Shuffling Cards via One-sided Transpositions

Stephen Connor

Joint work with Michael Bate and Oliver Matheau-Raven

UNIVERSITY of York

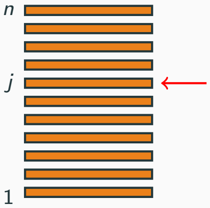
Introduction

Consider the following method for shuffling a pack of n cards:

Introduction

Consider the following method for shuffling a pack of n cards:

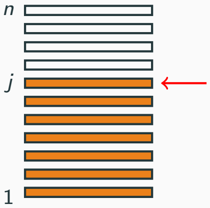
- Right hand chooses a card **uniformly at random**;



Introduction

Consider the following method for shuffling a pack of n cards:

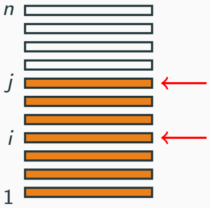
- Right hand chooses a card **uniformly at random**;
- Left hand chooses a card **uniformly** from **below** the Right hand;



Introduction

Consider the following method for shuffling a pack of n cards:

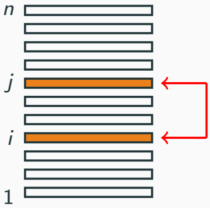
- Right hand chooses a card **uniformly at random**;
- Left hand chooses a card **uniformly** from **below** the Right hand;



Introduction

Consider the following method for shuffling a pack of n cards:

- Right hand chooses a card **uniformly at random**;
- Left hand chooses a card **uniformly** from **below** the Right hand;
- The two chosen cards are **transposed**.



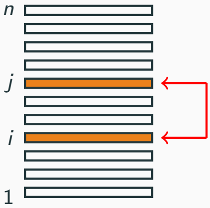
Introduction

Consider the following method for shuffling a pack of n cards:

- Right hand chooses a card **uniformly at random**;
- Left hand chooses a card **uniformly** from **below** the Right hand;
- The two chosen cards are **transposed**.

Natural question

How many shuffles does it take to "randomize" the deck?
(What is this shuffle's **mixing time**?)



Card shuffles as random walks

Most interesting card shuffles can be viewed as random walks on the symmetric group, S_n, with uniform stationary distribution π_n:

- top-to-random
- riffle shuffle
- random-to-random
- random k-cycles
- adjacent transpositions
- semi-random transpositions

Methods of bounding the rate of convergence to equilibrium include:

- coupling
- strong uniform times
- eigenanalysis
- representation theory

Card shuffles as random walks

Measure distance from equilibrium using the **total variation metric**:

$$d_n(t) = \sup_{B \subset S_n} \left(P_n^t(B) - \pi_n(B) \right) = \frac{1}{2} \sum_{\sigma \in S_n} \left| P_n^t(\sigma) - \pi_n(\sigma) \right| .$$

- takes values in $[0, 1]$
- in general, will depend upon the starting state, but not if (as here) the Markov chain is transitive.

Define the ε-**mixing time** to be

$$t_n^{\text{mix}}(\varepsilon) = \min\{ t \,:\, d_n(t) \le \varepsilon \} .$$

The cutoff phenomenon

Many shuffles exhibit somewhat surprising convergence behaviour...

Definition

The sequence of shuffles generated by $(P_n)_{n\in\mathbb{N}}$ exhibits a **cutoff** at time t_n with *window* of size w_n if $w_n = o(t_n)$ and:

$$\begin{aligned}\lim_{c\to\infty}\liminf_{n\to\infty} d_n(t_n - cw_n) &= 1 \\ \lim_{c\to\infty}\limsup_{n\to\infty} d_n(t_n + cw_n) &= 0\,.\end{aligned}$$

The cutoff phenomenon

Many shuffles exhibit somewhat surprising convergence behaviour...

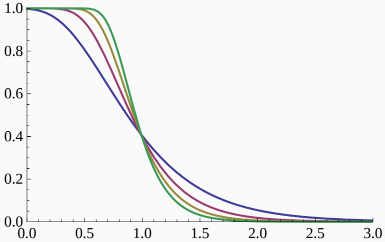
Definition

The sequence of shuffles generated by $(P_n)_{n\in\mathbb{N}}$ exhibits a **cutoff** at time t_n with *window* of size w_n if $w_n = o(t_n)$ and:

$$\begin{aligned} \lim_{c\to\infty} \liminf_{n\to\infty} d_n(t_n - cw_n) &= 1 \\ \lim_{c\to\infty} \limsup_{n\to\infty} d_n(t_n + cw_n) &= 0\,. \end{aligned}$$

Cutoff implies that

$$t_n^{\text{mix}}(\varepsilon) \sim t_n \text{ for all } \varepsilon > 0.$$



The cutoff phenomenon

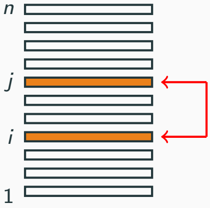
Previous results:

- top-to-random
  $\rightarrow$ cutoff at $n \log n$
- riffle shuffle
  $\rightarrow$ cutoff at $\frac{3}{2} \log_2 n$
- random-to-random
  $\rightarrow$ cutoff at $\frac{3}{4} n \log n$
- random k-cycles
  $\rightarrow$ cutoff at $\frac{n}{k} \log n$
- adjacent transpositions
  $\rightarrow$ cutoff at $\frac{n^2}{2\pi^2} \log n$
- semi-random transpositions
  $\rightarrow$ upper bound $O(n \log n)$

The one-sided transposition shuffle

Our shuffle transposes cards in positions (i, j) with probability

$$P_n(i,j) = \frac{1}{jn}, \text{ for all } 1 \leq i \leq j \leq n.$$



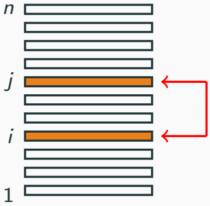
The one-sided transposition shuffle

Our shuffle transposes cards in positions (i, j) with probability

$$P_n(i,j) = \frac{1}{jn}, \text{ for all } 1 \leq i \leq j \leq n.$$

This differs significantly from previously studied shuffles which have been analysed using group representation theory:

- dependence between Left and Right hands
- generating set is entire conjugacy class of transpositions, but P_n is far from uniform on this set



Our main results

Theorem

The one-sided transposition shuffle exhibits a cutoff at $t_n = n \log n$.

Our main results

Theorem

The one-sided transposition shuffle exhibits a cutoff at $t_n = n \log n$.

Diaconis & Shahshahani (1981) showed that the *standard* transposition shuffle exhibits a cutoff at time $\frac{n}{2} \log n$.

Our main results

Theorem

The one-sided transposition shuffle exhibits a cutoff at $t_n = n \log n$.

Diaconis & Shahshahani (1981) showed that the *standard* transposition shuffle exhibits a cutoff at time $\frac{n}{2} \log n$.

By biasing the Right hand, we can recover this result as a special case of the following:

Theorem

Suppose that the Right hand chooses card j with probability proportional to j^{α}. Then we see a cutoff at time t_n:

α	$(-\infty, -1)$	-1	$(-1, 1]$	$(1, \infty)$
t_n	$\zeta(-\alpha) n^{-\alpha} \log n$	$n(\log n)^2$	$\frac{1}{1+\alpha} n \log n$	$\frac{\alpha}{1+\alpha} n \log n$

Upper bound

We use the classical ℓ^2 bound on total variation distance.

Lemma

Let the eigenvalues of P_n be $1 = \beta_1 > \beta_2 \geq \cdots \geq \beta_m > -1$. Then

$$d_n(t)^2 \leq \frac{1}{4} \sum_{i \neq 1} \beta_i^{2t} \,.$$

Our analysis is inspired by work of Dieker & Saliola (2018) and Bernstein & Nestoridi (2019) on the random-to-random shuffle.

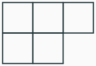
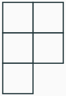
To get a handle on the eigenvalues of P_n we need to introduce the concept of **Young tableaux**.

Young tableaux

Definition

A **partition** of n is a decreasing tuple $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_l)$ such that $\sum_i \lambda_i = n$ and $\lambda_1 \geq \cdots \geq \lambda_l$. We denote this by $\lambda \vdash n$.

We may represent a partition using a **Young diagram**, e.g.

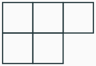
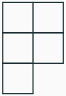
$(3, 2) =$  $(2, 2, 1) =$  $(5) =$ 

Young tableaux

Definition

A **partition** of n is a decreasing tuple $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_l)$ such that $\sum_i \lambda_i = n$ and $\lambda_1 \geq \cdots \geq \lambda_l$. We denote this by $\lambda \vdash n$.

We may represent a partition using a **Young diagram**, e.g.

$(3, 2) =$  $(2, 2, 1) =$  $(5) =$ 

A **standard Young tableau** (SYT) is an allocation of $1, \ldots, n$ to a Young diagram, such that rows and columns are increasing, e.g.

1	2	3
4	5	

1	2	4
3	5	

1	2	5
3	4	

1	3	4
2	5	

1	3	5
2	4	

The **dimension** of λ, d_λ, is the number of tableaux of shape λ.

Link to eigenvalues

Theorem

The eigenvalues of P_n are labelled by standard Young tableaux of size n, and the eigenvalue represented by a tableau of shape λ appears d_λ times.

Lemma

The eigenvalue corresponding to a tableau T is given by

$$\mathrm{eig}(T) = \frac{1}{n} \sum_{\substack{\text{boxes} \\ (i,j)\in T}} \frac{j-i+1}{T(i,j)} .$$

Example: if $T = \begin{array}{|c|c|c|}\hline 1 & 2 & 3 \\ \hline 4 & 5 & \\ \cline{1-2}\end{array}$ then $\mathrm{eig}(T) = \frac{1}{5}\left(\frac{1}{1} + \frac{2}{2} + \frac{3}{3} + \frac{0}{4} + \frac{1}{5}\right)$.

Main ideas:

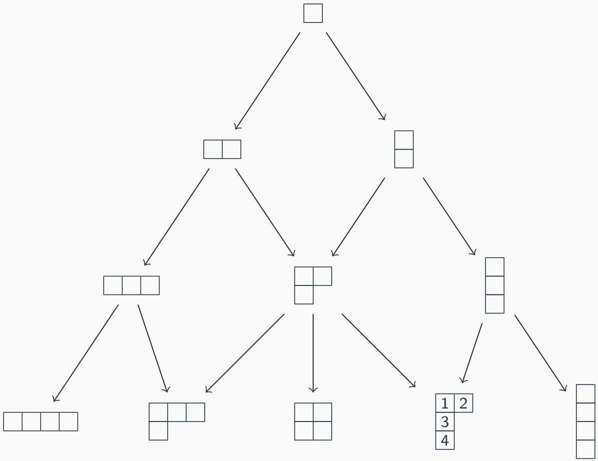
1. Natural recursive structure:
   - a deck of $(n+1)$ cards contains a deck labelled $1, \dots, n$;
   - this corresponds to a natural embedding of S_n inside S_{n+1};
   - we can obtain Young diagrams for partitions of $(n+1)$ by adding boxes to diagrams representing partitions of n.
2. Commutation relation between the operator on n cards and the operator on $(n+1)$ cards:
   - arises when we consider the difference between

     adding a card and shuffling

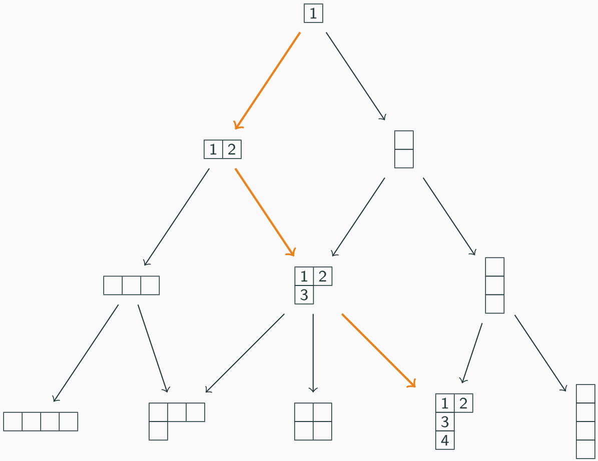
     versus

     shuffling and then adding a card.

Upshot: we may **lift** the eigenvalues of P_n to obtain those of P_{n+1} by following paths through **Young's lattice**.



Upshot: we may **lift** the eigenvalues of P_n to obtain those of P_{n+1} by following paths through **Young's lattice**.



Upper bound on the mixing time

Combining these results we obtain the bound:

$$d_n(t)^2 \leq \frac{1}{4}\sum_{i\neq 1}\beta_i^{2t} = \frac{1}{4}\sum_{\substack{\lambda\vdash n\\ \lambda\neq(n)}} d_\lambda \sum_{T\in\mathrm{SYT}(\lambda)} \mathrm{eig}(T)^{2t}$$

To establish how large t must be to make this small, we need to understand how the dimensions and eigenvalues behave for large n.

Upper bound on the mixing time

Combining these results we obtain the bound:

$$d_n(t)^2 \le \frac{1}{4}\sum_{i\neq 1}\beta_i^{2t} = \frac{1}{4}\sum_{\substack{\lambda\vdash n\\ \lambda\neq(n)}} d_\lambda \sum_{T\in\mathrm{SYT}(\lambda)} \mathrm{eig}(T)^{2t}$$

To establish how large t must be to make this small, we need to understand how the dimensions and eigenvalues behave for large n.

Theorem

For any $c > 0$,

$$\limsup_{n\to\infty} d_n(n\log n + cn) \le \sqrt{2}e^{-c}\,.$$

Two helpful types of monotonicity.

1. Fixed λ, different tableaux.

Construct $T_\lambda^{\downarrow}$ by filling boxes of λ from top to bottom, and $T_\lambda^{\rightarrow}$ by filling boxes from left to right.

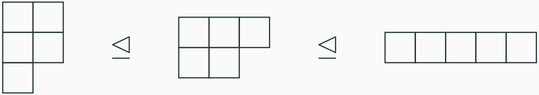
For any $T \in \mathrm{SYT}(\lambda)$,

$$\mathrm{eig}(T_\lambda^{\downarrow}) \leq \mathrm{eig}(T) \leq \mathrm{eig}(T_\lambda^{\rightarrow})$$

T	$\begin{smallmatrix}1&3&5\\2&4\end{smallmatrix}$	$\begin{smallmatrix}1&3&4\\2&5\end{smallmatrix}$	$\begin{smallmatrix}1&2&5\\3&4\end{smallmatrix}$	$\begin{smallmatrix}1&2&4\\3&5\end{smallmatrix}$	$\begin{smallmatrix}1&2&3\\4&5\end{smallmatrix}$
$\mathrm{eig}(T)$	0.503	0.523	0.57	0.59	0.64

2. Dominance ordering of partitions.

Write $\mu \trianglelefteq \lambda$ if we can form λ by moving boxes of μ up and right.



If $\mu \trianglelefteq \lambda$ then

$$\operatorname{eig}(T_\mu^{\rightarrow}) \leq \operatorname{eig}(T_\lambda^{\rightarrow}) \quad \text{and} \quad \operatorname{eig}(T_\mu^{\downarrow}) \leq \operatorname{eig}(T_\lambda^{\downarrow})$$

So it makes sense to deal with large ($\lambda_1 \geq 3n/4$) and small partitions separately, exploiting the above.

Upper bound insight: consider the partition $\lambda = (n-1, 1)$. There there are $(n-1)$ tableaux with this shape ($d_\lambda = n-1$), with the largest eigenvalue coming from the tableau $T_\lambda^{\rightarrow}$:

1	2	3	...	$n-2$	$n-1$
n					

The corresponding eigenvalue is $1 - \frac{1}{n}$, and so this partition makes a contribution to the upper bound of at most

$$d_\lambda \sum_{T \in \mathrm{SYT}(\lambda)} \mathrm{eig}(T)^{2t} \quad \leq \quad (n-1)^2 \left(1 - \frac{1}{n}\right)^{2t},$$

which at time $t = n \log n + cn$ is no greater than e^{-2c}.

Lower bound

Theorem

For any $c > 2$,

$$\liminf_{n\to\infty} d_n(n\log n - n\log\log n - cn) \geq 1 - \frac{\pi^2}{6(c-2)^2}.$$

Lower bound

Theorem

For any $c > 2$,

$$\liminf_{n\to\infty} d_n(n\log n - n\log\log n - cn) \geq 1 - \frac{\pi^2}{6(c-2)^2}\,.$$

Sketch proof

For any set of permutations $B_n \subset S_n$,

$$d_n(t) \geq P_n^t(B_n) - \pi_n(B_n)\,.$$

Focus on cards near the top of the deck, since intuitively these should take longer to mix.

Let

$$B_n = \{\rho \in S_n \,:\, \rho \text{ has} \geq 1 \text{ fixed point in top } n/\log n \text{ cards}\}.$$

Then

- $\pi_n(B_n) \leq 1/\log n$
- $P_n^t(B_n) \geq \mathbb{P}(\text{not all top } n/\log n \text{ cards touched by time } t)$

Let

$$B_n = \{\rho \in S_n \,:\, \rho \text{ has} \geq 1 \text{ fixed point in top } n/\log n \text{ cards}\}\,.$$

Then

- $\pi_n(B_n) \leq 1/\log n$
- $P_n^t(B_n) \geq \mathbb{P}\,(\text{not all top } n/\log n \text{ cards touched by time } t)$

Now estimate how many shuffles it takes for all top $n/\log n$ cards to be touched, by coupling with a counting process.

This is similar to the standard coupon-collector problem, but:

- the Right and Left hands don't "collect" cards independently
- the counting process can increment by either one or two.

□

Final remarks

- Our analysis yields an exact formula for all of the eigenvalues of the one-sided transposition shuffle;
- The results give the cutoff time and a bound on the cutoff window;
- Weighting the distribution of the Right hand shows that mixing is fastest when Right and Left hands are independent.

Going further

- Same ideas used by others to tackle different transposition shuffles (one-sided k-transpositions; generalized symmetric groups; Jucys–Murphy elements);
- More general one-sided shuffles?
- Is there a distributional approximation for the shuffle near cutoff?

References

ME Bate, SB Connor, and O Matheau-Raven.
Cutoff for a one-sided transposition shuffle.
Ann. Appl. Probab., 31(4), 2021.

M Bernstein and E Nestoridi.
Cutoff for random to random card shuffle.
Ann. Probab., 47(5), 2019.

P Diaconis and M Shahshahani.
Generating a random permutation with random transpositions.
Z. Wahrscheinlichkeit, 57(2), 1981.

AB Dieker and FV Saliola.
Spectral analysis of random-to-random Markov chains.
Adv. Math., 323, 2018.