Non-uniform permutations biased according to their records

Mathilde Bouvel
(Loria, CNRS, Univ. Lorraine)

talk based on joint work and work in progress with
Nicolas Auger, Cyril Nicaud and Carine Pivoteau

BIRS workshop on *Permutations and Probability*, September 2021

Non-uniform permutations

Context:

Analysis of algorithms working on arrays of numbers (sorting, ...)

Average-case analysis of algorithms:

- The uniform distribution on the data set is usually assumed.
- It provides a first answer, but it is not always realistic.
  E.g., sorting algorithms are often used on data which is already "almost sorted". (Ex. of TimSort [Auger, Jugé, Nicaud, Pivoteau, 2018])

$\Rightarrow$ Find non-uniform models with good balance between simplicity (so that we can study it) and accuracy (in terms of modeling data)

Some classical models for non-uniform permutations

- Ewens: $\mathbb{P}(\sigma)$ is proportional to $\theta^{\text{number of cycles of } \sigma}$
- Mallows: $\mathbb{P}(\sigma)$ is proportional to $\theta^{\text{number of inversions of } \sigma}$

Our record-biased permutations

Goal: A non-uniform distribution on permutations, which gives higher probabilities to permutations that are “almost sorted”.

Record-biased permutations:

- A record is an element larger than all those preceding it.
  Example: **3 4** 1 2 **6 8** 7 **9** 5 has 5 records.
- Roughly, a permutation with many records is “almost sorted”. More formally, the number of non-records is a measure of presortedness as defined by [Manilla, 1985], see [Auger, Bouvel, Pivoteau, Nicaud, 2016].
- In our model, $\mathbb{P}(\sigma)$ is proportional to $\theta^{\text{number of records of } \sigma}$.

Remark: Related to the Ewens distribution via Foata’s *fundamental bijection*, which sends number of cycles to number of records.
Example: $2\,4\,3\,1\,9\,6\,8\,7\,5 = (3)(4\,1\,2)(6)(8\,7)(9\,5) \rightarrow \mathbf{3\,4}\,1\,2\,\mathbf{6\,8}\,7\,\mathbf{9}\,5$

Outline of the talk

Goal: Describe properties of the model of record-biased permutations.
Applications to the analysis of algorithms will be discussed only a little.

Results obtained:

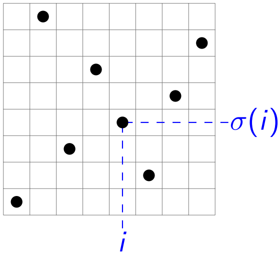
- Random sampling can be done in linear time, in several ways.
  - viewing permutations as words
  - or viewing permutations as *diagrams*
- Behavior of classical permutation statistics:
  - We obtain precise probabilities of elementary events.
  - We deduce their expected values and asymptotic distribution.
  - Applications to analysis of algorithms [ABNP, 2016]:
    - expected running time of InsertionSort,
    - expected number of mispredictions in MinMaxSearch
- What does a large record-biased permutation typically look like?
  - We describe the (deterministic) permuton limit for our model.

Before we dive in: Several ways of seeing a permutation

We can represent a permutation of size n, say $\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow & \downarrow \\ 1 & 8 & 3 & 6 & 4 & 2 & 5 & 7 \end{pmatrix}$

as

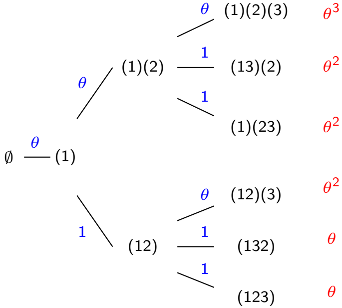
- a word (a.k.a. *one-line representation*): $\sigma = 1\ 8\ 3\ 6\ 4\ 2\ 5\ 7$
- a product of cycles: $\sigma = (1)\ (3)\ (8\ 7\ 5\ 4\ 6\ 2)$
- a diagram, *i.e.* an $n \times n$ grid with points at coordinates $(i, \sigma(i))$:

$\sigma =$ 

Linear random samplers

Random sampling combining Ewens and Foata

- Ewens-distributed permutations can be sampled in linear time using a variant of the Chinese restaurant process:
  - Insert i from 1 to n.
  - At step i, create a new cycle (i) with probability $\frac{\theta}{\theta+i-1}$, or insert i in an existing cycle, immediately after a previously inserted element, each with probability $\frac{1}{\theta+i-1}$.

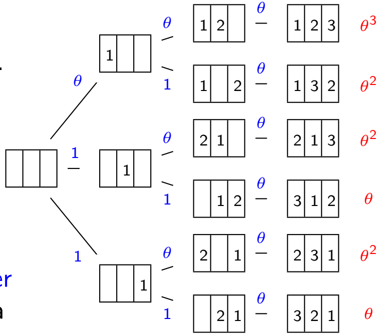


- Using appropriate data structures, we can implement Foata's transform in linear time,
  hence sampling record-biased permutation in linear time.
- We can also do it directly, with appropriate data structures.

Random sampling of permutations as words

Another sampling procedure for
record-biased permutations of size n:

- Start with an empty array of n cells.
- Insert i from 1 to n.
- At step i,
  - either insert i in the leftmost empty cell (this creates a record): with probability $\frac{\theta}{\theta+n-i}$;
  - or insert i in one of the $n-i$ other empty cells (this does not create a record): with probability $\frac{1}{\theta+n-i}$ for each such cell.

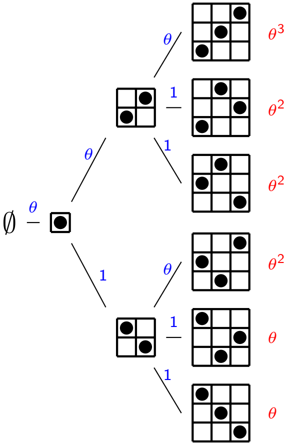


- Using appropriate data structures (one linked-list and two auxiliary arrays), we can implement this sampling procedure in linear time.

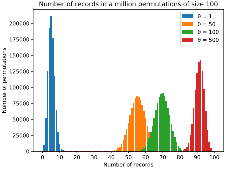
Random sampling of permutations as diagrams

Yet another sampling procedure for record-biased permutations of size n:

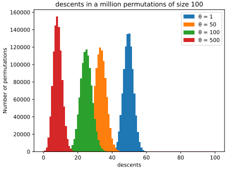
- Start with an empty diagram.
- For i from 1 to n, insert an i-th column and a new row, with a new point at their intersection:
  - with probability $\frac{\theta}{\theta+i-1}$, the new row is the topmost one (hence the new point a record);
  - for each $j < i$, with probability $\frac{1}{\theta+i-1}$, the new row is just under the point in column j (hence not a record).



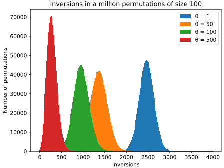
- Using appropriate data structures (a linked list with direct access to its cells), we can implement this sampling procedure in linear time.



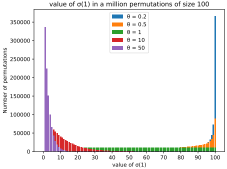
Number of records



Number of descents



Number of inversions



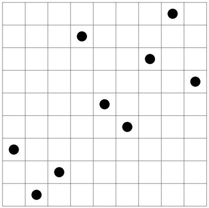
Value of the first element

Histograms are for 10^6 permutations, of size $n = 100$,
and for $\theta = 1, 50, 100$ and 500 (resp. $\theta = 0.2, 0.5, 1, 10$ and 50).

Playing with the samplers: a typical diagram arises

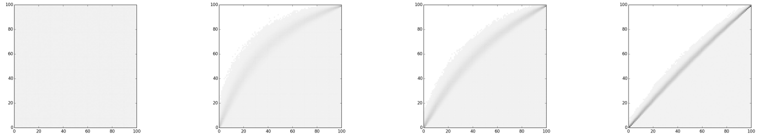
Recall that the diagram of a permutation σ of size n is the set of points at coordinates $(i, \sigma(i))$ for $1 \leq i \leq n$.

The normalized diagram of σ is the same picture, rescaled to the unit square.



$\sigma = 3\ 1\ 2\ 8\ 5\ 4\ 7\ 9\ 6$

Pictures obtained overlapping 10 000 permutations of size 100 sampled according to the record-biased model with $\theta = 1, 50, 100$ and 500:



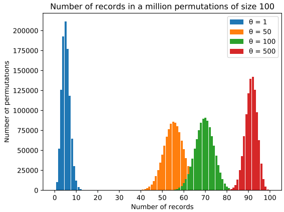
We explain it by describing the permuton limit of record-biased permutations.

Behavior of statistics

Recall that a record of a permutation σ is given by an index i such that $\sigma(i) > \sigma(j)$ for all $j < i$.

Results:

- The expected number of records in record-biased permutations of size n is $\sum_{i=1}^{n} \frac{\theta}{\theta+i-1}$.
- For fixed θ, it is $\sim \theta \log(n)$ as $n \to \infty$.
- For fixed θ, the distribution of the number of records in record-biased permutations is asymptotically Gaussian.



Histogram for 10^6 permutations, of size $n = 100$, and for $\theta = 1, 50, 100$ and 500.

Proof idea: *Via* the Foata bijection, records in record-biased permutations correspond to cycles in Ewens-distributed permutations.

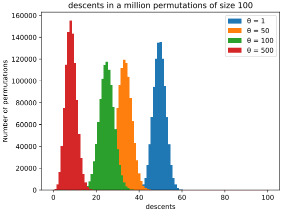
Remark: Expectation can also be derived from $\mathbb{P}(\text{record at } i) = \frac{\theta}{\theta+i-1}$, which is obvious from the random sampler of diagrams.

Number of descents

A descent of a permutation σ is given by an index i s.t. $\sigma(i-1) > \sigma(i)$.

Results:

- The expected number of descents in record-biased permutations of size n is $\frac{n(n-1)}{2(\theta+n-1)}$
- For fixed θ, it is $\sim \frac{n}{2}$ as $n \to \infty$.
- For fixed θ, the distribution of the number of descents in record-biased permutations is asymptotically Gaussian.



Histogram for 10^6 permutations, of size $n = 100$, and for $\theta = 1, 50, 100$ and 500.

Proof idea: Descents in record-biased permutations correspond to anti-exceedances in Ewens-distributed permutations. These are closely related to weak exceedances studied by [Féray, 2013].

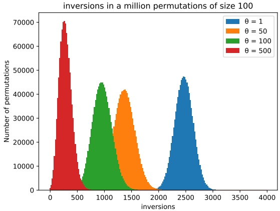
Remark: $\mathbb{P}$(descent at i) and hence the expectation can also be derived from the random sampler of diagrams.

Number of inversions: statements

An inversion of σ is given by a pair (i,j) s.t. $i<j$ and $\sigma(i)>\sigma(j)$.

Results:

- The expected number of inversions in record-biased permutations of size n is $\frac{n(n+1-2\theta)}{4}+\frac{\theta(\theta-1)}{2}\sum_{i=1}^{n}\frac{1}{\theta+i-1}$
- For fixed θ, it is $\sim \frac{n^2}{4}$ as $n\to\infty$.
- For fixed θ, the distribution of the number of inversions in record-biased permutations is asymptotically Gaussian.



Histogram for 10^6 permutations, of size $n=100$, and for $\theta=1,50,100$ and 500.

Remark: No known natural analogue on Ewens-distributed permutations.

Number of inversions: proof sketch

Let inv_j be the number of inversions of the form (i, j),
and $\mathrm{inv} = \sum_j \mathrm{inv}_j$ be the number of inversions.

Remarks: With the sampling procedure as diagrams

- inv_j is completely determined by step j of the procedure, and depends only on the height of the j-th point inserted;
- in particular, for $j \neq j'$, inv_j and $\mathrm{inv}_{j'}$ are independent.

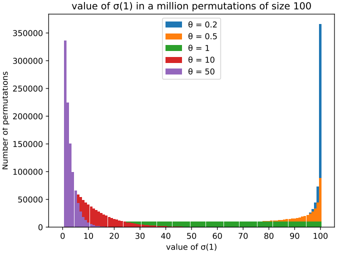
Expectation: The first remark gives $\mathbb{P}(\mathrm{inv}_j = k) = \begin{cases} \frac{\theta}{\theta+j-1} & \text{if } k = 0 \\ \frac{1}{\theta+j-1} & \text{if } k \neq 0 \end{cases}$,
from which we deduce expressions for
$\mathbb{E}(\mathrm{inv}_j) = \sum_k k \cdot \mathbb{P}(\mathrm{inv}_j = k)$ and $\mathbb{E}(\mathrm{inv}) = \sum_j \mathbb{E}(\mathrm{inv}_j)$.

Asymptotic normality: Follows from independence comparing the order of $\sum_j \mathbb{E}(\mathrm{inv}_j^3) = \Theta(n^4)$ and $\sqrt{\mathbb{V}(\mathrm{inv})}^3 = \Theta(n^{3/2})$.

Value of the first element: statements

Results:

- The expected value of $\sigma(1)$ in record-biased permutations of size n is $\frac{\theta+n}{\theta+1}$
- For fixed θ, it is $\sim \frac{n}{\theta+1}$ as $n \to \infty$.
- For fixed θ, asymptotically, the rescaled first value $\sigma(1)/n$ in a record-biased permutation of size n follows a beta distribution of parameters $(1, \theta)$.



Histogram for 10^6 perm., of size $n = 100$, and for $\theta = 0.2, 0.5, 1, 10$ and 50.

Remark: Corresponds to the minimum over all cycles of the maximal value in a cycle for Ewens-distributed permutations.
This statistics was not studied so far.

Value of the first element: proof sketch

Expectation: We use the sampling procedure as words.

- The first element is k when the first $k-1$ insertions do not create records but the k-th insertion creates a record.
- Therefore $\mathbb{P}(\sigma(1)=k) = \prod_{i=1}^{k-1} \frac{n-i}{\theta+n-i} \cdot \frac{\theta}{\theta+n-k} = \frac{(n-1)!\,\theta^{(n-k)}\theta}{(n-k)!\theta^{(n)}}$, where $x^{(m)} = x(x+1)\dots(x+m-1)$ is the rising factorial.
- (Magical?) simplifications arise giving $\mathbb{E}(\sigma(1)) = \frac{\theta+n}{\theta+1}$.

Asymptotic distribution: We compute moments of $\sigma(1)$ similarly.

- The computation of $\mathbb{E}(\sigma(1)^r)$ uses similar simplifications and involves Eulerian polynomials $A_r(z)$ (because $\sum_n n^r z^n = \frac{zA_r(z)}{(1-z)^{r+1}}$).
- We obtain $\mathbb{E}(\sigma(1)^r) \sim_{n\to\infty} \frac{r!n^r}{(\theta+1)^{(r)}}$.
- After normalization, we recognize the r-th moment $\frac{r!}{(\theta+1)^{(r)}}$ of a beta distribution of parameter $(1,\theta)$.

One remark: Various regimes for θ

For our four statistics, we have:

- formula (depending on θ and n) for its expectation, valid for θ fixed and $\theta = \theta(n)$;
- the asymptotic behavior of these expectations when θ is fixed;
- the limiting distribution when θ is fixed.

Asymptotic behavior of expectations in various regimes for θ:

	$\theta = 1$ (uniform)	fixed $\theta > 0$	$\theta = n^\epsilon$, $0 < \epsilon < 1$	$\theta = \lambda n$, $\lambda > 0$	$\theta = n^\delta$, $\delta > 1$
records	$\log n$	$\theta \cdot \log n$	$(1-\epsilon) \cdot n^\epsilon \log n$	$\lambda \log(1 + 1/\lambda) \cdot n$	n
descents	$n/2$	$n/2$	$n/2$	$n/2(\lambda + 1)$	$n^{2-\delta}/2$
inversions	$n^2/4$	$n^2/4$	$n^2/4$	$n^2/4 \cdot f(\lambda)$	$n^{3-\delta}/6$
first value	$n/2$	$n/(\theta + 1)$	$n^{1-\epsilon}$	$(\lambda + 1)/\lambda$	1

where $f(\lambda) = 1 - 2\lambda + 2\lambda^2 \log(1 + 1/\lambda)$.

In the last part of the talk, we will focus on the regime $\theta = \lambda n$.

Another remark: analysis of algorithms

InsertionSort:

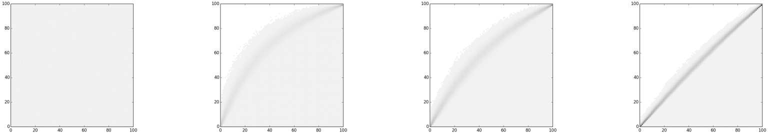
- For $i = 1, 2, \dots n$, swap i with the elements to its left until i reaches the i-th cell.
- The number of swaps is the number of inversions, whose expected behavior is known from the previous table.

MinMaxSearch:

- Several algorithms to find the min and the max in an array: naive version with $2n$ comparisons, clever version with $\frac{3}{2}n$ comparisons.
- But the naive algorithm is typically more efficient on uniform data! Why? Not only the comparisons count in practice.
- The *branch predictors* cause *mispredictions*, hence a slow-down. We quantify this by computing the average number of mispredictions.
- This also explains why the clever algorithm is more efficient on "almost sorted" data (in some regimes for θ).

Permuton limit of record-biased permutations

(in the regime $\theta = \lambda n$)



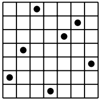
The framework of permutons

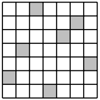
[Hoppen, Kohayakawa, Moreira, Rath, Sampaio, 2013]

Definition: A permuton μ is a probability measure on the unit square with uniform projections (or marginals):

$$\text{for all } a < b \text{ in } [0,1],\ \mu([a,b] \times [0,1]) = \mu([0,1] \times [a,b]) = b - a.$$

Remark: The normalized diagrams of permutations (denoted σ) are essentially permutons (denoted μ_σ)





Replacing each point $(i/n, \sigma(i)/n)$ by a little square $[(i-1)/n, i/n] \times [(\sigma(i)-1)/n, \sigma(i)/n]$, and distributing the mass 1 uniformly on these little squares

Convergence of a sequence of permutations (σ_n) to a permuton μ:

- inherited from the weak convergence of measures, namely:
- $\sigma_n \to \mu$ when $\sup_{R \text{ rectangle } \subset [0,1]^2} |\mu_{\sigma_n}(R) - \mu(R)| \to 0$ as $n \to +\infty$.
- If each σ_n has size n, taking R of the form $[0, i/n] \times [0, j/n]$ is enough.

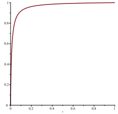
Permuton limit of record-biased permutations

Theorem:
Let σ_n be a random record-biased permutation of size n for $\theta = \lambda n$.
μ_{σ_n} converges in probability to $\mu = \mu_c + \mu_u$ defined below.

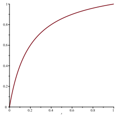
Letting $f_\lambda(x) = \frac{x(\lambda+1)}{\lambda+x}$, we define

- μ_u is the uniform measure
  of total mass $c_\lambda \int_0^1 f_\lambda$ for $c_\lambda = \frac{1}{\lambda+1}$
  on the area under the curve $y = f_\lambda(x)$;
- μ_c is the measure
  supported by the curve $y = f_\lambda(x)$
  with density $\frac{\lambda}{\lambda+x}$ with respect to Leb_c,
  defined by $Leb_c(x, f_\lambda(x)) = Lebesgue(x)$

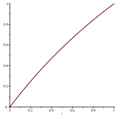
$f_{0.01}$:



$f_{0.2}$:



f_3:



Two steps towards this statement:
guessing μ and proving convergence.

Guessing the limit μ

The pictures suggest to decompose μ as $\mu_u + \mu_c$, with μ_c on a curve, and μ_u uniform under the curve. To determine are:

- the equation $y = f_\lambda(x)$ of the curve,
- how to distribute the mass between μ_c and μ_u.

To find the equation $y = f_\lambda(x)$ of the curve,

- we estimate $\mathbb{P}(\text{max before position } i \text{ is } j)$ for $i \approx xn$ and $j \approx yn$;
- we find the relation between x and y which makes this probability not larger than 1, and non-zero once summed over j.

To find the relative measures on the curve and below,

- we compute the measure of the records in σ_n and take the limit in n: this gives the measure $\int_0^1 \frac{\lambda}{\lambda+x} dx$ on the curve;
- we distribute uniformly the mass $c_\lambda \int_0^1 f_\lambda(x)dx$ below the curve, for c_λ s.t. $\int_a^b (\frac{\lambda}{\lambda+x} + c_\lambda f_\lambda(x))dx = b - a$.

Wrapping up

- We introduced a new model of non-uniform random permutations
  - with a bias toward sortedness *via* their records,
  - motivated by the analysis of algorithms,
  - and with applications there.
- Our model is however closely related to the Ewens model by Foata's bijection.
- We have several efficient procedures for sampling our record-biased permutations.
- We described properties of this model, namely
  - the behavior of some classical statistics
  - and the permuton limit

!! Thank you !!

Any questions or suggestions?