



# ANNALES

DE

# L'INSTITUT FOURIER

Joseph NAINUDEL & Ashkan NIKEGHBALI

**The distribution of eigenvalues of randomized permutation matrices**

Tome 63, n° 3 (2013), p. 773-838.

[http://aif.cedram.org/item?id=AIF\\_2013\\_\\_63\\_3\\_773\\_0](http://aif.cedram.org/item?id=AIF_2013__63_3_773_0)

© Association des Annales de l'institut Fourier, 2013, tous droits réservés.

L'accès aux articles de la revue « Annales de l'institut Fourier » (<http://aif.cedram.org/>), implique l'accord avec les conditions générales d'utilisation (<http://aif.cedram.org/legal/>). Toute reproduction en tout ou partie de cet article sous quelque forme que ce soit pour tout usage autre que l'utilisation à fin strictement personnelle du copiste est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

cedram

*Article mis en ligne dans le cadre du  
Centre de diffusion des revues académiques de mathématiques  
<http://www.cedram.org/>*

## THE DISTRIBUTION OF EIGENVALUES OF RANDOMIZED PERMUTATION MATRICES

by Joseph NAJNUDEL & Ashkan NIKEGHBALI

---

ABSTRACT. — In this article we study in detail a family of random matrix ensembles which are obtained from random permutations matrices (chosen at random according to the Ewens measure of parameter  $\theta > 0$ ) by replacing the entries equal to one by more general non-vanishing complex random variables. For these ensembles, in contrast with more classical models as the Gaussian Unitary Ensemble, or the Circular Unitary Ensemble, the eigenvalues can be very explicitly computed by using the cycle structure of the permutations. Moreover, by using the so-called virtual permutations, first introduced by Kerov, Olshanski and Vershik, and studied with a probabilistic point of view by Tsilevich, we are able to define, on the same probability space, a model for each dimension greater than or equal to one, which gives a meaning to the notion of almost sure convergence when the dimension tends to infinity. In the present paper, depending on the precise model which is considered, we obtain a number of different results of convergence for the point measure of the eigenvalues, some of these results giving a strong convergence, which is not common in random matrix theory.

RÉSUMÉ. — Dans cet article, nous étudions en détail une famille d'ensembles de matrices aléatoires qui sont obtenues à partir de matrices de permutation aléatoires en remplaçant les coefficients égaux à un par des variables aléatoires complexes non nulles plus générales. Pour ces ensembles, les valeurs propres peuvent être calculées très explicitement en utilisant la structure en cycles des permutations. De plus, en utilisant les permutations virtuelles, étudiées par Kerov, Olshanski, Vershik et Tsilevich, nous sommes capables de définir, sur le même espace de probabilité, un modèle pour chaque dimension supérieure ou égale à un, ce qui donne un sens à la notion de convergence presque sûre quand la dimension tend vers l'infini. Dans le présent article, selon le modèle précis qui est étudié, nous obtenons différents résultats de convergence pour la mesure ponctuelle des valeurs propres, certains de ces résultats donnant une convergence forte.

---

*Keywords:* Random matrix, permutation matrix, virtual permutation, convergence of eigenvalues.

*Math. classification:* 15A18, 15A52, 20B30.

## 1. Introduction

### 1.1. Random permutation matrices and outline of the paper

The distribution of the eigenvalues of random matrices and some related objects such as their characteristic polynomials have received much attention in the last few decades. They have been applied in such diverse branches as physics, number theory, analysis or probability theory as illustrated in the monographs [11], [10] and [1] or the survey paper [5]. The main matrix ensembles which have been studied are the Gaussian ensembles and some of their generalizations, the classical compact Lie groups  $U(N)$  (the group of unitary matrices),  $O(N)$  (the orthogonal group) and some of their continuous sub-groups, endowed with the Haar probability measure. It is also natural to investigate for the distribution of the eigenvalues of random permutation matrices, *i.e.* matrices which are canonically associated to a random element of a given finite symmetric group. Indeed it is well-known that the eigenvalues of a permutation matrix  $M_\sigma$  associated with a permutation  $\sigma$  are entirely determined by the cycle structure of  $\sigma$ , and hence one can hope to take advantage of the extensive literature on random permutations (see *e.g.* the book by Arratia, Barbour and Tavaré [2]) to describe completely the structure of the point process of the eigenvalues of random permutation matrices (*e.g.* the correlation measure of order  $q$ , the convergence of the normalized and non-normalized empirical spectral distribution, *etc.*). This has been shortly sketched out in the pioneering work by Diaconis and Shahshahani [6] and further developed by Wieand in [16], who studied the problem of counting how many eigenvalues lie in some fixed arc of the unit circle. Wieand compares the results obtained in this case with those obtained for the unitary group under the Haar measure and notices some similarities but also some differences when it comes to look at more refined aspects. Then it is suggested that one should try to compute finer statistics related to the eigenvalues in order to see how random permutation matrices fit into the random matrix picture. Of course one expects some drastic differences: for instance the point process associated with the eigenvalues of random permutation matrices should not be determinantal whereas it is determinantal for the unitary group. The goal of the present work is to continue the work initiated by Diaconis and Shahshahani in [6] and Wieand in [16]. Before mentioning more precisely the various directions in which we wish to extend the existing work, it should be mentioned that other works have been recently done on random permutation matrices, such as the paper by Hambly, Keevach, O'Connell and Stark [8] on

the characteristic polynomial of random unitary matrices, the papers by Zeindler ([18]) and by Dehay and Zeindler ([4]), or the works by Wieand ([17]) and Evans ([7]) on the eigenvalues of random wreath products (but in this latter case, the techniques that are involved are different and we shall also address this framework in a future work).

The ensembles of matrices we study in this paper are constructed from permutation matrices by replacing the entries equal to 1 by more general random variables. The advantage of this kind of ensembles is that it somehow combines the advantages of the ensembles of permutation matrices, and of the more classical ensembles of unitary matrices (as the Circular Unitary Ensemble): the very explicit computations made on permutation matrices can be naturally extended to our setting, and on the other hand, we have more flexibility on the possible distributions of the eigenvalues, and in particular their properties of absolute continuity: for example, the eigenvalues are not necessarily roots of unity, and the average distribution of the eigenvalues (one-point correlation function) can be absolutely continuous with respect to the Lebesgue measure.

We now describe in more detail how we shall continue and extend some of the previous works and shall postpone precise definitions to the next paragraph:

- We shall consider a larger ensemble of random matrices with more general distributions than for ensembles of permutation matrices; roughly speaking, we first pick a permutation of size  $N$  from  $\Sigma_N$  (the group of permutations of size  $N$ ) at random according to the Ewens measure of parameter  $\theta \geq 0$  (under this measure, the probability of a given permutation  $\sigma$  is proportional to  $\theta^n$ , where  $n$  is the number of cycles of  $\sigma$ ) and then consider the corresponding matrix; we then replace the 1's by a sequence of i.i.d. random variables  $z_1, \dots, z_N$  taking values in  $\mathbb{C}^*$ . This ensemble is a sub-group of the linear sub-group  $GL(N, \mathbb{C})$  and the classical random permutation matrices correspond to the case where  $\theta = 1$  and where the distribution of the  $z_i$ 's is the Dirac measure at 1. The choice of the Ewens measure is natural since it is a one parameter deformation of the uniform distribution which is coherent with the projections from  $\Sigma_{N'}$  onto  $\Sigma_N$  for  $N' \geq N$  (see the next paragraphs). If the  $z_i$ 's take their values on the unit circle, then our ensemble is a sub-group of  $U(N)$ .
- We shall also give a meaning to almost sure convergence for the probability empirical spectral distribution. Indeed, it is not a priori

obvious to define almost sure convergence in random matrix theory since the probability space changes with  $N$  and to the best of our knowledge there has not been, so far, a satisfactory way to overcome this difficulty. We propose to deal with this problem by considering the space of *virtual permutations*, which were introduced by Kerov, Olshanski and Vershik in [9]. A virtual permutation is a sequence of permutations  $(\sigma_N)_{N \geq 1}$  constructed in a coherent way, in the sense that for all  $N' \geq N \geq 1$ , the cycle structure of  $\sigma_N$  can be obtained from the cycle structure of  $\sigma_{N'}$  simply by removing the elements strictly larger than  $N$ . The virtual permutations satisfy the following property: if  $(\sigma_N)_{N \geq 1}$  is a virtual permutation, if  $N' \geq N \geq 1$  and if  $\sigma_{N'}$  is distributed according to the Ewens measure of parameter  $\theta$  on  $\Sigma_{N'}$ , then  $\sigma_N$  follows the Ewens measure of parameter  $\theta$  on  $\Sigma_N$ . We shall then take advantage of the work by Tsilevich [14], who proved almost sure convergence for the relative cycle lengths of virtual permutations. We shall also extensively use an algorithmic way to generate the Ewens measure on the space of virtual permutations.

- We shall study in detail the point process of the eigenvalues of a matrix  $M$  drawn from our matrix ensembles. For instance we establish various convergence results for the empirical spectral measure  $\mu(M) = \sum \delta_\lambda$ , or for  $\mu(M)/N$  ( $N$  being the dimension of the matrix) where the sum is over all eigenvalues  $\lambda$  of  $M$ , counted with multiplicity. We also consider the average measure obtained by taking the expectation of  $\mu(M)$ , as well as the correlation measure of order  $q$ . In the special case where the random variables  $z_i$ 's take their values on the unit circle, the point process associated with the eigenvalues has some more remarkable properties. For instance, when the  $z_i$ 's are uniformly distributed on the unit circle, the empirical spectral measure as well as its limit (in a sense to be made precise) are invariant by translation. Still in this case, the 1-correlation for fixed  $N$  and  $N = \infty$  is the Lebesgue measure, whereas for  $q \geq 2$ , and fixed  $N$ , the  $q$ -correlation is not absolutely continuous with respect to the Lebesgue measure anymore. For  $N = \infty$ , the pair correlation measure is still absolutely continuous with respect to the Lebesgue measure with an explicit density, but this result fails to hold for the correlations of order greater than or equal to 3. One can push further the analogy with the studies made for the classical compact continuous groups by characterizing the distribution function of the

smallest eigenangle (again when the  $z_i$ 's are uniformly distributed on the unit circle) as the solution of some integral equation.

## 1.2. Definitions and notation

In this section we describe the structure of the eigenvalues of the “generalized permutation matrices” mentioned above, independently of any probability measure, and then we define the family of probability measures which will be studied in detail in the sequel of this article. More precisely let  $\Sigma_N$  be the group of permutations of order  $N$ . By straightforward computations, one can check that the set of matrices  $M$  such that there exists a permutation  $\sigma \in \Sigma_N$  and complex numbers  $z_1, z_2, \dots, z_N$ , different from zero, satisfying  $M_{jk} = z_j \mathbf{1}_{j=\sigma(k)}$ , is a multiplicative group, denoted by  $\mathcal{G}(N)$  in this article, and which can be written as the wreath product of  $\mathbb{C}^*$  and  $\Sigma_N$ . The group  $\mathcal{G}_N$  can also be viewed as the group generated by the permutation matrices and the diagonal matrices of  $GL(N, \mathbb{C})$ . The elements of  $\mathcal{G}(N)$  such that  $|z_j| = 1$  for all  $j$ ,  $1 \leq j \leq N$ , form a subgroup denoted  $\mathcal{H}(N)$  of  $\mathcal{G}(N)$ , which can be viewed as the wreath product of  $\mathbb{U}$  (the set of complex numbers with modulus one) and  $\Sigma_N$ . One can also define, for all integers  $k \geq 1$ , the group  $\mathcal{H}_k(N)$  of elements of  $\mathcal{G}(N)$  such that  $z_j^k = 1$  for  $1 \leq j \leq N$ : the subgroup  $\mathcal{H}_k(N)$  of  $\mathcal{H}(N)$  is the wreath product of  $\mathbb{U}_k$  (group of  $k$ -th roots of unity) and  $\Sigma_N$ . Note that the group structure of  $\mathcal{G}$ ,  $\mathcal{H}$  or  $\mathcal{H}_N$  does not play a fundamental role in our work, however, this structure gives a parallel between this paper and the study of other groups of matrices, such as the orthogonal group, the unitary group or the symplectic group (we note that  $\mathcal{H}$  and  $\mathcal{H}_N$  are subgroups of the unitary group of dimension  $N$ ). The advantage of the study of matrices in  $\mathcal{G}(N)$  is the fact that the structure of their eigenvalues can be very explicitly described in function of the cycle structure of the corresponding permutations. More precisely, let  $M$  be a matrix satisfying  $M_{jk} = z_j \mathbf{1}_{j=\sigma(k)}$  for all  $1 \leq j, k \leq N$ , where  $\sigma$  is a permutation of order  $N$ , and  $z_1, \dots, z_N \in \mathbb{C}^*$ . If the supports of the cycles of  $\sigma$  are  $C_1, \dots, C_n$ , with corresponding cardinalities  $l_1, \dots, l_n$ , and if for  $1 \leq m \leq n$ ,  $R_m$  is the set of the roots of order  $l_m$  of the complex number

$$Z_m := \prod_{j \in C_m} z_j,$$

then the set of eigenvalues of  $M$  is the union of the sets  $R_m$ , and the multiplicity of any eigenvalue is equal to the number of sets  $R_m$  containing it. An example of explicit calculation is that of the trace: since for all

integers  $l \geq 2$ , the sum of the  $l$ -th roots of unity is equal to zero, one immediately deduces that

$$\mathrm{Tr}(M) = \sum_{j \in F} z_j,$$

where  $F$  is the set of fixed points of  $\sigma$ . More generally, one can compute the trace of all the powers of  $M$ . Indeed, for all integers  $k \geq 1$ , the eigenvalues of  $M^k$  can be computed by taking the  $k$ -th powers of the elements of the sets  $R_m$ . Therefore,

$$\mathrm{Tr}(M^k) = \sum_{m=1}^n \sum_{\omega^{l_m} = Z_m} \omega^k.$$

Now, if  $l_m$  is not a divisor of  $k$ , the last sum is equal to zero, and if  $l_m$  is a divisor of  $k$ , all the terms of the last sum are equal to  $Z_m^{k/l_m}$ . We deduce that

$$\mathrm{Tr}(M^k) = \sum_{l_m | k} l_m Z_m^{k/l_m}.$$

We see that the description of the eigenvalues and the computations above do not depend on any probability measure given on the space  $\mathcal{G}(N)$ . Let us now define a particular class of probability measures on  $\mathcal{G}(N)$  which will be studied in detail in this paper.

**DEFINITION 1.1.** — *Let  $\theta > 0$  and let  $\mathcal{L}$  be a probability distribution on  $\mathbb{C}^*$ . The probability measure  $\mathbb{P}(N, \theta, \mathcal{L})$  on  $\mathcal{G}(N)$  is the law of the matrix  $M(\sigma, z_1, \dots, z_N)$ , where:*

- *the permutation  $\sigma$  follows the Ewens measure of parameter  $\theta$  on  $\Sigma_N$ , i.e. the probability that  $\sigma$  is equal to a given permutation is proportional to  $\theta^n$ , where  $n$  is the number of cycles of  $\sigma$ .*
- *for all  $j$ ,  $1 \leq j \leq N$ ,  $z_j$  is a random variable following the probability law  $\mathcal{L}$ .*
- *the random permutation  $\sigma$  and the random variables  $z_1, \dots, z_N$  are all independent.*
- *$M(\sigma, z_1, \dots, z_N)$  is the matrix  $M \in \mathcal{G}(N)$  such that for all  $1 \leq j, k \leq N$ ,  $M_{jk} = z_j \mathbb{1}_{j=\sigma(k)}$ .*

In this paper we prove for a large class of probability distributions  $\mathcal{L}$ , the weak convergence of the law of the empirical measure of the eigenvalues, when the dimension  $N$  tends to infinity. In several particular cases we are also interested in almost sure convergences, and then we need to couple all the dimensions  $N$  on the same probability space. This can be done by introducing the so-called virtual permutations, which were first defined by

Kerov, Olshanski and Vershik in [9] and also studied by Tsilevich [14]. A virtual permutation is a sequence  $(\sigma_N)_{N \geq 1}$  of permutations, such that for all  $N \geq 1$ ,  $\sigma_N \in \Sigma_N$ , and the cycle structure of  $\sigma_N$  is obtained from the cycle structure of  $\sigma_{N+1}$ , simply by removing the element  $N + 1$  (for example if  $\sigma_8 = (13745)(28)(6)$ , then  $\sigma_7 = (13745)(2)(6)$ ). Now, for  $\theta > 0$ , it is possible to define on the space of virtual permutations the so-called Ewens measure of parameter  $\theta$  as the unique (by the monotone class theorem) probability measure, such that if  $(\sigma_N)_{N \geq 1}$  follows this measure,  $\sigma_N$  follows the Ewens( $\theta$ ) measure on  $\Sigma_N$ . Now we can introduce the following definition:

DEFINITION 1.2. — *Let  $\theta > 0$  and let  $\mathcal{L}$  be a probability law on  $\mathbb{C}^*$ . The probability measure  $\mathbb{P}(\infty, \theta, \mathcal{L})$ , defined on the product of the probability spaces  $\mathcal{G}(N)$ ,  $N \geq 1$ , is the law of a sequence of random matrices  $(M_N)_{N \geq 1}$ , such that  $M_N = M(\sigma_N, z_1, \dots, z_N)$ , where:*

- *the sequence  $(\sigma_N)_{N \geq 1}$  is a random virtual permutation following Ewens measure of parameter  $\theta$ .*
- *for all  $j \geq 1$ ,  $z_j$  is a random variable following the distribution  $\mathcal{L}$ .*
- *the virtual permutation  $(\sigma_N)_{N \geq 1}$  and the random variables  $(z_j)_{j \geq 1}$  are independent.*

It is easy to check that for all  $N \geq 1$ , the image of  $\mathbb{P}(\infty, \theta, \mathcal{L})$  by the  $N$ -th coordinate projection is the measure  $\mathbb{P}(N, \theta, \mathcal{L})$ . The properties on  $\mathbb{P}(N, \theta, \mathcal{L})$  or  $\mathbb{P}(\infty, \theta, \mathcal{L})$  which are obtained in this article depend on the probability distribution  $\mathcal{L}$  in an essential way.

In the next section, we shall review some properties of virtual permutations which are needed in the sequel of the paper, and then we give general results on the point process associated with the eigenvalues of random matrices from  $\mathcal{G}(N)$ . We finally refine some of these results in the case of  $\mathcal{H}(N)$ .

## 2. Generating the Ewens measure on the set of virtual permutations

The space of virtual permutations was first introduced by Kerov, Olshanski and Vershik in [9] in the context of representation theory; the interested reader can refer to the notes by Olshanski in [12] for more details and references. Here we shall mostly be concerned with the probabilistic aspects of virtual permutations which were studied in detail by Tsilevich in [14]. We now review a construction of the virtual permutations which is explained



in [14] and which is suitable for the probabilistic reasoning. We then show how to generate the Ewens measure of parameter  $\theta$  on the space of virtual permutations. This is already explained by other means by Tsilevich in her unpublished note [15]; here we provide a more elementary way to generate it and give all details since this is going to be at the source of many of our proofs and because also virtual permutations might not be so well-known. In the sequel, we shall assume that the reader is familiar with the GEM and Poisson-Dirichlet distributions (if not, one can refer to [3], pp. 40–48).

As already mentioned, one of the interests of the construction of virtual permutations is that it gives a natural explanation of convergences in law involved when one looks at the relative lengths of cycles of random permutations in the symmetric group  $\Sigma_N$  when  $N$  goes to infinity. Recall that for all integer  $N \geq 1$ , there exists a bijective map

$$\Phi_N : \prod_{j=1}^N \{1, 2, \dots, j\} \longrightarrow \Sigma_N$$

such that

$$\Phi_N ((m_j)_{1 \leq j \leq N}) = \tau_{N, m_N} \circ \tau_{N-1, m_{N-1}} \circ \dots \circ \tau_{2, m_2} \circ \tau_{1, m_1},$$

where  $\tau_{j,k}$  is the unique permutation such that  $\tau_{j,k}(j) = k$ ,  $\tau_{j,k}(k) = j$ , and  $\tau_{j,k}(l) = l$  for  $l$  different from  $j$  and  $k$  (if  $j \neq k$ ,  $\tau_{j,k}$  is a transposition, if  $j = k$ , it is the identity). If  $N_1 \geq N_2$ , the bijections  $\Phi_{N_1}$  and  $\Phi_{N_2}$  induce a natural surjective map  $\pi_{N_1, N_2}$  from  $\Sigma_{N_1}$  to  $\Sigma_{N_2}$ , defined in the following way: if  $\sigma \in \Sigma_{N_1}$ , there exists a unique sequence  $(m_j)_{1 \leq j \leq N_1}$ ,  $m_j \in \{1, \dots, j\}$  such that

$$\sigma = \Phi_{N_1} ((m_j)_{1 \leq j \leq N_1}),$$

and one then defines:

$$\pi_{N_1, N_2}(\sigma) := \Phi_{N_2} ((m_j)_{1 \leq j \leq N_2}).$$

Note that for  $N_1 \geq N_2 \geq N_3$ ,

$$\pi_{N_2, N_3} \circ \pi_{N_1, N_2} = \pi_{N_1, N_3}.$$

Now, a virtual permutation is a sequence of permutations  $(\sigma_N)_{N \geq 1}$ , such that  $\sigma_N \in \Sigma_N$  for all  $N$ , and which is consistent with respect to  $\pi$ , i.e. for all  $N_1 \geq N_2$ ,

$$\pi_{N_1, N_2}(\sigma_{N_1}) = \sigma_{N_2}.$$

We denote by  $\Sigma_\infty$  the (infinite) set of virtual permutations, and note that the group structure of  $\Sigma_N$  does not induce any group structure on  $\Sigma_\infty$ .

Moreover, there is a natural bijection,

$$\Phi_\infty : \prod_{j=1}^\infty \{1, 2, \dots, j\} \longrightarrow \Sigma_\infty$$

induced by the bijections  $\Phi_N$ . Indeed, for all infinite sequences  $(m_j)_{j \geq 1}$ ,  $m_j \in \{1, \dots, j\}$  one defines:

$$\Phi_\infty((m_j)_{j \geq 1}) := (\sigma_N)_{N \geq 1}$$

where for all  $N$ ,

$$\sigma_N = \Phi_N((m_j)_{1 \leq j \leq N}).$$

Here, one immediately checks that  $(\sigma_N)_{N \geq 1}$  is consistent. Moreover, one can also define a surjective map  $\pi_{\infty, N}$  from  $\Sigma_\infty$  to  $\Sigma_N$  for all  $N \geq 1$ , by setting:

$$\pi_{\infty, N}((\sigma_n)_{n \geq 1}) := \sigma_N,$$

and one can check the relation

$$\pi_{N_1, N_2} \circ \pi_{\infty, N_1} = \pi_{\infty, N_2}.$$

The cycle structure of a virtual permutation can be described by the so-called Chinese restaurant process, described for example by Pitman ([13]). More precisely, let

$$\sigma_\infty = (\sigma_N)_{N \geq 1} \in \Sigma_\infty$$

be a virtual permutation. There exists a unique sequence  $(m_j)_{j \geq 1}$ ,  $m_j \in \{1, \dots, j\}$  such that

$$\sigma_\infty = \Phi_\infty((m_j)_{j \geq 1}).$$

Then the cycle structure for  $\sigma_N$  can be described by induction on  $N$ :

- $\sigma_1$  is (of course!) the identity of  $\Sigma_1$ ;
- if  $m_{N+1} = N + 1$  for  $N \geq 1$ , the cycle structure of  $\sigma_{N+1}$  is obtained from the structure of  $\sigma_N$  by simply adding the fixed point  $N + 1$ ;
- if  $m_{N+1} \leq N$  for  $N \geq 1$ , the cycle structure of  $\sigma_{N+1}$  is obtained from the structure of  $\sigma_N$  by inserting  $N + 1$  just before  $m_{N+1}$ , in the cycle containing  $m_{N+1}$ .

For example, if  $\sigma_4 = (124)(3)$  and  $m_5 = 5$ , then  $\sigma_5 = (124)(3)(5)$ , and if  $\sigma_4 = (124)(3)$  and  $m_5 = 2$ , then  $\sigma_5 = (1524)(3)$ . As described in [14], one can define on  $\Sigma_\infty$  the so-called Ewens measure, which is the equivalent on virtual permutations of the Ewens measure on  $\Sigma_N$ . More precisely, let  $\theta \in \mathbb{R}_+$  be a parameter, and let  $(M_j)_{j \geq 1}$  be a sequence of independent random variables,  $M_j \in \{1, \dots, j\}$  such that for all  $j \geq 2$ :

$$\mathbb{P}[M_j = j] = \frac{\theta}{\theta + j - 1},$$

and

$$\mathbb{P}[M_j = k] = \frac{1}{\theta + j - 1}$$

for all  $k < j$ . On  $\Sigma_\infty$ , the Ewens measure  $\mu_\infty^{(\theta)}$  of parameter  $\theta$  is defined as the image of the law of  $(M_j)_{j \geq 1}$  by the map  $\Phi_\infty$ . The name of this measure is consistent, since the image of  $\mu_\infty^{(\theta)}$  by  $\pi_{\infty, N}$  is precisely the Ewens measure  $\mu_N^{(\theta)}$  of parameter  $\theta$  on  $\Sigma_N$ , under which the probability of a permutation  $\sigma_N$  is given by the expression:

$$\frac{\theta^{c(\sigma_N)-1}}{(\theta + 1)(\theta + 2) \cdots (\theta + N - 1)},$$

where  $c(\sigma_N)$  is the number of cycles of  $\sigma_N$ .

**Remark:** For  $\theta = 1$ , the Ewens measure on  $S_N$  is the uniform measure, and for  $\theta = 0$ , it is the uniform measure on permutations with a unique cycle.

In [14] it is proved that if  $(\sigma_N)_{N \geq 1}$  is a virtual permutation following the Ewens measure of parameter  $\theta$ , then for all  $k \geq 0$ , the  $k$ -th length of cycle (by decreasing order) corresponding to the permutation  $\sigma_N$ , divided by  $N$ , tends a.s. to a random variable  $x_k$  when  $N$  goes to infinity. Moreover, the decreasing sequence  $(x_k)_{k \geq 1}$  follows the Poisson-Dirichlet distribution of parameter  $\theta$ . This property can in fact be easily explained by the construction of the Ewens measure on  $\Sigma_\infty$  we give below.

Let  $\lambda = (\lambda_j)_{j \geq 1}$  be a decreasing sequence in  $\mathbb{R}_+$  and let us denote:

$$K(\lambda) := \inf\{k \geq 1, \lambda_k = 0\} \in \mathbb{N}^* \cup \{\infty\}.$$

The set  $E(\lambda)$  is defined as the disjoint union of circles  $(C_j)_{1 \leq j < K(\lambda)}$ , such that  $C_j$  has perimeter  $\lambda_j$ . Now let  $x = (x_k)_{k \geq 1}$  be a sequence of distinct points in  $E(\lambda)$ . One defines a virtual permutation  $\sigma_\infty(\lambda, x) = (\sigma_N(\lambda, x))_{N \geq 1}$  as follows: for  $N \geq 1$ ,  $k \in \{1, \dots, N\}$ , there exists a unique  $j$  such that the point  $x_k$  lies on the circle  $C_j$ . Let us follow the circle  $C_j$ , counterclockwise, starting from  $x_k$ : the image of  $k$  by  $\sigma_N(\lambda, x)$  is the index of the first point in  $\{x_1, \dots, x_N\}$  we encounter after  $x_k$  (for example, if  $x_k$  is the only point in  $C_j$  and  $\{x_1, \dots, x_N\}$ , then  $k$  is a fixed point of  $\sigma_N(\lambda, x)$ , because starting from  $x_k$ , we do a full turn of the circle  $C_j$ , before encountering  $x_k$  again). The cycle structure of  $\sigma_N(\lambda, x)$  is the following: two elements  $k$  and  $l$  in  $\{1, \dots, N\}$  are in the same cycle if and only if  $x_k$  and  $x_l$  lie on the same circle, and the order of the elements  $\{k_1, \dots, k_p\}$  in a given cycle corresponds to the counterclockwise order of the points  $x_{k_1}, \dots, x_{k_p}$ , which are on the same circle. Moreover, the cycle structure of  $\sigma_{N+1}(\lambda, x)$  can be obtained from the structure of  $\sigma_N(\lambda, x)$  by a ‘‘chinese restaurant’’ process:

- If  $x_{N+1}$  is on a circle which does not contain any of the points  $x_1, \dots, x_N$ , then one simply adds the fixed point  $N + 1$ ;
- If  $x_{N+1}$  is on a circle which contains some of the points  $x_1, \dots, x_N$ , and if  $x_{N+1}$  lies just before  $x_p$  if one follows this circle counter-clockwise, then one inserts  $N + 1$  in the cycle containing  $p$ , just before  $p$ .

This construction implies that  $(\sigma_N(\lambda, x))_{N \geq 1}$  is a consistent sequence of permutations, and then  $\sigma_\infty(\lambda, x)$  is a virtual permutation. Therefore a virtual permutation can be viewed as a chinese restaurant process with continuous tables (the circles  $(C_j)_{1 \leq j < K(\lambda)}$ ), and with an infinite number of customers (the points  $(x_k)_{k \geq 1}$ ): its component of index  $N$  is obtained by taking into account only the  $N$  first customers. Note that for the moment,  $\lambda$ , the sequence of lengths of the circles (or the tables!) plays a minor role in our construction. However, it becomes important when one introduces randomness. In the following proposition, we give an explicit construction of the Ewens measure on  $\Sigma_\infty$  (and therefore, on  $\Sigma_N$ , by using  $\pi_{\infty, N}$ ), which involves the notion of *uniform measure* on  $E(\lambda)$ , defined as follows:  $x \in E(\lambda)$  is uniform if and only if for  $j \geq 1$ ,  $x$  lies in  $C_j$  with probability  $\lambda_j$ , the perimeter of  $C_j$  (note that the sum of the perimeters is a.s. equal to one), and conditionally on  $x \in C_j$ ,  $x$  is uniform on the circle  $C_j$ .

**PROPOSITION 2.1.** — *Let  $\theta \in \mathbb{R}_+$ . If the random sequence  $\lambda$  follows the Poisson-Dirichlet distribution of parameter  $\theta$  (for  $\theta = 0$ , one sets  $\lambda_1 = 1$  and  $\lambda_k = 0$  for  $k \geq 1$ ), and if, conditionally on  $\lambda$ , the points  $(x_k)_{k \geq 1}$  are i.i.d., with distribution absolutely continuous with respect to the uniform measure on  $E(\lambda)$  (and hence, a.s. distinct), then for  $x := (x_k)_{k \geq 1}$ , the virtual permutation  $\sigma_\infty(\lambda, x)$  follows the Ewens measure of parameter  $\theta$ .*

*Proof.* — The probability law of a random virtual permutation is uniquely determined by its image by  $\Phi_\infty^{-1}$ , which is the probability law of a sequence  $(M_j)_{j \geq 1}$  of random variables ( $M_j \in \{1, \dots, j\}$ ). By the monotone class theorem, this law is uniquely determined by the sequence of laws of  $(M_j)_{1 \leq j \leq N}$ ,  $N \geq 1$ . Now, by applying  $\Phi_N$  for all  $N \geq 1$ , one deduces that the law of a virtual permutation is uniquely determined by the sequence of laws of its images by  $\pi_{\infty, N}$ ,  $N \geq 1$ . This property implies Proposition 2.1 if one shows that for all  $N \geq 1$ ,  $\sigma_N(\lambda, x)$  follows the Ewens measure of parameter  $\theta$ . To prove this, let us first observe that for all permutations  $\psi \in S_N$ , the law of  $(x_{\psi(k)})_{1 \leq k \leq N}$  is equal to the law of  $(x_k)_{1 \leq k \leq N}$ . Hence, the law of  $\sigma_N(\lambda, x)$  is invariant by conjugation. This implies that the probability of a given permutation, under this law, depends only on its cycle structure, as under the Ewens measure. Therefore, it is sufficient to prove that for all

partitions  $(l_1, \dots, l_p)$  of  $N$ , the probability that the supports of the cycles of  $\sigma_N(\lambda, x)$  are exactly the sets of the form

$$\{l_1 + \dots + l_n + 1, l_1 + \dots + l_n + 2, \dots, l_1 + l_2 + \dots + l_n + l_{n+1}\}$$

for  $0 \leq n \leq p - 1$ , is the same as under the Ewens measure. Now, conditionally on  $\lambda$ , this probability can be written as follows:

$$\sum_{i_1 \neq i_2 \neq \dots \neq i_p} \prod_{n=1}^p \lambda_{i_n}^{l_n}.$$

Hence one only needs to prove the equality:

$$(2.1) \quad \mathbb{E} \left[ \sum_{i_1 \neq i_2 \neq \dots \neq i_p} \prod_{n=1}^p \lambda_{i_n}^{l_n} \right] = \frac{\theta^{p-1}}{(\theta + 1) \dots (\theta + N - 1)} \prod_{n=1}^p (l_n - 1)!$$

since its right-hand side is the probability of the event described above, under the Ewens measure of parameter  $\theta$ . The equation (2.1) is a consequence of the Ewens sampling formula: for sake of completeness, we give a proof here. For  $\theta = 0$ , both sides are equal to one if  $p = 1$  and  $l_1 = N$ , and to zero otherwise: therefore one can assume that  $\theta > 0$ . Let  $(\mu_j)_{j \geq 1}$  be a random sequence following the GEM law of parameter  $\theta$ , and let  $(\nu_j)_{j \geq 1}$  be the sequence obtained by putting  $(\mu_j)_{j \geq 1}$  in decreasing order. One obviously has

$$\sum_{i_1 \neq i_2 \neq \dots \neq i_p} \prod_{n=1}^p \mu_{i_n}^{l_n} = \sum_{i_1 \neq i_2 \neq \dots \neq i_p} \prod_{n=1}^p \nu_{i_n}^{l_n}.$$

Now,  $(\nu_j)_{j \geq 1}$  and  $(\lambda_j)_{j \geq 1}$  have the same law, hence it is sufficient to prove that

$$(2.2) \quad \mathbb{E} \left[ \sum_{i_1 \neq i_2 \neq \dots \neq i_p} \prod_{n=1}^p \mu_{i_n}^{l_n} \right] = \frac{\theta^{p-1}}{(\theta + 1) \dots (\theta + N - 1)} \prod_{n=1}^p (l_n - 1)!$$

One needs the following lemma:

LEMMA 2.2. — *Let  $(\mu_j)_{j \geq 1}$  be a GEM process of parameter  $\theta > 0$ , and  $r, s \in \mathbb{R}_+$ . Then the quantity*

$$E(\theta, r, s) := \mathbb{E} \left[ \sum_{i=1}^{\infty} \mu_i^r \left( 1 - \sum_{j=1}^i \mu_j \right)^s \right]$$

satisfies the equality

$$E(\theta, r, s) = \frac{r!(s + \theta - 1)\theta}{(r + s + \theta - 1)!(r + s)}.$$

*Proof of Lemma 2.2.* — One can write

$$E(\theta, r, s) = \mathbb{E}[\mu_1^r(1 - \mu_1)^s] + \mathbb{E}\left[\left(1 - \mu_1\right)^{r+s} \mathbb{E}\left[\sum_{i=2}^{\infty} \left(\frac{\mu_i}{1 - \mu_1}\right)^r \left(1 - \sum_{j=2}^i \frac{\mu_j}{1 - \mu_1}\right)^s \middle| \mu_1\right]\right].$$

Now conditionally on  $\mu_1$ ,  $\left(\frac{\mu_{j+1}}{1 - \mu_1}\right)_{j \geq 1}$  is a GEM process of parameter  $\theta$ . Therefore:

$$\mathbb{E}\left[\sum_{i=2}^{\infty} \left(\frac{\mu_i}{1 - \mu_1}\right)^r \left(1 - \sum_{j=2}^i \frac{\mu_j}{1 - \mu_1}\right)^s \middle| \mu_1\right] = \mathbb{E}\left[\sum_{i=1}^{\infty} \mu_i^r \left(1 - \sum_{j=1}^i \mu_j\right)^s\right] = E(\theta, r, s).$$

which implies

$$E(\theta, r, s) = \mathbb{E}[\mu_1^r(1 - \mu_1)^s] + E(\theta, r, s)\mathbb{E}[(1 - \mu_1)^{r+s}],$$

and, since the density of the law of  $\mu_1$ , with respect to the Lebesgue measure is  $\theta(1 - x)^{\theta-1}$  on  $(0, 1)$ ,

$$\begin{aligned} E(\theta, r, s) &= \frac{\mathbb{E}[\mu_1^r(1 - \mu_1)^s]}{1 - \mathbb{E}[(1 - \mu_1)^{r+s}]} = \frac{\theta \int_0^1 x^r(1 - x)^{s+\theta-1} dx}{1 - \theta \int_0^1 (1 - x)^{r+s+\theta-1} dx} \\ &= \frac{\theta r!(s + \theta - 1)!(r + s + \theta)!}{1 - \theta/(r + s + \theta)} = \frac{\theta r!(s + \theta - 1)!(r + s + \theta)}{(r + s)(r + s + \theta)!} \end{aligned}$$

which implies Lemma 2.2. □

Now let us go back to the proof of Proposition 2.1. For all integers  $q \geq 0$ ,  $1 \leq j_1 < \dots < j_q$ , and  $r_1, \dots, r_q, r, s \geq 0$ :

$$\begin{aligned} \mathbb{E}\left[\sum_{j=j_q+1}^{\infty} \left(\prod_{p=1}^q \mu_{j_p}^{r_p}\right) \mu_j^r \left(1 - \sum_{i=1}^j \mu_i\right)^s \middle| (\mu_i)_{1 \leq i \leq j_q}\right] \\ = \left(\prod_{p=1}^q \mu_{j_p}^{r_p}\right) \left(1 - \sum_{i=1}^{j_q} \mu_i\right)^{r+s} \mathbb{E}\left[\sum_{j=j_q+1}^{\infty} \left(\frac{\mu_j}{1 - \sum_{i=1}^{j_q} \mu_i}\right)^r \left(1 - \sum_{i=j_q+1}^j \frac{\mu_i}{1 - \sum_{i=1}^{j_q} \mu_i}\right)^s \middle| (\mu_i)_{1 \leq i \leq j_q}\right]. \end{aligned}$$

Since conditionally on  $(\mu_i)_{1 \leq i \leq j_p}$ ,

$$\left(\frac{\mu_{j_q+j'}}{1 - \sum_{i=1}^{j_q} \mu_i}\right)_{j' \geq 1}.$$

is a GEM process of parameter  $\theta$ , the last conditional expectation is equal to  $E(\theta, r, s)$ . One deduces that

$$\begin{aligned} \mathbb{E} \left[ \sum_{j=j_q+1}^{\infty} \left( \prod_{p=1}^q \mu_{j_p}^{r_p} \right) \mu_j^r \left( 1 - \sum_{i=1}^j \mu_i \right)^s \right] \\ = E(\theta, r, s) \mathbb{E} \left[ \left( \prod_{p=1}^q \mu_{j_p}^{r_p} \right) \left( 1 - \sum_{i=1}^{j_q} \mu_i \right)^{r+s} \right]. \end{aligned}$$

By considering all the possible values of  $(j_p)_{1 \leq p \leq q}$ , and by adding the equalities, one obtains that

$$\begin{aligned} \mathbb{E} \left[ \sum_{j_1 < j_2 < \dots < j_{q+1}} \left( \prod_{p=1}^q \mu_{j_p}^{r_p} \right) \mu_{j_{q+1}}^r \left( 1 - \sum_{i=1}^{j_{q+1}} \mu_i \right)^s \right] \\ = E(\theta, r, s) \mathbb{E} \left[ \sum_{j_1 < j_2 < \dots < j_q} \left( \prod_{p=1}^q \mu_{j_p}^{r_p} \right) \left( 1 - \sum_{i=1}^{j_q} \mu_i \right)^{r+s} \right]. \end{aligned}$$

By applying recursively this equality, and using Lemma 2.2, one deduces that

$$\begin{aligned} \mathbb{E} \left[ \sum_{j_1 < j_2 < \dots < j_q} \left( \prod_{p=1}^q \mu_{j_p}^{r_p} \right) \right] &= \prod_{p=1}^q \left[ E \left( \theta, r_p, \sum_{m=p+1}^q r_m \right) \right] \\ &= \prod_{p=1}^q \frac{\theta(r_p)! \left[ \left( \sum_{m=p+1}^q r_m \right) + \theta - 1 \right]!}{\left( \sum_{m=p}^q r_m \right) \left[ \left( \sum_{m=p}^q r_m \right) + \theta - 1 \right]!} \\ &= \theta^q \left( \prod_{p=1}^q r_p! \right) \frac{(\theta - 1)!}{\left[ \left( \sum_{m=1}^q r_m \right) + \theta - 1 \right]!} \prod_{p=1}^q \frac{1}{\sum_{m=p}^q r_m}. \end{aligned}$$

We can now compute the left-hand side of (2.2):

$$\begin{aligned} \mathbb{E} \left[ \sum_{i_1 \neq i_2 \neq \dots \neq i_p} \prod_{n=1}^p \mu_{i_n}^{l_n} \right] &= \sum_{\sigma \in \Sigma_p} \mathbb{E} \left[ \sum_{i_{\sigma(1)} < \dots < i_{\sigma(p)}} \prod_{n=1}^p \mu_{i_{\sigma(n)}}^{l_{\sigma(n)}} \right] \\ &= \sum_{\sigma \in \Sigma_p} \mathbb{E} \left[ \sum_{i_1 < \dots < i_p} \prod_{n=1}^p \mu_{i_n}^{l_{\sigma(n)}} \right] \\ &= \sum_{\sigma \in \Sigma_p} \theta^p \left( \prod_{n=1}^p l_{\sigma(n)}! \right) \frac{(\theta - 1)!}{\left[ \left( \sum_{n=1}^p l_{\sigma(n)} \right) + \theta - 1 \right]!} \prod_{n=1}^p \frac{1}{\sum_{m=n}^p l_{\sigma(m)}} \end{aligned}$$

$$= \theta^p \left( \prod_{n=1}^p l_n! \right) \frac{(\theta - 1)!}{(\theta + N - 1)!} \sum_{\sigma \in \Sigma_p} \frac{1}{\prod_{n=1}^p \sum_{m=n}^p l_{\sigma(m)}}.$$

Therefore, (2.2) and then Proposition 2.1, is proved if one checks the equality:

$$\sum_{\sigma \in \Sigma_p} \frac{1}{\prod_{n=1}^p \sum_{m=n}^p l_{\sigma(m)}} = \frac{1}{\prod_{n=1}^p l_n}.$$

Now, since for all  $l > 0$ ,

$$\frac{1}{l} = \int_0^\infty e^{-lx} dx,$$

one deduces that

$$\begin{aligned} \sum_{\sigma \in \Sigma_p} \frac{1}{\prod_{n=1}^p \sum_{m=n}^p l_{\sigma(m)}} &= \sum_{\sigma \in \Sigma_p} \int_{x_1, \dots, x_p \geq 0} e^{-\sum_{n=1}^p x_n (\sum_{m=n}^p l_{\sigma(m)})} \prod_{n=1}^p dx_n \\ &= \sum_{\sigma \in \Sigma_p} \int_{x_1, \dots, x_p \geq 0} e^{-\sum_{m=1}^p l_{\sigma(m)} \sum_{n=1}^m x_n} \prod_{n=1}^p dx_n. \end{aligned}$$

By doing the change of variable

$$y_{\sigma(m)} = \sum_{n=1}^m x_n,$$

one obtains:

$$\begin{aligned} &\sum_{\sigma \in \Sigma_p} \frac{1}{\prod_{n=1}^p \sum_{m=n}^p l_{\sigma(m)}} \\ &= \sum_{\sigma \in \Sigma_p} \int_{y_1, \dots, y_p \geq 0} e^{-\sum_{m=1}^p l_{\sigma(m)} y_{\sigma(m)}} \mathbb{1}_{y_{\sigma(1)} \leq \dots \leq y_{\sigma(p)}} \prod_{n=1}^p dy_n \\ &= \int_{y_1, \dots, y_p \geq 0} e^{-\sum_{m=1}^p l_m y_m} \prod_{n=1}^p dy_n = \frac{1}{\prod_{n=1}^p l_n} \end{aligned}$$

which completes the proof of Proposition 2.1. □

As an illustration, we quickly show how this proposition implies the following almost sure convergence result for relative cycle lengths due to Tsilevich:

**PROPOSITION 2.3** (Tsilevich [14]). — *Let  $(\sigma_N)_{N \geq 1}$  be a virtual permutation following the Ewens probability measure with parameter  $\theta$ . One defines the sequence  $(\alpha_k^{(N)})_{k \geq 1}$  of normalized lengths of cycles of  $\sigma_N$  (i.e. lengths divided by  $N$ ), ordered by increasing smallest elements, and completed by zeros. Then, for all  $k \geq 1$ ,  $\alpha_k^{(N)}$  converges almost surely to a*



random variable  $\alpha_k^{(\infty)}$ , and  $(\alpha_k^{(\infty)})_{k \geq 1}$  follows a GEM distribution of parameter  $\theta$ . In particular, the law of  $(\alpha_k^{(N)})_{k \geq 1}$  converges weakly to the GEM( $\theta$ ) distribution. Moreover, if  $y_l^{(N)}$  denotes the  $l$ -th largest element of  $(\alpha_k^{(N)})_{k \geq 1}$  for all integers  $N \geq 1$  and  $N = \infty$ , then

$$y_l^{(N)} \xrightarrow{N \rightarrow \infty} y_l^{(\infty)}$$

a.s., and  $(y_l^{(\infty)})_{l \geq 1}$  follows a Poisson-Dirichlet distribution of parameter  $\theta$ . In particular, the law of  $(y_l^{(N)})_{l \geq 1}$  (i.e. the sequence of decreasing normalized lengths of cycles) tends to the PD( $\theta$ ) distribution.

*Proof.* — Let us construct  $(\sigma_N)_{N \geq 1}$  via Proposition 2.1. Since Proposition 2.3 is trivial for  $\theta = 0$ , one can suppose  $\theta > 0$ . This implies (with the notation of Proposition 2.1) that  $\lambda_j > 0$  for all  $j \geq 1$ , and a.s., there exists  $p$  such that  $x_p$  lies on the circle  $C_j$ . We define a sequence  $(j_n)_{n \geq 1}$  by the following recursive construction:

- The index  $j_1$  is given by:  $x_1 \in C_{j_1}$ ;
- For  $n \geq 1$ ,  $j_1, \dots, j_n$  already defined,  $j_{n+1}$  is given by:  $x_p \in C_{j_{n+1}}$ , where  $p$  is the smallest index such that  $x_p \notin C_{j_1} \cup \dots \cup C_{j_n}$  (this index a.s. exists).

It is easy to check that for  $k \geq 1$ :

$$\alpha_k^{(N)} = \frac{|C_{j_k} \cap \{x_1, \dots, x_N\}|}{N} = \frac{1}{N} \sum_{p=1}^N \mathbb{1}_{x_p \in C_{j_k}}.$$

Now, by the law of large numbers, it is almost sure that for all integers  $j \geq 1$ :

$$\frac{1}{N} \sum_{p=1}^N \mathbb{1}_{x_p \in C_j} \xrightarrow{N \rightarrow \infty} \lambda_j.$$

Then,  $\alpha_k^{(N)}$  tends almost surely to  $\lambda_{j_k}$ . Now, by construction of the sequence  $(j_k)_{k \geq 1}$ , we see that  $(\lambda_{j_k})_{k \geq 1}$  is the classical size-biased reordering of  $(\lambda_j)_{j \geq 1}$ , and hence a GEM process of parameter  $\theta$ . Now, it is obvious that  $(y_l^{(\infty)})_{l \geq 1}$  is the decreasing reordering of  $(\lambda_{j_k})_{k \geq 1}$ , i.e. the PD( $\theta$ ) process  $(\lambda_l)_{l \geq 1}$ . It remains to prove that  $y_l^{(N)} \rightarrow \lambda_l$  almost surely. Indeed, if  $l$  is fixed, there exists a.s. a (random) index  $N_0 > l$  such that:

$$\sum_{k=1}^{N_0} \alpha_k^{(\infty)} > 1 - \lambda_l/2.$$

Since  $\alpha_k^{(N)} \xrightarrow{N \rightarrow \infty} \alpha_k^{(\infty)}$  for all  $k \geq 1$ , there exists a.s. a (random)  $N_1$  such that for  $N \geq N_1$ :

$$\sum_{k=1}^{N_0} \alpha_k^{(N)} > 1 - \lambda_l/2.$$

Since the numbers  $(\alpha_k^{(\infty)})_{k \geq 1}$  are a.s. pairwise distinct, there exists a.s.  $N_2$  such that if  $N \geq N_2$ , the order of  $(\alpha_k^{(N)})_{1 \leq k \leq N_0}$  is the same as the order of  $(\alpha_k^{(\infty)})_{1 \leq k \leq N_0}$ . In particular (recall that  $N_0 > l$ ), the  $l$ -th largest element of  $(\alpha_k^{(N)})_{1 \leq k \leq N_0}$  has the same index  $r$  as the  $l$ -th largest element of  $(\alpha_k^{(\infty)})_{1 \leq k \leq N_0}$ . Now, since  $\alpha_k^{(\infty)} < \lambda_l/2$  for all  $k > N_0$ ,  $\alpha_r^{(\infty)}$  is also the  $l$ -th largest element of  $(\alpha_k^{(\infty)})_{k \geq 1}$ , i.e.  $\lambda_l$ . For  $N \geq \sup(N_1, N_2)$ ,  $\alpha_r^{(N)}$  is the  $l$ -th largest element of  $(\alpha_k^{(N)})_{1 \leq k \leq N_0}$  and  $\alpha_k^{(N)} < \lambda_l/2$  for  $N > N_0$ , hence the  $l$ -th largest element  $y_l^{(N)}$  of  $(\alpha_k^{(N)})_{k \geq 1}$  is included in the interval  $[\alpha_r^{(N)}, \alpha_r^{(N)} \vee \lambda_l/2]$ . Since  $\alpha_r^{(N)} \xrightarrow{N \rightarrow \infty} \alpha_r^{(\infty)} = \lambda_l$ , Proposition 2.3 is proved. □

### 3. The “non-unitary case”

#### 3.1. The normalized and non-normalized empirical eigenvalues distributions

Let  $M$  be a matrix in  $\mathcal{G}(N)$  for some  $N \geq 1$ . We associate with the point process of the eigenvalues of  $M$  the finite measure  $\mu(M)$  on  $\mathbb{C}$  defined by

$$\mu(M) := \sum_{\lambda \in E(M)} m_M(\lambda) \delta_\lambda,$$

where  $E(M)$  is the set of eigenvalues of  $M$ ,  $m_M(\lambda)$  is the multiplicity of  $\lambda$  as an eigenvalue of  $M$ , and  $\delta_\lambda$  is Dirac measure at  $\lambda$ . By the general description of eigenvalues given in Section 1.2, one has, for all  $\sigma \in \Sigma_N$ ,  $z_1, \dots, z_N \in \mathbb{C}^*$ :

$$\mu(M(\sigma, z_1, \dots, z_N)) = \sum_{m=1}^n \sum_{\omega^{l_m} = Z_m} \delta_\omega,$$

where  $l_1, l_2, \dots, l_n$  are the lengths of the cycles  $C_1, C_2, \dots, C_n$  of  $\sigma$ , and for  $1 \leq m \leq n$ :

$$Z_m = \prod_{j \in C_m} z_j.$$

Let us now suppose that the distribution of a sequence of random matrices  $(M_N)_{N \geq 1}$ ,  $M_N \in \mathcal{G}(N)$ , is of the form  $\mathbb{P}(\infty, \theta, \mathcal{L})$ . One has  $M_N =$

$M(\sigma_N, z_1, \dots, z_N)$  where  $(\sigma_N)_{N \geq 1}$  follows the Ewens( $\theta$ ) distribution, and is independent of the sequence  $(z_j)_{j \geq 1}$  of i.i.d. variables, which have law  $\mathcal{L}$ . Since for all  $N \geq 1$ , the cycle structure of  $\sigma_N$  can be deduced from the cycle structure of  $\sigma_{N+1}$  by removing  $N + 1$ , there exists a partition  $\Pi$  of  $\mathbb{N}^*$  such that for all  $N$ , the supports of the cycles of  $\sigma_N$  are obtained by intersecting the sets of  $\Pi$  with  $\{1, \dots, N\}$ . Moreover, under the Ewens( $\theta$ ) measure,  $\Pi$  contains a.s. an infinite number of sets (see Section 2): let us order them by increasing smallest elements, and denote them by  $(C_m)_{m \geq 1}$ . One then has

$$(3.1) \quad \mu(M_N) = \sum_{m=1}^{\infty} \mathbb{1}_{l_{N,m} > 0} \sum_{\omega^{l_{N,m}} = Z_{N,m}} \delta_{\omega},$$

where  $l_{N,m}$  is the cardinality of  $C_{N,m}$ , the intersection of  $C_m$  and  $\{1, \dots, N\}$ , and

$$Z_{N,m} = \prod_{j \in C_{N,m}} z_j.$$

The natural question one can now ask is the behaviour of the measure  $\mu(M_N)$  for large  $N$ . Since each cycle of  $\sigma_N$  gives a number of eigenvalues equal to its length, one can expect that  $\mu(M_N)$  is dominated by the large cycles of  $M_N$ . Moreover, the  $l$  eigenvalues corresponding to a cycle of length  $l$  form a regular polygon of order  $l$ , and the distance of their vertices to the origin is equal to the  $l$ -th root of the product of  $l$  i.i.d. random variables of law  $\mathcal{L}$ . If  $l$  is large and if one can apply a multiplicative version of the law of large numbers, one can expect that this distance does not vary too much. Then, it is natural to guess that under some well-chosen conditions on  $\mathcal{L}$ , the measure  $\mu(M_N)$ , which has total mass  $N$ , is close to  $N$  times the uniform measure on a circle centered at the origin. Indeed, we can prove the following statement:

**PROPOSITION 3.1.** — *Let  $(M_N)_{N \geq 1}$  be a sequence of matrices following the law  $\mathbb{P}(\infty, \theta, \mathcal{L})$  for some  $\theta > 0$  and some probability  $\mathcal{L}$  on  $\mathbb{C}^*$ . We suppose that if  $Z$  is a random variable which follows the distribution  $\mathcal{L}$ , then  $\log(|Z|)$  is integrable. Under these assumptions, almost surely, the probability measure  $\mu(M_N)/N$  converges weakly to the uniform distribution on the circle of center zero and radius  $\exp(\mathbb{E}[\log(|Z|)])$ .*

*Proof.* — Let  $f$  be a continuous and bounded function from  $\mathbb{C}$  to  $\mathbb{R}$ , and let  $R > 0$ . Then, there exists a constant  $A > 0$ , and a function  $\alpha$  from  $(0, R)$  to  $\mathbb{R}_+$ , tending to zero at zero, such that for all  $\epsilon \in (0, R)$ , for all

integers  $l \geq 1$ , and for all  $z \in \mathbb{C}$  such that  $|z|^{1/l} \in (R - \epsilon, R + \epsilon)$ :

$$(3.2) \quad \left| \sum_{\omega^l=z} f(\omega) - \frac{l}{2\pi} \int_0^{2\pi} f(Re^{i\lambda}) d\lambda \right| \leq \frac{A}{\epsilon} + l\alpha(\epsilon).$$

Indeed, let us define for all  $\delta > 0$ :

$$\eta(\delta) := \sup\{|f(y) - f(y')|, |y - y'| \leq \delta, |y|, |y'| \leq 2R\},$$

which tends to zero with  $\delta$  since  $f$  is uniformly continuous on any compact set. With this definition, we obtain:

$$\left| \sum_{\omega^l=z} f(\omega) - \sum_{\omega^l=z'} f(\omega) \right| \leq l\eta(\epsilon),$$

where  $z' := zR^l/|z|^l$  has modulus  $R^l$ . Now, there exists  $\lambda \in [0, 2\pi/l)$  such that:

$$\sum_{\omega^l=z'} f(\omega) = \sum_{\omega^l=1} f(R\omega e^{i\lambda}) =: \Phi(\lambda)$$

One has, for all  $\lambda, \lambda' \in [0, 2\pi/l)$ :

$$|\Phi(\lambda) - \Phi(\lambda')| \leq l\eta(R|\lambda - \lambda'|) \leq l\eta(2\pi R/l).$$

Moreover

$$\int_0^{2\pi/l} \Phi(\lambda) d\lambda = \int_0^{2\pi} f(Re^{i\lambda}) d\lambda,$$

and then, for all  $\lambda \in [0, 2\pi/l)$ ,

$$\left| \Phi(\lambda) - \frac{l}{2\pi} \int_0^{2\pi} f(Re^{i\lambda}) d\lambda \right| \leq l\eta(2\pi R/l),$$

which implies:

$$\left| \sum_{\omega^l=z} f(\omega) - \frac{l}{2\pi} \int_0^{2\pi} f(Re^{i\lambda}) d\lambda \right| \leq l[\eta(\epsilon) + \eta(2\pi R/l)].$$

If  $l \leq 2\pi R/\epsilon$ , one can majorize this quantity by  $4\pi R\eta(2\pi R)/\epsilon$ , and if  $l \geq 2\pi R/\epsilon$ , one can majorize it by  $2l\eta(\epsilon)$ . Hence we obtain (3.2), for  $A = 4\pi R\eta(2\pi R)$  and  $\alpha(\epsilon) = 2\eta(\epsilon)$ .

Since for  $B > 0$  depending only on  $f$ , the left-hand side of (3.2) can be trivially majorized by  $B$  for any  $z \in \mathbb{C}$ , we deduce, for  $\mu$  equal to the uniform measure on the circle of radius  $R$ :

$$\begin{aligned} & \left| \frac{1}{N} \int_{\mathbb{C}} f d\mu(M_N) - \int_{\mathbb{C}} f d\mu \right| \\ & \leq \sum_{m=1}^{\infty} \mathbb{1}_{l_{N,m} > 0} \left[ \frac{Bl_{N,m}}{N} \mathbb{1}_{|Z_{N,m}|^{1/l_{N,m}} \notin (R-\epsilon, R+\epsilon)} + \frac{A}{N\epsilon} + \frac{l_{N,m}}{N} \alpha(\epsilon) \right]. \end{aligned}$$

For now, let us take  $R := \exp\left(\mathbb{E}[\log(|Z|)]\right)$ , where  $Z$  is a random variable following the law  $\mathcal{L}$ . By the strong law of large numbers applied to the sequence  $(\log |z_j|)_{j \in C_m}$ , it is not difficult to check that a.s., for all  $m \geq 1$ :

$$\frac{Bl_{N,m}}{N} \mathbb{1}_{|Z_{N,m}|^{1/l_{N,m}} \notin (R-\epsilon, R+\epsilon)}$$

tends to zero when  $N$  goes to infinity. Moreover, independently of  $N$ , this quantity is dominated by  $Bs_m$ , where  $s_m$  is the supremum of  $l_{N,m}/N$  for  $N \geq 1$ . For the moment, let us assume that a.s.:

$$(3.3) \quad \sum_{m=1}^{\infty} s_m < \infty$$

In this case, one can apply dominated convergence and obtain:

$$\sum_{m=1}^{\infty} \mathbb{1}_{l_{N,m} > 0, |Z_{N,m}|^{1/l_{N,m}} \notin (R-\epsilon, R+\epsilon)} \frac{Bl_{N,m}}{N} \xrightarrow{N \rightarrow \infty} 0.$$

Moreover,

$$\sum_{m=1}^{\infty} \mathbb{1}_{l_{N,m} > 0} \frac{A}{N\epsilon} \xrightarrow{N \rightarrow \infty} 0$$

a.s., since the number of cycles of  $\sigma_N$  increases slowly with respect to  $N$  (the order of magnitude is  $\log(N)$ ), and

$$\sum_{m=1}^{\infty} \mathbb{1}_{l_{N,m} > 0} \frac{l_{N,m}}{N} \alpha(\epsilon) = \alpha(\epsilon).$$

Hence we deduce that

$$\limsup_{N \rightarrow \infty} \left| \frac{1}{N} \int_{\mathbb{C}} f d\mu(M_N) - \int_{\mathbb{C}} f d\mu \right| \leq \alpha(\epsilon),$$

and by taking  $\epsilon \rightarrow 0$ , we are done. It only remains to prove (3.3). This relation can be shown by looking carefully at the construction of the Ewens( $\theta$ ) measure on virtual permutations given in Section 2. Indeed, for  $N \geq 1$ , conditionally on  $(l_{K,m})_{m \geq 1, 1 \leq K \leq N}$ , with  $m_0 := \inf\{m \geq 1, l_{N,m} = 0\}$ , one has  $l_{N+1,m} = l_{N,m} + \mathbb{1}_{m=m_1}$  where  $m_1$  is a random index equal to  $m'$  with probability  $l_{N,m}/(N + \theta)$  for  $m' < m_0$  and to  $m_0$  with probability  $\theta/(N + \theta)$ . This implies quite easily that  $(l_{N,m}/(N + \theta))_{N \geq 1}$  is a nonnegative submartingale. By Doob's inequality one deduces that the expectation of  $s_m^2$  is dominated by a constant (depending only on  $\theta$ ) times the expectation of  $y_m^2$ , where  $y_m$  is the limit of  $l_{N,m}/N$ , which exists almost surely.

Now,  $(y_m)_{m \geq 0}$  is a GEM process of parameter  $\theta$ , hence the expectation of  $y_m^2$  decreases exponentially with  $m$ . Consequently,

$$(3.4) \quad \sum_{m=1}^{\infty} \mathbb{E}[s_m] < \infty,$$

which implies (3.3) almost surely. □

If we do not want to deal with virtual permutations, we can replace the a.s. convergence by a weak convergence in probability, as follows:

**COROLLARY 3.2.** — *Let  $(M_N)_{N \geq 1}$  be a sequence of matrices such that  $M_N \in \mathcal{G}(N)$  follows the distribution  $\mathbb{P}(N, \theta, \mathcal{L})$  for some  $\theta > 0$  and some probability  $\mathcal{L}$  on  $\mathbb{C}^*$ . We suppose that if  $Z$  is a random variable which follows the distribution  $\mathcal{L}$ , then  $\log(|Z|)$  is integrable. Under these assumptions, the probability measure  $\mu(M_N)/N$  converges weakly in probability to the uniform distribution on the circle of center zero and radius  $R := \exp(\mathbb{E}[\log(|Z|)])$ , i.e. for all continuous, bounded functions  $f$  from  $\mathbb{C}$  to  $\mathbb{R}$ :*

$$\frac{1}{N} \int_{\mathbb{C}} f d\mu(M_N) \xrightarrow{N \rightarrow \infty} \frac{1}{2\pi} \int_0^{2\pi} f(Re^{i\lambda}) d\lambda$$

in probability.

This convergence result means that most of the eigenvalues of a matrix  $M_N$  following  $\mathbb{P}(N, \theta, \mathcal{L})$  are concentrated around the circle of radius  $\exp(\mathbb{E}[\log(|Z|)])$ . Now, even for  $N$  large, there remain some eigenvalues which are far from this circle, since the law of large numbers involved in the proof of Proposition 3.1 does not apply for the small cycles of the permutation  $\sigma_N$  associated with  $M_N$ . In order to study the influence of the small cycles, let us suppose that  $(M_N)_{N \geq 1}$  follows the measure  $\mathbb{P}(\infty, \theta, \mathcal{L})$  (which is possible since the image of  $\mathbb{P}(\infty, \theta, \mathcal{L})$  by the  $N$ -th coordinate is  $\mathbb{P}(N, \theta, \mathcal{L})$ ). Then, one can write the measure  $\mu(M_N)$  in the following way:

$$\mu(M_N) = \sum_{k=1}^{\infty} \sum_{m \geq 1, l_{N,m}=k} \sum_{\omega^k = Z_{N,m}} \delta_{\omega},$$

with the same notation as in equation (3.1). This equality implies the following equality in distribution:

$$(3.5) \quad \mu(M_N) = \sum_{k=1}^{\infty} \sum_{p=1}^{a_{N,k}} \sum_{\omega^k = T_{k,p}} \delta_{\omega},$$

where for all  $k \geq 1$ ,  $a_{N,k}$  is the number of  $k$ -cycles of the permutation  $\sigma_N$  (which follows the Ewens( $\theta$ ) measure on  $\Sigma_N$ ), where for  $k, p \geq 1$ , the law of  $T_{k,p}$  is the multiplicative convolution of  $k$  copies of the distribution  $\mathcal{L}$ , and

where  $(a_{N,k})_{k \geq 1}$  and the variables  $T_{k,p}$ ,  $k, p \geq 1$  are independent. Now, the finite dimensional marginals of  $(a_{N,k})_{k \geq 1}$  converge, in distribution, to the corresponding marginals of  $(a_k)_{k \geq 1}$ , where the variables  $a_k$  are independent Poisson random variables, with  $\mathbb{E}[a_k] = \theta/k$  (see for instance [2]). One can then expect that in a sense which needs to be made precise, the law of  $\mu(M_N)$  converges to the distribution of

$$(3.6) \quad \mu_\infty := \sum_{k=1}^\infty \sum_{p=1}^{a_k} \sum_{\omega^k = T_{k,p}} \delta_\omega,$$

where all the variables  $a_k$  and  $T_{k,p}$  in sight are independent. Of course, one needs to be careful, because the measure  $\mu_\infty$  has an infinite total mass, which, under the assumptions of Proposition 3.1, is expected to concentrate around the circle of radius  $\exp(\mathbb{E}[\log(|Z|)])$ . One also remarks that the convergence expected here is very different from the convergence proved in Proposition 3.1; in particular, it involves the measure  $\mu(M_N)$  and not the probability  $\mu(M_N)/N$ . In order to state rigorously our result, let us give the following definition:

DEFINITION 3.3. — *Let  $X$  be a real, integrable, random variable. For  $q > 0$ , we say that  $X$  is in  $\mathcal{U}^q$  if and only if for a sequence  $(X_k)_{k \geq 1}$  of i.i.d. random variables with the same distribution as  $X$  and for all  $\epsilon > 0$ , there exists  $C > 0$  such that for all  $n \geq 1$ :*

$$\mathbb{P} \left[ \left| \left( \frac{1}{n} \sum_{k=1}^n X_k \right) - \mathbb{E}[X] \right| \geq \epsilon \right] \leq \frac{C}{n^q}.$$

Remark 3.4. — If  $q \geq 1$  is an integer, by expanding

$$\mathbb{E} \left[ \left( \frac{1}{n} \sum_{k=1}^n (X_k - \mathbb{E}[X]) \right)^{2q} \right]$$

and by using Markov’s inequality, one easily proves that a random variable in  $L^{2q}$  is also in  $\mathcal{U}^q$ .

We can now state a result about the convergence of  $\mu(M_N)$ .

PROPOSITION 3.5. — *Let  $(M_N)_{N \geq 1}$  be a sequence of matrices such that  $M_N \in \mathcal{G}(N)$  follows the distribution  $\mathbb{P}(N, \theta, \mathcal{L})$  for some  $\theta > 0$  and some probability  $\mathcal{L}$  on  $\mathbb{C}^*$ . We suppose that if  $Z$  is a random variable following the distribution  $\mathcal{L}$ , then  $\log(|Z|)$  is in  $\mathcal{U}^q$  for some  $q > 0$ . Under these assumptions, for all bounded, continuous functions  $f$  from  $\mathbb{C}$  to  $\mathbb{R}$ , such that  $f = 0$  on a neighborhood of the circle  $|z| = R$ , where  $R := \exp(\mathbb{E}[\log(|Z|)])$ ,  $f$  is a.s. integrable with respect to  $\mu_\infty$ , where the random measure  $\mu_\infty$  is*

given by (3.6). Moreover, the following convergence in distribution holds for such test functions  $f$ :

$$\int_{\mathbb{C}} f \, d\mu(M_N) \xrightarrow{N \rightarrow \infty} \int_{\mathbb{C}} f \, d\mu_{\infty}.$$

*Proof.* — Let  $(\xi_r)_{r \geq 1}$  be a sequence of independent Bernoulli random variables, such that the parameter of  $\xi_r$  is equal to  $\theta/(\theta + r - 1)$  (in particular,  $\xi_1 = 1$  almost surely). We suppose that  $(\xi_r)_{r \geq 1}$  is independent of  $(T_{k,p})_{k,p \geq 1}$ , and for all  $N, k \geq 1$ , we define  $b_{N,k}$  as the number of pairs of consecutive ones in the sequence  $(\xi_1, \xi_2, \dots, \xi_N, 1)$  which are separated by a distance of  $k$ , and  $b_k$  as the analog for the infinite sequence  $(\xi_r)_{r \geq 1}$ . In other words:

$$b_{N,k} = \mathbb{1}_{N+1-k \geq 1, \xi_{N+1-k}=1, \xi_{N+2-k}=\dots=\xi_N=0} + \sum_{1 \leq j \leq N-k} \mathbb{1}_{\xi_j=\xi_{j+k}=1, \xi_{j+1}=\dots=\xi_{j+k-1}=0},$$

and

$$b_k = \sum_{j \geq 1} \mathbb{1}_{\xi_j=\xi_{j+k}=1, \xi_{j+1}=\dots=\xi_{j+k-1}=0}.$$

By the classical properties of the Feller coupling (see for instance [2]),  $(b_k)_{k \geq 1}$  has the same distribution as  $(a_k)_{k \geq 1}$  and for all  $N \geq 1$ ,  $(b_{N,k})_{k \geq 1}$  has the same distribution as  $(a_{N,k})_{k \geq 1}$ . Therefore, in Proposition 3.5, one can replace  $\mu(M_N)$  by  $\nu_N$  and  $\mu_{\infty}$  by  $\nu_{\infty}$ , where:

$$\nu_N = \sum_{k=1}^{\infty} \sum_{p=1}^{b_{N,k}} \sum_{\omega^k = T_{k,p}} \delta_{\omega},$$

and

$$\nu_{\infty} := \sum_{k=1}^{\infty} \sum_{p=1}^{b_k} \sum_{\omega^k = T_{k,p}} \delta_{\omega}.$$

Let  $f$  be a continuous and bounded function from  $\mathbb{C}$  to  $\mathbb{R}$ , equal to zero in a neighborhood of the circle  $|z| = R$ . There exist  $A > 0$ ,  $0 < R_1 < R < R_2$ , depending only on  $f$ , such that  $|f(z)| \leq A \mathbb{1}_{|z| \notin (R_1, R_2)}$ , and then,

$$\int_{\mathbb{C}} |f| \, d\nu_{\infty} \leq A \sum_{k=1}^{\infty} \sum_{p=1}^{b_k} k \mathbb{1}_{|T_{k,p}|^{1/k} \notin (R_1, R_2)}.$$

The function  $f$  is a.s. integrable with respect to  $\nu_{\infty}$  (and then,  $\mu_{\infty}$ ). Indeed,

$$\int_{\mathbb{C}} |f| \, d\nu_{\infty} < \infty$$



if and only if

$$\sum_{k=1}^{\infty} \sum_{p=1}^{b_k} \mathbb{1}_{|T_{k,p}|^{1/k} \notin (R_1, R_2)} < \infty,$$

and the expectation of this quantity is

$$\theta \sum_{k=1}^{\infty} \frac{\mathbb{P}[|T_{k,1}|^{1/k} \notin (R_1, R_2)]}{k},$$

which is finite since  $\log(|Z|)$  is in  $\mathcal{U}^q$  for some  $q > 0$ . Let us now introduce the random measure:

$$\nu'_N = \sum_{k=1}^{\infty} \sum_{p=1}^{c_{N,k}} \sum_{\omega^k = T_{k,p}} \delta_{\omega},$$

where

$$(3.7) \quad c_{N,k} = \sum_{1 \leq j \leq N-k} \mathbb{1}_{\xi_j = \xi_{j+k} = 1, \xi_{j+1} = \dots = \xi_{j+k-1} = 0}.$$

One has  $b_{N,k} = c_{N,k}$ , except for  $k$  equal to the smallest integer  $k_0$  such that  $\xi_{N+1-k_0} = 1$ , in which case  $b_{N,k_0} = c_{N,k_0} + 1$ . Therefore,

$$\nu_N = \nu'_N + \nu''_N,$$

where

$$\nu''_N = \sum_{\omega^{k_0} = T_{k_0, c_{N, k_0} + 1}} \delta_{\omega}.$$

Since  $f$  is a.s. integrable with respect to  $\nu_{\infty}$ , and for all  $k \geq 1$ ,  $b_k$  is the increasing limit of  $c_{N,k}$ , one obtains, a.s.:

$$(3.8) \quad \int_{\mathbb{C}} f d\nu'_N \xrightarrow{N \rightarrow \infty} \int_{\mathbb{C}} f d\nu_{\infty}.$$

Moreover, there exist  $0 < R_1 < R < R_2$ , depending only on  $f$ , such that for all  $k' \geq 1$ :

$$\begin{aligned} \mathbb{P} \left[ \int_{\mathbb{C}} f d\nu''_N \neq 0 \right] &\leq \mathbb{P} \left[ |T_{k_0, c_{N, k_0} + 1}|^{1/k_0} \notin (R_1, R_2) \right] \\ &= \sum_{k \geq 1} \mathbb{P}[k_0 = k] \mathbb{P}[|T_{k,1}|^{1/k} \notin (R_1, R_2)] \\ &\leq \mathbb{P}[k_0 \leq k'] + \sup_{k'' > k'} \mathbb{P} \left[ |T_{k'',1}|^{1/k''} \notin (R_1, R_2) \right] \end{aligned}$$

Now,  $\mathbb{P}[k_0 \leq k']$  tends to zero when  $N$  goes to infinity (the order of magnitude is  $1/N$ ), and by taking  $N \rightarrow \infty$ , and then  $k' \rightarrow \infty$ , one deduces:

$$(3.9) \quad \mathbb{P} \left[ \int_{\mathbb{C}} f d\nu''_N \neq 0 \right] \xrightarrow{N \rightarrow \infty} 0.$$

Therefore, by taking (3.8) and (3.9) together:

$$\int_{\mathbb{C}} f d\nu'_N \xrightarrow{N \rightarrow \infty} \int_{\mathbb{C}} f d\nu_{\infty}$$

in probability, and a fortiori, in distribution. □

*Remark 3.6.* — In the proof of Proposition 3.5, we have used the Feller coupling in order to replace the convergence in law by a convergence in probability. However, this coupling does not correspond to the coupling used by considering the measure  $\mathbb{P}(\infty, \theta, \mathcal{L})$ . Moreover, if  $(M_N)_{N \geq 1}$  follows this measure, the number of cycles of a given length in  $\sigma_N$  does not converge when  $N$  goes infinity, and the support of these cycles is changing infinitely often. Hence, one cannot expect an almost sure convergence (or even a convergence in probability) in Proposition 3.5.

Propositions 3.1 and 3.5 apply for a large family of distributions  $\mathcal{L}$ , however, some integrability conditions need to be satisfied. One can ask what happens if these conditions do not hold. We are not able to prove a result for all the possible distributions  $\mathcal{L}$ , but it is possible to study some important particular cases, if the probability distributions involved in our problem can be explicitly computed. Here, the most important distributions we need to deal with are the probability measures  $\mathcal{L}_k$ ,  $k \geq 1$ , defined in the following way:  $\mathcal{L}_k$  is the unique measure on  $\mathbb{C}^*$ , invariant by multiplication by a  $k$ -th root of unity, and such that its image by the  $k$ -th power is the multiplicative convolution of  $k$  copies of  $\mathcal{L}$ . Intuitively,  $\mathcal{L}_k$  is the law of a random  $k$ -th root of the product of  $k$  independent random variables with law  $\mathcal{L}$ , chosen uniformly among the  $k$  possible roots. The measures  $\mathcal{L}_k$  are not easy to compute in general. One case where the computation is simplified is the case where  $\mathcal{L}$  has the radial symmetry. Indeed let  $(Y_p)_{p \geq 1}$  be a sequence of i.i.d. random variables with the same law as  $\log(|Z|)$ , where  $Z$  follows the distribution  $\mathcal{L}$ . The distribution  $\mathcal{L}_k$  is the law of  $e^{i\Theta + \frac{1}{k} \sum_{p=1}^k Y_p}$  where  $\Theta$  is independent of  $(Y_p)_{p \geq 1}$  and uniform on  $[0, 2\pi)$ . In particular, if  $\mathcal{L}$  is the law of  $e^{i\Theta + \rho S_{\alpha}}$ , where  $\rho$  is a strictly positive parameter,  $\Theta$  a uniform variable on  $[0, 2\pi)$  and  $S_{\alpha}$  ( $\alpha \in (0, 2]$ ) an independent standard symmetric stable random variable of index  $\alpha$ , then  $\mathcal{L}_k$  is the law of  $e^{i\Theta + \rho k^{(1-\alpha)/\alpha} S_{\alpha}}$ . Using this explicit description of  $\mathcal{L}_k$ , we can make a detailed study of the “stable case”. For  $\alpha > 1$  (and in particular for a log-normal modulus, corresponding to  $\alpha = 2$ ), Propositions 3.1 and 3.5 directly apply. Therefore, let us suppose  $\alpha \leq 1$ . For  $\alpha = 1$ , one has the following:

**PROPOSITION 3.7.** — *Let  $\rho > 0$ , and let  $\mathcal{L}$  be the law of  $e^{i\Theta + \rho S_1}$ , where  $\Theta$  is a uniform random variable on  $[0, 2\pi)$  and  $S_1$  an independent standard*

symmetric Cauchy random variable. For  $\theta > 0$ , let  $(M_N)_{N \geq 1}$  be a sequence of random matrices such that  $M_N$  follows the distribution  $\mathbb{P}(N, \theta, \mathcal{L})$ . Then the distribution of the random probability measure  $\mu(M_N)/N$  converges to the law of the random measure

$$\bar{\mu}_\infty := \sum_{m \geq 1} x_m \mu_{e^{\rho S_1^{(m)}}}^U,$$

where  $(x_m)_{m \geq 1}$  is a Poisson-Dirichlet process with parameter  $\theta$ ,  $(S_1^{(m)})_{m \geq 1}$  is an independent sequence of i.i.d. standard symmetric Cauchy variables, and for  $R > 0$ ,  $\mu_R^U$  is the uniform measure on the circle of center zero and radius  $R$ . This convergence has to be understood as follows: for all continuous and bounded functions  $f$  from  $\mathbb{C}$  to  $\mathbb{R}$ ,

$$\frac{1}{N} \int_{\mathbb{C}} f d\mu(M_N) \xrightarrow{N \rightarrow \infty} \int_{\mathbb{C}} f d\bar{\mu}_\infty$$

in distribution.

*Proof.* — One can suppose that  $(M_N)_{N \geq 1}$  follows the distribution  $\mathbb{P}(\infty, \theta, \mathcal{L})$ . With the same notation as in the proof of Proposition 3.1, one sees that  $\mu(M_N)/N$  has the same distribution as the random measure:

$$\bar{\mu}_N := \sum_{m=1}^{\infty} \frac{l_{N,m}}{N} \left( \frac{\mathbb{1}_{l_{N,m} > 0}}{l_{N,m}} \sum_{\omega \in \mathbb{U}_{l_{N,m}}} \delta_{\omega L_m} \right),$$

where  $(L_m)_{m \geq 1}$  is a sequence of i.i.d. random variables with law  $\mathcal{L}$  (recall that  $\mathcal{L}_k$  is equal to  $\mathcal{L}$  for all  $k \geq 1$ ). If  $f$  is a continuous, bounded function from  $\mathbb{C}$  to  $\mathbb{R}$ , one has:

$$\int_{\mathbb{C}} f d\bar{\mu}_N = \sum_{m=1}^{\infty} \frac{l_{N,m}}{N} \left( \frac{\mathbb{1}_{l_{N,m} > 0}}{l_{N,m}} \sum_{\omega \in \mathbb{U}_{l_{N,m}}} f(\omega L_m) \right).$$

Now, there exists a GEM process  $(y_m)_{m \geq 1}$  of parameter  $\theta$  and a sequence of random variables  $(s_m)_{m \geq 1}$ , such that almost surely,

$$\sum_{m=1}^{\infty} s_m < \infty$$

$$\frac{l_{N,m}}{N} \xrightarrow{N \rightarrow \infty} y_m,$$

for all  $m \geq 1$  and

$$\frac{l_{N,m}}{N} \leq s_m$$

for all  $N, m \geq 1$ . By the convergence of Riemann sums corresponding to the integral of continuous functions on a compact set, one deduces that almost surely, for all  $m \geq 1$ :

$$\frac{l_{N,m}}{N} \left( \frac{\mathbb{1}_{l_{N,m} > 0}}{l_{N,m}} \sum_{\omega \in \mathbb{U}_{l_{N,m}}} f(\omega L_m) \right) \xrightarrow{N \rightarrow \infty} \frac{y_m}{2\pi} \int_0^{2\pi} f(L_m e^{i\lambda}) d\lambda,$$

where the left-hand side is smaller than or equal to  $s_m \|f\|_\infty$ , independently of  $N$ . By dominated convergence:

$$\int_{\mathbb{C}} f d\bar{\mu}_N \xrightarrow{N \rightarrow \infty} \frac{1}{2\pi} \sum_{m \geq 1} y_m \int_0^{2\pi} f(L_m e^{i\lambda}) d\lambda,$$

almost surely. This implies Proposition 3.7. □

*Remark 3.8.* — Almost surely, the random measure  $\bar{\mu}_\infty$  is strictly positive for any nonempty set of  $\mathbb{C}$ . Therefore, for all continuous functions  $f$  from  $\mathbb{C}$  to  $\mathbb{R}_+$ , non-identically zero, and for all  $A \in \mathbb{R}_+$ :

$$\mathbb{P} \left[ \int_{\mathbb{C}} f d\mu(M_N) \leq A \right] \xrightarrow{N \rightarrow \infty} 0.$$

Hence, one cannot expect an analog of Proposition 3.5 in the case studied here.

For the case  $\alpha < 1$ , one intuitively expects that most of the eigenvalues become very large or very small. The precise statement is the following:

**PROPOSITION 3.9.** — *Let  $\rho > 0$ , and let  $\mathcal{L}$  be the law of  $e^{i\Theta + \rho S_\alpha}$ , where  $\Theta$  is a uniform random variable on  $[0, 2\pi)$  and  $S_\alpha$  an independent standard symmetric stable random variable, with index  $\alpha < 1$ . For  $\theta > 0$ , let  $(M_N)_{N \geq 1}$  be a sequence of random matrices such that  $M_N$  follows the distribution  $\mathbb{P}(N, \theta, \mathcal{L})$ . Then, the distribution of the random probability  $\mu(M_N)/N$  converges to the law of the random measure  $G_\theta \delta_0$ , where  $G_\theta$  is a beta random variable with parameters  $(\theta/2, \theta/2)$ , in the following sense: for all continuous functions  $f$  from  $\mathbb{C}$  to  $\mathbb{R}$ , with compact support,*

$$\frac{1}{N} \int_{\mathbb{C}} f d\mu(M_N) \xrightarrow{N \rightarrow \infty} G_\theta f(0)$$

*in distribution.*

*Remark 3.10.* — The total mass of the limit measure is a.s. strictly smaller than one. Intuitively, this is due to the fact that a large part of the total mass of  $\mu(M_N)$  is going far away from zero, when  $N$  is large. This mass is missing in the limiting measure in Proposition 3.9, because we consider functions  $f$  with compact support.

*Proof.* — We suppose that  $(M_N)_{N \geq 1}$  follows the distribution  $\mathbb{P}(\infty, \theta, \mathcal{L})$ . Let  $f$  be a continuous function from  $\mathbb{C}$  to  $\mathbb{R}$ , with compact support, and let us choose  $R > 1$  such that  $f(z) = 0$  for all  $z$  such that  $|z| > R$ . Let us define, for all  $z \in \mathbb{C}$ ,

$$g(z) := \mathbb{1}_{|z| \leq 1} f(0).$$

and for  $r \in (0, 1)$ :

$$\beta(r) := \sup\{|f(z) - f(0)|, |z| \leq r\},$$

which tends to zero with  $r$ . One checks that

$$\frac{1}{N} \int_{\mathbb{C}} |f - g| d\mu(M_N) \leq \beta(r) + 2\|f\|_{\infty} \frac{\mu(M_N)(\{z \in \mathbb{C}, r \leq |z| \leq R\})}{N},$$

which implies:

$$\begin{aligned} \mathbb{P}\left[\frac{1}{N} \int_{\mathbb{C}} |f - g| d\mu(M_N) > \beta(r) + (2\|f\|_{\infty} + 1)r\right] \\ \leq \mathbb{P}\left[\frac{\mu(M_N)(\{z \in \mathbb{C}, r \leq |z| \leq R\})}{N} \geq r\right] \\ \leq \frac{1}{r} \mathbb{E}\left[\frac{\mu(M_N)(\{z \in \mathbb{C}, r \leq |z| \leq R\})}{N}\right]. \end{aligned}$$

Now,

$$\begin{aligned} \mathbb{E}\left[\frac{\mu(M_N)(\{z \in \mathbb{C}, r \leq |z| \leq R\})}{N}\right] \\ = \sum_{m=1}^{\infty} \mathbb{E}\left[\frac{l_{N,m}}{N} \mathbb{1}_{l_{N,m} > 0, |Z_{N,m}|^{(1/l_{N,m})} \in (r, R)}\right]. \end{aligned}$$

From the independence of  $l_{N,m}$  and  $(z_j)_{j \geq 1}$ , and from the basic properties of stable random variables, one has

$$\mathbb{P}\left[l_{N,m} > 0, |Z_{N,m}|^{(1/l_{N,m})} \in (r, R) \mid l_{N,m}\right] = \Psi(l_{N,m}),$$

where the function  $\Psi$  (which can depend on  $\alpha, \rho, r$  and  $R$ , but not on  $N$ ) is bounded by one and tends to zero at infinity. Hence:

$$\mathbb{E}\left[\frac{\mu(M_N)(\{z \in \mathbb{C}, r \leq |z| \leq R\})}{N}\right] \leq \mathbb{E}\left[\sum_{m=1}^{\infty} \frac{l_{N,m} \Psi(l_{N,m})}{N}\right].$$

Now, for all  $m \geq 1$ , a.s.:

$$\frac{l_{N,m} \Psi(l_{N,m})}{N} \xrightarrow[N \rightarrow \infty]{} 0$$

and

$$\frac{l_{N,m} \Psi(l_{N,m})}{N} \leq s_m.$$

By dominated convergence,

$$\mathbb{E} \left[ \frac{\mu(M_N)(\{z \in \mathbb{C}, r \leq |z| \leq R\})}{N} \right] \xrightarrow{N \rightarrow \infty} 0$$

which implies:

$$\mathbb{P} \left[ \frac{1}{N} \int_{\mathbb{C}} |f - g| d\mu(M_N) > \beta(r) + (2\|f\|_{\infty} + 1)r \right] \xrightarrow{N \rightarrow \infty} 0.$$

By letting  $r$  go to zero, one deduces that

$$\frac{1}{N} \int_{\mathbb{C}} |f - g| d\mu(M_N) \xrightarrow{N \rightarrow \infty} 0$$

in probability. Therefore, it is sufficient to prove the conclusion of Proposition 3.9, with  $f$  replaced by  $g$ . Moreover, one can suppose  $f(0) = g(0) = 1$ .

In this case, one has:

$$\frac{1}{N} \int_{\mathbb{C}} g d\mu(M_N) = \sum_{m \geq 1} \frac{l_{N,m}}{N} \mathbb{1}_{l_{N,m} > 0, |Z_{N,m}| \leq 1}.$$

Now, by symmetry of the stable variables considered here and by the independence of  $(l_{N,m})_{m \geq 1}$  and  $(z_j)_{j \geq 1}$ ,

$$\frac{1}{N} \int_{\mathbb{C}} g d\mu(M_N) = \sum_{m \geq 1} \frac{l_{N,m}}{N} \epsilon_m$$

in distribution, where  $(\epsilon_m)_{m \geq 1}$  is a sequence of i.i.d Bernoulli variables of parameter  $1/2$ , independent of  $(l_{N,m})_{N,m \geq 1}$ . The parameter  $1/2$  is explained as follows:  $\epsilon_m = \mathbb{1}_{|Z_{N,m}| \leq 1}$ , where  $\log(|Z_{N,m}|)$  is a symmetric, stable random variable, and then

$$\mathbb{P}[\epsilon_m = 1] = \mathbb{P}[\log(|Z_{N,m}|) \leq 0] = \mathbb{P}[\log(|Z_{N,m}|) \geq 0] = 1/2.$$

Now, by dominated convergence (recall that the sum of the variables  $s_m$  is a.s. finite),

$$\sum_{m \geq 1} \frac{l_{N,m}}{N} \epsilon_m \xrightarrow{N \rightarrow \infty} \sum_{m \geq 1} y_m \epsilon_m,$$

which implies Proposition 3.9 if we check that

$$X := \sum_{m \geq 1} y_m \epsilon_m$$

is equal to  $G_{\theta}$  in distribution (recall that  $(y_m)_{m \geq 1}$  is a  $\text{GEM}(\theta)$  process, independent of the sequence  $(\epsilon_m)_{m \geq 1}$ ). This fact can be proved as follows: by self-similarity of the  $\text{GEM}(\theta)$  process,

$$X = VX + \epsilon(1 - V),$$

in distribution, where  $V$ ,  $X$ ,  $\epsilon$  are independent,  $V$  is a beta variable of parameters  $\theta$  and 1, and  $\epsilon$  is a Bernoulli variable of parameter  $1/2$ . With this identity, one can compute the moments of  $X$  by induction, and finally, one can identify its law.  $\square$

Again for the stable case with parameter  $\alpha < 1$ , one has an analog of Proposition 3.5, for the eigenvalues which are in a compact set not containing zero. The precise statement is:

PROPOSITION 3.11. — *Let  $\rho > 0$ , and let  $\mathcal{L}$  be the law of  $e^{i\Theta + \rho S_\alpha}$ , where  $\Theta$  is a uniform random variable on  $[0, 2\pi)$  and  $S_\alpha$  an independent standard symmetric stable random variable, with index  $\alpha < 1$ . For  $\theta > 0$ , let  $(M_N)_{N \geq 1}$  be a sequence of random matrices such that  $M_N$  follows the distribution  $\mathbb{P}(N, \theta, \mathcal{L})$ . Then, for all continuous functions  $f$  from  $\mathbb{C}$  to  $\mathbb{R}$ , with compact support, and such that  $f = 0$  on a neighborhood of zero,  $f$  is a.s. integrable with respect to  $\mu_\infty$ , where the random measure  $\mu_\infty$  is given by (3.6). Moreover, the following convergence in distribution holds with such test functions:*

$$\int_{\mathbb{C}} f d\mu(M_N) \xrightarrow{N \rightarrow \infty} \int_{\mathbb{C}} f d\mu_\infty.$$

*Proof.* — The proof is similar to the proof of Proposition 3.5. One only has to change the estimate:

$$\mathbb{P}[|T_{k,1}|^{1/k} \notin (R_1, R_2)] \leq Ck^{-q}$$

(for some  $C, q > 0$ ), by the estimate:

$$\mathbb{P}[|T_{k,1}|^{1/k} \in (r, R)] \leq Ck^{-q},$$

available for all  $r, R$  such that  $R > r > 0$ , by the classical properties of symmetric stable random variables.  $\square$

### 3.2. The average eigenvalues distributions

Another interesting problem is the study of the expectation  $\tilde{\mu}_N$  of the random measure  $\mu(M_N)$ , where  $M_N$  follows the distribution  $\mathbb{P}(N, \theta, \mathcal{L})$ . The measure  $\tilde{\mu}_N$  of a Borel set in  $\mathbb{C}$  represents the expected number of eigenvalues of  $M_N$  (with multiplicity) contained in this set, and it can be explicitly computed:

PROPOSITION 3.12. — Let  $N \geq 1$ ,  $\theta > 0$ ,  $\mathcal{L}$  a probability measure on  $\mathbb{C}^*$ . If  $M_N$  follows the distribution  $\mathbb{P}(N, \theta, \mathcal{L})$ , then the expectation  $\tilde{\mu}_N$  of  $\mu(M_N)$  can be represented as follows:

$$\tilde{\mu}_N = \theta \sum_{k=1}^N \frac{N(N-1)\cdots(N-k+1)}{(N-1+\theta)(N-2+\theta)\cdots(N-k+\theta)} \mathcal{L}_k.$$

*Proof.* — Let  $f$  be a Borel function from  $\mathbb{C}$  to  $\mathbb{R}_+$ . By taking the same notation as in (3.5), one has:

$$\int_{\mathbb{C}} f d\tilde{\mu}_N = \mathbb{E} \left[ \sum_{k=1}^{\infty} \sum_{p=1}^{a_{N,k}} \sum_{\omega^k = T_{k,p}} f(\omega) \right]$$

Now, for all  $k, p \geq 1$ :

$$\mathbb{E} \left[ \sum_{\omega^k = T_{k,p}} f(\omega) \right] = k \int_{\mathbb{C}} f d\mathcal{L}_k$$

Moreover,  $(T_{k,p})_{k,p \geq 1}$  is independent of  $(a_{N,k})_{k \geq 1}$ . One deduces that

$$\mathbb{E} \left[ \sum_{k=1}^{\infty} \sum_{p=1}^{a_{N,k}} \sum_{\omega^k = T_{k,p}} f(\omega) \mid (a_{N,k})_{k \geq 1} \right] = \sum_{k=1}^{\infty} k a_{N,k} \int_{\mathbb{C}} f d\mathcal{L}_k.$$

Therefore

$$\tilde{\mu}_N = \sum_{k=1}^{\infty} k \mathbb{E}[a_{N,k}] \mathcal{L}_k.$$

By doing explicit computations of  $\mathbb{E}[a_{N,k}]$  (see e.g. [2]), one deduces Proposition 3.12. □

*Remark 3.13.* — When the distribution of the  $z_j$ 's is the Dirac mass at 1, then  $\mathcal{L}_k$  is the uniform measure on the  $k$ -th roots of unity.

One has a similar result for the limiting random measure  $\mu_{\infty}$  of  $\mu(M_N)$ :

PROPOSITION 3.14. — Let  $\theta > 0$ ,  $\mathcal{L}$  a probability measure on  $\mathbb{C}^*$ ,  $(M_N)_{N \geq 1}$  a sequence of random matrices such that  $M_N$  follows the distribution  $\mathbb{P}(N, \theta, \mathcal{L})$ . Then the expectation of the random measure  $\mu_{\infty}$ , given by (3.6) (it can be considered, in some sense, as the limit of  $\mu(M_N)$  for  $N \rightarrow \infty$ ), is the measure:

$$\tilde{\mu}_{\infty} = \theta \sum_{k=1}^{\infty} \mathcal{L}_k.$$

*Proof.* — The proof of Proposition 3.14 is exactly similar to the proof of Proposition 3.12. □



Now, we have a sequence of finite measures  $\tilde{\mu}_N$ , defined as the expectation of  $\mu(M_N)$  and explicitly described, and an infinite measure  $\tilde{\mu}_\infty$ , defined as the expectation of  $\mu_\infty$ . Moreover, we know that, for a large class of probability laws  $\mathcal{L}$ , the random probability measure  $\mu(M_N)/N$  converges weakly in probability to the uniform measure on a circle, and in a sense which can be made precise,  $\mu(M_N)$  tends to  $\mu_\infty$ . Hence, we can expect analog convergences for the sequence of measures  $(\tilde{\mu}_N)_{N \geq 1}$ . One indeed has the following result:

**PROPOSITION 3.15.** — *Let  $\theta > 0$ ,  $\mathcal{L}$  a probability measure on  $\mathbb{C}^*$ ,  $(M_N)_{N \geq 1}$  a sequence of random matrices such that  $M_N$  follows the distribution  $\mathbb{P}(N, \theta, \mathcal{L})$ . We suppose that if  $Z$  follows the distribution  $\mathcal{L}$ , then  $\log(|Z|)$  is integrable, and we define  $R := \exp(\mathbb{E}[\log(|Z|)])$ . Then the probability measure  $\tilde{\mu}_N/N$  (which represents the probability distribution of a random eigenvalue of  $M_N$ , chosen uniformly among the  $N$  possible eigenvalues), converges weakly to the uniform distribution on the circle  $\{z = R\}$ . Moreover, if  $\log(|Z|)$  is in  $\mathcal{U}^2$ , then for  $0 < R_1 < R < R_2$ , the restriction of  $\tilde{\mu}_N$  to the set  $\{|z| \notin (R_1, R_2)\}$  converges weakly to the corresponding restriction of  $\tilde{\mu}_\infty$ , which is a finite measure.*

*Proof.* — Let  $f$  be a continuous and bounded function from  $\mathbb{C}$  to  $\mathbb{R}$ . By Proposition 3.1 (more precisely, by Corollary 3.2):

$$\frac{1}{N} \int_{\mathbb{C}} f d\mu(M_N) \xrightarrow{N \rightarrow \infty} \frac{1}{2\pi} \int_0^{2\pi} f(Re^{i\lambda}) d\lambda$$

in probability. Since  $f$  is uniformly bounded, one deduces the first part of Proposition 3.15 by taking the expectation. Let us now prove the second part. We now suppose that  $f$  is a bounded and continuous function from  $\mathbb{C}$  to  $\mathbb{R}$ , equal to zero in a neighborhood of the circle  $\{|z| = R\}$ . By Proposition 3.12, it is sufficient to prove that

$$\sum_{k=1}^{\infty} |t_{N,k,\theta} \mathbb{1}_{N \geq k} - 1| \int_{\mathbb{C}} |f| d\mathcal{L}_k \xrightarrow{N \rightarrow \infty} 0,$$

where

$$t_{N,k,\theta} = \frac{N(N-1) \cdots (N-k+1)}{(N-1+\theta)(N-2+\theta) \cdots (N-k+\theta)}.$$

Each term of the sum converges to zero when  $N$  goes to infinity. Hence by dominated convergence we are done if we prove that:

$$(3.10) \quad \sum_{k=1}^{\infty} v_{k,\theta} \int_{\mathbb{C}} |f| d\mathcal{L}_k < \infty,$$

where

$$v_{k,\theta} = 1 + \sup\{t_{N,k,\theta}, N \geq k\}.$$

Now, for  $\theta \geq 1$ ,  $v_{k,\theta} = 2$ , and for  $\theta < 1$ :

$$v_{k,\theta} = 1 + \frac{k!(\theta - 1)!}{(k - 1 + \theta)!},$$

which implies that  $v_{k,\theta}$  is dominated by  $k^{(1-\theta)_+}$  for fixed  $\theta$ . On the other hand, the fact that  $\log(|Z|)$  is in  $\mathcal{U}^2$  implies that the integral of  $|f|$  with respect to  $\mathcal{L}_k$  decreases with  $k$  at least as fast as  $1/k^2$ . This implies (3.10). □

As above, it is interesting to see what happens if  $\mathcal{L}$  is the distribution of  $e^{i\Theta + \rho S_\alpha}$ , where  $\rho > 0$ ,  $\Theta$  is a uniform random variable on  $[0, 2\pi)$  and  $S_\alpha$  is an independent standard symmetric stable random variable, with index  $\alpha \in (0, 2]$ . For the log-normal case  $\alpha = 2$ , Proposition 3.15 applies directly. For the case  $\alpha \in (1, 2)$ , one can apply the first part of Proposition 3.15, which gives the convergence of the average empirical measure  $\tilde{\mu}_N/N$ , but the second part cannot apply. Indeed, by using the classical tail estimates of stable random variables one checks that for all nonempty open sets  $A \subset \mathbb{C}$ :

$$\tilde{\mu}_\infty(A) = \theta \sum_{k=1}^\infty \mathcal{L}_k(A) = \infty.$$

For  $\alpha = 1$ , one has  $\mathcal{L}_k = \mathcal{L}$  for all  $k \geq 1$ , and then:

$$\tilde{\mu}_N = N\mathcal{L}$$

for all  $N \geq 1$ . The most interesting case is  $\alpha < 1$ .

**PROPOSITION 3.16.** — *Let  $\rho > 0$ , and let  $\mathcal{L}$  be the law of  $e^{i\Theta + \rho S_\alpha}$ , where  $\Theta$  is a uniform random variable on  $[0, 2\pi)$  and  $S_\alpha$  an independent standard symmetric stable random variable, with index  $\alpha < 1$ . For  $\theta > 0$ , let  $(M_N)_{N \geq 1}$  be a sequence of random matrices such that  $M_N$  follows the distribution  $\mathbb{P}(N, \theta, \mathcal{L})$ . Then the probability measure  $\tilde{\mu}_N/N$  (which represents the probability distribution of a random eigenvalue of  $M_N$ , chosen uniformly among the  $N$  possible eigenvalues), converges vaguely to half of the Dirac measure at zero. Moreover, for all  $r, R$  such that  $0 < r < R$ , the restriction of  $\tilde{\mu}_\infty$  to the set  $\{|z| \in (r, R)\}$  is infinite if  $\alpha \geq 1/2$ , finite if  $\alpha < 1/2$ , and for  $\alpha < 1/(2 \vee (3 - \theta))$  (in particular for  $\alpha \leq 1/3$ ), it is the weak limit of the corresponding restriction of  $\tilde{\mu}_N$  when  $N$  goes to infinity.*

*Proof.* — Let  $f$  be a continuous function from  $\mathbb{C}$  to  $\mathbb{R}$ , with compact support. By Proposition 3.9, one has the convergence in distribution:

$$\frac{1}{N} \int_{\mathbb{C}} f d\mu(M_N) \xrightarrow{N \rightarrow \infty} G_{\theta} f(0).$$

Since  $f$  is uniformly bounded, one can take the expectation and obtain:

$$\frac{1}{N} \int_{\mathbb{C}} f d\tilde{\mu}_N \xrightarrow{N \rightarrow \infty} \frac{1}{2} f(0),$$

which gives the first part of Proposition 3.16. Now, one checks that for all  $r, R$  such that  $0 < r < R$ ,  $\mathcal{L}_k(\{z \in \mathbb{C}, |z| \in (r, R)\})$  decreases as  $k^{1-1/\alpha}$  when  $k$  goes to infinity, which gives the condition for the finiteness of  $\tilde{\mu}_{\infty}$  restricted to  $\{z \in \mathbb{C}, |z| \in (r, R)\}$ . The proof of the convergence of the restriction of  $\tilde{\mu}_N$  to this set toward the restriction of  $\tilde{\mu}_{\infty}$  is similar to the proof of the second part of Proposition 3.15. One only needs to check that:

$$\sum_{k=1}^{\infty} v_{k,\theta} \mathcal{L}_k(\{z \in \mathbb{C}, |z| \in (r, R)\}) < \infty,$$

which is true, since  $v_{k,\theta}$  is dominated by  $k^{(1-\theta)_+}$  and  $\mathcal{L}_k(\{z \in \mathbb{C}, |z| \in (r, R)\})$  is dominated by  $k^{-1/\alpha}$ , for  $1/\alpha > 2 \vee (3 - \theta)$ . □

### 3.3. The $q$ -correlation

The measures  $\tilde{\mu}_N$  and  $\tilde{\mu}_{\infty}$  give the average repartition of the eigenvalues (and its limit for large dimension) of a matrix following the distribution  $\mathbb{P}(N, \theta, \mathcal{L})$ . It is interesting to generalize this study to the  $q$ -correlation of eigenvalues for all strictly positive integers  $q$ , i.e. the distribution of the possible sequences of  $q$  eigenvalues. More precisely, if  $M_N$  follows the distribution  $\mathbb{P}(N, \theta, \mathcal{L})$ , let us consider, for all  $q \geq 1$ , the random measure on  $\mathbb{C}^q$ :

$$\mu^{[q]}(M_N) = \sum_{j_1 \neq j_2 \neq \dots \neq j_q} \delta_{(\omega_{j_1}, \dots, \omega_{j_q})},$$

where  $(\omega_j)_{1 \leq j \leq N}$  is a sequence containing all the eigenvalues of  $M_N$ , with multiplicity. The “ $q$ -correlation measure” is defined as the average of  $\mu^{[q]}(M_N)$ : it is a finite measure  $\tilde{\mu}_N^{[q]}$  on  $\mathbb{C}^q$ , with total mass  $N!/(N - q)!$ . One checks that  $\mu^{[q]}(M_N)$  has (by taking the same notation as in (3.5)) the same distribution as the random measure

$$\nu_N := \sum_{(w_{k,p})_{k \geq 1, 1 \leq p \leq a_{N,k}} \in W(q, (a_{N,k})_{k \geq 1})} \sum_{(E_{k,p} \in S_{k, w_{k,p}}(T_{k,p}))_{k \geq 1, 1 \leq p \leq a_{N,k}}} \Delta((E_{k,p})_{k \geq 1, 1 \leq p \leq a_{N,k}}),$$

where  $W(q, (a_{N,k})_{k \geq 1})$  is the set of families  $(w_{k,p})_{k \geq 1, 1 \leq p \leq a_{N,k}}$  of nonnegative integers, with total sum  $q$ ,  $S_{k,w_{k,p}}(T_{k,p})$  is the family of subsets of  $k$ -th roots of  $T_{k,p}$ , with cardinality  $w_{k,p}$ , and  $\Delta((E_{k,p})_{k \geq 1, 1 \leq p \leq a_{N,k}})$  is defined by:

$$\Delta((E_{k,p})_{k \geq 1, 1 \leq p \leq a_{N,k}}) = \sum_{\sigma \in \Sigma_q} \delta_{(\omega_{\sigma(j)})_{1 \leq j \leq q}}$$

for a sequence  $(\omega_j)_{1 \leq j \leq q}$  containing all the elements of the sets  $E_{k,p}$ , each element appearing a number of times equal to the number of sets  $E_{k,p}$  containing it. Here,  $w_{k,p}$  represents the number of eigenvalues  $\omega_{j_r}$  ( $1 \leq r \leq q$ ) which come from the  $p$ -th  $k$ -cycle of the permutation, and  $E_{k,p}$  represents the set of these eigenvalues. The sum for  $\sigma \in \Sigma_q$  in the definition of  $\Delta$  is due to the fact that in the definition of  $\mu^{[q]}(M_N)$ , the same  $q$  eigenvalues are put into all the  $q!$  possible orders. Let us now take the conditional expectation with respect to  $(a_{N,k})_{k \geq 1}$ . One obtains the following random measure:

$$\mathbb{E}[\nu_N \mid (a_{N,k})_{k \geq 1}] = \sum_{(w_{k,p})_{k \geq 1, 1 \leq p \leq a_{N,k}} \in W(q, (a_{N,k})_{k \geq 1})} \mathcal{L}^{[q]}((w_{k,p})_{k \geq 1, 1 \leq p \leq a_{N,k}}),$$

for

$$\begin{aligned} &\mathcal{L}^{[q]}((w_{k,p})_{k \geq 1, 1 \leq p \leq a_{N,k}}) \\ &= \frac{1}{\prod_{k \geq 1, 1 \leq p \leq a_{N,k}} w_{k,p}!} \sum_{\sigma \in \Sigma_q} \sigma \cdot \tilde{\mathcal{L}}^{[q]}((w_{k,p})_{k \geq 1, 1 \leq p \leq a_{N,k}}), \end{aligned}$$

where

$$\sigma \cdot \tilde{\mathcal{L}}^{[q]}((w_{k,p})_{k \geq 1, 1 \leq p \leq a_{N,k}})$$

is the image, by the permutation  $\sigma$  of the coordinates, of

$$\tilde{\mathcal{L}}^{[q]}((w_{k,p})_{k \geq 1, 1 \leq p \leq a_{N,k}}) = \bigotimes_{k \geq 1, 1 \leq p \leq a_{N,k}} \mathcal{L}_k^{[w_{k,p}]},$$

for

$$\mathcal{L}_k^{[w_{k,p}]} = \sum_{z_1 \neq \dots \neq z_{w_{k,p}} \in \mathbb{U}_k} (z_1, z_2, \dots, z_{w_{k,p}}) \cdot \mathcal{L}_k,$$

where  $(z_1, z_2, \dots, z_{w_{k,p}}) \cdot \mathcal{L}_k$  is the image of  $\mathcal{L}_k$  by the application  $z \mapsto (zz_1, zz_2, \dots, zz_{w_{k,p}})$ . More intuitively, the measure  $\mathcal{L}_k^{[r]}$  is obtained by taking all the possible  $r$ -uples of  $k$ -th roots of a complex number following the distribution  $\mathcal{L}$ . The measure  $\tilde{\mathcal{L}}^{[q]}$  is obtained from the measures  $\mathcal{L}_k^{[w_{k,p}]}$  via a concatenation of the  $w_{k,p}$ -uples, each of them corresponding to a cycle of the random permutation involved in  $M_N$ . One then gets  $\mathcal{L}^{[q]}$  by taking

into account all the permutations of the  $q$  eigenvalues. Since the permutations of the eigenvalues associated to the  $p$ -th  $k$ -cycle were already taken into account in the definition of  $\mathcal{L}_k^{[w_{k,p}]}$ , one needs to compensate this by putting, for each cycle, a factor  $w_{k,p}!$  at the denominator. By defining  $\lambda_{k,r}$  as the number of indices  $p$  such that  $w_{k,p} = r$ , one then gets:

$$\mathbb{E}[\nu_N \mid (a_{N,k})_{k \geq 1}] = \sum_{(\lambda_{k,r})_{k \geq 1, r \geq 0} \in L(q, (a_{N,k})_{k \geq 1})} \frac{\prod_{k \geq 1} a_{N,k}!}{\prod_{k \geq 1, r \geq 0} [\lambda_{k,r}!(r!)^{\lambda_{k,r}}]} \cdots \left[ \sum_{\sigma \in \Sigma_q} \sigma \cdot \left[ \bigotimes_{k \geq 1, r \geq 0} (\mathcal{L}_k^{[r]})^{\otimes \lambda_{k,r}} \right] \right]$$

where  $L(q, (a_{N,k})_{k \geq 1})$  is the set of families  $(\lambda_{k,r})_{k \geq 1, r \geq 0}$  of nonnegative integers such that

$$\sum_{k \geq 1, r \geq 0} r \lambda_{k,r} = q$$

and for all  $k \geq 1$ :

$$\sum_{r \geq 0} \lambda_{k,r} = a_{N,k}.$$

Here, we have regrouped all the identical terms corresponding to the same values of  $(\lambda_{k,r})_{k \geq 1, r \geq 0}$ : the number of such identical terms is equal to the number of choices of  $(w_{k,p})_{k \geq 1, 1 \leq p \leq a_{N,k}}$  giving the same family  $(\lambda_{k,r})_{k \geq 1, r \geq 0}$ , i.e.

$$\frac{\prod_{k \geq 1} a_{N,k}!}{\prod_{k \geq 1, r \geq 0} \lambda_{k,r}!}.$$

Note that for all  $k \geq 1$ , the measure  $\mathcal{L}_k^{[0]}$  is trivial (it is the unique probability measure on a space with one element), hence, one can remove it in a tensor product. One deduces:

$$\mathbb{E}[\nu_N \mid (a_{N,k})_{k \geq 1}] = \sum_{(\lambda_{k,r})_{k, r \geq 1} \in \tilde{L}(q)} \prod_{k \geq 1} \cdots \frac{a_{N,k}!}{(a_{N,k} - \sum_{r \geq 1} \lambda_{k,r})! \prod_{r \geq 1} [\lambda_{k,r}!(r!)^{\lambda_{k,r}}]} \sum_{\sigma \in \Sigma_q} \sigma \cdot \left[ \bigotimes_{k \geq 1, r \geq 1} (\mathcal{L}_k^{[r]})^{\otimes \lambda_{k,r}} \right]$$

where  $\tilde{L}(q)$  is the set of families  $(\lambda_{k,r})_{k, r \geq 1}$  of nonnegative integers such that

$$\sum_{k \geq 1, r \geq 1} r \lambda_{k,r} = q$$

and where the inverse of the factorial of a strictly negative integer is considered to be equal to zero. Then, by removing the conditioning with respect to  $(a_{N,k})_{k \geq 1}$ , one gets:

$$\tilde{\mu}_N^{[q]} = \sum_{(\lambda_{k,r})_{k,r \geq 1} \in \tilde{L}(q)} u_{N,\theta}^{[\sum_{r \geq 1} \lambda_{k,r}]_{k \geq 1}} \frac{1}{\prod_{k,r \geq 1} [\lambda_{k,r}!(r!)^{\lambda_{k,r}}]} \cdots \sum_{\sigma \in \Sigma_q} \sigma \cdot \left[ \bigotimes_{k \geq 1, r \geq 1} (\mathcal{L}_k^{[r]})^{\otimes \lambda_{k,r}} \right],$$

where, for all sequences of nonnegative integers  $(\lambda_k)_{k \geq 1}$  for which the set of  $k$  such that  $\lambda_k > 0$  is finite:

$$u_{N,\theta}^{[\lambda_k]_{k \geq 1}} = \mathbb{E} \left[ \prod_{k \geq 1} \frac{a_{N,k}!}{(a_{N,k} - \lambda_k)!} \right]$$

(recall that this quantity depends on  $\theta$ , as the law of  $(a_{N,k})_{k \geq 1}$ ). By elementary combinatorial arguments, one can prove that

$$u_{N,\theta}^{[\lambda_k]_{k \geq 1}} = \frac{\prod_{j=1}^{\sum_{k \geq 0} \lambda_k} (N + 1 - j)}{\prod_{k \geq 1} k^{\lambda_k}} \cdot \mathcal{P},$$

where  $\mathcal{P}$  is the probability that in a random permutation following the Ewens( $\theta$ ) distribution, the integers from  $1 + \sum_{j < k} \lambda_j$  and  $\sum_{j \leq k} \lambda_j$  lie in different  $k$ -cycles, for all  $k \geq 1$ . By the Feller coupling, one deduces that

$$u_{N,\theta}^{[\lambda_k]_{k \geq 1}} = \mathbf{1}_{\sum_{k \geq 1} k \lambda_k \leq N} \left( \prod_{j=1}^{\sum_{k \geq 1} k \lambda_k} \frac{N + 1 - j}{N + \theta - j} \right) \frac{\theta^{\sum_{k \geq 1} \lambda_k}}{\prod_{k \geq 1} k^{\lambda_k}}.$$

Then by denoting  $\tilde{L}(N, q)$  the set of families  $(\lambda_{k,r})_{k,r \geq 1}$  of nonnegative integers such that

$$\sum_{k \geq 1, r \geq 1} r \lambda_{k,r} = q$$

and

$$\sum_{k \geq 1, r \geq 1} k \lambda_{k,r} \leq N,$$

one obtains the following result:

PROPOSITION 3.17. — *Let  $\theta > 0$  and let  $\mathcal{L}$  be a probability distribution on  $\mathbb{C}^*$ . If one takes the notation above, the  $q$ -correlation measure  $\tilde{\mu}_N^{[q]}$  associated with the eigenvalues of a random matrix following the distribution*

$\mathbb{P}(N, \theta, \mathcal{L})$  is given by the formula:

$$\begin{aligned} \tilde{\mu}_N^{[q]} = & \sum_{(\lambda_{k,r})_{k,r \geq 1} \in \tilde{L}(N,q)} \left( \prod_{j=1}^{\sum_{k,r \geq 1} k \lambda_{k,r}} \frac{N+1-j}{N+\theta-j} \right) \cdots \\ & \cdots \prod_{k,r \geq 1} \left[ \frac{1}{\lambda_{k,r}!} \left( \frac{\theta}{r!k} \right)^{\lambda_{k,r}} \right] \sum_{\sigma \in \Sigma_q} \sigma \cdot \left[ \bigotimes_{k \geq 1, r \geq 1} (\mathcal{L}_k^{[r]})^{\otimes \lambda_{k,r}} \right] \end{aligned}$$

*Remark 3.18.* — One can check that Proposition 3.12 is a particular case of Proposition 3.17 for  $q = 1$ . Note that for  $N, q \geq 2$ , the  $q$ -correlation measure  $\tilde{\mu}_N^{[q]}$  is not absolutely continuous with respect to Lebesgue measure. The situation is different from what one generally gets by studying classical ensembles of random matrices. For example, in many ensembles of hermitian and unitary matrices, the point process of the eigenvalues is determinantal, which means that the correlation measure is absolutely continuous with respect to the Lebesgue measure, with density at  $(x_1, \dots, x_q)$  given by  $\det(K(x_m, x_n)_{1 \leq m, n \leq q})$ , where  $K$  is a certain function from  $\mathbb{U}^2$  (for unitary matrices) or  $\mathbb{R}^2$  (for hermitian matrices) to  $\mathbb{R}$ .

In a similar way, we can define the  $q$ -correlation measure  $\tilde{\mu}_\infty^{[q]}$ , associated with the random point measure  $\mu_\infty$  defined by (3.6). We obtain the following:

**PROPOSITION 3.19.** — *Let  $\theta > 0$  and  $\mathcal{L}$  be a probability distribution on  $\mathbb{C}^*$ . Then, the  $q$ -correlation measure  $\tilde{\mu}_\infty^{[q]}$  associated with the point measure  $\mu_\infty$  defined by (3.6) is given by:*

$$\begin{aligned} \tilde{\mu}_\infty^{[q]} = & \sum_{(\lambda_{k,r})_{k,r \geq 1} \in \tilde{L}(q)} \prod_{k,r \geq 1} \left[ \frac{1}{\lambda_{k,r}!} \left( \frac{\theta}{r!k} \right)^{\lambda_{k,r}} \right] \cdots \\ & \cdots \sum_{\sigma \in \Sigma_q} \sigma \cdot \left[ \bigotimes_{k \geq 1, r \geq 1} (\mathcal{L}_k^{[r]})^{\otimes \lambda_{k,r}} \right]. \end{aligned}$$

Since we have an explicit expression for the correlations measures, we can expect some limit theorems when the dimension  $N$  goes to infinity. In fact, we have the following proposition:

**PROPOSITION 3.20.** — *Let  $\theta > 0, q \geq 1$  integer,  $\mathcal{L}$  a probability measure on  $\mathbb{C}^*$ ,  $(M_N)_{N \geq 1}$  a sequence of random matrices such that  $M_N$  follows the distribution  $\mathbb{P}(N, \theta, \mathcal{L})$ . We suppose that if  $Z$  follows the distribution  $\mathcal{L}$ , then  $\log(|Z|)$  is integrable, and we define  $R := \exp(\mathbb{E}[\log(|Z|)])$ . Then the*

probability measure  $\frac{1}{N(N-1)\cdots(N-q+1)}\tilde{\mu}_N^{[q]}$  (which represents the probability distribution of a random sequence of  $q$  eigenvalues of  $M_N$ , chosen uniformly among the  $N(N-1)\cdots(N-q+1)$  possible sequences), converges weakly to the  $q$ -th power of the uniform distribution on the circle  $\{z = R\}$ . Moreover, if  $\log(|Z|)$  is in  $\mathcal{U}^{q+1}$ , then for  $0 < R_1 < R < R_2$ , the restriction of  $\tilde{\mu}_N^{[q]}$  to the set  $\{(z_1, \dots, z_q) \in \mathbb{C}^q, \forall r \leq q, |z_r| \notin (R_1, R_2)\}$ , converges weakly to the corresponding restriction of  $\tilde{\mu}_\infty^{[q]}$ .

*Proof.* — Recall that  $\tilde{\mu}_N^{[q]}$  is the average of the random measure:

$$\mu^{[q]}(M_N) = \sum_{j_1 \neq j_2 \neq \dots \neq j_q} \delta_{(\omega_{j_1}, \dots, \omega_{j_q})},$$

where  $(\omega_j)_{1 \leq j \leq N}$  is a sequence containing all the eigenvalues of  $M_N$ , with multiplicity. Let  $f_1, \dots, f_q$  be bounded, continuous functions from  $\mathbb{C}$  to  $\mathbb{R}_+$ . One has:

$$\begin{aligned} I &:= \int_{\mathbb{C}^q} f_1(z_1) \cdots f_q(z_q) d\tilde{\mu}_N^{[q]}(z_1, \dots, z_q) \\ &= \mathbb{E} \left[ \sum_{j_1 \neq j_2 \neq \dots \neq j_q} f_1(\omega_{j_1}) \cdots f_q(\omega_{j_q}) \right], \end{aligned}$$

which implies,

$$(3.11) \quad J - \|f_1\|_\infty \cdots \|f_q\|_\infty [N^q - N(N-1)\cdots(N-q+1)] \leq I \leq J,$$

for

$$J := \mathbb{E} \left[ \prod_{r=1}^q \left( \sum_{j=1}^N f_r(\omega_j) \right) \right]$$

or equivalently

$$J = \mathbb{E} \left[ \prod_{r=1}^q \left( \int_{\mathbb{C}} f_r d\mu(M_N) \right) \right].$$

Now, by Proposition 3.1 and Corollary 3.2,

$$\left( \frac{1}{N} \int_{\mathbb{C}} f_r d\mu(M_N) \right)_{1 \leq r \leq q} \xrightarrow{N \rightarrow \infty} \left( \frac{1}{2\pi} \int_0^{2\pi} f_r(Re^{i\lambda}) d\lambda \right)_{1 \leq r \leq q}$$

in probability. By applying the bounded, continuous function from  $\mathbb{R}^q$  to  $\mathbb{R}$ :

$$(x_1, \dots, x_q) \mapsto (|x_1| \cdots |x_q|) \wedge (\|f_1\|_\infty \cdots \|f_q\|_\infty),$$

one deduces that

$$\frac{J}{N^q} \xrightarrow{N \rightarrow \infty} \frac{1}{(2\pi)^q} \int_{[0, 2\pi]^q} f_1(Re^{i\lambda_1}) \cdots f_q(Re^{i\lambda_q}) d\lambda_1 \cdots d\lambda_q.$$



By using the inequalities (3.11), one obtains that

$$\frac{I}{N(N-1)\cdots(N-q+1)} \xrightarrow{N \rightarrow \infty} \frac{1}{(2\pi)^q} \int_{[0,2\pi]^q} f_1(Re^{i\lambda_1}) \cdots f_q(Re^{i\lambda_q}) d\lambda_1 \cdots d\lambda_q,$$

which gives the first part of Proposition 3.20. In order to prove the second part, let us suppose that  $0 < R_1 < R < R_2$  and that  $f$  is a bounded, continuous function from  $\mathbb{C}^q$  to  $\mathbb{R}$ , vanishing if one of the coordinates has a modulus between  $R_1$  and  $R_2$ . It is sufficient to prove:

$$\sum_{(\lambda_{k,r})_{k,r \geq 1} \in \tilde{L}(q)} \left| \mathbb{1}_{N \geq \sum_{k,r \geq 1} k\lambda_{k,r}} t_{N, \sum_{k,r \geq 1} k\lambda_{k,r}, \theta}^{-1} \right| \prod_{k,r \geq 1} \left[ \frac{1}{\lambda_{k,r}!} \left( \frac{\theta}{r!k} \right)^{\lambda_{k,r}} \right] \sum_{\sigma \in \Sigma_q} \int_{\mathbb{C}^q} f d \left( \sigma \cdot \left[ \bigotimes_{k \geq 1, r \geq 1} \left( \mathcal{L}_k^{[r]} \right)^{\otimes \lambda_{k,r}} \right] \right) \xrightarrow{N \rightarrow \infty} 0,$$

with the notation of the proof of Propositions 3.15 and 3.17. By dominated convergence, we are done if we check:

$$\sum_{(\lambda_{k,r})_{k,r \geq 1} \in \tilde{L}(q)} v_{\sum_{k,r \geq 1} k\lambda_{k,r}, \theta} \prod_{k,r \geq 1} \left[ \frac{1}{\lambda_{k,r}!} \left( \frac{\theta}{r!k} \right)^{\lambda_{k,r}} \right] \cdots \sum_{\sigma \in \Sigma_q} \int_{\mathbb{C}^q} f d \left( \sigma \cdot \left[ \bigotimes_{k \geq 1, r \geq 1} \left( \mathcal{L}_k^{[r]} \right)^{\otimes \lambda_{k,r}} \right] \right) < \infty.$$

Now, for  $(\lambda_{k,r})_{k,r \geq 1} \in \tilde{L}(q)$ , by the estimates obtained in the proof of Proposition 3.15:

$$\begin{aligned} v_{\sum_{k,r \geq 1} k\lambda_{k,r}, \theta} &\leq C(\theta) \left( 1 + \sum_{k,r \geq 1} k\lambda_{k,r} \right)^{(1-\theta)_+} \\ &\leq 2qC(\theta) [1 \vee \sup\{k \geq 1, \exists r \geq 1, \lambda_{k,r} > 0\}]^{(1-\theta)_+} \\ &\leq 2qC(\theta) \prod_{k,r \geq 1} k^{\lambda_{k,r}(1-\theta)_+}, \end{aligned}$$

where  $C(\theta) > 0$  depends only on  $\theta$ . By assumption, if  $Z$  is a random variable which follows the distribution  $\mathcal{L}$ , then  $\log(Z) \in \mathcal{U}^{q+1}$ , which implies that

$$\mathcal{L}_k \{z \in \mathbb{C}, |z| \notin (R_1, R_2)\}$$

decreases at least as fast as  $1/k^{q+1}$  when  $k$  goes to infinity, with a constant depending only on  $\mathcal{L}$ ,  $R_1$  and  $R_2$ . Since  $f$  vanishes when a coordinate has

modulus in  $(R_1, R_2)$ , one deduces:

$$\int_{\mathbb{C}^q} f d \left[ \bigotimes_{k \geq 1, r \geq 1} \left( \mathcal{L}_k^{[r]} \right)^{\otimes \lambda_{k,r}} \right] \leq C \prod_{k \geq 1, r \geq 1} \left( \frac{k(k-1) \cdots (k-r+1)}{k^{q+1}} \right)^{\lambda_{k,r}}$$

where  $C$  can depend on  $\mathcal{L}, R_1, R_2, f, q$ . Hence, one only needs to check that:

$$\sum_{(\lambda_{k,r})_{k,r \geq 1} \in \tilde{L}(q)} \prod_{k,r \geq 1} \left[ \frac{1}{\lambda_{k,r}!} \left( \frac{\theta C_k^r}{k^\beta} \right)^{\lambda_{k,r}} \right] < \infty,$$

where  $\beta > q + 1$  and where by convention, the binomial coefficient  $C_k^r$  is equal to zero for  $k < r$ . Now, for  $(\lambda_{k,r})_{k,r \geq 1} \in \tilde{L}(q)$ :

$$C_k^r \leq k^r$$

and

$$\theta \sum_{k,r \geq 1} \lambda_{k,r} \leq (1 + \theta)^q.$$

Hence it is sufficient to have

$$\sum_{(\lambda_{k,r})_{k,r \geq 1} \in \tilde{L}(q)} \prod_{k,r \geq 1} k^{\lambda_{k,r}(r-\beta)} < \infty.$$

Since

$$\sum_{k,r \geq 1} r \lambda_{k,r} = q,$$

one necessarily has  $\lambda_{k,r} = 0$  for  $r > q$ . Hence, in any case,

$$k^{\lambda_{k,r}(r-\beta)} \leq k^{\lambda_{k,r}\gamma}$$

for  $\gamma = q - \beta < -1$ , and Proposition 3.20 is proved if

$$\sum_{(\lambda_{k,r})_{k,r \geq 1} \in \tilde{L}(q)} \prod_{k,r \geq 1} k^{\gamma \lambda_{k,r}} < \infty.$$

This last estimate is easy to check, by replacing  $\tilde{L}(q)$  by the (larger) set of families  $(\lambda_{k,r})_{k,r \geq 1}$  of nonnegative integers such that  $\lambda_{k,r} = 0$  for  $r > q$  and  $\lambda_{k,r} \leq q$  in any case. □

### 4. The point process of eigenangles and its scaling limit in the unitary case

In this section we do a more precise study of the point process of eigenvalues, in the unitary case. In particular, we obtain a scaling limit for the eigenangles if they are properly renormalized. This limit is dominated by the

large cycles of permutations, hence, we can expect almost sure convergence results if we consider virtual permutations. More precisely, let  $(M_N)_{N \geq 1}$  be a sequence of random matrices following the distribution  $\mathbb{P}(\infty, \theta, \mathcal{L})$ , where  $\theta > 0$  and  $\mathcal{L}$  is a probability distribution on the unit circle. Recall that the measure  $\mu(M_N)$ , representing the point process of eigenvalues can be written as:

$$\mu(M_N) = \sum_{m=1}^{\infty} \mathbb{1}_{l_{N,m} > 0} \sum_{\omega^{l_{N,m}} = Z_{N,m}} \delta_{\omega}.$$

Here  $l_{N,m}$  is the cardinality of the intersection  $C_{N,m}$  of  $C_m$  and  $\{1, \dots, N\}$ , and

$$Z_{N,m} = \prod_{j \in C_{N,m}} z_j,$$

where  $(C_m)_{m \geq 1}$  is the partition of  $\mathbb{N}^*$  (ordered by increasing smallest element) given by the cycle structure of a virtual permutation following the Ewens( $\theta$ ) measure, and  $(z_j)_{j \geq 1}$  is an independent sequence of i.i.d. variables with law  $\mathcal{L}$ . Since all the eigenvalues are on the unit circle, it can be more practical to consider the eigenangles. The corresponding point process can be described as follows: one takes the point  $x$  if and only if  $e^{ix}$  is an eigenvalue of  $M_N$ , with a multiplicity equal to the multiplicity of the eigenvalue. Note that this process is  $2\pi$ -periodic, and the corresponding random measure  $\tau(M_N)$  can be written as:

$$\tau(M_N) = \sum_{m=1}^{\infty} \mathbb{1}_{l_{N,m} > 0} \sum_{x \equiv \Theta_{N,m} \pmod{2\pi/l_{N,m}}} \delta_x,$$

where

$$\Theta_{N,m} = \frac{1}{il_{N,m}} \sum_{j \in C_{N,m}} \log(z_j)$$

is real and well-defined modulo  $2\pi/l_{N,m}$ , for  $l_{N,m} > 0$ . Note that the construction of the measure  $\tau(M_N)$  implies immediately that  $\tau(M_N)([0, 2\pi)) = N$ , in other words, the average spacing of two consecutive points of the corresponding point process is equal to  $2\pi/N$ . If we want to expect a convergence for  $N$  going to infinity, we need to rescale  $\tau(M_N)$  in order to have a constant average spacing, say, one. That is why we introduce the rescaled measure  $\tau_N(M_N)$ , defined as the image of  $\tau(M_N)$  by the multiplication by  $N/2\pi$ : the corresponding point process contains the points  $x$  such that  $e^{2i\pi x/N}$  is an eigenvalue of  $M_N$ . One checks that

$$\tau_N(M_N) = \sum_{m=1}^{\infty} \mathbb{1}_{y_{N,m} > 0} \sum_{k \in \mathbb{Z}} \delta_{(\gamma_{N,m+k})/y_{N,m}},$$

where  $y_{N,m} := l_{N,m}/N$  is the  $m$ -th renormalized cycle length of the permutation  $\sigma_N$  associated with  $M_N$ , and:

$$\gamma_{N,m} := \frac{1}{2i\pi} \sum_{j \in C_{N,m}} \log(z_j)$$

is well-defined modulo 1. Now the general results on virtual permutations (see [14] or Section 2) imply that for all  $m \geq 1$ ,  $y_{N,m}$  converges a.s. to a random variable  $y_m$ , where  $(y_m)_{m \geq 1}$  is a GEM process of parameter  $\theta$ . Then, if  $\gamma_{N,m}$  is supposed to be equal to zero, one can expect that the measure  $\tau_N(M_N)$  converges to

$$\tau_\infty((M_N)_{N \geq 1}) := \sum_{m=1}^\infty \sum_{k \in \mathbb{Z}} \delta_{k/y_m}.$$

Of course, one needs to be careful since  $\tau_\infty((M_N)_{N \geq 1})$  has an infinite Dirac mass at zero. Moreover, the condition  $\gamma_{N,m} = 0$  is satisfied if and only if  $\mathcal{L}$  is  $\delta_1$ , the Dirac measure at one. That is why we state the following result:

PROPOSITION 4.1. — *Let  $(M_N)_{N \geq 1}$  be a sequence of random matrices, which follows the distribution  $\mathbb{P}(\infty, \theta, \delta_1)$ , for  $\theta > 0$ . Then, with the notation above, the random measure  $\tau_N(M_N)$  converges a.s. to the random measure  $\tau_\infty((M_N)_{N \geq 1})$ , in the following sense: for all continuous functions  $f$  from  $\mathbb{R}$  to  $\mathbb{R}_+$ , with compact support,*

$$\int_{\mathbb{R}} f d\tau_N(M_N) \xrightarrow{N \rightarrow \infty} \int_{\mathbb{R}} f d\tau_\infty((M_N)_{N \geq 1})$$

almost surely.

Remark 4.2. — In Proposition 4.1, the matrix  $M_N$  is simply a permutation matrix, associated to a permutation  $\sigma_N \in \Sigma_N$ . Moreover, the positivity of  $f$  is needed in order to deal with the infinite Dirac mass at zero of  $\tau_\infty((M_N)_{N \geq 1})$ .

Proof. — If  $f(0) > 0$ , the integral of  $f$  with respect to  $\tau_N(M_N)$  is greater than or equal to  $f(0)$  times the number of  $m$  such that  $y_{N,m} > 0$ , i.e. the number of cycles of  $\sigma_N$ . Since this number tends a.s. to infinity with  $N$ , we are done. Then, we can suppose  $f(0) = 0$ , which implies:

$$\int_{\mathbb{R}} f d\tau_N(M_N) = \sum_{m=1}^\infty \sum_{k \in \mathbb{Z} \setminus \{0\}} f(k/y_{N,m})$$

and

$$\int_{\mathbb{R}} f d\tau_\infty((M_N)_{N \geq 1}) = \sum_{m=1}^\infty \sum_{k \in \mathbb{Z} \setminus \{0\}} f(k/y_m)$$

where we take the convention  $f(k/y_{N,m}) = 0$  for  $y_{N,m} = 0$ . Now, if the support of  $f$  is included in  $[-A, A]$  and if  $s_m$  is the supremum of  $y_{N,m}$  for all  $N \geq 1$ , it is clear that  $f(k/y_{N,m}) = f(k/y_m) = 0$  for all  $k, m$  such that  $|k| > As_m$ . One deduces that

$$\int_{\mathbb{R}} f d\tau_N(M_N) = \sum_{(m,k) \in S} f(k/y_{N,m})$$

and

$$\int_{\mathbb{R}} f d\tau_{\infty}((M_N)_{N \geq 1}) = \sum_{(m,k) \in S} f(k/y_m)$$

where  $S$  is the set (independent of  $N$ ) of couples of integers  $(m, k)$  such that  $m \geq 1$  and  $0 < |k| \leq As_m$ . Now

$$\sum_{m \geq 1} s_m < \infty$$

almost surely (see (3.3)), and then  $S$  is a.s. a finite set. Since  $f(k/y_{N,m})$  tends a.s. to  $f(k/y_m)$  for all  $k, m$  (recall that  $f$  is continuous), we are done. □

The a.s. weak convergence given in Proposition 4.1 cannot be directly generalized if the law  $\mathcal{L}$  is not the Dirac measure at one, because  $\gamma_{N,m}$  changes in a non-negligible way for each  $N \in C_m$ . However, one can expect a weaker convergence. A natural candidate for the corresponding limit distribution would be the law of a random measure defined by:

$$(4.1) \quad \sum_{m=1}^{\infty} \sum_{k \in \mathbb{Z}} \delta_{(k+\chi_m)/y_m},$$

where we recall that  $(y_m)_{m \geq 1}$  is a GEM( $\theta$ ) process, and where  $(\chi_m)_{m \geq 1}$  is an independent i.i.d. sequence in  $\mathbb{R}/\mathbb{Z}$ . The distribution of  $\chi_1$  needs to be, in a sense which has to be made precise, close to the distribution of:

$$\frac{1}{2i\pi} \sum_{j=1}^K \log(z_j),$$

where  $K$  is a large integer (recall that  $(z_j)_{j \geq 1}$  is an i.i.d. sequence of random variables with law  $\mathcal{L}$ ). From the following result, we deduce a good candidate:

LEMMA 4.3. — *Let  $\mathcal{L}$  be a probability measure on  $\mathbb{U}$ , and for  $k \geq 1$ , let  $\mathcal{L}^{*k}$  be the multiplicative convolution of  $k$  copies of  $\mathcal{L}$ . Moreover, let  $r(\mathcal{L})$  be the infimum of the integers  $r$  such that  $\mathcal{L}$  is carried by the set  $\mathbb{U}_r$  of  $r$ -th*

roots of unity. Then for all sequences  $(d_k)_{k \geq 1}$  of strictly positive integers, tending to infinity with  $k$ , the probability measure:

$$\frac{1}{d_k} \sum_{p=k}^{k+d_k-1} \mathcal{L}^{*p}$$

converges weakly to  $\mathcal{L}^*$ , where  $\mathcal{L}^*$  is the uniform distribution on  $\mathbb{U}_r$  if  $r(\mathcal{L}) < \infty$ , and the uniform distribution on  $\mathbb{U}$  if  $r(\mathcal{L}) = \infty$ .

*Proof.* — Let us define, for all  $q \in \mathbb{Z}$ , and for all probability measures  $\mathcal{M}$  on  $\mathbb{U}$ :

$$\widehat{\mathcal{M}}(q) := \int_{\mathbb{U}} z^q d\mathcal{M}.$$

Moreover, let us set, for  $k \geq 1$ :

$$\mathcal{M}_k := \frac{1}{d_k} \sum_{p=k}^{k+d_k-1} \mathcal{L}^{*p}$$

Then, one has:

$$\widehat{\mathcal{M}}_k(q) = \frac{1}{d_k} \sum_{p=k}^{k+d_k-1} [\widehat{\mathcal{L}}(q)]^p,$$

which is equal to one if  $\widehat{\mathcal{L}}(q) = 1$  and which tends to zero when  $k \rightarrow \infty$ , if  $\widehat{\mathcal{L}}(q) \neq 1$ . Now,  $\widehat{\mathcal{L}}(q) = 1$  if and only if  $q$  is divisible by  $r(\mathcal{L})$ , for  $r(\mathcal{L}) < \infty$ , and if and only if  $q = 0$ , for  $r(\mathcal{L}) = \infty$ . Hence:

$$\widehat{\mathcal{L}}^*(q) = \mathbf{1}_{\widehat{\mathcal{L}}(q)=0},$$

and finally,

$$\widehat{\mathcal{M}}_k(q) \xrightarrow[k \rightarrow \infty]{} \widehat{\mathcal{L}}^*(q).$$

□

Because of Lemma 4.3, the law  $D(\mathcal{L})$  of  $\chi_1$ , in equation (4.1), is chosen as follows: if  $r(\mathcal{L}) < \infty$ , it is the uniform distribution on the classes  $\{0, 1/r, \dots, (r-1)/r\}$  modulo 1, and if  $r(\mathcal{L}) = \infty$ , it is the uniform measure on  $\mathbb{R}/\mathbb{Z}$ .

*Remark 4.4.* — If  $r(\mathcal{L})$  is finite, it is the smallest integer  $r \geq 1$  such that  $M_N \in \mathcal{H}_r(N)$  a.s., for  $M_N$  following the distribution  $\mathcal{P}(N, \theta, \mathcal{L})$ .

We are now able to state our convergence result for any distribution  $\mathcal{L}$  on  $\mathbb{U}$ :

PROPOSITION 4.5. — Let  $(M_N)_{N \geq 1}$  be a sequence of random matrices, such that for all  $N \geq 1$ ,  $M_N$  follows the distribution  $\mathbb{P}(N, \theta, \mathcal{L})$ , where  $\theta > 0$  and where  $\mathcal{L}$  is a probability measure on  $\mathbb{U}$ . We suppose that  $\mathcal{L}$  satisfies one of the two following conditions:

- the measure is carried by  $\mathbb{U}_r$  for some integer  $r \geq 1$ .
- there exists  $v > 1$  such that for  $\epsilon > 0$  small enough, and for all arcs  $\mathcal{A}$  in  $\mathbb{U}$  of size  $\epsilon$ ,  $\mathcal{L}(\mathcal{A}) \leq |\log(\epsilon)|^{-v}$ .

Let  $\tau_\infty(\theta, \mathcal{L})$  be the random measure defined by:

$$\tau_\infty(\theta, \mathcal{L}) := \sum_{m=1}^{\infty} \sum_{k \in \mathbb{Z}} \delta_{(k+\chi_m)/x_m},$$

where  $(x_m)_{m \geq 1}$  is a Poisson-Dirichlet process of parameter  $\theta$ , and  $(\chi_m)_{m \geq 1}$  is an independent sequence of i.i.d. random variables on  $\mathbb{R}/\mathbb{Z}$ , following the distribution  $D(\mathcal{L})$  defined above. Then, with the previous notation, the distribution of the random measure  $\tau_N(M_N)$  converges to the distribution of the random measure  $\tau_\infty(\theta, \mathcal{L})$ , in the following sense: for all continuous functions  $f$  from  $\mathbb{R}$  to  $\mathbb{R}_+$ , with compact support,

- if  $f(0) > 0$  and  $r(\mathcal{L}) < \infty$ , then

$$\int_{\mathbb{R}} f d\tau_\infty(\theta, \mathcal{L}) = \infty$$

a.s., and for all  $A > 0$ :

$$\mathbb{P} \left[ \int_{\mathbb{R}} f d\tau_N(M_N) \leq A \right] \longrightarrow 0$$

when  $N \rightarrow \infty$ ;

- if  $f(0) = 0$  or  $r(\mathcal{L}) = \infty$ , then:

$$\int_{\mathbb{R}} f d\tau_\infty(\theta, \mathcal{L}) < \infty$$

a.s., and

$$\int_{\mathbb{R}} f d\tau_N(M_N) \xrightarrow{N \rightarrow \infty} \int_{\mathbb{R}} f d\tau_\infty(\theta, \mathcal{L})$$

in distribution.

Remark 4.6. — In Proposition 4.5, the Poisson-Dirichlet distribution can be replaced by a GEM distribution, since it does not change the law of the random measure  $\tau_\infty(\theta, \mathcal{L})$ .

*Proof.* — Without changing the laws of the random measures  $\tau_N(M_N)$  and  $\tau_\infty(\theta, \mathcal{L})$ , one can suppose that  $(M_N)_{N \geq 1}$  follows the distribution  $\mathbb{P}(\infty, \theta, \mathcal{L})$  and that, with the notation above:

$$\tau_\infty(\theta, \mathcal{L}) = \sum_{m=1}^\infty \sum_{k \in \mathbb{Z}} \delta_{(k+\chi_m)/y_m},$$

where  $(\chi_m)_{m \geq 1}$  is supposed to be independent of  $(y_m)_{m \geq 1}$ . Recall that

$$\tau_N(M_N) = \sum_{m=1}^\infty \mathbb{1}_{y_{N,m} > 0} \sum_{k \in \mathbb{Z}} \delta_{(k+\gamma_{N,m})/y_{N,m}},$$

where, conditionally on  $(y_{N,m})_{m \geq 1}$ ,  $(\gamma_{N,m})_{m \geq 1}$  is a sequence of independent random variables on  $\mathbb{R}/\mathbb{Z}$ , and for all  $m \geq 1$ ,  $\gamma_{N,m}$  has the same law as  $\frac{\log Z}{2i\pi}$ , where  $Z$  follows the distribution  $\mathcal{L}^{*N}y_{N,m} = \mathcal{L}^{*l_{N,m}}$ . Let us now suppose  $f(0) > 0$  and  $r(\mathcal{L}) < \infty$ . Since  $\chi_m$  is a.s. equal to zero for infinitely many  $m$  (because  $D(\mathcal{L})(\{0\}) > 0$ ),  $\tau_\infty(\theta, \mathcal{L})$  has an infinite Dirac mass at zero, and then a.s.:

$$\int_{\mathbb{R}} f d\tau_\infty(\theta, \mathcal{L}) = \infty.$$

Moreover

$$\int_{\mathbb{R}} f d\tau_N(M_N) \geq f(0) \sum_{m=1}^\infty \mathbb{1}_{y_{N,m} > 0, \gamma_{N,m} = 0}.$$

Now, if  $(Z_{p,q})_{p,q \geq 1}$  is a family of independent random variables, independent of  $\sigma_N$  (the permutation associated with  $M_N$ ), and such that for all  $p, q \geq 1$ ,  $Z_{p,q}$  has distribution  $\mathcal{L}^{*p}$ , then

$$\sum_{m=1}^\infty \mathbb{1}_{y_{N,m} > 0, \gamma_{N,m} = 0} = \sum_{p=1}^\infty \sum_{q=1}^{a_{N,p}} \mathbb{1}_{Z_{p,q} = 1}$$

in distribution, where  $a_{N,p}$  denotes the number of  $m$  such that  $l_{N,m} = p$  (i.e. the number of  $p$ -cycles in  $\sigma_N$ ). Therefore:

$$\int_{\mathbb{R}} f d\tau_N(M_N) \succeq f(0) \sum_{p=1}^\infty \sum_{q=1}^{c_{N,p}} \mathbb{1}_{Z_{p,q} = 1},$$

where  $\succeq$  denotes the stochastic domination, and where  $(c_{N,p})_{p \geq 1}$ , defined by the Feller coupling as in (3.7), is supposed to be independent of  $(Z_{p,q})_{p,q \geq 1}$ . Since for all  $p \geq 1$ ,  $c_{N,p}$  increases a.s. to  $b_p$ , Proposition 4.5 is proved for  $f(0) > 0$  and  $r(\mathcal{L}) < \infty$ , if we show that a.s.,

$$\sum_{p=1}^\infty \sum_{q=1}^{b_p} \mathbb{1}_{Z_{p,q} = 1} = \infty.$$



and a fortiori, if we prove that

$$\sum_{p=1}^{\infty} \mathbb{1}_{\exists q \leq b_p, Z_{p,q}=1} = \infty.$$

Since all the variables  $(b_p)_{p \geq 1}$  and  $(Z_{p,q})_{p,q \geq 1}$  are independent, by the Borel-Cantelli lemma it is sufficient to have

$$\sum_{p=1}^{\infty} \mathbb{P}[\exists q \leq b_p, Z_{p,q} = 1] = \infty.$$

Now, if for  $p \geq 1$ , we define

$$P(p) = \mathcal{L}^{*p}(\{1\}),$$

we have:

$$\mathbb{P}[\exists q \leq b_p, Z_{p,q} = 1 \mid b_p] = 1 - [1 - P(p)]^{b_p},$$

and, since  $b_p$  is a Poisson random variable of parameter  $\theta/p$ ,

$$\mathbb{P}[\exists q \leq b_p, Z_{p,q} = 1] = 1 - e^{-\theta P(p)/p},$$

and then we are done if we prove:

$$(4.2) \quad \sum_{p \geq 1} P(p)/p = \infty.$$

Now, by Lemma 4.3:

$$\frac{1}{2^k} \sum_{p=2^k}^{2^{k+1}-1} \mathcal{L}^{*p}$$

converges weakly to the uniform distribution on  $\mathbb{U}_{r(\mathcal{L})}$ , for  $k \rightarrow \infty$ . Since all the measures involved here are supported by the finite set  $\mathbb{U}_{r(\mathcal{L})}$ , one deduces that

$$\frac{1}{2^k} \sum_{p=2^k}^{2^{k+1}-1} P(p) \xrightarrow[k \rightarrow \infty]{} 1/r(\mathcal{L}),$$

which implies (4.2). We can now suppose  $f(0) = 0$  or  $r(\mathcal{L}) = \infty$ . In order to prove Proposition 4.5 in this case, we need the following result:

LEMMA 4.7. — *Let  $\theta > 0$ ,  $r \in \mathbb{N}^*$ , and let  $(l_{N,m})_{m \geq 1}$  be the sequence of cycle lengths (ordered by increasing smallest elements, and completed by zeros) of a random permutation in  $\Sigma_N$ , following the Ewens measure with parameter  $\theta$ . Then, there exists a function  $\Psi_{r,\theta}$  from  $\mathbb{N}^*$  to  $\mathbb{R}^+$ , tending to*

zero at infinity, and such that for all strictly positive integers  $(l_m)_{1 \leq m \leq r}$ ,  $(l'_m)_{1 \leq m \leq r}$ ,  $N$ , satisfying:

$$\sum_{m=1}^r l_m < N - N^{2/3}, \quad \sum_{m=1}^r l'_m < N - N^{2/3},$$

$$\forall m \in \{1, \dots, r\}, \exists q \in \mathbb{N}^*, q^2 \leq l_m, l'_m < (q + 1)^2,$$

one has

$$\frac{\mathbb{P}[\forall m \in \{1, \dots, r\}, l_{N,m} = l_m]}{\mathbb{P}[\forall m \in \{1, \dots, r\}, l_{N,m} = l'_m]} \in (1 - \Psi_{r,\theta}(N), 1 + \Psi_{r,\theta}(N)).$$

*Proof.* — For  $0 \leq m \leq r$ , let us define:

$$L_m = \sum_{p=1}^m l_p, \quad L'_m = \sum_{p=1}^m l'_p,$$

By using the Feller coupling, one obtains the following expression:

$$\mathbb{P}[\forall m \in \{1, \dots, r\}, l_{N,m} = l_m] = \prod_{m=1}^r \left[ \frac{\theta}{N - L_m + \theta} \prod_{p=L_{m-1}+1}^{L_m-1} \frac{N - p}{N - p + \theta} \right]$$

$$= \left( \prod_{p=1}^{L_r} \frac{N - p}{N - p + \theta} \right) \left( \prod_{m=1}^r \frac{\theta}{N - L_m} \right).$$

Hence, if  $L_r \geq L'_r$

$$\frac{\mathbb{P}[\forall m \in \{1, \dots, r\}, l_{N,m} = l_m]}{\mathbb{P}[\forall m \in \{1, \dots, r\}, l_{N,m} = l'_m]} = \left( \prod_{p=L'_r+1}^{L_r} \frac{N - p}{N - p + \theta} \right) \left( \prod_{m=1}^r \frac{N - L'_m}{N - L_m} \right).$$

For  $1 \leq m \leq r$ , there exists an integer  $q \geq 1$  such that  $q^2 \leq l_m, l'_m < (q + 1)^2$ , and  $q \leq \sqrt{N}$ , since  $l_m, l'_m < N$ . Hence,  $|l_m - l'_m| \leq 2q \leq 2\sqrt{N}$ , and  $|L_m - L'_m| \leq 2r\sqrt{N}$ . Moreover,  $N - L_m, N - L'_m > N^{2/3}$  by assumption. Therefore, one has the majorization:

$$\left| \log \left( \frac{\mathbb{P}[\forall m \in \{1, \dots, r\}, l_{N,m} = l_m]}{\mathbb{P}[\forall m \in \{1, \dots, r\}, l_{N,m} = l'_m]} \right) \right|$$

$$\leq 2r\sqrt{N} \log \left( 1 + \frac{\theta}{N^{2/3}} \right) + r \log \left( 1 + \frac{2r\sqrt{N}}{N^{2/3}} \right).$$

This implies Lemma 4.7 for:

$$\Psi_{r,\theta}(N) = \left( 1 + \frac{\theta}{N^{2/3}} \right)^{2r\sqrt{N}} \left( 1 + \frac{2r\sqrt{N}}{N^{2/3}} \right)^r + \frac{1}{N} - 1.$$

□

A consequence of Lemma 4.7 is the following:

LEMMA 4.8. — *Let  $\theta > 0$  and let  $\mathcal{L}$  be a probability law on the unit circle. Let  $(l_{N,m})_{m \geq 1}$  be the sequence of cycle lengths (ordered by increasing smallest elements, and completed by zeros) of a random permutation in  $\Sigma_N$ , following Ewens measure with parameter  $\theta > 0$ . Let  $(\gamma_{N,m})_{m \geq 1}$  be a sequence of random variables on  $\mathbb{R}/\mathbb{Z}$  such that conditionally on  $(l_{N,m})_{m \geq 1}$ ,  $(\gamma_{N,m})_{m \geq 1}$  are independent and for all  $m \geq 1$ ,  $\gamma_{N,m}$  has the same law as  $\frac{\log Z}{2i\pi}$ , where  $Z$  follows the multiplicative convolution  $\mathcal{L}^{*l_{N,m}}$  of  $l_{N,m}$  copies of  $\mathcal{L}$ . Then, for all  $r \geq 1$ , the family  $(l_{N,m}/N, \gamma_{N,m})_{m \leq r}$  converges in distribution to  $(y_m, \chi_m)_{m \leq r}$ , where  $(y_m)_{m \geq 1}$  is a GEM process of parameter  $\theta$ , and  $(\chi_m)_{m \geq 1}$  is an independent sequence of i.i.d. random variables, with law  $D(\mathcal{L})$ , defined above.*

*Proof.* — Let  $q_{N,m}$  be the integer part of the square root of  $l_{N,m}$ . For all continuous functions  $g$  from  $(\mathbb{R}/\mathbb{Z})^r$  to  $\mathbb{R}_+$ , one has:

$$\begin{aligned} & \mathbb{E}[g((\gamma_{N,m})_{m \leq r}) \mid (q_{N,m})_{m \leq r}] \\ &= \int_{\mathbb{U}^r} g(\log z_1/2i\pi, \dots, \log z_r/2i\pi) d\mathcal{N}_{(q_{N,m})_{m \leq r}}(z_1, \dots, z_r), \end{aligned}$$

where  $\mathcal{N}_{(q_{N,m})_{m \leq r}}$  is defined as the conditional expectation of a random measure, more precisely:

$$\mathcal{N}_{(q_{N,m})_{m \leq r}} = \mathbb{E} \left[ \bigotimes_{m=1}^r \mathcal{L}^{*l_{N,m}} \mid (q_{N,m})_{m \leq r} \right].$$

Let us suppose that  $q_{N,m} \geq N^{1/3}$  for all  $m \leq r$  and:

$$N - \sum_{m=1}^r (q_{N,m} + 1)^2 > N^{2/3}.$$

By Lemma 4.7, the conditional law of  $(l_{N,m})_{m \leq r}$  given  $(q_{N,m})_{m \leq r}$  is a linear combination of Dirac measures on the  $r$ -uples  $(l_m)_{m \leq r}$  of integers such that  $q_{N,m}^2 \leq l_m < (q_{N,m} + 1)^2$ , and the quotient between two coefficients of this combination lies on the interval  $(1 - \Psi_{r,\theta}(N), 1 + \Psi_{r,\theta}(N))$ . One deduces that

$$(1 - \Psi_{r,\theta}(N)) \bigotimes_{m=1}^r \mathcal{L}'_{q_{N,m}} \leq \mathcal{N}_{(q_{N,m})_{m \leq r}} \leq (1 + \Psi_{r,\theta}(N)) \bigotimes_{m=1}^r \mathcal{L}'_{q_{N,m}}$$

where, for  $q \geq 1$ :

$$\mathcal{L}'_q = \frac{1}{2q + 1} \sum_{l=q^2}^{q^2+2q} \mathcal{L}^{*l}.$$

Hence:

$$\begin{aligned} & \mathbb{E}[g((\gamma_{N,m})_{m \leq r}) \mid (q_{N,m})_{m \leq r}] \\ &= C \int_{\mathbb{U}^r} g(\log z_1/2i\pi, \dots, \log z_r/2i\pi) \prod_{m=1}^r \mathcal{L}'_{q_{N,m}}(dz_m) \end{aligned}$$

for

$$1 - \Psi_{r,\theta}(N) \leq C \leq 1 + \Psi_{r,\theta}(N).$$

Now, by Lemma 4.3,  $\mathcal{L}'_q$  converges to  $\mathcal{L}^*$  when  $q$  goes to infinity, hence, there exists a function  $K_g$  from  $\mathbb{R}_+$  to  $\mathbb{R}_+$ , tending to zero at infinity, and such that for all  $q \geq q_0 \geq 1$ ,  $q$  integer:

$$\begin{aligned} & \left| \int_{\mathbb{U}^r} g(\log z_1/2i\pi, \dots, \log z_r/2i\pi) \prod_{m=1}^r \mathcal{L}'_{q_{N,m}}(dz_m) \right. \\ & \quad \left. - \int_{\mathbb{U}^r} g(\log z_1/2i\pi, \dots, \log z_r/2i\pi) \prod_{m=1}^r \mathcal{L}^*(dz_m) \right| \leq K_g(q_0). \end{aligned}$$

Therefore, if  $N$  is large enough, if  $q_{N,m} \geq N^{1/3}$  for all  $m \leq r$  and if

$$N - \sum_{m=1}^r (q_{N,m} + 1)^2 > N^{2/3},$$

then:

$$\mathbb{E}[g((\gamma_{N,m})_{m \leq r}) \mid (q_{N,m})_{m \leq r}] \in [A_1, B_1],$$

for

$$\begin{aligned} A_1 = (1 - \Psi_{r,\theta}(N)) & \left[ -K_g(N^{1/3}) + \int_{\mathbb{U}^r} g(\log z_1/2i\pi, \dots \right. \\ & \quad \left. \dots, \log z_r/2i\pi) \prod_{m=1}^r D(\mathcal{L})(dz_m) \right] \end{aligned}$$

and

$$\begin{aligned} B_1 = (1 + \Psi_{r,\theta}(N)) & \left[ K_g(N^{1/3}) + \int_{\mathbb{U}^r} g(\log z_1/2i\pi, \dots \right. \\ & \quad \left. \dots, \log z_r/2i\pi) \prod_{m=1}^r D(\mathcal{L})(dz_m) \right]. \end{aligned}$$

Now, let  $f$  be a continuous function from  $[0, 1]^r$  to  $\mathbb{R}_+$ . One has:

$$\mathbb{E}[f((q_{N,m}^2/N)_{m \leq r})g((\gamma_{N,m})_{m \leq r})] \in [A_2, B_2],$$

for

$$A_2 = A_1 \mathbb{E}[f((q_{N,m}^2/N)_{m \leq r}) \mathbb{1}_{\mathcal{E}}]$$

and

$$B_2 = B_1 \mathbb{E} [f((q_{N,m}^2/N)_{m \leq r}) \mathbf{1}_{\mathcal{E}}] + \|f\|_{\infty} \|g\|_{\infty} \mathbb{P}[\mathcal{E}^c],$$

where

$$\mathcal{E} := \{ \forall m \leq r, q_{N,m} \geq N^{1/3}, N - \sum_{m=1}^r (q_{N,m} + 1)^2 > N^{2/3} \}$$

and  $\mathcal{E}^c$  is the complement of the event  $\mathcal{E}$ . If we suppose that  $(l_{N,m})_{N,m \geq 1}$  are the cycle lengths associated with a virtual permutation following the Ewens( $\theta$ ) measure, then for all  $m \leq r$ ,  $q_{N,m}^2/N$  tends a.s. to  $y_m$ , for a GEM( $\theta$ ) processes  $(y_m)_{m \geq 1}$ , and the event  $\mathcal{E}^c$  holds for finitely many values of  $N$ . Therefore:

$$\mathbb{E} [f((q_{N,m}^2/N)_{m \leq r}) \mathbf{1}_{\mathcal{E}}] \xrightarrow{N \rightarrow \infty} \mathbb{E} [f((y_m)_{m \leq r})]$$

and

$$\mathbb{P}[\mathcal{E}^c] \xrightarrow{N \rightarrow \infty} 0.$$

One deduces that

$$\mathbb{E} [f((q_{N,m}^2/N)_{m \leq r}) g((\gamma_{N,m})_{m \leq r})] \xrightarrow{N \rightarrow \infty} \mathbb{E} [f((y_m)_{m \leq r}) g((\chi_m)_{m \leq r})]$$

(recall that  $(y_m)_{m \leq r}$  and  $(\chi_m)_{m \leq r}$  are supposed to be independent). Now,

$$\left| \frac{q_{N,m}^2}{N} - \frac{l_{N,m}}{N} \right| \leq \frac{2}{\sqrt{N}},$$

hence, by the uniform continuity of  $f$  and the boundedness of  $g$ :

$$\mathbb{E} [f(l_{N,m}/N)_{m \leq r} g((\gamma_{N,m})_{m \leq r})] \xrightarrow{N \rightarrow \infty} \mathbb{E} [f((y_m)_{m \leq r}) g((\chi_m)_{m \leq r})],$$

which proves Lemma 4.8. □

Now, let  $f$  be a continuous, nonnegative function with compact support, and  $\Phi$  a continuous function from  $\mathbb{R}_+$  to  $[0, 1]$ , such that  $\Phi(x) = 0$  for  $x \leq 1$  and  $\Phi(x) = 1$  for  $x \geq 2$ . Let us suppose  $f(0) = 0$  or  $r(\mathcal{L}) = \infty$ . We first remark that for all integers  $r, s \geq 1$ :

$$\sum_{m=1}^r \Phi(s y_{N,m}) \sum_{k \in \mathbb{Z}} f[(k + \gamma_{N,m})/y_{N,m}] \xrightarrow{N \rightarrow \infty} \sum_{m=1}^r \Phi(s y_m) \sum_{k \in \mathbb{Z}} f[(k + \chi_m)/y_m]$$

in distribution. This is a consequence of Lemma 4.8 and the fact that

$$\Phi(sy) \sum_{k \in \mathbb{Z}} f[(k + \chi)/y]$$

is continuous with respect to  $(y, \chi) \in [0, 1] \times (\mathbb{R}/\mathbb{Z})$ . Now, if  $f$  is supported by  $[-A, A]$  (for  $A > 0$ ), one has the inequalities:

$$\sum_{k \in \mathbb{Z}} f[(k + \gamma_{N,m})/y_{N,m}] \leq (2A + 1)\|f\|_\infty$$

and

$$\sum_{k \in \mathbb{Z}} f[(k + \chi_m)/y_m] \leq (2A + 1)\|f\|_\infty.$$

Therefore, for  $\lambda \in \mathbb{R}_+$ :

$$\begin{aligned} & \left| \mathbb{E} \left[ e^{i\lambda \sum_{m=1}^r \mathbb{1}_{y_{N,m} > 0} \sum_{k \in \mathbb{Z}} f[(k + \gamma_{N,m})/y_{N,m}]} \right] \right. \\ & \quad \left. - \mathbb{E} \left[ e^{i\lambda \sum_{m=1}^r \sum_{k \in \mathbb{Z}} f[(k + \chi_m)/y_m]} \right] \right| \\ & \leq \left| \mathbb{E} \left[ e^{i\lambda \sum_{m=1}^r \Phi(sy_{N,m}) \sum_{k \in \mathbb{Z}} f[(k + \gamma_{N,m})/y_{N,m}]} \right] \right. \\ & \quad \left. - \mathbb{E} \left[ e^{i\lambda \sum_{m=1}^r \Phi(sy_m) \sum_{k \in \mathbb{Z}} f[(k + \chi_m)/y_m]} \right] \right| \\ & \quad + (2A + 1)\lambda \|f\|_\infty \sum_{m=1}^r (\mathbb{P}[0 < y_{N,m} \leq 2/s] + \mathbb{P}[y_m \leq 2/s]). \end{aligned}$$

The first term in the right-hand side of this inequality tends to zero when  $N$  goes to infinity. Hence:

$$\begin{aligned} & \limsup_{N \rightarrow \infty} \left| \mathbb{E} \left[ e^{i\lambda \sum_{m=1}^r \mathbb{1}_{y_{N,m} > 0} \sum_{k \in \mathbb{Z}} f[(k + \gamma_{N,m})/y_{N,m}]} \right] \right. \\ & \quad \left. - \mathbb{E} \left[ e^{i\lambda \sum_{m=1}^r \sum_{k \in \mathbb{Z}} f[(k + \chi_m)/y_m]} \right] \right| \\ & \leq (2A + 1)\lambda \|f\|_\infty \sum_{m=1}^r \left( \limsup_{N \rightarrow \infty} \mathbb{P}[y_{N,m} \leq 2/s] + \mathbb{P}[y_m \leq 2/s] \right). \end{aligned}$$

Now if we assume that  $(y_{N,m})_{N, m \geq 1}$  is the family of renormalized cycle lengths of a virtual permutation following the Ewens( $\theta$ ) measure, we obtain, by Fatou's lemma:

$$\liminf_{N \rightarrow \infty} \mathbb{P}[y_{N,m} > 2/s] \geq \mathbb{E} \left[ \liminf_{N \rightarrow \infty} \mathbb{1}_{y_{N,m} > 2/s} \right].$$

Since  $y_{N,m}$  tends to  $y_m$  for  $N$  going to infinity:

$$\limsup_{N \rightarrow \infty} \mathbb{P}[y_{N,m} \leq 2/s] \leq \mathbb{P}[y_m \leq 2/s].$$

By taking  $s \rightarrow \infty$ , we deduce that:

$$(4.3) \quad \left| \mathbb{E} \left[ e^{i\lambda \sum_{m=1}^r \mathbb{1}_{y_{N,m} > 0} \sum_{k \in \mathbb{Z}} f[(k + \gamma_{N,m})/y_{N,m}]} \right] - \mathbb{E} \left[ e^{i\lambda \sum_{m=1}^r \sum_{k \in \mathbb{Z}} f[(k + \chi_m)/y_m]} \right] \right|$$

tends to zero when  $N$  goes to infinity. Now, for all  $\chi \in [0, 1)$  and  $y \in (0, 1]$ :

$$\sum_{k \in \mathbb{Z}} f[(k + \chi)/y] \leq (2A + 1)\|f\|_\infty (\mathbb{1}_{\chi \leq Ay} + \mathbb{1}_{1-\chi \leq Ay})$$

and if  $f(0) = 0$ ,  $\chi = 0$ :

$$\sum_{k \in \mathbb{Z}} f[(k + \chi)/y] \leq (2A + 1)\|f\|_\infty \mathbb{1}_{Ay \geq 1}.$$

Therefore, if  $\gamma_{N,m}$  is identified with the unique element of  $[0, 1)$  in its congruence class modulo 1, then:

$$\begin{aligned} (4.4) \quad \mathbb{E} & \left[ \sum_{m=r+1}^\infty \mathbb{1}_{y_{N,m} > 0} \sum_{k \in \mathbb{Z}} f[(k + \gamma_{N,m})/y_{N,m}] \right] \\ & \leq (2A + 1)\|f\|_\infty \sum_{m=r+1}^\infty \left[ \mathbb{P}(y_{N,m} > 0, 0 < \gamma_{N,m} \leq Ay_{N,m}) \right. \\ & \quad \left. + \mathbb{P}(y_{N,m} > 0, 1 - \gamma_{N,m} \leq Ay_{N,m}) + \mathbb{P}(\gamma_{N,m} = 0, Ay_{N,m} \geq 1) \right] \end{aligned}$$

if  $f(0) = 0$ . In fact, this inequality remains true for  $f(0) > 0$ . Indeed, in this case, one has, by assumption,  $r(\mathcal{L}) = \infty$ , and then by the conditions given in Proposition 4.5,  $\mathcal{L}$  has no atom. Since

$$\begin{aligned} \mathbb{E} & \left[ \sum_{m=r+1}^\infty \mathbb{1}_{y_{N,m} > 0} \sum_{k \in \mathbb{Z}} f[(k + \gamma_{N,m})/y_{N,m}] \right] \\ & \leq (2A + 1)\|f\|_\infty \sum_{m=r+1}^\infty \left[ \mathbb{P}(y_{N,m} > 0, \gamma_{N,m} \leq Ay_{N,m}) \right. \\ & \quad \left. + \mathbb{P}(y_{N,m} > 0, 1 - \gamma_{N,m} \leq Ay_{N,m}) \right], \end{aligned}$$

we also have (4.4). Now, for all  $\rho \in (0, 1)$ :

$$\begin{aligned} & \mathbb{P}[y_{N,m} > 0, 0 < \gamma_{N,m} \leq Ay_{N,m}] \\ & = \mathbb{P}[y_{N,m} > 0, 0 < \gamma_{N,m} \leq \rho^m] + \mathbb{P}[y_{N,m} \geq \rho^m/A]. \end{aligned}$$

Since conditionally on  $\{y_{N,m} > 0\}$ , the law of  $\gamma_{N,m}$  is a convex combination of iterated convolutions of  $\mathcal{L}$ , it satisfies the same assumptions as  $\mathcal{L}$ , with the same underlying constants. One deduces that there exist  $v > 1$ ,  $m_0 \geq 1$ ,  $C > 0$ , independent of  $N$ , such that for  $m \geq m_0$ :

$$(4.5) \quad \mathbb{P}[y_{N,m} > 0, 0 < \gamma_{N,m} \leq \rho^m] \leq Cm^{-v}.$$

Now, if the normalized lengths of cycles  $(y_{N,m})_{N,m \geq 1}$  are associated with a virtual permutation following the Ewens( $\theta$ ) measure, and if for all  $m \geq 1$ ,  $s_m$  is the supremum of  $y_{N,m}$  for all  $N \geq 1$ , then the expectation of  $s_m$

decreases exponentially with  $m$  (this result is contained in our proof of (3.4) above). One deduces that there exist  $\rho \in (0, 1)$  and  $m_1 \geq 1$ , independent of  $N$  and such that for  $m \geq m_0$ :

$$(4.6) \quad \mathbb{P}[y_{N,m} \geq \rho^m / A] \leq \rho^m.$$

From (4.5) and (4.6),

$$\sup_{N \geq 1} \sum_{m=r+1}^{\infty} \mathbb{P}[y_{N,m} > 0, 0 < \gamma_{N,m} \leq Ay_{N,m}] \xrightarrow{r \rightarrow \infty} 0.$$

In a similar way, one can prove:

$$\sup_{N \geq 1} \sum_{m=r+1}^{\infty} \mathbb{P}[y_{N,m} > 0, 1 - \gamma_{N,m} \leq Ay_{N,m}] \xrightarrow{r \rightarrow \infty} 0$$

and

$$\sup_{N \geq 1} \sum_{m=r+1}^{\infty} \mathbb{P}[\gamma_{N,m} = 0, Ay_{N,m} \geq 1] \xrightarrow{r \rightarrow \infty} 0.$$

Therefore, by (4.4):

$$(4.7) \quad \sup_{N \geq 1} \mathbb{E} \left[ \sum_{m=r+1}^{\infty} \mathbf{1}_{y_{N,m} > 0} \sum_{k \in \mathbb{Z}} f[(k + \gamma_{N,m})/y_{N,m}] \right] \xrightarrow{r \rightarrow \infty} 0.$$

By replacing  $\gamma_{N,m}$  by  $\chi_m$  and  $y_{N,m}$  by  $y_m$ , one obtains:

$$(4.8) \quad \mathbb{E} \left[ \sum_{m=r+1}^{\infty} \mathbf{1}_{y_m > 0} \sum_{k \in \mathbb{Z}} f[(k + \chi_m)/y_m] \right] \xrightarrow{r \rightarrow \infty} 0,$$

since the law of  $\chi_m$  satisfies the same assumptions as  $\mathcal{L}$  and the expectation of  $y_m$  decreases exponentially with  $m$ . Since the quantity given by (4.3) tends to zero when  $N$  goes to infinity, one easily deduces from (4.7) and (4.8) that it is also the case for:

$$\left| \mathbb{E} \left[ e^{i\lambda \sum_{m=1}^{\infty} \mathbf{1}_{y_{N,m} > 0} \sum_{k \in \mathbb{Z}} f[(k + \gamma_{N,m})/y_{N,m}]} \right] - \mathbb{E} \left[ e^{i\lambda \sum_{m=1}^{\infty} \sum_{k \in \mathbb{Z}} f[(k + \chi_m)/y_m]} \right] \right|,$$

which proves Proposition 4.5. □

## 5. The uniform case

### 5.1. Eigenvalues distributions and correlation measures

In this section, we focus on the uniform case, i.e. the case where for  $N \geq 1$ , the random matrix  $M_N$  follows the law  $\mathbb{P}(N, \theta, \mathcal{L})$ , where  $\theta > 0$



and  $\mathcal{L}$  is the uniform distribution on the unit circle. Let  $(y_{N,m})_{N,m \geq 1}$  be the family of renormalized cycle lengths corresponding to a virtual permutation following the Ewens( $\theta$ ) measure,  $y_m$  the limit of  $y_{N,m}$  for  $N$  going to infinity, and  $(\chi_m)_{m \geq 1}$  a sequence of i.i.d. uniform random variables on  $[0, 1)$ , independent of  $(y_{N,m})_{N,m \geq 1}$ . The random measure  $\tau_N(M_N)$  has the same law as

$$\bar{\tau}_N := \sum_{m=1}^{\infty} \mathbb{1}_{y_{N,m} > 0} \sum_{k \in \mathbb{Z}} \delta_{(\chi_m + k)/y_{N,m}}$$

and  $\tau_\infty(\theta, \mathcal{L})$  is equal to

$$\bar{\tau}_\infty := \sum_{m=1}^{\infty} \sum_{k \in \mathbb{Z}} \delta_{(\chi_m + k)/y_m}$$

in distribution. This description of the law of  $\tau_N(M_N)$  and  $\tau_\infty(\theta, \mathcal{L})$  implies the following remarkable property:

**PROPOSITION 5.1.** — *Under the assumptions given above, the distributions of the random measures  $\tau_N(M_N)$  and  $\tau_\infty(\theta, \mathcal{L})$  are invariant by translation.*

*Proof.* — It is enough to prove it for  $\bar{\tau}_N$  and  $\bar{\tau}_\infty$ . The image of  $\bar{\tau}_N$  by a translation of  $A \in \mathbb{R}$  is equal to

$$\sum_{m=1}^{\infty} \mathbb{1}_{y_{N,m} > 0} \sum_{k \in \mathbb{Z}} \delta_{(\chi'_m + k)/y_{N,m}},$$

where  $\chi'_m$  is the fractional part of  $\chi_m + Ay_{N,m}$ . Now one easily sees that conditionally on  $(y_{N,m})_{m \geq 1}$ , the sequence  $(\chi'_m)_{m \geq 1}$  is an i.i.d. sequence of uniform variables on  $[0, 1)$ , as  $(\chi_m)_{m \geq 1}$ . This implies the invariance by translation of  $\bar{\tau}_N$ . For  $\bar{\tau}_\infty$ , the proof is exactly similar.  $\square$

The main interest of the introduction of the measures  $\bar{\tau}_N$  and  $\bar{\tau}_\infty$  is the following: we cannot expect an a.s. convergence of the random measure  $\tau_N(M_N)$  when  $N$  goes to infinity, even if  $(M_N)_{N \geq 1}$  follows the distribution  $\mathbb{P}(\infty, \theta, \mathcal{L})$  because, with the notation of the previous section, the “shifts”  $\gamma_{N,m}$  (which are equal to  $\chi_m$  in distribution), do not converge a.s. when  $N$  goes to infinity; however, since we take the same variables  $\chi_m$  in the definitions of  $\bar{\tau}_N$  and  $\bar{\tau}_\infty$ , we can expect such a convergence for these new measures. More precisely, we have the following:

**PROPOSITION 5.2.** — *Almost surely, with the notation above, the random measure  $\bar{\tau}_N$  converges vaguely to  $\bar{\tau}_\infty$ , which is locally finite.*

*Proof.* — The fact that  $\bar{\tau}_\infty$  is locally finite is a consequence of Proposition 4.5 (one has  $r(\mathcal{L}) = \infty$ ). Now, let  $f$  be a continuous function from  $\mathbb{R}$  to  $\mathbb{R}$ , with support included in  $[-A, A]$  (for some  $A > 0$ ). One has:

$$\int_{\mathbb{R}} f d\bar{\tau}_N = \sum_{m=1}^{\infty} \sum_{k \in \mathbb{Z}} \mathbb{1}_{y_{N,m} > 0} f((\chi_m + k)/y_{N,m})$$

and

$$\int_{\mathbb{R}} f d\bar{\tau}_\infty = \sum_{m=1}^{\infty} \sum_{k \in \mathbb{Z}} f((\chi_m + k)/y_m).$$

Let  $s_m$  be the supremum of  $y_{N,m}$  for  $N \geq 1$ . For  $|k| > A + 1$  or  $As_m < \chi_m < 1 - As_m$ , one has:

$$\mathbb{1}_{y_{N,m} > 0} f((\chi_m + k)/y_{N,m}) = f((\chi_m + k)/y_m) = 0,$$

and since

$$\mathbb{E} \left[ \sum_{m \geq 1} s_m \right] < \infty,$$

there exists a.s. a (random) finite subset  $S$  of  $\mathbb{N}^* \times \mathbb{Z}$  such that:

$$\int_{\mathbb{R}} f d\bar{\tau}_N = \sum_{(m,k) \in S} \mathbb{1}_{y_{N,m} > 0} f((\chi_m + k)/y_{N,m})$$

and

$$\int_{\mathbb{R}} f d\bar{\tau}_\infty = \sum_{(m,k) \in S} f((\chi_m + k)/y_m).$$

Since a.s., for all  $m \geq 1$ ,  $k \in \mathbb{Z}$ :

$$\mathbb{1}_{y_{N,m} > 0} f((\chi_m + k)/y_{N,m}) \xrightarrow[N \rightarrow \infty]{} f((\chi_m + k)/y_m),$$

one deduces Proposition 5.2. □

One can now study the correlation measures associated with the point processes studied above. More precisely, if

$$\tau_N(M_N) = \sum_{j \geq 1} \delta_{x_j},$$

where  $(x_j)_{j \geq 1}$  is a sequence of real numbers, one can define, for all  $q \geq 1$  the random measure on  $\mathbb{R}^q$ :

$$\tau_N^{[q]}(M_N) = \sum_{j_1 \neq j_2 \neq \dots \neq j_q} \delta_{(x_{j_1}, \dots, x_{j_q})}.$$

Similarly, one can define  $\tau_\infty^{[q]}(\theta, \mathcal{L})$ ,  $\bar{\tau}_N^{[q]}$  and  $\bar{\tau}_\infty^{[q]}$ . Of course, one has the following equalities in distribution:

$$\tau_N^{[q]}(M_N) = \bar{\tau}_N^{[q]}$$

and

$$\tau_\infty^{[q]}(\theta, \mathcal{L}) = \bar{\tau}_\infty^{[q]}.$$

Proposition 5.2 can be generalized as follows:

PROPOSITION 5.3. — For all  $q \geq 1$  the random measure  $\bar{\tau}_N^{[q]}$  converges vaguely towards  $\bar{\tau}_\infty^{[q]}$ , which is locally finite.

*Proof.* — The fact that  $\bar{\tau}_\infty^{[q]}$  is locally finite is a consequence of the local finiteness of  $\bar{\tau}_\infty$ . Now, let  $f$  be a continuous function from  $\mathbb{R}^q$  to  $\mathbb{R}$ , with support included in  $[-A, A]^q$  (for some  $A > 0$ ). One has:

$$\int_{\mathbb{R}} f d\bar{\tau}_N^{[q]} = \sum_{(m_1, k_1) \neq \dots \neq (m_q, k_q) \in \mathbb{N}^* \times \mathbb{Z}} \mathbb{1}_{y_{N, m_1}, \dots, y_{N, m_q} > 0} f((\chi_{m_1} + k_1)/y_{N, m_1}, \dots, (\chi_{m_q} + k_q)/y_{N, m_q})$$

and

$$\int_{\mathbb{R}} f d\bar{\tau}_\infty^{[q]} = \sum_{(m_1, k_1) \neq \dots \neq (m_q, k_q) \in \mathbb{N}^* \times \mathbb{Z}} f((\chi_{m_1} + k_1)/y_{m_1}, \dots, (\chi_{m_q} + k_q)/y_{m_q})$$

Let  $s_m$  be the supremum of  $y_{N, m}$  for  $N \geq 1$ . If for some  $j \leq q$ ,  $|k_j| > A + 1$  or  $As_{m_j} < \chi_{m_j} < 1 - As_{m_j}$ , one has:

$$\mathbb{1}_{y_{N, m_1}, \dots, y_{N, m_q} > 0} f((\chi_{m_1} + k_1)/y_{N, m_1}, \dots, (\chi_{m_q} + k_q)/y_{N, m_q}) = 0$$

and

$$f((\chi_{m_1} + k_1)/y_{m_1}, \dots, (\chi_{m_q} + k_q)/y_{m_q}) = 0.$$

Since

$$\mathbb{E} \left[ \sum_{m \geq 1} s_m \right] < \infty,$$

there exists a.s. a (random) finite subset  $S$  of  $\mathbb{N}^* \times \mathbb{Z}$  such that:

$$\int_{\mathbb{R}} f d\bar{\tau}_N^{[q]} = \sum_{(m_1, k_1) \neq \dots \neq (m_q, k_q) \in S} \mathbb{1}_{y_{N, m_1}, \dots, y_{N, m_q} > 0} f((\chi_{m_1} + k_1)/y_{N, m_1}, \dots, (\chi_{m_q} + k_q)/y_{N, m_q})$$

and

$$\int_{\mathbb{R}} f d\bar{\tau}_\infty^{[q]} = \sum_{(m_1, k_1) \neq \dots \neq (m_q, k_q) \in S} f((\chi_{m_1} + k_1)/y_{m_1}, \dots, (\chi_{m_q} + k_q)/y_{m_q})$$

Since a.s., for all  $m_1, \dots, m_q \geq 1, k_1, \dots, k_q \in \mathbb{Z}$ :

$$\mathbb{1}_{y_{N, m_1}, \dots, y_{N, m_q} > 0} f((\chi_{m_1} + k_1)/y_{N, m_1}, \dots, (\chi_{m_q} + k_q)/y_{N, m_q})$$

tends to

$$f((\chi_{m_1} + k_1)/y_{m_1}, \dots, (\chi_{m_q} + k_q)/y_{m_q})$$

when  $N$  goes to infinity, one deduces Proposition 5.3. □

Now, for  $q \geq 1$ , we define the “ $q$ -correlation measure”  $\tilde{\tau}_N^{[q]}$  associated with the point process  $\tau_N(M_N)$  (or equivalently,  $\bar{\tau}_N$ ), as the average of the measure  $\tau_N^{[q]}(M_N)$  (or  $\bar{\tau}_N^{[q]}$ ). Similarly, we define the measure  $\tilde{\tau}_\infty^{[q]}$  as the average of  $\tau_\infty^{[q]}(\theta, \mathcal{L})$  of  $\bar{\tau}_\infty^{[q]}$ . Both  $\tilde{\tau}_N^{[q]}$  and  $\tilde{\tau}_\infty^{[q]}$  are positive measures on  $\mathbb{R}^q$ . From the convergence of  $\tilde{\tau}_N^{[q]}$  toward  $\tilde{\tau}_\infty^{[q]}$ , one deduces the following result:

**PROPOSITION 5.4.** — *For all  $q \geq 1$ , the measures  $\tilde{\tau}_N^{[q]}$  ( $N \geq 1$ ) and  $\tilde{\tau}_\infty^{[q]}$  are locally finite, and  $\tilde{\tau}_N^{[q]}$  converges vaguely towards  $\tilde{\tau}_\infty^{[q]}$  when  $N$  goes to infinity.*

*Proof.* — Let  $f$  be a continuous function from  $\mathbb{R}$  to  $\mathbb{R}$ , with compact support. One has,

$$\begin{aligned} & \left| \int_{\mathbb{R}} f d\tilde{\tau}_N^{[q]} \right| \\ & \leq \|f\|_\infty \sum_{(m_1, k_1) \neq \dots \neq (m_q, k_q) \in \mathbb{N}^* \times \mathbb{Z}} \mathbb{1}_{(|k_j| \leq A+1)_{j \leq q}, (\chi_{m_j} \notin (As_{m_j}, 1 - As_{m_j}))_{j \leq q}} \end{aligned}$$

which is independent of  $N$  and has a finite expectation. Indeed, this quantity is bounded by a constant times

$$\left[ \sum_{m=1}^\infty (\mathbb{1}_{\chi_m \notin (As_m, 1 - As_m)}) \right]^q,$$

and

$$\begin{aligned} & \mathbb{E} \left[ \left( \sum_{m=1}^\infty (\mathbb{1}_{\chi_m \notin (As_m, 1 - As_m)}) \right)^q \right] \\ & \leq \sum_{m_1, \dots, m_q \geq 1} \mathbb{P}[(\chi_{m_j} \notin (As_j, 1 - As_j))_{j \geq 1}] \\ & \leq \sum_{m \geq 1} \sum_{\max\{m_j, 1 \leq j \leq q\} = m} \mathbb{P}[\chi_m \notin (As_m, 1 - As_m)] \\ & \leq 2A \sum_{m \geq 1} m^q \mathbb{E}[s_m] < \infty, \end{aligned}$$

since the expectation of  $s_m$  decreases exponentially with  $m$  (see the proof of (3.4)). Almost surely, by proposition 5.2,

$$\int_{\mathbb{R}} f d\tilde{\tau}_N^{[q]} \xrightarrow{N \rightarrow \infty} \int_{\mathbb{R}} f d\tilde{\tau}_\infty^{[q]},$$

and then one obtains, by taking expectation and applying dominated convergence:

$$\int_{\mathbb{R}} f d\tilde{\tau}_N^{[q]} \xrightarrow{N \rightarrow \infty} \int_{\mathbb{R}} f d\tilde{\tau}_\infty^{[q]},$$

where all these integrals are finite. This proves Proposition 5.4.  $\square$

For  $q = 1$ , the  $q$ -correlation measure of a subset of  $\mathbb{R}$  is simply the average number of points lying on this set. It can be very simply expressed:

PROPOSITION 5.5. — *The 1-correlation measures  $\tilde{\tau}_N^{[1]}$  ( $N \geq 1$ ) and  $\tilde{\tau}_\infty^{[1]}$  are equal to Lebesgue measure on  $\mathbb{R}$ .*

*Proof.* — Let  $f$  be a nonnegative, continuous function from  $\mathbb{R}$  to  $\mathbb{R}$ . One has:

$$\begin{aligned} \mathbb{E} \left[ \int_{\mathbb{R}} f d\tilde{\tau}_N \mid (y_{N,m})_{m \geq 1} \right] &= \sum_{m \geq 1} \mathbb{1}_{y_{N,m} > 0} \sum_{k \in \mathbb{Z}} \int_0^1 f((k+x)/y_{N,m}) dx \\ &= \sum_{m \geq 1} y_{N,m} \int_{\mathbb{R}} f(z) dz \\ &= \int_{\mathbb{R}} f(z) dz, \end{aligned}$$

which proves Proposition 5.5 for  $\tilde{\tau}_N^{[1]}$ . The proof for  $\tilde{\tau}_\infty^{[1]}$  is similar.  $\square$

We remark that for  $q \geq 2$ , the correlation measure  $\tilde{\tau}_N^{[q]}$  is not absolutely continuous with respect to the Lebesgue measure. Indeed, for all integers  $l$ ,  $1 \leq l \leq N$ , the probability that the point process associated with  $\tilde{\tau}_N$  has two points separated by an interval of exactly  $N/l$  is not equal to zero (this event holds if the corresponding permutation has a cycle of length  $l$ ). Similarly, for  $q \geq 3$ ,  $\tilde{\tau}_\infty^{[q]}$  is not absolutely continuous with respect to the Lebesgue measure, since the point process associated with  $\tilde{\tau}_\infty$  has almost surely three distinct points  $x, y, z \in \mathbb{R}$  such that  $y - x = z - y$ .

However, despite the fact that  $\tilde{\tau}_N^{[2]}$  is not absolutely continuous with respect to the Lebesgue measure for all  $N \geq 1$ , its limit  $\tilde{\tau}_\infty^{[2]}$  is absolutely continuous with respect to the Lebesgue measure. More precisely one has the following:

PROPOSITION 5.6. — *The measure  $\tilde{\tau}_\infty^{[2]}$  on  $\mathbb{R}^2$  (which depends on the parameter  $\theta$ ) has a density  $\rho$  with respect to the Lebesgue measure, which is called “2-correlation function”, and which is given by:*

$$\rho(x, y) = \phi_\theta(x - y),$$

where the function  $\phi_\theta$  from  $\mathbb{R}$  to  $\mathbb{R}$  is defined by:

$$\phi_\theta(x) = \frac{\theta}{\theta + 1} + \frac{\theta}{x^2} \sum_{a \in \mathbb{N}^*, a \leq |x|} a \left(1 - \frac{a}{|x|}\right)^{\theta-1}.$$

*Proof.* — One can write:

$$\begin{aligned} \bar{\tau}_\infty^{[2]} &= \sum_{m \neq m' \geq 1} \sum_{k, k' \in \mathbb{Z}} \delta_{((\chi_m+k)/y_m, (\chi_{m'}+k')/y_{m'})} \\ &\quad + \sum_{m \geq 1} \sum_{k \neq k' \in \mathbb{Z}} \delta_{((\chi_m+k)/y_m, (\chi_m+k')/y_m)}, \end{aligned}$$

which implies that for all nonnegative and continuous functions  $f$  from  $\mathbb{R}^2$  to  $\mathbb{R}$ ,

$$\begin{aligned} (5.1) \quad \mathbb{E} \left[ \int_{\mathbb{R}^2} f d\bar{\tau}_\infty^{[2]} \Big| (y_m)_{m \geq 1} \right] &= \sum_{m \neq m' \geq 1} \sum_{k, k' \in \mathbb{Z}} \int_0^1 \int_0^1 f((k+x)/y_m, (k'+x')/y_{m'}) dx dx' \\ &\quad + \sum_{m \geq 1} \sum_{k \neq k' \in \mathbb{Z}} \int_0^1 f((k+x)/y_m, (k'+x)/y_m) dx \\ &= \left( \sum_{m \neq m' \geq 1} y_m y_{m'} \right) \int_{\mathbb{R}^2} f(x, x') dx dx' \\ &\quad + \sum_{m \geq 1} \sum_{a \in \mathbb{Z} \setminus \{0\}} y_m \int_{\mathbb{R}} f(x, x + a/y_m) dx. \end{aligned}$$

The expectation of the first term of (5.1) is equal to

$$\left( 1 - \mathbb{E} \left[ \sum_{m \geq 1} y_m^2 \right] \right) \int_{\mathbb{R}^2} f(x, x') dx dx' = \frac{\theta}{\theta + 1} \int_{\mathbb{R}^2} f(x, x') dx dx'.$$

In order to compute the expectation of the second term, let us fix  $m$  and  $a$ , and let us denote by  $d_m$  the density of the probability distribution of  $y_m$  (recall that  $d_m(x) = 0$  for all  $x \notin [0, 1]$ ). One has:

$$\begin{aligned} \mathbb{E}[y_m f(x, x + a/y_m)] &= \int_0^1 t d_m(t) f(x, x + a/t) dt \\ &= \int_{\mathbb{R}} \frac{a^2}{|u|^3} d_m(a/u) f(x, x + u) du \end{aligned}$$

and then:

$$\mathbb{E} \left[ y_m \int_{\mathbb{R}} f(x, x + a/y_m) dx \right] = \int_{\mathbb{R}^2} \frac{a^2}{|x' - x|^3} d_m(a/(x' - x)) f(x, x') dx dx'.$$

Finally, the expectation of the second term of (5.1) is:

$$\sum_{a \in \mathbb{Z} \setminus \{0\}} \sum_{m \geq 1} \int_{\mathbb{R}^2} \frac{a^2}{|x' - x|^3} d_m(a/(x' - x)) f(x, x') dx dx'$$

which proves Proposition 5.6 with

$$\phi_\theta(x) = \frac{\theta}{\theta + 1} + \frac{1}{|x|^3} \sum_{a \geq 1} a^2 \sum_{m \geq 1} d_m(a/|x|).$$

Now, for all continuous functions  $g$  from  $[0, 1]$  to  $\mathbb{R}_+$ , one has:

$$\mathbb{E} \left[ \sum_{m \geq 1} y_m g(y_m) \right] = \mathbb{E}[g(y_M)],$$

where conditionally on  $(y_m)_{m \geq 1}$ , the random index  $M$  is chosen in a size-biased way, *i.e.*  $M = m$  with probability  $y_m$ . By classical properties of GEM and Poisson-Dirichlet processes,  $y_M$  is equal to  $y_1$  in distribution, and then its density at  $x \in [0, 1]$  is  $\theta(1-x)^{\theta-1}$ . Hence, one deduces that

$$\int_0^1 xg(x) \sum_{m \geq 1} d_m(x) = \theta \int_0^1 (1-x)^{\theta-1} g(x) dx,$$

and then, for almost every  $x \in [0, 1]$ ,

$$\sum_{m \geq 1} d_m(x) = \frac{\theta(1-x)^{\theta-1}}{x},$$

which implies Proposition 5.6. □

## 5.2. The smallest eigenangle

Another interesting problem about the point process associated with  $\tau_N(M_N)$  is the estimation of its smallest positive point. This point corresponds (after scaling the eigenangle by  $N$ ) to the first eigenvalue of  $M_N$  obtained by starting from 1 and by turning counterclockwise on the unit circle. Its distribution has a limit by the following result:

**PROPOSITION 5.7.** — *With the notation above, the smallest positive point corresponding to the random measure  $\bar{\tau}_N$  tends a.s. to the smallest positive point corresponding to  $\bar{\tau}_\infty$ .*

*Proof.* — One has, for all  $m \geq 1$ ,

$$\chi_m/y_{N,m} \xrightarrow{N \rightarrow \infty} \chi_m/y_m,$$

if by convention,  $\chi_m/y_{N,m} = +\infty$  for  $y_{N,m} = 0$ . One deduces that for all  $m_0 \geq 1$ ,

$$\begin{aligned} \limsup_{N \rightarrow \infty} \inf \{ \chi_m/y_{N,m}, m \geq 1 \} &\leq \limsup_{N \rightarrow \infty} \inf \{ \chi_m/y_{N,m}, 1 \leq m \leq m_0 \} \\ &= \inf \{ \chi_m/y_m, 1 \leq m \leq m_0 \} \end{aligned}$$

and then, by taking  $m_0 \rightarrow \infty$ :

$$\limsup_{N \rightarrow \infty} \inf \{ \chi_m/y_{N,m}, m \geq 1 \} \leq \inf \{ \chi_m/y_m, m \geq 1 \}.$$

On the other hand, for all  $A > 0$ , there exists a.s.  $m_1 \geq 1$  such that  $\chi_m/s_m \geq A$  for all  $m \geq m_1$ , which implies that  $\chi_m/y_{N,m} \geq A$  and  $\chi_m/y_m \geq A$ . Consequently

$$\begin{aligned} \liminf_{N \rightarrow \infty} \inf \{ \chi_m/y_{N,m}, m \geq 1 \} &\geq A \wedge \liminf_{N \rightarrow \infty} \inf \{ \chi_m/y_{N,m}, 1 \leq m \leq m_1 \} \\ &\geq A \wedge \inf \{ \chi_m/y_m, 1 \leq m \leq m_1 \} \\ &\geq A \wedge \inf \{ \chi_m/y_m, m \geq 1 \}. \end{aligned}$$

By taking  $A \rightarrow \infty$ , one obtains:

$$\liminf_{N \rightarrow \infty} \inf \{ \chi_m/y_{N,m}, m \geq 1 \} \geq \inf \{ \chi_m/y_m, m \geq 1 \},$$

and finally,

$$\inf \{ \chi_m/y_{N,m}, m \geq 1 \} \xrightarrow{N \rightarrow \infty} \inf \{ \chi_m/y_m, m \geq 1 \},$$

which proves Proposition 5.7. □

One immediately deduces the following:

**COROLLARY 5.8.** — *The smallest positive point of the random measure  $\tau_N(M_N)$  converges in distribution to the smallest positive point of  $\tau_\infty(\theta, \mathcal{L})$ .*

At this stage, one naturally seeks for the explicit computation of the distribution of the smallest positive point of  $\tau_\infty(\theta, \mathcal{L})$ . We can remark the similarity between this process and the limit point process (a determinantal process with sine kernel) obtained from the scaled eigenangles (with the same scaling by  $N$ ) of a random unitary matrix following the Haar measure. For the Haar measure, the law of the smallest positive point satisfies a certain Painlevé differential equation (see for instance [1] for more details and references). In our case, one can also obtain this probability distribution as a solution of some integral equation. More precisely, one has the following result:



PROPOSITION 5.9. — For all  $x \geq 0$ , let  $G(x)$  be the probability that the point process  $\tau_\infty(\theta, \mathcal{L})$  has no point in the interval  $(0, x)$ , and for all  $x \in \mathbb{R}$ , let us set:

$$H(x) := \mathbb{1}_{x>0} x^{\theta-1} G(x).$$

Then  $H$  is integrable and satisfies the following equation:

$$(5.2) \quad xH(x) = \theta \int_0^1 (1-y)H(x-y)dy.$$

Moreover, if the Fourier transform  $\widehat{H}$  of  $H$  is given by

$$\widehat{H}(\lambda) = \int_{-\infty}^{\infty} e^{-i\lambda x} H(x) dx,$$

then it satisfies the equation:

$$(5.3) \quad \widehat{H}(\lambda) = \widehat{H}(0) \exp\left(-i\theta \int_0^\lambda \frac{1 - e^{-i\mu} - i\mu}{\mu^2} d\mu\right),$$

for all  $\lambda \in \mathbb{R}$ .

*Proof.* — The probability  $G(x)$  can be expressed as follows:

$$G(x) = \mathbb{P} \left[ \inf_{m \geq 1} \{\chi_m / y_m\} \geq x \right],$$

where we recall that  $(y_m)_{m \geq 1}$  is a GEM process of parameter  $\theta$ . Conditionally on  $y_1$ , the sequence  $(y'_m := y_{m+1} / (1 - y_1))_{m \geq 1}$  is also a GEM process of parameter  $\theta$ . Therefore,

$$\begin{aligned} \mathbb{P} \left[ \inf_{m \geq 2} \{\chi_m / y_m\} \geq x \mid \chi_1, y_1 \right] &= \mathbb{P} \left[ \inf_{m \geq 1} \{\chi_{m+1} / y'_m\} \geq x(1 - y_1) \mid \chi_1, y_1 \right] \\ &= G(x(1 - y_1)). \end{aligned}$$

By taking the expectation, one obtains:

$$\begin{aligned} G(x) &= \mathbb{E} \left[ \mathbb{1}_{\chi_1 / y_1 \geq x} G(x(1 - y_1)) \right] \\ &= \mathbb{E} \left[ (1 - xy_1)_+ G(x(1 - y_1)) \right] \\ &= \int_0^{1 \wedge 1/x} \theta(1 - y)^{\theta-1} (1 - yx) G(x(1 - y)) dy \end{aligned}$$

since the law of  $y_1$  has density  $\theta(1 - y)^{\theta-1}$  with respect to Lebesgue measure. Hence:

$$G(x) = \frac{\theta}{x^\theta} \int_0^1 (x - y)^{\theta-1} (1 - y) G(x - y) dy$$

where by convention, we set  $G(y) = 0$  for all strictly negative  $y$ . This implies the equation (5.2) in Proposition 5.9. Note that for all  $x \geq 0$ ,

$$H(x) \leq x^{\theta-1}.$$

Moreover, for  $x \geq 2$ ,  $0 \leq y \leq 1$ :

$$\begin{aligned} H(x-y) &= (x-y)^{\theta-1} G(x-y) \leq (x-y)^{\theta-1} G(x-1) \\ &\leq (x-1)^{\theta-1} 2^{(\theta-1)+} G(x-1) \\ &\leq 2^{(\theta-1)+} H(x-1), \end{aligned}$$

and by (5.2),

$$H(x) \leq \frac{2^{(\theta-1)+\theta}}{2x} H(x-1).$$

Hence, at infinity,  $H$  decreases faster than exponentially, and  $H$ ,  $x \mapsto xH(x)$  are in  $L^1$ . One deduces that  $\hat{H}$  is well-defined and differentiable. By (5.2) one has for all  $\lambda \in \mathbb{R}$ :

$$\hat{H}'(\lambda) = -i\theta \hat{H}(\lambda) \hat{K}(\lambda),$$

where  $K$  is the function defined by  $K(y) = (1-y)\mathbb{1}_{0 \leq y \leq 1}$ . Therefore:

$$\hat{H}(\lambda) = \hat{H}(0) \exp\left(-i\theta \int_0^\lambda \hat{K}(\mu) d\mu\right),$$

which implies 5.3. □

## BIBLIOGRAPHY

- [1] G. W. ANDERSON, A. GUIONNET & O. ZEITOUNI, *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University Press, Cambridge, 2010, xiv+492 pages.
- [2] R. ARRATIA, A. D. BARBOUR & S. TAVARÉ, *Logarithmic combinatorial structures: a probabilistic approach*, EMS Monographs in Mathematics, European Mathematical Society (EMS), Zürich, 2003, xii+363 pages.
- [3] 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, x+277 pages.
- [4] P.-O. DEHAYE & D. ZEINDLER, "On averages of randomized class functions on the symmetric group and their asymptotics", <http://arxiv.org/pdf/0911.4038>, 2009.
- [5] P. DIACONIS, "Patterns in eigenvalues: the 70th Josiah Willard Gibbs lecture", *Bull. Amer. Math. Soc. (N.S.)* **40** (2003), no. 2, p. 155-178.
- [6] P. DIACONIS & M. SHAHSHAHANI, "On the eigenvalues of random matrices", *J. Appl. Probab.* **31A** (1994), p. 49-62, Studies in applied probability.
- [7] S. N. EVANS, "Eigenvalues of random wreath products", *Electron. J. Probab.* **7** (2002), p. 1-15.
- [8] B. M. HAMBLY, P. KEEVASH, N. O'CONNELL & D. STARK, "The characteristic polynomial of a random permutation matrix", *Stochastic Process. Appl.* **90** (2000), no. 2, p. 335-346.
- [9] S. KEROV, G. OLSHANSKI & A. VERSHIK, "Harmonic analysis on the infinite symmetric group. A deformation of the regular representation", *C. R. Acad. Sci. Paris Sér. I Math.* **316** (1993), no. 8, p. 773-778.
- [10] M. L. MEHTA, *Random matrices*, third ed., Pure and Applied Mathematics (Amsterdam), vol. 142, Elsevier/Academic Press, Amsterdam, 2004, xviii+688 pages.

- [11] F. MEZZADRI & N.-C. E. SNAITH, *Recent perspectives in random matrix theory and number theory*, London Mathematical Society Lecture Note Series, vol. 322, Cambridge University Press, 2005.
- [12] G. OLSHANSKI, *Asymptotic combinatorics with applications to mathematical physics*, Lecture Notes in Mathematics, vol. 1815, Springer, 2003.
- [13] J. PITMAN, *Combinatorial stochastic processes*, Lecture Notes in Mathematics, vol. 1875, Springer-Verlag, Berlin, 2006, Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard, x+256 pages.
- [14] N. V. TSILEVICH, “Distribution of cycle lengths of infinite permutations”, *J. Math. Sci.* **87** (1997), no. 6, p. 4072-4081.
- [15] ———, “Stationary measures on the space of virtual permutations for an action of the infinite symmetric group”, PDMI Preprint, 1998.
- [16] K. WIEAND, “Eigenvalue distributions of random permutation matrices”, *Ann. Probab.* **28** (2000), no. 4, p. 1563-1587.
- [17] ———, “Permutation matrices, wreath products, and the distribution of eigenvalues”, *J. Theoret. Probab.* **16** (2003), no. 3, p. 599-623.
- [18] D. ZEINDLER, “Permutation matrices and the moments of their characteristic polynomial”, *Electron. J. Probab.* **15** (2010), no. 34, p. 1092-1118.

Manuscrit reçu le 6 décembre 2010,  
accepté le 15 mars 2012.

Joseph NAJNUDEL & Ashkan NIKEGHBALI  
Universität Zürich  
Institut für Mathematik  
Winterthurerstrasse 190  
8057-Zürich( Switzerland)  
joseph.najnudel@math.uzh.ch  
ashkan.nikeghbali@math.uzh.ch