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 pattern. 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. Several statistics have been widely studied: total number of cycles, number of cycles of a given length, of descents, inversions, exceedances 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.

A probabilistic point of view on the topic raises other questions. Let us consider, for each N, a probability measure μ_N on permutations of size N. 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 [16]. It is defined by

$$\mu_N(\{\sigma\}) = \frac{\theta^{\#(\sigma)}}{\theta(\theta+1)\cdots(\theta+N-1)}, \tag{1.1}$$

*LaBRI, CNRS, Université Bordeaux 1, France. E-mail: valentin.feray@labri.fr

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

Having chosen a sequence of probability disribution of S_N, any statistic on permutations 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 purpose of this article is to introduce a new general approach to this family of problems, based on the method of moments.

We then use it to determine the second-order fluctuations of a large family of statistics on permutations: occurrences of dashed patterns (Theorem 1.8).

Random permutations, either with uniform or Ewens distribution, are well-studied objects. Giving a complete list of references is impossible. In Section 1.5, we compare our results with the literature.

1.2 Motivating examples

Let us begin by describing a few examples of results, covered by our method.

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 [22] in the case of uniform measure and by G.A. Watterson [30, Theorem 5] in the framework of a general Ewens distribution (see also [1, Theorem 5.1]).

Theorem 1.1 ([30]). *Let p be a positive integer. When N tends to infinity, $\Gamma_p^{(N)}$ converges in distribution towards a Poisson law with parameter θ/p. Moreover, the sequences of random variables $(\Gamma_{p'}^{(N)})_{N \geq 1}$ for $p' \leq p$ are asymptotically independent.*

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

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

When σ is a Ewens random permutation, this random variable is distributed according to a Bernoulli law with parameter $\frac{i+\theta}{N+\theta-1}$ (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

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

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 sections 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 and σ a random Ewens permutation of size N. Then, almost surely,*

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

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 $x_1, \ldots, x_r$, the vector $(Z_\sigma^{(N)}(x_1), \ldots, Z_\sigma^{(N)}(x_r))$ converges towards a Gaussian vector $(G(x_1), \ldots, G(x_r))$ with covariance matrix $(K(x_i, x_j))_{1 \le i,j \le r}$, for some explicit function K (see Section 5.4).

Although we have no interpretation for that, let us note that the limit of $F_\sigma^{(N)}(x)$ is the cumulative distribution function of a β variable with parameters 1 and 2.

With this formulation, Theorem 1.2 is new, but the first part is quite easy while the second is a consequence of [15, Appendix A] (see Section 5). We also refer to an article of A. Barbour and S. Janson [5], where the case of the uniform measure is addressed with another method.

Adjacencies. We consider here only uniform random permutations, that is the case $\theta = 1$. 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 $B_i^{\mathrm{ad},N}$ which takes value 1 if i is an adjacency and 0 otherwise. Then $B_i^{\mathrm{ad},N}$ is distributed according to a Bernoulli law with 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 $A^{(N)} = \sum_{i=1}^{N-1} B_i^{\mathrm{ad},N}$.

Theorem 1.3 ([32]). *$A^{(N)}$ converges in distribution towards a Poisson variable with parameter 2.*

This result first appeared in papers of J. Wolfowitz and I. Kaplansky [32, 21] and was rediscovered recently in the context of genomics (see [33] and also [11, Theorem 10]).

In these three examples, the underlying variables behave asymptotically as 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 of the three results presented above in a uniform way. Besides, our proofs follow the intuition that events involving the image of different integers are almost inedependent.

1.3 The main lemma

From now on, N is a positive integer and σ a random Ewens permutation in S_N. We shall use the standard notation $[N]$ for the set of the first N positive integers.

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 $\sigma(i) = s$. Despite its simple definition, this collection of events allows to reconstruct the permutation and thus generates the full algebra of observables (we call them *elementary events*).

For random variables $X_1, \ldots, X_\ell$ on the same probability space (with expectation denoted $\mathbb{E}$), we define their joint cumulant

$$\kappa(X_1, \ldots, X_\ell) = [t_1 \ldots t_\ell] \ln\left(\mathbb{E}\big(\exp(t_1 X_1 + \cdots + t_\ell X_\ell)\big)\right). \tag{1.2}$$

As usual, $[t_1 \ldots t_\ell]F$ stands for the coefficient of $t_1 \ldots t_\ell$ in the series expansion of F in positive powers of $t_1, \ldots, t_\ell$. Joint cumulants have been introduced by Leonov and Shiryaev [23]. For a summary of their most useful properties, see [20, Proposition 6.16].

Our main lemma is a bound on joint cumulants of products of elementary events. To state it, we introduce the following notations. Consider two lists of positive integers of the same length $\mathbf{i} = (i_1, \ldots, i_r)$ and $\mathbf{s} = (s_1, \ldots, s_r)$ 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 $\mathrm{CC}(G)$. Besides, if Π is a set-partition Π, we write $\#(\Pi)$ for its number of parts. In particular, $\#(\mathrm{CC}(G))$ is the number of connected components of G.

Finally, if π_1 and π_2, we denote $\pi_1 \vee \pi_2$ the finest partition which is coarser than π_1 and π_2 (here, *fine* and *coarse* refer to the refinement order; see Section 1.7).

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 $\tau = (\tau_1, \ldots, \tau_\ell)$ of $[r]$, any $N \geq 1$ and lists $\mathbf{i} = (i_1, \ldots, i_r)$ and $\mathbf{s} = (s_1, \ldots, s_r)$ of integers in $[N]$, one has:*

$$\left| \kappa \left(\prod_{j \in \tau_1} B_{i_j, s_j}^{(N)}, \ldots, \prod_{j \in \tau_\ell} B_{i_j, s_j}^{(N)} \right) \right| \leq C_r N^{-\#\left(\mathrm{CC}(G_1(\mathbf{i}, \mathbf{s})) \right) - \#\left(\mathrm{CC}(G_2(\mathbf{i}, \mathbf{s})) \vee \tau \right) + 1}. \tag{1.3}$$

Note that the integer $\#\left(\mathrm{CC}(G_1(\mathbf{i}, \mathbf{s})) \right)$ is the number of *different* pairs (i_j, s_j). The second quantity involved in the theorem $\#\left(\mathrm{CC}(G_2(\mathbf{i}, \mathbf{s})) \vee \tau \right)$ does not have a similar interpretation. However, it admits an equivalent description. Consider the graph G_2', obtained from $G_2(\mathbf{i}, \mathbf{s})$ by merging vertices corresponding to elements in the same part of τ. Then $\#\left(\mathrm{CC}(G_2(\mathbf{i}, \mathbf{s})) \vee \tau \right)$ is the number of connected components of G_2'.

As an example, let us consider the *distinct case*, that is we assume that the entries in the lists $\mathbf{i}$ and $\mathbf{s}$ are distinct. We shall use the standard notation for falling factorials $(x)_a = x(x-1)\cdots(x-a+1)$. In this case, the expectation of a product of $B_{i,s}^{(N)}$ is simply $1/(N + \theta - 1)_a$, where a is the number of factors (the case $\theta = 1$ is obvious, while the general case is explained in Lemma 2.1). Joint cumulants can be expressed in terms of joint moments – see [20, Proposition 6.16 (vi)] –, so the left-hand side of (1.3) can be written as an explicit rational function in N of degree $-r$. According to our main lemma, 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 $B_{i_j, s_j}^{(N)}$ behave asymptotically as independent (joint cumulants vanish when the set of variables can be split into two mutually independent sets).

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. This may help to find a similar behaviour (that is random variables with small cumulants) in completely different contexts, see Section 1.6.*

1.4 Applications

Recall that, if $Y^{(1)}, \ldots, Y^{(m)}$ are random variables such that the law of the m-tuple $(Y^{(1)}, \ldots, Y^{(m)})$ is entirely determined by its joint moments, then the two following statements are equivalent (see [6, Theorem 30.2] for the analogous property in terms of moments).

- For any ℓ and any list $i_1, \ldots, i_\ell$ in $[m]$,

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

- The sequence of vectors $(X_n^{(1)}, \ldots, X_n^{(m)})$ converges in distribution towards the vector $(Y^{(1)}, \ldots, Y^{(m)})$.

As Gaussian and Poisson variables are determined by their moments (see *e.g.* the criterion [6, Theorem 30.1]), cumulants can be used to prove convergence in distribution towards Gaussian or Poisson variables, such as the results of Section 1.2.

Theorem 1.4 can be used to give new uniform 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.

To give more evidence that our approach is quite general, we study the number of occurrences of *dashed patterns*. This notion has been introduced[1] in 2000 by E. Babson and E. Steingrímsson [3].

Definition 1.6. *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 $i_1 < \cdots < i_p$ such that:*

- *for any $x \in X$, one has $i_{x+1} = i_x + 1$.*

- *$\sigma(i_1), \ldots, \sigma(i_p)$ is in the same relative order as $\tau(1), \ldots, \tau(p)$.*

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

Example 1.7. *$O_{21,\emptyset}^{(N)}$ is the number of inversions, 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 patten or as a linear combination of such statistics, see [3, Section 2].*

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 1.8. *Let (τ, X) be a dashed pattern of size p (see definition 1.6) and σ_N a sequence of random Ewens permutations. 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:*

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

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

This theorem is proved in Section 6.3 using Theorem 1.4.

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 factor). We formulate it as a conjecture.

Conjecture 1.9. *For any dashed pattern (τ, X), one has $V_{\tau,X} > 0$.*

The following partial result has been proved by M. Bóna [9, Propositions 1 and 2] (M. Bóna works with the uniform distribution, but it should not be too hard to show that $V_{\tau,X}$ does not depend on θ).

Proposition 1.10. *For any $k \geq 1$, $\tau = \mathrm{Id}_k$ and $X = \emptyset$ or $X = [k-1]$, Conjecture 1.9 holds true.*

The proof relies on an expression of $V_{\tau,X}$ as a signed sum of products of binomial coefficients. This expression can be extended to the general case and we have checked by computer that Conjecture 1.9 holds true for all patterns of size 8 or less.

---

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

1.5 Comparison with other methods

There is a huge literature on random permutations. While we will not make a comprehensive survey of the subject, we shall try to present the existing methods and results related to our paper.

Our Poisson convergence results have been obtained previously by the moment method in the articles [21] and [30]. Our cumulant approach is not really different from these proofs. Yet, we have chosen to present these examples for two reasons:

- first, it illustrates the fact that our approach can prove in a uniform ways convergence towards different distributions ;

- second, the combinatorics is simpler in the Poisson cases, so they serve as toy model to explain the general structure of the proofs.

Let us mention also the existence of a powerful method, called the Stein-Chen method, that proves Poisson convergence, together with precise bounds on total variation distances – see, *e.g.*, [4, Chapter 4].

Let us now consider our normal approximation results. For uniform permutations, both are already known or could be obtained easily with methods existing in the literature.

- Theorem 1.2 has been proved by A. Barbour and S. Janson [5], who established a functional version of a combinatorial central limit theorem from Hoeffding [19]. This theorem deals with statistics of the form

$$\sum_{1 \leq i,j \leq N} a_{i,j}^{(N)} B_{i,j}^{(N)}$$

where $A^{(N)}$ is a sequence of deterministic $N \times N$ matrices.

- Theorem 1.8 has been proved for some particular patterns using dependency graphs and cumulants: see [9, Theorems 10 and 17] and [18, Section 6]. The case of a general pattern (under uniform distribution) can be handled with the same arguments.

These methods are very different one from each other and none of them can be used to prove both results in a uniform way. Note also that they only work in the uniform case. Yet, going from the uniform model to a general Ewens distribution should be doable using the *chinese restaurant process* [1, Example 2.4] (with this coupling, an Ewens random permutation differs from a uniform random permutation by $O(2|\theta - 1| \log(n))$ values).

To conclude, while less powerful in the Poisson case, our method has the advantage of providing a uniform proof for all these results and to extend directly to a general Ewens distribution.

1.6 Future work

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

It would be interesting to describe which permutation statistics can be (asymptotically) studied with our approach. This problem is discussed in Section 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 family of objects. The extension to colored permutations should be straightforward. A promising

direction is the following: consider a graph G with vertex set $[n]$ and take some random subset S of its vertices, uniformly among all subsets of size p. If p grows linearly with n, then the events "i *lies in S*" (for $1 \le i \le n$) have small joint cumulants (this is easy to see with the material of Section 2).

1.7 Preliminaries: set partitions

The combinatorics of set partitions is central in the theory of cumulants and will be important in this article. We recall here some well-known facts about them.

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 $\pi \le \pi'$) 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 $\mathrm{rk}(\pi)$ 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 any two set partitions π and π',

$$\mathrm{rk}(\pi \vee \pi') \le \mathrm{rk}(\pi) + \mathrm{rk}(\pi'). \tag{1.4}$$

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:

$$\mu(\pi, \{S\}) = (-1)^{\#(\pi)-1}(\#(\pi) - 1)!. \tag{1.5}$$

1.8 Outline of the paper

The paper is organized as follows. Section 2 contains the proof of the 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 cycles, Section 5 for exceedances and finally, Section 6 for generalized patterns (including adjacencies and dashed patterns).

2 Proof of the main lemma

This section is devoted to the proof of Theorem 1.4. It is organized as follows. First, we give a simple formula for the joint moments of the elementary events $(B_{i,s})$. Second, we establish a general criterion based on joint moments that implies that the corresponding variables have small joint cumulants. Third, this criterion is used to prove Theorem 1.4 in the case of distinct indices. The general case finally can be deduced from this particular case, as shown in the last part of this section.

2.1 Joint moments

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

Note that $(B_{i,s}^{(N)})^2 = B_{i,s}^{(N)}$, while $B_{i,s}^{(N)} B_{i,s'}^{(N)} = 0$ if $s \ne s'$ and $B_{i,s}^{(N)} B_{i',s}^{(N)} = 0$ if $i \ne i'$. 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 distinct indices (some entry of the list $\mathbf{i}$ can be equal to an entry of $\mathbf{s}$).

We see these two lists as a *partial permutation*

$$\widetilde{\sigma}_{\mathbf{i},\mathbf{s}} = \left(\begin{array}{ccc} i_1 & \dots & i_r \\ s_1 & \dots & s_r \end{array} \right),$$

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 $\widetilde{\sigma}_{\mathbf{i},\mathbf{s}}$ is denoted $\#(\widetilde{\sigma}_{\mathbf{i},\mathbf{s}})$.

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 random Ewens permutation σ of size $N - 1$. Write it as a product of cycles and apply the following random transformation.

- With probability $\theta/(N+\theta-1)$, 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 $1/(N+\theta-1)$, 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 Ewens permutation of size N. Iterating this, one obtains a linear time and space algorithm to pick a random Ewens permutation.

Let us come back now to the computation of joint moments. The following lemma may be known, but the author has not been able to find it in the literature.

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

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

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

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

Proof. As Ewens measure is constant on conjugacy classes of S_N, one can assume without loss of generality that $i_1 = N - r + 1$, $i_2 = N - r + 2$, ..., $i_r = N$. 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 $\widetilde{\sigma}_{\mathbf{i},\mathbf{s}}(i_j)$, $\widetilde{\sigma}_{\mathbf{i},\mathbf{s}}^2(i_j)$ 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 $\widetilde{\sigma}_{\mathbf{i},\mathbf{s}}$. 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

$$\frac{\theta^{\#(\widetilde{\sigma}_{\mathbf{i},\mathbf{s}})}}{(N+\theta-1)\dots(N-r+\theta)}.$$

$\square$

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: for a subset $\Delta = \{j_1,\dots,j_h\}$ of $[\ell]$, we write

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

We also introduce the auxiliary quantity $U_{A,\Delta}^{(N)}$ implicitly defined by the property: for any subset $\Delta \subseteq [\ell]$,

$$\prod_{\delta \subseteq \Delta} U_{A,\delta}^{(N)} = M_{A,\Delta}^{(N)}.$$

Using Möbius inversion on the boolean lattice, we have explicitly: for any subset $\Delta \subseteq [\ell]$,

$$U_{A,\Delta}^{(N)} = \prod_{\delta \subseteq \Delta}\left(M_{A,\delta}^{(N)}\right)^{(-1)^{|\delta|}}$$

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*

$$U_{A,\Delta}^{(N)} = 1 + O(N^{-|\Delta|+1}); \tag{2.1}$$

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

$$\kappa_{A,\Delta}^{(N)} = O(N^{-|\Delta|+1}). \tag{2.2}$$

Proof. Let us consider the implication *I* $\Rightarrow$ *II*. We denote $T_\Delta^{(N)} = U_{A,\Delta}^{(N)} - 1$ 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 well-known relation between joint moments and cumulants [20, Proposition 6.16 (vi)]:

$$\kappa_{A,[\ell]}^{(N)} = \sum_{\pi \in \mathcal{P}([\ell])} \mu(\pi, \{[\ell]\}) \prod_{C \in \pi} M_{A,C}^{(N)}. \tag{2.3}$$

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 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 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,

$$\kappa_{A,[\ell]}^{(N)} = \sum_{\substack{\Delta_1,\ldots,\Delta_m \\ \text{distinct}}} T_{\Delta_1}^{(N)} \ldots T_{\Delta_m}^{(N)} \left(\sum_{\substack{\pi \in \mathcal{P}([\ell]) \\ \text{for all } i, \ \pi \geq \Pi(\Delta_i)}} \mu(\pi, \{[\ell]\}) \right). \tag{2.4}$$

The condition on π can be rewritten as

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

Hence, by definition of the Möbius function, the sum in the parenthesis is equal to 0, unless $\Pi(\Delta_1) \vee \cdots \vee \Pi(\Delta_m) = \{[\ell]\}$ (in other terms, unless the hypergraph with edges $(\Delta_i)_{1 \leq i \leq m}$ is connected). On the one hand, by Equation (1.4), it may happen only if:

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

On the other hand, one has

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

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 $A_1^{(N)}, \ldots, A_\ell^{(N)}$ such that *II* holds. By induction hypothesis, one gets immediately that

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

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

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

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

$$M_{A,[\ell]}^{(N)} = \prod_{\Delta \subsetneq [\ell]} (M_{A,\Delta}^{(N)})^{(-1)^{\ell-1-|\Delta|}} + O(N^{-\ell+1}). \tag{2.5}$$

Consider ℓ sequences of random variables $B_1^{(N)}, \ldots, B_\ell^{(N)}$ such that, for $\Delta \subsetneq [\ell]$, the equality $M_{B,\Delta}^{(N)} = M_{A,\Delta}^{(N)}$ holds, and such that Equation (2.5) 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 moments, 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 (2.5) and, hence, so does family A. $\qquad\square$

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

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

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

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

correspond to different power sums of $z_1, \ldots, z_T$. Thus we have to find a set $\{z_1, \ldots, z_T\}$ of complex numbers with specified power sums up to 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 (equation (2.3)). 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 [27].

2.3 Case of distinct indices

Recall that, in the statement of Theorem 1.4, we fix a set-partition τ and two lists $\mathbf{i}$ and $\mathbf{s}$ and we want to bound the quantity

$$\left| \kappa\left(\prod_{j\in\tau_1} B_{i_j,s_j}^{(N)}, \ldots, \prod_{j\in\tau_\ell} B_{i_j,s_j}^{(N)} \right) \right|.$$

We first consider 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 $N \geq 1$:

$$A_h^{(N)} = (N + \theta - 1)_{a_j} \prod_{j\in\tau_h} B_{i_j,s_j}^{(N)},$$

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

Let us prove that $A_1^{(N)}, \ldots, A_\ell^{(N)}$ fulfills property *I* of this lemma. Of course, the case $\Delta = [\ell]$ is generic. Thanks to Lemma 2.1, the joint moments of the family A have in this case an explicit expression: for $\delta \subseteq [\ell]$,

$$M_{A,\delta}^{(N)} = \frac{\displaystyle\prod_{j\in\delta}(N + \theta - 1)_{a_j}}{(N + \theta - 1)_{\sum_{j\in\delta} a_j}}.$$

Therefore, we have to prove that the quantity

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

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

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 $Q_{a_1,\ldots,a_{\ell-1},0} = 1$ and the statement holds.

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

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

Setting $X = N + \theta - 1 - a_\ell$, the second factor becomes

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

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

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

Using the terminology of lemma 2.2, it means that the list $A_1^{(N)},\ldots,A_\ell^{(N)}$ of sequences of random variables has the quasi-factorisation property. Thus it also has the small cumulant property and in particular

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

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

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

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

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

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

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

when X *is a positive number going to infinity.*

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

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

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

$$\left\{ \begin{array}{c} \text{lists of subsets} \\ \delta_1, \ldots, \delta_m \text{ of even size such} \\ \text{that, for all } h \leq m, j_h \in \delta_h \end{array} \right\} \quad \rightarrow \quad \left\{ \begin{array}{c} \text{lists of subsets} \\ \delta_1, \ldots, \delta_m \text{ of odd size such} \\ \text{that, for all } h \leq m, j_h \in \delta_h \end{array} \right\}$$
$$(\delta_1, \ldots, \delta_m) \quad \mapsto \quad (\delta_1 \nabla \{j_0\}, \ldots, \delta_m \nabla \{j_0\}),$$

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

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

Remark 2.5. *Thanks to a result of Leonov and Shiryaev that expresses cumulants of products of random variables as product of cumulants (see [23] or [28, Theorem 4.4]), it would have been enough to prove our result for $a_1 = \cdots = a_\ell = 1$. But, as our proof uses an induction on a_ℓ, we have not made this choice.*

Remark 2.6. *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 section is equivalent to Lemma 4.8 of paper [28]. 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 $A_1^{(N)}, \ldots, A_\ell^{(N)}$ 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

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

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

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

Lemma 2.7. *Let $A_1^{(N)}, \ldots, A_\ell^{(N)}$ 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 $\pi_0; \pi_1, \pi_2, \ldots$ are set partitions of $[\ell]$,*

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

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 (2.4):

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

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

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

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

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

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

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

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

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

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,

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

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, ..., C_h

$$\mathbb{E}(A_{C_1}^{(N)} \cdots A_{C_h}^{(N)}) = \frac{1}{(N + \theta - 1)(N + \theta - 2) \ldots (N + \theta - h)}.$$

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

$$\left((N + \theta - 1) A_C^{(N)} \right)_{C \in \pi_1}. \tag{2.6}$$

By the same argument as in Section 2.3, this 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 Section 2.1) that the corresponding product of moments 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 $i_j \neq i_h$;

- θ^γ 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,

$$Y_{\pi_1, \pi_2}^{(N)} = \frac{\alpha_{\pi_1, \pi_2}}{(N + \theta - 1)^{\#(\pi_1)}} \sum_{\substack{\pi \geq \tau \\ \pi \wedge CC(G_1(\mathbf{i}, \mathbf{s})) = \pi_1 \\ \pi \wedge CC(G_2(\mathbf{i}, \mathbf{s})) = \pi_2}} \prod_{C \in \pi} \mathbb{E} \left(\prod_{\substack{C' \in \pi_1 \\ C' \subseteq C}} (N + \theta - 1) A_{C'} \right). \tag{2.7}$$

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 (2.7) above is a truncated cumulant of the family (2.6) 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 $CC(G_1(\mathbf{i}, \mathbf{s}))$. $\qquad \square$

Remark 2.8. *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 on 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 graph quotients. These lemmas may already have appeared in the literature, though the author has not been able to find a reference. They will be useful in the next sections for applications of Theorem 1.4.
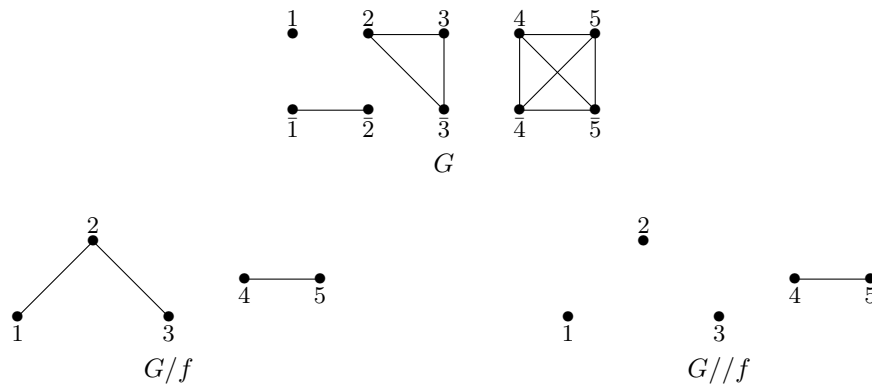
Figure 1: An example of a graph, its quotient and strong quotient.

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 f be a surjective map from V to another set W. Then the *quotient* 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 on the top 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.

3.2 Connected components of quotients

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

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

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 spanning 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 spanning 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

$$\#(\mathrm{CC}(G)) \le |V| - |F| = |W| - |F_{G/f}| + \sum_{w \in W} (|f^{-1}(W)| - 1 - |F_w|)$$

$$\le \#(\mathrm{CC}(G/f)) + \sum_{w \in W} (\#(\mathrm{CC}(G[f^{-1}(w)])) - 1). \quad \square$$

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

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

3.3 Fibers of size 2

In this section, we further assume that $V = W \sqcup W$ and that f is the canonical application $W \sqcup W \to W$ consisting in forgetting to which copy of W the element belongs. Throughout the paper, for simplicity of notation, we will use overlined letters for elements of the second copy of W.

In this context, in addition to the quotient 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 *strong quotient* of G.

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

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

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

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

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

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 of edges of G obtained as follows: for each edge of F_1, we take the two corresponding edges in G and for each edge of $F_2 \backslash F_1$, we take the corresponding edge in G (if there is several corresponding edges, choose one arbitrarily).

We will prove by contradiction that F is acyclic. Suppose that F contains a cycle C. Each edge of C projects on an edge in F_2 and thus the projection of C is a list $S = (e_1, \ldots, e_h)$ of *consecutive* edges in F_2 (*consecutive* means that we can orient the edges so 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 contains some pairs of edges of the form

$$\left(\{w, w'\}, \{\overline{w}, \overline{w'}\} \right)$$

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. 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 is acyclic.

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

$$\#(\mathrm{CC}(G)) \leq 2|W| - |F| = (|W| - |F_1|) + (|W| - |F_2|) = \#(\mathrm{CC}(G_1)) + \#(\mathrm{CC}(G_2)). \quad \square$$

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 using the variables $B_{i,s}^{(N)}$: here,

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

where $B_{(i_1, \ldots, i_p)}^{\mathrm{c}, N} = B_{i_1, i_2}^{(N)} \ldots B_{i_{p-1}, i_p}^{(N)} B_{i_p, i_1}^{(N)}$ is the indicator function of the event "$(i_1, \ldots, i_p)$ is a cycle of σ". Therefore, one has

$$\kappa_\ell(\Gamma_p^{(N)}) = \sum_{\substack{i_1^1 < i_2^1, i_3^1, \ldots, i_p^1 \\ \vdots \\ i_1^\ell < i_2^\ell, i_3^\ell, \ldots, i_p^\ell}} \kappa\big(B_{i_1^1, i_2^1}^{(N)} \ldots B_{i_p^1, i_1^1}^{(N)}, \cdots, B_{i_1^\ell, i_2^\ell}^{(N)} \ldots B_{i_p^\ell, i_1^\ell}^{(N)}\big). \tag{4.1}$$

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

Now, we would like to apply our main lemma to every summand of equation (4.1). To this purpose, one has to understand what is the exponent of N in the upper bound given by Theorem 1.4.

For a matrix

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

we denote:

- $M(\mathbf{i}) = |\{(i_j^r, i_{j+1}^r); 1 \leq j \leq p, 1 \leq r \leq \ell\}|$ the number of different entries in the matrix of pairs (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

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

- $t(\mathbf{i})$ the number of distinct entries.

Clearly, $M(\mathbf{i})$ is always at least equal to $t(\mathbf{i})$. 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:

$$\big|\kappa\big(B_{i_1^1, i_2^1}^{(N)} \ldots B_{i_p^1, i_1^1}^{(N)}, \cdots, B_{i_1^\ell, i_2^\ell}^{(N)} \ldots B_{i_p^\ell, i_1^\ell}^{(N)}\big)\big| \leq C_{p\ell} N^{-M(\mathbf{i}) - Q(\mathbf{i}) + 1}$$

$$\leq C_{p\ell} N^{-M(\mathbf{i})} \leq C_{p\ell} N^{-t(\mathbf{i})}. \tag{4.2}$$

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

As the number of summands in Equation (4.1) depends on N, we can not use directly inequality (4.2). We need a bound on the number of matrices $\mathbf{i}$ with a given value of $t(\mathbf{i})$.

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*

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

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 in t blocks 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 $[L]$. $\qquad\square$

Step 4: conclude.

By inequality (4.2) 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 $(i_1^{r_1}, \ldots, i_p^{r_1})$ and $(i_1^{r_2}, \ldots, i_p^{r_2})$ 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, then $G(\mathbf{i})$ is the complete graph and we get that i_j^r does not depend on r.

Finally

$$\kappa_\ell(\Gamma_p^{(N)}) = \sum_{i_1 < i_2, i_3, \ldots, i_p} \kappa_\ell\big(B_{i_1, i_2}^{(N)} \ldots B_{i_p, i_1}^{(N)}\big) + O(N^{-1}). \qquad (4.3)$$

But each $B_{i_1, i_2}^{(N)} \ldots B_{i_p, i_1}^{(N)}$ is a Bernoulli variable with parameter $\theta/(N + \theta - 1)_p$. Therefore their moments are all equal to $\theta/(N + \theta - 1)_p$ and by formula (2.3), their cumulants are $\theta/(N + \theta - 1)_p + O(N^{-2p})$. Finally, as there are $(N)_p/p$ terms in equation (4.3),

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

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

Moreover, a simple adaptation of the proof of Equation (4.3) implies that

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

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 $(\Gamma_1^{(N)}, \ldots, \Gamma_p^{(N)})$ tends in distribution towards a vector $(P_1, \ldots, P_p)$ where the P_i are independent Poisson-distributed random variables with respective parameters θ/i. $\qquad\square$

5 Number of exceedances

In this section, we look at our second motivating problem, the number of exceedances in random Ewens permutations. The first two subsections make a link between a physical statistics model and this problem, justifying our work. The last two subsections 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) from 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) from the right with probability $\frac{\delta}{N+1}$ (resp. $\frac{\beta}{N+1}$);

- with the remaining probability (we suppose $\alpha, \beta, \gamma, \delta < 1$ so 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 [14].

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 (or stationary distribution) 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 $\alpha = 1$, $\gamma, \delta = 0$. In this case, thanks to a result of S. Corteel and L. Williams [13], 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 [26, 31]. They have also been widely studied from a purely combinatorial point of view [29, 12, 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
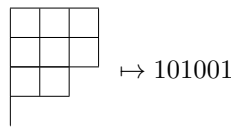
Figure 2: From the shape of a permutation tableau to a word in $\{0;1\}^{N-1}$.

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
$$\left\{ \left(n_1(w^T) - N F_{w^T}^{(N)}(x), -N(x - F_{w^T}^{(N)}(x)) - 1 \right) : x \in [0;1] \right\},$$
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 section. 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 [13, Theorem 3.1] but we do not know how to deal with the extra parameters there).

Theorem 5.1. *[13] 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 [12, 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:
$$w^T = (\delta_2(\sigma), \delta_3(\sigma), \ldots, \delta_{N+1}(\sigma)),$$
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 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 transformation on permutation [24, § 10.2], sends one of this statistics to the other. Take a permutation σ, written in its cycle notation so that:

- its cycles end with their minima;

- 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 exceedances in σ (a similar statement is given in [24, Theorem 10.2.3]).

From now on, we assume $\beta \cdot \theta = 1$. The properties above imply that $\mu_N^T(\beta)$ is the push-forward of the Ewens measure with 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

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

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 as $F_\sigma^{(N+1)}$, where σ is a random Ewens permutation of size N and $F_\sigma^{(N+1)}$ is the function defined in Section 1.2.

This was our original motivation to study $F_\sigma^{(N+1)}$.

5.3 Bounds for cumulants

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

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 $i_1, \ldots, i_\ell$ and $s_1, \ldots, s_\ell$ of integers in $[N]$,*

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

where C_ℓ is the constant defined by Theorem 1.4.

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

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

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

- its vertex set is $[\ell] \sqcup [\ell] = \{1, \bar{1}, \ldots, \ell, \bar{\ell}\}$;

- there is an edge between j and k (resp. j and $\bar{k}$, $\bar{j}$ and $\bar{k}$) 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 $G(\mathbf{i}, \mathbf{s})$ ($G_1(\mathbf{i}, \mathbf{s})$ and $G_2(\mathbf{i}, \mathbf{s})$ are respectively its strong and usual quotients). $\qquad\square$

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 $x_1, \ldots, x_\ell$, one has*

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

Proof. To simplify the notations, we suppose that $Nx_1, \ldots, Nx_\ell$ are integers, so that

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

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):

$$(N-1)^\ell \kappa(F_\sigma^{(N)}(x_1), \ldots, F_\sigma^{(N)}(x_\ell)) = \sum_{\substack{2 \leq i_1 \leq Nx_1 \\ \vdots \\ 2 \leq i_\ell \leq Nx_\ell}} \sum_{\substack{s_1 \geq i_1 \\ \vdots \\ s_\ell \geq i_\ell}} \kappa(B_{i_1,s_1}^{(N)}, \ldots, B_{i_\ell,s_\ell}^{(N)}). \qquad (5.1)$$

We apply Lemma 4.1 to the list $i_1, \ldots, i_\ell, s_1, \ldots, s_\ell$ and get that the number of pairs of lists $(\mathbf{i}, \mathbf{s})$ such that $|\{i_1, \ldots, i_\ell, s_1, \ldots, s_\ell\}|$ 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 pairs of lists $(\mathbf{i}, \mathbf{s})$ with $|\{i_1, \ldots, i_\ell, s_1, \ldots, s_\ell\}| = t$ to the right-hand side of (5.1) is smaller than $C'_{2\ell} C_\ell N$, which ends the proof of Proposition 5.3 (step 4). $\qquad \square$

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 $G(\mathbf{i}, \mathbf{s})$ associated to this pair of lists 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 $G(\mathbf{i}, \mathbf{s}) = G$. 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, so that they fulfill some inequalities. Finally, their total contribution to (5.1) is smaller than $C_5 N^{-1}$.

Comparison with a result of B. Derrida, J.L. Lebowitz and E.R. Speer. In [15, Appendix A], a *long range correlation phenomenon* for the *SSEP* is proved. Rewritten in terms of Ewens random permutations *via* the material of the previous section and with mathematical terminology, it asserts that, for $i_1 < \cdots < i_\ell$,

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

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 that is what our approach does.

5.4 Convergence results

In this section, 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:

$$\mathbb{E}(B_i^{\text{ex},N}) = \frac{N - i + \theta}{N + \theta - 1};$$

$$\text{Var}(B_i^{\text{ex},N}) = \frac{(i-1)(N - i + \theta)}{(N + \theta - 1)^2};$$

$$\text{Cov}(B_i^{\text{ex},N}, B_j^{\text{ex},N}) = -\frac{(n - j + \theta)(i - 1)}{(N + \theta - 1)^2(N + \theta - 2)} \text{ for } i < j,$$

from which we get the limits:

$$\lim_{N\to\infty} \mathbb{E}(F_\sigma^{(N)}(x)) = \int_0^x (1-t)dt + o(1) = \frac{1-(1-x)^2}{2}; \qquad (5.2)$$

$$\lim_{N\to\infty} N\operatorname{Cov}(F_\sigma^{(N)}(x), F_\sigma^{(N)}(y)) = \int_0^{\min(x,y)} t(1-t)dt \qquad (5.3)$$

$$- \int_{\substack{0\le t\le x\\ 0\le u\le y}} \min(t,u)(1-\max(t,u))dtdu.$$

We call $K(x,y)$ the right-hand side of the second equation. We begin with 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 (5.2) and (5.3). For the almost-sure convergence, we have to study the fourth centered moment.

From moment-cumulant formula (2.3) and using the fact that all cumulants 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,

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

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

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

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

Let us consider the second statement. Proposition 5.3 implies that, for any list $j_1,\dots,j_\ell$ 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 $(K(x_i,x_j))$. $\qquad \square$

The previous theorem deals with pointwise convergence. It is also possible to get some results for the random functions $(F_\sigma^{(N)})_{N\ge 1}$. 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 $(x \mapsto Z_\sigma^{(N)}(x))_{N\ge 1}$ converges in distribution towards the centered Gaussian process $x \mapsto G(x)$ with covariance function $\operatorname{Cov}(G(x), G(y)) = K(x,y)$.

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 [25, 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 $(x_i)_{0 \leq i \leq k}$ with $0 = x_0 < x_1 < \cdots < x_k = 1$, one has

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

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 [7, 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* [7, Section 5, Theorems 5.1 and 7.1], that is, for each $\epsilon > 0$ there exists some constant M such that:

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

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 (2.3) and the fact that $Z_\sigma^{(N)}(s)$ and $Z_\sigma^{(N)}(s')$ are centered, one has:

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

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,

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

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

We can now apply Theorem 10.2 of Billingsley's book [7] 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 (Subsection 6.2) and dashed patterns (Subsection 6.3). These two statistics belong in fact to the same general framework and we discuss in Subsection 6.4 the possibility of unifying our results.

The proofs in this section are a little bit more technical than the ones before and in particular we need a new lemma for step 3, given in Subsection 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 $i_1, \ldots, i_L$ 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 family of sets $(D_{\{j,k\}})_{1 \leq j < k \leq L}$, there exists a constant $C''_{L,\boldsymbol{D}}$ with the following property. For any $N \geq 1$ and $t \leq L$, the number of sequences $i_1, \ldots, i_L$ with entries in $[N]$, whose corresponding graph has exactly t connected components is bounded from above by $C''_{L,\boldsymbol{D}} N^t$.*

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, so 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 $\boldsymbol{D}$. $\square$

6.2 Adjacencies

In this section, we prove the following extension of Theorem 1.3.

Theorem 6.2. *Let σ_N be a sequence of random Ewens permutations, such that σ_N has size N. Then the number $A^{(N)}$ of adjacencies in σ_N converges in distribution towards a Poisson variable with 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):

$$\kappa_\ell(A^{(N)}) = \sum_{\substack{1 \leq i_1, s_1, \ldots, i_\ell, s_\ell \leq N \\ \epsilon_1, \ldots, \epsilon_\ell = \pm 1}} \kappa\left(B_{i_1, s_1}^{(N)} B_{i_1+1, s_1+\epsilon_1}^{(N)}, \cdots, B_{i_\ell, s_\ell}^{(N)} B_{i_\ell+1, s_\ell+\epsilon_\ell}^{(N)} \right). \tag{6.1}$$

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

- 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 has vertex set $[\ell] \sqcup [\ell]$ and has an edge between j and k (resp. j and $\bar{k}$, $\bar{j}$ and $\bar{k}$) if $|i_j - i_k| \le 2$ (resp. $|i_j - s_k| \le 2$, $|s_j - s_k| \le 2$)

We will use Theorem 1.4 to give a bound for

$$\left| \kappa\big(B^{(N)}_{i_1,s_1} B^{(N)}_{i_1+1,s_1+\epsilon_1}, \cdots, B^{(N)}_{i_\ell,s_\ell} B^{(N)}_{i_\ell+1,s_\ell+\epsilon_\ell}\big) \right|$$

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

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

is at least equal to $2\#(\mathrm{CC}(H_1)) \ge \#(\mathrm{CC}(H_1)) + 1$. Besides, in this case, the graph G'_2 introduced in Section 1.3 has the same vertex set as H_2 and fewer edges. Hence it has more connected components. Therefore, Theorem 1.4 implies (step 2):

$$\left| \kappa\big(B^{(N)}_{i_1,s_1} B^{(N)}_{i_1+1,s_1+\epsilon_1}, \cdots, B^{(N)}_{i_\ell,s_\ell} B^{(N)}_{i_\ell+1,s_\ell+\epsilon_\ell}\big) \right| \le C_{2\ell} N^{-\#(\mathrm{CC}(H_1)) - \#(\mathrm{CC}(H_2))}.$$

But, using the terminology of Section 3.3, the graphs H_1 and H_2 are the strong and usual quotients of H. Therefore, by Lemma 3.2, one has:

$$\#(\mathrm{CC}(H)) \le \#(\mathrm{CC}(H_1)) + \#(\mathrm{CC}(H_2)). \tag{6.2}$$

Besides, Lemma 6.1 implies the number of lists $\mathbf{i}$ and $\mathbf{s}$ with entries in $[N]$ such that H has exactly t connected components is bounded from above by $C''_{2\ell,\boldsymbol{D}} N^t$ for $\boldsymbol{D}$ well-chosen (step 3). In particular the constant $C''_{2\ell,\boldsymbol{D}}$ does not depend on N. Therefore, the total contribution of these lists to equation (6.1) is bounded from above by $C_{2\ell} N^{-t} \cdot C''_{2\ell,\boldsymbol{D}} N^t = C_{2\ell} \cdot C''_{2\ell,\boldsymbol{D}}$.

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

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

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

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

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

- if $s = i; \epsilon = -1$ (here $2 \le i \le N-1$) or $s = i+1; \epsilon = -1$ (here $1 \le i \le N-1$) or $s = i+2; \epsilon = -1$ (here $1 \le i \le N-2$), then the parameter is $\frac{\theta}{(N+\theta-1)(N+\theta-2)}$ ($3N-5$ cases);

- otherwise, the parameter is $\frac{1}{(N+\theta-1)(N+\theta-2)}$.

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

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

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

6.3 Dashed patterns

In this section, we prove Theorem 1.8, which describes, 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,

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

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

$$\kappa_\ell(O_{\tau,X}^{(N)}) = \sum_{(i_j^r)} \sum_{(s_j^r)} \kappa\left(B_{i_1^1,s_1^1}^{(N)} \ldots B_{i_p^1,s_p^1}^{(N)}, \ldots, B_{i_1^\ell,s_1^\ell}^{(N)} \ldots B_{i_p^\ell,s_p^\ell}^{(N)} \right). \tag{6.3}$$

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, $i_1^r < \cdots < i_p^r$ (resp. $s_{\tau^{-1}(1)}^r < \cdots < s_{\tau^{-1}(p)}^r$);

- 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 **i** and **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 has vertex set $(\overline{[p] \times [\ell]}) \sqcup ([p] \times [\ell])$ and has an edge between (j, r) and (k, t) (resp. (j, r) and $\overline{(k, t)}$; $\overline{(j, r)}$ and $\overline{(k, t)}$) 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 strong and usual quotients of H, as defined in Section 3. Therefore, one has, by Lemma 3.2:

$$\#(\mathrm{CC}(H)) \leq \#(\mathrm{CC}(H_1)) + \#(\mathrm{CC}(H_2)).$$

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

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

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

$$\#(\mathrm{CC}(H)) \leq \#(\mathrm{CC}(H_1)) + \#(\mathrm{CC}(H_2')) + (p - q - 1)\ell. \tag{6.4}$$

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

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

In this case, the number of different pairs 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 Section 1.3 has the same vertex set, but fewer edges than H_2'. Hence, it has more connected components and we obtain:

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

Using inequality (6.4) above, this can be rewritten as (step 2)

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

As in the previous section, Lemma 6.1 asserts that the number of pairs of lists $((i_j^r), (s_j^r))$ such that $\#(\mathrm{CC}(H)) = t$ is smaller than $C_{2p\ell,\boldsymbol{D}}'' N^t$ for a well chosen $\boldsymbol{D}$ (step 3). Hence their total contribution to Equation (6.3) is bounded from above by $C_{p\ell} C_{p\ell,\boldsymbol{D}}'' N^{(p-q-1)\ell+1}$. Finally, one has:

$$\kappa_\ell(O_{(X,\tau)}^{(N)}) = O(N^{(p-q-1)\ell+1}), \tag{6.5}$$

or equivalently $\kappa_\ell(Z_{(X,\tau)}^{(N)}) = O(N^{-\ell/2+1})$. As in Section 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 $\#(\mathrm{CC}(H_1)) = p$ and $\#(\mathrm{CC}(H_2')) = 1$. Therefore, if we want an equality in Equation (6.4), we need $\#(\mathrm{CC}(H)) = 2p - q$, which implies that all entries in the lists $\mathbf{i}$ and $\mathbf{s}$ are distinct. For these lists, one has (Lemma 2.1)

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

But the number of lists with distinct entries in the index set of equation (6.3) is asymptotically $\frac{N^{2p-q}}{p!(p-q)!}$. Finally,

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

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 (6.5) 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 (6.3). We can split the sum depending on the graph H associated to the matrices $\mathbf{i}$ and $\mathbf{s}$ and on the actual value $\delta_e(\mathbf{i}, \mathbf{s})$ 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. Then the fact that $\kappa_\ell(O_{(X,\tau)}^{(N)})$ is a rational function is an immediate consequence of the following points:

- the numbers of graphs H and of possible values for the differences $\delta_e(\mathbf{i}, \mathbf{s})$ (for $e \in E_H$) are finite;

- the cumulant $\kappa\left(B_{i_1^1,s_1^1}^{(N)} \dots B_{i_p^1,s_p^1}^{(N)}, \dots, B_{i_1^\ell,s_1^\ell}^{(N)} \dots B_{i_p^\ell,s_p^\ell}^{(N)} \right)$ is a rational function in N which depends only on the graph H and values of $\delta_e(\mathbf{i}, \mathbf{s})$ (for $e \in E_H$);

- 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. $\qquad\square$

6.4 Conclusion: local statistics

Recently, several authors have further generalized the notion of dashed patterns into the notion of *bivincular* patterns [10, Section 2]. The idea is roughly that, in an occurrence of a bivincular pattern, one can ask that some values are consecutive (and not only some places as in dashed patterns). This new notion is very natural as occurrences of bivincular patterns in the inverse of a permutation correspond to occurrences of bivincular patterns in the permutation itself (which is not true for dashed patterns).

It would be interesting to give a general theorem on the asymptotic behavior of the number of occurrences of a given bivincular 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 bivincular patterns and converge towards a Poisson distribution.

- The dashed patterns are special cases of bivincular patterns. As we have seen in the previous section, their number of occurrences converges, after normalization, towards a Gaussian law (at least for patterns of size smaller than 9, the general case relies on Conjecture 1.9). Other bivincular patterns exhibit the same behavior, for example the one considered in [10].

- 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 [10]), has an expectation of order n, but a probability of being 0 with a positive lower bound.

Unfortunately, we have not been able to give a general statement. Let us however emphasize the fact that our approach unifies the first two cases.

More generally, our approach seems suitable to study what could be called a *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 $i_1, \ldots, i_p$ and $s_1, \ldots, s_p$ satisfying the constraints in S and such that $\sigma(i_j) = s_j$ for all j in $[p]$. For instance, the number of d-descents studied in [8] is a local statistic.

We call any linear combination of statistics $O_{p,S}^{(N)}$ a local statistic. The number of occurrences of a bivincular patterns, but also the number of exceedances 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.

References

[1] R. Arratia, A. Barbour, and S. Tavare. *Logarithmic combinatorial structures: a probabilistic approach*. EMS Monographs in Mathematics. EMS, Zürich, 2003. MR-2032426

[2] J. Aval, A. Boussicault, and P. Nadeau. Tree-like tableaux. In *Proceedings of FPSAC*, volume AO, pages 63–74. DMTCS, 2011. MR-2820698

[3] E. Babson and E. Steingrímsson. Generalized permutation patterns and a classification of the Mahonian statistics. *Sém. Lothar. Combin.*, 44:117–134, 2000. MR-1758852

[4] A. Barbour, L. Holst, and S. Janson. Poisson approximation. volume 2 of *Oxford studies in probability*. Oxford University Press, 1992. MR-1163825

[5] A. Barbour and S. Janson. A functional combinatorial central limit theorem. *Elec. J. Prob.,* 14:2352–2370, 2009. MR-2556014

[6] P. Billingsley. *Probability and measure*. Wiley, 1995. 3rd edition. MR-1324786

[7] P. Billingsley. *Convergence of probability measures*. Wiley, New York, 1999. 2nd edition. MR-1700749

[8] M. Bóna. Generalized descents and normality. *Elec. J. Comb.,* 15(1):N21, 7 pages, 2008. MR-2426166

[9] M. Bóna. On three different notions of monotone subsequences. In *Permutation Patterns,* volume 376 of *London Math. Soc. Lecture Note Series*, pages 89–113. Cambridge University Press, 2010. MR-2732825

[10] 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. MR-2652101

[11] 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. MR-2855213

[12] S. Corteel and P. Nadeau. Bijections for permutation tableaux. *Eur. J.Comb.,* 30(1):295–310, 2009. MR-2460235

[13] S. Corteel and L. Williams. Tableaux combinatorics for the asymmetric exclusion process. *Adv. Appl. Math.,* 39(3):293–310, 2007. MR-2352041

[14] 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. MR-2275686

[15] B. Derrida, J. Lebowitz, and E. Speer. Entropy of open lattice systems. *J. Stat. Phys.,* 126(4):1083–1108, 2007. MR-2311899

[16] W. Ewens. The sampling theory of selectively neutral alleles. *Theoretical Population Biology*, 3:87–112, 1972. MR-0325177

[17] V. Goncharov. Some facts from combinatorics. *Izvestia Akad. Nauk. SSSR, Ser. Mat*, 8:3–48, 1944.

[18] P. Hitczenko and S. Janson. Asymptotic normality of statistics on permutation tableaux. *Contemporary Math*, 520:83–104, 2010. MR-2681856

[19] W. Hoeffding. A combinatorial central limit theorem. *The Annals of Mathematical Statistics*, 22(4):558–566, 1951. MR-0044058

[20] S. Janson, T. Łuczak, and A. Ruciński. *Random graphs*, volume 45 of *Wiley Series in Discrete Mathematics and Optimization*. Wiley-Interscience, 2000. MR-1782847

[21] I. Kaplansky. The asymptotic distribution of runs of consecutive elements. *Ann. Math. Stat.,* 16(2):200–203, 1945. MR-0013847

[22] V. Kolchin. A problem of the allocation of particles in cells and cycles of random permutations. *Theory of Probability and its Applications*, 16:74–90, 1971.

[23] V. Leonov and A. Shiryaev. On a method of calculation of semi-invariants. *Theory Prob. Appl.,* 4:319–329, 1959. MR-0123345

[24] M. Lothaire. *Combinatorics on words*, volume 17 of *Encyclopedia of Math. and its Appl.* Addison-Wesley, 1983. MR-0675953

[25] J. Marckert. One more approach to the convergence of the empirical process to the Brownian bridge. *Electronic Journal of Statistics*, 2:118–126, 2008. MR-2386089

[26] A. Postnikov. Total positivity, Grassmannians, and networks. arXiv preprint arXiv:math/0609764v1, 2006.

[27] G. Rota and B. Taylor. The classical umbral calculus. *SIAM Journal on Mathematical Analysis*, 25:694, 1994. MR-1266584

[28] P. Śniady. Gaussian fluctuations of characters of symmetric groups and of Young diagrams. *Probab. Theory Related Fields*, 136(2):263–297, 2006. MR-2240789

[29] E. Steingrímsson and L. Williams. Permutation tableaux and permutation patterns. *J. Combin. Theory Ser. A*, 114(2):211–234, 2007. MR-2293088

[30] G. Watterson. The sampling theory of selectively neutral alleles. *Adv. Appl. Prob.*, 6:463–488, 1974. MR-0351504

[31] L. K. Williams. Enumeration of totally positive Grassmann cells. *Adv. Math.*, 190(2):319–342, 2005. MR-2102660

[32] J. Wolfowitz. Note on runs of consecutive elements. *Ann. Math. Stat.*, 15(1):97–98, 1944. MR-0010341

[33] W. Xu, B. Alain, and D. Sankoff. Poisson adjacency distributions in genome comparison: multichromosomal, circular, signed and unsigned cases. *Bioinformatics*, 24(16):i146, 2008.