 107289, 53 Zbl 1473.60058 MR 4130460

[37] R. Eldan, Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations. *Geom. Funct. Anal.* **28** (2018), no. 6, 1548–1596 Zbl 1428.60045 MR 3881829

[38] N. Enriquez and L. Ménard, Spectra of large diluted but bushy random graphs. *Random Structures Algorithms* **49** (2016), no. 1, 160–184 Zbl 1348.05189 MR 3521277

[39] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, Spectral statistics of Erdős–Rényi graphs II: Eigenvalue spacing and the extreme eigenvalues. *Comm. Math. Phys.* **314** (2012), no. 3, 587–640 Zbl 1251.05162 MR 2964770

[40] L. Erdős, A. Knowles, H.-T. Yau, and J. Yin, The local semicircle law for a general class of random matrices. *Electron. J. Probab.* **18** (2013), Paper No. 59 Zbl 1373.15053 MR 3068390

[41] L. Erdős, B. Schlein, and H.-T. Yau, Wegner estimate and level repulsion for Wigner random matrices. *Int. Math. Res. Not. IMRN* **2010** (2010), no. 3, 436–479 Zbl 1204.15043 MR 2587574

[42] L. Erdős, B. Schlein, and H.-T. Yau, Universality of random matrices and local relaxation flow. *Invent. Math.* **185** (2011), no. 1, 75–119 Zbl 1225.15033 MR 2810797

[43] A. Guionnet and J. Husson, Large deviations for the largest eigenvalue of Rademacher matrices. *Ann. Probab.* **48** (2020), no. 3, 1436–1465 Zbl 1444.60021 MR 4112720

[44] A. Guionnet and O. Zeitouni, Concentration of the spectral measure for large matrices. *Electron. Comm. Probab.* **5** (2000), 119–136 Zbl 0969.15010 MR 1781846

[45] A. Guionnet and O. Zeitouni, Large deviations asymptotics for spherical integrals. *J. Funct. Anal.* **188** (2002), no. 2, 461–515 Zbl 1002.60021 MR 1883414

[46] Y. He, Bulk eigenvalue fluctuations of sparse random matrices. *Ann. Appl. Probab.* **30** (2020), no. 6, 2846–2879 Zbl 1482.15037 MR 4187130

[47] Y. He and A. Knowles, Fluctuations of extreme eigenvalues of sparse Erdős–Rényi graphs. *Probab. Theory Related Fields* **180** (2021), no. 3-4, 985–1056 Zbl 1468.05155 MR 4288336

[48] J. Huang, B. Landon, and H.-T. Yau, Bulk universality of sparse random matrices. *J. Math. Phys.* **56** (2015), no. 12, Paper No. 123301 Zbl 1329.05262 MR 3429490

[49] J. Huang, B. Landon, and H.-T. Yau, Transition from Tracy-Widom to Gaussian fluctuations of extremal eigenvalues of sparse Erdős–Rényi graphs. *Ann. Probab.* **48** (2020), no. 2, 916–962 Zbl 1440.05181 MR 4089498

[50] K. Johansson, On fluctuations of eigenvalues of random Hermitian matrices. *Duke Math. J.* **91** (1998), no. 1, 151–204 Zbl 1039.82504 MR 1487983

[51] D. Jonsson, Some limit theorems for the eigenvalues of a sample covariance matrix. *J. Multivariate Anal.* **12** (1982), no. 1, 1–38 Zbl 0491.62021 MR 650926

[52] O. Khorunzhy, M. Shcherbina, and V. Vengerovsky, Eigenvalue distribution of large weighted random graphs. *J. Math. Phys.* **45** (2004), no. 4, 1648–1672 Zbl 1068.05062 MR 2043849

[53] J. O. Lee and K. Schnelli, Local law and Tracy–Widom limit for sparse random matrices. *Probab. Theory Related Fields* **171** (2018), no. 1-2, 543–616 Zbl 1429.60012 MR 3800840

[54] A. E. Litvak and K. E. Tikhomirov, Singularity of sparse Bernoulli matrices. *Duke Math. J.* **171** (2022), no. 5, 1135–1233 Zbl 07513357 MR 4402560

[55] M. L. Mehta, *Random Matrices*. 3rd edn., Pure Appl. Math. (Amsterdam) 142, Elsevier/Academic Press, Amsterdam, 2004 Zbl 1107.15019 MR 2129906

[56] L. Pastur and M. Shcherbina, *Eigenvalue Distribution of Large Random Matrices*. Math. Surveys Monogr. 171, American Mathematical Society, Providence, RI, 2011 Zbl 1244.15002 MR 2808038

[57] M. Rudelson and R. Vershynin, No-gaps delocalization for general random matrices. *Geom. Funct. Anal.* **26** (2016), no. 6, 1716–1776 Zbl 1375.60027 MR 3579707

[58] J. Salez, Every totally real algebraic integer is a tree eigenvalue. *J. Combin. Theory Ser. B* **111** (2015), 249–256 Zbl 1307.05150 MR 3315609

[59] M. Shcherbina and B. Tirozzi, Central limit theorem for fluctuations of linear eigenvalue statistics of large random graphs. *J. Math. Phys.* **51** (2010), no. 2, 023523, 20 Zbl 1309.05163 MR 2605074

[60] A. Soshnikov, Universality at the edge of the spectrum in Wigner random matrices. *Comm. Math. Phys.* **207** (1999), no. 3, 697–733 Zbl 1062.82502 MR 1727234

[61] M. Talagrand, Concentration of measure and isoperimetric inequalities in product spaces. *Publ. Math. Inst. Hautes Études Sci.* **81** (1995), 73–205 Zbl 0864.60013 MR 1361756

[62] T. Tao and V. Vu, The Wigner–Dyson–Mehta bulk universality conjecture for Wigner matrices. *Electron. J. Probab.* **16** (2011), no. 77, 2104–2121 Zbl 1245.15041 MR 2851058

[63] E. Tarquini, G. Biroli, and M. Tarzia, Level statistics and localization transitions of Lévy matrices. *Phys. Rev. Lett.* **116** (2016), no. 1, Paper No. 010601 Zbl 1356.15018 MR 3555687

[64] K. Tikhomirov, Singularity of random Bernoulli matrices. *Ann. of Math. (2)* **191** (2020), no. 2, 593–634 Zbl 1458.15023 MR 4076632

[65] K. Tikhomirov and P. Youssef, Outliers in spectrum of sparse Wigner matrices. *Random Structures Algorithms* **58** (2021), no. 3, 517–605 MR 4234995

[66] C. A. Tracy and H. Widom, Introduction to random matrices. In *Geometric and Quantum Aspects of Integrable Systems (Scheveningen, 1992)*, pp. 103–130, Lecture Notes in Phys. 424, Springer, Berlin, 1993 Zbl 0791.15017 MR 1253763

[67] C. A. Tracy and H. Widom, On orthogonal and symplectic matrix ensembles. *Comm. Math. Phys.* **177** (1996), no. 3, 727–754 Zbl 0851.60101 MR 1385083

[68] V. H. Vu, Recent progress in combinatorial random matrix theory. *Probab. Surv.* **18** (2021), 179–200 Zbl 07367835 MR 4260513

[69] E. P. Wigner, On the distribution of the roots of certain symmetric matrices. *Ann. of Math. (2)* **67** (1958), 325–327 Zbl 0085.13203 MR 95527

[70] I. Zakharevich, A generalization of Wigner's law. *Comm. Math. Phys.* **268** (2006), no. 2, 403–414 Zbl 1147.82334 MR 2259200

Alice Guionnet
Ecole Normale Supérieure de Lyon, Université de Lyon, CNRS, 69342 Lyon, France;
alice.guionnet@ens-lyon.fr