Comportement asymptotique de statistiques dans des permutations aléatoires

Valentin Féray

LaBRI, CNRS, Bordeaux

Journées ALÉA
Luminy, 6 mars 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. Description of the method

Example: number of fixed points

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

Example: number of fixed points

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

By inclusion-exclusion, the number of *derangements* (= permutations without fixed points) of n is:

$$D(n) = \sum_{0 \leq i \leq n} (-1)^i \binom{n}{i} (n-i)!$$

Example: number of fixed points

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

By inclusion-exclusion, the number of *derangements* (= permutations without fixed points) of n is:

$$D(n) = \sum_{0 \le i \le n} (-1)^i \binom{n}{i} (n-i)!$$

whence the probability of having no fixed points:

$$P(X_n = 0) = \frac{D(n)}{n!} = \sum_{0 \le i \le n} \frac{(-1)^i}{i!} \longrightarrow_{n \to \infty} e^{-1}$$

Example: number of fixed points

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

By inclusion-exclusion, the number of *derangements* (= permutations without fixed points) of n is:

$$D(n) = \sum_{0 \le i \le n} (-1)^i \binom{n}{i} (n-i)!$$

whence the probability of having no fixed points:

$$P(X_n = 0) = \frac{D(n)}{n!} = \sum_{0 \le i \le n} \frac{(-1)^i}{i!} \longrightarrow_{n\to\infty} e^{-1}$$

The probability of having k fixed points follows:

$$P(X_n = k) = \frac{1}{n!} \binom{n}{k} D(n-k) = \frac{1}{k!} \frac{D(n-k)}{(n-k)!} \longrightarrow_{n\to\infty} \frac{e^{-1}}{k!}$$

Example: number of fixed points

We have just proved:

Theorem

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

Remark. We could also have used generating series (see *Analytic combinatorics*, example IX.4).

Example: number of fixed points

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 Bernouilli variable of parameter $1/n$, (F_i takes value 1 if $\sigma(i) = i$).

Example: number of fixed points

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* Bernouilli variables of parameter $1/n$ converges toward a Poisson law of parameter 1.

Example: number of fixed points

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 Bernouilli variable of parameter $1/n$, (F_i takes value 1 if $\sigma(i) = i$).

Reminder: law of small numbers

The sum of n *independent* Bernouilli 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

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* Bernouilli 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!) and use it to reprove the theorem.

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

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

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

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

Covariance of the F_i

Back to fixed points and uniform measure:

Easy computation: if $i \neq j$,

$$\begin{aligned}\mathrm{Cov}(F_i, F_j) &= \mathbb{E}(F_i F_j) - \mathbb{E}(F_i)\mathbb{E}(F_j) \\ &= \frac{1}{n(n-1)} - \left(\frac{1}{n}\right)^2 = \frac{1}{n^2(n-1)}\end{aligned}$$

Covariance of the F_i

Back to fixed points and uniform measure:

Easy computation: if $i \neq j$,

$$\begin{aligned}\mathrm{Cov}(F_i, F_j) &= \mathbb{E}(F_i F_j) - \mathbb{E}(F_i)\mathbb{E}(F_j) \\ &= \frac{1}{n(n-1)} - \left(\frac{1}{n}\right)^2 = \frac{1}{n^2(n-1)}\end{aligned}$$

Remark. $\mathrm{Cov}(F_i, F_j) \ll \mathbb{E}(F_i F_j), \mathbb{E}(F_i)\mathbb{E}(F_j)$.
Confirms the intuition of *almost independence*.

Covariance of the F_i

Back to fixed points and uniform measure:

Easy computation: if $i \neq j$,

$$\begin{aligned}\mathrm{Cov}(F_i, F_j) &= \mathbb{E}(F_i F_j) - \mathbb{E}(F_i)\mathbb{E}(F_j) \\ &= \frac{1}{n(n-1)} - \left(\frac{1}{n}\right)^2 = \frac{1}{n^2(n-1)}\end{aligned}$$

Remark. $\mathrm{Cov}(F_i, F_j) \ll \mathbb{E}(F_i F_j), \mathbb{E}(F_i)\mathbb{E}(F_j)$.
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 fixed points

Recall: F_i is the characteristic function of the event $\sigma(i) = i$.
If h, i and j are pairwise distinct,

$$\kappa_3(F_h, F_i, F_j) = \frac{1}{n(n-1)(n-2)} - 3\frac{1}{n^2(n-1)} + 2\frac{1}{n^3}$$

Cumulants of fixed points

Recall: F_i is the characteristic function of the event $\sigma(i) = i$.
If h, i and j are pairwise distinct,

$$\begin{aligned}\kappa_3(F_h, F_i, F_j) &= \frac{1}{n(n-1)(n-2)} - 3\frac{1}{n^2(n-1)} + 2\frac{1}{n^3} \\ &= \frac{4}{n^3(n-1)(n-2)} = O(n^{-5})\end{aligned}$$

Cumulants of fixed points

Recall: F_i is the characteristic function of the event $\sigma(i) = i$.
If h, i and j are pairwise distinct,

$$\begin{aligned}\kappa_3(F_h, F_i, F_j) &= \frac{1}{n(n-1)(n-2)} - 3\frac{1}{n^2(n-1)} + 2\frac{1}{n^3} \\ &= \frac{4}{n^3(n-1)(n-2)} = O(n^{-5})\end{aligned}$$

In general,

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

where t is the number of distinct values in the list $i_1, \dots, i_\ell$.

Remark. A priori, it is a rational function of degree $-t$. It is quite technical to prove that it has in fact degree $-2t+1$.

Cumulants and convergence in distribution

Our goal: show that $\sum_i F_i$ converges in distribution towards a Poisson law.

Cumulants and convergence in distribution

Our goal: show that $\sum_i F_i$ converges in distribution towards a Poisson law.

Cumulants are a good tool to prove convergence in distribution

Theorem

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

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

then, in distribution,

$$X_n \longrightarrow X.$$

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

Asymptotic analysis of cumulants

Recall $X_n = \sum_{1 \le i \le n} F_i$. By multilinearity,

$$\kappa_\ell(X_n, \dots, X_n) = \sum_{1 \le i_1, \dots, i_\ell \le n} \kappa_\ell(F_{i_1}, \dots, F_{i_\ell})$$

Asymptotic analysis of cumulants

Recall $X_n = \sum_{1 \le i \le n} F_i$. By multilinearity,

$$\kappa_\ell(X_n, \dots, X_n) = \sum_{1 \le i_1, \dots, i_\ell \le n} \kappa_\ell(F_{i_1}, \dots, F_{i_\ell})$$

Fix some positive integer $t \le \ell$.

- There are $S(\ell, t) n(n-1) \dots (n-t+1)$ lists $(i_1, \dots, i_\ell)$ with exactly t distinct values.

Notation: $S(\ell, t)$ is the number of set partitions of $[\ell]$ with t parts.

Asymptotic analysis of cumulants

Recall $X_n = \sum_{1 \leq i \leq n} F_i$. By multilinearity,

$$\kappa_\ell(X_n, \dots, X_n) = \sum_{1 \leq i_1, \dots, i_\ell \leq n} \kappa_\ell(F_{i_1}, \dots, F_{i_\ell})$$

Fix some positive integer $t \leq \ell$.

- There are $S(\ell, t) n(n-1) \dots (n-t+1)$ lists $(i_1, \dots, i_\ell)$ with exactly t distinct values.
- For each one of these sequences, $\kappa_\ell(F_{i_1}, \dots, F_{i_\ell}) = O(n^{-2t+1})$.

See previous slide: moreover, the O is *uniform* (depends only on ℓ).

Asymptotic analysis of cumulants

Recall $X_n = \sum_{1 \le i \le n} F_i$. By multilinearity,

$$\kappa_\ell(X_n, \dots, X_n) = \sum_{1 \le i_1, \dots, i_\ell \le n} \kappa_\ell(F_{i_1}, \dots, F_{i_\ell})$$

Fix some positive integer $t \le \ell$.

- There are $S(\ell, t) n(n-1) \dots (n-t+1)$ lists $(i_1, \dots, i_\ell)$ with exactly t distinct values.
- For each one of these sequences, $\kappa_\ell(F_{i_1}, \dots, F_{i_\ell}) = O(n^{-2t+1})$.
- Hence, the *total* contribution of these lists is $O(n^{-t+1})$.

See previous slide: moreover, the O is *uniform* (depends only on ℓ).

Asymptotic analysis of cumulants

Recall $X_n = \sum_{1 \leq i \leq n} F_i$. By multilinearity,

$$\kappa_\ell(X_n, \dots, X_n) = \sum_{1 \leq i_1, \dots, i_\ell \leq n} \kappa_\ell(F_{i_1}, \dots, F_{i_\ell})$$

Fix some positive integer $t \leq \ell$.

- There are $S(\ell, t)n(n-1)\dots(n-t+1)$ lists $(i_1, \dots, i_\ell)$ with exactly t distinct values.
- For each one of these sequences, $\kappa_\ell(F_{i_1}, \dots, F_{i_\ell}) = O(n^{-2t+1})$.
- Hence, the *total* contribution of these lists is $O(n^{-t+1})$.

Finally, we get:

$$\kappa_\ell(X_n, \dots, X_n) = \sum_{1 \leq i \leq n} \kappa_\ell(F_i, \dots, F_i) + O(N^{-1}).$$

End of the proof

We proved

$$\kappa_\ell(X_n, \dots, X_n) = \sum_{1 \le i \le n} \kappa_\ell(F_i, \dots, F_i) + O(N^{-1}).$$

End of the proof

We proved

$$\kappa_\ell(X_n, \ldots, X_n) = \sum_{1 \leq i \leq n} \kappa_\ell(F_i, \ldots, F_i) + O(N^{-1}).$$

that is: X_n has *asymptotically* the same cumulant than a sum of n *independent* Bernouilli variables of parameter $1/n$.

End of the proof

We proved

$$\kappa_\ell(X_n, \dots, X_n) = \sum_{1 \leq i \leq n} \kappa_\ell(F_i, \dots, F_i) + O(N^{-1}).$$

that is: X_n has *asymptotically* the same cumulant than a sum of n *independent* Bernouilli variables of parameter $1/n$.

$\Longrightarrow$ it converges in distribution towards a Poisson law of parameter 1 (law of small numbers).

Main steps of the proof for dashed patterns

Notation: $B_{i,s}(\sigma) = \begin{cases} 1 & \text{if } \sigma(i) = s; \\ 0 & \text{else.} \end{cases}$

Note: the number of occurrences of any dashed pattern writes as a sum of products of such variables.

Main steps of the proof for dashed patterns

Notation: $B_{i,s}(\sigma) = \begin{cases} 1 & \text{if } \sigma(i) = s; \\ 0 & \text{else.} \end{cases}$

Note: the number of occurrences of any dashed pattern writes as a sum of products of such variables.

Hence, its cumulants are *huge* sums of cumulants of product of $B_{i,s}$.

Main steps of the proof for dashed patterns

Notation: $B_{i,s}(\sigma) = \begin{cases} 1 & \text{if } \sigma(i) = s; \\ 0 & \text{else.} \end{cases}$

Note: the number of occurrences of any dashed pattern writes as a sum of products of such variables.

Hence, its cumulants are *huge* sums of cumulants of product of $B_{i,s}$.

We need a bound for joint cumulants of products of $B_{i,s}$ (next slide).

Main steps of the proof for dashed patterns

Notation: $B_{i,s}(\sigma) = \begin{cases} 1 & \text{if } \sigma(i) = s; \\ 0 & \text{else.} \end{cases}$

Note: the number of occurrences of any dashed pattern writes as a sum of products of such variables.

Hence, its cumulants are *huge* sums of cumulants of product of $B_{i,s}$.

We need a bound for joint cumulants of products of $B_{i,s}$ (next slide).

Then, one has to determine which summands have the biggest contribution to cumulants (not easy!)...

The general bound for cumulants

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

The general bound for cumulants

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

Denote:

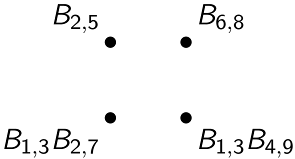
- t the number of different couples (i, s), here 5.

The general bound for cumulants

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

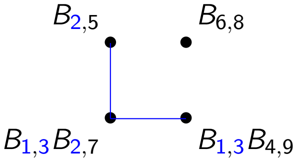


The general bound for cumulants

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, $m = 2$)

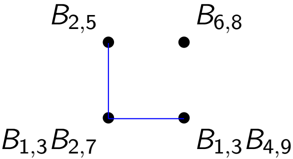


The general bound for cumulants

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

Future work

- More statistics: Generalized patterns (with some adjacencies in places and values) or even more general setting (where we can add equalities/inequalities between some places and values).
- More objects: random graphs, ...
- More precise results: speed of convergence, local limit laws, large deviation...