The Dirichlet Markov Ensemble

Djalil Chafaï

Preprint – September 2007. Revised – August 2009, October 2009.

Abstract

We equip the polytope of $n \times n$ Markov matrices with the normalized trace of the Lebesgue measure of $\mathbb{R}^{n^2}$. This probability space provides random Markov matrices, with i.i.d. rows following the Dirichlet distribution of mean $(1/n, \ldots, 1/n)$. We show that if $\mathbf{M}$ is such a random matrix, then the empirical distribution built from the singular values of $\sqrt{n}\mathbf{M}$ tends as $n \to \infty$ to a Wigner quarter–circle distribution. Some computer simulations reveal striking asymptotic spectral properties of such random matrices, still waiting for a rigorous mathematical analysis. In particular, we believe that with probability one, the empirical distribution of the complex spectrum of $\sqrt{n}\mathbf{M}$ tends as $n \to \infty$ to the uniform distribution on the unit disc of the complex plane, and that moreover, the spectral gap of $\mathbf{M}$ is of order $1 - 1/\sqrt{n}$ when n is large.

AMS 2000 Mathematical Subject Classification: 15A52; 15A51; 15A42; 60F15; 62H99.
Keywords: Random matrices; Markov matrices; Dirichlet laws; Spectral gap.

1 Introduction

Markov chains constitute an essential tool for the modelling of stochastic phenomena in Biology, Computer Science, Engineering, and Physics. It is nowadays well known that the trend to the equilibrium of ergodic Markov chains is related to the spectral decomposition of their Markov transition matrix, see for instance [Sen06, SC97, CSC08]. The corresponding literature is very rich, and many statisticians including for instance the famous Persi Diaconis contributed to this subject, by providing quantitative bounds for various concrete specific Markov chains. But how a Markov chain behaves when its Markov transition matrix is taken arbitrarily in the set of Markov matrices? The present paper aims to provide some partial answers to this natural concrete question. From the statistical point of view, one can think about considering random Markov matrices following the "uniform law" over the set of Markov matrices, which corresponds to a maximum entropy distribution or Bayesian prior, see for example [DR06]. Recall that a $n \times n$ square real matrix $\mathbf{M}$ is Markov if and only if its entries are non–negative and each row sums up to 1, i.e. if and only if each row of $\mathbf{M}$ belongs to the simplex

$$\Lambda_n = \{(x_1, \ldots, x_n) \in [0,1]^n \text{ such that } x_1 + \cdots + x_n = 1\} \tag{1}$$

which is the portion of the unit $\|\cdot\|_1$-sphere of $\mathbb{R}^n$ with non–negative coordinates. The spectrum of a Markov matrix lies in the unit disc $\{z \in \mathbb{C}; |z| \leqslant 1\}$, contains 1, and is symmetric with respect to the real axis in the complex plane.

Uniform distribution on Markov matrices

Let $\mathcal{M}_n$ be the set of $n \times n$ Markov matrices. We need to give a precise meaning to the notion of uniform distribution on $\mathcal{M}_n$. This set is a convex compact polytope with $n(n-1)$ degrees of freedom if $n > 1$. It has zero Lebesgue measure in $\mathbb{R}^{n^2}$.

Since $\mathcal{M}_n$ is a polytope of $\mathbb{R}^{n^2}$ (i.e. intersection of half spaces), the trace of the Lebesgue measure on it makes sense and coincides with a cone measure[1], despite its zero Lebesgue measure in $\mathbb{R}^{n^2}$. Since $\mathcal{M}_n$ is additionally compact, the trace of the Lebesgue measure can be normalized into a probability distribution. We thus define *the uniform distribution* $\mathcal{U}(\mathcal{M}_n)$ on $\mathcal{M}_n$ as the normalized trace of the Lebesgue measure of $\mathbb{R}^{n^2}$. The following theorem relates $\mathcal{U}(\mathcal{M}_n)$ to the Dirichlet distribution.

Theorem 1.1 (Dirichlet Markov Ensemble)**.** *We have $\mathbf{M} \sim \mathcal{U}(\mathcal{M}_n)$ if and only if the rows of $\mathbf{M}$ are i.i.d. and follow the Dirichlet law of mean $(\frac{1}{n}, \ldots, \frac{1}{n})$. The probability distribution $\mathcal{U}(\mathcal{M}_n)$ is invariant by permutations of rows and columns.*

Corollary 1.2. *If $\mathbf{M} \sim \mathcal{U}(\mathcal{M}_n)$ then for every $1 \leqslant i, j \leqslant n$, $\mathbf{M}_{i,j} \sim \mathrm{Beta}(1, n-1)$ and for every $1 \leqslant i, i', j, j' \leqslant n$,*

$$\mathrm{Cov}(\mathbf{M}_{i,j}, \mathbf{M}_{i',j'}) = \begin{cases} 0 & \text{if } i \neq i' \\ \frac{n-1}{n^2(n+1)} & \text{if } i = i' \text{ and } j = j' \\ -\frac{1}{n^2(n+1)} & \text{if } i = i' \text{ and } j \neq j'. \end{cases}$$

Moreover $\mathbf{M}_{i,j}$ and $\mathbf{M}_{i',j'}$ are independent if and only if $i \neq i'$.

The set $\mathcal{M}_n$ is also a compact semi–group for the matrix product. The following two theorems concern the translation invariance of $\mathcal{U}(\mathcal{M}_n)$ and the question of the existence of an idempotent probability distribution on $\mathcal{M}_n$.

Theorem 1.3 (Translation invariance)**.** *For every $\mathbf{T} \in \mathcal{M}_n$, the law $\mathcal{U}(\mathcal{M}_n)$ is invariant by the left translation $\mathbf{M} \mapsto \mathbf{TM}$ if and only if $\mathbf{T}$ is a permutation matrix. The same holds true for the right translation $\mathbf{M} \mapsto \mathbf{MT}$.*

Theorem 1.4 (Idempotent distributions)**.** *There is no probability distribution on $\mathcal{M}_n$, absolutely continuous with respect to $\mathcal{U}(\mathcal{M}_n)$, with full support, and which is invariant by every left translations $\mathbf{M} \mapsto \mathbf{TM}$ where $\mathbf{T}$ runs over $\mathcal{M}_n$. The same holds true for right translations.*

The proofs of theorems 1.1, 1.3, 1.4 and corollary 1.2 are given in section 2.

[1] Actually, one can define the trace of the Lebesgue measure and then the uniform distribution on many compact subsets of the Euclidean space, by using the notion of Hausdorff measure [Fal03]. See also [CPSV09] for an approximate simulation method based on billiards and random reflections.

Asymptotic behavior of singular values and eigenvalues

The spectral properties of large dimensional random matrices are connected to many areas of mathematics, see for instance the books [Meh04, HP00, BS06, AGZ09, For09, ER05] and the survey [Bai99]. If $\mathbf{M} \sim \mathcal{U}(\mathcal{M}_n)$, then almost surely, the real matrix $\mathbf{M}$ is invertible, non–normal, with neither independent nor centered entries. The singular values of certain large dimensional centered random matrices with independent rows is considered for instance in [Aub06] and [MP06, PP07].

For any square $n \times n$ matrix $\mathbf{A}$ with real or complex entries, let the complex eigenvalues $\lambda_1(\mathbf{A}), \ldots, \lambda_n(\mathbf{A})$ of $\mathbf{A}$ be labeled so that $|\lambda_1(\mathbf{A})| \geqslant \cdots \geqslant |\lambda_n(\mathbf{A})|$. The *spectral radius* of $\mathbf{A}$ is thus given by $|\lambda_1(\mathbf{A})| = \max_{1 \leqslant k \leqslant n} |\lambda_k(\mathbf{A})|$. The *empirical spectral distribution* (ESD) of $\mathbf{A}$ is the discrete probability distribution on $\mathbb{C}$ with at most n atoms defined by

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

The *singular values* $s_1(\mathbf{A}) \geqslant \cdots \geqslant s_n(\mathbf{A}) \geqslant 0$ of $\mathbf{A}$ are the eigenvalues of the positive semi–definite Hermitian matrix $(\mathbf{A}\mathbf{A}^*)^{1/2}$ where

$$\mathbf{A}^* = \overline{\mathbf{A}}^\top$$

denotes the conjugate transpose of $\mathbf{A}$. Namely, for every $1 \leqslant k \leqslant n$,

$$s_k(\mathbf{A}) = \lambda_k(\sqrt{\mathbf{A}\mathbf{A}^*}) = \sqrt{\lambda_k(\mathbf{A}\mathbf{A}^*)}.$$

Note that $\mathbf{A}\mathbf{A}^*$ and $\mathbf{A}^*\mathbf{A}$ share the same spectrum. The atoms of the ESD of $\sqrt{\mathbf{A}\mathbf{A}^*}$ are $s_1(\mathbf{A}), \ldots, s_n(\mathbf{A})$. The singular values of $\mathbf{A}$ have a clear geometrical interpretation: the linear operator $\mathbf{A}$ maps the unit ball to an ellipsoid, and the singular values of $\mathbf{A}$ are exactly the half–lengths of its principal axes. In particular, $s_1(\mathbf{A}) = \max_{\|x\|_2=1} \|\mathbf{A}x\|_2 = \|\mathbf{A}\|_{2\to2}$ while $s_n(\mathbf{A}) = \min_{\|x\|_2=1} \|\mathbf{A}x\|_2 = \|\mathbf{A}^{-1}\|_{2\to2}^{-1}$. Moreover, $\mathbf{A}$ has exactly $\mathrm{rank}(\mathbf{A})$ non zero singular values. The relationship between the eigenvalues and the singular values are captured by the Weyl–Horn inequalities

$$\forall k \in \{1, \ldots, n\}, \quad \prod_{i=1}^{k} |\lambda_i(\mathbf{A})| \leqslant \prod_{i=1}^{k} s_i(\mathbf{A}) \quad \text{with equality when } k = n,$$

see [Hor54, Wey49]. If $\mathbf{A}$ is normal, i.e. $\mathbf{A}\mathbf{A}^* = \mathbf{A}^*\mathbf{A}$, then $s_k(\mathbf{A}) = |\lambda_k(\mathbf{A})|$ for every $1 \leqslant k \leqslant n$. Back to our Dirichlet Markov Ensemble, if $\mathbf{M} \sim \mathcal{U}(\mathcal{M}_n)$ then $\mathbf{M}$ is almost surely a non–normal matrix, and thus one cannot express the singular values of $\mathbf{M}$ in terms of the eigenvalues of $\mathbf{M}$. The following theorem gives the asymptotic behavior of the empirical distribution built from the singular values of $\mathbf{M}$.

Theorem 1.5 (Singular values for Dirichlet Markov Ensemble)**.** *Let $(X_{i,j})_{1 \leqslant i,j < \infty}$ be an infinite array of i.i.d. exponential random variables of unit mean. For every n, let $\mathbf{M}$ be the $n \times n$ random matrix defined for every $1 \leqslant i, j \leqslant n$ by*

$$\mathbf{M}_{i,j} = \frac{X_{i,j}}{\sum_{k=1}^{n} X_{i,k}}.$$

Probability distribution name	Support	Lebesgue density
Circle or circular law $\mathcal{C}_\sigma$	$\{z \in \mathbb{C}; \lvert z\rvert \leqslant \sigma\} \subset \mathbb{C}$	$z \mapsto (\pi\sigma^2)^{-1}$
Wigner semi–circle distribution $\mathcal{W}_\sigma$	$[-2\sigma, +2\sigma] \subset \mathbb{R}$	$x \mapsto (2\pi\sigma^2)^{-1}\sqrt{4\sigma^2 - x^2}$
Wigner quarter–circle distribution $\mathcal{Q}_\sigma$	$[0, 2\sigma] \subset \mathbb{R}$	$x \mapsto (\pi\sigma^2)^{-1}\sqrt{4\sigma^2 - x^2}$
Marchenko–Pastur distribution $\mathcal{P}_\sigma$	$[0, 4\sigma^2] \subset \mathbb{R}$	$x \mapsto (2\pi\sigma^2 x)^{-1}\sqrt{x(4\sigma^2 - x)}$

Table 1: Some of the remarkable probability distributions in random matrices.

Then $\mathbf{M} \sim \mathcal{U}(\mathcal{M}_n)$ *and*

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n}\delta_{\lambda_k(n\mathbf{M}\mathbf{M}^\top)} \xrightarrow[n\to\infty]{w} \mathcal{P}_1\right) = 1$$

where $\xrightarrow{w}$ *denotes the weak convergence of probability distributions and* $\mathcal{P}_1$ *the Marchenko–Pastur distribution defined in table 1. In other words,*

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n}\delta_{s_k(\sqrt{n}\mathbf{M})} \xrightarrow[n\to\infty]{w} \mathcal{Q}_1\right) = 1$$

where $\mathcal{Q}_1$ *denotes the Wigner quarter–circle distribution defined in table 1.*

Following the notations of table 1, for every real fixed parameter $\sigma > 0$, every real random variable W, and every complex random variable $Z = U + \sqrt{-1}V$ with $U = \mathrm{RealPart}(Z)$ and $V = \mathrm{ImaginaryPart}(Z)$, we have, by a change of variables,

$$\big(W^2 \sim \mathcal{P}_\sigma \Leftrightarrow |W| \sim \mathcal{Q}_\sigma\big) \quad \text{and} \quad \big(W \sim \mathcal{W}_\sigma \Rightarrow W^2 \sim \mathcal{P}_\sigma \text{ and } |W| \sim \mathcal{Q}_\sigma\big).$$

Moreover, we have, simply by using the Cramér-Wold theorem,

$$Z \sim \mathcal{C}_{2\sigma} \;\Leftrightarrow\; \Big(\mathrm{RealPart}(e^{\sqrt{-1}\theta}Z) \sim \mathcal{W}_\sigma \text{ for every } \theta \in [0, 2\pi)\Big).$$

In particular, we have

$$Z \sim \mathcal{C}_{2\sigma} \;\Rightarrow\; U \sim \mathcal{W}_\sigma \text{ and } V \sim \mathcal{W}_\sigma.$$

Beware however that U and V are not independent random variables! Furthermore, if $\mathbb{P}(|Z| = \sigma; V \geqslant 0) = 1$ then Z follows the uniform distribution over the upper half circle of radius σ if and only if U follows the so–called arc–sine distribution on $[-\sigma, +\sigma] \subset \mathbb{R}$ with Lebesgue density $x \mapsto (\pi\sqrt{\sigma^2 - x^2})^{-1}$.

The proof of theorem 1.5 is given in section 3. Since $|\lambda_1(\mathbf{A})| \leqslant s_1(\mathbf{A})$ for any square matrix $\mathbf{A}$, and since $\lambda_1(\mathbf{M}) = 1$, we have for every $n \geqslant 1$

$$s_1(\mathbf{M}) \geqslant |\lambda_1(\mathbf{M})| = 1.$$

However, theorem 1.5 implies in particular that almost surely

$$\frac{1}{n}\,\mathrm{Card}\left\{1 \leqslant k \leqslant n \text{ such that } s_k(\mathbf{M}) > \frac{2}{\sqrt{n}}\right\} \underset{n\to\infty}{\longrightarrow} 0.$$

Random Q-matrices

Bryc, Dembo, and Jiang studied in [BDJ06] the limiting spectral distribution of *random Hankel, Markov, and Toeplitz matrices*. Let us explain briefly what they mean by "random Markov matrices". They proved the following theorem (see [BDJ06, th. 1.3] and also [BS08]) : let $(\mathbf{X}_{i,j})_{1<i<j<\infty}$ be an infinite triangular array of i.i.d. real random variables of mean 0 and variance 1. Let $\mathbf{Q}$ be the symmetric $n \times n$ random matrix defined for every $1 \leqslant i \leqslant j \leqslant n$ by $\mathbf{Q}_{i,j} = \mathbf{Q}_{j,i} = X_{i,j}$ if $i < j$, and

$$\mathbf{Q}_{i,i} = -\sum_{\substack{1 \leqslant k \leqslant n \\ k \neq i}} \mathbf{Q}_{i,k} \quad \text{for every } 1 \leqslant i \leqslant n.$$

Then, almost surely, the ESD of $n^{-1/2}\,\mathbf{Q}$ converges as $n \to \infty$ to the free convolution[2] of a semi–circle law and a standard Gaussian law.

This result gives an answer to a precise question raised by Bai in his 1999 review article [Bai99, sec. 6.1.1]. The matrix $\mathbf{Q}$ is not Markov. However, it looks like a Markov generator, i.e. a Q-matrix, since its rows sum up to 0. Unfortunately, the assumptions do not allow the off–diagonal entries of $\mathbf{Q}$ to have non–negative support, and thus $\mathbf{Q}$ cannot be almost surely a Markov generator. In particular, if $\mathbf{I}$ stands for the identity matrix of size $n \times n$, the symmetric matrix $\mathbf{M} = \mathbf{Q} + \mathbf{I}$ cannot be almost surely Markov.

Eigenvalues and the circular law

If $\mathbf{M}$ is as in theorem 1.5, then $\lambda_1(\sqrt{n}\,\mathbf{M}) = \sqrt{n}$ goes to $+\infty$ as $n \to \infty$ while its weight in the ESD is $1/n$. Thus, it does not contribute to the limiting spectral distribution of $\sqrt{n}\,\mathbf{M}$. Numerical simulations (see figure 1) suggest that the empirical distribution of the rest of the spectrum tends as $n \to \infty$ to the uniform distribution on the unit disc. One can formulate this conjecture as follows.

Conjecture 1.6 (Circle law for the Dirichlet Markov Ensemble)**.** *If* $\mathbf{M}$ *is as in theorem 1.5, then*

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=1}^{n} \delta_{\lambda_k(\sqrt{n}\,\mathbf{M})} \xrightarrow[n\to\infty]{w} \mathcal{C}_1\right) = 1$$

where $\overset{w}{\to}$ *denotes the weak convergence of probability distributions and* $\mathcal{C}_1$ *the uniform distribution over the unit disc* $\{z \in \mathbb{C}; |z| \leqslant 1\}$ *as defined in table 1.*

The main difficulty in conjecture 1.6 lies in the fact that $\mathbf{M}$ is non–normal with non i.i.d. entries. The limiting spectral distributions of non–normal random matrices is a notoriously difficult subject, see for instance [TVK08]. The method used for the singular values for the proof of theorem 1.5 fails for the eigenvalues, due to the lack of variational formulas for the eigenvalues. In contrast to singular values, the eigenvalues of non–normal matrices are very sensitive to perturbations, a phenomenon captured by the notion of pseudo–spectrum [TE05]. The reader may find in [Cha08] a more general version of theorem 3.1 which goes beyond the exponential case, and some partial answers to conjecture 1.6.

[2]This limiting spectral distribution is a symmetric law on $\mathbb{R}$ with smooth bounded density of unbounded support. See [HP00] or [Bia97] for Voiculescu's free convolution.

Sub–dominant eigenvalue

The fact that non–centered entries produce an explosive extremal eigenvalue was already noticed in various situations, see for instance [And90], [Sil94], [BDJ06, th. 1.4], [BS07], and [Cha07]. It is natural to ask about the asymptotic behavior (convergence and fluctuations) of the sub–dominant eigenvalue $\lambda_2(\mathbf{M})$ when $\mathbf{M} \sim \mathcal{U}(\mathcal{M}_n)$. The reader may find some answers in [GN03, GONS00], and may forge new conjectures from our simulations (see figures 2 and 3). For instance, by analogy with the Complex Ginibre Ensemble [Kos92, Rid03], one can state the following:

Conjecture 1.7 (Behavior of sub–dominant eigenvalue and spectral gap). *If $\mathbf{M}$ is as in theorem 1.5, then $\lambda_1(\mathbf{M}) = 1$ while*

$$\mathbb{P}\Big(\lim_{n\to\infty} \sqrt{n}\,|\lambda_2(\mathbf{M})| = 1\Big) = 1.$$

In particular, the spectral gap $1 - |\lambda_2(\mathbf{M})|$ of $\mathbf{M}$ is of order $1 - 1/\sqrt{n}$ for large n. Moreover, there exist deterministic sequences (a_n) and (b_n) and a probability distribution $\mathcal{G}$ on $\mathbb{R}$ such that

$$b_n(|\lambda_2(\mathbf{M})| - a_n) \xrightarrow[n\to\infty]{d} \mathcal{G}$$

where $\xrightarrow{d}$ denotes the convergence in law.

There is not clear indication that $\mathcal{G}$ is a Gumbel distribution as for the Complex Ginibre Ensemble. Moreover, our simulations suggest that the sub–dominant eigenvalue is real with positive probability (depends on n), which is not surprising knowing [Ede97, EKS94]. Note that Goldberg and Neumann have shown [GN03] that if $\mathbf{X}$ is an $n\times n$ random matrix with i.i.d. rows such that for every $1 \leqslant i, j, j' \leqslant n$,

$$\mathbb{E}[\mathbf{X}_{i,j}] = \frac{1}{n}, \quad \text{and} \quad \mathrm{Var}(\mathbf{X}_{i,j}) = O\left(\frac{1}{n^2}\right), \quad \text{and} \quad |\mathrm{Cov}(\mathbf{X}_{i,j}, \mathbf{X}_{i,j'})| = O\left(\frac{1}{n^3}\right)$$

then $\mathbb{P}(|\lambda_2(\mathbf{X})| \leqslant r) \geqslant p$ for any $p \in (0,1)$, any $0 < r < 1$, and large enough n. This is the case if we set $\mathbf{X} = \mathbf{M}$.

Other distributions

The Dirichlet distribution of dimension n and mean $(\frac{1}{n}, \dots, \frac{1}{n})$ is the uniform distribution on the simplex Λ_n defined by (1). One can replace the uniform distribution by a Dirichlet distribution of dimension n and arbitrary mean. The argument used in the proof of theorem 1.5 remains the same due to the very similar construction of Dirichlet distributions by projection from i.i.d. Gamma random variables. One can also replace the $\|\cdot\|_1$-norm by any other $\|\cdot\|_p$-norm, and investigate the limiting spectral distribution of the corresponding random matrices. This case can be handled with the construction of the uniform distribution by projection proposed in [SZ90]. Replacing the non–negative portion of spheres by the non–negative portion of balls is also possible by using [BGMN05]. More generally, one can consider random matrices with independent rows. The case of the uniform distribution on the whole unit $\|\cdot\|_p$–ball of $\mathbb{R}^n$ is considered for instance by in [Aub06] by using [BGMN05] together with random matrices results for i.i.d. centered entries. It is crucial here to have an explicit construction of the distribution from an i.i.d. array. For the link with the sampling of convex bodies, see [Aub07]. The case of matrices with i.i.d. rows following a log-concave isotropic distribution is considered in the recent work [PP07], by using recently developed results on log-concave measures. The reader may find a universal version of theorem 3.1 in [Cha08], where the exponential law is replaced by an arbitrary law.

Doubly Stochastic matrices

The Birkhoff or transportation polytope is the set of $n\times n$ doubly stochastic matrices, i.e. matrices which are Markov and have a Markov transpose. Each $n \times n$ doubly stochastic matrix corresponds to a transportation map of n unit masses into n boxes of unit mass (matching), and conversely, each transportation map of this kind is a $n \times n$ doubly stochastic matrix. Geometrically, the Birkhoff polytope is a convex compact subset of $\mathcal{M}_n$ of zero Lebesgue measure in $\mathbb{R}^{n^2}$ and $(n-1)^2$ degrees of freedom if $n > 1$. As for $\mathcal{M}_n$, one can define the uniform distribution as the normalized trace of the Lebesgue measure. However, we ignore if this distribution has a probabilistic representation that allows exact simulation as for $\mathcal{U}(\mathcal{M}_n)$. The spectral behavior of random doubly stochastic matrices was considered in the Physics literature, see for instance [Ber01]. On the purely discrete side, the Birkhoff polytope is also related to magic squares, transportation polytopes and contingency tables, see [DE87, DE85] and [DG95]. Note also that if $\mathbf{M}$ is Markov, then $\mathbf{M}\mathbf{M}^\top$ and $\frac{1}{2}(\mathbf{M}+\mathbf{M}^\top)$ are not Markov in general. However, this is the case when $\mathbf{M}$ is doubly stochastic. The Birkhoff-von Neumann theorem states that the extremal points of the Birkhoff polytope are exactly the permutation matrices. The reader may find nice spectral results on random uniform permutation matrices in [HKOS00, Wie00] and references therein.

Another interesting polytope of matrices is the set of symmetric $n \times n$ Markov matrices, which is a convex compact polytope of zero Lebesgue measure in $\mathbb{R}^{n^2}$ with $\frac{1}{2}n(n-1)$ degrees of freedom if $n > 1$. As for $\mathcal{M}_n$, one can define the uniform distribution as the normalized trace of the Lebesgue measure. However, we ignore if this distribution has a probabilistic representation that allows simulation as for $\mathcal{U}(\mathcal{M}_n)$. One can ask about the spectral properties of the corresponding random symmetric Markov matrices. Note that these matrices are doubly stochastic, but the converse is false except when $n = 1$ or $n = 2$. Our construction of $\mathcal{U}(\mathcal{M}_n)$ in theorem 1.5 corresponds in the Markovian probabilistic jargon to a random conductance model on the complete oriented graph. The study of the spectral properties of random reversible Markov conductance models on the complete non–oriented graph can be found in [Cha09, BCC08, BCC09]. For other graphs, the reader may find some clues in [BDPX05].

Let $\mathbf{M}$ be as in theorem 1.5. Numerical simulations suggest that almost surely, the ESD of the symmetric matrix $\frac{1}{2}(\mathbf{M}+\mathbf{M}^\top)$ tends, as $n \to \infty$, to a semi–circle Wigner distribution.

If $\mathbf{U}$ is an $n \times n$ unitary matrix, then $(|\mathbf{U}_{i,j}|^2)_{1\leqslant i,j\leqslant n}$ is a doubly stochastic matrix. These doubly stochastic matrices are called *uni–stochastic* or *unitary-*

stochastic. There exists doubly stochastic matrices which are not uni–stochastic, see [BEK$^+$05] and [Tan01]. However, every permutation matrix is orthogonal and thus uni–stochastic. The Haar measure on the unitary group induces a probability distribution on the set of uni–stochastic matrices. How about the asymptotic spectral properties of the corresponding random matrices?

Perron–Frobenius eigenvector (invariant vector)

If $\mathbf{M} \sim \mathcal{U}(\mathcal{M}_n)$, then almost surely, all the entries of $\mathbf{M}$ are non-zero, and in particular, $\mathbf{M}$ is almost surely recurrent irreducible and aperiodic. By a theorem of Perron and Frobenius [Sen06], it follows that almost surely, the eigenspace of $\mathbf{M}^\top$ associated to the eigenvalue 1 is of dimension 1 and contains a unique vector with non–negative entries and unit $\|\cdot\|_1$-norm. One can ask about the asymptotic behavior of this vector as $n \to \infty$. For a fixed n, the distribution of this vector is the distribution of the rows of the infinite product of random matrices $\lim_{k\to\infty} \mathbf{M}^k$.

2 Structure of the Dirichlet Markov Ensemble

Let Λ_n be as in (1). For any $a \in (0,\infty)^n$, the Dirichlet distribution $\mathcal{D}_n(a_1,\dots,a_n)$, supported by Λ_n, is defined as the distribution of

$$\frac{1}{\|G\|_1} G = \left(\frac{G_1}{G_1+\cdots+G_n},\dots,\frac{G_n}{G_1+\cdots+G_n}\right)$$

where G is a random vector of $\mathbb{R}^n$ with independent entries with $G_i \sim \mathrm{Gamma}(1,a_i)$ for every $1 \leqslant i \leqslant n$. Here $\mathrm{Gamma}(\lambda,a)$ has density

$$t \mapsto \frac{\lambda^a}{\Gamma(a)} t^{a-1} e^{-\lambda t}\, \mathrm{I}_{(0,\infty)}(t),$$

where $\Gamma(a) = \int_0^\infty t^{a-1} e^{-t}\,dt$ is the Euler Gamma function. Let $P \sim \mathcal{D}_n(a_1,\dots,a_n)$. For every partition $I_1,\dots,I_k$ of $\{1,\dots,n\}$ into k non empty subsets, we have

$$\left(\sum_{i\in I_1} P_i,\dots,\sum_{i\in I_k} P_i\right) \sim \mathcal{D}_k\left(\sum_{i\in I_1} a_i,\dots,\sum_{i\in I_k} a_i\right).$$

The mean and covariance matrix of $\mathcal{D}_n(a_1,\dots,a_n)$ are given by

$$\frac{1}{\|a\|_1} a \quad \text{and} \quad \frac{1}{\|a\|_1^2(1+\|a\|_1)}(\|a\|_1 \mathrm{diag}(a) - aa^\top)$$

where $a = (a_1,\dots,a_n)^\top$ and $\mathrm{diag}(a)$ is the diagonal matrix with diagonal given by a. For any non-empty subset I of $\{1,\dots,n\}$, we have

$$\sum_{i\in I} P_i \sim \mathrm{Beta}\left(\sum_{i\in I} a_i, \sum_{i\notin I} a_i\right),$$

where $\mathrm{Beta}(\alpha,\beta)$ denotes the Euler Beta distribution on $[0,1]$ of Lebesgue density

$$t \mapsto \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} t^{\alpha-1}(1-t)^{\beta-1}\,\mathrm{I}_{[0,1]}(t).$$

If $P_I = (P_i)_{i\in I}$, $P_{I^c} = (P_i)_{i\notin I}$, $a_I = (a_i)_{i\in I}$, and $|I| = \mathrm{card}(I)$, then

$$\frac{1}{\sum_{i\in I} P_i} P_I \text{ and } P_{I^c} \text{ are independent and } \frac{1}{\sum_{i\in I} P_i} P_I \sim \mathcal{D}_{|I|}(a_I),$$

For any $\alpha > 0$, the Dirichlet distribution $\mathcal{D}_n(\alpha,\ldots,\alpha)$ is exchangeable, with negatively correlated components. More generally, if $P \sim \mu$ where μ is an exchangeable probability distribution on the simplex Λ_n with $n > 1$, then

$$0 = \mathrm{Var}(1) = \mathrm{Var}(P_1 + \cdots + P_n) = n\mathrm{Var}(P_1) + n(n-1)\mathrm{Cov}(P_1,P_2).$$

Consequently, $\mathrm{Cov}(P_1,P_2) = -(n-1)^{-1}\mathrm{Var}(P_1)$ and in particular $\mathrm{Cov}(P_1,P_2) \leqslant 0$.

We refer for instance to [Wil62] for other properties of Dirichlet distributions. Corollary 1.2 follows immediately from theorem 1.1 together with the basic properties of the Dirichlet distributions mentioned above.

Proof of theorem 1.1. As a subset of $\mathbb{R}^n$, the simplex Λ_n defined by (1) is of zero Lebesgue measure. However, by considering Λ_n as a convex subset of the hyper-plane of equation $x_1+\cdots+x_n = 1$ or by using the general notion of Hausdorff measure, one can see that in fact, the Dirichlet distribution $\mathcal{D}_n(1,\ldots,1)$ is the normalized trace of the Lebesgue measure of $\mathbb{R}^n$ on the simplex Λ_n. In other words, $\mathcal{D}_n(1,\ldots,1)$ can be seen as the uniform distribution on Λ_n, see [SZ90].

We identify $\mathcal{M}_n$ with $(\Lambda_n)^n = \Lambda_n \times \cdots \times \Lambda_n$ where Λ_n is repeated n times. The trace of the Lebesgue measure of $\mathbb{R}^{n^2} = (\mathbb{R}^n)^n$ on $(\Lambda_n)^n$ is the n-tensor product of the trace of the Lebesgue measure of $\mathbb{R}^n$ on Λ_n, i.e. the n-tensor product measure $\mathcal{D}_n(1,\ldots,1)^{\otimes n}$. Consequently, for every positive integer n,

$$(\mathcal{M}_n, \mathcal{U}(\mathcal{M}_n)) = ((\Lambda_n)^n, \mathcal{D}_n(1,\ldots,1)^{\otimes n}).$$

This gives the invariance of $\mathcal{U}(\mathcal{M}_n)$ by permutation of rows. If $\mathbf{M} \sim \mathcal{U}(\mathcal{M}_n)$, then the rows of $\mathbf{M}$ are i.i.d. and follow the Dirichlet distribution $\mathcal{D}_n(1,\ldots,1)$. Finally, the invariance of $\mathcal{U}(\mathcal{M}_n)$ by permutation of columns comes from the exchangeability of the Dirichlet distribution $\mathcal{D}_n(1,\ldots,1)$. □

Recursive simulation

The simulation of $\mathcal{U}(\mathcal{M}_n)$ follows from the simulation of n i.i.d. realizations of $\mathcal{D}_n(1,\ldots,1)$ by using n^2 i.i.d. exponential random variables. The elements of Dyson's classical Gaussian ensembles GUE and GOE can be simulated recursively by adding a new independent line/column. It is natural to ask about a recursive method for the Dirichlet Markov Ensemble. If

$$X \sim \mathcal{D}_{n-1}(a_2,\ldots,a_n) \quad \text{and} \quad Y \sim \mathrm{Beta}(a_1, a_2+\cdots+a_n)$$

are independent, then

$$(Y,(1-Y)X) \sim \mathcal{D}_n(a_1,\ldots,a_n).$$

This recursive simulation of Dirichlet distributions is known as the *stick–breaking* algorithm [Set94]. It allows to simulate $\mathcal{U}(\mathcal{M}_n)$ recursively on n. Namely, if $\mathbf{M}$ is such that $\mathbf{M} \sim \mathcal{U}(\mathcal{M}_n)$, then

$$\begin{pmatrix} Y & (1-Y)\cdot\mathbf{M} \\ Z_1 & Z_2 \;\cdots\; Z_n \end{pmatrix} \sim \mathcal{U}(\mathcal{M}_{n+1})$$

where Z is a random row vector of $\mathbb{R}^{n+1}$ with $Z \sim \mathcal{D}_{n+1}(1,\ldots,1)$ and Y is a random column vector of $\mathbb{R}^n$ with i.i.d. entries of law $\mathrm{Beta}(1,n)$, with $\mathbf{M}, Y, Z$ independent. Here $((1-Y)\cdot\mathbf{M})_{i,j} := (1-Y)_i\mathbf{M}_{i,j}$ for every $1 \leqslant i,j \leqslant n$.

Asymptotic behavior of the rows

Let $\mathbf{M}$ and $(X_{i,j})_{1\leqslant i,j<\infty}$ be as in theorem 1.5. Let us fix $k \geqslant 1$ and $n \geqslant i \geqslant 1$. The k^{th} moment $m_{n,i,k}$ of the discrete probability distribution $\frac{1}{n}\sum_{j=1}^n \delta_{n\mathbf{M}_{i,j}}$ is given by

$$\begin{aligned} m_{n,i,k} &= \frac{1}{n}\sum_{j=1}^n (n\mathbf{M}_{i,j})^k \\ &= \sum_{j=1}^n \frac{n^k}{n}\frac{X_{i,j}^k}{(X_{i,1}+\cdots+X_{i,n})^k} \\ &= \frac{n^k}{(X_{i,1}+\cdots+X_{i,n})^k}\frac{X_{i,1}^k+\cdots+X_{i,n}^k}{n}. \end{aligned}$$

Therefore, by using twice the strong law of large numbers, we get that almost surely,

$$\lim_{n\to\infty} m_{n,i,k} = \frac{\mathbb{E}[X_{1,1}^k]}{\mathbb{E}[X_{1,1}]^k} = \mathbb{E}[X_{1,1}^k].$$

As a consequence, almost surely, for any fixed $i \geqslant 1$ and every $k \geqslant 1$,

$$\lim_{n\to\infty} W_k\left(\frac{1}{n}\sum_{j=1}^n \delta_{n\mathbf{M}_{i,j}}\,;\,\mathcal{E}_1\right) = 0,$$

where $\mathcal{E}_1 = \mathcal{L}(X_{1,1})$ is the exponential law on unit mean and where $W_k(\cdot\,;\,\cdot)$ is the so called Wasserstein–Mallows coupling distance of order k (see for instance [Vil03] or [Rac91]). This result is a special case of a more general well known phenomenon (sometimes referred as the Poincaré observation) concerning the coordinates of a uniformly distributed random point on the unit $\|\cdot\|_p$–sphere of $\mathbb{R}^n$ with $1 \leqslant p < \infty$ when $n \to \infty$, see for instance [NR03], [Jia09], and references therein.

Semi–group structure and translation invariance

The set $\mathcal{M}_n$ is a semi–group for the usual matrix product. In particular, for every $\mathbf{T} \in \mathcal{M}_n$, the set $\mathcal{M}_n$ is stable by the left translation $\mathbf{M} \mapsto \mathbf{TM}$ and the right translation $\mathbf{M} \mapsto \mathbf{MT}$. When $\mathbf{T}$ is a permutation matrix, then these translations are bijective maps, and the left translation (respectively right) translation corresponds to rows (respectively columns) permutations.

For some fixed $\mathbf{T} \in \mathcal{M}_n$, let us consider the left translation $\mathbf{M} \mapsto \mathbf{TM}$, where $\mathbf{M} \sim \mathcal{U}(\mathcal{M}_n)$. By linearity, we have

$$\mathbb{E}[\mathbf{TM}] = \mathbf{T}\mathbb{E}[\mathbf{M}] = \mathbf{T}\frac{1}{n}\mathbf{1} = \frac{1}{n}\mathbf{1}$$

where $\mathbf{1}$ is the $n \times n$ matrix full of ones. Thus, the left translation by $\mathbf{T}$ leaves the mean invariant.

Proof of theorem 1.3. First of all, the case $n = 1$ is trivial and one can assume that $n > 1$ in the rest of the proof. A probability distribution μ on $\mathcal{M}_n$ is invariant by the left translation $\mathbf{M} \mapsto \mathbf{PM}$ for every permutation matrix $\mathbf{P}$ of size $n \times n$ if and only if μ is row exchangeable. Similarly, μ is invariant by the right translation $\mathbf{M} \mapsto \mathbf{MP}$ for every permutation matrix $\mathbf{P}$ of size $n \times n$ if and only if μ is column exchangeable. Theorem 1.1 gives then the invariance of $\mathcal{U}(\mathcal{M}_n)$ by left and right translations with respect to permutation matrices[3].

Conversely, let us assume that the law $\mathcal{U}(\mathcal{M}_n)$ is invariant by the left translation $\mathbf{M} \mapsto \mathbf{TM}$ for some $\mathbf{T} \in \mathcal{M}_n$. If $\mathbf{M} \sim \mathcal{U}(\mathcal{M}_n)$, and since the components of the first column $\mathbf{M}_{\cdot,1}$ of $\mathbf{M}$ are i.i.d. we have

$$\begin{aligned}
\mathrm{Var}((\mathbf{TM})_{1,1}) &= \mathrm{Var}\left(\sum_{k=1}^{n} \mathbf{T}_{1,k}\mathbf{M}_{k,1}\right) \\
&= \sum_{k=1}^{n} (\mathbf{T}_{1,k})^2 \mathrm{Var}(\mathbf{M}_{k,1}) \\
&= \mathrm{Var}(\mathbf{M}_{1,1}) \sum_{k=1}^{n} (\mathbf{T}_{1,k})^2.
\end{aligned}$$

The invariance hypothesis implies in particular that $\mathrm{Var}(\mathbf{M}_{1,1}) = \mathrm{Var}((\mathbf{TM})_{1,1})$. Since $\mathrm{Var}(\mathbf{M}_{1,1}) = (n-1)/(n^2(n+1)) > 0$, we get $1 = \sum_{k=1}^{n} (\mathbf{T}_{1,k})^2$. Now, $\mathbf{T}$ is Markov and thus $\sum_{k=1}^{n} \mathbf{T}_{1,k} = 1$, which gives

$$\sum_{k=1}^{n} (\mathbf{T}_{1,k} - (\mathbf{T}_{1,k})^2) = 0.$$

Since $\mathbf{T}$ is Markov, its entries are in $[0,1]$ and hence $\mathbf{T}_{1,k} \in \{0,1\}$ for every $1 \leqslant k \leqslant n$. The condition $\sum_{k=1}^{n} \mathbf{T}_{1,k} = 1$ gives then that the first line of $\mathbf{T}$ is an element of the canonical basis of $\mathbb{R}^n$. The same argument used for $(\mathbf{TM})_{k,1}$ for every $1 \leqslant k \leqslant n$

[3]However that as a law over $\mathbb{R}^{n^2}$, $\mathcal{U}(\mathcal{M}_n)$ is not exchangeable. The permutation of rows and columns correspond to a proper subset of the group of permutations of the n^2 entries.

shows that every line of $\mathbf{T}$ is an element of the canonical basis, and thus $\mathbf{T}$ is a binary matrix with exactly a unique 1 on each row. Since $\mathbf{TM} \sim \mathcal{U}(\mathcal{M}_n)$, it has independent rows, and thus the position of the 1's on the rows of $\mathbf{T}$ are pairwise different, which means that $\mathbf{T}$ is a permutation matrix as expected.

Let us consider now the case where the law $\mathcal{U}(\mathcal{M}_n)$ is invariant by the right translation $\mathbf{M} \mapsto \mathbf{MT}$ for some $\mathbf{T} \in \mathcal{M}_n$. If $\mathbf{M} \sim \mathcal{U}(\mathcal{M}_n)$, we can first take a look at the mean. Namely, $\mathbb{E}[\mathbf{MT}] = \mathbb{E}[\mathbf{M}]\mathbf{T} = \frac{1}{n}\mathbf{S}$ where $\mathbf{S}$ is defined by

$$\mathbf{S}_{i,j} = \sum_{k=1}^{n} \mathbf{T}_{k,j}$$

for every $1 \leqslant i, j \leqslant n$. Now, the invariance hypothesis gives on the other hand

$$\mathbb{E}[\mathbf{MT}] = \mathbb{E}[\mathbf{M}] = \frac{1}{n}\mathbf{1}$$

and thus $\mathbf{S} = \mathbf{1}$, which means that $\mathbf{T}$ is doubly stochastic, i.e. both $\mathbf{T}$ and $\mathbf{T}^\top$ are Markov. The invariance hypothesis implies also that

$$\mathrm{Var}((\mathbf{MT})_{1,1}) = \mathrm{Var}(\mathbf{M}_{1,1}) = \frac{n-1}{n^2(n+1)}.$$

But since the first line $\mathbf{M}_{1,\cdot}$ of $\mathbf{M}$ is $\mathcal{D}_n(1,\ldots,1)$ distributed,

$$\begin{aligned}\mathrm{Var}((\mathbf{MT})_{1,1}) &= \sum_{1\leqslant i,j\leqslant n} \mathbf{T}_{i,1}\mathbf{T}_{j,1}\mathrm{Cov}(\mathbf{M}_{1,i};\mathbf{M}_{1,j}) \\ &= \frac{n-1}{n^2(1+n)}\sum_{i=1}^{n}(\mathbf{T}_{i,1})^2 - \frac{2}{n^2(n+1)}\sum_{1\leqslant i<j\leqslant n}\mathbf{T}_{i,1}\mathbf{T}_{j,1}.\end{aligned}$$

Since $\mathbf{T}$ is doubly stochastic, we have $1 = \sum_{i=1}^{n} \mathbf{T}_{i,1}$ and thus

$$(n-1)\sum_{i=1}^{n}(\mathbf{T}_{i,1} - (\mathbf{T}_{i,1})^2) = -2\sum_{1\leqslant i<j\leqslant n}\mathbf{T}_{i,1}\mathbf{T}_{j,1}.$$

The terms of the left and right hand side have opposite signs, which gives that $\mathbf{T}_{i,1} \in \{0,1\}$ for every $1 \leqslant i \leqslant n$. The same method used for $(\mathbf{MT})_{1,k}$ for every $1 \leqslant k \leqslant n$ shows that $\mathbf{T}$ is a binary matrix. Since $\mathbf{T}$ is doubly stochastic, it follows that $\mathbf{T}$ is actually a permutation matrix, as expected. $\square$

The set of $n \times n$ permutation matrices is a discrete subgroup of the orthogonal group of $\mathbb{R}^n$, isomorphic to the symmetric group Σ_n. The group of permutation matrices plays for the Dirichlet Markov Ensemble the role played by the orthogonal group for Dyson's GOE or COE, and the role played by the unitary group for Dyson's GUE or CUE. In some sense, we replaced an L^2 Gaussian structure by an L^1 Dirichlet structure while maintaining the permutation invariance.

A very natural question is to ask about the existence of a convolution idempotent probability distribution on the compact semi–group $\mathcal{M}_n$. Recall that a probability distribution μ on a semi–group $\mathfrak{S}$ is idempotent if and only if $\mu * \mu = \mu$. Here the convolution $\mu * \nu$ of two probability distributions μ and ν on $\mathfrak{S}$ is defined, for every bounded continuous $f : \mathfrak{S} \to \mathbb{R}$, by

$$\int_{\mathfrak{S}} f(s) d(\mu * \nu)(s) = \int_{\mathfrak{S}} \left(\int_{\mathfrak{S}} f(s_l s_r)\, d\mu(s_l) \right) d\nu(s_r).$$

Actually, the structure of compact semi–groups and their idempotent measures was deeply investigated in the 1960's, see [Ros71, p. 158-160] for a historical account. In particular, one can find in [Ros71, lem. 3] the following result.

Lemma 2.1. *Let μ be a regular probability distribution over a compact Hausdorff semi–group $\mathfrak{S}$ such that the support of μ generates $\mathfrak{S}$. Then the mass of the convolution sequence μ^{*n} concentrates on the kernel $K(\mathfrak{S})$ of $\mathfrak{S}$. More precisely, for every open set O containing K and every $\varepsilon > 0$, there exists a positive integer n_ε such that $\mu^{*n}(O) > 1 - \varepsilon$ for every $n \geqslant n_\varepsilon$.*

Here μ^{*n} denotes the convolution product $\mu * \cdots * \mu$ of n copies of μ. If μ^{*n} tends to μ as $n \to \infty$ then μ is convolution idempotent, that is $\mu * \mu = \mu$. The kernel $K(\mathfrak{S})$ of $\mathfrak{S}$ is the sub–semi–group of $\mathfrak{S}$ obtained by taking the intersection of the family of two sided ideals of $\mathfrak{S}$, see [Ros71, th. 1]. A direct consequence of lemma 2.1 is the absence of a translation invariant probability measure μ on $\mathfrak{S}$ with full support such that the kernel of $\mathfrak{S}$ is a μ–proper sub–semi–group of $\mathfrak{S}$. By μ–proper sub–semi–group here we mean that its μ-measure is < 1. This result can be easily understood intuitively since the translation associated to a non invertible element of $\mathfrak{S}$ gives a strict contraction of the support.

Proof of theorem 1.4. The kernel of the semi–group $\mathcal{M}_n$ is constituted by the $n \times n$ Markov matrices with equal rows, which are the $n \times n$ idempotent Markov matrices (i.e. $\mathbf{M}^2 = \mathbf{M}$). The reader may find more details in [Ros71, p. 146]. Since the kernel of $\mathcal{M}_n$ is a $\mathcal{U}(\mathcal{M}_n)$–proper sub–semi–group of $\mathcal{M}_n$, lemma 2.1 implies the absence of any convolution idempotent probability distribution on $\mathcal{M}_n$, absolutely continuous with respect to $\mathcal{U}(\mathcal{M}_n)$ and with full support. The proof is finished by noticing that if a probability distribution on $\mathcal{M}_n$ is invariant by every left (or right) translation, then it is convolution idempotent. Note by the way that the Wedderburn matrix $\frac{1}{n}\mathbf{1}$ belongs to the kernel of $\mathcal{M}_n$, and also that this kernel is equal to $\{\lim_{k\to\infty} \mathbf{M}^k; \mathbf{M} \in \mathcal{A}_n\}$ where $\mathcal{A}_n$ is the collection of irreducible aperiodic elements of $\mathcal{M}_n$. The reader may find in [Ros71, ch. 5] the structure of non fully supported idempotent probability distributions on compact semi–groups and in particular on $\mathcal{M}_n$. ☐

3 Proofs of theorem 1.5

The following theorem can be found for instance in [BS06, th. 3.6].

Theorem 3.1 (Singular values of large dimensional non–centered random arrays)**.** *Let $(X_{i,j})_{1 \leqslant i,j < \infty}$ be an infinite array of i.i.d. real random variables with mean m and variance $\sigma^2 \in (0, \infty)$. If $\mathbf{X} = (X_{i,j})_{1 \leqslant i,j \leqslant n}$, then*

$$\mathbb{P}\left(\frac{1}{n} \sum_{k=1}^{n} \delta_{s_k(n^{-1/2}\mathbf{X})} \xrightarrow[n\to\infty]{w} \mathcal{Q}_\sigma \right) = 1$$

where $\xrightarrow{w}$ *denotes the weak converge of probability distributions and* $\mathcal{Q}_\sigma$ *is the Wigner quarter–circle distribution defined in table 1. Moreover,*

$$\mathbb{P}\Big(\lim_{n\to\infty} s_1(n^{-1/2}\,\mathbf{X}) = 2\sigma^2\Big) = 1 \quad \textit{if and only if} \quad \mathbb{E}[\mathbf{X}_{1,1}] = 0 \textit{ and } \mathbb{E}[|\mathbf{X}_{1,1}|^4] < \infty.$$

The following lemma is a consequence of [BY93, le. 2] (see also [BS06, le. 5.13]).

Lemma 3.2 (Uniform law of large numbers)**.** *If* $(X_{i,j})_{1\leqslant i,j<\infty}$ *is an infinite array of i.i.d. random variables of mean* m*, then by denoting* $S_{i,n} = \sum_{j=1}^n X_{i,j}$*,*

$$\max_{1\leqslant i\leqslant n}\left|\frac{S_{i,n}}{n} - m\right| \xrightarrow[n\to\infty]{a.s.} 0$$

and in the case where $m \neq 0$*, we have also*

$$\max_{1\leqslant i\leqslant n}\left|\frac{n}{S_{i,n}} - \frac{1}{m}\right| \xrightarrow[n\to\infty]{a.s.} 0.$$

The following lemma is a consequence of the Courant–Fischer variational formulas for singular values, see [HJ90]. Also, we leave the proof to the reader.

Lemma 3.3 (Singular values of diagonal multiplicative perturbations)**.** *For every* $n\times n$ *matrix* $\mathbf{A}$*, every* $n\times n$ *diagonal matrix* $\mathbf{D}$*, and every* $1\leqslant k\leqslant n$*,*

$$s_n(\mathbf{D})s_k(\mathbf{A}) \leqslant s_k(\mathbf{DA}) \leqslant s_1(\mathbf{D})s_k(\mathbf{A}).$$

We are now able to prove theorem 1.5.

Proof of theorem 1.5. We have $\mathbf{M} = \mathbf{DE}$ where $\mathbf{E} = (X_{i,j})_{1\leqslant i,j\leqslant n}$ and $\mathbf{D}$ is the $n\times n$ diagonal matrix given for every $1\leqslant i\leqslant n$ by

$$\mathbf{D}_{i,i} = \frac{1}{\sum_{j=1}^n X_{i,j}}.$$

The fact that $\mathbf{M}\sim\mathcal{U}(\mathcal{M}_n)$ follows immediately from theorem 1.1 combined with the construction of the Dirichlet distribution $\mathcal{D}_n(1,\ldots,1)$ from i.i.d. exponential random variables. It remains to prove the convergence of the ESD of $\sqrt{n}\mathbf{M}\mathbf{M}^\top$ as $n\to\infty$ to the Wigner quarter–circle distribution $\mathcal{Q}_1$. For such, we use the method of Aubrun [Aub06], by replacing the unit $\|\cdot\|_1$–ball by the portion of the unit $\|\cdot\|_1$–sphere with non–negative coordinates. If suffices to show that almost surely, the discrete measure $\frac{1}{n}\sum_{k=2}^n \delta_{s_k(\sqrt{n}\mathbf{M})}$ tends weakly to the Wigner quarter–circle distribution $\mathcal{Q}_1$.

We first observe that $\mathbf{E}$ is a rank one additive perturbation of the centered random matrix $\mathbf{E}-\mathbb{E}\mathbf{E}$. Also, a standard interlacing inequality gives

$$s_2(\mathbf{E}) \leqslant s_1(\mathbf{E}-\mathbb{E}\mathbf{M}).$$

Now by the second part of theorem 3.1 we have $s_1(\mathbf{E}-\mathbb{E}\mathbf{E}) = O(\sqrt{n})$ almost surely. Consequently, $s_2(n^{-1/2}\,\mathbf{E}) = O(1)$ almost surely. In particular, almost surely, the sequence $(\frac{1}{n}\sum_{k=2}^n \delta_{s_k(n^{-1/2}\,\mathbf{E})})_{n\geqslant 1}$ remains in a compact set. The desired result follows then from the combination of the first part of theorem 3.1 with lemmas 3.3 and 3.2. This proof does not rely on the exponential nature of the $X_{i,j}$'s and remains actually valid for more general laws, see [Cha08]. □

There is no equivalent of lemma 3.3 for the eigenvalues instead of the singular values, and thus the method used to prove theorem 1.5 fails for conjecture 1.6. Note that by lemma 3.2 used with the exponential distribution of mean $m = 1$,

$$\|n\mathbf{D} - \mathbf{I}\|_{2\to 2} = \max_{1\leqslant i\leqslant n} \left| \frac{n}{\sum_{j=1}^n X_{i,j}} - 1 \right| \underset{n\to\infty}{\longrightarrow} 0 \quad \text{a.s.}$$

If $\mathbf{A}$ is diagonal, then we simply have $\|\mathbf{A}\|_{2\to 2} = s_1(\mathbf{A}) = \max_{1\leqslant k\leqslant n} |\mathbf{A}_{k,k}|$, and when $\mathbf{A}$ is diagonal and invertible, $\|\mathbf{A}^{-1}\|_{2\to 2}^{-1} = s_n(\mathbf{A}) = \min_{1\leqslant k\leqslant n} |\mathbf{A}_{k,k}|$. Now, by the circular law theorem for non–central random matrices [Cha07], we get that almost surely, the ESD of $n^{-1/2}\,\mathbf{E}$ converges, as $n \to \infty$, to the uniform distribution $\mathcal{C}_1$ (see table 1). It is then natural to decompose $\sqrt{n}\,\mathbf{M}$ as

$$\sqrt{n}\,\mathbf{M} = n\mathbf{D}n^{-1/2}\,\mathbf{E} = (n\mathbf{D} - \mathbf{I})n^{-1/2}\,\mathbf{E} + n^{-1/2}\,\mathbf{E}.$$

Unfortunately, since $m = 1 \neq 0$, we have almost surely (see [Cha07])

$$\left\|n^{-1/2}\,\mathbf{E}\right\|_{2\to 2} = s_1(n^{-1/2}\,\mathbf{E}) \underset{n\to\infty}{\longrightarrow} +\infty.$$

This suggests that $\sqrt{n}\,\mathbf{M}$ cannot be seen as a perturbation of $n^{-1/2}\,\mathbf{E}$ with a matrix of small norm. Actually, even if it was the case, the relation between the two spectra is unknown since $\mathbf{E}$ is not normal. One can think about using logarithmic potentials to circumvent the problem. The strength of the logarithmic potential approach is that it allows to study the asymptotic behavior of the ESD (i.e. eigenvalues) of non–normal matrices via the singular values of a family of matrices indexed by $z \in \mathbb{C}$. The details are given in [Cha07] for instance. The logarithmic potential of the ESD of $\sqrt{n}\,\mathbf{M}$ at point z is

$$\begin{aligned} U_n(z) &= -\frac{1}{n}\log\left|\det(\sqrt{n}\,\mathbf{M} - z\mathbf{I})\right| \\ &= -\frac{1}{n}\log|\det(n\mathbf{D})| - \frac{1}{n}\log\left|\det(n^{-1/2}\,\mathbf{E} - z(n\mathbf{D})^{-1})\right|. \end{aligned}$$

Now, by lemma 3.2,

$$\frac{1}{n}\log|\det(n\mathbf{D})| \underset{n\to\infty}{\longrightarrow} 0 \quad \text{a.s.}$$

By the circular law theorem for non–central random matrices [Cha07] and the lower envelope theorem [ST97], almost surely, for quasi-every[4] $z \in \mathbb{C}$, the quantity

$$\liminf_{n\to\infty} -\frac{1}{n}\log\left|\det(n^{-1/2}\,\mathbf{E} - z\mathbf{I})\right|$$

is equal to the logarithmic potential at point z of the uniform distribution $\mathcal{C}_1$ on the unit disc $\{z \in \mathbb{C}; |z| \leqslant 1\}$. It is thus enough to show that almost surely, for every $z \in \mathbb{C}$,

$$\frac{1}{n}\log\left|\det(n^{-1/2}\,\mathbf{E} - z(n\mathbf{D})^{-1})\right| - \frac{1}{n}\log\left|\det(n^{-1/2}\,\mathbf{E} - z\mathbf{I})\right| \underset{n\to\infty}{\longrightarrow} 0.$$

[4]This means "except on a subset of zero capacity", in the sense of potential theory, see [ST97].

Unfortunately, we ignore how to prove that. A possible alternative beyond potential theoretic tools is to adapt the method developed in [TVK08] by Tao and Vu involving a "replacement principle". The reader may find some progresses in [Cha08].

Acknowledgements. Part of this work was done during two visits to Laboratoire Jean Dieudonné in Nice, France. The author would like to thank Pierre Del Moral and Persi Diaconis for their kind hospitality there. Many thanks to Zhidong Bai, Franck Barthe, Włodzimierz Bryc, Mireille Capitaine, Delphine Féral, Michel Ledoux, and Gérard Letac for exchanging some ideas on the subject. This work benefited from many stimulating discussions with Neil O'Connell when he visited the Institut de Mathématiques de Toulouse.

References

[AGZ09] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An Introduction to Random Matrices*, Cambridge Studies in Advanced Mathematics, Cambridge University Press, 2009.

[And90] A. L. Andrew, *Eigenvalues and singular values of certain random matrices*, J. Comput. Appl. Math. **30** (1990), no. 2, 165–171.

[Aub06] G. Aubrun, *Random points in the unit ball of* ℓ_p^n, Positivity **10** (2006), no. 4, 755–759.

[Aub07] ———, *Sampling convex bodies: a random matrix approach*, Proc. Amer. Math. Soc. **135** (2007), no. 5, 1293–1303 (electronic).

[Bai99] 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.

[BCC08] Ch. Bordenave, P. Caputo, and D. Chafaï, *Spectrum of large random reversible Markov chains: two examples*, preprint `arXiv:0811.1097 [math.PR]`, 2008.

[BCC09] ———, *Spectrum of large random reversible Markov chains - heavy-tailed weights on the complete graph*, preprint `arXiv:0903.3528 [math.PR]`, 2009.

[BDJ06] 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.

[BDPX05] S. Boyd, P. Diaconis, P. Parrilo, and L. Xiao, *Symmetry analysis of reversible Markov chains*, Internet Math. **2** (2005), no. 1, 31–71.

[BEK$^+$05] I. Bengtsson, Å. Ericsson, M. Kuś, W. Tadej, and K. Życzkowski, *Birkhoff's polytope and unistochastic matrices,* $N = 3$ *and* $N = 4$, Comm. Math. Phys. **259** (2005), no. 2, 307–324.

[Ber01] G. Berkolaiko, *Spectral gap of doubly stochastic matrices generated from equidistributed unitary matrices*, J. Phys. A **34** (2001), no. 22, L319–L326.

[BGMN05] F. Barthe, O. Guédon, Sh. Mendelson, and A. Naor, *A probabilistic approach to the geometry of the* ℓ_p^n *ball*, The Annals of Probability **33** (2005), no. 2, 480–513.

[Bia97] Ph. Biane, *On the free convolution with a semi-circular distribution*, Indiana Univ. Math. J. **46** (1997), no. 3, 705–718.

[BS06] Z. D. Bai and J. W. Silverstein, *Spectral Analysis of Large Dimensional Random Matrices*, Mathematics Monograph Series 2, Science Press, Beijing, 2006.

[BS07] A. Bose and A. Sen, *Spectral norm of random large dimensional noncentral Toeplitz and Hankel matrices*, Electron. Comm. Probab. **12** (2007), 29–35 (electronic).

[BS08] ———, *Another look at the moment method for large dimensional random matrices*, Electron. J. Probab. **13** (2008), no. 21, 588–628.

[BY93] Z. D. Bai and Y. Q. Yin, *Limit of the smallest eigenvalue of a large-dimensional sample covariance matrix*, Ann. Probab. **21** (1993), no. 3, 1275–1294.

[Cha07] D. Chafaï, *Circular law for non-central random matrices*, preprint `arXiv:0709.0036 [math.PR]`, 2007.

[Cha08] ______, *Circular Law Theorem for Random Markov Matrices*, unpublished notes `arXiv:0808.1502 [math.PR]`, 2008.

[Cha09] ______, *Aspects of large random Markov kernels*, Stochastics (2009), no. 81, 415–429.

[CPSV09] Fr. Comets, S. Popov, G. M. Schütz, and M. Vachkovskaia, *Billiards in a general domain with random reflections*, Arch. Ration. Mech. Anal. **191** (2009), no. 3, 497–537.

[CSC08] G.-Y. Chen and L. Saloff-Coste, *The cutoff phenomenon for ergodic Markov processes*, Electron. J. Probab. **13** (2008), no. 3, 26–78.

[DE85] P. Diaconis and B. Efron, *Testing for independence in a two-way table: new interpretations of the chi-square statistic*, Ann. Statist. **13** (1985), no. 3, 845–913, With discussions and with a reply by the authors.

[DE87] ______, *Probabilistic-geometric theorems arising from the analysis of contingency tables*, Contributions to the theory and application of statistics, Academic Press, Boston, MA, 1987, pp. 103–125.

[DG95] P. Diaconis and A. Gangolli, *Rectangular arrays with fixed margins*, Discrete probability and algorithms (Minneapolis, MN, 1993), IMA Vol. Math. Appl., vol. 72, Springer, New York, 1995, pp. 15–41.

[DR06] P. Diaconis and S. W. W. Rolles, *Bayesian analysis for reversible Markov chains*, Ann. Statist. **34** (2006), no. 3, 1270–1292.

[Ede97] A. Edelman, *The probability that a random real Gaussian matrix has k real eigenvalues, related distributions, and the circular law*, J. Multivariate Anal. **60** (1997), no. 2, 203–232.

[EKS94] A. Edelman, E. Kostlan, and M. Shub, *How many eigenvalues of a random matrix are real?*, J. Amer. Math. Soc. **7** (1994), no. 1, 247–267.

[ER05] A. Edelman and N. R. Rao, *Random matrix theory*, Acta Numer. **14** (2005), 233–297.

[Fal03] K. Falconer, *Fractal geometry*, second ed., John Wiley & Sons Inc., Hoboken, NJ, 2003, Mathematical foundations and applications.

[For09] P. Forrester, *Log-gases and Random matrices*, Book draft available on the author web page, 2009.

[GN03] G. Goldberg and M. Neumann, *Distribution of subdominant eigenvalues of matrices with random rows*, SIAM J. Matrix Anal. Appl. **24** (2003), no. 3, 747–761 (electronic).

[GONS00] G. Goldberg, P. Okunev, M. Neumann, and H. Schneider, *Distribution of subdominant eigenvalues of random matrices*, Methodol. Comput. Appl. Probab. **2** (2000), no. 2, 137–151.

[HJ90] R. A. Horn and Ch. R. Johnson, *Matrix analysis*, Cambridge University Press, Cambridge, 1990, Corrected reprint of the 1985 original.

[HKOS00] B. M. Hambly, P. Keevash, N. O'Connell, and D. Stark, *The characteristic polynomial of a random permutation matrix*, Stochastic Process. Appl. **90** (2000), no. 2, 335–346.

[Hor54] A. Horn, *On the eigenvalues of a matrix with prescribed singular values*, Proc. Amer. Math. Soc. **5** (1954), 4–7.

[HP00] 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.

[Jia09] T. Jiang, *Approximation of Haar distributed matrices and limiting distributions of eigenvalues of Jacobi ensembles*, Probab. Theory Related Fields **144** (2009), no. 1-2, 221–246.

[Kos92] E. Kostlan, *On the spectra of Gaussian matrices*, Linear Algebra Appl. **162/164** (1992), 385–388, Directions in matrix theory (Auburn, AL, 1990).

[Meh04] M. L. Mehta, *Random matrices*, third ed., Pure and Applied Mathematics (Amsterdam), vol. 142, Elsevier/Academic Press, Amsterdam, 2004.

[MP06] Sh. Mendelson and A. Pajor, *On singular values of matrices with independent rows*, Bernoulli **12** (2006), no. 5, 761–773.

[NR03] A. Naor and D. Romik, *Projecting the surface measure of the sphere of ℓ_p^n*, Ann. Inst. H. Poincaré Probab. Statist. **39** (2003), no. 2, 241–261.

[PP07] A. Pajor and L. Pastur, *On the Limiting Empirical Measure of the sum of rank of one matrices with log-concave distribution*, preprint `arXiv:0710.1346 [math.PR]`, september 2007.

[Rac91] 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.

[Rid03] B. Rider, *A limit theorem at the edge of a non-Hermitian random matrix ensemble*, J. Phys. A **36** (2003), no. 12, 3401–3409, Random matrix theory.

[Ros71] M. Rosenblatt, *Markov processes. Structure and asymptotic behavior*, Springer-Verlag, New York, 1971, Die Grundlehren der mathematischen Wissenschaften, Band 184.

[SC97] 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.

[Sen06] E. Seneta, *Non-negative matrices and Markov chains*, Springer Series in Statistics, Springer, New York, 2006, Revised reprint of the second (1981) edition.

[Set94] J. Sethuraman, *A constructive definition of Dirichlet priors*, Statist. Sinica **4** (1994), no. 2, 639–650.

[Sil94] J. W. Silverstein, *The spectral radii and norms of large-dimensional non-central random matrices*, Comm. Statist. Stochastic Models **10** (1994), no. 3, 525–532.

[ST97] E. B. Saff and V. Totik, *Logarithmic potentials with external fields*, Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], vol. 316, Springer-Verlag, Berlin, 1997, Appendix B by Thomas Bloom.

[SZ90] G. Schechtman and J. Zinn, *On the volume of the intersection of two ℓ_n^p balls*, Proc. Amer. Math. Soc. **110** (1990), 217–224.

[Tan01] G. Tanner, *Unitary-stochastic matrix ensembles and spectral statistics*, J. Phys. A **34** (2001), no. 41, 8485–8500.

[TE05] L. N. Trefethen and M. Embree, *Spectra and pseudospectra*, Princeton University Press, Princeton, NJ, 2005, The behavior of nonnormal matrices and operators.

[TVK08] T. Tao, V. Vu, and M Krishnapur, *Random matrices: Universality of ESDs and the circular law*, preprint `arXiv:0807.4898v5 [math.PR]`, 2008.

[Vil03] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics, vol. 58, American Mathematical Society, Providence, RI, 2003.

[Wey49] H. Weyl, *Inequalities between the two kinds of eigenvalues of a linear transformation*, Proc. Nat. Acad. Sci. U. S. A. **35** (1949), 408–411.

[Wie00] K. Wieand, *Eigenvalue distributions of random permutation matrices*, Ann. Probab. **28** (2000), no. 4, 1563–1587.

[Wil62] S. S. Wilks, *Mathematical statistics*, A Wiley Publication in Mathematical Statistics, John Wiley & Sons Inc., New York, 1962.

Djalil Chafaï
Laboratoire d'Analyse et de Mathématiques Appliquées (UMR CNRS 8050)
Université Paris-Est Marne-la-Vallée
5 boulevard Descartes, Champs-sur-Marne, F-77454, Cedex 2, France.
E-mail: djalil(at)chafai.net **Web:** http://djalil.chafai.net/

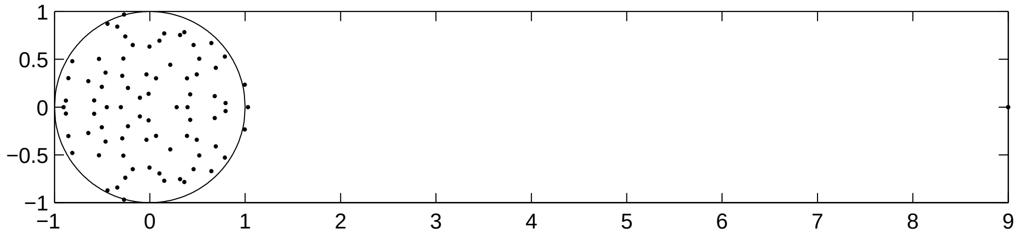

Figure 1: Plot of the spectrum of a single realization of $\sqrt{n}\,\mathbf{M}$ where $\mathbf{M} \sim \mathcal{U}(\mathcal{M}_n)$ with $n = 81$. We see one isolated eigenvalue $\lambda_1(\sqrt{n}\,\mathbf{M}) = \sqrt{n} = 9$ while the rest of the spectrum remains near the unit disc and seems uniformly distributed, in accordance with conjecture 1.6.

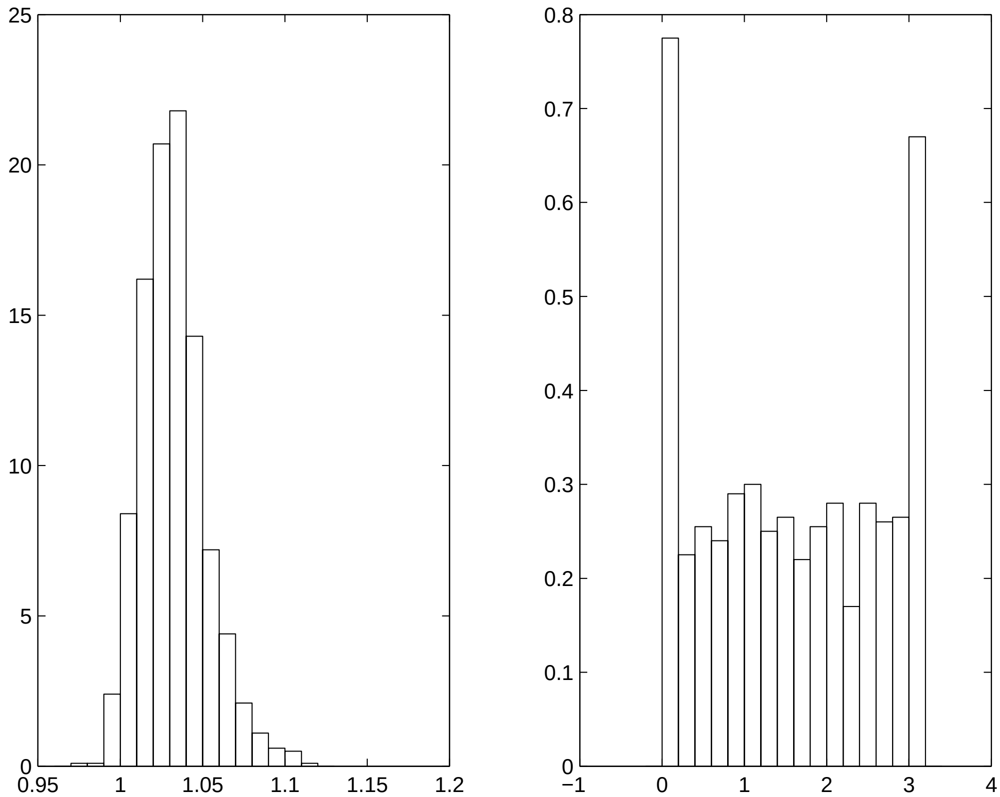

Figure 2: Here 1000 i.i.d. realizations of $\sqrt{n}\,\mathbf{M}$ where simulated where $\mathbf{M} \sim \mathcal{U}(\mathcal{M}_n)$ with $n = 300$. The first plot is the histogram of $|\lambda_2(\sqrt{n}\,\mathbf{M})|$, i.e. the module of the sub–dominant eigenvalue $\lambda_2(\sqrt{n}\,\mathbf{M})$. The second plot is the histogram of $|\mathrm{Phase}(\lambda_2(\sqrt{n}\,\mathbf{M}))|$. Recall that the spectrum is symmetric with respect to the real axis since the matrices are real.

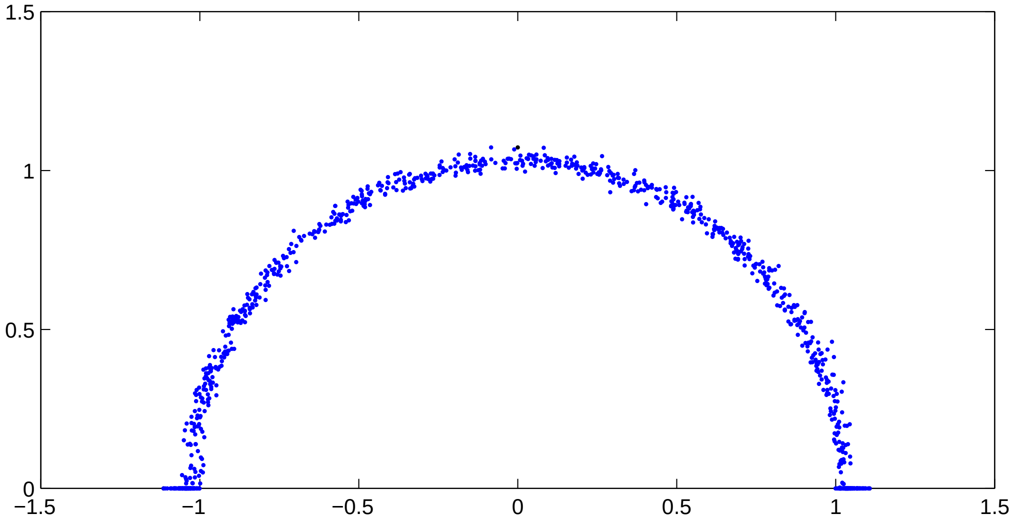

Figure 3: Here we reused the sample used for figure 2. The graphic is a plot of the 1000 i.i.d. realizations of the sub–dominant eigenvalue $\lambda_2(\sqrt{n}\,\mathbf{M})$. Since we deal with real matrices, the spectrum is symmetric with respect to the real axis, and we plotted $(\mathrm{RealPart}(\lambda_2), |\mathrm{ImaginaryPart}(\lambda_2)|)$ in the complex plane.