



Spectrum of large random reversible Markov chains

Charles Bordenave, Pietro Caputo, Djalil Chafai

► To cite this version:

Charles Bordenave, Pietro Caputo, Djalil Chafai. Spectrum of large random reversible Markov chains. 2008. hal-00337409v1

HAL Id: hal-00337409

<https://hal.science/hal-00337409v1>

Preprint submitted on 6 Nov 2008 (v1), last revised 18 Mar 2010 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Spectrum of large random reversible Markov chains

Charles BORDENAVE, Pietro CAPUTO, Djalil CHAFAÏ

Preprint, November 2008

Abstract

In this work, we adopt a Random Matrix Theory point of view to study the spectrum of large reversible Markov chains in random environment. As the number of states tends to infinity, we consider both the almost sure global behavior of the spectrum, and the local behavior at the edge including the so called spectral gap. We study presently two simple models. The first one is on the complete graph while the second is on the chain graph (birth-and-death dynamics). These two models exhibit different scalings and limiting objects. The first model is related to the semi-circle law and Wigner's theorem. It contains as a special case a natural reversible Dirichlet Markov Ensemble. The second model is related to homogenization and also to asymptotics for the roots of random orthogonal polynomials. A special case gives rise to the arc-sine law as in a theorem by Erdős & Turán. This work raises several open problems.

Keywords: random matrices, random graphs, reversible Markov chains, random walks, random environment, spectral gap, Wigner semi-circle law, arc-sine law, tridiagonal matrices, birth-and-death processes, spectral analysis, orthogonal polynomials, weak and strong limit theorems, homogenization.

AMS-MS: 15A52; 60K37; 60F15; 62H99; 33C47; 37H10; 47B36.

1 Introduction

The spectral analysis of large dimensional random matrices is a very active domain of research, connected to a remarkable number of areas of Mathematics and their applications, see e.g. [51, 24, 35, 29, 2, 10, 28]. On the other hand, it is well known that the spectrum of reversible Markov chains provides useful information on their trend to equilibrium, see e.g. [40, 15, 37]. In the present work, we propose to study the spectrum of random reversible Markov chains with very large finite state space. The randomness of these chains, which appears as a random environment, is constructed by putting random weights on the edges of a finite graph. This approach raises a collection of stimulating open problems, lying at the interface between random matrix theory, random walks in random environment, and random graphs. We focus here on two elementary models, on the complete graph and on the chain graph. These two models exhibit different scalings and limiting objects. The study of spectral aspects of certain random Markov chains or random walks in random environment is not new, see for instance [20, 9, 53, 14, 13, 11, 45] and references therein. Here we adopt a particular Random Matrix Theory point of view.

More precisely, consider a finite connected undirected graph $G = (V, E)$, with vertex set V and edge set E , together with a set of weights, given by positive random variables

$$\mathbf{U} = \{U_{i,j}; \{i,j\} \in E\}.$$

Since the graph G is undirected we set $U_{i,j} = U_{j,i}$. On the network (G, \mathbf{U}) , we consider the random walk in random environment with state space V and transition probabilities

$$K_{i,j} = \frac{U_{i,j}}{\rho_i} \quad \text{where} \quad \rho_i = \sum_{j: \{i,j\} \in E} U_{i,j}. \quad (1)$$

The Markov kernel K is reversible with respect to the measure $\rho = \{\rho_i, i \in V\}$ in that

$$\rho_i K_{i,j} = \rho_j K_{j,i}$$

for all $i, j \in V$. Positivity of the weights implies that K is irreducible and ρ_i is its unique invariant distribution, up to normalization. When the variables \mathbf{U} are identically equal this is just the standard simple random walk on G , and $K - I$ is the associated Laplacian.

The construction of reversible Markov kernels from graphs with weighted edges as in (1) is classical in the Markovian literature, see e.g. [15, 23]. As for the choice of the graph $G = (V, E)$, we shall work with the simplest cases, namely the complete graph or a one-dimensional chain graph. Before passing to the precise description of models and results, let us briefly recall some broad facts.

By labeling the $n = |V|$ vertices of G and putting $K_{i,j} = 0$ if $\{i, j\} \notin E$, one can see K as a random $n \times n$ Markov matrix. The entries of K belong to $[0, 1]$ and each row sums up to 1. The spectrum of K does not depend on the way we label V . In general, even if the random weights \mathbf{U} are i.i.d. the random matrix K has non-independent entries due to the normalizing sums ρ_i . Note that K is in general non-symmetric, but by reversibility, it is symmetric in $L^2(\rho)$ and its spectrum $\sigma(K)$ is real. Moreover,

$$1 \in \sigma(K) \subset [-1, +1]$$

and it is convenient to denote the eigenvalues of K by

$$-1 \leq \lambda_n(K) \leq \dots \leq \lambda_1(K) = 1.$$

Since K is irreducible, the eigenspace of the largest eigenvalue 1 is one-dimensional and thus $\lambda_2(K) < 1$. Moreover, since K is reversible, the period of K is 1 or 2, and this last case is equivalent to $\lambda_n(K) = -1$ (the spectrum of K is in fact symmetric when K has period 2). For all these Markovian spectral aspects, see e.g. [41]. The *bulk behavior* of $\sigma(K)$ is studied via the Empirical Spectral Distribution (ESD)

$$\mu_K = \frac{1}{n} \sum_{k=1}^n \delta_{\lambda_k(K)}.$$

Since K is Markov, its ESD contains a probabilistic pathwise information on the corresponding random walk. Namely, the moments of the ESD μ_K satisfy, for any $\ell \in \mathbb{Z}_+$

$$\int_{-1}^{+1} x^\ell \mu_K(dx) = \frac{1}{n} \text{Tr}(K^\ell) = \frac{1}{n} \sum_{i \in V} r_\ell^{\mathbf{U}}(i) \quad (2)$$

where $r_\ell^{\mathbf{U}}(i)$ denotes the probability that the random walk on (G, \mathbf{U}) started at i returns to i after ℓ steps. The averaged version of (2) obtained by taking the expectation of both sides is also at the heart of the so called *trace-moments method* in Random Matrix Theory, see e.g. [2, 29].

The *edge behavior* of $\sigma(K)$ corresponds to the extreme eigenvalues $\lambda_2(K)$ and $\lambda_n(K)$, or more generally, to the k -extreme values $\lambda_2(K), \dots, \lambda_{k+1}(K)$ and $\lambda_n(K), \dots, \lambda_{n-k+1}(K)$. The geometric decay to the equilibrium measure ρ of the *continuous time* random walk with semigroup $(e^{t(K-I)})_{t \geq 0}$ generated by $K - I$ is governed by the so called *spectral gap*

$$\text{gap}(K - I) = 1 - \lambda_2(K).$$

When the period of K is 1, the relevant quantity for the *discrete time* random walk with kernel K is

$$\varsigma(K) = 1 - \max_{\substack{\lambda \in \sigma(K) \\ \lambda \neq 1}} |\lambda| = 1 - \max(-\lambda_n(K), \lambda_2(K)).$$

In that case, for any fixed value of n , we have $(K^\ell)_{i,\cdot} \rightarrow \rho$ as $\ell \rightarrow \infty$, for every $1 \leq i \leq n$. The reader may find an introduction to such aspects in [40].

Complete graph model

Here we set $V = \{1, \dots, n\}$ and $E = \{\{i, j\}; i, j \in V\}$. Note that we have self-loops at any vertex. The weights $U_{i,j}$, $1 \leq i \leq j \leq n$ are i.i.d. random variables with common law \mathcal{L} supported on $(0, \infty)$ (with no atom at 0). The law \mathcal{L} is independent of n . Without loss of generality, we assume that the marks \mathbf{U} come from the truncation of a single infinite triangular array $(U_{i,j})_{1 \leq i \leq j}$ of i.i.d. random variables of law \mathcal{L} . This defines a common probability space, which is convenient for almost sure convergence as $n \rightarrow \infty$.

Since all the entries of K are positive, the kernel K is irreducible and aperiodic (i.e. the period is 1). As a consequence, $\lambda_n(K) > -1$ and thus $\sigma(K)$ is of the form

$$-1 < \lambda_n(K) \leq \dots \leq \lambda_2(K) < \lambda_1(K) = 1.$$

As in the case of matrices with i.i.d. entries, the spectral properties of K will depend mostly on the tail of \mathcal{L} at infinity. When \mathcal{L} has finite mean $\int_0^\infty x \mathcal{L}(dx) = m$ we set $m = 1$. This is no loss of generality since K is invariant under the linear scaling $t \rightarrow t U_{i,j}$. If \mathcal{L} has a finite second moment we write $\sigma^2 = \int_0^\infty (x - 1)^2 \mathcal{L}(dx)$ for the variance. The rows of K are equally distributed (but not independent) and follow an exchangeable law on \mathbb{R}^n . Since each row sums up to one, we get by exchangeability that for every $1 \leq i, j \neq j' \leq n$,

$$\mathbb{E}(K_{i,j}) = \frac{1}{n} \quad \text{and} \quad \text{Cov}(K_{i,j}, K_{i,j'}) = -\frac{1}{n-1} \text{Var}(K_{1,1}).$$

However, the common variance of the entries of K may depend on \mathcal{L} . When \mathcal{L} is exponential then the rows of K are Dirichlet distributed. The following theorem states that if \mathcal{L} has finite positive variance $0 < \sigma^2 < \infty$, then the bulk of the spectrum of $\sqrt{n}K$ behaves as in the case of Wigner matrices, i.e. as if we had truly i.i.d. entries. The ESD of $\sqrt{n}K$ is

$$\mu_{\sqrt{n}K} = \frac{1}{n} \sum_{k=1}^n \delta_{\sqrt{n}\lambda_k(K)}.$$

Theorem 1.1 (Bulk behavior). *If \mathcal{L} has finite positive variance $0 < \sigma^2 < \infty$ then*

$$\mu_{\sqrt{n}K} \xrightarrow[n \rightarrow \infty]{w} \mathcal{W}_{2\sigma}$$

almost surely, where “ \xrightarrow{w} ” stands for weak convergence of probability measures and $\mathcal{W}_{2\sigma}$ is the Wigner semi-circle law with Lebesgue density

$$x \mapsto \frac{1}{2\pi\sigma^2} \sqrt{4\sigma^2 - x^2} \mathbb{1}_{[-2\sigma, +2\sigma]}(x). \quad (3)$$

The proof of Theorem 1.1, given in Section 2, relies on a uniform strong law of large numbers which allows to estimate $\rho_i = n(1 + o(1))$ and therefore yields a comparison of $\sqrt{n}K$ with a suitable Wigner matrix with i.i.d. entries. Note that even though

$$\lambda_1(\sqrt{n}K) = \sqrt{n} \rightarrow \infty \quad \text{as } n \rightarrow \infty, \quad (4)$$

the weak limit of $\mu_{\sqrt{n}K}$ is not affected since $\lambda_1(\sqrt{n}K)$ has weight $1/n$ in $\mu_{\sqrt{n}K}$. Theorem 1.1 implies that the bulk of $\sigma(K)$ collapses globally in a weak sense at speed $n^{-1/2}$. This result is too weak to provide the almost sure convergence of the extremal eigenvalues $\lambda_n(\sqrt{n}K)$ and $\lambda_2(\sqrt{n}K)$ to the edge of the limiting support $[-2\sigma, +2\sigma]$. We only get from Theorem 1.1 that almost surely, for every fixed $k \in \mathbb{Z}_+$,

$$\liminf_{n \rightarrow \infty} \sqrt{n}\lambda_{n-k}(K) \leq -2\sigma \quad \text{and} \quad \limsup_{n \rightarrow \infty} \sqrt{n}\lambda_{k+2}(K) \geq +2\sigma.$$

The weak convergence stated in Theorem 1.1 is relative to continuous bounded functions and cannot be used directly for polynomials. In particular the convergence of the moments of the ESD $\mu_{\sqrt{n}K}$ to the moments of $\mathcal{W}_{2\sigma}$ is not provided immediately by Theorem 1.1. Also, the identity (2) is useless for now. The control of the moments of the ESD $\mu_{\sqrt{n}K}$ requires the control of the extremal eigenvalues of $\sqrt{n}K$. Our next result below expresses that the convergence of the extremal eigenvalues of $\sqrt{n}K$ to the edge of the limiting support $[-2\sigma, +2\sigma]$ holds provided that \mathcal{L} has finite fourth moment (i.e. $\mathbb{E}(U_{1,1}^4) < \infty$).

Theorem 1.2 (Edge behavior). *If \mathcal{L} has finite positive variance $0 < \sigma^2 < \infty$ and finite fourth moment then almost surely, for any fixed $k \in \mathbb{Z}_+$,*

$$\lim_{n \rightarrow \infty} \sqrt{n}\lambda_{n-k}(K) = -2\sigma \quad \text{and} \quad \lim_{n \rightarrow \infty} \sqrt{n}\lambda_{k+2}(K) = +2\sigma.$$

In particular, almost surely,

$$\text{gap}(K - I) = 1 - \frac{2\sigma}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right) \quad \text{and} \quad \varsigma(K) = 1 - \frac{2\sigma}{\sqrt{n}} + o\left(\frac{1}{\sqrt{n}}\right). \quad (5)$$

The proof of Theorem 1.2, given in Section 2, relies on a rank one reduction (referred as a “Wedderburn rank one reduction” [50, p. 69] by G. H. Golub in Matrix Analysis) which will allow us to compare for instance $\lambda_2(\sqrt{n}K)$ with the largest eigenvalue of a Wigner matrix with centered entries. The approach requires refined versions of the uniform law of large numbers used in the proof of Theorem 1.1.

The edge behavior of Theorem 1.2 allows to reinforce the convergence of the bulk of Theorem 1.1 by providing the convergence of moments. Recall that for any integer $p \geq 1$,

the weak convergence together with the convergence of moments up to order p is equivalent to the convergence in Wasserstein W_p distance, see e.g. [49, 39]. For every real $p \geq 1$, the Wasserstein distance $W_p(\mu, \nu)$ between two probability measures μ, ν on \mathbb{R} is defined by

$$W_p(\mu, \nu) = \inf_{\Pi} \left(\int_{\mathbb{R} \times \mathbb{R}} |x - y|^p \Pi(dx, dy) \right)^{1/p} \quad (6)$$

where the infimum runs over the convex set of probability measures on $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$ with marginals μ and ν . Let $\tilde{\mu}_{\sqrt{n}K}$ be the trimmed ESD defined by

$$\tilde{\mu}_{\sqrt{n}K} = \frac{1}{n-1} \sum_{k=2}^n \delta_{\sqrt{n}\lambda_k(K)} = \frac{n}{n-1} \mu_{\sqrt{n}K} - \frac{1}{n-1} \delta_{\sqrt{n}}.$$

We have then the following Corollary of theorems 1.1 and 1.2, proved in Section 2.

Corollary 1.3 (Strong convergence). *If \mathcal{L} has finite positive variance and finite fourth moment then almost surely, for every $p \geq 1$,*

$$\lim_{n \rightarrow \infty} W_p(\tilde{\mu}_{\sqrt{n}K}, \mathcal{W}_{2\sigma}) = 0 \quad \text{while} \quad \lim_{n \rightarrow \infty} W_p(\mu_{\sqrt{n}K}, \mathcal{W}_{2\sigma}) = \begin{cases} 0 & \text{if } p < 2 \\ 1 & \text{if } p = 2 \\ \infty & \text{if } p > 2. \end{cases}$$

Recall that for every $k \in \mathbb{Z}_+$, the k^{th} moment of the Wigner semi-circle law $\mathcal{W}_{2\sigma}$ is zero if k is odd and is σ^k times the $(k/2)^{\text{th}}$ Catalan number if k is even. On the other hand, from (2), we know that for every $k \in \mathbb{Z}_+$, the k^{th} moment of the ESD $\mu_{\sqrt{n}K}$ writes

$$\int_{\mathbb{R}} x^k \mu_{\sqrt{n}K}(dx) = \frac{1}{n} \text{Tr} \left((\sqrt{n}K)^k \right) = n^{-1+\frac{k}{2}} \sum_{i=1}^n r_k^{\text{U}}(i)$$

where $r_k^{\text{U}}(i)$ be the probability that the random walk on V with kernel K started at i returns to i after k steps. Additionally, from (4) we get

$$\int_{\mathbb{R}} x^k \mu_{\sqrt{n}K}(dx) = n^{-1+\frac{k}{2}} + \left(1 - \frac{1}{n}\right) \int_{\mathbb{R}} x^k \tilde{\mu}_{\sqrt{n}K}(dx)$$

where $\tilde{\mu}_{\sqrt{n}K}$ is the trimmed ESD defined earlier. We can then state the following.

Corollary 1.4 (Return probabilities). *Let $r_k^{\text{U}}(i)$ be the probability that the random walk on V with kernel K started at i returns to i after k steps. If \mathcal{L} has finite positive variance $0 < \sigma^2 < \infty$ and finite fourth moment then almost surely, for every $k \in \mathbb{Z}_+$,*

$$\lim_{n \rightarrow \infty} n^{-1+\frac{k}{2}} \left(\sum_{i=1}^n r_k^{\text{U}}(i) - 1 \right) = \begin{cases} 0 & \text{if } k \text{ is odd} \\ \frac{\sigma^k}{\frac{k}{2}+1} \binom{k}{k/2} & \text{if } k \text{ is even.} \end{cases} \quad (7)$$

The sequence of Catalan numbers appears in the enumeration of many combinatorial objects. Here it is of particular interest that the r^{th} Catalan number $\frac{1}{r+1} \binom{2r}{r}$ is the number of non-negative simple paths of length r that start and end at 0.

We end up our analysis of the complete graph model with the behavior of the invariant probability distribution $\hat{\rho}$ of K , obtained by normalizing the invariant vector ρ as

$$\hat{\rho} = (\rho_1 + \cdots + \rho_n)^{-1}(\rho_1\delta_1 + \cdots + \rho_n\delta_n).$$

Let us denote by $\mathcal{U} = n^{-1}(\delta_1 + \cdots + \delta_n)$ the uniform law on $\{1, \dots, n\}$. We recall that the *total variation distance* $\|\mu - \nu\|_{\text{TV}}$ between two probability measures $\mu = \sum_{k=1}^n \mu_k \delta_k$ and $\nu = \sum_{k=1}^n \nu_k \delta_k$ on $\{1, \dots, n\}$ is given by

$$\|\mu - \nu\|_{\text{TV}} = \frac{1}{2} \sum_{k=1}^n |\mu_k - \nu_k|.$$

Theorem 1.5 (Invariant probability measure). *If \mathcal{L} has finite second moment, then a.s.*

$$\lim_{n \rightarrow \infty} \|\hat{\rho} - \mathcal{U}\|_{\text{TV}} = 0. \quad (8)$$

The proof of Theorem 1.5, given in Section 2, relies as before on a uniform law of large numbers. The speed of convergence and fluctuation of $\|\hat{\rho} - \mathcal{U}\|_{\text{TV}}$ depends on the tail of \mathcal{L} . The reader can find in Lemma 2.3 of Section 2 some estimates in this direction.

Recall that for a Markov matrix, the law \mathcal{U} is invariant iff the matrix is bistochastic, and is reversible iff the matrix is symmetric. Also, from this point of view, the random reversible Markov kernel K behaves asymptotically like a symmetric (and thus bistochastic) matrix. This is compatible with the Wigner-like behavior stated by Theorem 1.1.

Chain graph model (birth-and-death)

The complete graph model discussed earlier provides a random reversible Markov kernel which is irreducible and aperiodic. One of the key feature of this model lies in the fact that the degree of each vertex is n , which goes to infinity as $n \rightarrow \infty$. This property allows to use a law of large numbers to control the normalization ρ_i . The method will roughly still work if we replace the complete graphs sequence by a sequence of graphs for which the degrees are of order n . To go beyond this framework, it is natural to consider *local* models (i.e. *finite range* models) for which the degrees are uniformly bounded. We shall focus on a simple birth-and-death Markov kernel $K = (K_{i,j})_{1 \leq i,j \leq n}$ on $\{1, \dots, n\}$ given by

$$K_{i,i+1} = b_i, \quad K_{i,i} = a_i, \quad K_{i,i-1} = c_i$$

where $(a_i)_{1 \leq i \leq n}$, $(b_i)_{1 \leq i \leq n}$, $(c_i)_{1 \leq i \leq n}$ are in $[0, 1]$ with $c_1 = b_n = 0$, $b_i + a_i + c_i = 1$ for every $1 \leq i \leq n$, and $c_{i+1} > 0$ and $b_i > 0$ for every $1 \leq i \leq n-1$. In other words, we have

$$K = \begin{pmatrix} a_1 & b_1 & & & & \\ c_2 & a_2 & b_2 & & & \\ & c_3 & a_3 & b_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & c_{n-1} & a_{n-1} & b_{n-1} \\ & & & & c_n & a_n \end{pmatrix}. \quad (9)$$

The kernel K is irreducible, reversible, and every vertex has degree ≤ 3 . For an arbitrary $\rho_1 > 0$, the measure $\rho = \rho_1 \delta_1 + \dots + \rho_n \delta_n$ defined for every $2 \leq i \leq n$ by

$$\rho_i = \rho_1 \prod_{k=1}^{i-1} \frac{b_k}{c_{k+1}} = \rho_1 \frac{b_1 \cdots b_{i-1}}{c_2 \cdots c_i}$$

is invariant and reversible for K . In other words, for every $1 \leq i, j \leq n$,

$$\rho_i K_{i,j} = \rho_j K_{j,i}.$$

For every $1 \leq i \leq n$, the i^{th} row (c_i, a_i, b_i) of K belongs to the 3-dimensional simplex

$$\Lambda_3 = \{v \in [0, 1]^3; v_1 + v_2 + v_3 = 1\}.$$

For every $v \in \Lambda_3$, we define the left and right “reflections” $v^- \in \Lambda_3$ and $v^+ \in \Lambda_3$ of v by

$$v_- = (v_1 + v_3, v_2, 0) \quad \text{and} \quad v_+ = (0, v_2, v_1 + v_3).$$

The following result provides a general answer for the behavior of the bulk.

Theorem 1.6 (Global behavior for ergodic environment). *Let $\mathbf{p} : \mathbb{Z} \rightarrow \Lambda_3$ be an ergodic random field. Let K be the random birth-and-death kernel (9) on $\{1, \dots, n\}$ obtained from \mathbf{p} by taking for every $1 \leq i \leq n$*

$$(c_i, a_i, b_i) = \begin{cases} \mathbf{p}(i) & \text{if } 2 \leq i \leq n-1 \\ \mathbf{p}(1)_+ & \text{if } i = 1 \\ \mathbf{p}(n)_- & \text{if } i = n. \end{cases}$$

Then there exists a non-random probability measure μ on $[-1, +1]$ such that almost surely,

$$\lim_{n \rightarrow \infty} W_p(\mu_K, \mu) = 0$$

for every $p \geq 1$, where W_p is the Wasserstein distance (6). Moreover, for every $\ell \geq 0$,

$$\int_{-1}^{+1} x^\ell \mu(dx) = \mathbb{E}[r_\ell^{\mathbf{p}}(0)]$$

where $r_\ell^{\mathbf{p}}(0)$ is the probability of return to 0 in ℓ steps for the random walk on \mathbb{Z} with random environment \mathbf{p} . The expectation is taken with respect to the environment \mathbf{p} .

The proof of Theorem 1.6, given in Section 3, is a simple consequence of the ergodic theorem, see also [9] for an application to random conductance models. The reflective boundary condition is not necessary for this result on the bulk of the spectrum, and essentially any boundary condition (e.g. Dirichlet or periodic) produces the same limiting law, with essentially the same proof. Moreover, this result is not limited to the one-dimensional random walks and it remains valid for any *finite range reversible random walk with ergodic random environment on \mathbb{Z}^d* . However, and as we shall see below, a more precise analysis is possible for certain type of environments when $d = 1$.

Observe that a random conductance model on the chain graph can be obtained by defining K with (1) by putting i.i.d. positive weights \mathbf{U} of law \mathcal{L} on the edges of the chain

graph. For instance, if we remove the self-loops from the chain graph, this corresponds to define K by (9) with $a_1 = \dots = a_n = 0$, $b_1 = c_n = 1$, and, for every $2 \leq i \leq n-1$,

$$b_i = 1 - c_i = V_i = \frac{U_{i,i+1}}{U_{i,i+1} + U_{i,i-1}}.$$

where $(U_{i,i+1})_{i \geq 1}$ are i.i.d. random variables of law \mathcal{L} supported in $(0, \infty)$. The random variables V_1, \dots, V_n are dependent here. When \mathcal{L} is exponential, the V_i 's have a Beta law.

Let us consider now an alternative simple way to make K random. Namely, we use a sequence $(V_i)_{i \geq 1}$ of i.i.d. random variables on $[0, 1]$ of common law \mathcal{L} and we define the random birth-and-death Markov kernel K by (9) with

$$b_1 = c_n = 1 \quad \text{and} \quad b_i = 1 - c_i = V_i \quad \text{for every } 2 \leq i \leq n-1.$$

In other words, the random Markov kernel K is of the form

$$K = \begin{pmatrix} 0 & 1 & & & & \\ 1 - V_2 & 0 & V_2 & & & \\ & 1 - V_3 & 0 & V_3 & & \\ & & \ddots & \ddots & \ddots & \\ & & & 1 - V_{n-1} & 0 & V_{n-1} \\ & & & & 1 & 0 \end{pmatrix}. \quad (10)$$

This is not a random conductance model. However, this kernel is a particular case of the one appearing in Theorem 1.6, corresponding to the i.i.d. environment given by

$$\mathbf{p}(i) = (1 - V_i, 0, V_i)$$

for every $i \geq 1$. This allows to state the following Corollary of Theorem 1.6, which can be obtained alternatively by using the trace-moment method used in [38] for random symmetric tridiagonal matrices with i.i.d. entries, see Section 3. If \mathcal{L} has no atom at 0 and 1, then K is irreducible, reversible, 2-periodic, with spectrum $\sigma(K)$ of the form

$$-1 = \lambda_n(K) < \lambda_{n-1}(K) \leq \dots \leq \lambda_2(K) < \lambda_1(K) = 1, \quad (11)$$

and moreover $\sigma(K)$ is symmetric with respect to 0, with $0 \in \sigma(K)$ when n is odd.

Corollary 1.7 (Global behavior for i.i.d. environment). *Let K be the random birth-and-death Markov kernel (10) where $(V_i)_{i \geq 2}$ are i.i.d. of law \mathcal{L} on $[0, 1]$. Then there exists a non-random probability distribution μ on $[-1, +1]$ such that almost surely,*

$$\lim_{n \rightarrow \infty} W_p(\mu_K, \mu) = 0$$

for every $p \geq 1$, where W_p is the Wasserstein distance as in (6). The limiting spectral distribution μ is fully characterized by its sequence of moments, given for every $k \geq 1$ by

$$\int_{-1}^{+1} x^{2k-1} \mu(dx) = 0 \quad \text{and} \quad \int_{-1}^{+1} x^{2k} \mu(dx) = \sum_{\gamma \in D_k} \prod_{i \in \mathbb{Z}} \mathbb{E} \left(V^{N_\gamma(i)} (1 - V)^{N_\gamma(i-1)} \right)$$

where V is a random variable of law \mathcal{L} and where

$$D_k = \{\gamma = (\gamma_0, \dots, \gamma_{2k}) : \gamma_0 = \gamma_{2k} = 0, \text{ and } |\gamma_\ell - \gamma_{\ell+1}| = 1 \text{ for every } 0 \leq \ell \leq 2k-1\}$$

is the set of loop paths of length $2k$ of the simple random walk on \mathbb{Z} (walk bridge), and

$$N_\gamma(i) = \sum_{\ell=0}^{2k-1} \mathbb{1}_{\{(\gamma_\ell, \gamma_{\ell+1})=(i, i+1)\}}$$

is the number of times γ crosses the horizontal line $y = i + \frac{1}{2}$ in the increasing direction.

When the random variables $(V_i)_{i \geq 2}$ are only stationnary and ergodic, Corollary 1.7 remains valid provided that we adapt slightly the formula for the even moments of μ (more precisely, we just have to put the product inside the expectation).

For the complete graph model, Theorem 1.1 states that the weak limit of the ESD μ_K is trivial and that one has to consider instead the dilated ESD $\mu_{\sqrt{n}K}$. In contrast, for the models on the chain graph considered in Theorem 1.6 and Corollary 1.7, the ESD μ_K does not need to be scaled in order to converge to a non-trivial object.

Remark 1.8 (From Dirac masses to arc-sine laws). Corollary 1.7 gives a formula for the moments of μ . This formula is a series involving the “Beta-moments” of \mathcal{L} . We ignore how to explicitly compute these series when \mathcal{L} is an arbitrary law on $[0, 1]$. However, in the deterministic case where $\mathcal{L} = \delta_{1/2}$, we have, for every integer $k \geq 1$,

$$\int_{-1}^{+1} x^{2k} \mu(dx) = \sum_{\gamma \in D_k} 2^{-\sum_i N_\gamma(i) - \sum_i N_\gamma(i-1)} = 2^{-2k} \binom{2k}{k} = \int_{-1}^{+1} x^{2k} \frac{dx}{\pi \sqrt{1-x^2}}$$

which shows that μ is the arc-sine law on $[-1, +1]$ (see also Figure 4). More generally, a very similar computation reveals that if $\mathcal{L} = \delta_p$ with $0 < p < 1$ then μ is the arc-sine law on $[-2\sqrt{p(1-p)}, +2\sqrt{p(1-p)}]$. We give in figures 2-3-4 some simulations illustrating Corollary 1.7 for various other choices of \mathcal{L} .

Remark 1.9 (Universality). The law μ in Corollary 1.7 is not universal, in the sense that it depends on many “Beta-moments” of \mathcal{L} , in contrast with the complete graph case where the limiting spectral distribution depends on \mathcal{L} only via its first two moments.

We turn now to the edge behavior of $\sigma(K)$ where K is as in (10). The structure of $\sigma(K)$ given by (11) suggest to seek for the almost sure behavior of $\lambda_2(K)$ and $\lambda_{n-1}(K)$ as n goes to infinity. Since the limiting spectral distribution μ is symmetric, the convex hull of its support is of the form $[-\alpha_\mu, +\alpha_\mu]$ for some $0 \leq \alpha_\mu \leq 1$. The following result gives an interesting information on α_μ . More generally, the reader may forge many conjectures in the same spirit for the map $\mathcal{L} \mapsto \mu$ from the simulations given by figures 2-3-4.

Theorem 1.10 (Edge behavior for i.i.d. environment). Let K be the random birth-and-death Markov kernel (10) where $(V_i)_{i \geq 2}$ are i.i.d. of law \mathcal{L} on $[0, 1]$. Let μ be the symmetric limiting spectral distribution on $[-1, +1]$ which appears in Corollary 1.7. Let $[-\alpha_\mu, +\alpha_\mu]$ be the convex hull of the support of μ . If \mathcal{L} has a positive density at $1/2$ then $\alpha_\mu = 1$. Consequently, almost surely,

$$\lambda_2(K) = -\lambda_{n-1}(K) = 1 + o(1).$$

Additionally, if \mathcal{L} is supported on $[0, t]$ with $0 < t < 1/2$ or on $[t, 1]$ with $1/2 < t < 1$ then almost surely $\limsup_{n \rightarrow \infty} \lambda_2(K) < 1$ and therefore $\alpha_\mu < 1$.

The proof of Theorem 1.10 is given in Section 3. The speed of convergence of $\lambda_2(K) - 1$ to 0 is highly dependent on the choice of the law \mathcal{L} . More precisely, if

$$\mathbb{E} \left[\log \frac{V}{1-V} \right] = 0 \quad \text{and} \quad \mathbb{E} \left[\left(\log \frac{V}{1-V} \right)^2 \right] > 0$$

where $V \sim \mathcal{L}$, then K is the so called Sinai random walk on $\{1, \dots, n\}$, see e.g. [14]. In this case, by a slight modification of the analysis of [14], one can prove that almost surely,

$$-\infty < \liminf_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \log(1 - \lambda_2(K)) \leq \limsup_{n \rightarrow \infty} \frac{1}{\sqrt{n}} \log(1 - \lambda_2(K)) < 0.$$

Thus, the convergence to the edge here occurs exponentially fast in \sqrt{n} . On the other hand, if for instance $\mathcal{L} = \delta_{1/2}$ (simple reflected random walk on $\{1, \dots, n\}$) then it is known that $1 - \lambda_2(K)$ decays as n^{-2} only. Even slower decay can be obtained with $\mathcal{L} = (\delta_{1/2} + \delta_{3/4})/2$.

We end up with an interpretation of Corollary 1.7 in terms of the roots of certain random orthogonal polynomials. The proof of Corollary 1.11 is given in Section 3.

Corollary 1.11 (Random orthogonal polynomials). *Let $(V_n)_{n \geq 2}$ be a sequence of i.i.d. random variables with law \mathcal{L} supported in $(0, 1)$. Set $V_1 = 1$ and let $(Q_n)_{n \geq 0}$ be the random sequence of polynomials defined by the recursive relation*

$$Q_{n+1}(x) = xQ_n(x) - V_n(1 - V_{n+1})Q_{n-1}(x)$$

for every $n \geq 1$, with initial values $Q_0 \equiv 1$ and $Q_1(x) = x$. Let $(x_{n,k})_{1 \leq k \leq n}$ be the roots of Q_n and $\mu_{Q_n} = \frac{1}{n} \sum_{k=1}^n \delta_{x_{n,k}}$. Then almost surely,

$$\lim_{n \rightarrow \infty} W_p(\mu_{Q_n}, \mu) = 0$$

for every $p \geq 1$, where W_p is the Wasserstein distance as in (6) and where μ is the non-random probability measure on $[-1, +1]$ depending on \mathcal{L} and given by Corollary 1.7.

From the classical theory of orthogonal polynomials, $(Q_n)_{n \geq 0}$ is the sequence of orthogonal polynomials of a compactly supported law which is random, see Section 3. In the simple case where $\mathcal{L} = \delta_{1/2}$, it turns out that $(Q_n)_{n \geq 0}$ are the Chebyshev polynomials of the first kind, which are orthogonal with respect to the arc-sine law on $[-1, +1]$. In this case, the roots $(x_{n,i})_{1 \leq i \leq n}$ of Q_n are explicitly known:

$$x_{n,i} = \cos \left(\frac{2i-1}{n} \frac{\pi}{2} \right), \quad 1 \leq i \leq n.$$

On the other hand, we know by Remark 1.8 that μ is the arc-sine law on $[-1, +1]$. In this example, the limit of μ_{Q_n} coincides with the law for which $(Q_n)_{n \geq 0}$ are orthogonal, and the arc-sine law on $[-1, +1]$ appears as a sort of fixed point. This is not the case for more general \mathcal{L} , see Section 3 for more explanations.

Further remarks and open problems

We have defined random reversible Markov kernels with n states by using the complete or the chain graph with n vertices and random weights on the edges. The underlying graph appears as a prescribed skeleton for these Markov kernels. One can explore several other topologies. One can also think about making the skeleton itself random, by using a random graph such as random k -regular graphs. We mention below some other open problems together with remarks on related models considered in the literature.

Fluctuations at the edge

An interesting problem concerns the fluctuations of $\lambda_2(\sqrt{n}K)$ around its limiting value 2σ in the complete graph model. Under suitable conditions on \mathcal{L} , one may seek for a deterministic sequence (a_n) , and a probability distribution \mathcal{D} on \mathbb{R} such that

$$a_n (\lambda_2(\sqrt{n}K) - 2\sigma) \xrightarrow[n \rightarrow \infty]{d} \mathcal{D}$$

where “ \xrightarrow{d} ” stands for convergence in distribution. The same may be asked for the random variable $\lambda_n(\sqrt{n}K) + 2\sigma$. Computer simulations (see Figure 1) suggest that $a_n \approx n^{2/3}$ and that \mathcal{D} is close to a Tracy-Widom distribution. The heuristics here is that $\lambda_2(\sqrt{n}K)$ behaves like the λ_1 of a centered Gaussian random symmetric matrix. The difficulty is that the entries of K are not i.i.d., not centered, and of course not Gaussian. At the time of writing, the fluctuations of the extremal eigenvalues of random matrices is still an active domain of research, even in the i.i.d. case, see e.g. [26, 42, 47].

Example of symmetric Markov generator

Rather than considering the random walk with infinitesimal generator $K - I$ on the complete graph as we did, one may start with the *symmetric* infinitesimal generator G defined by $G_{i,j} = G_{j,i} = U_{i,j}$ for every $1 \leq i < j \leq n$ and $G_{i,i} = -\sum_{j \neq i} G_{i,j}$ for every $1 \leq i \leq n$. Here $(U_{i,j})_{1 \leq i < j}$ is a triangular array of i.i.d. real random variables of law \mathcal{L} . For this model, the uniform probability measure \mathcal{U} is reversible and invariant. Such kind of random matrices have been investigated in [16]. In particular, if \mathcal{L} has first moment $m > 0$ and variance $\sigma^2 = 1$ then from [16, Theorem 1.3] we get that a.s. the limiting spectral distribution of the random symmetric matrix

$$n^{-1/2}(G - (n-2)mI - m1)$$

is the free convolution of a Wigner semi-circle law with a standard Gaussian law. Here 1 is the $n \times n$ matrix full of 1's. This limiting spectral distribution is a symmetric law on \mathbb{R} with smooth bounded density of unbounded support. This is not a complete surprise since the diagonal of G is asymptotically independent of the off-diagonal entries of G , and gives rise to a Gaussian law by the central limit theorem. By using perturbative arguments (rank lemma), one may show that the same result holds for the empirical spectral distribution

$$\delta_{-n^{-1/2}(n-2)m} * \frac{1}{n} \sum_{k=1}^n \delta_{\lambda_k(n^{-1/2}G)}.$$

Non-reversible Markov ensembles

A non-reversible model is obtained when the underlying complete graph is oriented. That is each vertex i has now (besides the self-loop) $n - 1$ outgoing edges (i, j) and $n - 1$ incoming edges (j, i) . On each of these edges we place an independent positive weight $V_{i,j}$ with law \mathcal{L} , and on each self-loop an independent positive weight $V_{i,i}$ with law \mathcal{L} . This gives us a non-reversible stochastic matrix

$$\tilde{K}_{i,j} = \frac{V_{i,j}}{\sum_{k=1}^n V_{i,k}}.$$

The spectrum of \tilde{K} is now complex. If \mathcal{L} is exponential, then the matrix \tilde{K} describes the Dirichlet Markov Ensemble considered in [18]. Numerical simulations suggest that if \mathcal{L} has say finite positive variance then with probability one, the complex ESD of $n^{1/2}\tilde{K}$ converges weakly as $n \rightarrow \infty$ to the uniform law on the unit disc of the complex plane (known as the circle or circular law). At the time of writing, this conjecture is still open, see e.g. [19]. It was shown in [17] that if \mathcal{L} has positive variance and finite fourth moment, then with probability one, the complex ESD of $(n^{-1/2}V_{i,j})_{1 \leq i,j \leq n}$ converges weakly as $n \rightarrow \infty$ to the circular law. In fact, for this i.i.d. case, a finite positive variance is enough, as shown recently in [46].

Geometric point of view

The set of Markov kernels on $V = \{1, \dots, n\}$ is a convex and compact polytope of \mathbb{R}^{n^2} with $n(n-1)$ degrees of freedom. Additionally, K belongs to the interior of this polytope iff K has positive entries. Such kernels are irreducible. By a famous reversibility criterion due to Kolmogorov, see e.g. [34, Theorem 1.7 page 21], these kernels are reversible iff they belong to the intersection of polynomial hypersurfaces. This shows that considering random reversible Markov kernels corresponds to defining a probability measure on a non-linear non-convex compact set. Note however that the set of Markov kernels which are reversible for a fixed probability measure is convex.

Heavy-tailed weights

Remarkable works has been devoted in the recent years to the spectral analysis of large dimensional symmetric random matrices with heavy-tailed i.i.d. entries, see e.g. [43, 1, 5, 52, 8]. Similarly, on the complete graph, one may consider the bulk and edge behavior of the random reversible Markov kernels constructed by (1) when the law \mathcal{L} of the weights is heavy-tailed (i.e. with at least an infinite second moment). In that case, and in contrast with Theorem 1.1, the scaling is not \sqrt{n} and the limiting spectral distribution is not the Wigner semi-circle law. We address such heavy-tailed models in a forthcoming paper in preparation [12]. Another interesting model is the so called *trap model* which corresponds to put heavy-tailed weights only on the diagonal of \mathbf{U} (holding times), see e.g. [13] for some recent advances.

2 Proofs for the complete graph model

Here we prove theorems 1.1, 1.2, 1.5 and Corollary 1.3. Recall that K is self-adjoint as an operator from $L^2(\rho)$ to $L^2(\rho)$. Here $L^2(\rho)$ denotes \mathbb{R}^n equipped with the scalar product

$$\langle x, y \rangle_\rho = \sum_{i=1}^n \rho_i x_i y_i.$$

Let us also denote by $L^2(1)$ the usual Euclidean Hilbert space \mathbb{R}^n with the scalar product

$$\langle x, y \rangle = \sum_{i=1}^n x_i y_i.$$

The following simple lemma allows us to work with symmetric matrices when needed.

Lemma 2.1 (Spectral equivalence). *Almost surely, the spectrum of the reversible Markov matrix K coincides with the spectrum of the symmetric matrix S defined by*

$$S_{i,j} = \sqrt{\frac{\rho_i}{\rho_j}} K_{i,j} = \frac{U_{i,j}}{\sqrt{\rho_i \rho_j}}.$$

Moreover, the corresponding eigenspaces dimensions also coincide.

Proof. It suffices to observe that the map $x \mapsto \hat{x}$ defined by

$$\hat{x} = (x_1 \sqrt{\rho_1}, \dots, x_n \sqrt{\rho_n})$$

is an isometry from $L^2(\rho)$ to $L^2(1)$ and that for any $x, y \in \mathbb{R}^n$ and $1 \leq i \leq n$, we have

$$(Kx)_i = \sum_{j=1}^n K_{i,j} x_j$$

and

$$\langle Kx, y \rangle_\rho = \sum_{i,j=1}^n K_{i,j} x_j y_i \rho_i = \sum_{i,j=1}^n U_{i,j} x_j y_i = \sum_{i,j=1}^n S_{i,j} \hat{x}_i \hat{y}_j = \langle S\hat{x}, \hat{y} \rangle.$$

□

The random symmetric matrix S has non-centered, non-independent entries. Each entry of S is bounded and belongs to the interval $[0, 1]$, since for every $1 \leq i, j \leq n$, we have $S_{i,j} \leq U_{i,j} / \sqrt{U_{i,j} U_{j,i}} = 1$. The off-diagonal entries of S are distributed as

$$\frac{\xi_0}{\sqrt{(\xi_{-(n-1)} + \dots + \xi_0)(\xi_0 + \dots + \xi_{n-1})}},$$

where $\xi_{-(n-1)}, \dots, \xi_0, \dots, \xi_{n-1}$ are $2n-1$ i.i.d. random variables with law \mathcal{L} . The diagonal entries $S_{i,i} = K_{i,i}$ are distributed as $\xi_1 / (\xi_1 + \dots + \xi_n)$ where ξ_1, \dots, ξ_n are n i.i.d. random variables with law \mathcal{L} . In the sequel, for any $n \times n$ real symmetric matrix A , we denote by

$$\lambda_n(A) \leq \dots \leq \lambda_1(A)$$

its ordered spectrum. We shall also denote by $\|A\|$ the operator norm of A , defined by

$$\|A\|^2 = \max_{x \in \mathbb{R}^n} \frac{\langle Ax, Ax \rangle}{\langle x, x \rangle}.$$

Clearly, $\|A\| = \max(\lambda_1(A), -\lambda_n(A))$. To prove Theorem 1.1 we shall compare the symmetric random matrix $\sqrt{n}S$ with the symmetric $n \times n$ random matrices

$$W_{i,j} = \frac{U_{i,j} - 1}{\sqrt{n}} \quad \text{and} \quad \widetilde{W}_{i,j} = \frac{U_{i,j}}{\sqrt{n}}. \quad (12)$$

Note that W defines a so called Wigner matrix, i.e. W is symmetric and it has centered i.i.d. entries with finite positive variance. We shall also need the non-centered matrix \widetilde{W} . It is well known that under the sole assumption $\sigma^2 \in (0, \infty)$ on \mathcal{L} , almost surely,

$$\mu_W \xrightarrow[n \rightarrow \infty]{w} \mathcal{W}_{2\sigma} \quad \text{and} \quad \mu_{\widetilde{W}} \xrightarrow[n \rightarrow \infty]{w} \mathcal{W}_{2\sigma}$$

where μ_W and $\mu_{\widetilde{W}}$ are the ESD of W and \widetilde{W} , see e.g. [2, Theorems 2.1 and 2.12]. Note that \widetilde{W} is a rank one perturbation of W , which implies that the spectra of W and \widetilde{W} are interlaced (Weyl-Poincaré inequalities, see e.g. [31, 2]). Moreover, under the assumption of finite fourth moment on \mathcal{L} , it is known that almost surely

$$\lambda_n(W) \rightarrow -2\sigma \quad \text{and} \quad \lambda_1(W) \rightarrow +2\sigma.$$

In particular, almost surely,

$$\|W\| = 2\sigma + o(1). \quad (13)$$

On the other hand, and still under the finite fourth moment assumption, almost surely,

$$\lambda_1(\widetilde{W}) \rightarrow +\infty \quad \text{while} \quad \lambda_2(\widetilde{W}) \rightarrow +2\sigma \quad \text{and} \quad \lambda_n(\widetilde{W}) \rightarrow -2\sigma$$

see e.g. [3, 27, 2]. Heuristically, when n is large, the law of large numbers implies that ρ_i is close to n (recall that here \mathcal{L} has mean 1), and thus $\sqrt{n}S$ is close to \widetilde{W} . The main tools needed for comparison of the matrix $\sqrt{n}S$ with \widetilde{W} are listed in the following lemmas. We begin with a Kolmogorov-Marcinkiewicz-Zygmund strong uniform law of large numbers, connected to Baum-Katz type theorems.

Lemma 2.2 (Uniform law of large numbers). *Let $(A_{i,j})_{i,j \geq 1}$ be a symmetric array of i.i.d. random variables. For any reals $a > 1/2$, $b \geq 0$, and $M > 0$, if $\mathbb{E}(|A_{1,1}|^{(1+b)/a}) < \infty$ then*

$$\max_{1 \leq i \leq Mn^b} \left| \sum_{j=1}^n (A_{i,j} - c) \right| = o(n^a) \quad \text{a.s.} \quad \text{where} \quad c = \begin{cases} \mathbb{E}(A_{1,1}) & \text{if } a \leq 1 \\ \text{any number} & \text{if } a > 1. \end{cases}$$

Proof. This Kolmogorov-Marcinkiewicz-Zygmund strong uniform law of large numbers is proved in [4, Lemma 2] for a non-symmetric array. The symmetry makes the random variables $(\sum_{j=1}^n A_{i,j})_{i \geq 1}$ dependent, but a careful analysis of the argument shows that this is not a problem except for a sort of converse, see [4, Lemma 2] for details. \square

Lemma 2.3 (Almost sure bounds). *If \mathcal{L} has finite moment of order $\kappa \in [1, 2]$ then*

$$\max_{1 \leq i \leq n^{\kappa-1}} \left| \frac{\rho_i}{n} - 1 \right| = o(1) \quad (14)$$

almost surely, and in particular, if \mathcal{L} has finite second moment, then almost surely

$$\max_{1 \leq i \leq n} \left| \frac{\rho_i}{n} - 1 \right| = o(1). \quad (15)$$

Moreover if \mathcal{L} has finite moment of order κ with $2 \leq \kappa < 4$, then almost surely

$$\max_{1 \leq i \leq n} \left| \frac{\rho_i}{n} - 1 \right| = o(n^{\frac{2-\kappa}{\kappa}}). \quad (16)$$

Additionally, if \mathcal{L} has finite fourth moment, then almost surely

$$\sum_{i=1}^n \left(\frac{\rho_i}{n} - 1 \right)^2 = O(1). \quad (17)$$

Proof. Proof of (14-15-16). The result (14) follows from Lemma 2.2 with

$$A_{i,j} = U_{i,j}, \quad a = M = 1, \quad b = \kappa - 1.$$

We recover the standard strong law of large numbers with $\kappa = 1$. The result (16) follows from Lemma 2.2 with this time (we recover (15) with $\kappa = 2$)

$$A_{i,j} = U_{i,j}, \quad a = 2/\kappa, \quad b = M = 1.$$

Proof of (17). We set $\epsilon_i = n^{-1}\rho_i - 1$ for every $1 \leq i \leq n$. Since \mathcal{L} has finite fourth moment, the result (13) for the centered Wigner matrix W defined by (12) gives that

$$\sum_{i=1}^n \epsilon_i^2 = \frac{\langle W1, W1 \rangle}{\langle 1, 1 \rangle} \leq \|W\|^2 = 4\sigma^2 + o(1) = O(1)$$

almost surely. Here $\langle \cdot, \cdot \rangle$ is the Euclidean dot product of \mathbb{R}^n . \square

We are now able to give a proof of Theorem 1.5.

Proof of Theorem 1.5. Since \mathcal{L} has finite first moment, by the strong law of large numbers,

$$\rho_1 + \cdots + \rho_n = \sum_{i=1}^n U_{i,i} + 2 \sum_{1 \leq i < j \leq n} U_{i,j} = n^2(1 + o(1))$$

almost surely. For every fixed $i \geq 1$, we have also $\rho_i = n(1 + o(1))$ almost surely. As a consequence, for every fixed $i \geq 1$, almost surely,

$$\hat{\rho}_i = \frac{\rho_i}{\rho_1 + \cdots + \rho_n} = \frac{n(1 + o(1))}{n^2(1 + o(1))} = \frac{1}{n}(1 + o(1)).$$

Moreover, since \mathcal{L} has finite second moment, the $o(1)$ in the right hand side above is uniform over $1 \leq i \leq n$ thanks to (15) of Lemma 2.3. This achieves the proof. \square

Note that $\hat{\rho}_i = n^{-1}(1 + O(\delta))$ for every $1 \leq i \leq n$, where

$$\delta = \max_{1 \leq i \leq n} |\epsilon_i| = o(1) \quad \text{with} \quad \epsilon_i = n^{-1}\rho_i - 1. \quad (18)$$

We will repeatedly use these notations in the sequel. Let us consider now Theorem 1.1. Lemma 2.1 reduces Theorem 1.1 to the study of the ESD of $\sqrt{n}S$. Any couple of entries of S are dependent. One can find in the literature many extensions of the Wigner theorem for random symmetric matrices with non-i.i.d. entries. However, these results are irrelevant for our model, or do not apply directly, or do not provide the almost-sure convergence.

Proof of Theorem 1.1. We first recall a standard fact about comparison of spectral densities of symmetric matrices. Let $L(F, G)$ denote the Lévy distance between two cumulative distribution functions F and G on \mathbb{R} , defined by

$$L(F, G) = \inf\{\varepsilon > 0 \text{ such that } F(\cdot - \varepsilon) - \varepsilon \leq G \leq F(\cdot + \varepsilon) + \varepsilon\}.$$

It is well known [6] that the Lévy distance is a metric for weak convergence of probability distributions on \mathbb{R} . If F_A and F_B are the cumulative distribution functions of the empirical

spectral distributions of two hermitian $n \times n$ matrices A and B , we have the following bound for the third power of $L(F_A, F_B)$ in terms of the trace of $(A - B)^2$:

$$L^3(F_A, F_B) \leq \frac{1}{n} \text{Tr}((A - B)^2) = \frac{1}{n} \sum_{i,j=1}^n (A_{i,j} - B_{i,j})^2. \quad (19)$$

The proof of this estimate is a consequence of the Hoffman-Wielandt inequality [30], see also [2, Lemma 2.3]. By Lemma 2.1, we have $\sqrt{n}\lambda_k(K) = \lambda_k(\sqrt{n}S)$ for every $1 \leq k \leq n$. We shall use the bound (19) for the matrices $A = \sqrt{n}S$ and $B = \widetilde{W}$, where \widetilde{W} is defined in (12). We will show that a.s.

$$\frac{1}{n} \sum_{i,j=1}^n (A_{i,j} - B_{i,j})^2 = O(\delta^2), \quad (20)$$

where $\delta = \max_i |\epsilon_i|$ as in (18). Since \mathcal{L} has finite positive variance, we know that the ESD of B tends weakly as $n \rightarrow \infty$ to the Wigner semi-circle law on $[-2\sigma, +2\sigma]$. Therefore the bound (20), with (19) and the fact that $\delta \rightarrow 0$ as $n \rightarrow \infty$ is sufficient to prove the theorem. We turn to a proof of (20). For every $1 \leq i, j \leq n$, we have

$$A_{i,j} - B_{i,j} = \frac{U_{i,j}}{\sqrt{n}} \left(\frac{n}{\sqrt{\rho_i \rho_j}} - 1 \right).$$

Set, as usual $\rho_i = n(1 + \epsilon_i)$ and define $\psi_i = (1 + \epsilon_i)^{-\frac{1}{2}} - 1$. Note that by Lemma 2.3, almost surely, $\psi_i = O(\delta)$ uniformly in $i = 1, \dots, n$. Also,

$$\frac{n}{\sqrt{\rho_i \rho_j}} - 1 = (1 + \psi_i)(1 + \psi_j) - 1 = \psi_i + \psi_j + \psi_i \psi_j.$$

In particular, $\frac{n}{\sqrt{\rho_i \rho_j}} - 1 = O(\delta)$. Therefore

$$\frac{1}{n} \sum_{i,j=1}^n (A_{i,j} - B_{i,j})^2 \leq O(\delta^2) \left(\frac{1}{n^2} \sum_{i,j=1}^n U_{i,j}^2 \right).$$

By the strong law of large numbers, $\frac{1}{n^2} \sum_{i,j=1}^n U_{i,j}^2 \rightarrow \sigma^2 + 1$ a.s., which implies (20). \square

Next, we turn to the proof of Theorem 1.2 which concerns the edge of $\sigma(\sqrt{n}S)$.

Proof of Theorem 1.2. Thanks to Lemma 2.1 and the global behavior proven in Theorem 1.1, it is enough to show that, almost surely,

$$\limsup_{n \rightarrow \infty} \sqrt{n} \max(|\lambda_2(S)|, |\lambda_n(S)|) \leq 2\sigma.$$

The eigenspace of S of the eigenvalue 1 is almost surely of dimension 1, and is given by $\mathbb{R}(\sqrt{\rho_1}, \dots, \sqrt{\rho_n})$. Let P be the orthogonal projector on $\mathbb{R}\sqrt{\rho}$. The matrix P is $n \times n$ symmetric of rank 1, and for every $1 \leq i, j \leq n$,

$$P_{i,j} = \frac{\sqrt{\rho_i \rho_j}}{\sum_{k=1}^n \rho_k}.$$

The spectrum of the symmetric matrix $S - P$ is

$$\{\lambda_n(S), \dots, \lambda_2(S)\} \cup \{0\}.$$

By subtracting P from S we remove the largest eigenvalue 1 from the spectrum, without touching the remaining eigenvalues. Let \mathcal{V} be the random set of vectors of unit Euclidean norm of \mathbb{R}^n which are orthogonal to $\sqrt{\rho}$ for the scalar product $\langle \cdot, \cdot \rangle$ of \mathbb{R}^n . We have then

$$\sqrt{n} \max(|\lambda_2(S)|, |\lambda_n(S)|) = \max_{v \in \mathcal{V}} |\langle \sqrt{n} S v, v \rangle| = \max_{v \in \mathcal{V}} |\langle \tilde{A} v, v \rangle|$$

where \tilde{A} is the $n \times n$ random symmetric matrix defined by

$$\tilde{A}_{i,j} = \sqrt{n}(S - P)_{i,j} = \sqrt{n} \left(\frac{U_{i,j}}{\sqrt{\rho_i \rho_j}} - \frac{\sqrt{\rho_i \rho_j}}{\sum_{k=1}^n \rho_k} \right).$$

In Lemma 2.4 below we establish that almost surely $\langle v, (\tilde{A} - W)v \rangle = O(\delta) + O(n^{-1/2})$ uniformly in $v \in \mathcal{V}$, where W is defined in (12) and δ is given by (18). Thus, using (13),

$$|\langle W v, v \rangle| \leq \max(|\lambda_1(W)|, |\lambda_n(W)|) = 2\sigma + o(1),$$

we obtain that almost surely, uniformly in $v \in \mathcal{V}$,

$$|\langle \tilde{A} v, v \rangle| \leq |\langle W v, v \rangle| + |\langle (\tilde{A} - W)v, v \rangle| = 2\sigma + o(1) + O(\delta).$$

Thanks to Lemma 2.3 we know that $\delta = o(1)$ and the theorem follows. \square

Lemma 2.4. *Almost surely, uniformly in $v \in \mathcal{V}$, we have, with $\delta := \max_i |\epsilon_i|$,*

$$\langle v, (\tilde{A} - W)v \rangle = O(\delta) + O(n^{-1/2}).$$

Proof. We start by rewriting the matrix

$$\tilde{A}_{i,j} = \frac{\sqrt{n} U_{i,j}}{\sqrt{\rho_i \rho_j}} - \frac{\sqrt{n} \sqrt{\rho_i \rho_j}}{\sum_k \rho_k}$$

by expanding around the law of large numbers. We set $\rho_i = n(1 + \epsilon_i)$ and we define

$$\varphi_i = \sqrt{1 + \epsilon_i} - 1 \quad \text{and} \quad \psi_i = \frac{1}{\sqrt{1 + \epsilon_i}} - 1.$$

Observe that φ_i and ψ_i are of order ϵ_i and by Lemma 2.3, cf. (17) we have a.s.

$$\langle \varphi, \varphi \rangle = O(1) \quad \text{and} \quad \langle \psi, \psi \rangle = O(1). \quad (21)$$

We expand

$$\sqrt{\rho_i \rho_j} = n(1 + \epsilon_i)^{\frac{1}{2}}(1 + \epsilon_j)^{\frac{1}{2}} = n(1 + \varphi_i)(1 + \varphi_j).$$

Similarly, we have

$$\frac{1}{\sqrt{\rho_i \rho_j}} = n^{-1}(1 + \psi_i)(1 + \psi_j).$$

Moreover, writing

$$\sum_{k=1}^n \rho_k = n^2 \left(1 + \frac{1}{n} \sum_k \epsilon_k \right)$$

and setting $\gamma := (1 + \frac{1}{n} \sum_k \epsilon_k)^{-1} - 1$ we see that

$$\left(\sum_{k=1}^n \rho_k \right)^{-1} = \frac{1}{n^2} (1 + \gamma).$$

Note that $\gamma = O(\delta)$. Using these expansions we obtain

$$\frac{\sqrt{n} U_{i,j}}{\sqrt{\rho_i \rho_j}} = \frac{1}{\sqrt{n}} U_{i,j} (1 + \psi_i)(1 + \psi_j)$$

and

$$\frac{\sqrt{n} \sqrt{\rho_i \rho_j}}{\sum_k \rho_k} = \frac{1}{\sqrt{n}} (1 + \varphi_i)(1 + \varphi_j)(1 + \gamma).$$

From these expressions, with the definitions

$$\Phi_{i,j} = \varphi_i + \varphi_j + \varphi_i \varphi_j \quad \text{and} \quad \Psi_{i,j} = \psi_i + \psi_j + \psi_i \psi_j,$$

we obtain

$$\tilde{A}_{i,j} = W_{i,j} (1 + \Psi_{i,j}) + \frac{1}{\sqrt{n}} [\Psi_{i,j} - \Phi_{i,j} (1 + \gamma) + \gamma].$$

Therefore, we have

$$\langle v, (W - \tilde{A})v \rangle = - \sum_{i,j} v_i W_{i,j} \Psi_{i,j} v_j + \frac{1 + \gamma}{\sqrt{n}} \langle v, \Phi v \rangle - \frac{1}{\sqrt{n}} \langle v, \Psi v \rangle - \frac{\gamma}{\sqrt{n}} \langle v, 1 \rangle^2.$$

Let us first show that

$$\langle v, 1 \rangle = O(1). \tag{22}$$

Indeed, $v \in \mathcal{V}$ implies that for any $c \in \mathbb{R}$,

$$\langle v, 1 \rangle = \langle v, 1 - c\sqrt{\rho} \rangle.$$

Taking $c = 1/\sqrt{n}$ we see that

$$1 - c\sqrt{\rho_i} = 1 - \sqrt{1 + \epsilon_i} = -\varphi_i.$$

Thus, Cauchy–Schwarz’ inequality implies

$$\langle v, 1 \rangle^2 \leq \langle v, v \rangle \langle \varphi, \varphi \rangle$$

and (22) follows from (21) above. Next, we show that

$$\langle v, \Phi v \rangle = O(1). \tag{23}$$

Note that

$$\langle v, \Phi v \rangle = 2\langle v, 1 \rangle \langle v, \varphi \rangle + \langle v, \varphi \rangle^2.$$

Since $\langle v, \varphi \rangle^2 \leq \langle v, v \rangle \langle \varphi, \varphi \rangle$ we see that (23) follows from (21) and (22). In the same way we obtain that $\langle v, \Psi v \rangle = O(1)$. So far we have obtained the estimate

$$\langle v, (W - \tilde{A})v \rangle = - \sum_{i,j} v_i W_{i,j} \Psi_{i,j} v_j + O(n^{-1/2}). \quad (24)$$

To bound the first term above we observe that

$$\begin{aligned} \sum_{i,j} v_i W_{i,j} \Psi_{i,j} v_j &= 2 \sum_i \psi_i v_i (Wv)_i + \sum_{i,j} \psi_i v_i W_{i,j} \psi_j v_j \\ &= 2 \langle \hat{\psi}, Wv \rangle + \langle \hat{\psi}, W\hat{\psi} \rangle, \end{aligned}$$

where $\hat{\psi}$ denotes the vector $\hat{\psi}_i := \psi_i v_i$. Note that

$$\langle \hat{\psi}, \hat{\psi} \rangle = \sum_i \psi_i^2 v_i^2 \leq O(\delta^2) \langle v, v \rangle = O(\delta^2).$$

Therefore, by definition of the norm $\|W\|$

$$|\langle \hat{\psi}, W\hat{\psi} \rangle| \leq \sqrt{\langle \hat{\psi}, \hat{\psi} \rangle} \sqrt{\langle W\hat{\psi}, W\hat{\psi} \rangle} \leq \|W\| \langle \hat{\psi}, \hat{\psi} \rangle = O(\delta^2) \|W\|.$$

Similarly, we have

$$|\langle \hat{\psi}, Wv \rangle| \leq \sqrt{\langle \hat{\psi}, \hat{\psi} \rangle} \sqrt{\langle Wv, Wv \rangle} \leq O(\delta) \|W\| \sqrt{\langle v, v \rangle} = O(\delta) \|W\|.$$

From (13), $\|W\| = 2\sigma + o(1) = O(1)$. Therefore, going back to (24) we have obtained

$$\langle v, (W - \tilde{A})v \rangle = O(\delta) + O(n^{-1/2}).$$

□

We end this section with the proof of Corollary 1.3.

Proof of Corollary 1.3. By Theorem 1.2, almost surely, and for any compact subset C of \mathbb{R} containing strictly $[0, 2\sigma]$, the law $\tilde{\mu}_{\sqrt{n}K}$ is supported in C for large enough n . On the other hand, since $\mu_{\sqrt{n}K} = (1 - n^{-1})\tilde{\mu}_{\sqrt{n}K} + n^{-1}\delta_{\sqrt{n}}$, we get from Theorem 1.1 that almost surely, $\tilde{\mu}_{\sqrt{n}K}$ tends weakly to $\mathcal{W}_{2\sigma}$ as $n \rightarrow \infty$. Now, for sequences of probability measures supported in a common compact set, by the Weierstrass theorem, weak convergence is equivalent to Wasserstein convergence W_p for every $p \geq 1$. Consequently, almost surely,

$$\lim_{n \rightarrow \infty} W_p(\tilde{\mu}_{\sqrt{n}K}, \mathcal{W}_{2\sigma}) = 0. \quad (25)$$

for every $p \geq 1$. It remains to study $W_p(\mu_{\sqrt{n}K}, \mathcal{W}_{2\sigma})$. Recall that if ν_1 and ν_2 are two probability measures on \mathbb{R} with cumulative distribution functions F_{ν_1} and F_{ν_2} with respective generalized inverses $F_{\nu_1}^{-1}$ and $F_{\nu_2}^{-1}$, then, for every real $p \geq 1$, we have, according to e.g. [49, Remark 2.19 (ii)],

$$W_p(\nu_1, \nu_2)^p = \int_0^1 |F_{\nu_1}^{-1}(t) - F_{\nu_2}^{-1}(t)|^p dt. \quad (26)$$

Let us take $\nu_1 = \mu_{\sqrt{n}K} = (1 - n^{-1})\tilde{\mu}_{\sqrt{n}K} + n^{-1}\delta_{\sqrt{n}}$ and $\nu_2 = \mathcal{W}_{2\sigma}$. Theorem 1.2 gives $\lambda_2(\sqrt{n}K) < \infty$ a.s. Also, a.s., for large enough n , and for every $t \in (0, 1)$,

$$F_{\nu_1}^{-1}(t) = F_{\mu_{\sqrt{n}K}}^{-1}(t) = \sqrt{n} \mathbb{1}_{[1-n^{-1}, 1)}(t) + F_{\tilde{\mu}_{\sqrt{n}K}}^{-1}(t + n^{-1}) \mathbb{1}_{(0, 1-n^{-1})}(t).$$

The desired result follows then by plugging this identity in (26) and by using (25). □

3 Proofs for the chain graph model

Here we prove theorem 1.6 and 1.10 and corollaries 1.7 and 1.11.

Proof of Theorem 1.6. Since μ_K is supported in the compact set $[-1, +1]$ which does not depend on n , the Weierstrass theorem implies that the weak convergence of μ_K as $n \rightarrow \infty$ is equivalent to the convergence of all moments, and is also equivalent to the convergence in Wasserstein distance W_p for every $p \geq 1$. Thus, by a Prohorov tightness argument, it suffices to show that a.s. for any $\ell \geq 0$, the ℓ^{th} moment of μ_K converges to $\mathbb{E}[r_\ell^{\mathbf{P}}(0)]$ as $n \rightarrow \infty$. The sequence $(\mathbb{E}[r_\ell^{\mathbf{P}}(0)])_{\ell \geq 0}$ will be then necessarily the sequence of moments of a probability measure μ on $[-1, +1]$ which is the unique adherence value of μ_K as $n \rightarrow \infty$.

For any $\ell \geq 0$ and $i \geq 1$ let $r_\ell^{\mathbf{P},n}(i)$ be the probability of return to i after ℓ steps for the random walk on $\{1, \dots, n\}$ with kernel K (this probability is a random variable since K is random). Clearly, $r_\ell^{\mathbf{P},n}(i) = r_\ell^{\mathbf{P}}(i)$ whenever $1 + \ell < i < n - \ell$. Therefore, for every fixed ℓ , the ergodic theorem implies that almost surely,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n r_\ell^{\mathbf{P},n}(i) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n r_\ell^{\mathbf{P}}(i) = \mathbb{E}[r_\ell^{\mathbf{P}}(0)].$$

This ends the proof. \square

Proof of Corollary 1.7. The desired convergence follows immediately from Theorem 1.6 with $\mathbf{p}(i) = (1 - V_i, 0, V_i)$ for every $i \geq 1$. The expression of the moments of μ is immediate. Let us explain now how the result can be deduced also from the trace-moment method. Let us take $a_1 = \dots = a_n$ in (9). The Markov matrix K defined by (9) is not symmetric in general, but we have an analog of Lemma 2.1. Namely, the spectrum of K is identical to the spectrum of the symmetric tridiagonal matrix S defined for any $1 \leq i, j \leq n$ by

$$S_{i,j} = \sqrt{\frac{\rho_i}{\rho_j}} K_{i,j} = \sqrt{b_i c_{i+1}} \delta_{j,i+1} + \sqrt{b_{i-1} c_i} \delta_{j,i-1}$$

where $\delta_{k,l} = \mathbb{1}_{\{k=l\}}$ is the Kronecker symbol. In other words

$$S = \begin{pmatrix} 0 & \sqrt{b_1 c_2} & & & \\ \sqrt{b_1 c_2} & 0 & \sqrt{b_2 c_3} & & \\ & \sqrt{b_2 c_3} & 0 & \sqrt{b_3 c_4} & \\ & & \ddots & \ddots & \\ & & & \sqrt{b_{n-2} c_{n-1}} & 0 & \sqrt{b_{n-1} c_n} \\ & & & & \sqrt{b_{n-1} c_n} & 0 \end{pmatrix}. \quad (27)$$

Moreover, the eigenspaces dimensions are also identical. The random reversible Markov kernel K defined by (10) is spectrally equivalent to the random symmetric tridiagonal matrix obtained from (27) by taking $b_i = 1 - c_i = V_i$ for every $2 \leq i \leq n - 1$ and $b_1 = c_n = 1$. The entries of this matrix are dependent. Nevertheless, the desired result follows from the trace-moments method and appears as a variant of the proof used by Popescu for models of random symmetric tridiagonal matrices with independent entries [38, Theorem 2.8]. We leave the details to the reader. \square

Remark 3.1 (Computation of the moments of μ for Beta environments). As noticed in Remark 1.8, the limiting spectral distribution μ is the arc-sine law when $\mathcal{L} = \delta_{1/2}$. Assume now that \mathcal{L} is uniform on $[0, 1]$. Then for every integers $m \geq 0$ and $n \geq 0$,

$$\mathbb{E}(V^m(1-V)^n) = \int_0^1 u^m(1-u)^n du = \text{Beta}(n+1, m+1) = \frac{\Gamma(n+1)\Gamma(m+1)}{\Gamma(n+m+2)}$$

which gives

$$\mathbb{E}(V^m(1-V)^n) = \frac{n!m!}{(n+m+1)!} = \frac{1}{(n+m+1)\binom{n+m}{m}}.$$

The law of $\binom{n+m}{m}V^m(1-V)^n$ is the law of the probability of having m success in $n+m$ tosses of a coin with a probability of success p uniformly distributed in $[0, 1]$. One may think about the trivial case (Uniform-Binomial) of the Dirichlet-Multinomial Bose-Einstein distribution. Similar formulas may be obtained when \mathcal{L} is an Euler Beta law $\text{Beta}(\alpha, \beta)$.

Proof of Theorem 1.10. Proof of the first statement. It is enough to show that for every $0 < a < 1$, there exists an integer k_a such that for all $k \geq k_a$,

$$\int_{-1}^{+1} x^{2k} \mu(dx) \geq a^{2k}. \quad (28)$$

By assumption, there exists $C > 0$ and $0 < t_0 < 1/2$ such that for all $0 < t < t_0$,

$$\mathbb{P}(V \in [1/2 - t, 1/2 + t]) \geq Ct$$

where V is random variable of law \mathcal{L} . In particular, for all $0 < t < t_0$,

$$\mathbb{E} \left[V^{N_\gamma(i)} (1-V)^{N_\gamma(i-1)} \right] \geq Ct \left(\frac{1}{2} - t \right)^{N_\gamma(i) + N_\gamma(i-1)},$$

and, if

$$\|\gamma\|_\infty = \max\{i \geq 0 : \max(N_\gamma(i), N_\gamma(-i)) \geq 1\}$$

then

$$\begin{aligned} \int_{-1}^{+1} x^{2k} \mu(dx) &\geq \sum_{\gamma \in D_k} \prod_{i \in \mathbb{Z}} Ct \left(\frac{1}{2} - t \right)^{N_\gamma(i) + N_\gamma(i-1)} \\ &\geq \sum_{\gamma \in D_k} (Ct)^{2\|\gamma\|_\infty} \left(\frac{1}{2} - t \right)^{\sum_i N_\gamma(i) + N_\gamma(i-1)} \\ &\geq \left(\frac{1}{2} - t \right)^{2k} \sum_{\gamma \in D_k} (Ct)^{2\|\gamma\|_\infty} \\ &\geq \left(\frac{1}{2} - t \right)^{2k} |D_{k,\alpha}| (Ct)^{2k^\alpha}, \end{aligned}$$

where $D_{k,\alpha} = \{\gamma \in D_k : \|\gamma\|_\infty \leq k^\alpha\}$. Now, from the Brownian Bridge version of Donsker's Theorem, for all $\alpha > 1/2$,

$$\lim_{k \rightarrow \infty} \frac{|D_{k,\alpha}|}{|D_k|} = 1.$$

Since $|D_k| = \text{Card}(D_k) = \binom{2k}{k}$, Stirling's formula gives $|D_k| \sim 4^k(\pi k)^{-1/2}$, and thus

$$\int_{-1}^{+1} x^{2k} \mu(dx) \geq (\pi k)^{-1/2} (1-2t)^{2k} (Ct)^{2k\alpha} (1+o(1)).$$

We then deduce the desired result (28) by taking t small enough such that $1-2t > a$ and $1/2 < \alpha < 1$. This achieves the proof of the first statement.

Proof of the second statement. One can observe that if $\mathcal{L} = \delta_p$ for some $p \in (0, 1)$ with $p \neq 1/2$, an explicit computation of the spectrum will provide the desired result, in accordance with Remark 1.8. For the general case, we get from [36], for any $2 \leq k \leq n-1$,

$$1 - \lambda_2(K) \geq \frac{1}{4 \max(B_k^+, B_k^-)}$$

where

$$B_k^+ = \max_{i > k} \left[\left(\sum_{j=k+1}^i \frac{1}{\rho_j(1-V_j)} \right) \sum_{j \geq i} \rho_j \right] \quad \text{and} \quad B_k^- = \max_{i < k} \left[\left(\sum_{j=i}^{k-1} \frac{1}{\rho_j V_j} \right) \sum_{j \leq i} \rho_j \right]$$

with the convention $V_1 = 1 - V_n = 1$. Here we have fixed the value of n and ρ is any invariant (reversible) measure for K . It is convenient to take $\rho_1 = 1$ and for every $2 \leq i \leq n$

$$\rho_i = \frac{V_2 \cdots V_{i-1}}{(1-V_2) \cdots (1-V_i)}.$$

By symmetry, it suffices to consider the case where \mathcal{L} is supported in $[0, t]$ with $0 < t < 1/2$. Let us take $k = 2$. In this case, $B_2^- = 1$, and the desired result will follow if we show that B_2^+ is bounded above by a constant independent of n . To this end, we remark first that for any $\ell > j$ we have $\rho_\ell = \rho_j \prod_{m=j}^{\ell-1} (V_m/(1-V_{m+1}))$. Therefore, setting $e^{-\gamma} = t/(1-t) < 1$, we have $\rho_\ell \leq \rho_j e^{-\gamma(\ell-j)}$. It follows that, for any $k < i$,

$$\begin{aligned} \sum_{j=k+1}^i \sum_{\ell \geq i} \frac{\rho_\ell}{\rho_j(1-V_j)} &\leq \frac{1}{1-t} \sum_{j=k+1}^i e^{-\gamma(i-j)} \sum_{\ell \geq i} e^{-\gamma(\ell-i)} \\ &\leq \frac{(1-e^{-\gamma})^{-2}}{1-t} = \frac{1-t}{(1-2t)^2}. \end{aligned}$$

In particular, $B_2^+ \leq (1-t)/(1-2t)^2$, which concludes the proof. \square

Relation to random orthogonal polynomials

There are deep links between orthogonal polynomials, symmetric tridiagonal matrices, and the Hamburger moments problem, see e.g. [48, 44, 21, 22]. See also [33, 32, 48, 22] for the relationship with birth-and-death processes on \mathbb{Z}_+ . More precisely, fix a prescribed bounded sequence $(\gamma_n)_{n \geq 0}$ of positive reals and define the symmetric tridiagonal matrix

$$\Gamma = \begin{pmatrix} 0 & \gamma_1 & & & \\ \gamma_1 & 0 & \gamma_2 & & \\ & \gamma_2 & 0 & \gamma_3 & \\ & & \ddots & \ddots & \ddots \\ & & & \gamma_{n-2} & 0 & \gamma_{n-1} \\ & & & & \gamma_{n-1} & 0 \end{pmatrix} \quad (29)$$

with eigenvalues $(\lambda_{n,k})_{1 \leq k \leq n}$ and ESD

$$\mu_\Gamma = \frac{1}{n} \sum_{k=1}^n \delta_{\lambda_{n,k}}.$$

The characteristic polynomial P_n of Γ is the last term of the second order recursion

$$P_{k+1}(x) = xP_k(x) - \gamma_k^2 P_{k-1}(x)$$

for every $0 \leq k \leq n-1$, with $P_{-1} \equiv 0$ and $P_0 \equiv 1$. From the classical theory [21, 44, 48], $(P_n)_{n \geq 0}$ is the sequence of orthogonal polynomials for a compactly supported probability measure ν on \mathbb{R} . Namely, for every $i, j \geq 0$,

$$\int_{\mathbb{R}} P_i(x) P_j(x) \nu(dx) = \delta_{i,j}.$$

The law ν is unique and solves a Hamburger moments problem. The matrix Γ and the law ν are known respectively as the *Jacobi matrix* and the *spectral measure* of $(P_n)_{n \geq 0}$. The roots $(\lambda_{n,k})_{1 \leq k \leq n}$ of P_n are real, distinct, belong to the support of ν , and there exists one and only one root of P_{n+1} between two roots of P_n (interlacement). The Stieltjes transform of ν is given by the continuous fraction (for every $z \in \mathbb{C}$ with $\Im(z) > 0$)

$$\int_{\mathbb{R}} \frac{1}{z-x} \nu(dx) = \frac{1}{z - \frac{\gamma_1^2}{z - \frac{\gamma_2^2}{z - \ddots}}}.$$

The n^{th} approximant of this continuous fraction is the Stieltjes transform of the Gauss-Christoffel quadrature ν_n of ν . In other words, ν_n is the discrete law on \mathbb{R} with atoms $(\lambda_{n,k})_{1 \leq k \leq n}$ which is equal to ν as a linear form on all polynomials of degree $\leq n$. If $(v_{n,k})_{1 \leq k \leq n}$ are the orthogonal eigenvectors of the symmetric matrix Γ associated to the eigenvalues $(\lambda_{n,k})_{1 \leq k \leq n}$, then from [29, Proposition 1.1.9] we get for every real $p \geq 1$

$$\lim_{n \rightarrow \infty} W_p(\nu_n, \nu) = 0 \quad \text{where} \quad \nu_n = \frac{1}{n} \sum_{k=1}^n \langle v_{n,k}, e_1 \rangle^2 \delta_{\lambda_{n,k}}.$$

It turns out that ν_n is the so called *distribution measure at vector e_1* of the self-adjoint operator Γ , while ν is the *distribution measure at vector e_1* of the self-adjoint operator $\bar{\Gamma} : \ell^2(\mathbb{N}) \rightarrow \ell^2(\mathbb{N})$ where $\mathbb{N} = \{1, 2, \dots\}$, see e.g. [29, Example 1.1.1 and Proposition 1.1.9]. Here $\bar{\Gamma}$ is the operator which coincides with Γ on $\ell^2(\{1, \dots, n\})$ for every $n \geq 1$. In other words, for every $k \geq 0$,

$$\langle \bar{\Gamma}^k e_1, e_1 \rangle_{\ell^2(\mathbb{N})} = \lim_{n \rightarrow \infty} \langle \Gamma^k e_1, e_1 \rangle_{\ell^2(\{1, \dots, n\})} = \lim_{n \rightarrow \infty} \int_{\mathbb{R}} x^k \nu_n(dx) = \int_{\mathbb{R}} x^k \nu(dx).$$

One should not confuse ν_n with μ_Γ . In general, μ_Γ does not converge weakly to ν . Actually, a famous theorem by Erdős & Turán [25] states that if ν is supported on $[-1, +1]$ with almost-everywhere positive density, then μ_Γ tends weakly as $n \rightarrow \infty$ to the arc-sine

distribution on $[-1, +1]$, see e.g. [48, p. 24–25]. In fact the convergence holds for Wasserstein distance W_p for every $p \geq 1$. In the special case where $\gamma_n = 1$ for every $n \geq 1$, then ν turns out to be a Wigner semi-circle law on $[-1, +1]$ and $(P_n)_{n \geq 0}$ are the Jacobi polynomials with parameters $\alpha = \beta = 1/2$, also known as the Chebyshev polynomials of the second kind. From the Erdős-Turán theorem, we get then a simple example for which ν (Wigner semi-circle law on $[-1, +1]$) is not the weak limit μ of μ_Γ (arc-sine law on $[-1, +1]$). On the other hand, it is well known that the arc-sine law on $[-1, +1]$ is a fixed point of the map $\nu \mapsto \mu$ and $(P_n)_{n \geq 0}$ are in this case the Chebyshev polynomials of the first kind.

Recall that the recursive relations of the Chebyshev polynomials of the first and second kind differ only on the first coefficients. We are now ready for the proof of Corollary 1.11.

Proof of Corollary 1.11. Observe first that $\mu_{Q_n} = \mu_\Gamma$ where Γ is the random symmetric tridiagonal matrix defined from (29) with $\gamma_n = \sqrt{V_n(1 - V_{n+1})}$ for every $n \geq 1$. On the other hand, let S be the symmetric tridiagonal matrix (27) considered in the proof of Corollary 1.7. It coincides with the matrix (29) with this time $\gamma_k = \sqrt{b_k c_{k+1}}$ for every $1 \leq k \leq n-1$. Due to the reflection boundary conditions $b_1 = c_n = 1$, the finite sequence $\gamma_1, \dots, \gamma_{n-1}$ is not the first $n-1$ terms of a fixed infinite sequence. Also, the roots of Q_n are not the eigenvalues of the random reversible kernel K of Corollary 1.7. However, the empirical distribution μ_{Q_n} of the roots of Q_n and the ESD of K share the same weak adherence values since the roots of Q_{n-1} are interlaced with the roots of the characteristic polynomial of K . The desired result follows then from Corollary 1.7. \square

Remark 1.8 and the numerical experiments of figures 2-3-4 shows that μ is not the arc-sine law on $[-1, +1]$ for many choices of \mathcal{L} . In such cases, the hypothesis on ν of the Erdős-Turán theorem are necessarily almost-surely violated. Note that the arc-sine law appears also in the literature as the limit of the empirical distribution of the roots of orthogonal polynomials associated to random measures defined with uniform random canonical moments, see e.g. [7] and references therein.

Remark 3.2 (Yet another sequence of random orthogonal polynomials). *It is quite natural to consider the empirical measure of the roots μ_{Q_n} associated to the sequence of random orthogonal polynomials $(Q_n)_{n \geq 0}$ defined by the recursive relation*

$$Q_{n+1}(x) = xQ_n(x) - W_n Q_{n-1}(x)$$

for every $n \geq 1$ with initial conditions $Q_0 \equiv 1$ and $Q_1(x) = x$, where this time $(W_n)_{n \geq 1}$ are i.i.d. positive random variables. For every $n \geq 1$, the random polynomial Q_n is the characteristic polynomial of the random matrix

$$\begin{pmatrix} 0 & \sqrt{W_1} & & & & \\ \sqrt{W_1} & 0 & \sqrt{W_2} & & & \\ & \sqrt{W_2} & 0 & \sqrt{W_3} & & \\ & & \ddots & \ddots & \ddots & \\ & & & \sqrt{W_{n-2}} & 0 & \sqrt{W_{n-1}} \\ & & & & \sqrt{W_{n-1}} & 0 \end{pmatrix}.$$

This matrix is the so called Jacobi matrix of the sequence $(Q_n)_{n \geq 0}$, and its ESD is μ_{Q_n} . Since this matrix is symmetric tridiagonal with i.i.d. entries, the result of Popescu in [38]

implies that a.s. μ_{Q_n} tends as $n \rightarrow \infty$ to a non-random probability measure which depends on the common law of the W_i 's via its moments. Here again, the limit of μ_{Q_n} as $n \rightarrow \infty$ does not coincide in general with the (random) law for which $(Q_n)_{n \geq 0}$ are orthogonal.

Acknowledgements. The second author would like to thank the *Équipe de Probabilités et Statistique de l'Institut de Mathématiques de Toulouse* for kind hospitality. The last author would like to thank Delphine FÉRAL and Sandrine PÉCHÉ for interesting discussions on the extremal eigenvalues of symmetric non-central random matrices with i.i.d entries.

References

- [1] A. Auffinger, G. Ben Arous, and S. Péché, *Poisson convergence for the largest eigenvalues of Heavy Tailed Random Matrices*, preprint, [arXiv:0710.3132 \[math.PR\]](#), 2007.
- [2] Z. D. Bai, *Methodologies in spectral analysis of large-dimensional random matrices, a review*, Statist. Sinica **9** (1999), no. 3, 611–677, With comments by G. J. Rodgers and J. W. Silverstein; and a rejoinder by the author.
- [3] Z. D. Bai and Y. Q. Yin, *Necessary and sufficient conditions for almost sure convergence of the largest eigenvalue of a Wigner matrix*, Ann. Probab. **16** (1988), no. 4, 1729–1741.
- [4] ———, *Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix*, Ann. Probab. **21** (1993), no. 3, 1275–1294.
- [5] G. Ben Arous and A. Guionnet, *The spectrum of heavy tailed random matrices*, Comm. Math. Phys. **278** (2008), no. 3, 715–751.
- [6] P. Billingsley, *Convergence of probability measures*, second ed., Wiley Series in Probability and Statistics: Probability and Statistics, John Wiley & Sons Inc., New York, 1999, A Wiley-Interscience Publication.
- [7] M. Birke and H. Dette, *A note on random orthogonal polynomials on a compact interval*, preprint, [arXiv.math:0809.4936](#), 2008.
- [8] G. Biroli, J.-P. Bouchaud, and M. Potters, *On the top eigenvalue of heavy-tailed random matrices*, Europhys. Lett. EPL **78** (2007), no. 1, Art. 10001, 5.
- [9] D. Boivin and J. Depauw, *Spectral homogenization of reversible random walks on \mathbb{Z}^d in a random environment*, Stochastic Process. Appl. **104** (2003), no. 1, 29–56.
- [10] B. Bollobás, *Random graphs*, second ed., Cambridge Studies in Advanced Mathematics, vol. 73, Cambridge University Press, Cambridge, 2001.
- [11] E. Bolthausen and A.-S. Sznitman, *Ten lectures on random media*, DMV Seminar, vol. 32, Birkhäuser Verlag, Basel, 2002. MR MR1890289 (2003f:60183)
- [12] Ch. Bordenave, P. Caputo, and D. Chafaï, *Spectrum of large random reversible Markov chains – Heavy-tailed weights on the complete graph*, in preparation, 2008.
- [13] A. Bovier and A. Faggionato, *Spectral characterization of aging: the REM-like trap model*, Ann. Appl. Probab. **15** (2005), no. 3, 1997–2037.
- [14] ———, *Spectral analysis of Sinai's walk for small eigenvalues*, Ann. Probab. **36** (2008), no. 1, 198–254.
- [15] S. Boyd, P. Diaconis, P. Parrilo, and L. Xiao, *Symmetry analysis of reversible Markov chains*, Internet Math. **2** (2005), no. 1, 31–71.
- [16] W. Bryc, A. Dembo, and T. Jiang, *Spectral measure of large random Hankel, Markov and Toeplitz matrices*, Ann. Probab. **34** (2006), no. 1, 1–38.
- [17] D. Chafaï, *Circular law for non-central random matrices*, preprint [arXiv:0709.0036 \[math.PR\]](#), 2007.
- [18] ———, *The Dirichlet Markov Ensemble*, preprint [arXiv:0709.4678 \[math.PR\]](#), 2007.
- [19] ———, *Circular Law Theorem for Random Markov Matrices*, partial results, [arXiv:0808.1502 \[math.PR\]](#), 2008.

- [20] D. Cheliotis and B. Virag, *The spectrum of the random environment and localization of noise*, preprint, [arXiv:math:0804.4814](#), 2008.
- [21] T. S. Chihara, *An introduction to orthogonal polynomials*, Gordon and Breach Science Publishers, New York, 1978, Mathematics and its Applications, Vol. 13. MR MR0481884 (58 #1979)
- [22] H. Dette and W. J. Studden, *The theory of canonical moments with applications in statistics, probability, and analysis*, Wiley Series in Probability and Statistics: Applied Probability and Statistics, John Wiley & Sons Inc., New York, 1997, A Wiley-Interscience Publication. MR MR1468473 (98k:60020)
- [23] P. G. Doyle and J. L. Snell, *Random walks and electric networks*, Carus Mathematical Monographs, vol. 22, Mathematical Association of America, Washington, DC, 1984.
- [24] F. J. Dyson, *Statistical theory of the energy levels of complex systems. Parts I, II, III*, J. Mathematical Phys. **3** (1962), 140–175.
- [25] P. Erdős and P. Turán, *On interpolation. III. Interpolatory theory of polynomials*, Ann. of Math. (2) **41** (1940), 510–553. MR MR0001999 (1,333e)
- [26] D. Féral and S. Péché, *The largest eigenvalue of rank one deformation of large Wigner matrices*, Comm. Math. Phys. **272** (2007), no. 1, 185–228.
- [27] Z. Füredi and J. Komlós, *The eigenvalues of random symmetric matrices*, Combinatorica **1** (1981), no. 3, 233–241.
- [28] A. Guionnet, *Lectures on Random Matrices: Macroscopic Asymptotics*, Notes de cours de l'École d'été de Saint-Flour 2006, to appear in Lecture Notes in Mathematics, Springer-Verlag, 2008.
- [29] F. Hiai and D. Petz, *The semicircle law, free random variables and entropy*, Mathematical Surveys and Monographs, vol. 77, American Mathematical Society, Providence, RI, 2000. MR MR1746976 (2001j:46099)
- [30] A. J. Hoffman and H. W. Wielandt, *The variation of the spectrum of a normal matrix*, Duke Math. J. **20** (1953), 37–39.
- [31] R. Horn and C. Johnson, *Topics in matrix analysis*, Cambridge University Press, Cambridge, 1991.
- [32] S. Karlin and J. McGregor, *The classification of birth and death processes*, Trans. Amer. Math. Soc. **86** (1957), 366–400. MR MR0094854 (20 #1363)
- [33] ———, *Random walks*, Illinois J. Math. **3** (1959), 66–81. MR MR0100927 (20 #7352)
- [34] F. P. Kelly, *Reversibility and stochastic networks*, John Wiley & Sons Ltd., Chichester, 1979, Wiley Series in Probability and Mathematical Statistics.
- [35] M. L. Mehta, *Random matrices*, third ed., Pure and Applied Mathematics (Amsterdam), vol. 142, Elsevier/Academic Press, Amsterdam, 2004. MR MR2129906 (2006b:82001)
- [36] L. Miclo, *An example of application of discrete Hardy's inequalities*, Markov Process. Related Fields **5** (1999), no. 3, 319–330. MR MR1710983 (2000h:60081)
- [37] R. Montenegro and P. Tetali, *Mathematical Aspects of Mixing Times in Markov Chains*, Foundations and Trends in Theoretical Computer Science, vol. 1:3, Now Publishers, 2006.
- [38] I. Popescu, *General tridiagonal random matrix models, limiting distributions and fluctuations*, [arXiv:math/0610827](#) [math.PR] to appear in Probab. Theory Related Fields, 2008.
- [39] S. T. Rachev, *Probability metrics and the stability of stochastic models*, Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics, John Wiley & Sons Ltd., Chichester, 1991. MR MR1105086 (93b:60012)
- [40] L. Saloff-Coste, *Lectures on finite Markov chains*, Lectures on probability theory and statistics (Saint-Flour, 1996), Lecture Notes in Math., vol. 1665, Springer, Berlin, 1997, pp. 301–413.
- [41] E. Seneta, *Non-negative matrices and Markov chains*, Springer Series in Statistics, Springer, New York, 2006, Revised reprint of the second (1981) edition [Springer-Verlag, New York; MR0719544]. MR MR2209438
- [42] A. Soshnikov, *Universality at the edge of the spectrum in Wigner random matrices*, Comm. Math. Phys. **207** (1999), no. 3, 697–733.
- [43] ———, *Poisson statistics for the largest eigenvalues of Wigner random matrices with heavy tails*, Electron. Comm. Probab. **9** (2004), 82–91 (electronic).

- [44] G. Szegő, *Orthogonal polynomials*, fourth ed., American Mathematical Society, Providence, R.I., 1975, American Mathematical Society, Colloquium Publications, Vol. XXIII. MR MR0372517 (51 #8724)
- [45] A.-S. Sznitman, *Topics in random walks in random environment*, School and Conference on Probability Theory, ICTP Lect. Notes, XVII, Abdus Salam Int. Cent. Theoret. Phys., Trieste, 2004, pp. 203–266 (electronic). MR MR2198849 (2007b:60247)
- [46] T. Tao and V. Vu, *Random matrices: Universality of ESDs and the circular law*, preprint [arXiv:0807.4898](https://arxiv.org/abs/0807.4898) [math.PR], 2008.
- [47] C. A. Tracy and H. Widom, *The distribution of the largest eigenvalue in the Gaussian ensembles: $\beta = 1, 2, 4$* , Calogero-Moser-Sutherland models (Montréal, QC, 1997), CRM Ser. Math. Phys., Springer, New York, 2000, pp. 461–472.
- [48] W. Van Assche, *Asymptotics for orthogonal polynomials*, Lecture Notes in Mathematics, vol. 1265, Springer-Verlag, Berlin, 1987. MR MR903848 (88i:42035)
- [49] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003. MR MR1964483 (2004e:90003)
- [50] J. H. M. Wedderburn, *Lectures on matrices*, Dover Publications Inc., New York, 1964.
- [51] E. P. Wigner, *The collected works of Eugene Paul Wigner. Part A. The scientific papers. Vol. II. Nuclear physics*, Springer-Verlag, Berlin, 1996, Annotated by Herman Feshbach, Edited and with a preface by Arthur S. Wightman and Jagdish Mehra.
- [52] I. Zakharevich, *A generalization of Wigner’s law*, Comm. Math. Phys. **268** (2006), no. 2, 403–414.
- [53] O. Zeitouni, *Random walks in random environment*, Lectures on probability theory and statistics, Lecture Notes in Math., vol. 1837, Springer, Berlin, 2004, pp. 189–312. MR MR2071631 (2006a:60201)

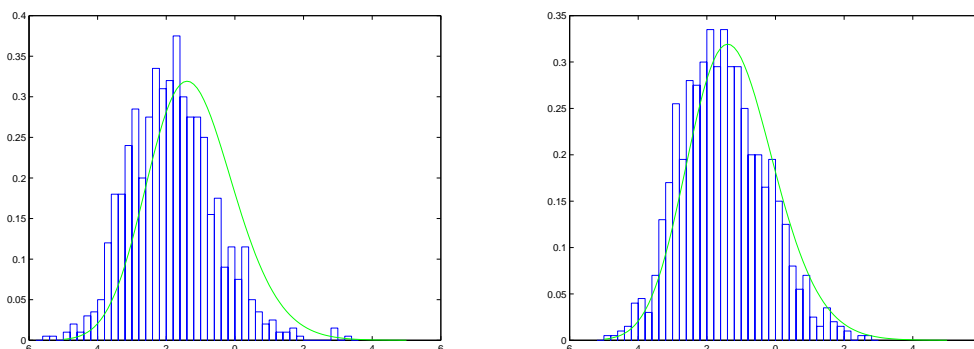


Figure 1: These plots correspond to the fluctuation of the sub-dominant eigenvalue of K in Theorem 1.2 (complete graph) with \mathcal{L} uniform on $[0, 1]$. In both plots, the green solid bell shaped curve is the density of the Tracy-Widom distribution, for the λ_{\max} of the so called Beta Ensemble with $\beta = 1$, which is the Gaussian Orthogonal Ensemble GOE, see e.g. [47], plotted by using the Matlab package RMLab version 0.02 by Momar Dieng. Both histograms were obtained with a sample of size 1000 of $n^{2/3}(\frac{m}{\sigma}\lambda_2(\sqrt{n}K) - 2)$, where m is the first moment of \mathcal{L} . The left hand side histogram corresponds to $n = 50$, whereas the right hand side histogram corresponds to $n = 500$. One can notice in both graphics a negative bias, which is smaller for $n = 500$ than for $n = 50$. Is it asymptotically zero?

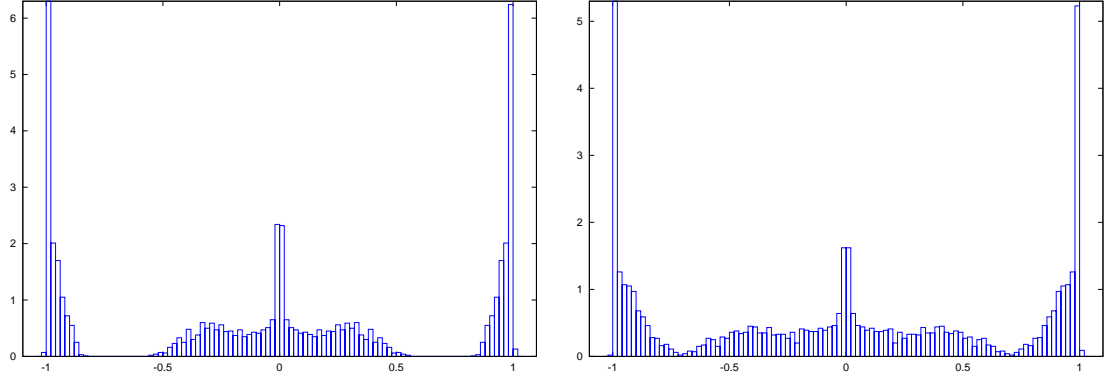


Figure 2: These plots illustrate Corollary 1.7. Each histogram corresponds to the spectrum of a single realization of K with $n = 5000$, for various choices of \mathcal{L} . From left to right \mathcal{L} is the uniform law on $[0, t] \cup [1 - t, 1]$ for $t = 1/8$, $t = 1/4$.

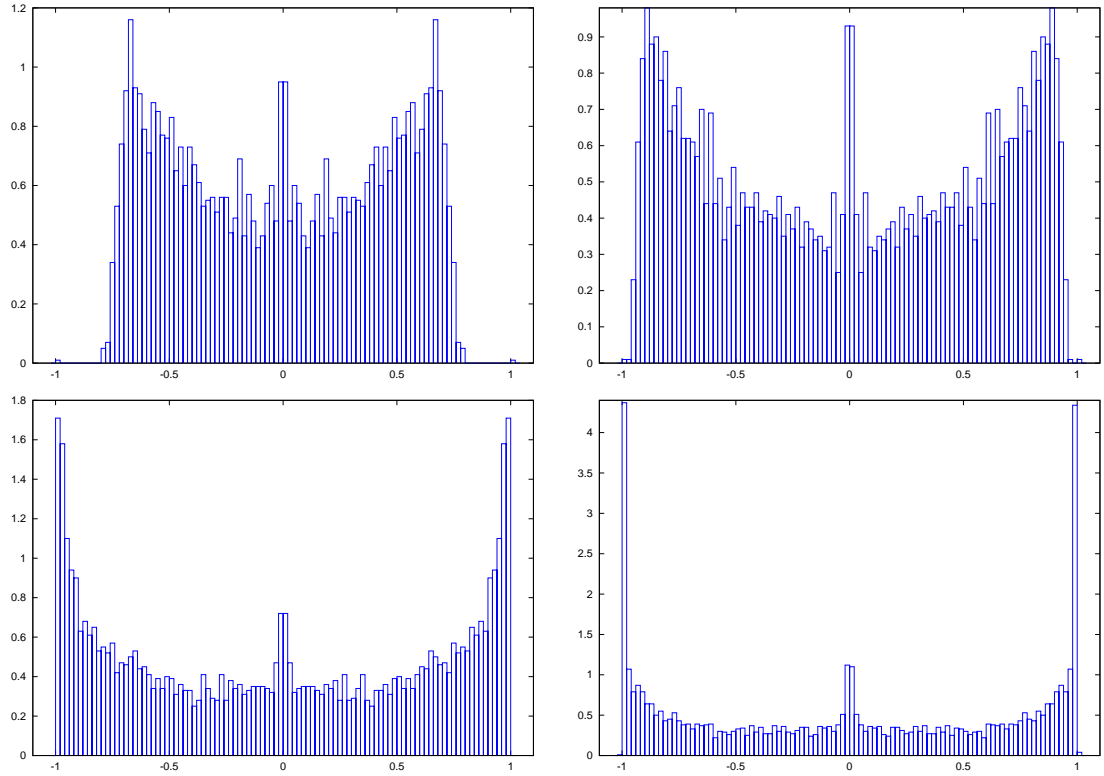


Figure 3: These plots illustrate Corollary 1.7 and the second statement of Theorem 1.10. Each histogram corresponds to the spectrum of a single realization of K with $n = 5000$, for various choices of \mathcal{L} . From left to right and top to bottom, \mathcal{L} is uniform on $[0, t]$ with $t = 1/8$, $t = 1/4$, $t = 1/2$, and $t = 1$.

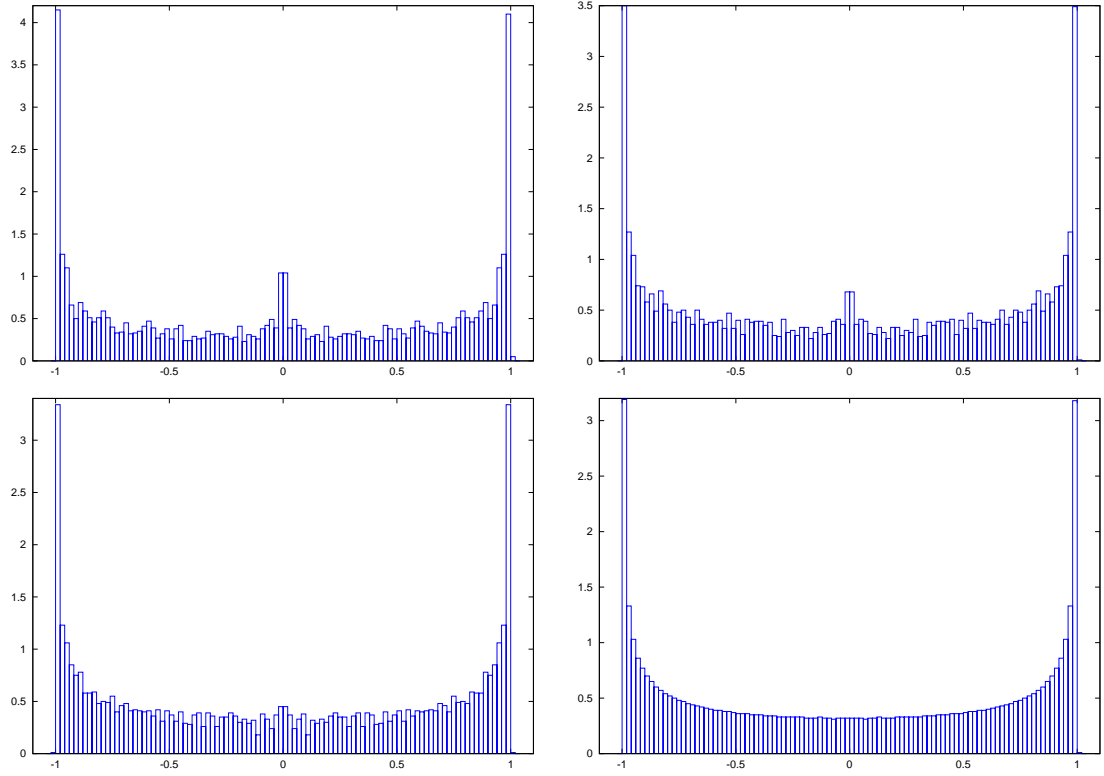


Figure 4: These plots illustrate Corollary 1.7. Each histogram corresponds to the spectrum of a single realization of K with $n = 5000$, for various choices of \mathcal{L} . From left to right and top to bottom, \mathcal{L} is uniform on $[t, 1 - t]$ with $t = 0$, $t = 1/8$, $t = 1/4$, $t = 1/2$. The last case corresponds to the arc-sine limiting spectral distribution mentioned in Remark 1.8.

Charles BORDENAVE e-mail: `charles.bordenave(AT)math.univ-toulouse.fr`
UMR5219 CNRS & INSTITUT DE MATHÉMATIQUES DE TOULOUSE
UNIVERSITÉ TOULOUSE III, FRANCE.

Pietro CAPUTO e-mail: `caputo(AT)mat.uniroma3.it`
DIPARTIMENTO DI MATEMATICA
UNIVERSITÀ ROMA TRE, ITALY.

Djalil CHAFAÏ – corresponding author e-mail: `chafai(AT)math.univ-toulouse.fr`
UMR181 INRA ENVT & INSTITUT DE MATHÉMATIQUES DE TOULOUSE
UNIVERSITÉ DE TOULOUSE, FRANCE.

Compiled November 7, 2008