Asymptotic behavior of some statistics in Ewens random permutations[†]

Valentin Féray[‡]

LaBRI, CNRS, Université de Bordeaux, 351 cours de la Libération, 33 400 Talence, France

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 considered here is Ewens sampling model, which generalizes uniform random permutations. We describe the asymptotic behavior of a large family of statistics, including the number of occurrences of any given 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.

Keywords: random permutations, *SSEP*, cumulants, dashed patterns

1 Introduction

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, number of descents, inversions, excedances or, more recently, 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 positive integer N, a probability measure μ_N of permutations of size N. Then any statistics above can be interpreted as a sequence of random variables $(X_N)_{N\geq 1}$. The natural question is now: what is the asymptotic behavior (possibly after normalization) of $(X_N)_{N\geq 1}$?

The simplest model of random permutations is of course the uniform random permutations (for each N, μ_N is the uniform distribution on the symmetric group S_N). A generalization of this model has been introduced by W.J. Ewens in the context of population dynamics [6]. It is defined by

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

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

[†]This paper is an extended abstract of paper [7], which will be submitted elsewhere.

[‡]partially supported by ANR grant PSYCO ANR-11-JS02-001.

a small abuse of language and use the expression *Ewens random permutation* for a random permutation distributed with Ewens measure.

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

The purpose of this article is to introduce a new general approach to this family of problems, based on the moment method. Our motivation is the following intuition: events which involve the images of different integers by the permutation σ are, informally speaking, *almost* independent. We will give a precise statement of this *almost* independence, that is an upper bound on joint cumulants (Lemma 2.2). This bound allows to study the asymptotic behavior of some statistics in Ewens random permutation (even some statistics whose bivariate generating series is not known). In this extended abstract, we have chosen to illustrate this method on two simple problems.

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 [11] and V.F. Kolchin [13] in the case of uniform measure and by G.A. Watterson [16, Theorem 5] in the framework of a general Ewens distribution (see also [1, Theorem 5.1]).

Theorem 1.1 ([16]) *Let p be a positive integer. When N tends to infinity, $\Gamma_p^{(N)}$ converges in distribution towards a Poisson law of parameter $(1+\theta)/p$. Moreover, the sequences of random variables $(\Gamma_{p'}^{(N)})_{N\geq 1}$ for $p' \leq p$ are asymptotically independent.*

In section 4, we illustrate our method by giving a new proof of this result.

Excedances. A (weak) excedance of a permutation σ in S_N is an integer i such that $\sigma(i) \geq i$. Let $B_i^{\text{ex},N}$ be the random variable equal to 1 if i is an excedance and 0 otherwise.

Let σ a permutation of size N. We define the function $F_\sigma^{(N)}$ as follows: if Nx is an integer, then

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

and the function $F_\sigma^{(N)}$ is an affine function between i/N and $(i+1)/n$ for each integer $i \leq N-1$.

Our interest in this quantity comes from the bijection between permutations and permutation tableaux (see, *e.g.*, [4]). The function $F_\sigma^{(N)}$ describes the shape of the permutation tableau associated to σ. It is also meaningful in a related model of statistical physics called *symmetric simple exclusion process* (SSEP). For more details, see the full version of the paper [7, paragraphs 5.1 and 5.2].

Theorem 1.2 ([5]) *Let x be a real number between 0 and 1. Then, $F_\sigma^{(N)}(x)$ tends almost surely towards $(1-(1-x)^2)/2$. 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 $0 \leq x_1, \dots, x_r \leq 1$, the vector $(Z_\sigma^{(N)}(x_1), \dots, Z_\sigma^{(N)}(x_r))$ converges towards a Gaussian vector $(G(x_1), \dots, G(x_r))$ of covariance matrix $(K(x_i,x_j))_{1\leq i,j\leq r}$, where K is an explicit function (given in paragraph 5.2).

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

Other statistics. The same method can be applied to other statistics. Unfortunately, we did not manage to obtain a general theorem, including all our examples. Nevertheless, we have quite general results, as the description of the limiting law of the renormalized number of occurrences of a given dashed pattern: see section 6 for the definition of this notion and the statement of this result (the proof is omitted in this extended abstract).

Comparison with other methods. Of course, the use of joint cumulants to determine limit laws of renormalized statistics is not original, see for example [12, Example 6.19] for an application in graph theory.

Nevertheless, in this example, the considered statistics are sums of Bernoulli variables, most of them being independent from each other (this information is encoded in the so-called dependency graph). In our work, *none* of the Bernoulli variables involved are independent, we are only able to give sharp bounds for their joint cumulants (Lemma 2.2).

These bounds on cumulants and the use of joint cumulants for *pairwise dependent* variables (*i.e.* whose dependency graph is the complete graph) are, to our knowledge, original.

Outline of the paper. Section 2 is devoted to our main lemma, a bound on joint cumulants of elementary events. Before proving the stated theorems, we give a graph-theoretical lemma in section 3. Then sections 4 and 5 consist respectively of the proofs of Theorems 1.1 and 1.2. Finally section 6 deals with dashed patterns.

2 Moments and cumulants of elementary events

2.1 Cumulants: definition and basic properties

Joint cumulants of random variables form a classical tool to quantify correlation. They generalize the notion of covariance. Most of this material can be found in Leonov's and Shiryaev's paper [14].

Definition of cumulants

Consider a sequence $X_1, \ldots, X_\ell$ of random variables on the same probability space (denote $\mathbb{E}$ the expectation on this space) indexed by the set $[\ell]$ (standard abbreviation for $\{1, 2, \ldots, \ell\}$). Let us define

$$\kappa(X_1, \ldots, X_\ell) = [t_1 \ldots t_\ell] \ln \Big(\mathbb{E}\big(\exp(t_1 X_1 + \cdots + t_\ell X_\ell) \big) \Big). \tag{2}$$

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 $t_1, \ldots, t_\ell$. Note that joint cumulants are multilinear functions. In the case where all the X_i are equal, we recover the ℓ-th cumulant $\kappa_\ell(X)$ of a single variable [8].

For small values of ℓ, the joint cumulants of one or two variables are respectively the mean of a single random variable (*i.e.*, $\kappa(X_1) = \mathbb{E}(X_1)$) and the covariance of a couple of random variables (*i.e.*, $\kappa(X_1, X_2) = \mathbb{E}(X_1 X_2) - \mathbb{E}(X_1)\mathbb{E}(X_2)$). For three variables, one has

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

In general, moments and cumulants are related by:

$$\kappa(X_1,\dots,X_\ell)=\mathbb{E}\big(X_1\cdots X_\ell\big)+\text{reminder}, \tag{3}$$

where the reminder is a linear combination of products of smaller order joint moments.

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

$$\kappa(X_1,\dots,X_\ell)=[t_1\dots t_\ell]\ln\left(\mathbb{E}\big(\exp(\sum_{i\in A}t_iX_i)\big)\right)+[t_1\dots t_\ell]\ln\left(\mathbb{E}\big(\exp(\sum_{i\in B}t_iX_i)\big)\right)=0.$$

Because of this strong property, joint cumulants can be seen as a quantification of the dependence of random variables.

Convergence in distribution using cumulants.
Consider now m sequences of random variables: $(X_n^{(i)})_{n\ge1}$ for $i\in[m]$. A consequence of Equation (3) is that the convergence of all joint cumulants

$$\kappa\left(X_n^{(i_1)},\dots,X_n^{(i_\ell)}\right); \ell\ge1, 1\le i_1,\dots,i_\ell\le m$$

is equivalent to the convergence of all joint moments

$$\mathbb{E}\left(X_n^{(i_1)}\cdots X_n^{(i_\ell)}\right); \ell\ge1, 1\le i_1,\dots,i_\ell\le m.$$

In particular, if $Y^{(1)},\dots,Y^{(m)}$ are random variables such that the law of the m-tuple $(Y^{(1)},\dots,Y^{(m)})$ is entirely determined by its joint moments, then the two following statements are equivalent (see [3, Theorem 30.2] for the same property in terms of moments).

- For any ℓ and any list $i_1,\dots,i_\ell$ in $[m]$,
$$\lim_{n\to\infty}\kappa\left(X_n^{(i_1)},\dots,X_n^{(i_\ell)}\right)=\kappa\left(Y^{(i_1)},\dots,Y^{(i_\ell)}\right).$$
- The sequence of m-tuples $(X_n^{(1)},\dots,X_n^{(m)})$ converges in distribution towards $(Y^{(1)},\dots,Y^{(m)})$.

Recall that Gaussian and Poisson variables are determined by their moments, see *e.g.* the criterion [3, Theorem 30.1]. Hence, cumulants can be used to prove convergence in distribution towards Gaussian or Poisson variables, such as the results of the previous paragraph.

2.2 Joint moments of elementary events

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

If i and s are two integers in $[N]$, we consider the Bernoulli variable $B_{i,s}^{(N)}$ which takes value 1 if and only if $\sigma(i)=s$. 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*).

In this paragraph, we compute the joint moments of the family of random variables $(B_{i,s}^{(N)})_{1\le i,s\le N}$. Indeed, our method is based on the fact that these joint moments have an explicit closed form.

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

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

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

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

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

where $(x)_r$ is a short notation for the falling factorial $x(x-1)\dots(x-r+1)$.

Proof: Omitted for brevity. It relies on two important properties of the Ewens measure. First, it is conjugacy-invariant. Second, a random sampling can easily be obtained by an inductive procedure, called *Chinese restaurant process* (see, *e.g.* [1, Example 2.19] for details). □

2.3 The main lemma

Our main tool is a bound on joint cumulants of products of elementary events $(B_{i,s}^{(N)})$. To state it, we introduce the following notations. Consider two lists $\mathbf{i} = (i_1,\dots,i_r)$ and $\mathbf{s} = (s_1,\dots,s_r)$ of positive integers of the same length r, as well as a set partition τ of $[r]$ with ℓ parts. Then we define the graphs $G_1(\mathbf{i},\mathbf{s})$ and $G_2^\tau(\mathbf{i},\mathbf{s})$ as follows.

- The vertex set of $G_1(\mathbf{i},\mathbf{s})$ is $[r]$ and j and j' are linked in $G_1(\mathbf{i},\mathbf{s})$ iff $i_j = i_{j'}$ *and* $s_j = s_{j'}$.
- For $G_2^\tau(\mathbf{i},\mathbf{s})$, first consider the graph $G_2(\mathbf{i},\mathbf{s})$ whose vertex set is $[r]$ and j and j' are linked in $G_2(\mathbf{i},\mathbf{s})$ iff $i_j = i_{j'}$ *or* $s_j = s_{j'}$. Then $G_2^\tau(\mathbf{i},\mathbf{s})$ is obtained from $G_2(\mathbf{i},\mathbf{s})$ by merging vertices corresponding to elements in the same part of τ.

The connected components of a graph G form a set partition of its vertex set that we denote $\mathrm{CC}(G)$. We use also the symbol $\#(\mathrm{CC}(G))$ for the number of connected components of G.

Theorem 2.2 *Fix a positive integer r. There exists a constant C_r, depending on r, such that for any set partition $\tau = (\tau_1,\dots,\tau_\ell)$ of $[r]$, any $N \ge 1$ and lists $\mathbf{i} = (i_1,\dots,i_r)$ and $\mathbf{s} = (s_1,\dots,s_r)$ of integers in $[N]$, one has:*

$$\left|\kappa\left(\prod_{j\in\tau_1} B_{i_j,s_j}^{(N)},\dots,\prod_{j\in\tau_\ell} B_{i_j,s_j}^{(N)}\right)\right| \le C_r N^{-\#\left(\mathrm{CC}(G_1(\mathbf{i},\mathbf{s}))\right)-\#\left(\mathrm{CC}(G_2^\tau(\mathbf{i},\mathbf{s}))\right)+1}. \tag{4}$$

Note that the integer $\#\big(\mathrm{CC}(G_1(\mathbf{i},\mathbf{s}))\big)$ is the number of *different* couples (i_j, s_j). The second quantity does not admit such a simple description.

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

$$\sum_{\pi \in \mathcal{P}([\ell])} (-1)^{|\pi|-1}(|\pi|-1)! \prod_{C \in \pi} \frac{1}{(N+\theta)_{\sum_{j\in C} a_j}}.$$

Note that each summand has degree $-\sum_{j=1}^{r} a_j = -\ell$. According to our theorem, the sum has degree at most $-\ell - r + 1$, which means that many simplifications are happening. This reflects the fact that the variables $B_{i_j,s_j}^{(N)}$ are very weakly correlated.

Remark 2.3 *In this extended abstract, we do not give the proof of this theorem, which is rather long and technical (even in the case presented above, where the cumulant is a totally explicit rational function). The proof can of course be found in the full version [7, Section 2].*

3 Graph-theoretical lemmas

In this section, we present an easy lemma on the number of connected components on contraction of graphs. This lemma will be useful in the asymptotic study of the number of excedances. It also appears when considering other statistics (not presented in this extended abstract), that is why we state it separately.

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

Let us further assume that $V = W \sqcup W$ and 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 notations, we will use overlined letters for elements of the second copy of W.

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

Example. Consider the graph G of Figure 1. Its vertex set is the set $V = \{1,2,3,4,5,\bar{1},\bar{2},\bar{3},\bar{4},\bar{5}\}$. Consider the application f from V to $W = \{1,2,3,4,5\}$, consisting in forgetting the bar. The contraction G/f and the surcontration $G//f$ are drawn on Figure 1.

Lemma 3.1 *Let G and f be as above. Then*

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

Proof: Omitted for brevity. It relies on easy considerations on maximal spanning forests of G. □

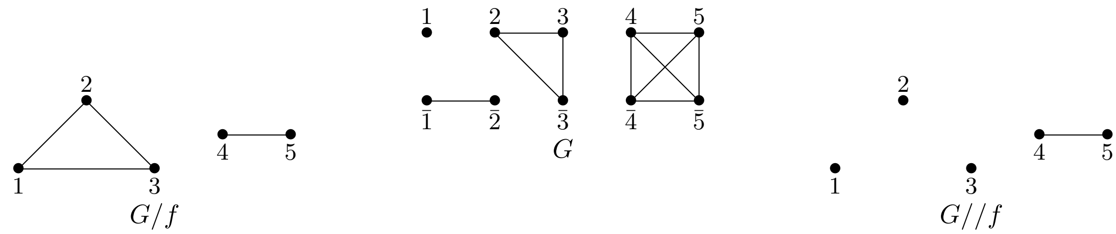

Fig. 1: An example of a graph, its contraction and surcontraction.

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 Lemma 2.2 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.

Expanding the cumulants of the considered statistic.

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

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

where $B^{\mathrm{c},N}_{(i_1,\dots,i_p)} = B^{(N)}_{i_1,i_2} \dots B^{(N)}_{i_{p-1},i_p} B^{(N)}_{i_p,i_1}$ is 1 if $(i_1, \dots, i_p)$ is a cycle of σ and 0 otherwise. By multilinearity of cumulants, we get:

$$\kappa_\ell(\Gamma_p^{(N)}) = \sum_{\substack{i_1^1 < i_2^1, i_3^1, \dots, i_p^1 \\ \vdots \\ i_1^\ell < i_2^\ell, i_3^\ell, \dots, i_p^\ell}} \kappa\big(B^{(N)}_{i_1^1,i_2^1} \dots B^{(N)}_{i_p^1,i_1^1}, \cdots, B^{(N)}_{i_1^\ell,i_2^\ell} \dots B^{(N)}_{i_p^\ell,i_1^\ell}\big). \tag{5}$$

An upper bound for the elementary cumulants.

Now, we would like to apply our main lemma to every summand of equation (5). For this, one has to understand what is the exponent of N in the upper bound given by Lemma 2.2.

For a matrix $\mathbf{i} = (i_j^r)_{\substack{1\leq j\leq p \\ 1 \leq r \leq \ell}}$, we denote:

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

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

- $t(\mathbf{i})$ 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, Lemma 2.2 writes as:

$$\left|\kappa\big(B^{(N)}_{i^1_1,i^1_2}\dots B^{(N)}_{i^1_p,i^1_1},\cdots,B^{(N)}_{i^\ell_1,i^\ell_2}\dots B^{(N)}_{i^\ell_p,i^\ell_1}\big)\right| \le C_{p\ell}N^{-M(\mathbf{i})-Q(\mathbf{i})+1} \le C_{p\ell}N^{-M(\mathbf{i})} \le C_{p\ell}N^{-t(\mathbf{i})}. \quad (6)$$

An upper bound for the number of lists.

As the number of summands in Equation (5) depends on N, we need a bound on this number to use inequality (6).

This bound comes from the following simple lemma:

Lemma 4.1 *For each $\ell \ge 1$, there exists a constant C'_ℓ with the following property. For any $N \ge 1$ and $t \in [\ell]$, the number of lists $\mathbf{i}$ of length ℓ with entries in $[N]$ such that $|\{i_1,\dots,i_\ell\}| = t$ is bounded by $C'_\ell N^t$.*

Proof: If we specify which indices correspond to entries with the same values (that is a set partition of the set of indices), the number of corresponding lists is $\binom{N}{t}$ and hence is bounded by N^t. But the number C'_ℓ of possible partitions of the set $[\ell]$ depends only on ℓ, whence the result. □

End of proof of the theorem.

By inequality (6) and Lemma 4.1, the contribution of lists (i^r_j) taking exactly t different values is bounded by $C'_{p\ell}C_{p\ell}$ and hence, for all $\ell \ge 1$, one has $\kappa_\ell(\Gamma^{(N)}_p) = O(1)$.

To compute the component of order 1, let us make the following remark: by the argument above, the total contribution of lists (i^r_j) 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

$$\{i^{r_1}_j; 1 \le j \le p\} \cap \{i^{r_2}_j; 1 \le j \le p\} \neq \emptyset,$$

the cyclic words $(i^{r_1}_1,\dots,i^{r_1}_p)$ and $(i^{r_2}_1,\dots,i^{r_2}_p)$ are equals. As i^r_1 is always the minimum of the i^r_j, the two words are in fact always equal in this case. In particular $G(\mathbf{i})$ is a disjoint union of cliques. If we further assume $Q(\mathbf{i}) = 1$, *i.e.* $G(\mathbf{i})$ is connected, $G(\mathbf{i})$ is the complete graph and we get that i^r_j does not depend on r. Therefore,

$$\kappa_\ell(\Gamma^{(N)}_p) = \sum_{i_1<i_2,i_3,\dots,i_p} \kappa_\ell\big(B^{(N)}_{i_1,i_2}\dots B^{(N)}_{i_p,i_1}\big) + O(N^{-1}). \quad (7)$$

But each $B^{(N)}_{i_1,i_2}\dots B^{(N)}_{i_p,i_1}$ is a Bernoulli variable of parameter $(1+\theta)/(N+\theta)_p$. Therefore its moments are all equal to $(1+\theta)/(N+\theta)_p$ and by formula (3), its cumulants are $(1+\theta)/(N+\theta)_p + O(N^{-2p})$. Finally, as there are $(N)_p/p$ terms in equation (7),

$$\kappa_\ell(\Gamma^{(N)}_p) = \frac{1+\theta}{p} + O(N^{-1}).$$

This is valid for any $\ell \ge 1$, so $\Gamma^{(N)}_p$ converges in distribution towards a Poisson law of parameter $\frac{1+\theta}{p}$.

Moreover, a simple adaptation of the proof of Equation (7) implies that

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

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

Remark. After equation (7), one could have finished the proof without computation by the following argument: $\Gamma_p^{(N)}$ has asymptotically the same cumulants as a virtual variable X_N, which writes as a sum of *independent* random variables with the same distribution as the $B^{c,N}_{(i_1,\dots,i_p)}$. As each $B^{c,N}_{(i_1,\dots,i_p)}$ is a Bernoulli variable of expectation $(1+\theta)/(N+\theta)_p$ and as there are $(N)_p/p$ such variables, X_N tends in distribution towards a Poisson law of parameter $(1+\theta)/p$. And so does $\Gamma_p^{(N)}$.

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

5 Number of excedances

In this section, we prove Theorem 1.2. The proof roughly follows the same guideline as the one of Theorem 1.1 in the previous paragraph.

5.1 Bounds for cumulants

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

Let us begin by the following bound:

Proposition 5.1 *For each $\ell \geq 1$, there exists a constant C_ℓ with the following property. For any $N \geq 1$ and lists $i_1, \dots, i_\ell$ and $s_1, \dots, s_\ell$ of integers in $[N]$,*

$$\kappa(B^{(N)}_{i_1,s_1}, \dots, B^{(N)}_{i_\ell,s_\ell}) \leq C_\ell N^{-|\{i_1,\dots,i_\ell,s_1,\dots,s_\ell\}|+1}. \tag{8}$$

Proof: Using Lemma 2.2 for $\tau = \big\{\{1\}, \dots, \{\ell\}\big\}$, we only have to prove that

$$-\#\big(\operatorname{CC}(G_1(\mathbf{i},\mathbf{s}))\big) - \#\big(\operatorname{CC}(G_2^\tau(\mathbf{i},\mathbf{s}))\big) \geq -|\{i_1, \dots, i_\ell, s_1, \dots, s_\ell\}|. \tag{9}$$

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

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

Then, inequality (9) is simply Lemma 3.1 applied to the graph $G_3(\mathbf{i},\mathbf{s})$ ($G_1(\mathbf{i},\mathbf{s})$ and $G_2^\tau(\mathbf{i},\mathbf{s})$ are respectively its contraction and surcontraction). □

We can now prove the following bound:

Proposition 5.2 *There exists a constant C''_ℓ such that, for any integer $N \geq 1$ and real numbers $x_1, \dots, x_\ell$, one has*

$$|\kappa(F_\sigma^{(N)}(x_1), \dots, F_\sigma^{(N)}(x_\ell))| \leq C''_\ell N^{-\ell+1}. \tag{10}$$

Proof: To simplify the notations, we suppose that $Nx_1, \dots, Nx_\ell$ 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 $B_i^{\mathrm{ex},N}$ can be written as $B_i^{\mathrm{ex},N} = \sum_{s \geq i} B_{i,s}^{(N)}$. Finally, by multilinearity,

$$(N-1)^\ell \kappa(F_\sigma^{(N)}(x_1), \dots, F_\sigma^{(N)}(x_\ell)) = \sum_{\substack{2 \leq i_1 \leq Nx_1 \\ \vdots \\ 2 \leq i_\ell \leq Nx_\ell}} \sum_{\substack{s_1 \geq i_1 \\ \vdots \\ s_\ell \geq i_\ell}} \kappa(B_{i_1,s_1}^{(N)}, \dots, B_{i_\ell,s_\ell}^{(N)}). \tag{11}$$

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

Combining this with Proposition 5.1, we get that the total contribution of couples of lists $(\mathbf{i}, \mathbf{s})$ with $|\{i_1, \dots, i_\ell, s_1, \dots, s_\ell\}| = t$ to the right-hand side of (11) is smaller than $C'_{2\ell} C_\ell N$, which ends the proof. □

Illustration of the proof. Set $\ell = 5$ and consider the lists $\mathbf{i} = (5, 2, 2, 7, 7)$ and $\mathbf{s} = (8, 8, 2, 7, 7)$. The graph $G_3(\mathbf{i}, \mathbf{s})$ associated to this couple of sequences is the graph G drawn of Figure 1. It follows immediately that $G_1(\mathbf{i}, \mathbf{s}) = G//f$ has 4 connected components while $G_2^\tau(\mathbf{i}, \mathbf{s}) = G/f$ has 2. Therefore, by Lemma 2.2,

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

The same bound is valid for all sequences $\mathbf{i}$ and $\mathbf{s}$ such that $G_3(\mathbf{i}, \mathbf{s}) = G$. There are fewer than N^4 such sequences: to construct such a sequence, one has to choose distinct values for the four connected components of G, such that they fulfill some inequalities. Finally, their total contribution to (11) is smaller than $C_5 N^{-1}$.

Remark. This result is closely related to a result of B. Derrida, J.L. Lebowitz and E.R. Speer [5, Appendix A], as explained in the full version [7, paragraph 5.3].

5.2 Convergence results

In this paragraph, we explain how one can deduce from the bound on cumulants, some results on the convergence of $F_\sigma^{(N)}$.

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

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

It is easily deduced that:

$$\lim_{N\to\infty} \mathbb{E}(F_\sigma^{(N)}(x)) = \int_0^x (1-t)dt + o(1) = \frac{1-(1-x)^2}{2}; \tag{12}$$

$$\lim_{N\to\infty} N \operatorname{Cov}(F_\sigma^{(N)}(x), F_\sigma^{(N)}(y)) = \int_0^{\min(x,y)} t(1-t)dt - \int_0^x \int_0^y \min(t,u)(1-\max(t,u))dtdu. \tag{13}$$

We call $K(x,y)$ the right-hand side of the second equation.

Proof of Theorem 1.2: Consider the first statement. The convergence in probability follows immediately from equations (12) and (13). The almost-sure convergence can be obtained by studying the fourth centered moment: this is a classical trick not detailed here.

Let us consider the second statement. Proposition 5.2 implies that, for any list $j_1,\dots,j_\ell$ of integers in $[r]$, one has

$$\kappa(Z_\sigma^{(N)}(x_{j_1}), \dots, Z_\sigma^{(N)}(x_{j_\ell})) = O(N^{-r/2+1}). \tag{14}$$

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 Gaussian vector. The covariance matrix is the limit of the covariance of the $Z_\sigma^{(N)}(x_i)$, that is $(K(x_i,x_j))$. □

The theorem gives the first and second order asymptotics for the values of $F_\sigma^{(N)}$ and $Z_\sigma^{(N)}$ at some fixed point(s). It is also possible with the same method to prove some stronger more technical result: the sequence of random functions $(Z_\sigma^{(N)})_{N\geq 1}$ converges in distribution towards a Gaussian process in the space of continuous real-valued functions on [0; 1] endowed with uniform norm (see [7, paragraph 5.4]).

6 Dashed patterns

We now present without proof a general result for the number of occurences of dashed patterns in large Ewens random permutations. This notion has been introduced by E. Babson and E. Steingrimsson [2].

Definition 6.1 *A dashed pattern of size p is the data of a permutation $\tau \in S_p$ and a subset X of $[p-1]$. An occurrence of the dashed pattern (τ, X) in a permutation $\sigma \in S_N$ is a list $i_1 < \dots < i_p$ such that:*

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

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

Example 6.2 *$O_{21,\emptyset}^{(N)}$ is the number of inversion, while $O_{21,\{1\}}^{(N)}$ is the number of descents. Many classical statistics on permutations can be written as the number of occurrences of a given dashed patten or as a linear combination of such statistics, see [2, section 2].*

With the same method as above, it is proven in the full version of the paper [7, paragraph 6.3] the following general result for the asymptotic behavior of number of occurences of dashed patterns.

Theorem 6.3 *Let (τ, X) be a dashed pattern of size p (see definition 6.1) and σ_N a sequence of random permutations, each σ_N being of size N distributed with Ewens measure. We denote $q = |X|$. Then, $\frac{O^{(N)}_{\tau,X}(\sigma_N)}{N^{p-q}}$, 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:*

$$\sqrt{N}\left(\frac{O^{(N)}_{\tau,X}(\sigma_N)}{N^{p-q}} - \frac{1}{p!(p-q)!}\right) \to \mathcal{N}(0, V_{\tau,X}),$$

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

The asymptotic number of occurrences of some patterns (or *motifs*) have already been studied in the framework of random words. In particular, results similar to Theorem 6.3 have been obtained by P. Flajolet and co-authors [15, 10]. Note that the methods used by these authors can *a priori* not be extended to the framework of permutation patterns for the following reasons. First, it is not known how to compute the generating functions of the number of occurrences of a given pattern in permutations. Second, except in the case $\theta = 0$ (uniform random permutations), the occurrences of patterns in disjoint places are not independent events.

References

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

[2] E. Babson and E. Steingrımsson. Generalized permutation patterns and a classification of the Mahonian statistics. *Sém. Lothar. Combin.*, 44:117–134, 2000.

[3] P. Billingsley. *Probability and measure*. Wiley, 1995. 3rd edition.

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

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

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

[7] V. Féray. Asymptotic behavior of some statistics in Ewens random permutations. arXiv:1201.2157v1, 2011.

[8] R. Fisher. Moments and product moments of sampling distributions. *Proc. Lond. Math. Soc. Series 2*, 30:199–238, 1929.

[9] P. Flajolet and R. Sedgewick. *Analytic Combinatorics*. Cambridge University Press, 2009.

[10] P. Flajolet, W. Szpankowski, and B. Vallée. Hidden word statistics. *J. ACM*, 53(1):147–183, 2006.

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

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

[13] 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.

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

[15] P. Nicodème, B. Salvy, and P. Flajolet. Motif statistics. In *European Symposium on Algorithm*, volume 1643 of *Lecture Notes in Computer Science*, pages 194–211. Springer, 1999.

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