Cycles in Mallows random permutations

Teun Verstraaten
Groningen University

(joint work with Tobias Müller)

Bernoulli Institute for Mathematics, Computer Science and Artificial Intelligence
Groningen

12th of May 2022

The Mallows model

- $$\text{inv}(\pi) := |\{(i,j) : i < j \text{ and } \pi(i) > \pi(j)\}|,$$

  denotes the number of *inversions* of a permutation π

The Mallows model

- $$\mathrm{inv}(\pi) := |\{(i,j) : i < j \text{ and } \pi(i) > \pi(j)\}|,$$
  denotes the number of *inversions* of a permutation π
- We sample a permutation $\Pi_n \in S_n$ according to:
  $$\mathbb{P}(\Pi_n = \pi) = \frac{q^{\mathrm{inv}(\pi)}}{\sum_{\sigma \in S_n} q^{\mathrm{inv}(\sigma)}}.$$
  where $q > 0$ is a parameter

The Mallows model

- $$\operatorname{inv}(\pi) := |\{(i,j) : i < j \text{ and } \pi(i) > \pi(j)\}|,$$

  denotes the number of *inversions* of a permutation π
- We sample a permutation $\Pi_n \in S_n$ according to:

  $$\mathbb{P}(\Pi_n = \pi) = \frac{q^{\operatorname{inv}(\pi)}}{\sum_{\sigma \in S_n} q^{\operatorname{inv}(\sigma)}}.$$

  where $q > 0$ is a parameter
- **Notation** : $\Pi_n \sim \text{Mallows}(n, q)$

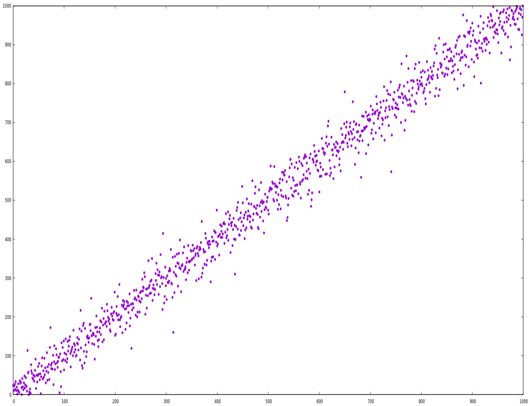
The Mallows model

- $$\mathrm{inv}(\pi) := |\{(i,j) : i < j \text{ and } \pi(i) > \pi(j)\}|,$$
  denotes the number of *inversions* of a permutation π
- We sample a permutation $\Pi_n \in S_n$ according to:
  $$\mathbb{P}(\Pi_n = \pi) = \frac{q^{\mathrm{inv}(\pi)}}{\sum_{\sigma \in S_n} q^{\mathrm{inv}(\sigma)}}.$$
  where $q > 0$ is a parameter
- **Notation** : $\Pi_n \sim \mathrm{Mallows}(n, q)$
- Setting $q = 1$ we retrieve the uniform distribution on S_n

The Mallows model

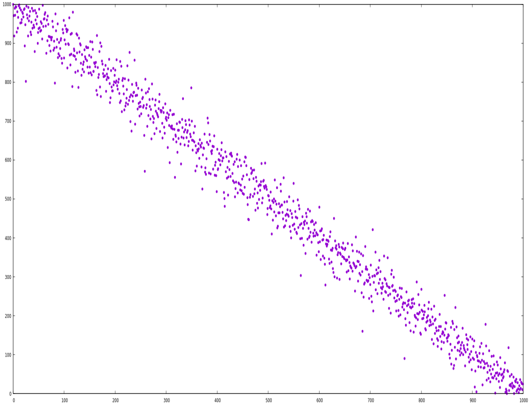
- $$\mathrm{inv}(\pi) := |\{(i,j) : i < j \text{ and } \pi(i) > \pi(j)\}|,$$
  denotes the number of *inversions* of a permutation π
- We sample a permutation $\Pi_n \in S_n$ according to:
  $$\mathbb{P}(\Pi_n = \pi) = \frac{q^{\mathrm{inv}(\pi)}}{\sum_{\sigma \in S_n} q^{\mathrm{inv}(\sigma)}}.$$
  where $q > 0$ is a parameter
- **Notation** : $\Pi_n \sim \mathrm{Mallows}(n, q)$
- Setting $q = 1$ we retrieve the uniform distribution on S_n
- Intuition:
  - when $0 < q < 1$ we stay "close to the identity $i \mapsto i$",
  - when $q > 1$ we stay "close to reverse map $i \mapsto n + 1 - i$".

A simulation



$(n = 1000, q = .95.)$

A simulation



$(n = 1000, q = 1.05.)$

Background

Introduced by C.L. Mallows in 1957 in the context of "statistical ranking theory".

Since then studied in connection with mixing times of Markov chains, finitely dependent colorings of the integers, stable matchings, random binary search trees, learning theory, q-analogs of exchangeability, determinantal point processes, statistical physics, genomics.

Cycle structure

- A cycle of π is a sequence of distinct elements $(x_0, \ldots, x_{k-1})$ such that $\pi(x_i) = x_{i+1 \mod k}$ for $i = 0, \ldots, k-1$

Cycle structure

- A cycle of π is a sequence of distinct elements $(x_0, \ldots, x_{k-1})$ such that $\pi(x_i) = x_{i+1 \mod k}$ for $i = 0, \ldots, k-1$
- If Π_n selected uniformly from S_n then $C_1(\Pi_n) \to \mathrm{Pois}(1)$ as $n \to \infty$

Cycle structure

- A cycle of π is a sequence of distinct elements $(x_0, \ldots, x_{k-1})$ such that $\pi(x_i) = x_{i+1 \mod k}$ for $i = 0, \ldots, k-1$
- If Π_n selected uniformly from S_n then $C_1(\Pi_n) \to \text{Pois}(1)$ as $n \to \infty$
- In general if Π_n selected uniformly

$$(C_1(\Pi_n), \ldots, C_\ell(\Pi_n)) \to (\text{Pois}(1), \text{Pois}(1/2), \ldots, \text{Pois}(1/\ell))$$

(Gontcharoff [1941] and Kolchin [1976])

We are interested in the cycle structure

$$(C_1(\Pi_n), C_2(\Pi_n), \ldots, C_\ell(\Pi_n))$$

where $\Pi_n \sim \text{Mallows}(n, q)$ and $q \neq 1$

Cycle structure when $q < 1$

Cycle structure when $q < 1$

Theorem (He 2021+, Müller+V 2022+)

Fix $0 < q < 1$ and let $\Pi_n \sim \mathrm{Mallows}(n, q)$. There exist positive constants $m_1, m_2, \dots$ and an infinite matrix $P \in \mathbb{R}^{\mathbb{N}\times\mathbb{N}}$ such that for all $\ell \geq 1$ we have

$$\frac{1}{\sqrt{n}}\left(C_1(\Pi_n) - m_1 n, \dots, C_\ell(\Pi_n) - m_\ell n\right) \quad \xrightarrow[n\to\infty]{d} \quad \mathcal{N}_\ell(\underline{0}, P_\ell),$$

where $\mathcal{N}_\ell(\cdot,\cdot)$ denotes the ℓ–dimensional multivariate normal distribution and P_ℓ is the submatrix of P on the indices $[\ell] \times [\ell]$.

Cycle structure when $q < 1$

Theorem (He 2021+, Müller+V 2022+)

Fix $0 < q < 1$ and let $\Pi_n \sim \mathrm{Mallows}(n, q)$. There exist positive constants $m_1, m_2, \ldots$ and an infinite matrix $P \in \mathbb{R}^{\mathbb{N}\times\mathbb{N}}$ such that for all $\ell \geq 1$ we have

$$\frac{1}{\sqrt{n}}\left(C_1(\Pi_n) - m_1 n, \ldots, C_\ell(\Pi_n) - m_\ell n\right) \quad \xrightarrow[n\to\infty]{d} \quad \mathcal{N}_\ell(\underline{0}, P_\ell),$$

where $\mathcal{N}_\ell(\cdot,\cdot)$ denotes the ℓ–dimensional multivariate normal distribution and P_ℓ is the submatrix of P on the indices $[\ell] \times [\ell]$.

- Different behaviour from $q = 1$

Cycle structure when $q < 1$

Theorem (He 2021+, Müller+V 2022+)

Fix $0 < q < 1$ and let $\Pi_n \sim \mathrm{Mallows}(n, q)$. There exist positive constants $m_1, m_2, \ldots$ and an infinite matrix $P \in \mathbb{R}^{\mathbb{N}\times\mathbb{N}}$ such that for all $\ell \geq 1$ we have

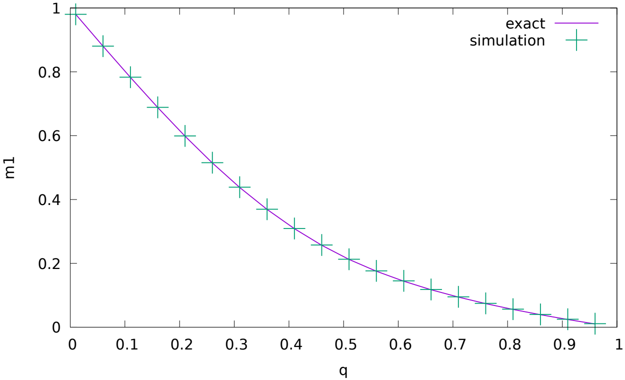
$$\frac{1}{\sqrt{n}}\left(C_1(\Pi_n) - m_1 n, \ldots, C_\ell(\Pi_n) - m_\ell n\right) \quad \xrightarrow[n\to\infty]{d} \quad \mathcal{N}_\ell(\underline{0}, P_\ell),$$

where $\mathcal{N}_\ell(\cdot,\cdot)$ denotes the ℓ–dimensional multivariate normal distribution and P_ℓ is the submatrix of P on the indices $[\ell] \times [\ell]$.

- Different behaviour from $q = 1$: Expect linear number of i–cycles as $n \to \infty$

Cycle structure when $q < 1$

m_1 is the fraction of fixed points



$m_1 \to 1$ as $q \downarrow 0$ (tends to identity perm.)
$m_1 \to 0$ as $q \uparrow 1$

$q > 1$ (cycles of even length)

Theorem (He 2021+, Müller+V 2022+)

For $q > 1$ there exist positive constants $\mu_2, \mu_4, \ldots$ and an infinite matrix $Q \in \mathbb{R}^{\mathbb{N}\times\mathbb{N}}$ such that for all $\ell \geq 1$ we have

$$\frac{1}{\sqrt{n}}(C_2(\Pi_n) - \mu_2 n, \ldots, C_{2\ell}(\Pi_n) - \mu_{2\ell} n) \xrightarrow{d} \mathcal{N}_\ell(\underline{0}, Q_\ell),$$

where $\mathcal{N}_\ell(\cdot,\cdot)$ denotes the ℓ–dimensional multivariate normal distribution and Q_ℓ is the submatrix of Q on the indices $[\ell] \times [\ell]$.

$q > 1$ (cycles of odd length)

Theorem (He 2021+, Müller+V 2022+)

For $q > 1$, *let* $\Pi_n \sim \mathrm{Mallows}(n, q)$

$$(C_1(\Pi_{2n+1}), C_3(\Pi_{2n+1}), \ldots) \quad \textit{converges to some distribution,}$$

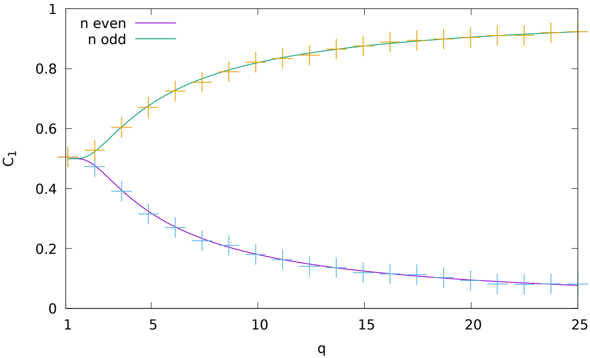
and

$$(C_1(\Pi_{2n}), C_3(\Pi_{2n}), \ldots) \quad \textit{converges to some distribution.}$$

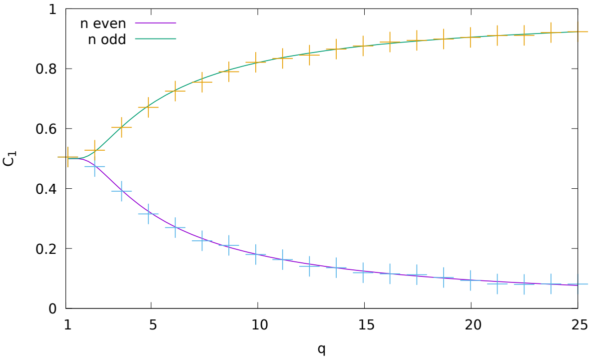
Moreover, the two limiting distributions above are distinct for all $q > 1$.

- We give explicit limiting distributions in terms of an extension of $\mathrm{Mallows}(n, q)$ to $\mathbb{Z}$ due to Gnedin and Olshanski

Fixed points for $q > 1$



Fixed points for $q > 1$

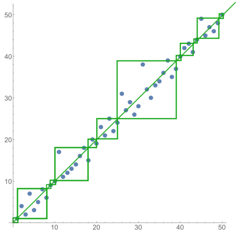


- $\mathbb{E}C_1 \to \frac{1}{2}$ as $q \downarrow 1$
- $\mathbb{E}C_1(\Pi_{2n+1}) \to 1$ as $q \to \infty$
- $\mathbb{E}C_1(\Pi_{2n}) \to 0$ as $q \to \infty$

Renewal structure if $q < 1$

Renewal structure if $q < 1$

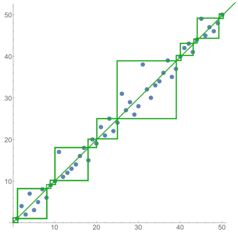
- $\Pi_n \sim \text{Mallows}(n, q)$ has a "renewal structure":



(Plot taken from [Basu+Bhatnagar 2017].)

Renewal structure if $q < 1$

- $\Pi_n \sim \text{Mallows}(n, q)$ has a "renewal structure":

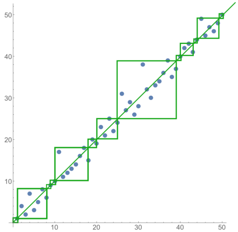


(Plot taken from [Basu+Bhatnagar 2017].)

- For large n there are many T_i such that $\Pi_n[[T_i]] = [T_i]$

Renewal structure if $q < 1$

- $\Pi_n \sim \text{Mallows}(n, q)$ has a "renewal structure":



(Plot taken from [Basu+Bhatnagar 2017].)

- For large n there are many T_i such that $\Pi_n[[T_i]] = [T_i]$
- The 'renewal blocks' are approximately i.i.d.

For $q < 1$ then

$$\frac{1}{\sqrt{n}} \left(C_1(\Pi_n) - m_1 n, \ldots, C_\ell(\Pi_n) - m_\ell n \right) \quad \xrightarrow[n \to \infty]{d} \quad \mathcal{N}_\ell(\underline{0}, P_\ell),$$

follows from results on renewal processes by Gut+Janson 1983

So why not linearly many fixed points when $q > 1$?

So why not linearly many fixed points when $q > 1$?

If r_n denotes the map $i \mapsto n+1-i$ then

$$\mathrm{inv}(r_n \circ \pi) = \binom{n}{2} - \mathrm{inv}(\pi)$$

(π flips order of i and j iff $r_n \circ \pi$ does not)

So why not linearly many fixed points when $q > 1$?

If r_n denotes the map $i \mapsto n + 1 - i$ then

$$\mathrm{inv}(r_n \circ \pi) = \binom{n}{2} - \mathrm{inv}(\pi)$$

(π flips order of i and j iff $r_n \circ \pi$ does not)
Then if $\Pi_n \sim \mathrm{Mallows}(n, q)$ we have

$$\mathbb{P}(r_n \circ \Pi_n = \pi) = \frac{q^{\binom{n}{2} - \mathrm{inv}\,\pi}}{\sum_{\sigma \in S_n} q^{\mathrm{inv}(\sigma)}} = \frac{(1/q)^{\mathrm{inv}\,\pi}}{F(n, q)}$$

So why not linearly many fixed points when $q > 1$?

If r_n denotes the map $i \mapsto n + 1 - i$ then

$$\mathrm{inv}(r_n \circ \pi) = \binom{n}{2} - \mathrm{inv}(\pi)$$

(π flips order of i and j iff $r_n \circ \pi$ does not)
Then if $\Pi_n \sim \mathrm{Mallows}(n, q)$ we have

$$\mathbb{P}(r_n \circ \Pi_n = \pi) = \frac{q^{\binom{n}{2} - \mathrm{inv}\,\pi}}{\sum_{\sigma \in S_n} q^{\mathrm{inv}(\sigma)}} = \frac{(1/q)^{\mathrm{inv}\,\pi}}{F(n,q)}$$

That is

$$r_n \circ \Pi_n \overset{d}{=} \mathrm{Mallows}(n, 1/q) \quad \text{if} \quad \Pi_n \sim \mathrm{Mallows}(n, q)$$

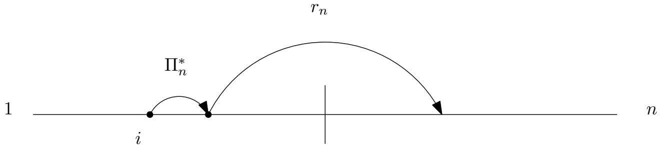
Fixed points for $q > 1$

Fixed points for $q > 1$

- Let $q > 1$. Sample first $\Pi_n^* \sim \text{Mallows}(n, 1/q)$ where $1/q < 1$

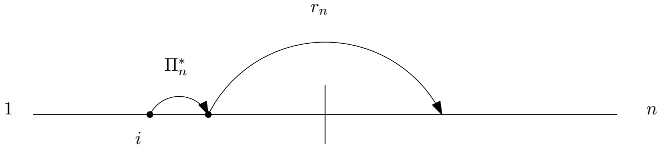
Fixed points for $q > 1$

- Let $q > 1$. Sample first $\Pi_n^* \sim \text{Mallows}(n, 1/q)$ where $1/q < 1$



Fixed points for $q > 1$

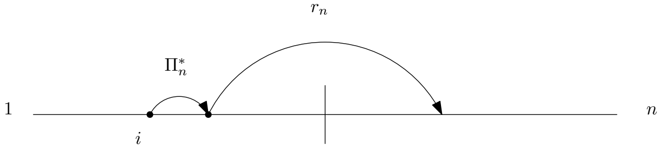
- Let $q > 1$. Sample first $\Pi_n^* \sim \text{Mallows}(n, 1/q)$ where $1/q < 1$



- Π_n^* 'close to identity' $\implies$ $\Pi_n^*(i)$ close to i

Fixed points for $q > 1$

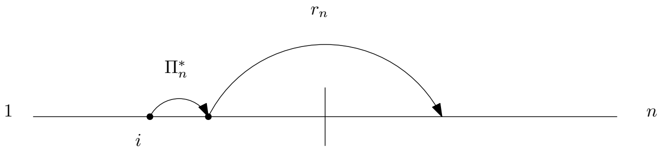
- Let $q > 1$. Sample first $\Pi_n^* \sim \text{Mallows}(n, 1/q)$ where $1/q < 1$



- Π_n^* 'close to identity' $\implies$ $\Pi_n^*(i)$ close to i
- $r_n \circ \Pi_n^*(i)$ far from i if i far from middle of $[n]$

Fixed points for $q > 1$

- Let $q > 1$. Sample first $\Pi_n^* \sim \text{Mallows}(n, 1/q)$ where $1/q < 1$



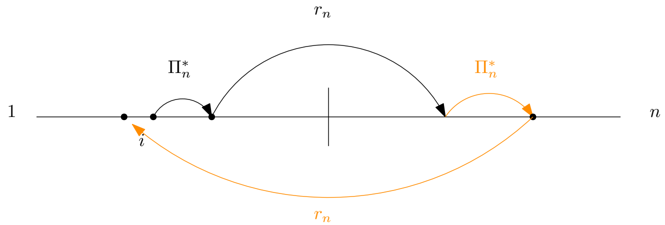
- Π_n^* 'close to identity' $\implies$ $\Pi_n^*(i)$ close to i
- $r_n \circ \Pi_n^*(i)$ far from i if i far from middle of $[n]$
- Fixed points for $q > 1$ must occur close to the mid point of $[n]$

$q > 1$ even cycles

We do not see the same obstruction for even cycles that we saw for odd cycles:

$q > 1$ even cycles

We do not see the same obstruction for even cycles that we saw for odd cycles:
Take again $\Pi_n^* \sim \text{Mallows}(n, 1/q)$ and $\Pi_n := r_n \circ \Pi_n^*$



$q \to \infty$

- $\mathbb{E}C_1(\Pi_{2n+1}) \to 1 \qquad$ as $q \to \infty$
- $\mathbb{E}C_1(\Pi_{2n}) \to 0 \qquad$ as $q \to \infty$

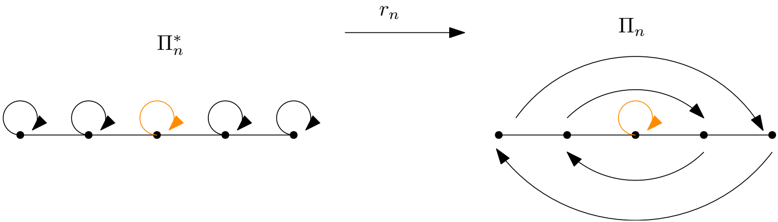
Is this reasonable?

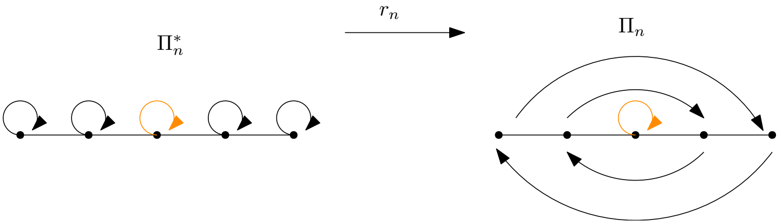
$q \to \infty$

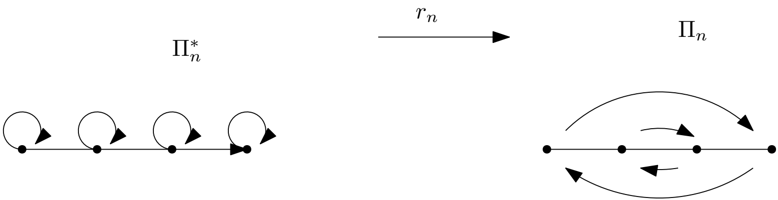
- $\mathbb{E}C_1(\Pi_{2n+1}) \to 1 \qquad$ as $q \to \infty$
- $\mathbb{E}C_1(\Pi_{2n}) \to 0 \qquad$ as $q \to \infty$

Is this reasonable?

- As $q \to \infty$ we have $1/q \to 0$ so that $\Pi_n^* \sim \text{Mallows}(n, 1/q)$ 'gets very close to the identity'


r_n
Π_n^*
Π_n


Π_n^*
r_n
Π_n


Π_n^*
r_n
Π_n

Thanks for your attention. Questions?