



HAL
open science

Asymptotic behavior of some statistics in Ewens random permutations

Valentin Féray

► **To cite this version:**

Valentin Féray. Asymptotic behavior of some statistics in Ewens random permutations. 2012. hal-00658364v1

HAL Id: hal-00658364

<https://hal.science/hal-00658364v1>

Preprint submitted on 10 Jan 2012 (v1), last revised 13 Aug 2013 (v5)

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

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

ASYMPTOTIC BEHAVIOR OF SOME STATISTICS IN EWENS RANDOM PERMUTATIONS

VALENTIN FÉRAY

ABSTRACT. The purpose of this article is to present a general method to find limiting laws for some renormalized statistics on random permutations. The model of random permutations considered here is Ewens sampling model, which generalizes uniform random permutations. Under this model, we describe the asymptotic behavior of some statistics, including the number of occurrences of any dashed patterns. Our approach is based on the method of moments and relies on the following intuition: two events involving the images of different integers are *almost* independent.

1. INTRODUCTION

1.1. **Background.** Permutations are one of the most classical objects in enumerative combinatorics. A lot of statistics on permutations have been widely studied: total number of cycles, number of cycles of a given length, of descents, inversions, excedances or more recently, of occurrences of a given (generalized) pattern... A classical question in enumerative combinatorics consists in computing the (multivariate) generating series of permutations with respect to some of these statistics. It would be impossible to do an exhaustive list of the different statistics which have been considered and the results which have been obtained.

A probabilistic point of view on the topic raises other questions. Let us consider, for each N , a probability measure μ_N of permutations of size N . Then any statistic above can be interpreted as a sequence of random variables $(X_N)_{N \geq 1}$. The natural question is now: what is the asymptotic behavior (possibly after normalization) of $(X_N)_{N \geq 1}$?

The simplest model of random permutations is of course the uniform random permutations (for each N , μ_N is the uniform distribution on the symmetric group S_N). A generalization of this model has been introduced by W.J. Ewens in the context of population dynamics [13]. It is defined by

$$(1) \quad \mu_N(\{\sigma\}) = \frac{(1 + \theta)^{\#(\sigma)}}{(\theta + 1)(\theta + 2) \cdots (\theta + N)},$$

where $\theta > -1$ is a fixed real parameter and $\#(\sigma)$ stands for the number of cycles of the permutation σ . Of course, when $\theta = 0$, we recover the uniform distribution. From now on, we will allow ourselves a small abuse of language and use the

1991 *Mathematics Subject Classification.* 05A16, 05A05.

Key words and phrases. random permutations, *SSEP*, cumulants, dashed patterns.

Partially supported by ANR grant PSYCO.

expression *Ewens random permutation* for a random permutation distributed with Ewens measure.

Random permutations, either with uniform or Ewens distribution, are well-studied objects. A lot of examples in the seminal book of P. Flajolet and R. Sedgewick deal with the asymptotic behavior of some statistics on uniform random permutations [15, Examples IX.3, IX.4, IX.5, IX.9]. A survey of important techniques and results in the area can be found in the book [1]. Most of them involve explicit formulae for generating series or algorithms to generate a random permutation.

The purpose of this article is to introduce a new general approach to this family of problems, based on the method of moments.

1.2. Motivating examples. Let us begin by describing a few examples of results, which suggest that a more uniform and intuitive approach could be found.

Number of cycles of a given length p . Let $\Gamma_p^{(N)}$ be the random variable given by the number of cycles of length p in an Ewens random permutation σ in S_N . The asymptotic distribution of $\Gamma_p^{(N)}$ has been studied by V.L. Goncharov [17] and V.F. Kolchin [19] in the case of uniform measure and by G.A. Watterson [28, Theorem 5] for the framework of a general Ewens distribution (see also [1, Theorem 5.1]).

Theorem 1.1 ([28]). *Let p be a positive integer. When N tends to infinity, $\Gamma_p^{(N)}$ converges in distribution towards a Poisson law of parameter $(1 + \theta)/p$. Moreover, the sequences of random variables $(\Gamma_{p'}^{(N)})_{N \geq 1}$ for $p' \leq p$ are asymptotically independent.*

Let us give an intuitive (but false) explanation of the first part of the result, assuming that some non-independent variables are independent.

If i_1, \dots, i_p is a list of pairwise distinct integers between 1 and N such that its minimum is i_1 (there are $(N)_p/p$ such lists, where $(N)_p$ is the usual notation for the falling factorial $(N)_p = N(N-1) \dots (N-p+1)$), we define

$$(2) \quad B_{(i_1, \dots, i_p)}^{c, N}(\sigma) = \begin{cases} 1 & \text{if } (i_1 \dots i_p) \text{ is a cycle of } \sigma; \\ 0 & \text{otherwise.} \end{cases}$$

Each $B_{(i_1, \dots, i_p)}^{c, N}$ is distributed according to a Bernoulli law of parameter $(1+\theta)/(N)_p$ (see Lemma 2.1). These variables are *not* independent. Nevertheless the sum $\Gamma_p^{(N)}$ of these $(N)_p/p$ Bernoulli variables of parameter $(1 + \theta)/(N)_p$ converges in distribution towards a Poisson variable of parameter $(1 + \theta)/p$.

Excedances. A (weak) excedance of a permutation σ in S_N is an integer i such that $\sigma(i) \geq i$. Let $B_i^{\text{ex}, N}$ be the random variable defined by:

$$B_i^{\text{ex}, N}(\sigma) = \begin{cases} 0 & \text{if } \sigma(i) < i; \\ 1 & \text{if } \sigma(i) \geq i. \end{cases}$$

When σ is a Ewens random permutation, this random variable is distributed according to a Bernoulli law of parameter $\frac{i+1+\theta}{N+\theta}$ (see Lemma 2.1).

Let x be a fixed real number in $[0; 1]$ and σ a permutation of size N . When Nx is an integer, we define

$$F_\sigma^{(N)}(x) := \frac{\sum_{i=1}^{Nx} B_i^{\text{ex},N}(\sigma)}{N}$$

and we extend the function $F_\sigma^{(N)}$ by linearity between the points i/N and $(i+1)/N$ (for $1 \leq i \leq N-1$). In paragraphs 5.1 and 5.2, we explain why we are interested in this quantity: it is related to a statistical physics model, the *symmetric simple exclusion process* (SSEP), and to permutation tableaux, some combinatorial objects which have been intensively studied in the last few years.

We show the following.

Theorem 1.2. *Let x be a real number between 0 and 1. Then, almost surely,*

$$\lim_{N \rightarrow \infty} F_\sigma^{(N)}(x) = \frac{1 - (1-x)^2}{2}.$$

Moreover, if we define the rescaled fluctuations

$$Z_\sigma^{(N)}(x) := \sqrt{N} \left(F_\sigma^{(N)}(x) - \mathbb{E}(F_\sigma^{(N)}(x)) \right),$$

then, for any x_1, \dots, x_r , the vector $(Z_\sigma^{(N)}(x_1), \dots, Z_\sigma^{(N)}(x_r))$ converge towards a Gaussian vector $(G(x_1), \dots, G(x_r))$ of covariance matrix $(K(x_i, x_j))_{1 \leq i, j \leq r}$, for some explicit function K (see paragraph 5.4).

If $i \neq j$, the variables $B_i^{\text{ex},N}$ and $B_j^{\text{ex},N}$ are *not* independent (their covariance is computed explicitly in paragraph 5.4). Nevertheless, the limit and the Gaussian fluctuations correspond to what would happen with independent variables (only the actual value of the covariance matrix $K(x_i, x_j)$ is different).

With this formulation, Theorem 1.2 is new, but the first part is quite easy while the second is a consequence of [12, Appendix A] (see section 5).

Adjacencies. We consider here only uniform random permutations, that is the case $\theta = 0$. An adjacency of a permutation σ in S_N is an integer i such that $\sigma(i+1) = \sigma(i) \pm 1$. As above, we introduce the random variable $B_i^{\text{ad},N}$ which takes value 1 if i is an adjacency and 0 otherwise. Then $B_i^{\text{ad},N}$ is distributed according to a Bernoulli law of parameter $\frac{2}{N}$. An easy computation shows that they are *not* independent.

We are interested in the total number of adjacencies in σ , that is the random variable on S_N defined by $A^{(N)} = \sum_{i=1}^{N-1} B_i^{\text{ad},N}$.

Theorem 1.3 ([30]). *$A^{(N)}$ converges in distribution towards a Poisson variable of parameter 2.*

This result first appeared in papers of J. Wolfowitz and I. Kaplansky [30, 18] and was rediscovered recently in the context of genomics (see [31] and also [8,

Theorem 10]). Note that it corresponds exactly to what would have been obtained if the variables $B_i^{\text{ad},N}$ were independent.

Of course, as the Bernoulli random variable considered in each of these examples are *not* independent, the explanations given for these results are not rigorous proofs. Nevertheless, the considered events involve (most of the time) the images of different integers by the permutation σ . Therefore, speaking informally, they are *almost* independent. The main lemma of this paper is a precise statement of this *almost* independence, that is an upper bound on joint cumulants. This result allows us to give new proofs and generalizations of the three results presented above.

1.3. Set partitions. The combinatorics of set partitions is central in the theory of cumulants (as explained below) and will be important in this article.

A *set partition* of a set S is a (non-ordered) family of non-empty disjoint subsets of S (called parts of the partition), whose union is S .

Denote $\mathcal{P}(S)$ the set of set partitions of a given set S . Then $\mathcal{P}(S)$ may be endowed with a natural partial order: the *refinement* order. We say that π is *finer* than π' or π' *coarser* than π (and denote $\pi \leq \pi'$) if every part of π is included in a part of π' .

Endowed with this order, $\mathcal{P}(S)$ is a complete lattice, which means that each family F of set partitions admits a join (the finest set partition which is coarser than all set partitions in F , denoted with \vee) and a meet (the coarsest set partition which is finer than all set partitions in F , denoted with \wedge). In particular, there is a maximal element $\{S\}$ (the partition in only one part) and a minimal element $\{\{x\}, x \in S\}$ (the partition in singletons).

Moreover, this lattice is ranked: the rank $\text{rk}(\pi)$ of a set partition π is $|S| - \#(\pi)$, where $\#(\pi)$ denotes the number of parts of π . The rank is compatible with the lattice structure in the following sense: for all set partitions π and π' ,

$$(3) \quad \text{rk}(\pi \vee \pi') \leq \text{rk}(\pi) + \text{rk}(\pi').$$

Lastly, denote μ the Möbius function of the partition lattice $\mathcal{P}(S)$. In this paper, we only use evaluations of μ at pairs $(\pi, \{S\})$ (that is the second argument is the maximum element of $\mathcal{P}(S)$). In this case, the value of the Möbius function is given by:

$$(4) \quad \mu(\pi, \{S\}) = (-1)^{\#(\pi)-1} (\#(\pi) - 1)!.$$

1.4. Cumulants. Joint cumulants of random variables form a classical tool to quantify correlation. They generalize the notion of covariance. We present in this paragraph their definition and basic properties.

Definition.

They are defined as follows: if X_1, \dots, X_ℓ are random variables on the same probability space (denote \mathbb{E} the expectation on this space), then

$$(5) \quad \kappa(X_1, \dots, X_\ell) = [t_1 \dots t_\ell] \ln \left(\mathbb{E} \left(\exp(t_1 X_1 + \dots + t_\ell X_\ell) \right) \right).$$

As usual, $[t_1 \dots t_\ell]F$ stands for the coefficient of $t_1 \dots t_\ell$ in the series expansion of F in positive powers of t_1, \dots, t_ℓ . Note that joint cumulants are multilinear functions. In the case where all the X_i are equal, we recover the ℓ -th cumulant $\kappa_\ell(X)$ of a single variable [14].

Joint cumulants can be expressed in terms of joint moments, and *vice-versa* [26]. Denote $[\ell]$ the set $\{1, \dots, \ell\}$.

$$(6) \quad \mathbb{E}(X_1 \cdots X_\ell) = \sum_{\pi \in \mathcal{P}([\ell])} \prod_{C \in \pi} \kappa(X_i; i \in C);$$

$$(7) \quad \kappa(X_1, \dots, X_\ell) = \sum_{\pi \in \mathcal{P}([\ell])} \mu(\pi, \{[\ell]\}) \prod_{C \in \pi} \mathbb{E} \left(\prod_{i \in C} X_i \right).$$

Recall that $\mu(\pi, \{[\ell]\})$ has an explicit expression given by Equation (4). For example the joint cumulants of one or two variables are simply the mean of a single random variable ($\kappa(X_1) = \mathbb{E}(X_1)$) and the covariance of a couple of random variables ($\kappa(X_1, X_2) = \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2)$). For three variables, one has

$$\begin{aligned} \kappa(X_1, X_2, X_3) &= \mathbb{E}(X_1 X_2 X_3) - \mathbb{E}(X_1 X_2)\mathbb{E}(X_3) - \mathbb{E}(X_1 X_3)\mathbb{E}(X_2) \\ &\quad - \mathbb{E}(X_2 X_3)\mathbb{E}(X_1) + 2\mathbb{E}(X_1)\mathbb{E}(X_2)\mathbb{E}(X_3). \end{aligned}$$

Cumulants of independent random variables.

An interesting property of cumulants is the following: if the set of variables $\{X_i, 1 \leq i \leq \ell\}$ can be split into two sets $\{X_i, i \in A\}$ and $\{X_i, i \in B\}$ (with $A \sqcup B = [\ell]$) such that the variables from the first set are independent from the variables from the second, then

$$\begin{aligned} \kappa(X_1, \dots, X_\ell) &= [t_1 \dots t_\ell] \ln \left(\mathbb{E} \left(\exp \left(\sum_{i \in A} t_i X_i \right) \right) \right) \\ &\quad + [t_1 \dots t_\ell] \ln \left(\mathbb{E} \left(\exp \left(\sum_{i \in B} t_i X_i \right) \right) \right) = 0. \end{aligned}$$

Informally speaking, this means that joint cumulants of independent random variables vanish. Therefore, joint cumulants can be seen as a quantification of the dependence of random variables.

Convergence in distribution using cumulants.

Consider now m sequences of random variables: $(X_n^{(i)})_{n \geq 1}$ for $i \in [m]$. A consequence of Equations (6) and (7) is that the convergence of all joint cumulants

$$\kappa \left(X_n^{(i_1)}, \dots, X_n^{(i_\ell)} \right); \ell \geq 1, 1 \leq i_1, \dots, i_\ell \leq m$$

is equivalent to the convergence of all joint moments

$$\mathbb{E} \left(X_n^{(i_1)} \cdots X_n^{(i_\ell)} \right); \ell \geq 1, 1 \leq i_1, \dots, i_\ell \leq m.$$

In particular, if $Y^{(1)}, \dots, Y^{(m)}$ are random variables such that the law of the m -tuple $(Y^{(1)}, \dots, Y^{(m)})$ is entirely determined by its joint moments, then the two

following statements are equivalent (see [4, Theorem 30.2] for the same property in terms of moments).

- For any ℓ and any list i_1, \dots, i_ℓ in $[m]$,

$$\lim_{n \rightarrow \infty} \kappa \left(X_n^{(i_1)}, \dots, X_n^{(i_\ell)} \right) = \kappa \left(Y^{(i_1)}, \dots, Y^{(i_\ell)} \right).$$

- The sequence of m -tuples $(X_n^{(1)}, \dots, X_n^{(m)})$ converges in distribution towards $(Y^{(1)}, \dots, Y^{(m)})$.

Recall that Gaussian and Poisson variables are determined by their moments, see *e.g.* the criterion [4, Theorem 30.1]. Hence, cumulants can be used to prove convergence in distribution towards Gaussian or Poisson variables, such as the results of the previous paragraph.

1.5. The main lemma. From now on, N is a positive integer and σ a random Ewens permutation in S_N .

If i and s are two integers in $[N]$, we consider the Bernoulli variable $B_{i,s}^{(N)}$ which takes value 1 if and only if $\sigma(i) = s$. It is the characteristic function of one of the simplest event we can think of in the symmetric group.

Our main lemma is a bound on joint cumulants of products of such variables. To state it, we introduce the following notations. Consider two lists of positive integers of the same length $\mathbf{i} = (i_1, \dots, i_r)$ and $\mathbf{s} = (s_1, \dots, s_r)$ and define the graphs $G_1(\mathbf{i}, \mathbf{s})$ and $G_2(\mathbf{i}, \mathbf{s})$ as follows:

- the vertex set of $G_1(\mathbf{i}, \mathbf{s})$ is $[r]$ and j and h are linked in $G_1(\mathbf{i}, \mathbf{s})$ if and only if $i_j = i_h$ and $s_j = s_h$.
- the vertex set of $G_2(\mathbf{i}, \mathbf{s})$ is also $[r]$ and j and h are linked in $G_2(\mathbf{i}, \mathbf{s})$ if and only if $\{i_j, s_j\} \cap \{i_h, s_h\} \neq \emptyset$.

The connected components of a graph G form a set partition of its vertex set that we denote $\text{CC}(G)$. In particular, $\#(\text{CC}(G))$ is the number of connected components of G .

Theorem 1.4 (main lemma). *Fix a positive integer r . There exists a constant C_r , depending on r , such that for any set partition $\tau = (\tau_1, \dots, \tau_\ell)$ of $[r]$, any $N \geq 1$ and lists $\mathbf{i} = (i_1, \dots, i_r)$ and $\mathbf{s} = (s_1, \dots, s_r)$ of integers in $[N]$, one has:*

$$(8) \quad \left| \kappa \left(\prod_{j \in \tau_1} B_{i_j, s_j}^{(N)}, \dots, \prod_{j \in \tau_\ell} B_{i_j, s_j}^{(N)} \right) \right| \leq C_r N^{-\#(\text{CC}(G_1(\mathbf{i}, \mathbf{s}))) - \#(\text{CC}(G_2(\mathbf{i}, \mathbf{s})) \vee \tau) + 1}.$$

Note that the integer $\#(\text{CC}(G_1(\mathbf{i}, \mathbf{s})))$ is the number of *different* couples (i_j, s_j) . The second quantity involved in the theorem $\#(\text{CC}(G_2(\mathbf{i}, \mathbf{s})) \vee \tau)$ does not have a similar interpretation. However, it admits an equivalent description. Consider the graph G_2' , whose vertex set is indexed by the parts of τ and in which τ_h is linked with $\tau_{h'}$ if

$$\left(\{i_j; j \in \tau_h\} \cup \{s_j; j \in \tau_h\} \right) \cap \left(\{i_j; j \in \tau_{h'}\} \cup \{s_j; j \in \tau_{h'}\} \right) \neq \emptyset.$$

Then $\#(\text{CC}(G_2(\mathbf{i}, \mathbf{s})) \vee \tau)$ is the number of connected components of G'_2 .

As an example, let us consider the case where the entries in the lists \mathbf{i} and \mathbf{s} are pairwise distinct. In this case, the joint moment of products of $B_{i,s}^{(N)}$ is simply $1/(N+\theta)_a$, where a is the number of factors (the case $\theta = 0$ is obvious, the general case is explained in Lemma 2.1). Thus, if we denote $a_j = |\tau_j|$, the left-hand side is the following explicit rational function

$$\sum_{\pi \in \mathcal{P}([\ell])} \mu(\pi, [\ell]) \prod_{\pi_i \in \pi} \frac{1}{(N+\theta)_{\sum_{j \in \pi_i} a_j}}.$$

Note that each summand has degree $-\sum_{j=1}^r a_j = -\ell$. According to our theorem, the sum has degree at most $-\ell - r + 1$, which means that many simplifications are happening (they are not at all trivial to explain!). This reflects the fact that the variables $B_{i_j, s_j}^{(N)}$ are very weakly correlated.

Remark 1.5. It is worth noticing that our proof of the main lemma goes through a very general criterion for a family of sequences of random variables to have small cumulants: see Lemma 2.2.

1.6. Applications. Theorem 1.4 can be used to give new proofs of Theorems 1.1, 1.2 and 1.3. Moreover, we get an extension of Theorem 1.3 to any value of the parameter θ . The result is the same (the limit law is a Poisson law of parameter 2), but the usual approach *via* generating series for the uniform case seems hard to extend to this framework: it would need an explicit form for the bivariate generating series of permutations with respect to their number of cycles and of excedances and such a result does not exist to our knowledge.

We must confess that our proofs of these results are quite technical. However, an important part of the difficulty is contained in the proof of Theorem 1.4 and hence must not be done again for each application. Moreover, these proofs are natural in the following sense: they are based on the idea that, when σ is a uniform random permutation, $\sigma(i)$ and $\sigma(j)$ are almost independent. Besides, although the problems may seem quite different (in particular the limit law is not always the same), these proofs all follow roughly the same guidelines.

To give more evidence that our approach is quite general, we study the number of occurrences of *dashed patterns*. This notion has been introduced¹ in 2000 by E. Babson and E. Steingrímsson, because it gives a general setting which includes a lot of usual statistics of permutations [3].

Thanks to our main lemma, we describe the second order asymptotics of the number of occurrences of any given dashed pattern in a random Ewens permutation.

Theorem 1.6. *Let (τ, X) be a dashed pattern of size p (see definition 6.3) and σ_N a sequence of random permutations, each σ_N being of size N distributed with*

¹In the paper of Babson and Steingrímsson, they are called generalized patterns. But, as some more general *generalized patterns* have been introduced since (see next paragraph), we prefer to use *dashed patterns*.

Ewens measure. We denote $q = |X|$. Then, $\frac{O_{\tau,X}^{(N)}(\sigma_N)}{N^{p-q}}$, that is the renormalized number of occurrences of (τ, X) , tends almost surely towards $\frac{1}{p!(p-q)!}$. Besides, one has the following central limit theorem:

$$Z_{(X,\tau)}^{(N)} := \sqrt{N} \left(\frac{O_{\tau,X}^{(N)}}{N^{p-q}} - \frac{1}{p!(p-q)!} \right) \rightarrow \mathcal{N}(0, V_{\tau,X}),$$

where the arrow denotes a convergence in distribution and $V_{\tau,X}$ is some nonnegative real number.

This theorem is proved in paragraph 6.3 using Theorem 1.4. Note that an approach with generating series would probably fail as it seems unlikely that there exists a general formula for the generating series of the number of occurrences of any given dashed pattern.

Unfortunately, we are not able to show in general that the constant $V_{\tau,X}$ is positive ($V_{\tau,X} = 0$ would mean that we have not chosen the good normalization). It is nevertheless possible to write $V_{\tau,X}$ as a signed sum of products of binomial coefficients and we have checked by computer the following conjecture for all patterns of size 8 or less.

Conjecture 1.7. *For any dashed pattern (τ, X) , one has $V_{\tau,X} > 0$.*

Remark 1.8. The asymptotic number of occurrences of some patterns (or *motifs*) have already been studied in the framework of random words, where results similar to Theorem 1.6 have been obtained [22, 16, 6]. Note that the methods used by these authors can *a priori* not be extended to the framework of permutation patterns for the following reasons. First, it is not known how to compute the generating functions of the number of occurrences of a given pattern in permutations. Second, except in the case $\theta = 0$ (uniform random permutations), the occurrences of patterns in disjoint places are not independent events (see remark 6.5 for details).

1.7. Future work. In addition to the conjecture above, we mention three directions for further research on the topic.

The notion of dashed patterns has been further extended to the notion of *generalized patterns* in a recent paper of M. Bousquet-Mélou, A. Claesson, M. Dukes and S. Kitaev [7, Section 2]. Unfortunately, we have not been able to obtain a general result for the asymptotic number of occurrences of generalized patterns. Finding such a result is, in the author's opinion, a challenging open problem. One could even consider a more general framework, see paragraph 6.4.

Another direction consists in refining our convergence results (speed of convergence, large deviations, local limit laws) by following the same guideline.

Finally, it is natural to wonder if the method can be extended to other measures and objects. The extension to colored permutations should be straightforward. But it would be interesting to obtain some result on other objects (for example other families of rook placements) with the same ideas.

1.8. Outline of the paper. The paper is organized as follows. In section 2, we prove our main lemma. Then, in section 3, we give two easy lemmas on connected components of graphs, which appear in all our applications. The three last sections are devoted to the different applications: section 4 for the cycles, section 5 for the excedances and finally, section 6 for the generalized patterns (including the adjacencies and the dashed patterns).

2. PROOF OF THE MAIN LEMMA

2.1. Joint moments. The first step of the proof consists in computing the joint moments of the family of random variables $(B_{i,s}^{(N)})_{1 \leq i,s \leq N}$.

Note that $(B_{i,s}^{(N)})^2 = B_{i,s}^{(N)}$, while $B_{i,s}^{(N)} B_{i,s'}^{(N)} = 0$ if $s \neq s'$ and $B_{i,s}^{(N)} B_{i',s}^{(N)} = 0$ if $i \neq i'$. Therefore, we can restrict ourselves to the computation of the joint moment $\mathbb{E} \left(B_{i_1, s_1}^{(N)} \cdots B_{i_r, s_r}^{(N)} \right)$, in the case where $\mathbf{i} = (i_1, \dots, i_r)$ and $\mathbf{s} = (s_1, \dots, s_r)$ are two lists of pairwise distinct indices (some entry in the list \mathbf{i} can be equal to an entry of \mathbf{s}).

We see these two lists as a *partial permutation*

$$\tilde{\sigma}_{\mathbf{i}, \mathbf{s}} = \begin{pmatrix} i_1 & \dots & i_r \\ s_1 & \dots & s_r \end{pmatrix},$$

which sends i_j to s_j . The notion of cycles of a permutation can be naturally extended to partial permutations: $(i_{j_1}, \dots, i_{j_\gamma})$ is a cycle of the partial permutation if $s_{j_1} = i_{j_2}$, $s_{j_2} = i_{j_3}$ and so on until $s_{j_\gamma} = i_{j_1}$. Note that a partial permutation does not necessarily have cycles. The number of cycles of $\tilde{\sigma}_{\mathbf{i}, \mathbf{s}}$ is denoted $\#(\tilde{\sigma}_{\mathbf{i}, \mathbf{s}})$.

The computation of $\mathbb{E} \left(B_{i_1, s_1}^{(N)} \cdots B_{i_r, s_r}^{(N)} \right)$ relies on two important properties of the Ewens measure. First, it is conjugacy-invariant. Second, a random sampling can be obtained inductively by the following procedure (see, *e.g.* [1, Example 2.19]).

Suppose that we have a permutation σ of size $N - 1$ taken with this distribution. Write it as a product of cycles and apply the following transformation.

- With probability $(1 + \theta)/(N + \theta)$, add N as a fixed point. More precisely, σ' is defined by:

$$\begin{cases} \sigma'(i) = \sigma(i) & \text{for } i < N; \\ \sigma'(N) = N. \end{cases}$$

- For each j , with probability $1/(N + \theta)$, add N just before j in its cycle. More precisely, σ' is defined by:

$$\begin{cases} \sigma'(i) = \sigma(i) & \text{for } i \neq \sigma^{-1}(j), N; \\ \sigma'(N) = j; \\ \sigma'(\sigma^{-1}(j)) = N. \end{cases}$$

Then σ' is a random permutation of S_N distributed with Ewens measure. Iterating this, one obtains a linear time and space algorithm to pick a random permutation distributed with Ewens measure.

Let us come back now to the computation of joint moments.

Lemma 2.1. *Let σ be a random permutation taken with Ewens distribution. Then one has*

$$\mathbb{E} \left(B_{i_1, s_1}^{(N)} \cdots B_{i_r, s_r}^{(N)} \right) = \frac{(1 + \theta)^{\#(\tilde{\sigma}_{\mathbf{i}, \mathbf{s}})}}{(N + \theta) \cdots (N + \theta - r + 1)}.$$

For example, the parameter of the Bernoulli variables $B_{i, s}^{(N)}$ are given by

$$\mathbb{E}(B_{i, s}^{(N)}) = \begin{cases} \frac{1 + \theta}{N + \theta} & \text{if } i = s; \\ \frac{1}{N + \theta} & \text{if } i \neq s. \end{cases}$$

Proof. As Ewens measure is constant on conjugacy classes of S_N , one can assume without loss of generality that $i_1 = N - r + 1, i_2 = N - r + 2, \dots, i_r = N$. Then permutations of S_N with $\sigma(i_j) = s_j$ are obtained in the previous algorithm as follows:

- Choose any permutation in S_{N-r} .
- For $1 \leq j \leq r$, add i_j in the place given by the following rule: if $s_j < i_j$, add i_j just before s_j in its cycle. Otherwise, look at $\tilde{\sigma}_{\mathbf{i}, \mathbf{s}}(i_j), \tilde{\sigma}_{\mathbf{i}, \mathbf{s}}^2(i_j)$ and so on until you find an element smaller than i_j and place i_j before it. If there is no such element, then i_j is a minimum of a cycle of $\tilde{\sigma}_{\mathbf{i}, \mathbf{s}}$. In this case, put it in a new cycle.

It is easy to check with the description of the construction of a permutation under Ewens measure that these choices of places happen with a probability

$$\frac{(1 + \theta)^{\#(\tilde{\sigma}_{\mathbf{i}, \mathbf{s}})}}{(N + \theta) \cdots (N - r + 1 + \theta)}. \quad \square$$

2.2. A general criterion for small cumulants. Let $A_1^{(N)}, \dots, A_\ell^{(N)}$ be ℓ sequences of random variables. We introduce the following notation for joint moments and cumulants of subsets of these variables: if $\Delta = \{j_1, \dots, j_h\}$ is a subset of $[\ell]$, then

$$M_{A, \Delta}^{(N)} = \mathbb{E} \left(A_{j_1}^{(N)} \cdots A_{j_h}^{(N)} \right), \quad \kappa_{A, \Delta}^{(N)} = \kappa \left(A_{j_1}^{(N)}, \dots, A_{j_h}^{(N)} \right).$$

Lemma 2.2. *Let $A_1^{(N)}, \dots, A_\ell^{(N)}$ be a list of sequences of random variables with normalized expectations, that is, for any N and j , $\mathbb{E}(A_j^{(N)}) = 1$. Then the following statements are equivalent:*

I. Quasi-factorization property: for any subset $\Delta \subseteq [\ell]$ of size at least 2, one has

$$(9) \quad \prod_{\delta \subset \Delta} \left(M_{A, \delta}^{(N)} \right)^{(-1)^{|\delta|}} = 1 + O(N^{-|\Delta|+1});$$

II. Small cumulant property: for any subset $\Delta \subseteq [\ell]$ of size at least 2, one has

$$(10) \quad \kappa_{A, \Delta}^{(N)} = O(N^{-|\Delta|+1}).$$

Proof. Let us consider the implication $I \Rightarrow II$. We denote $1 + T_{\Delta}^{(N)}$ the left-hand side of Equation (9) and assume that $T_{\Delta}^{(N)} = O(N^{-|\Delta|+1})$ for any $\Delta \subseteq [\ell]$ of size at least 2. The goal is to prove that $\kappa_{A, [\ell]}^{(N)} = O(N^{-\ell+1})$. Indeed, this corresponds to the case $\Delta = [\ell]$ of II , but the same proof will work for any $\Delta \subseteq [\ell]$.

Recall the relation between moments and cumulants (Equation (7)):

$$\kappa_{A, [\ell]}^{(N)} = \sum_{\pi \in \mathcal{P}([\ell])} \mu(\pi, \{[\ell]\}) \prod_{C \in \pi} M_{A, C}^{(N)}.$$

But joint moments can be expressed in terms of T :

$$M_{A, C}^{(N)} = \prod_{\substack{\Delta \subseteq C \\ |\Delta| \geq 2}} (1 + T_{\Delta}^{(N)}) = \sum_{\Delta_1, \dots, \Delta_m} T_{\Delta_1}^{(N)} \dots T_{\Delta_m}^{(N)},$$

where the sum runs over all finite lists of pairwise distinct (but not necessarily disjoint) subsets of C of size at least 2 (in particular, the length m of the list is not fixed). When we multiply this over all blocks C of a set partition π , we obtain the sum of $T_{\Delta_1}^{(N)} \dots T_{\Delta_m}^{(N)}$ over all lists of pairwise distinct subsets of $[\ell]$ of size at least 2 such that each Δ_i is contained in a block of π . In other terms, for each $i \in [m]$, π must be coarser than the partition $\Pi(\Delta_i)$, which, by definition, has Δ_i and singletons as blocks. Finally,

$$(11) \quad \kappa_{A, [\ell]}^{(N)} = \sum_{\substack{\Delta_1, \dots, \Delta_m \\ \text{pairwise distinct}}} T_{\Delta_1}^{(N)} \dots T_{\Delta_m}^{(N)} \left(\sum_{\substack{\pi \in \mathcal{P}([\ell]) \\ \text{for all } i, \pi \geq \Pi(\Delta_i)}} \mu(\pi, \{[\ell]\}) \right).$$

The condition on π can be rewritten as

$$\pi \geq \Pi(\Delta_1) \vee \dots \vee \Pi(\Delta_m).$$

Hence, by definition of the Möbius function, the sum in the parenthesis is equal to 0, unless $\Pi(\Delta_1) \vee \dots \vee \Pi(\Delta_m) = \{[\ell]\}$. On the one hand, by Equation (3), it may happen only if:

$$\sum_{i=1}^m \text{rk}(\Pi(\Delta_i)) = \sum_{i=1}^m (|\Delta_i| - 1) \geq \text{rk}([\ell]) = \ell - 1.$$

On the other hand, one has

$$T_{\Delta_1}^{(N)} \dots T_{\Delta_m}^{(N)} = O\left(N^{-\sum_{i=1}^m (|\Delta_i| - 1)}\right).$$

Hence only summands of order of magnitude $N^{-\ell+1}$ or less survive and one has

$$\kappa_{A, [\ell]}^{(N)} = O(N^{-\ell+1})$$

which is exactly what we wanted to prove.

Let us now consider the converse statement. We proceed by induction on ℓ and we assume that, for all ℓ' smaller than a given $\ell \geq 2$, the theorem holds.

Consider some sequences of random variables $A_1^{(N)}, \dots, A_\ell^{(N)}$ such that *II* holds. By induction hypothesis, one gets immediately that

$$\text{for all } \Delta \subsetneq [\ell], \prod_{\delta \subset \Delta} (M_{A,\delta}^{(N)})^{(-1)^{|\delta|}} = 1 + O(N^{-|\Delta|+1}).$$

Note that an immediate induction shows that the joint moment fulfills

$$\text{for all } \Delta \subsetneq [\ell], M_{A,\Delta}^{(N)} = O(1) \text{ and } (M_{A,\Delta}^{(N)})^{-1} = O(1).$$

It remains to prove that

$$\prod_{\Delta \subseteq [\ell]} (M_{A,\Delta}^{(N)})^{(-1)^{|\Delta|}} = 1 + O(N^{-\ell+1}).$$

Thanks to the estimate above for joint moment, this can be rewritten as

$$(12) \quad M_{A,[\ell]}^{(N)} = \prod_{\Delta \subsetneq [\ell]} (M_{A,\Delta}^{(N)})^{(-1)^{\ell-1-|\delta|}} + O(N^{-\ell+1}).$$

Consider ℓ sequences of random variables $B_1^{(N)}, \dots, B_\ell^{(N)}$ such that the equality $M_{B,\Delta}^{(N)} = M_{A,\Delta}^{(N)}$ holds for $\Delta \subsetneq [\ell]$ and such that Equation (12) is fulfilled when A is replaced by B (the reader may wonder whether such a family B exists; let us temporarily ignore this problem, which will be addressed in Remark 2.3). By definition, the family B of sequences of random variables fulfills condition *I* of the theorem and, hence, using the first part of the proof, has also property *II*. In particular:

$$\kappa_{B,[\ell]}^{(N)} = O(N^{-\ell+1}).$$

But, by hypothesis,

$$\kappa_{A,[\ell]}^{(N)} = O(N^{-\ell+1}).$$

As A and B have the same joint moment, except for $M_{A,[\ell]}^{(N)}$ and $M_{B,[\ell]}^{(N)}$, this implies that

$$M_{A,[\ell]}^{(N)} - M_{B,[\ell]}^{(N)} = \kappa_{A,[\ell]}^{(N)} - \kappa_{B,[\ell]}^{(N)} = O(N^{-\ell+1}).$$

But the family B fulfills Equation (12) and, hence, so does family A . \square

Remark 2.3. Let ℓ be a fixed integer and I a finite subset of $(\mathbb{N}_{>0})^\ell$. Then, for any list $(m_i)_{i \in I}$ of numbers, one can find some *complex-valued* random variables X_1, \dots, X_ℓ such that

$$\mathbb{E}(X_1^{i_1} \dots X_\ell^{i_\ell}) = m_{i_1, \dots, i_\ell}.$$

Indeed, one can look for a solution where X_1 is uniform on a finite set $\{z_1, \dots, z_T\}$ and $X_j = X_1^{dj-1}$, where d is an integer bigger than all coordinates of all vectors in I . Then the quantities

$$\{T \cdot \mathbb{E}(X_1^{i_1} \dots X_\ell^{i_\ell}), \mathbf{i} \in I\}$$

correspond to different power sums of z_1, \dots, z_T . Thus we have to find a family $\{z_1, \dots, z_T\}$ of complex number with specified power sums until degree d^j . This exists as soon as $T \geq d^j$, because \mathbb{C} is algebraically closed. In particular, the family B considered in the proof above exists.

However, we do not really need that this family exists. Indeed, during the whole proof, we are doing manipulations on the sequences of moments and cumulants using only the relations between them (equations (6) and (7)). We never consider the underlying random variables. Therefore, everything could be done even if the random variables did not exist, as it is often done in umbral calculus [24].

2.3. Case with distinct indices. We consider here the case where all entries in the sequences \mathbf{i} and \mathbf{s} are distinct. To be in the situation of Lemma 2.2, we set, for $h \in [\ell]$ and $N \geq 1$:

$$A_h^{(N)} = (N + \theta)_{a_j} \prod_{j \in \tau_h} B_{i_j, s_j}^{(N)},$$

where $a_j = |\tau_j|$. The normalization factor has been chosen such that $\mathbb{E}(A_h^{(N)}) = 1$. Hence, we will be able to apply Lemma 2.1.

Let us prove that $A_1^{(N)}, \dots, A_\ell^{(N)}$ fulfills property *I* of this lemma. Of course, the case $\Delta = [\ell]$ is generic. The joint moments of the family A have in this case an explicit expression: for $\delta \subseteq [\ell]$,

$$M_{A, \delta}^{(N)} = \frac{\prod_{j \in \delta} (N + \theta)_{a_j}}{(N + \theta)_{\sum_{j \in \delta} a_j}}.$$

Therefore, we have to prove that the quantity

$$Q_{a_1, \dots, a_\ell} := \prod_{\substack{\delta \subseteq [\ell] \\ |\delta| \geq 2}} (M_\delta^A)^{(-1)^{|\delta|}} = \prod_{\delta \subseteq [\ell]} \left((N + \theta)_{\sum_{j \in \delta} a_j} \right)^{(-1)^{|\delta|+1}}$$

write as $1 + O(N^{-\ell+1})$.

We proceed by induction over a_ℓ . If $a_\ell = 0$, for any $\delta \subseteq [\ell - 1]$, the factors corresponding to δ and $\delta \sqcup \{\ell\}$ cancel each other. Thus $Q_{a_1, \dots, a_{\ell-1}, 0} = 1$ and the statement holds.

If $a_\ell > 0$, the quantity Q_{a_1, \dots, a_ℓ} can be written as

$$Q_{a_1, \dots, a_\ell} = Q_{a_1, \dots, a_{\ell-1}} \cdot \prod_{\substack{\delta \subseteq [\ell] \\ \ell \in \delta}} \left(N + \theta - \sum_{j \in \delta} a_j \right)^{(-1)^{|\delta|+1}}.$$

Setting $X = N + \theta - a_\ell$, the second factor becomes

$$R_{a_1, \dots, a_{\ell-1}}(X) := \prod_{\delta \subseteq [\ell-1]} \left(X - \sum_{j \in \delta} a_j \right)^{(-1)^{|\delta|}}.$$

We will prove below (Lemma 2.4) that $R_{a_1, \dots, a_{\ell-1}}(X) = 1 + O(X^{-\ell+1})$. Besides, the induction hypothesis implies that $Q_{a_1, \dots, a_{\ell-1}} = 1 + O(N^{-\ell+1})$ and hence

$$Q_{a_1, \dots, a_\ell} = 1 + O(N^{-\ell+1})$$

Finally, the family $A_1^{(N)}, \dots, A_\ell^{(N)}$ of sequences of random variables has the quasi-factorisation property of Lemma 2.2. Thus it also has the small cumulant property and in particular

$$\kappa(A_1^{(N)}, \dots, A_\ell^{(N)}) = O(N^{-\ell+1}).$$

Using the definition of the $A_h^{(N)}$, this can be rewritten:

$$\kappa \left(\prod_{j \in \tau_1} B_{i_j, s_j}^{(N)}, \dots, \prod_{j \in \tau_\ell} B_{i_j, s_j}^{(N)} \right) = O(N^{-r-\ell+1}),$$

which is Theorem 1.4 in the case of distinct indices. \square

Here is the technical lemma that we left behind in the proof:

Lemma 2.4. *For any positive integers $a_1, \dots, a_{\ell-1}$,*

$$\prod_{\delta \subseteq [\ell-1]} \left(X - \sum_{j \in \delta} a_j \right)^{(-1)^{|\delta|}} = 1 + O(X^{-\ell+1}).$$

Proof. Define R_{ev} (resp. R_{odd}) as

$$\prod_{\delta} \left(X - \sum_{j \in \delta} a_j \right),$$

where the product runs over subsets of $[\ell-1]$ of even (resp. odd) size. Expanding the product, one gets

$$R_{\text{ev}} = \sum_{m \geq 0} \sum_{\delta_1, \dots, \delta_m} \sum_{j_1 \in \delta_1, \dots, j_m \in \delta_m} (-1)^m a_{j_1} \dots a_{j_m} X^{2^{\ell-2}-m}.$$

The index set of the second summation symbol is the set of lists of m distinct (but not necessarily disjoint) subsets of $[\ell-1]$ of even size. Of course, a similar formula with subsets of odd size holds for R_{odd} .

Let us fix an integer $m < \ell-1$ and a list j_1, \dots, j_m . Denote j_0 the smallest integer in $[\ell-1]$ different from j_1, \dots, j_m (as $m < \ell-1$, such an integer necessarily exists). Then one has a bijection:

$$\left\{ \begin{array}{l} \text{lists of subsets} \\ \delta_1, \dots, \delta_m \text{ of even size such} \\ \text{that, for all } h \leq m, j_h \in \delta_h \end{array} \right\} \rightarrow \left\{ \begin{array}{l} \text{lists of subsets} \\ \delta_1, \dots, \delta_m \text{ of odd size such} \\ \text{that, for all } h \leq m, j_h \in \delta_h \end{array} \right\}$$

$$(\delta_1, \dots, \delta_m) \mapsto (\delta_1 \nabla \{j_0\}, \dots, \delta_m \nabla \{j_0\}),$$

where ∇ is the symmetric difference operator. This bijection implies that the summand $(-1)^m a_{j_1} \dots a_{j_m} X^{2^{\ell-2}-m}$ appears as many times in R_{ev} than in R_{odd} . Finally, all terms corresponding to values of m smaller than $\ell-1$ cancel in the difference $R_{\text{ev}} - R_{\text{odd}}$ and one has

$$R_{\text{ev}} - R_{\text{odd}} = O(X^{2^{\ell-2}-\ell+1}). \quad \square$$

Remark 2.5. We would like to point out the fact that our result is closely related to a result of P. Śniady. Indeed, thanks to our multiplicative criterion to have small cumulants, the computation in this paragraph is equivalent to Lemma 4.8 of paper [25]. However, Śniady's proof relies on a non trivial theory of cumulants of observables of Young diagrams. Therefore, it seems to us that it is worth giving an alternative argument.

2.4. General case. Let $A_1^{(N)}, \dots, A_\ell^{(N)}$ be some sequences of random variables. We introduce some *truncated cumulants*: if π_0, π_1, π_2 and so on, are set partitions of $[\ell]$, we set

$$k_A^{(N)}(\pi_0) = \sum_{\substack{\pi \in \mathcal{P}([\ell]) \\ \pi \geq \pi_0}} \mu(\pi, \{[\ell]\}) \prod_{C \in \pi} M_{A,C}^{(N)}$$

$$k_A^{(N)}(\pi_0; \pi_1, \pi_2, \dots) = \sum_{\substack{\pi \in \mathcal{P}([\ell]) \\ \pi \geq \pi_0 \\ \pi \not\geq \pi_1, \pi_2, \dots}} \mu(\pi, \{[\ell]\}) \prod_{C \in \pi} M_{A,C}^{(N)}$$

In the context of Lemma 2.2, it is also possible to bound the truncated cumulants.

Lemma 2.6. *Let $A_1^{(N)}, \dots, A_\ell^{(N)}$ be some sequences of random variables as in Lemma 2.2, fulfilling property I (or equivalently property II).*

- If π_0 is a set partition of $[\ell]$,

$$k_A^{(N)}(\pi_0) = O(N^{-\#(\pi_0)+1}).$$

- More generally, if $\pi_0; \pi_1, \pi_2, \dots$ are set partitions of $[\ell]$,

$$k_A^{(N)}(\pi_0; \pi_1, \pi_2, \dots) = O(N^{-\#(\pi_0 \vee \pi_1 \vee \pi_2 \dots)+1}).$$

Proof. For the first statement, the proof is similar to the one of $I \Rightarrow II$ of Lemma 2.2. One can write an analogue of equation (11):

$$k_A^{(N)}(\pi_0) = \sum_{\substack{\Delta_1, \dots, \Delta_m \\ \text{pairwise distinct}}} T_{\Delta_1}^{(N)} \dots T_{\Delta_m}^{(N)} \left(\sum_{\substack{\pi \in \mathcal{P}([\ell]) \\ \pi \geq (\pi_0 \vee \pi(\Delta_1) \vee \dots)}} \mu(\pi, \{[\ell]\}) \right).$$

The same argument as above says that only terms corresponding to lists such that $\pi_0 \vee \pi(\Delta_1) \vee \dots = [\ell]$ survives. Such lists fulfills

$$\sum_{i=1}^m |\Delta_i| - 1 \geq \text{rk}([\ell]) - \text{rk}(\pi_0) = \#(\pi_0) - 1.$$

The first item of the Lemma follows because, by hypothesis,

$$T_{\Delta_1}^{(N)} \dots T_{\Delta_m}^{(N)} = O(N^{-\sum_i (|\Delta_i| - 1)}).$$

For the second statement, we use an inclusion/exclusion:

$$k_A^{(N)}(\pi_0; \pi_1, \dots, \pi_h) = \sum_{I \subseteq [h]} (-1)^{|I|} k_A^{(N)} \left(\pi_0 \vee \left(\bigvee_{i \in I} \pi_i \right) \right).$$

Then the second item follows from the first. \square

Let us come back to the proof of Theorem 1.4. We fix two lists \mathbf{i} and \mathbf{s} of length r , as well as a set partition τ of r . We want to find a bound for

$$\kappa \left(\prod_{j \in \tau_1} B_{i_j, s_j}^{(N)}, \dots, \prod_{j \in \tau_\ell} B_{i_j, s_j}^{(N)} \right) = \sum_{\substack{\pi \in \mathcal{P}([r]) \\ \pi \geq \tau}} \prod_{C \in \pi} \mathbb{E} \left(\prod_{i \in C} B_{i_j, s_j}^{(N)} \right).$$

We split the sum according to the values of the partitions $\pi_1 = \pi \wedge CC(G_1(\mathbf{i}, \mathbf{s}))$ and $\pi_2 = \pi \wedge CC(G_2(\mathbf{i}, \mathbf{s}))$. More precisely,

$$\kappa \left(\prod_{j \in \tau_1} B_{i_j, s_j}^{(N)}, \dots, \prod_{j \in \tau_\ell} B_{i_j, s_j}^{(N)} \right) = \sum_{\substack{\pi_1 \leq CC(G_1(\mathbf{i}, \mathbf{s})) \\ \pi_2 \leq CC(G_2(\mathbf{i}, \mathbf{s}))}} Y_{\pi_1, \pi_2}^{(N)},$$

where

$$Y_{\pi_1, \pi_2}^{(N)} = \sum_{\substack{\pi \geq \tau \\ \pi \wedge CC(G_1(\mathbf{i}, \mathbf{s})) = \pi_1 \\ \pi \wedge CC(G_2(\mathbf{i}, \mathbf{s})) = \pi_2}} \prod_{C \in \pi} \mathbb{E} \left(\prod_{i \in C} B_{i_j, s_j}^{(N)} \right).$$

We call the summation index the slice determined by π_1 and π_2 .

Let us fix some partitions π_1 and π_2 . For each block C of π_1 , we consider some sequence of random variables $(A_C^{(N)})_{N \geq 1}$ such that: for each list of distinct blocks C_1, \dots, C_h

$$\mathbb{E}(A_{C_1}^{(N)} \cdots A_{C_h}^{(N)}) = \frac{1}{(N + \theta)(N + \theta - 1) \cdots (N + \theta - h + 1)}.$$

For readers which wonder whether such variables exist, we refer to Remark 2.3, which remains valid here. Consider the family

$$(13) \quad \left((N + \theta) A_C^{(N)} \right)_{C \in \pi_1}.$$

Its joint moment are the same than the ones of the $B_{i, s}^{(N)}$ in the previous paragraph. It has been proven that such a family has the quasi-factorization property and, hence, its cumulants and truncated cumulants are small (Lemma 2.2).

But, if π is in the slice determined by π_1 and π_2 , one can check easily (see the description of joint moments in paragraph 2.1) that the corresponding product of moment is given by:

$$\prod_{C \in \pi} \mathbb{E} \left(\prod_{i \in C} B_{i_j, s_j}^{(N)} \right) = \alpha_{\pi_1, \pi_2} \prod_{C \in \pi} \mathbb{E} \left(\prod_{\substack{C' \in \pi_1 \\ C' \subseteq C}} A_{C'}^{(N)} \right),$$

where α_{π_1, π_2} depends only on π_1 and π_2 and is given by:

- 0 if π_2 contains in the same block two indices j and h such that $i_j = i_h$ but $s_j \neq s_h$ or $s_j = s_h$ but $i_j \neq i_h$;

- $(1 + \theta)^\gamma$ otherwise, where γ is the number of cycles of the partial permutation (\mathbf{i}, \mathbf{s}) , whose indices are all contained in the same block of π_2 .

As a consequence,

$$(14) \quad Y_{\pi_1, \pi_2}^{(N)} = \frac{\alpha_{\pi_1, \pi_2}}{(N + \theta)^{\#(\pi_1)}} \sum_{\substack{\pi \geq \tau \\ \pi \wedge CC(G_1(\mathbf{i}, \mathbf{s})) = \pi_1 \\ \pi \wedge CC(G_2(\mathbf{i}, \mathbf{s})) = \pi_2}} \prod_{C \in \pi} \mathbb{E} \left(\prod_{\substack{C' \in \pi_1 \\ C' \subseteq C}} (N + \theta) A_{C'} \right).$$

But the condition $\pi \wedge CC(G_1(\mathbf{i}, \mathbf{s})) = \pi_1$ can be rewritten as follows: $\pi \geq \pi_1$ and $\pi \not\geq \pi'$ for any $\pi_1 \leq \pi' \leq CC(G_1(\mathbf{i}, \mathbf{s}))$. A similar rewriting can be performed for the condition $\pi \wedge CC(G_2(\mathbf{i}, \mathbf{s})) = \pi_2$. Finally, the sum in equation (14) above is a truncated cumulant of the family (13) and is bounded from above by $O(N^{-|CC(G_2(\mathbf{i}, \mathbf{s})) \vee \tau| + 1})$. This implies

$$Y_{\pi_1, \pi_2}^{(N)} = O(N^{-\#(\pi_1) - |CC(G_2(\mathbf{i}, \mathbf{s})) \vee \tau| + 1}),$$

which ends the proof of Theorem 1.4 because π_1 has necessarily at least as many parts as $CC(G_1(\mathbf{i}, \mathbf{s}))$. \square

Remark 2.7. So far, we have considered the lists \mathbf{i} and \mathbf{s} as fixed. Therefore, the constant hidden in the Landau symbol O may depend of these lists. However, the quantity for which we establish an upper bound depends only on the partition τ and on which entries of the lists \mathbf{i} and \mathbf{s} coincide. For a fixed r , the number of partitions and of possible equalities is finite. Therefore, we can choose a constant depending only on r , as it is done in the statement of Theorem 1.4.

3. GRAPH-THEORETICAL LEMMAS

In this section, we present two quite easy lemmas on the number of connected components on contractions of graphs. These lemmas will be useful in the next sections for applications of Theorem 1.4.

3.1. Notations. Let us consider a graph G with vertex set V and edge set E . By definition, if V' is a subset of V , the graph $G[V']$ induced by G on V' has vertex set V' and edge set $E[V']$, where $E[V']$ is the subset of E consisting of edges having both their extremities in V' .

Let us consider a surjective map f from V to another set W . Then the *contraction* of G by f is the graph G/f with vertex set W and which has an edge between w and w' if, in G , there is at least one edge between a vertex of $f^{-1}(w)$ and a vertex of $f^{-1}(w')$.

Example. Consider the graph G of figure 1. Its vertex set is the 10-element set $V = \{1, 2, 3, 4, 5, \bar{1}, \bar{2}, \bar{3}, \bar{4}, \bar{5}\}$. Consider the application f from V to the set $W = \{1, 2, 3, 4, 5\}$, consisting in forgetting the bar (if any). The contracted graph G/f is drawn on the bottom left picture of Figure 1.

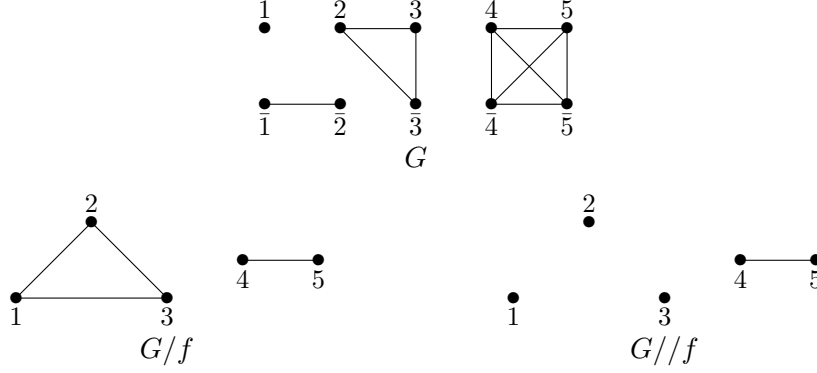


FIGURE 1. An example of a graph, its contraction and surcontraction.

3.2. Connected components of contractions.

Lemma 3.1. *Let G be a graph with vertex set V and f a surjective map from V to another set W . Then*

$$\#(\text{CC}(G)) \leq \#(\text{CC}(G/f)) + \sum_{w \in W} (\#(\text{CC}(G[f^{-1}(w)])) - 1).$$

Proof. For each edge (w, w') in G/f , we choose arbitrarily an edge (v, v') in G such that $f(v) = w$ and $f(v') = w'$ (by definition of G/f , such an edge exists but is not necessarily unique). Thereby, to each edge of G/f or of $G[f^{-1}(w)]$ (for any w in W) corresponds canonically an edge in G .

Take covering forests $F_{G/f}$ and $(F_w)_{w \in W}$ of graphs G/f and $G[f^{-1}(w)]$ for $w \in W$. With the remark above, to each covering forest corresponds a set of edges in G . Consider the union F of these sets. It is an acyclic set of edges of G . Indeed, if it contained a cycle, it must be contained in one of the fibers $f^{-1}(w)$, otherwise it would induce a cycle in $F_{G/f}$. But, in this case, all edges of the cycles belong to F_w , which is impossible, since F_w is a forest.

Finally, F is an acyclic set of edges in G and

$$\begin{aligned} \#(\text{CC}(G)) &\leq |V| - |F| = |W| - |F_{G/f}| + \sum_{w \in W} (|f^{-1}(W)| - 1 - |F_w|) \\ &\leq \#(\text{CC}(G/f)) + \sum_{w \in W} (\#(\text{CC}(G[f^{-1}(w)])) - 1). \quad \square \end{aligned}$$

Continuing the example. All fibers $f^{-1}(i)$ (for $i = 1, 2, 3, 4, 5$) are of size 2. Three of them contains one edge (for $i = 3, 4, 5$) and hence are connected, while the other two have two connected components. Finally, the sum in the lemma is equal to 2, which is equal to the difference

$$\#(\text{CC}(G)) - \#(\text{CC}(G/f)) = 4 - 2 = 2.$$

3.3. Fibers of size 2. In this paragraph, we consider the particular case where $V = W \sqcup W$ and f is the canonical application $W \sqcup W \rightarrow W$ consisting in forgetting to which copy of W the element belongs. Throughout the paper, for simplicity of notations, we will use overlined letters for elements of the second copy of W .

In this context, in addition to the contraction G/f , one can consider another graph with vertex set W . By definition, $G//f$ has an edge between w and w' if, in G , there is an edge between w and w' and an edge between \bar{w} and \bar{w}' . We call this graph the *surcontraction* of G .

Continuing the example. The graph G and the function f in the example above fit in the context described in this paragraph. The surcontraction $G//f$ is drawn on Figure 1 (bottom right picture).

Lemma 3.2. *Let G and f be as above. Then*

$$\#(\text{CC}(G)) \leq \#(\text{CC}(G/f)) + \#(\text{CC}(G//f)).$$

Proof. Set $G_1 = G/f$, $G_2 = G//f$ and $G_3 = G$.

By definition, an edge in G_1 between j and k corresponds to two edges in G_3 . In contrast, an edge (i, j) in G_2 corresponds to at least one edge in G_3 .

Consider a spanning forest F_1 in G_1 . As the set of edges of G_1 is smaller than the one of G_2 , F_1 can be completed into a spanning forest F_2 of G_2 . We consider the subset F_3 of edges of G_3 obtained as follows: for each edge of F_1 , we take the two corresponding edges in G_3 and for each edge of $F_2 \setminus F_1$, we take the corresponding edge in G_3 (if there is several corresponding edges, choose one arbitrarily).

We will prove by contradiction that F_3 is acyclic. Suppose that F_3 contains a cycle C_3 . Each edge of C_3 projects on an edge in F_2 and thus the projection of C_3 is a list $S = (e_1, \dots, e_h)$ of *consecutive* edges in F_2 (*consecutive* means that we can orient the edges such that, for each $\ell \in [h]$, the end point of e_ℓ is the starting point of $e_{\ell+1}$, with the convention $e_{h+1} = e_1$). This list is not necessarily a cycle because it can contain twice the same edges (either in the same direction or in different directions). Indeed, F_3 contains some pairs of edges of the form

$$(\{w, w'\}, \{\bar{w}, \bar{w}'\})$$

which project on the same edge in G_2 . But as edges from these pairs have no extremities in common, they can not appear consecutively in the cycle C_3 . Therefore, the same edge can not appear twice in a row in the list S . This implies that the list S contains a cycle C_2 as a factor. We have reached a contradiction as the edges in C_2 are edges of the forest F_2 . Thus F_3 is acyclic.

The number of edges in F_3 is clearly $2|F_1| + |F_2 \setminus F_1| = |F_1| + |F_2|$. Therefore

$$\begin{aligned} \#(\text{CC}(G_3)) &\leq 2|W| - |F_3| = (|W| - |F_1|) + (|W| - |F_2|) \\ &= \#(\text{CC}(G_1)) + \#(\text{CC}(G_2)). \quad \square \end{aligned}$$

4. TOY EXAMPLE: NUMBER OF CYCLES OF A GIVEN LENGTH p

In this section, we are interested in the number $\Gamma_p^{(N)}$ of cycles of length p in a random Ewens permutation of size N . The asymptotic behavior of $\Gamma_p^{(N)}$ is easy to determine (see Theorem 1.1), as its generating series is explicit and quite simple. We will give another proof which relies on Theorem 1.4 and does not use an explicit expression for the generating series of $\Gamma_p^{(N)}$.

The main steps of the proof are the same in the other examples, so let us emphasize them here.

Step 1: expand the cumulants of the considered statistic.

In this step, one has to express the statistic we are interested in in terms of the variables $B_{i,s}^{(N)}$: here,

$$\Gamma_p^{(N)} = \sum_{1 \leq i_1 < i_2, i_3, \dots, i_p \leq N} B_{(i_1, \dots, i_p)}^{c, N},$$

where $B_{(i_1, \dots, i_p)}^{c, N} = B_{i_1, i_2}^{(N)} \dots B_{i_{p-1}, i_p}^{(N)} B_{i_p, i_1}^{(N)}$ is defined by equation (2). Therefore, one has

$$(15) \quad \kappa_\ell(\Gamma_p^{(N)}) = \sum_{\substack{i_1^1 < i_2^1, i_3^1, \dots, i_p^1 \\ \vdots \\ i_1^\ell < i_2^\ell, i_3^\ell, \dots, i_p^\ell}} \kappa(B_{i_1^1, i_2^1}^{(N)} \dots B_{i_{p-1}^1, i_p^1}^{(N)}, \dots, B_{i_1^\ell, i_2^\ell}^{(N)} \dots B_{i_{p-1}^\ell, i_p^\ell}^{(N)}).$$

Step 2: Give an upper bound for the elementary cumulants.

Now, we would like to apply our main lemma to every summand of equation (15). For this, one has to understand what is the exponent of N in the upper bound given by Theorem 1.4.

For a matrix

$$(i_j^r)_{\substack{1 \leq j \leq p \\ 1 \leq r \leq \ell}},$$

we denote:

- $M(\mathbf{i}) = |\{(i_j^r, i_{j+1}^r); 1 \leq j \leq p, 1 \leq r \leq \ell\}|$ is the number of different entries in the matrix of couples (i_j^r, i_{j+1}^r) (by convention, $i_{p+1}^r = i_1^r$);
- $Q(\mathbf{i})$ the number of connected components of the graph $G(\mathbf{i})$ on $[\ell]$ where r_1 is linked with r_2 if

$$\{i_j^{r_1}; 1 \leq j \leq p\} \cap \{i_j^{r_2}; 1 \leq j \leq p\} \neq \emptyset.$$

In the case where τ has ℓ blocks of size p and where the list \mathbf{s} is obtained by a cyclic rotation of the list \mathbf{i} in each block, Theorem 1.4 writes as:

$$(16) \quad \left| \kappa(B_{i_1^1, i_2^1}^{(N)} \dots B_{i_{p-1}^1, i_p^1}^{(N)}, \dots, B_{i_1^\ell, i_2^\ell}^{(N)} \dots B_{i_{p-1}^\ell, i_p^\ell}^{(N)}) \right| \leq C_{p\ell} N^{-M(\mathbf{i}) - Q(\mathbf{i}) + 1} \leq C_{p\ell} N^{-M(\mathbf{i})}.$$

Step 3: give an upper bound for the number of lists.

As the number of summands in Equation (15) depends on N , we can not use directly inequality (16). We need a bound on the number of matrices \mathbf{i} with a given value of $M(\mathbf{i})$.

This bound comes from the following simple lemma:

Lemma 4.1. *For each $L \geq 1$, there exists a constant C'_L with the following property. For any $N \geq 1$ and $t \in [L]$, the number of lists \mathbf{i} of length L with entries in $[N]$ such that*

$$|\{i_1, \dots, i_L\}| = t$$

is bounded from above by $C'_L N^t$.

Proof. If we specify which indices correspond to entries with the same values (that is a set partition of the set of indices), the number of corresponding lists is $\binom{N}{t}$ and hence is bounded from above by N^t . This implies the lemma, with C'_L being equal to the number of set partitions of $[L]$. \square

Step 4: conclude.

For a matrix (i_j^r) , we denote $t(\mathbf{i})$ the number of distinct entries. Clearly, $M(\mathbf{i})$ is always at least equal to $t(\mathbf{i})$. Therefore, using inequality (16) and Lemma 4.1, for each $t \in [p \cdot \ell]$, the contribution of lists (i_j^r) taking exactly t different values is bounded from above by $C'_{p\ell} C_{p\ell}$ and hence

$$\text{for all } \ell \geq 1, \kappa_\ell(\Gamma_p^{(N)}) = O(1).$$

To compute the component of order 1, let us make the following remark: by the argument above, the total contribution of lists (i_j^r) with $M(\mathbf{i}) > t(\mathbf{i})$ or $Q(\mathbf{i}) > 1$ is $O(N^{-1})$.

But $M(\mathbf{i}) = t(\mathbf{i})$ implies that, as soon as

$$\{i_j^{r_1}; 1 \leq j \leq p\} \cap \{i_j^{r_2}; 1 \leq j \leq p\} \neq \emptyset,$$

the cyclic words $(i_1^{r_1}, \dots, i_p^{r_1})$ and $(i_1^{r_2}, \dots, i_p^{r_2})$ are equal. As i_1^r is always the minimum of the i_j^r , the two words are in fact always equal in this case. In particular $G(\mathbf{i})$ is a disjoint union of cliques. If we further assume $Q(\mathbf{i}) = 1$, i.e. $G(\mathbf{i})$ is connected, $G(\mathbf{i})$ is the complete graph and we get that i_j^r does not depend on r .

Finally

$$(17) \quad \kappa_\ell(\Gamma_p^{(N)}) = \sum_{i_1 < i_2, i_3, \dots, i_p} \kappa_\ell(B_{i_1, i_2}^{(N)} \dots B_{i_p, i_1}^{(N)}) + O(N^{-1}).$$

But each $B_{i_1, i_2}^{(N)} \dots B_{i_p, i_1}^{(N)}$ is a Bernoulli variable of parameter $(1 + \theta)/(N + \theta)_p$. Therefore their moments are all equal to $(1 + \theta)/(N + \theta)_p$ and by formula (7), their cumulants are $(1 + \theta)/(N + \theta)_p + O(N^{-2p})$. Finally, as there are $(N)_p/p$ terms in equation (17),

$$\kappa_\ell(\Gamma_p^{(N)}) = \frac{1 + \theta}{p} + O(N^{-1}),$$

which implies that $\Gamma_p^{(N)}$ converges in distribution towards a Poisson law of parameter $\frac{1 + \theta}{p}$.

Moreover, a simple adaptation of the proof of Equation (17) implies that

$$\kappa(\Gamma_{p_1}^{(N)}, \dots, \Gamma_{p_\ell}^{(N)}) = O(N^{-1})$$

as soon as two of the p_r 's are different. Indeed, no matrices $(i_j^r)_{\substack{1 \leq r \leq \ell \\ 1 \leq j \leq p_r}}$ with rows of different sizes fulfill simultaneously $M(\mathbf{i}) = t(\mathbf{i})$ and $Q(\mathbf{i}) = 1$. Finally, for any $p \geq 1$, the vector $(\Gamma_1^{(N)}, \dots, \Gamma_p^{(N)})$ tends in distribution towards a vector (P_1, \dots, P_p) where the P_i are independent Poisson-distributed random variables with respective parameters $(1 + \theta)/i$. \square

Remark. After equation (17), one could have finished the proof without computation by the following argument: $\Gamma_p^{(N)}$ has asymptotically the same cumulants as a virtual variable X_N , which writes as a sum of *independent* random variables with the same distribution as the $B_{(i_1, \dots, i_p)}^{c, N}$. As each $B_{(i_1, \dots, i_p)}^{c, N}$ is a Bernoulli variable of expectation $(1 + \theta)/(N + \theta)_p$ and as there are $(N)_p/p$ such variables, X_N converges in distribution towards a Poisson law of parameter $(1 + \theta)/p$. And so does $\Gamma_p^{(N)}$.

As promised in the introduction, this argument follows the idea that *everything happens as if the variables $B_{(i_1, \dots, i_p)}^{c, N}$ were independent*.

5. NUMBER OF EXCEDANCES

In this section, we look at our second motivating problem, the number of excedances in random permutations. The first two paragraphs make a link between a physical statistics model and this problem, justifying our work. The last two paragraphs are devoted to the proof of Theorem 1.2 and related results.

5.1. Symmetric simple exclusion process. The symmetric simple exclusion process (*SSEP* for short) is a model of statistical physics: we consider particles on a discrete line with N sites. No two particles can be in the same site at the same moment. The system evolves as follows:

- if its neighboring site is empty, a particle can jump to its left or its right with probability $\frac{1}{N+1}$;
- if the left-most site is empty (resp. occupied), a particle can enter (resp. leave) by the left with probability $\frac{\alpha}{N+1}$ (resp. $\frac{\gamma}{N+1}$);
- if the right-most site is empty (resp. occupied), a particle can enter (resp. leave) by the right with probability $\frac{\delta}{N+1}$ (resp. $\frac{\beta}{N+1}$);
- otherwise (we suppose $\alpha, \beta, \gamma, \delta < 1$ such that, in a given state, the sum of the probabilities of the events which may occur is smaller than 1), nothing happens.

Mathematically, this defines an irreducible aperiodic Markov chain on the finite set $\{0; 1\}^N$ (a state of the *SSEP* can be encoded as a word in 0 and 1 of length N , where the entries with value 1 correspond to the positions of the occupied sites).

This model is quite popular among physicists because, despite its simplicity, it exhibits interesting phenomenons like the existence of different phases. For a comprehensive introduction on the subject and a survey of results, see [11].

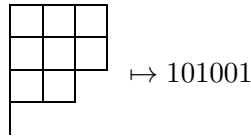


FIGURE 2. From permutation tableaux to words in $\{0; 1\}^{N-1}$.

A good way to describe a state τ of the *SSEP* is the function $F_\tau^{(N)}$ defined as follows: when Nx is an integer,

$$F_\tau^{(N)}(x) = \frac{1}{N} \cdot \sum_{i=1}^{Nx} \tau_i$$

and, for each $i \in [N]$, the function $F_\tau^{(N)}$ is affine between $(i-1)/N$ and i/N . One should see $F_\tau^{(N)}$ as the integral of the density of particles in the system.

We are interested in the steady state of the *SSEP*, that is the unique probability measure μ_N on $\{0; 1\}^N$, which is invariant by the dynamics. More precisely, we want to study asymptotically the properties of the random function $F_\tau^{(N)}$, where τ is distributed with μ_N and N tends to infinity.

5.2. Link with permutation tableaux and Ewens measure. From now on, we restrict to the case $\alpha, \gamma, \delta = 1$. In this case, thanks to a result of S. Corteel and L. Williams [10], the measure μ_N is related to some combinatorial objects, called permutation tableaux.

The latter are fillings of Young diagrams (which can have empty rows, but no empty columns) with 0 and 1 respecting some rules, the details of which will not be important here. The Young diagram is called the shape of the permutation tableau. The size of a permutation tableau is its number of rows, plus its number of columns (and not the number of boxes!).

In addition with their link with statistical physics, permutation tableaux also appear in algebraic geometry: they index the cells of some canonical decomposition of the totally positive part of the Grassmannian [23, 29]. They have also been widely studied from a purely combinatorial point of view [27, 9, 2].

To a permutation tableau T of size $N + 1$, one can associate a word w^T in $\{0; 1\}^N$ as follows: we label the steps of the border of the tableau starting from the North-East corner to the South-West corner. The first step is always a South step. For the other steps, we set $w_i^T = 1$ if and only if the $i + 1$ -th step is a south step. Clearly, the word w^T depends only on the shape of the tableau T . This procedure is illustrated on figure 2.

With this definition, the border of a tableau T of size $N + 1$ is the parametric broken line

$$\{(n_1(w^T) - NF_{w^T}^{(N)}(x), -N(x - F_{w^T}^{(N)}(x)) - 1) : x \in [0; 1]\},$$

where $n_1(w^T)$ is the number of 1 in w^T and $F_{w^T}^{(N)}$ the function associated to the word w^T as defined in the previous paragraph. Hence, $F_{w^T}^{(N)}$ is a good way to encode the shape of the permutation tableau T .

S. Corteel and L. Williams also introduced a statistics on permutation tableaux called *number of unrestricted rows* and denoted $u(T)$. If β is a positive real parameter, this statistics induces a measure $\mu_N^T(\beta)$ on permutation tableaux of size N , for which the probability to pick a tableau T is proportional to $\beta^{-u(T)}$. This measure is related to the *SSEP* by the following result (which is in fact a particular case of [10, Theorem 3.1] but we do not know how to deal with the extra parameters there).

Theorem 5.1. [10] *The steady state of the SSEP μ_N is the push-forward by the application $T \mapsto w_T$ of the probability measure $\mu_{N+1}^T(\beta)$.*

It turns out that this measure can also be described using random permutations. Indeed, S. Corteel and P. Nadeau [9, Theorem 1 and Section 3] have exhibited a simple bijection Φ between permutations of $N + 1$ and permutation tableaux of size $N + 1$, which satisfies:

- If a permutation σ is mapped to a tableau $T = \Phi(\sigma)$, then:

$$w^T = (\delta_2(\sigma), \delta_3(\sigma), \dots, \delta_{N+1}(\sigma)),$$

where $\delta_i = 1$ if i is an ascent, that is if $\sigma(i) < \sigma(i + 1)$ (by convention $\delta_{\sigma(N+1)}(\sigma) = 1$).

- The number of unrestricted rows of a tableau $T = \Phi(\sigma)$ is the number of right-to-left minima of σ : recall that i is a right-to-left minimum of σ if $\sigma_\ell > i$ for any $\ell > \sigma^{-1}(i)$.

We are rather interested in the number of cycles of permutations rather than their number of right-to-left minima. The following bijection, which is a variant of the first fundamental transformations on permutation [20, § 10.2], sends one of this statistics to the other. Take a permutation σ , written in its cycle notation such that:

- its cycles ends with their minimum;
- the minima of the cycles are in increasing order.

For example, $\sigma = (3\ 5\ 1)(7\ 4\ 2)(6)$. Now, erase the parenthesis: we obtain the word notation of a permutation $\Psi(\sigma)$.

The application Ψ is a bijection from S_N to S_N . Besides, the minima of the cycles of σ are the right-to-left minima of $\Psi(\sigma)$, while the ascents in $\Psi(\sigma)$ are the weak excedances in σ , that is the integers i such that $\sigma(i) \geq i$ (a similar statement is given in [20, Theorem 10.2.3]).

From now on, we assume $\beta \cdot (1 + \theta) = 1$. The properties above imply that $\mu_N^T(\beta)$ is the push-forward of the Ewens measure of parameter θ by the application $\Phi \circ \Psi$. Combining this with Theorem 5.1, the steady state of the *SSEP* μ_N is the push-forward of Ewens measure by the application $\sigma \mapsto w^{\Phi(\Psi(\sigma))}$. But this application admits an easy direct description

$$\begin{aligned} S_{N+1} &\rightarrow \{0; 1\}^N \\ \sigma &\mapsto (\delta_{\sigma(2) \geq 2}, \delta_{\sigma(3) \geq 3}, \dots, \delta_{\sigma(N+1) \geq N+1}). \end{aligned}$$

Recall that, as explained above, we are interested in the random function $F_\tau^{(N)}$, where τ is distributed according to the measure μ_{N-1} . The results above imply that this random function has the same distribution than $F_\sigma^{(N+1)}$, where σ is a random permutation of size N distributed with Ewens measure of parameter θ and $F_\sigma^{(N+1)}$ is the function defined in paragraph 1.2.

5.3. Bounds for cumulants. Let us fix some real numbers x_1, \dots, x_ℓ in $[0; 1]$. In this paragraph, we will give some bounds on the joint cumulants of the random variables $(F_\sigma^{(N)}(x_1), \dots, F_\sigma^{(N)}(x_\ell))$.

Let us begin by the following bound (step 2 of the proof, according to the division done in section 4).

Proposition 5.2. *For any $\ell \geq 1$, any $N \geq 1$ and any lists i_1, \dots, i_ℓ and s_1, \dots, s_ℓ of integers in $[N]$,*

$$\kappa(B_{i_1, s_1}^{(N)}, \dots, B_{i_\ell, s_\ell}^{(N)}) \leq C_\ell N^{-|\{i_1, \dots, i_\ell, s_1, \dots, s_\ell\}|+1},$$

where C_ℓ is the constant defined by Theorem 1.4.

Proof. Using Theorem 1.4 for $\tau = \{\{1\}, \dots, \{\ell\}\}$, we only have to prove that

$$-\#(\text{CC}(G_1(\mathbf{i}, \mathbf{s}))) - \#(\text{CC}(G_2(\mathbf{i}, \mathbf{s}))) \geq -|\{i_1, \dots, i_\ell, s_1, \dots, s_\ell\}|.$$

The last quantity $|\{i_1, \dots, i_\ell, s_1, \dots, s_\ell\}|$ can be seen as the number of connected component of the graphs $G_3(\mathbf{i}, \mathbf{s})$ defined as follows:

- its vertex set is $[\ell] \sqcup [\ell] = \{1, \bar{1}, \dots, \ell, \bar{\ell}\}$;
- there is an edge between j and k (resp. j and \bar{k} , \bar{j} and k) if and only if $i_j = i_k$ (resp. $i_j = s_k, s_j = s_k$).

The inequality above is simply Lemma 3.2 applied to the graph $G_3(\mathbf{i}, \mathbf{s})$ ($G_1(\mathbf{i}, \mathbf{s})$ and $G_2(\mathbf{i}, \mathbf{s})$ are respectively its surcontraction and contraction). \square

We can now prove the following bound:

Proposition 5.3. *There exists a constant C_ℓ'' such that, for any integer $N \geq 1$ and real numbers x_1, \dots, x_ℓ , one has*

$$|\kappa(F_\sigma^{(N)}(x_1), \dots, F_\sigma^{(N)}(x_\ell))| \leq C_\ell'' N^{-\ell+1}.$$

Proof. To simplify the notations, we suppose that Nx_1, \dots, Nx_ℓ are integers, so that

$$(N-1) \cdot F_\sigma^{(N)}(x_i) = \sum_{i=2}^{Nx_i} B_i^{\text{ex}, N}(\sigma).$$

But the Bernoulli variable $B_i^{\text{ex}, N}$ can be written as $B_i^{\text{ex}, N} = \sum_{s \geq i} B_{i, s}^{(N)}$. Finally, by multilinearity, one has (step 1):

$$(18) \quad (N-1)^\ell \kappa(F_\sigma^{(N)}(x_1), \dots, F_\sigma^{(N)}(x_\ell)) = \sum_{\substack{2 \leq i_1 \leq Nx_1 \\ \vdots \\ 2 \leq i_\ell \leq Nx_\ell}} \sum_{\substack{s_1 \geq i_1 \\ \vdots \\ s_\ell \geq i_\ell}} \kappa(B_{i_1, s_1}^{(N)}, \dots, B_{i_\ell, s_\ell}^{(N)}).$$

We apply Lemma 4.1 to the list $i_1, \dots, i_\ell, s_1, \dots, s_\ell$ and get that the number of couples of lists (\mathbf{i}, \mathbf{s}) such that $|\{i_1, \dots, i_\ell, s_1, \dots, s_\ell\}|$ is equal to a given number t is bounded from above by $C'_{2\ell} N^t$ (step 3).

Combining this with Proposition 5.2, we get that the total contribution of couples of lists (\mathbf{i}, \mathbf{s}) with $|\{i_1, \dots, i_\ell, s_1, \dots, s_\ell\}| = t$ to the right-hand side of (18) is smaller than $C'_{2\ell} C_\ell N$, which ends the proof of Proposition 5.3 (step 4). \square

Illustration of the proof. Set $\ell = 5$ and consider the lists $\mathbf{i} = (5, 2, 2, 7, 7)$ and $\mathbf{s} = (8, 8, 2, 7, 7)$. The graph $G_3(\mathbf{i}, \mathbf{s})$ associated to this couple of sequences is the graph G drawn of Figure 1. It follows immediately that $G_1(\mathbf{i}, \mathbf{s}) = G/f$ has 4 connected components while $G_2(\mathbf{i}, \mathbf{s}) = G/f$ has 2. Therefore, by Theorem 1.4,

$$\kappa(B_{5,8}^{(N)}, B_{2,8}^{(N)}, B_{2,2}^{(N)}, B_{7,7}^{(N)}, B_{7,7}^{(N)}) \leq C_5 N^{-5}.$$

The same bound is valid for all sequences \mathbf{i} and \mathbf{s} such that $G_3(\mathbf{i}, \mathbf{s}) = G$. There are fewer than N^4 such sequences: to construct such a sequence, one has to choose distinct values for the four connected components of G , such that they fulfill some inequalities. Finally, their total contribution to (18) is smaller than $C_5 N^{-1}$.

Comparison with a result of B. Derrida, J.L. Lebowitz and E.R. Speer. In [12, Appendix A], it is proven a *long range correlation phenomenon* for the *SSEP*. Rewritten in terms of Ewens random permutations *via* the material of the previous paragraph, it asserts that, for $i_1 < \dots < i_\ell$,

$$\kappa(B_{i_1}^{\text{ex},N}, \dots, B_{i_\ell}^{\text{ex},N}) = O(N^{-\ell+1}).$$

In fact, their result is more general because it corresponds to the *SSEP* with all parameters. This bound on cumulants can be obtained easily using our Proposition 5.2 and Lemma 4.1. A slight generalization of it (taking into account the case where some i 's can be equal) implies directly Proposition 5.3. Therefore, our method does not give some new results on the *SSEP*. Nevertheless, it was natural to try to understand the long range correlation phenomenon directly in terms of random permutations and it is what our approach does.

5.4. Convergence results. In this paragraph, we explain how one can deduce from the bound on cumulants, some results on the convergence of the random function $F_\sigma^{(N)}$, in particular Theorem 1.2.

In addition to the bounds above, we need equivalents for the first and second joint cumulants of the $F_\sigma^{(N)}(x)$. An easy computation gives:

$$\begin{aligned} \mathbb{E}(B_i^{\text{ex},N}) &= \frac{N - i + 1 + \theta}{N + \theta}; \\ \text{Var}(B_i^{\text{ex},N}) &= \frac{(i - 1)(N - i + 1 + \theta)}{(N + \theta)^2}; \\ \text{Cov}(B_i^{\text{ex},N}, B_j^{\text{ex},N}) &= -\frac{(n - j + 1 + \theta)(i - 1)}{(N + \theta^2)(N + \theta - 1)} \text{ for } i < j, \end{aligned}$$

from which we get the limits:

$$(19) \quad \lim_{N \rightarrow \infty} \mathbb{E}(F_\sigma^{(N)}(x)) = \int_0^x (1-t)dt + o(1) = \frac{1 - (1-x)^2}{2};$$

$$(20) \quad \lim_{N \rightarrow \infty} N \operatorname{Cov}(F_\sigma^{(N)}(x), F_\sigma^{(N)}(y)) = \int_0^{\min(x,y)} t(1-t)dt \\ - \int_{\substack{0 \leq t \leq x \\ 0 \leq u \leq y}} \min(t, u)(1 - \max(t, u))dtdu.$$

We call $K(x, y)$ the right-hand side of the second equation. We begin by a proof of Theorem 1.2, which describes the asymptotic behavior of $F_\sigma^{(N)}(x)$, for fixed value(s) of x .

Proof. Consider the first statement. The convergence in probability of $F_\sigma^{(N)}(x)$ towards $1/2 \cdot (1 - (1-x)^2)$ follows immediately from equations (19) and (20). For the almost-sure convergence, we have to study the fourth centered moment.

From moment-cumulant formula (7) and using the fact that all cumulant but the first are invariant by a shift of the variable,

$$\mathbb{E} \left((F_\sigma^{(N)}(x) - \mathbb{E}(F_\sigma^{(N)}(x)))^4 \right) = \kappa_4(F_\sigma^{(N)}(x)) + 3(\kappa_2(F_\sigma^{(N)}(x)))^2.$$

By proposition 5.3, this quantity is bounded from above by $O(N^{-2})$ and, in particular,

$$\sum_{N \geq 1} \mathbb{E} \left((F_\sigma^{(N)}(x) - \mathbb{E}(F_\sigma^{(N)}(x)))^4 \right) < \infty.$$

The end of the proof is classical. First, we inverse the summation and expectation symbols (all quantities are nonnegative). As its expectation is finite, the random variable

$$\sum_{N \geq 1} (F_\sigma^{(N)}(x) - \mathbb{E}(F_\sigma^{(N)}(x)))^4$$

is almost surely finite and hence its general term $((F_\sigma^{(N)}(x) - \mathbb{E}(F_\sigma^{(N)}(x)))^4)_{N \geq 1}$ tends almost surely to 0.

Let us consider the second statement. Proposition 5.3 implies that, for any list j_1, \dots, j_ℓ of integers in $[r]$, one has

$$\kappa(Z_\sigma^{(N)}(x_{j_1}), \dots, Z_\sigma^{(N)}(x_{j_\ell})) = O(N^{-r/2+1}).$$

In particular, for $r > 2$ the left-hand side tends to 0. As the variables $Z_\sigma^{(N)}(x_i)$ are centered, this implies that $(Z_\sigma^{(N)}(x_1), \dots, Z_\sigma^{(N)}(x_r))$ tends towards a centered Gaussian vector. The covariance matrix is the limit of the covariance of the $Z_\sigma^{(N)}(x_i)$, that is $(K(x_i, x_j))$. \square

It is also possible to obtain some results concerning the sequence of random function $(F_\sigma^{(N)})_{N \geq 1}$. In the following statement, we consider convergence in the functional space $(C([0; 1]), \|\cdot\|_\infty)$, that is uniform convergence of continuous functions.

Theorem 5.4. *Almost surely, the function $F_\sigma^{(N)}$ converges towards the function $x \mapsto 1/2 \cdot (1 - (1 - x)^2)$.*

Moreover, the sequence of random functions $(x \mapsto Z_\sigma^{(N)}(x))_{N \geq 1}$ converges in distribution towards the Gaussian process $x \mapsto G(x)$, whose finite dimension laws are Gaussian vectors with covariance matrices given by $(K(x_i, x_j))_{1 \leq i, j \leq r}$.

Proof. As, for any $N \geq 1$ and any $\sigma \in S_N$, the function $x \mapsto F_\sigma^{(N)}(x)$ is non-decreasing, the first statement follows easily from the convergence at any fixed x . The argument can be found for example in a paper of J.F. Marckert [21, first page], but it is so short and simple that we copy it here. By monotonicity of $F_\sigma^{(N)}$ and F , for any list $0 = x_0 < x_1 < \dots < x_k = 1$, one has

$$\begin{aligned} & \sup_{x \in [0;1]} |F_\sigma^{(N)}(x) - F(x)| \\ & \leq \max_{0 \leq j < k} \max (|F_\sigma^{(N)}(x_{j+1}) - F(x_j)|, |F_\sigma^{(N)}(x_j) - F(x_{j+1})|) \\ & \xrightarrow{\text{a.s.}} \max_{0 \leq j < k} |F(x_j) - F(x_{j+1})|, \end{aligned}$$

which may be chosen as small as wanted.

Consider the second statement. If the sequence of random function $x \mapsto Z_\sigma^{(N)}(x)$ has a limit, its finite-dimensional laws are necessarily the limits of the ones of $Z_\sigma^{(N)}$, that is, by Theorem 1.2, Gaussian vectors with covariance matrices given by $(K(x_i, x_j))_{1 \leq i, j \leq r}$. As a probability measure on $\mathcal{C}([0; 1])$ is entirely determined by its finite dimensional laws [5, Example 1.2], one just has to prove that the sequence $x \mapsto Z_\sigma^{(N)}(x)$ has indeed a limit. To do this, it is enough to prove that it is *tight* [5, Section 5, Theorems 5.1 and 7.1], that is, for each $\epsilon > 0$ there exists some constant M such that:

$$\text{for all } N > 0, \text{ one has } \text{Prob}(\|Z_\sigma^{(N)}\|_\infty > M) \leq \epsilon.$$

Once again, this follows from a careful analysis of the fourth moment.

Let $N \geq 1$ and $s \neq s'$ in $[0; 1]$ such that Ns and Ns' are integers. Using equation (7) and the fact that $Z_\sigma^{(N)}(s)$ and $Z_\sigma^{(N)}(s')$ are centered, one has:

$$\begin{aligned} & \mathbb{E} \left((Z_\sigma^{(N)}(s) - Z_\sigma^{(N)}(s'))^4 \right) \\ & = \kappa_4(Z_\sigma^{(N)}(s) - Z_\sigma^{(N)}(s')) + 3\kappa_2(Z_\sigma^{(N)}(s) - Z_\sigma^{(N)}(s'))^2 \\ & = N^2(\kappa_4(F_\sigma^{(N)}(s) - F_\sigma^{(N)}(s')) + 3\kappa_2(F_\sigma^{(N)}(s) - F_\sigma^{(N)}(s'))^2). \end{aligned}$$

A simple adaptation of the proof of Proposition 5.3 shows that

$$\kappa_\ell(F_\sigma^{(N)}(s) - F_\sigma^{(N)}(s')) \leq C_\ell N^{-\ell+1} |s - s'|.$$

Indeed, in Lemma 4.1, if we ask that at least one entry of the list \mathbf{i} is between Ns and Ns' then the number of lists is bounded from above by $C'_L N^t |s - s'|$. Finally,

$$\begin{aligned} \mathbb{E} \left((Z_\sigma^{(N)}(s) - Z_\sigma^{(N)}(s'))^4 \right) &\leq (N^2(C_4 N^{-3} |s - s'| + 3C_2^2 N^{-2} |s - s'|^2)) \\ &\leq (C_4 + 3C_2^2) |s - s'|^2. \end{aligned}$$

The last inequality has been deduced from $|s - s'| \geq N^{-1}$.

We can now apply Th. 10.2 of Billingsley's book [5] with $S_i = Z_\sigma^{(N)}(i/N)$ (for $0 \leq i \leq N$), $\alpha = \beta = 1$ and $u_\ell = (C_4 + 3C_2^2)^{1/2}/N$ (see equation (10.11) of the same book). We get that there exists some constant K such that

$$\text{Prob} \left(\max_{0 \leq i \leq N} |S_i| \geq M \right) \leq KM^{-4},$$

which proves that the sequence $Z_\sigma^{(N)}$ is tight. \square

6. GENERALIZED PATTERNS

This section is devoted to the applications of our method to adjacencies (paragraph 6.2) and dashed patterns (paragraph 6.3). These two statistics belong in fact to the same general framework and we discuss in paragraph 6.4 the possibility of unifying our results.

The proofs in this paragraph are a little bit more technical than the ones before and in particular we need a new lemma for step 3, given in paragraph 6.1.

6.1. Preliminaries. Let $L \geq 1$ be an integer. For each pair $\{j, k\} \subset [L]$, we choose a *finite* set of integers $D_{\{j, k\}}$.

Consider a list i_1, \dots, i_L of integers. For each pair $e = \{j, k\} \subset [L]$ (with $j < k$), we denote $\delta_e(\mathbf{i})$ the difference $i_k - i_j$. Then we associate to \mathbf{i} a graph of vertex set $[L]$ and edge set $\{e : \delta_e(\mathbf{i}) \in D_e\}$.

The following lemma is a slight generalization of Lemma 4.1

Lemma 6.1. *For each L and families of sets $(D_{\{j, k\}})_{1 \leq j < k \leq L}$, there exists a constant $C''_{L, \mathbf{D}}$ with the following property. For any $N \geq 1$ and $t \leq L$, the number of sequences i_1, \dots, i_L with entries in $[N]$, whose corresponding graph has exactly t connected components is bounded from above by $C''_{L, \mathbf{D}} N^t$.*

Proof. If we fix a graph G with vertex set L and t connected components and if we fix also, for each edge e of the graph, the actual value of $\delta_e(\mathbf{i})$, then the corresponding number of lists \mathbf{i} is smaller than N^t . Indeed, the sequence will be determined by the choice of one value per connected component of G (with some constraints, such that no extra edges appear). But the number of graphs and of values on edges are finite (the sets $D_{j, k}$ are finite) and depend only on L and on the family \mathbf{D} . \square

6.2. Adjacencies. In this paragraph, we prove the following extension of Theorem 1.3.

Theorem 6.2. *Let σ_N be a sequence of random permutations, such that σ_N has size N and is distributed with respect to Ewens measure of parameter θ . Then the number $A^{(N)}$ of adjacencies in σ_N converges in distribution towards a Poisson variable of parameter 2.*

Proof. As before, we write $A^{(N)}$ in terms of the $B_{i,s}^{(N)}$ (we use the convention $B_{i,s}^{(N)} = 0$ if i or s is not in $[N]$):

$$A^{(N)} = \sum_{\substack{1 \leq i, s \leq N \\ \epsilon = \pm 1}} B_{i,s}^{(N)} B_{i+1, s+\epsilon}^{(N)}.$$

Hence, for $\ell \geq 1$, its ℓ -th cumulant writes as (step 1):

$$(21) \quad \kappa_\ell(A^{(N)}) = \sum_{\substack{1 \leq i_1, s_1, \dots, i_\ell, s_\ell \leq N \\ \epsilon_1, \dots, \epsilon_\ell = \pm 1}} \kappa \left(B_{i_1, s_1}^{(N)} B_{i_1+1, s_1+\epsilon_1}^{(N)}, \dots, B_{i_\ell, s_\ell}^{(N)} B_{i_\ell+1, s_\ell+\epsilon_\ell}^{(N)} \right).$$

Given two lists \mathbf{i} and \mathbf{s} of positive integers, we consider the three following graph:

- H_1 has vertex set $[\ell]$ and has an edge between j and k if $|i_j - i_k| \leq 2$ and $|s_j - s_k| \leq 2$;
- H_2 has vertex set $[\ell]$ and has an edge between j and k if

$$\{i_j, i_j \pm 1, s_j, s_j \pm 1\} \cap \{i_k, i_k \pm 1, s_k, s_k \pm 1\} \neq \emptyset.$$
- H_3 has vertex set $[\ell] \sqcup [\ell]$ and has an edge between j and k (resp. j and \bar{k} , \bar{j} and k) if $|i_j - i_k| \leq 2$ (resp. $|i_j - s_k| \leq 2$, $|s_j - s_k| \leq 2$)

We will use Theorem 1.4 to give a bound for

$$\left| \kappa(B_{i_1, s_1}^{(N)} B_{i_1+1, s_1+\epsilon_1}^{(N)}, \dots, B_{i_\ell, s_\ell}^{(N)} B_{i_\ell+1, s_\ell+\epsilon_\ell}^{(N)}) \right|$$

Clearly, the number $M(\mathbf{i}, \mathbf{s})$ of different couples in the set

$$\{(i_j, s_j); 1 \leq j \leq \ell\} \cup \{(i_j + 1, s_j + \epsilon_j); 1 \leq j \leq \ell\}$$

is at least equal to $2\#(\text{CC}(H_1)) \geq \#(\text{CC}(H_1)) + 1$. Besides, in this case, the graph G'_2 introduced in paragraph 1.5 has the same vertex set as H_2 and fewer edges. Hence it has more connected components. Therefore, Theorem 1.4 implies (step 2):

$$\left| \kappa(B_{i_1, s_1}^{(N)} B_{i_1+1, s_1+\epsilon_1}^{(N)}, \dots, B_{i_\ell, s_\ell}^{(N)} B_{i_\ell+1, s_\ell+\epsilon_\ell}^{(N)}) \right| \leq C_{2\ell} N^{-\#(\text{CC}(H_1)) - \#(\text{CC}(H_2))}.$$

But, using the terminology of paragraph 3.3, the graphs H_1 and H_2 are the surcontraction and contraction of H_3 . Therefore, by Lemma 3.2, one has:

$$(22) \quad \#(\text{CC}(H_3)) \leq \#(\text{CC}(H_1)) + \#(\text{CC}(H_2)).$$

Besides, Lemma 6.1 implies the number of lists \mathbf{i} and \mathbf{s} with entries in $[N]$ such that H_3 has exactly t connected components is bounded from above by $C_{2\ell, \mathcal{D}}'' N^t$

for D well-chosen (step 3). In particular the constant $C''_{2\ell, D}$ does not depend on N . Therefore, the total contribution of these lists to equation (21) is bounded from above by $C_{2\ell} N^{-t} \cdot C''_{2\ell, D} N^t = C_{2\ell} \cdot C''_{2\ell, D}$.

Finally,

$$\kappa_\ell(A^{(N)}) = O(1).$$

Moreover, only lists such that $M(\mathbf{i}, \mathbf{s}) = 2$ and $\#(CC(H_1)) = 1$ contribute to the term of order 1. But this implies that the lists \mathbf{i} , \mathbf{s} and ϵ are constant. In other words,

$$\kappa_\ell(A^{(N)}) = \sum_{\substack{i, s \geq 1 \\ \epsilon = \pm 1}} \kappa_\ell(B_{i, s}^{(N)} B_{i+1, s+\epsilon}^{(N)}) + O(N^{-1}).$$

The $2(N-1)^2$ variables $B_{i, s}^{(N)} B_{i+1, s+\epsilon}^{(N)}$ are Bernoulli variables, whose parameters are given by:

- if $s = i \in [N-1]$ and $\epsilon = 1$, then the parameter is $\frac{\theta^2}{(N+\theta)(N+\theta-1)}$ ($N-1$ cases);
- if $s = i; \epsilon = -1$ (here $2 \leq i \leq N-1$) or $s = i+1; \epsilon = -1$ (here $1 \leq i \leq N-1$) or $s = i+2; \epsilon = -1$ (here $1 \leq i \leq N-2$), then the parameter is $\frac{\theta}{(N+\theta)(N+\theta-1)}$ ($3N-5$ cases);
- otherwise, the parameter is $\frac{1}{(N+\theta)(N+\theta-1)}$.

Recall that the cumulants of a sequence of Bernoulli variables $X^{(N)}$ of parameters $(p_N)_{N \geq 1}$ with $p_N \rightarrow 0$ are asymptotically given by $k_\ell(X^{(N)}) = p_N + O(p_N^2)$. Hence,

$$k_\ell(A^N) = 2(N-1)^2 \frac{1}{(N+\theta)(N+\theta-1)} + O(N^{-1}) = 2 + O(N^{-1}).$$

Finally, the cumulants of A^N converges towards those of a Poisson variable of parameter 2, which implies the convergence of A^N in distribution. \square

6.3. Dashed patterns. Let us recall the definition of dashed patterns in a permutation, as introduced by E. Babson and E. Steingrimsson [3].

Definition 6.3. A dashed pattern of size p is the data of a permutation $\tau \in S_p$ and a subset X of $[p-1]$. An occurrence of the dashed pattern (τ, X) in a permutation $\sigma \in S_N$ is a list $i_1 < \dots < i_p$ such that:

- for any $x \in X$, one has $i_{x+1} = i_x$.
- $\sigma(i_1), \dots, \sigma(i_p)$ is in the same relative order than $\tau(1), \dots, \tau(p)$.

The number of occurrences of the pattern (τ, X) will be denoted $O_{\tau, X}^{(N)}(\sigma)$.

Example 6.4. $O_{21, \emptyset}^{(N)}$ is the number of inversion, while $O_{21, \{1\}}^{(N)}$ is the number of descents. Many classical statistics on permutations can be written as the number of occurrences of a given dashed pattern or as a linear combination of such statistics, see [3, section 2].

In this paragraph, we prove Theorem 1.6, which gives, for any given dashed pattern (τ, X) , the asymptotic behavior of the sequence $(O_{\tau, X}^{(N)})_{N \geq 1}$ of random variables.

Proof. As in the previous examples, we write the quantity we want to study in terms of the variables $B_{i,s}^{(N)}$. Here,

$$O_{\tau, X}^{(N)} = \sum_{\substack{i_1 < \dots < i_p \\ \text{for all } x \in X, i_{x+1} = i_x + 1}} \sum_{\substack{s_1, \dots, s_p \\ s_{\tau-1(1)} < \dots < s_{\tau-1(p)}}} B_{i_1, s_1}^{(N)} \dots B_{i_p, s_p}^{(N)}.$$

Expanding its cumulants by multilinearity, we get (step 1)

$$(23) \quad k_\ell(O_{\tau, X}^{(N)}) = \sum_{(i_j^r)} \sum_{(s_j^r)} \kappa \left(B_{i_1^1, s_1^1}^{(N)} \dots B_{i_p^1, s_p^1}^{(N)}, \dots, B_{i_1^\ell, s_1^\ell}^{(N)} \dots B_{i_p^\ell, s_p^\ell}^{(N)} \right).$$

The first (resp. second) summation index is the set of matrices (i_j^r) (resp. (s_j^r)) with $(j, r) \in [p] \times [\ell]$ such that:

- for all r , $i_1^r < \dots < i_p^r$ (resp. $s_{\tau-1(1)}^r < \dots < s_{\tau-1(p)}^r$);
- for all r , for all $x \in X$, $i_{x+1}^r = i_x^r + 1$ (resp. no extra condition on the s 's).

Given such lists \mathbf{i} and \mathbf{s} , we consider the four following graphs:

- H_1 has vertex set $[p] \times [\ell]$ and has an edge between (j, r) and (k, t) if $|i_j^r - i_k^t| \leq 1$ and $s_j^r = s_k^t$;
- H_2 has vertex set $[p] \times [\ell]$ and has an edge between (j, r) and (k, t) if

$$\{i_j^r, i_j^r + 1, s_j^r\} \cap \{i_k^t, i_k^t + 1, s_k^t\} \neq \emptyset.$$

- H_3 has vertex set $([p] \times [\ell]) \sqcup ([p] \times [\ell])$ and has an edge between (j, r) and (k, t) (resp. (j, r) and $(\overline{k}, \overline{t})$; $(\overline{j}, \overline{r})$ and $(\overline{k}, \overline{t})$) if $|i_j^r - i_k^t| \leq 1$ (resp. $s_k^t - i_j^r = 0$ or 1 ; $s_j^r = s_k^t$).
- H_2' has vertex set $[\ell]$ and has an edge between r and t if

$$\left(\bigcup_{1 \leq j \leq p} \{i_j^r, i_j^r + 1, s_j^r\} \right) \cap \left(\bigcup_{1 \leq k \leq p} \{i_k^t, i_k^t + 1, s_k^t\} \right) \neq \emptyset.$$

The graphs H_1 and H_2 are respectively the surcontraction and contraction of H_3 , as defined in Section 3. Therefore, one has, by Lemma 3.2:

$$\#(\text{CC}(H_3)) \leq \#(\text{CC}(H_1)) + \#(\text{CC}(H_2)).$$

But one can further contract H_2 by the map $f : [p] \times [\ell] \rightarrow [\ell]$ defined by $f(j, r) = r$ and we obtain H_2' . With the notation of Section 3, it implies:

$$\#(\text{CC}(H_2)) \leq \#(\text{CC}(H_2')) + \sum_{r=1}^{\ell} [\#(\text{CC}(H_2[[p] \times \{r\}])) - 1].$$

But each induced graph $H_2[[p] \times \{r\}]$ (for $1 \leq r \leq \ell$) contains at least an edge between (x, r) and $(x+1, r)$ for each $x \in X$ (because we assumed that $i_{x+1}^r =$

$i_x^r + 1$). Thus it has at most $p - q$ connected components. Finally,

$$(24) \quad \#(\text{CC}(H_3)) \leq \#(\text{CC}(H_1)) + \#(\text{CC}(H_2')) + (p - q - 1)\ell.$$

Let us apply the main lemma (Theorem 1.4) to obtain a bound for

$$\left| \kappa \left(B_{i_1^1, s_1^1}^{(N)} \cdots B_{i_p^1, s_p^1}^{(N)}, \dots, B_{i_1^\ell, s_1^\ell}^{(N)} \cdots B_{i_p^\ell, s_p^\ell}^{(N)} \right) \right|.$$

In this case, the number of different couples in the indices of the Bernoulli variables is at least the number of connected components of H_1 . Besides, the graph G_2' introduced in paragraph 1.5 has the same vertex set, but fewer edges than H_2' . Hence, it has more connected components and we obtain:

$$\left| \kappa \left(B_{i_1^1, s_1^1}^{(N)} \cdots B_{i_p^1, s_p^1}^{(N)}, \dots, B_{i_1^\ell, s_1^\ell}^{(N)} \cdots B_{i_p^\ell, s_p^\ell}^{(N)} \right) \right| \leq C_{p\ell} N^{-\#(\text{CC}(H_1)) - \#(\text{CC}(H_2')) + 1}.$$

Using inequality above, this can be rewritten as (step 2)

$$\left| \kappa \left(B_{i_1^1, s_1^1}^{(N)} \cdots B_{i_p^1, s_p^1}^{(N)}, \dots, B_{i_1^\ell, s_1^\ell}^{(N)} \cdots B_{i_p^\ell, s_p^\ell}^{(N)} \right) \right| \leq C_{p\ell} N^{-\#(\text{CC}(H_3)) + (p-q-1)\ell + 1}.$$

As in the previous paragraph, Lemma 6.1 asserts that the number of couples of lists $((i_j^r), (s_j^r))$ such that $\#(\text{CC}(H_3)) = t$ is smaller than $C_{2p\ell, \mathbf{D}}'' N^t$ for a well chosen \mathbf{D} (step 3). Hence their total contribution to Equation (23) is bounded from above by the quantity $C_{p\ell} C_{p\ell, \mathbf{D}}'' N^{(p-q-1)\ell + 1}$. Finally, one has:

$$(25) \quad \kappa_\ell(O_{(X, \tau)}^{(N)}) = O(N^{(p-q-1)\ell + 1}),$$

or equivalently $\kappa_\ell(Z_{(X, \tau)}^{(N)}) = O(N^{-\ell/2 + 1})$. As in paragraph 5.4, the theorem follows from this bound and from the limits of the normalized expectation and variance.

For the expectation, we have to consider the case $\ell = 1$. In this case, one has $\#(\text{CC}(H_1)) = p$ and $\#(\text{CC}(H_2')) = 1$. Therefore, if we want an equality in Equation (24), we need $\#(\text{CC}(H_3)) = 2p - q$, which implies that all entries in the lists \mathbf{i} and \mathbf{s} are distinct. For these lists, one has (Lemma 2.1)

$$\kappa(B_{i_1^1, s_1^1}^{(N)} \cdots B_{i_p^1, s_p^1}^{(N)}) = \mathbb{E}(B_{i_1^1, s_1^1}^{(N)} \cdots B_{i_p^1, s_p^1}^{(N)}) = \frac{1}{(N + \theta)_p}.$$

But the number of lists with distinct entries in the index set of equation (23) is asymptotically $\frac{N^{2p-q}}{p!(p-q)!}$. Finally,

$$\lim_{N \rightarrow \infty} \frac{1}{N^{p-q}} \mathbb{E}(O_{(X, \tau)}^{(N)}) = \frac{1}{p!(p-q)!}.$$

It remains to prove that the renormalized variance $N^{-2(p-q)+1} \kappa_2(O_{(X, \tau)}^{(N)})$ has a limit $V_{\tau, X} \geq 0$, when N tends to infinity. But this follows from the bound (25) and the fact that any $\kappa_\ell(O_{(X, \tau)}^{(N)})$ is a rational function in N . Let us explain the latter fact.

Recall that $\kappa_\ell(O_{(X, \tau)}^{(N)})$ is given by equation (23). We can split the sum depending on the graph H_3 associated to the matrices \mathbf{i} and \mathbf{s} and on the actual value $\delta_\epsilon(\mathbf{i}, \mathbf{s})$

of $i_j^r - i_k^t$ (or $s_k^t - i_j^r$ and $s_k^t - s_j^r$ respectively) for each edge e of H_3 . Then the fact that $\kappa_\ell(O_{(X,\tau)}^{(N)})$ is a rational function is an immediate consequence of the following points:

- the number of graphs H_3 and possible values for the differences $\delta_e(\mathbf{i}, \mathbf{s})$ (for $e \in E_{H_3}$) are finite;
- the cumulant $\kappa(B_{i_1^1, s_1^1}^{(N)} \cdots B_{i_p^1, s_p^1}^{(N)}, \dots, B_{i_1^\ell, s_1^\ell}^{(N)} \cdots B_{i_p^\ell, s_p^\ell}^{(N)})$ is a rational function in N which depends only on the graph H_3 and values of $\delta_e(\mathbf{i}, \mathbf{s})$ (for $e \in E_{H_3}$);
- the number of matrices \mathbf{i} and \mathbf{s} corresponding to a given graph G and given values $\delta_e(\mathbf{i}, \mathbf{s})$ is a polynomial in N . \square

Remark 6.5. Fix a dashed pattern (τ, X) and consider a *uniform* random permutation σ . Let us denote A_{i_1, \dots, i_p} the event : i_1, \dots, i_p is an occurrence of (τ, X) in σ . Then the events A_{i_1, \dots, i_p} and A_{j_1, \dots, j_p} are independent as soon as the lists \mathbf{i} and \mathbf{j} are disjoint. This allows to build a simpler proof of the result above, still using the moment method (see [16] for a proof of a similar result in the context of random words). However, it does not hold in the case of Ewens random permutations with a general parameter θ .

6.4. Generalized patterns and local statistics. The notion of dashed patterns has been recently generalized by several authors in [7, Section 2]. The idea is roughly that, in an occurrence of a generalized pattern, one can ask that some values are consecutive (and not only some places as in dashed patterns). It would be interesting to give a general theorem on the asymptotic behavior of the number of occurrences of a given generalized pattern. This seems to be a hard problem as many different behavior can occur:

- The number of adjacencies is the sum of the number of occurrences of two different generalized patterns and converge towards a Poisson distribution.
- The dashed patterns are special cases of generalized patterns. As we have seen in the previous paragraph, their number of occurrences converges, after normalization, towards a Gaussian law. Other generalized patterns exhibit the same behavior, for example the one considered in [7] (the proof is the same as for dashed patterns; note that Remark 6.5 does not hold for occurrences of this pattern).
- Other behaviors can occur: for example, it is easy to see that the number of occurrences of the pattern $(123, \{1\}, \{1\})$ (we use the notations of [7]), has an expectation of order n , but a probability of being 0 with a positive lower bound.

Even if we have not been able to give a general statement, our approach unifies the first two cases.

The notion of generalized patterns can be further extended to the one of *local statistic*. Fix a integer $p \geq 1$ and a set S of constraints: a constraint is an equality or inequality (large or strict) whose members are of the form $i_j + d$ or $s_j + d$ where j belongs to $[p]$ and d is some integer. Then, for a permutation σ of S_N , we define

$O_{p,S}^{(N)}(\sigma)$ as the number of lists i_1, \dots, i_p and s_1, \dots, s_p satisfying the constraints in S and such that $\sigma(i_j) = s_j$ for all j in $[p]$.

We call any linear combination of statistics $O_{p,S}^{(N)}$ a local statistic. The number of occurrences of a generalized patterns, but also the number of excedances or of cycles of a given length p , are examples of local statistics. The method presented in this article is suitable for the asymptotic study of joint vectors of local statistics. We have failed to find a general statement, but we are convinced that our approach can be adapted to many more examples than the ones studied in this article.

However, the method does not seem appropriate to global statistics, such as the total number of cycles of the permutation or the length of the longest cycle.

ACKNOWLEDGEMENTS

I would like to thank Jean-Christophe Aval, Adrien Boussicault and Philippe Nadeau for sharing their interest in permutation tableaux, Mireille Bousquet-Mélou and Mathilde Bouvel for pointing out some references and, last but not least, Jean-François Marckert for a lot of explanations and references on convergence of random functions and Gaussian processes (in particular, his notes from lectures given in Graz, available online, have been very useful for me).

REFERENCES

- [1] R. Arratia, A. Barbour, and S. Tavaré. *Logarithmic combinatorial structures: a probabilistic approach*. EMS Monographs in Mathematics. EMS, Zürich, 2003.
- [2] J. Aval, A. Boussicault, and P. Nadeau. Tree-like tableaux. In *Proceedings of FPSAC*, volume AO, pages 63–74. DMTCS, 2011.
- [3] E. Babson and E. Steingrímsson. Generalized permutation patterns and a classification of the Mahonian statistics. *Sém. Lothar. Combin.*, 44:117–134, 2000.
- [4] P. Billingsley. *Probability and measure*. Wiley, 1995. 3rd edition.
- [5] P. Billingsley. *Convergence of probability measures*. Wiley, New York, 1999. 2nd edn.
- [6] J. Bourdon and B. Vallée. Pattern matching statistics on correlated sources. In *LATIN 2006*, volume 3887 of *Lecture Notes in Computer Science*, pages 224–237. Springer, 2006.
- [7] M. Bousquet-Mélou, A. Claesson, M. Dukes, and S. Kitaev. $(2+2)$ -free posets, ascent sequences and pattern avoiding permutations. *J. Comb. Th., Ser. A*, 117(7):884–909, 2010.
- [8] M. Bouvel, C. Chauve, M. Mishna, and D. Rossin. Average-case analysis of perfect sorting by reversals. *Disc. Math., Algorithms and Applications*, 3(3):369–392, 2011.
- [9] S. Corteel and P. Nadeau. Bijections for permutation tableaux. *Eur. J. Comb.*, 30(1):295–310, 2009.
- [10] S. Corteel and L. Williams. Tableaux combinatorics for the asymmetric exclusion process. *Adv. Appl. Math.*, 39(3):293–310, 2007.
- [11] B. Derrida. Matrix ansatz and large deviations of the density in exclusion processes. In *Proceedings of the International Congress of Mathematicians, Madrid*, pages 367–382, 2006.
- [12] B. Derrida, J. Lebowitz, and E. Speer. Entropy of open lattice systems. *J. Stat. Phys.*, 126(4):1083–1108, 2007.
- [13] W. Ewens. The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, 3:87–112, 1972.
- [14] R. Fisher. Moments and product moments of sampling distributions. *Proc. Lond. Math. Soc. Series 2*, 30:199–238, 1929.
- [15] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

- [16] P. Flajolet, W. Szpankowski, and B. Vallée. Hidden word statistics. *J. ACM*, 53(1):147–183, 2006.
- [17] V. Goncharov. Some facts from combinatorics. *Izvestia Akad. Nauk. SSSR, Ser. Mat.*, 8:3–48, 1944.
- [18] I. Kaplansky. The asymptotic distribution of runs of consecutive elements. *Ann. Math. Stat.*, 16(2):200–203, 1945.
- [19] V. Kolchin. A problem of the allocation of particles in cells and cycles of random permutations. *Theory of Probability and its Applications*, 16:74, 1971.
- [20] M. Lothaire. *Combinatorics on words*, volume 17 of *Encyclopedia of Math. and its Appl.* Addison-Wesley, 1983.
- [21] J. Marckert. One more approach to the convergence of the empirical process to the brownian bridge. *Electronic Journal of Statistics*, 2:118–126, 2008.
- [22] P. Nicodème, B. Salvy, and P. Flajolet. Motif statistics. In *European Symposium on Algorithm*, volume 1643 of *Lecture Notes in Computer Science*, pages 194–211. Springer, 1999.
- [23] A. Postnikov. Total positivity, Grassmannians, and networks. arXiv preprint math/0609764v1, 2006.
- [24] G. Rota and B. Taylor. The classical umbral calculus. *SIAM Journal on Mathematical Analysis*, 25:694, 1994.
- [25] P. Śniady. Gaussian fluctuations of characters of symmetric groups and of Young diagrams. *Probab. Theory Related Fields*, 136(2):263–297, 2006.
- [26] T. P. Speed. Cumulants and partition lattices I. *Austr. Jour. Stat.*, 25(2):378–388, 1983.
- [27] E. Steingrímsson and L. Williams. Permutation tableaux and permutation patterns. *J. Combin. Theory Ser. A*, 114(2):211–234, 2007.
- [28] G. Watterson. The sampling theory of selectively neutral alleles. *Adv. Appl. Prob.*, 6:463–488, 1974.
- [29] L. K. Williams. Enumeration of totally positive Grassmann cells. *Adv. Math.*, 190(2):319–342, 2005.
- [30] J. Wolfowitz. Note on runs of consecutive elements. *Ann. Math. Stat.*, 15(1):97–98, 1944.
- [31] W. Xu, B. Alain, and D. Sankoff. Poisson adjacency distributions in genome comparison: multichromosomal, circular, signed and unsigned cases. *Bioinformatics*, 24(16):i146, 2008.

LABRI, UNIVERSITÉ BORDEAUX 1, 351 COURS DE LA LIBÉRATION, 33 400 TALENCE, FRANCE

E-mail address: feray@labri.fr