Non-uniform permutations biased according to their records

Mathilde Bouvel
(Loria, CNRS, Univ. Lorraine)

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

Dagstuhl workshop on *Pattern Avoidance, Statistical Mechanics and Computational Complexity*, March 2023

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: $2\,4\,3\,1\,9\,6\,8\,7\,5 = (3)(4\,1\,2)(6)(8\,7)(9\,5) \rightarrow \mathbf{3}\,\mathbf{4}\,1\,2\,\mathbf{6}\,\mathbf{8}\,7\,\mathbf{9}\,5$

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 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.

Additional result: about the height of binary search trees associated with record-biased permutations [Corsini, 2022]

Linear random samplers

Some remarks about these random samplers

Sampling relying on Ewens and Foata: I will present two samplers that generate record-biased permutations directly. But it is possible to sample (in linear time) random permutations that are Ewens-distributed, *e.g.*

- using a variant of the Chinese restaurant process,
- or using the branching process known as Feller coupling.

Then, implementing Foata's bijection (in linear time) provides (linear time) random samplers for record-biased permutations.

Several uses of random samplers:

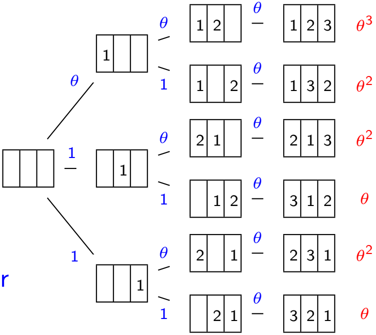
- In practice: to observe your objects!
- In theory: to prove properties of your objects, relying on the underlying process that generates your objects.

For the second item, it is much more convenient to sample record-biased permutations directly, rather than going through Ewens and Foata.

Random sampling of permutations as words

A 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.

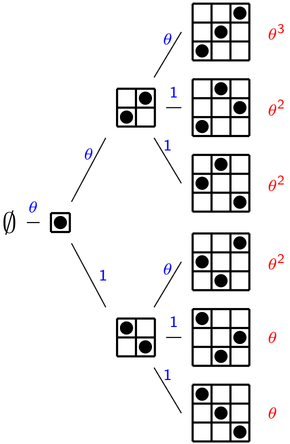


- Using appropriate data structures (one linked-list and two auxiliary arrays), we can implement this sampling procedure in linear time.

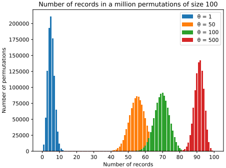
Random sampling of permutations as diagrams

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

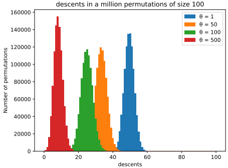
- Start with an empty diagram.
- For i from 1 to n, insert an i-th column and a new row, with a new point at their intersection:
  - with probability $\frac{\theta}{\theta+i-1}$, the new row is the topmost one (hence the new point a record);
  - for each $j < i$, with probability $\frac{1}{\theta+i-1}$,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



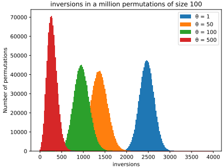
- Using appropriate data structures (a linked list with direct access to its cells), we can implement this sampling procedure in linear time.



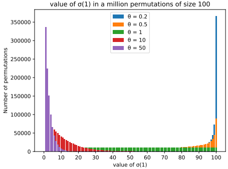
Number of records



Number of descents



Number of inversions



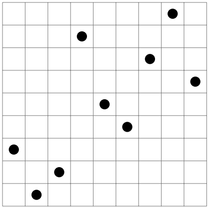
Value of the first element

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

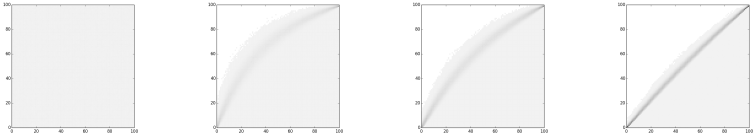
Recall that the diagram of a permutation σ of size n is the set of points at coordinates $(i, \sigma(i))$ for $1 \leq i \leq n$.

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:



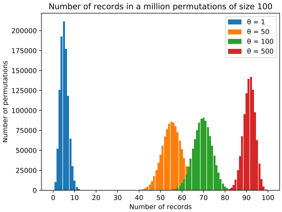
We explain it by describing the permuton limit of record-biased permutations (which is a deterministic permuton).

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 $n \to \infty$.



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

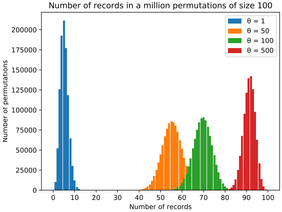
Remark: Expectation can also be derived from $\mathbb{P}(\text{record at } i) = \frac{\theta}{\theta+i-1}$, which is obvious from the random sampler of diagrams.

Number of records

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.

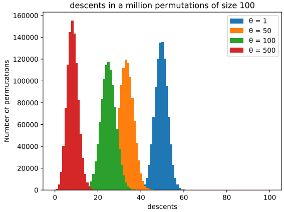
Remark: Expectation can also be derived from $\mathbb{P}(\text{record at } i) = \frac{\theta}{\theta+i-1}$, which is obvious from the random sampler of diagrams.

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 $n \to \infty$.



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

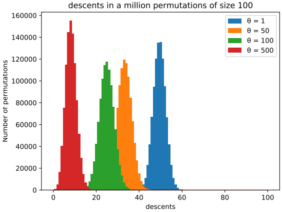
Remark: $\mathbb{P}$(descent at i) and hence the expectation can also be derived from the random sampler of diagrams.

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].

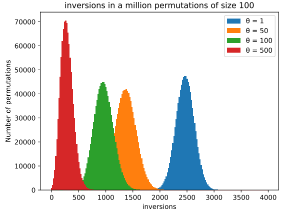
Remark: $\mathbb{P}(\text{descent at } i)$ and hence the expectation can also be derived from the random sampler of diagrams.

Number of inversions

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.

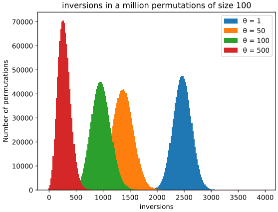
Remark: No known natural analogue on Ewens-distributed permutations.

Number of inversions

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.



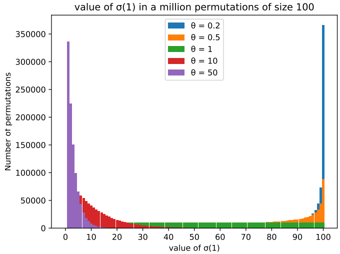
Histogram for 10^6 permutations, of size $n=100$, and for $\theta=1,50,100$ and 500.

Proof ingredients: Writing the number of inversions as $\sum_j \mathrm{inv}_j$ where inv_j is the number of inversions of the form (i,j), use the sampling procedure as diagrams to compute the distribution of each inv_j and show that they are independent.

Value of the first element

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.

Proof ingredients: The sampling procedure as words, and (magical) computations. But is there a simple proof that $\mathbb{E}(\sigma(1)) = \frac{\theta+n}{\theta+1}$???

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:

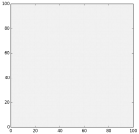
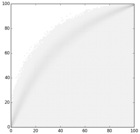
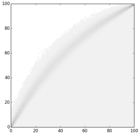
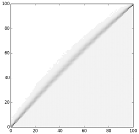
- For $i = 1, 2, \dots n$, swap i with the elements to its left until i reaches the i-th cell.
- The number of swaps is the number of inversions, whose expected behavior is known from the previous table.

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

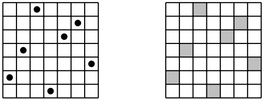
[Hoppen, Kohayakawa, Moreira, Rath, Sampaio, 2013]

Informally, a permuton is the rescaled diagram of an infinite permutation.

(Formal) 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 $\displaystyle\sup_{R \text{ rectangle } \subset [0,1]^2} |\mu_{\sigma_n}(R) - \mu(R)| \to 0$ as $n \to +\infty$.

Permuton limit of record-biased permutations

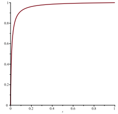
Theorem:
Let σ_n be a random record-biased permutation of size n for $\theta = \lambda n$.
μ_{σ_n} converges in probability to $\mu = \mu_c + \mu_u$ defined below.

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

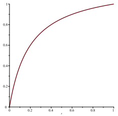
- μ_u is the uniform measure
  of total mass $c_\lambda \int_0^1 f_\lambda$ for $c_\lambda = \frac{1}{\lambda+1}$
  on the area under the curve $y = f_\lambda(x)$;
- μ_c is the measure
  supported by the curve $y = f_\lambda(x)$
  with density $\frac{\lambda}{\lambda+x}$ with respect to Leb_c,
  defined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

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

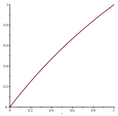
$f_{0.01}$:



$f_{0.2}$:



f_3:



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 $\mathbb{P}($max before position i 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?