Asymptotic behaviour of some statistics in Ewens random permutations

Valentin Féray

LaBRI, CNRS, Bordeaux

Conference on *Analysis of algorithms*
Montréal, June 19th, 2012

What is this talk about?

- S_n: set of permutations of n

  ex: $4253716 \in S_7$.

What is this talk about?

- S_n: set of permutations of n

  ex: $4253716 \in S_7$.

- Choose your favorite statistics on permutations $X : \bigsqcup_{n \geq 1} S_n \to \mathbb{R}$:

  ex: number of fixed points, number of cycles, number of occurrences of a given (generalized) pattern, ...

What is this talk about?

- S_n: set of permutations of n

  ex: $4253716 \in S_7$.

- Choose your favorite statistics on permutations $X : \bigsqcup_{n \geq 1} S_n \to \mathbb{R}$:

  ex: number of fixed points, number of cycles, number of occurrences of a given (generalized) pattern, ...

- Consider the uniform measure of S_n

  $\to X$ can be seen as a sequence of random variables $X_n : S_n \to \mathbb{R}$.

What is this talk about?

- S_n: set of permutations of n

  ex: $4253716 \in S_7$.

- Choose your favorite statistics on permutations $X : \bigsqcup_{n \geq 1} S_n \to \mathbb{R}$:

  ex: number of fixed points, number of cycles, number of occurrences of a given (generalized) pattern, ...

- Consider the uniform measure of S_n

  $\to X$ can be seen as a sequence of random variables $X_n : S_n \to \mathbb{R}$.

- Problem: Asymptotic behaviour of X_n??

  i.e. does X_n (after suitable renormalization) converge in distribution?

What is this talk about?

- S_n: set of permutations of n

  ex: $4253716 \in S_7$.

- Choose your favorite statistics on permutations $X : \bigsqcup_{n \geq 1} S_n \to \mathbb{R}$:

  ex: number of fixed points, number of cycles, number of occurrences of a given (generalized) pattern, ...

- Consider the uniform measure of S_n

  $\to X$ can be seen as a sequence of random variables $X_n : S_n \to \mathbb{R}$.

- Problem: Asymptotic behaviour of X_n??

  i.e. does X_n (after suitable renormalization) converge in distribution?

- Goal of the talk: give a quite general method to answer this question.

Outline of the talk

1. Introduction
   - Intuition on an example
   - More general results
2. The method on an example

Example: number of fixed points

$$X(\sigma) = |\{i : \sigma(i) = i\}|$$

Theorem

$(X_n)_{n \geq 1}$ converges in distribution towards a Poisson law of parameter 1.

Proof. inclusion/exclusion or generating series.

Example: number of fixed points

$$X(\sigma) = |\{i : \sigma(i) = i\}|$$

Theorem

$(X_n)_{n \geq 1}$ converges in distribution towards a Poisson law of parameter 1.

Remark. $X_n = \sum_{i=1}^{n} F_i$, where F_i is a Bernoulli variable of parameter $1/n$, (F_i takes value 1 if $\sigma(i) = i$).

Example: number of fixed points

$$X(\sigma) = |\{i : \sigma(i) = i\}|$$

Theorem

$(X_n)_{n \geq 1}$ *converges in distribution towards a Poisson law of parameter* 1.

Remark. $X_n = \sum_{i=1}^{n} F_i$, where F_i is a Bernoulli variable of parameter $1/n$, (F_i takes value 1 if $\sigma(i) = i$).

Reminder: law of small numbers

The sum of n *independent* Bernoulli variables of parameter $1/n$ converges toward a Poisson law of parameter 1.

Example: number of fixed points

$$X(\sigma) = |\{i : \sigma(i) = i\}|$$

Theorem

$(X_n)_{n \geq 1}$ converges in distribution towards a Poisson law of parameter 1.

Remark. $X_n = \sum_{i=1}^{n} F_i$, where F_i is a Bernoulli variable of parameter $1/n$,
(F_i takes value 1 if $\sigma(i) = i$).

Reminder: law of small numbers

The sum of n *independent* Bernoulli variables of parameter $1/n$ converges toward a Poisson law of parameter 1.

But the F_i are not independent!

Example: number of fixed points

$$X(\sigma) = |\{i : \sigma(i) = i\}|$$

Theorem

$(X_n)_{n \geq 1}$ *converges in distribution towards a Poisson law of parameter* 1.

Remark. $X_n = \sum_{i=1}^{n} F_i$, where F_i is a Bernoulli variable of parameter $1/n$, (F_i takes value 1 if $\sigma(i) = i$).

Reminder: law of small numbers

The sum of n *independent* Bernoulli variables of parameter $1/n$ converges toward a Poisson law of parameter 1.

But the F_i are not independent! We will show that they are *almost independent* in some sense.

Second example: excedances

$$X(\sigma) = |\{i : \sigma(i) \geq i\}|$$

Second example: excedances

$$X(\sigma) = |\{i : \sigma(i) \geq i\}|$$

Theorem

After renormalization, $(X_n)_{n\geq 1}$ converges in distribution towards a Gaussian law.

Proof. Also by generating series, but we will give another method which works in more generality.

Second example: excedances

$$X(\sigma) = |\{i : \sigma(i) \geq i\}|$$

Theorem

After renormalization, $(X_n)_{n\geq 1}$ converges in distribution towards a Gaussian law.

Proof. Also by generating series, but we will give another method which works in more generality.

Remark. $X_n = \sum_{i \leq s} B_{i,s}$, where $B_{i,s}$ is a Bernoulli variable of parameter $1/n$, ($B_{i,s}$ takes value 1 if $\sigma(i) = s$).

It would be easier if the $B_{i,s}$ were independent. Intuitively, such variables with distinct indices are almost independent.

The method is in fact much more general! (1/2)

Theorem

Let X be the number of occurrences of a given dashed pattern (or a linear combination of those).

linear combination of occurrences dashed patterns include:

> *numbers of inversions, descents, double descents, peaks, increasing runs or subsequences of a given length,...*

The method is in fact much more general! (1/2)

Theorem

Let X be the number of occurrences of a given dashed pattern (or a linear combination of those).
We consider a permutation σ_n of size n distributed with Ewens measure.

Ewens measure: a one-parameter deformation of uniform distribution

$$P(\{\sigma\}) \propto \theta^{\#\text{cycles}(\sigma)}.$$

The method is in fact much more general! (1/2)

Theorem

Let X be the number of occurrences of a given dashed pattern (or a linear combination of those).
We consider a permutation σ_n of size n distributed with Ewens measure.

Remark. The first-order asymptotic is easy: in probability,

$$X(\sigma_n) \sim c_1 n^{c_2},$$

with some constants c_1 and c_2 depending on X.

The method is in fact much more general! (1/2)

Theorem

Let X be the number of occurrences of a given dashed pattern (or a linear combination of those).
We consider a permutation σ_n of size n distributed with Ewens measure.

Then the fluctuations of order $1/\sqrt{n}$ of $\frac{X(\sigma_n)}{n^{c_2}}$ are asymptotically Gaussian.

The method is in fact much more general! (2/2)

Fix $p \in [0;1]$.
Model of random graph G_n of size n:

- $V(G_n) = [n]$;
- $E(G_n)$ is chosen uniformly among all sets of pairs of size $k = \lfloor p\binom{n}{2} \rfloor$.

The method is in fact much more general! (2/2)

Fix $p \in [0;1]$.
Model of random graph G_n of size n:

- $V(G_n) = [n]$;
- $E(G_n)$ is chosen uniformly among all sets of pairs of size $k = \lfloor p\binom{n}{2} \rfloor$.

Theorem (Janson, 1994)

The fluctuations of the number of triangles in G_n are asymptotically Gaussian.

Covariance of the $B_{i,s}$

Back to excedances and uniform measure:Recall: $B_{i,s}$ is the characteristic

function of the event $\sigma(i) = s$.

Easy computation: if $i \neq j$ and $s \neq t$,

$$\begin{aligned}\mathrm{Cov}(B_{i,s}, B_{j,t}) &= \mathbb{E}(B_{i,s}B_{j,t}) - \mathbb{E}(B_{i,s})\mathbb{E}(B_{j,t}) \\ &= \frac{1}{n(n-1)} - \left(\frac{1}{n}\right)^2 = \frac{1}{n^2(n-1)}\end{aligned}$$

Covariance of the $B_{i,s}$

Back to excedances and uniform measure:Recall: $B_{i,s}$ is the characteristic

function of the event $\sigma(i) = s$.

Easy computation: if $i \neq j$ and $s \neq t$,

$$\begin{aligned} \mathrm{Cov}(B_{i,s}, B_{j,t}) &= \mathbb{E}(B_{i,s} B_{j,t}) - \mathbb{E}(B_{i,s})\mathbb{E}(B_{j,t}) \\ &= \frac{1}{n(n-1)} - \left(\frac{1}{n}\right)^2 = \frac{1}{n^2(n-1)} \end{aligned}$$

Remark. $\mathrm{Cov}(B_{i,s}, B_{j,t}) \ll \mathbb{E}(B_{i,s} B_{j,t}), \mathbb{E}(B_{i,s})\mathbb{E}(B_{j,t})$.
Confirms the intuition of *almost independence*.

Covariance of the $B_{i,s}$

Back to excedances and uniform measure:Recall: $B_{i,s}$ is the characteristic

function of the event $\sigma(i) = s$.

Easy computation: if $i \neq j$ and $s \neq t$,

$$\mathrm{Cov}(B_{i,s}, B_{j,t}) = \mathbb{E}(B_{i,s}B_{j,t}) - \mathbb{E}(B_{i,s})\mathbb{E}(B_{j,t})$$
$$= \frac{1}{n(n-1)} - \left(\frac{1}{n}\right)^2 = \frac{1}{n^2(n-1)}$$

Remark. $\mathrm{Cov}(B_{i,s}, B_{j,t}) \ll \mathbb{E}(B_{i,s}B_{j,t}), \mathbb{E}(B_{i,s})\mathbb{E}(B_{j,t})$.
Confirms the intuition of *almost independence*.

Not very convincing: some dependent variables have null covariance.
$\longrightarrow$ we will compute joint cumulants.

What are joint cumulants?

$$\kappa_1(X) = \mathbb{E}(X), \quad \kappa_2(X, Y) = \mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

$$\kappa_3(X, Y, Z) = \mathbb{E}(XYZ) - \mathbb{E}(XY)\mathbb{E}(Z) - \mathbb{E}(XZ)\mathbb{E}(Y) - \mathbb{E}(YZ)\mathbb{E}(X) + 2\mathbb{E}(X)\mathbb{E}(Y)\mathbb{E}(Z).$$

What are joint cumulants?

$$\kappa_1(X) = \mathbb{E}(X), \quad \kappa_2(X,Y) = \mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

$$\kappa_3(X,Y,Z) = \mathbb{E}(XYZ) - \mathbb{E}(XY)\mathbb{E}(Z) - \mathbb{E}(XZ)\mathbb{E}(Y) - \mathbb{E}(YZ)\mathbb{E}(X) + 2\mathbb{E}(X)\mathbb{E}(Y)\mathbb{E}(Z).$$

In general, $\kappa_\ell(X_1,\dots,X_\ell) = \mathbb{E}(X_1 \cdots X_\ell)$ + homogeneous sum of products of joint moments of smaller degree (explicit description in terms of set partitions).

What are joint cumulants?

$$\kappa_1(X) = \mathbb{E}(X), \quad \kappa_2(X,Y) = \mathrm{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$
$$\kappa_3(X,Y,Z) = \mathbb{E}(XYZ) - \mathbb{E}(XY)\mathbb{E}(Z) - \mathbb{E}(XZ)\mathbb{E}(Y) - \mathbb{E}(YZ)\mathbb{E}(X) + 2\mathbb{E}(X)\mathbb{E}(Y)\mathbb{E}(Z).$$

In general, $\kappa_\ell(X_1, \dots, X_\ell) = \mathbb{E}(X_1 \cdots X_\ell)$ + homogeneous sum of products of joint moments of smaller degree (explicit description in terms of set partitions).

Nice behaviour with respect to independence**:*

$A, B, C, \dots$ are *independent* $\Leftrightarrow$

all joint cumulants $\kappa_\ell(A, \dots, A, B, \dots, B, C, \dots, C, \dots)$ *vanish (as soon as they involve at least two different variables).*

* if A, B, C, … have joint moments of all orders and the joint law is determined by its joint moments (easy criterion on moments of marginal laws).

Cumulants of $B_{i,s}$

If h, i and j (resp. s,t and u) are pairwise distinct,

$$\kappa_3(B_{h,u}, B_{i,s}, B_{j,t}) = \frac{1}{n(n-1)(n-2)} - 3\frac{1}{n^2(n-1)} + 2\frac{1}{n^3}$$

Cumulants of $B_{i,s}$

If h, i and j (resp. s,t and u) are pairwise distinct,

$$\kappa_3(B_{h,u}, B_{i,s}, B_{j,t}) = \frac{1}{n(n-1)(n-2)} - 3\frac{1}{n^2(n-1)} + 2\frac{1}{n^3}$$
$$= \frac{4}{n^3(n-1)(n-2)} = O(n^{-5})$$

Cumulants of $B_{i,s}$

If h, i and j (resp. s,t and u) are pairwise distinct,

$$\begin{aligned}\kappa_3(B_{h,u}, B_{i,s}, B_{j,t}) &= \frac{1}{n(n-1)(n-2)} - 3\frac{1}{n^2(n-1)} + 2\frac{1}{n^3} \\ &= \frac{4}{n^3(n-1)(n-2)} = O(n^{-5})\end{aligned}$$

In general,

$$\kappa_\ell(B_{i_1,s_1}, \dots, B_{i_\ell,s_\ell}) = O(n^{-2\ell+1}),$$

for two lists **i** and **s** of pairwise distinct integers.
Remark. A priori, it is a rational function of degree $-\ell$. It is quite technical to prove that it has in fact degree $-2\ell+1$.

Cumulants and convergence in distribution

Our goal: show that $\sum_{i \leq s} B_{i,s}$ converges in distribution towards a Poisson law.

Cumulants and convergence in distribution

Our goal: show that $\sum_{i \leq s} B_{i,s}$ converges in distribution towards a Poisson law.

Cumulants are a good tool to prove convergence in distribution

Theorem

Let Y be a random variable* and $(Y_n)_{n \geq 1}$ a sequence of random variables such that

$$\text{for any } \ell \geq 1, \ \lim_{n \to \infty} \kappa_\ell(Y_n, \dots, Y_n) = \kappa_\ell(Y, \dots, Y),$$

then, in distribution,

$$Y_n \longrightarrow Y.$$

* We assume that Y has moments of all orders and that its law is determined by its moments.

Expanding cumulants

Recall $X_n = \sum_{i \leq s} B_{i,s}$. By multilinearity,

$$\kappa_\ell(X_n, \dots, X_n) = \sum_{i_1 \leq s_1, \cdots, i_\ell \leq s_\ell} \kappa_\ell(B_{i_1,s_1}, \dots, B_{i_\ell,s_\ell}).$$

Expanding cumulants

Recall $X_n = \sum_{i \leq s} B_{i,s}$. By multilinearity,

$$\kappa_\ell(X_n, \dots, X_n) = \sum_{i_1 \leq s_1, \cdots, i_\ell \leq s_\ell} \kappa_\ell(B_{i_1,s_1}, \dots, B_{i_\ell,s_\ell}).$$

Let us compute the *total contribution* of lists **i** and **s** with distinct entries:

- for each of these lists, the cumulant $\kappa_\ell(B_{i_1,s_1}, \dots, B_{i_\ell,s_\ell})$ has the same value $f(n)$, which is $O(n^{-2\ell+1})$;
- there are fewer than $n^{2\ell}$ such lists.

Expanding cumulants

Recall $X_n = \sum_{i \leq s} B_{i,s}$. By multilinearity,

$$\kappa_\ell(X_n, \dots, X_n) = \sum_{i_1 \leq s_1, \cdots, i_\ell \leq s_\ell} \kappa_\ell(B_{i_1,s_1}, \dots, B_{i_\ell,s_\ell}).$$

Let us compute the *total contribution* of lists **i** and **s** with distinct entries:

- for each of these lists, the cumulant $\kappa_\ell(B_{i_1,s_1}, \dots, B_{i_\ell,s_\ell})$ has the same value $f(n)$, which is $O(n^{-2\ell+1})$;
- there are fewer than $n^{2\ell}$ such lists.

$\Rightarrow$ Their total contribution is $O(n)$.

A general bound for cumulants of $B_{i,s}$

Consider $\kappa(B_{1,3}B_{2,7}, B_{2,5}, B_{1,3}B_{4,9}, B_{6,8})$.

A general bound for cumulants of $B_{i,s}$

Consider $\kappa(B_{1,3}B_{2,7}, B_{2,5}, B_{1,3}B_{4,9}, B_{6,8})$.

Denote:

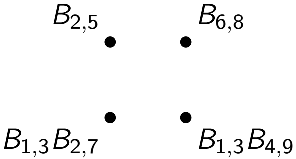
- t the number of different couples (i, s), here 5.

A general bound for cumulants of $B_{i,s}$

Consider $\kappa(B_{1,3}B_{2,7}, B_{2,5}, B_{1,3}B_{4,9}, B_{6,8})$.

Denote:

- t the number of different couples (i,s), here 5.
- m the number of connected components of the following graph

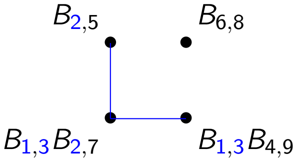


A general bound for cumulants of $B_{i,s}$

Consider $\kappa(B_{1,3}B_{2,7}, B_{2,5}, B_{1,3}B_{4,9}, B_{6,8})$.

Denote:

- t the number of different couples (i, s), here 5.
- m the number of connected components of the following graph (here, $m = 2$)

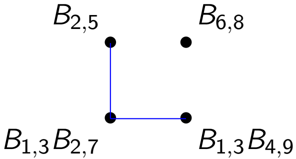


A general bound for cumulants of $B_{i,s}$

Consider $\kappa(B_{1,3}B_{2,7}, B_{2,5}, B_{1,3}B_{4,9}, B_{6,8})$.

Denote:

- t the number of different couples (i, s), here 5.
- m the number of connected components of the following graph (here, $m = 2$)



Then $\kappa(B_{1,3}B_{2,7}, B_{2,5}, B_{1,3}B_{4,9}, B_{6,8}) = O(n^{-t-m+1}) = O(n^{-5})$.

Asymptotic analysis of cumulants

Recall that $\kappa_\ell(X_n, \dots, X_n) = \sum_{i_1 \leq s_1, \cdots, i_\ell \leq s_\ell} \kappa_\ell(B_{i_1,s_1}, \dots, B_{i_\ell,s_\ell})$.

Asymptotic analysis of cumulants

Recall that $\kappa_\ell(X_n,\dots,X_n) = \sum_{i_1\le s_1,\cdots,i_\ell\le s_\ell} \kappa_\ell(B_{i_1,s_1},\dots,B_{i_\ell,s_\ell})$.

Fix some equality conditions between $i_1,\dots,i_\ell,s_1,\dots,s_\ell$ (for example $i_2 = i_5 = s_4$ and $i_3 = s_7$), that is a partition Π of $[\ell]\cup[\ell]$.

Asymptotic analysis of cumulants

Recall that $\kappa_\ell(X_n, \dots, X_n) = \sum_{i_1 \le s_1, \cdots, i_\ell \le s_\ell} \kappa_\ell(B_{i_1,s_1}, \dots, B_{i_\ell,s_\ell})$.

Fix some equality conditions between $i_1, \dots, i_\ell, s_1, \dots, s_\ell$ (for example $i_2 = i_5 = s_4$ and $i_3 = s_7$), that is a partition Π of $[\ell] \cup [\ell]$.

Let us compute the *total contribution* of lists **i** and **s** which fulfills the conditions above:

- for each of these lists, the cumulant $\kappa_\ell(B_{i_1,s_1}, \dots, B_{i_\ell,s_\ell})$ has the same value $f(n)$, which is $O(n^{-m(\Pi)-t(\Pi)+1})$;
- there are fewer than $n^{\#(\Pi)}$ (the number of blocks of Π) such lists.

Asymptotic analysis of cumulants

Recall that $\kappa_\ell(X_n, \dots, X_n) = \sum_{i_1 \le s_1, \cdots, i_\ell \le s_\ell} \kappa_\ell(B_{i_1,s_1}, \dots, B_{i_\ell,s_\ell})$.

Fix some equality conditions between $i_1, \dots, i_\ell, s_1, \dots, s_\ell$ (for example $i_2 = i_5 = s_4$ and $i_3 = s_7$), that is a partition Π of $[\ell] \cup [\ell]$.

Let us compute the *total contribution* of lists **i** and **s** which fulfills the conditions above:

- for each of these lists, the cumulant $\kappa_\ell(B_{i_1,s_1}, \dots, B_{i_\ell,s_\ell})$ has the same value $f(n)$, which is $O(n^{-m(\Pi)-t(\Pi)+1})$;
- there are fewer than $n^{\#(\Pi)}$ (the number of blocks of Π) such lists.

Lemma: $m(\Pi) + t(\Pi) \le (\Pi)$.
$\Rightarrow$ Their total contribution is $O(n)$.

End of the proof

We proved

$$\kappa_\ell(X_n, \dots, X_n) = O(N).$$

End of the proof

We proved

$$\kappa_\ell(X_n, \dots, X_n) = O(N).$$

First order asymptotic

Hence $\mathrm{Var}(X_n/n) = 0(n^{-1})$.
As $\mathbb{E}(X_n/n) \to 1/2$, the variable X_n/n tends to $1/2$.

End of the proof

We proved

$$\kappa_\ell(X_n, \dots, X_n) = O(N).$$

First order asymptotic

Hence $\mathrm{Var}(X_n/n) = 0(n^{-1})$.
As $\mathbb{E}(X_n/n) \to 1/2$, the variable X_n/n tends to $1/2$.

Second order asymptotic

Define $Y_n = n^{-1/2}(X_n - \mathbb{E}(X_n))$. One can show that $\mathrm{Var}(Y_n)$ tends to a constant $c > 0$.

Besides, $\kappa_\ell(Y_n) = n^{-\ell/2}\kappa_\ell(X_n)$ tends to O for $\ell \geq 3$.

$\Rightarrow$ Y_n converges in distribution towards a Gaussian law.

Further work

- introduce a notion of *quantified dependency graph* to systematize this kind of proof.
- large deviation, local limit laws,. . .
- other objects: random graphs $G(n, M)$, orthogonal random matrices,. . .