Card-cyclic-to-random shuffling with relabeling

Johan Jonasson[*†]

March 11, 2014

Abstract

The card-cyclic-to-random shuffle is the card shuffle where the n cards are labeled $1, \ldots, n$ according to their starting positions. Then the cards are mixed by first picking card 1 from the deck and reinserting it at a uniformly random position, then repeating for card 2, then for card 3 and so on until all cards have been reinserted in this way. Then the procedure starts over again, by first picking the card with label 1 and reinserting, and so on. Morris, Ning and Peres [3] recently showed that the order of the number of shuffles needed to mix the deck in this way is $n \log n$. In the present paper, we consider a variant of this shuffle with relabeling, i.e. a shuffle that differs from the above in that after one round, i.e. after all cards have been reinserted once, we relabel the cards according to the positions in the deck that they now have. The relabeling is then repeated after each round of shuffling. It is shown that even in this case, the correct order of mixing is $n \log n$.

AMS Subject classification : 60J10
Key words and phrases: mixing time, stability of eigenvalues

1 Introduction

The subject of mixing times for Markov chains an important and exciting research field that has attracted a lot of attention in recent decades. An outstanding subclass

[*]Chalmers University of Technology and University of Gothenburg
[†]Research supported by the Knut and Alice Wallenberg Foundation

of Markov chains that has been studied extensively is card shuffling, i.e. Markov chains on the symmetric group $\mathbb{S}_n$ of permutations of n items that one can think of as the cards of a deck.

One of the early card shuffles to be studied was the random transpositions shuffle, where each step of shuffle is made by picking two cards uniformly and independently at random and then swapping them. It was shown by Diaconis and Shahshahani [2] that the mixing time of this shuffle has a sharp threshold at $\frac{1}{2}n \log n$ shuffles. It is easy to see that at least order of $n \log n$ shuffles is required, since, by the coupon collector's problem, it takes this order of shuffles until most cards have been touched at all. Closely related to the random transpositions shuffle is the top-to-random shuffle where at each step the card presently in position one is moved to a uniform random position. The sharp threshold for this shuffle is $n \log n$ and again is is easy to see that at least order of $n \log n$ steps is required for mixing, for similar reasons.

In recent years some more systematic variants of these shuffles have been proposed and analyzed. Mossel, Peres and Sinclair [4] and Saloff-Coste and Zuniga [6] analyzed the cyclic-to-random shuffle, where at time t the card presently in position t mod n is swapped with a uniformly random card. Clearly at least once per n steps, each card will be touched and one of the interesting questions about this shuffle was if $O(n)$ shuffles is also sufficient to mix the whole deck. The answer turns out to be negative; indeed the mixing time is still of order $n \log n$. Pinsky [5] later introduced the card-cyclic-to-random transpositions shuffle (CCR shuffle), where at time t the card with *label* t mod n (i.e. the card that started out in position t mod n) is moved to a uniformly random position. Again it is obvious that every card will be touched once every n steps and again one main question was if this way of systematically randomizing the cards, suffices to mix the whole deck in $O(n)$, or at least $o(n \log n)$, steps. Again the answer turns out to be negative; Morris, Ning and Peres [3] prove that $n \log n$ is still the correct order. In this paper we investigate the *card-cyclic-to-random shuffle with relabeling* (the CCRR shuffle for short). For $k = 1, 2, \ldots$ let *round* k consist of steps $kn + 1, kn + 2, \ldots, kn + n$ of shuffling. The CCRR shuffle is the shuffle that is exactly as the card-cyclic-to-random shuffle for the first round. After that however, the cards are relabeled $1, \ldots, n$ according to their positions after the first round. Next a new round of CCR shuffling is carried out according to the new labels. After that the cards are relabeled again and a new round of CCR is done, and so on. The main result of this paper is that relabeling does not help to speed up mixing either, at least not more than by a constant.

Theorem 1.1 *The mixing time of the card-cyclic-to-random transpositions with relabeling is of order* $n \log n$.

Here, the mixing time is given by

$$\tau_{\text{mix}} := \min\{t : \|\mathbb{P}(X_t \in \cdot) - \pi\|_{TV} \leq \frac{1}{4}\}$$

where $X_t \in \mathbb{S}_n$ is the state of the deck of cards after t steps of shuffling, π is the uniform distribution on $\mathbb{S}_n$ and $\|\cdot\|_{TV}$ is the total variation norm, given in general by

$$\|\mu\|_{TV} := \frac{1}{2}\sum_{x \in S} |\mu(x)| = \max\{\mu(A) : A \subset S\}$$

for a signed measure μ on a finite space S.

2 Proof of the main result

For the upper bound on τ_{mix} is suffices to note that the proof in [3] for the CCR shuffle goes through exactly as it stands there. Hence we will focus entirely on the lower bound. The idea of the proof of the lower bound draws on the idea behind Wilson's technique introduced in [8] and [9], namely to use an eigenvector of the transition matrix for the movement of a single card to build a test function. However since estimating the variance of the test function will in fact be quite simple here, we will not need Wilson's Lemma explicitly.

Because of the cyclic structure of the shuffle, the movement of a single card is not time-homogenous if we consider individual steps of the shuffle. However in terms of *rounds*, the movement of a given card is indeed a time-homogenous Markov chain. Let $A = A(n)$ denote the transition matrix of this chain on n cards. It is difficult to come up with a closed-form expression for A, but the action of A can be probabilistically described as follows. Consider a card that starts a round in position $a \in [n]$. Let us refer to cards $1, \ldots, a-1$ as white cards and to cards $a+1, \ldots, n$ as black cards. Now in a first stage the $a-1$ white cards are sequentially picked out and reinserted at independent uniform positions. During this stage a certain number of cards will be reinserted above card a in the deck whereas the others will be uniformly spread out among the black cards below card a. The cards that in this stage end up above card a will form a well-mixed layer of white cards. Note that during stage 1, card a will move gradually higher

up in the deck. (Here we say that if $i < j$, then position i is higher up than, or above, position j.)

Next, after stage 1, card a itself is picked out and reinserted at a uniformly random position $U \in [n]$; this is stage 2. In the third and final stage, the black cards are picked out and reinserted. If card a was reinserted in the white layer at the top, then card a will move gradually down in the deck during the whole of this stage, whereas if not, then stage 3 divides into the two sub-stages where in the first of these, the black cards above card a are reinserted and a moves upwards and in the second, the black cards below card a are reinserted and a moves down the deck. Even though we will not need the exact distribution of where card a ends up under this procedure, we will still need some degree of control. The following two lemmas will be useful for that.

Lemma 2.1 *Let the sequence $Y_0, Y_1, \ldots, Y_{n-y_0}$ be recursively defined by $Y_0 = y_0$ and $Y_{t+1} = Y_t + 1$ with probability Y_t/n and $Y_{t+1} = Y_t$ with probability $1 - Y_t/n$ (where these probabilities are conditionally independent of $Y_1 - Y_0, \ldots, Y_t - Y_{t-1}$ given Y_t). Then*

$$\mathbb{E}[Y_t] = \left(1 + \frac{1}{n}\right)^t y_0$$

and

$$\mathbb{V}\mathrm{ar}(Y_t) \leq \frac{y_0}{n} \sum_{j=t}^{2t} \left(1 + \frac{1}{n}\right)^j - \frac{y_0^2}{n^2} t \left(1 + \frac{1}{n}\right)^{2t}.$$

In particular, for all t,

$$\mathbb{V}\mathrm{ar}(Y_t) < \frac{11}{20} n.$$

Proof. By conditioning on Y_t we get that

$$\mathbb{E}[Y_{t+1}] = \mathbb{E}\left[\frac{Y_t}{n}(Y_t + 1) + \left(1 - \frac{Y_t}{n}\right) Y_t\right] = \mathbb{E}\left[\left(1 + \frac{1}{n}\right) Y_t\right]$$

which proves the expression for the expectation. For the variance part, write $v_t := \mathbb{V}\mathrm{ar}(Y_t)$. Then $v_0 = 0$ and recursively

$$\mathbb{V}\mathrm{ar}(Y_{t+1}) = \mathbb{E}[\mathbb{V}\mathrm{ar}(Y_{t+1}|Y_t)] + \mathbb{V}\mathrm{ar}(\mathbb{E}[Y_{t+1}|Y_t]).$$

By definition of the Y_t's, $\mathbb{V}\mathrm{ar}(Y_{t+1}|Y_t) = Y_t(n-Y_t)/n^2$ and by the above $\mathbb{E}[Y_{t+1}|Y_t] = (1+1/n)Y_t$. Hence

$$\begin{aligned}\mathbb{E}[\mathbb{V}\mathrm{ar}(Y_{t+1}|Y_t)] &= \frac{\mathbb{E}[Y_t]}{n} - \frac{\mathbb{E}[Y_t^2]}{n^2} \\ &= \frac{(1+1/n)^t y_0}{n} - \frac{1}{n}\left(v_t + y_0^2\left(1+\frac{1}{n}\right)^{2t}\right).\end{aligned}$$

Adding the second term and writing $c := 1 + 1/n$ gives

$$\begin{aligned}v_{t+1} &= \left(c^2 - \frac{1}{n}\right)v_t + \frac{c^t y_0}{n}\left(1 - \frac{c^t y_0}{n}\right) \\ &< c^2 v_t + \frac{1}{n}c^t y_0 - \frac{1}{n^2}c^{2t}y_0^2.\end{aligned}$$

This recursion is readily solved and gives

$$\begin{aligned}v_t &< \frac{y_0}{n}\sum_{j=t}^{2t} c^j - \frac{y_0^2}{n^2}tc^{2t} \\ &< ne\alpha\left(e^{1-2\alpha} - e^{-\alpha} - \alpha^2(1-\alpha)e^{1-2\alpha}\right) \\ &< \frac{11}{20}n,\end{aligned}$$

where $\alpha = y_0/n$, the first inequality follows from that $t \le n - y_0 = n(1-\alpha)$ and the last inequality from standard optimization over α.

□

Lemma 2.2 *Let $X \in L^2(\mathbb{R})$ be a random variable and $f : \mathbb{R} \to \mathbb{R}$ be contractive, i.e. $|f(x) - f(y)| \le |x - y|$ for all $x, y \in \mathbb{R}$. Then*

$$\mathbb{V}\mathrm{ar}(f(X)) \le \mathbb{V}\mathrm{ar}(X).$$

Proof. Let X_1 and X_2 be two independent copies of X. Then

$$\begin{aligned}\mathbb{V}\mathrm{ar}(X) &= \frac{1}{2}\mathbb{V}\mathrm{ar}(X_1 - X_2) \\ &= \frac{1}{2}\mathbb{E}[|X_1 - X_2|^2] \\ &\ge \frac{1}{2}\mathbb{E}[|f(X_1) - f(X_2)|^2] \\ &= \mathbb{V}\mathrm{ar}(f(X)).\end{aligned}$$

□

Let Z be the position that card a ends up in after one round of shuffling. We want to estimate the expectation and variance of Z. Let W be the number of white cards that go to the top layer of white cards in stage 1. We will start by estimating the variance of Z given $W = w$ and $U = u$. If $u \leq w$, so that stage 2 moves card a to the top white layer, then by Lemma 2.1

$$\mathbb{E}[Z|U = u, W = w] = \left(1 + \frac{1}{n}\right)^{n-a} u$$

and

$$\mathbb{V}\mathrm{ar}(Z|U = u, W = w) < \frac{11}{20}n.$$

The case $u > w$ takes some more work. In order to not overly burden the notation, we suppress until further notice the conditioning on $\{W = w, U = u\}$. Let S be the number of black cards in positions $w + 1, \ldots, u - 1$; these are the black cards that get reinserted in the first part of stage 3. Let N be the number of cards below card a after these S black cards have been reinserted. Note that when the first part of stage 3 starts, then card a is in u and at that point, the number of black cards below a is $n - a - S$. Hence by Lemma 2.1,

$$\mathbb{E}[Z|N, S] = \left(1 + \frac{1}{n}\right)^{n-a-S} (n - N)$$

and

$$\mathbb{E}[N|S] = \left(1 + \frac{1}{n}\right)^{S} (n - u).$$

Hence

$$\begin{aligned} \mathbb{E}[Z|S] &= (n - \mathbb{E}[N|S]) \left(1 + \frac{1}{n}\right)^{n-a-S} \\ &= n \left(1 + \frac{1}{n}\right)^{n-a-S} - \left(1 + \frac{1}{n}\right)^{n-a} (n - u). \end{aligned}$$

It follows that

$$\mathbb{E}[Z] = n\mathbb{E}\left[\left(1 + \frac{1}{n}\right)^{n-a-S}\right] - \left(1 + \frac{1}{n}\right)^{n-a} (n - u). \tag{1}$$

We also get that

$$\begin{aligned}\mathbb{V}\mathrm{ar}(Z|S) &= \mathbb{E}\left[\mathbb{V}\mathrm{ar}(Z|N,S)|S\right] + \mathbb{V}\mathrm{ar}\left(\mathbb{E}[Z|N,S]|S\right) \\ &< \frac{11}{20}n + \mathbb{V}\mathrm{ar}\left(\left(1+\frac{1}{n}\right)^{n-a-S}(n-N)|S\right) \\ &= \frac{11}{20}n + \left(1+\frac{1}{n}\right)^{2(n-a-S)}\frac{11}{20}n.\end{aligned}$$

Therefore

$$\begin{aligned}\mathbb{V}\mathrm{ar}(Z) &= \mathbb{E}\left[\mathbb{V}\mathrm{ar}(Z|S)\right] + \mathbb{V}\mathrm{ar}\left(\mathbb{E}[Z|S]\right) \\ &< \frac{11}{20}(1+e^2)n + \mathbb{V}\mathrm{ar}\left(n\left(1+\frac{1}{n}\right)^{n-a-S}\right) \\ &\leq \frac{11}{20}(1+e^2)n + \mathbb{V}\mathrm{ar}(S) < \left(\frac{11}{20}+\frac{4}{5}e^2\right)n\end{aligned}$$

by Lemma 2.2, since the map $S \to (n/e)(1+1/n)^{n-a-S}$ is contractive and S is hypergeometric with variance at most $n/4$. Let us now bring back the conditioning on W and U into the notation. What we have just shown is among other things, that $\mathbb{V}\mathrm{ar}(Z|W,U) < Cn$ with $C := 11/20 + 4e^2/5$. Thus $\mathbb{V}\mathrm{ar}(Z|U) = \mathbb{E}[\mathbb{V}\mathrm{ar}(Z|W,U)|U] + \mathbb{V}\mathrm{ar}(\mathbb{E}[Z|W,U]|U) < Cn + \mathbb{V}\mathrm{ar}(\mathbb{E}[Z|W,U]|U)$. However, by (2)

$$\begin{aligned}&\left|\mathbb{E}\left[Z|W=w,U\right] - E\left[Z|W=w-1,U\right]\right| \\ &= \left|\mathbb{E}\left[n\left(1+\frac{1}{n}\right)^{n-a-S}\Big|W=w,U\right] - \mathbb{E}\left[n\left(1+\frac{1}{n}\right)^{n-a-S}\Big|W=w-1,U\right]\right| \\ &\leq \mathbb{E}\left[n\left(1+\frac{1}{n}\right)^{n-a-S}\left(\left(1+\frac{1}{n}\right)-1\right)\Big|W=w,U\right] \leq e,\end{aligned}$$

where the first inequality uses that the conditional distributions of S given $W=w$ and $W=w-1$ respectively, can easily be coupled so that the realizations do not differ by more than 1. It now follows that

$$\mathbb{V}\mathrm{ar}\left(E[Z|W,U]\Big|U\right) \leq e^2\mathbb{V}\mathrm{ar}(W) < \frac{11}{20}e^2 n$$

by Lemma 2.1. Hence

$$\mathbb{V}\mathrm{ar}(Z|U) < \left(\frac{11+27e^2}{20}\right)n.$$

This allows us to write $Z = \mathbb{E}[Z|U] + D$, where $D = Z - \mathbb{E}[Z|U]$ has

$$\mathbb{V}\mathrm{ar}(D) = \mathbb{E}\left[\mathbb{V}\mathrm{ar}(D|U)\right] = \mathbb{E}\left[\mathbb{V}\mathrm{ar}(Z|U)\right] < 11n. \tag{2}$$

Let us now write $1/n, 2/n \ldots, 1$ instead of $1, 2, \ldots, n$ for the positions in the deck and let $n \to \infty$. Then, by the above, the position of a card starting from $a \in [0,1]$ converges in distribution to that of $G_a(U)$ where U is uniform on $[0,1]$ and

$$G_a(u) = \begin{cases} e^{1-a}u, & u \le u_0(a) := 1 - (1-a)e^a \\ e^{e^{-a}(1-u)} - (1-u)e^{1-a}, & u > u_0(a) \end{cases} \tag{3}$$

Also, by (1), Lemma 2.1, the fact that S has variance at most $n/4$ and a Maclaurin expansion of e^{-x}, give that

$$\mathbb{E}[n^{-1}Z|U = u] \in \left(1 \pm \frac{\sqrt{11}}{n}\right) G_a(u). \tag{4}$$

Recall that we write $A = A(n)$ for the transition matrix of the movement of a card under one round of CCRR. Write $B = B(n) = [b_{ij}]$ for the transition matrix of a card that moves according to $G_a(x)$. More precisely, let U be uniform on $[0,1]$ and let b_{ij} be the probability that $G_a(U) \in (j - 1/n, j)$, $i, j \in \{1/n, 2/n, \ldots, 1\}$, where a is chosen uniformly at random in $(i-1, i)$. The next lemma states that the matrix B has a nontrivial eigenvalue bounded away from 0.

Lemma 2.3 *The transition matrix $B(n)$ has a (possibly complex) second eigenvalue λ such that $|\lambda| > 0.08$.*

Remark. Matlab evaluations up to $n = 10^5$ strongly suggest that the second eigenvalue is real and in the interval $(0.21, 0.22)$.

Proof. Write $B = S + D$ where S is the symmetric matrix $(B + B^T)/2$ and D is the skew-symmetric matrix $(B - B^T)/2$. We claim the following.

Lemma 2.4 *The second largest eigenvalue of S is at least 0.21*

Lemma 2.5 *The (purely imaginary) eigenvalues λ of D satisfy $|\lambda| < 0.13$. In particular, $\|D\|_{2\to 2} < 0.13$.*

It is well known that eigenvalues are stable in the sense that if C is normal (i.e. $CC^T = C^TC$), in particular if C is symmetric, with an eigenvalue λ_0 and E is a matrix with $\|E\|_{2\to 2} = \epsilon$, then $B + E$ has an eigenvalue λ such that $|\lambda - \lambda_0| \leq \epsilon$. See e.g. [7] for an elementary proof of this fact. This means that Lemma 2.3 follows immediately from Lemmas 2.4 and 2.5. □

Proof of Lemma 2.4 and Lemma 2.5. In the proof of these lemmas, it will be convenient to use the following convention: when a function f is defined on $\{1/n, 2/n, \ldots, 1\}$ we will identify it with its extension to $[0, 1]$ defined by $f(a) = f(n^{-1}\lceil na\rceil)$. By this convention, $\|f\|_2$ of the unextended n-dimensional vector f is $\sqrt{n}$ times $\|f\|_2$ of the extended f as a function in $L^2[0, 1]$.

Let us first study S. That (λ, ϕ) is an eigenvalue/eigenvector pair for S means that $\mathbb{E}[\phi(X_1)|X_0 = a] = \lambda\phi(a)$ for all $a = 1/n, 2/n, \ldots, 1$, where $X_1 = X_1(n)$ is the position of a card after one move according to S, starting from X_0. Write $Y = Y(n)$ for a random variable distributed as the position after one move according to $B(n)$ and let $Y^*(n)$ be distributed according to the position after one step of $B(n)^T$. (Note that B is doubly stochastic, so that B^T is the transition matrix of the reversed CCRR.) Thus X_1 is the (uniform) convex combination of Y and Y^*. The idea now is to find (κ, ψ) close enough to an eigenvalue/eigenvector pair to allow us to draw the desired conclusion. We do this with the aid of Matlab. Some more details on the Matlab computations, in particular the code, can be found in the appendix.

We use Matlab to calculate the eigenvalue $\kappa = 0.2293...$ and corresponding eigenvector χ with $n = 10^4$, scaled so that $\|\chi\|_2 = 1$. Now let $n = 10^5$ and extend χ to ψ, the linear interpolation of (a slightly smoothed out version (see the appendix) of) χ. Then we find that

$$\|\mathbb{E}[\psi(X_1(n))|X_0(n) = \cdot] - \kappa\psi(\cdot))\|_2 < 0.0012. \tag{5}$$

To arrive at the desired conclusion, a bound on the norm of the difference between $\mathbb{E}[\psi(X_1(m))|X_0 = \cdot]$ and $\mathbb{E}[\psi(X_1(n))|X_0 = \cdot]$ for $m > n = 10^5$ must also be established. This will be done by investigating the difference between the distributions of $G_a(U)$ and $G_{a+1/m}(U)$. Note that the distribution function of $G_a(U)$ is G_a^{-1}. We claim that $|G_a^{-1}(x) - G_{a+1/m}^{-1}(x)|$ is maximized when either $x = x_0 := G_{a+1/m}(u_0(a + 1/m))$ or $x = G_a(u_0(a))$ (recall that $u_0(a) = 1 - (1 - a)e^a$ is the breakpoint in the expression for $G_a(u)$).

Write $b := a+1/m$. To prove the claim, it suffices to show that $(d/dx)(G_a^{-1}(x) - G_b^{-1}(x))$ is negative for $x < G_a(u_0(a))$ and $x > G_b(u_0(b))$ and positive for $G_a(u_0(a)) < x < G_b(u_0(b))$, $m > 10^5$. This is equivalent to showing that

$G'_a(G_a^{-1}(x)) - G'_b(G_b^{-1}(x))$ is positive for $x < G_a(u_0(a))$ and $x > G_b(u_0(b))$ and negative for x between the two bounds. For $x < G_a(u_0(a))$, the difference of the derivatives is constantly $e^{1-a} - e^{1-a-1/m} > 0$. When $G_a(u_0(a)) < x < G_b(u_0(b))$, $G'_b(G_b^{-1}(x)) = e^{1-b}$, whereas $G'_a(G_a^{-1}(x)) \leq e^{1-a} - 1$, which is obviously smaller.

For $x > G_b(u_0(b))$, let $z = G_a^{-1}(x)$ and $y = G_b^{-1}(x)$. Then, since $G''_a > 0$ so that G'_a is increasing, we have $z - y \geq (G_b(z) - G_a(z))/G'_b(z)$. Bounding this from below gives

$$\begin{aligned}
\frac{G_b(z) - G_a(z)}{G_b(z)} &= \frac{e^{e^{-b}(1-z)} - e^{1-b} - e^{e^{-a}(1-z)} + e^{1-a}}{e^{1-b} - e^{e^{-b}(1-z)}} \\
&\geq \frac{m^{-1}e^{1-b}(1-z) + e^{e^{-b}(1-z)}(1 - e^{m^{-1}e^{-b}(1-z)})}{e^{1-b} - e^{e^{-b}(1-z)}} \\
&\geq (1-z)m^{-1}.
\end{aligned}$$

Therefore

$$\begin{aligned}
G'_a(z) - G'_b(y) &= e^{1-a} - e^{1-a-1/m} - e^{-a}e^{e^{-a}(1-z)(1-1/m)} \\
&\quad + e^{-a-1/m}e^{e^{-a-1/m}(1-z)} \\
&\geq e^{-a}\Big(em^{-1} - e^{e^{-a}(1-1/m)(1-z)} \\
&\quad + (1-m^{-1})e^{e^{-a}(1-1/m)(1-z)} - \frac{e}{2}m^{-1}\Big) \\
&\geq e^{-a}m^{-1}\left(e - e^{e^{-a}(1-1/m)(1-z)} - \frac{e}{2}m^{-1}\right) \\
&\geq e^{-a}m^{-1}\left(e - e^{1-a} - \frac{e}{2}m^{-1}\right) \geq 0,
\end{aligned}$$

since $a \geq 1/m$. This proves the claim.

Now, we have that

$$\begin{aligned}
0 &\leq G_{a+1/m}^{-1}(G_a(u_0(a))) - G_a^{-1}(G_a(u_0(a))) \\
&= (e^{a+1/m-1} - e^{a-1})G_a(u_0(a)) \\
&\leq m^{-1}.
\end{aligned}$$

Also

$$0 \leq G_a^{-1}(x_0) - G_{a+1/m}^{-1}(x_0) = G_a^{-1}(x_0) - u_0(a + m^{-1}).$$

Now $G_a(u_0(a)) = e^{1-a} - e(1-a)$ from which it readily follows that $G_{a+1/m}(u_0(a+1/m)) - G_a(u_0(a)) \leq 2ea/m$. The derivative of G_a is given by

$$G'_a(u) = \begin{cases} e^{1-a}, & u < u_0(a) \\ e^{1-a} - e^{-a}e^{e^{-a}(1-u)}, & u > u_0(a) \end{cases}$$

This is minimized for $u = u_0(a)$ and then becomes $e^{1-a} - e^{1-2a} > a/2$. Since $(d/dx)G_a^{-1}(x) = 1/G_a'(G_a^{-1}(x))$, it follows that

$$G_a^{-1}(x_0) \leq u_0(a) + 4em^{-1}.$$

Since $u_0(a + 1/m) > u_0(a)$, it follows that

$$G_a^{-1}(x_0) - G_{a+1/m}^{-1}(x_0) < 4em^{-1}.$$

Hence, for $a \in \{1/n, 2/n, \ldots, 1 - 1/n\}$ and all l,

$$\begin{aligned} |\sum_{j=1}^{l} b_{a+1/m,j} - \sum_{j=1}^{l} b_{a,j}| &\leq |G_{a+1/m}^{-1}(l) - G_a^{-1}(l)| \\ &< 4em^{-1} < 11m^{-1}. \end{aligned}$$

This means that the total variation distance between the distributions of two cards making a move according to B, starting from i and $i + 1/m$ respectively, is bounded by $11/m$. Writing $Y_a(m)$ for a random variable distributed according to the position after one round of CCRR for a card that starts in position a, a consequence of this is that one can construct a coupling of $Y_a(m)$ and $Y_{a+1/m}(m)$ such that $\mathbb{P}(Y_a(m) \neq Y_{a+1/m}(m)) < 11/m$. More generally, for $k < m$, one can couple so that $\mathbb{P}(Y_a(m) \neq Y_{a+k/m}(m)) < 11k/m$. This entails, with $\hat{\psi} = \max_x \psi(x) - \min_x \psi(x) < 4.5$, that

$$|\mathbb{E}[\psi(Y_a(m))] - \mathbb{E}[\psi(Y_{a+k/m}(m))]| < 50km^{-1}. \tag{6}$$

Next we give a corresponding bound for B^T. Note that $G_a'(j) = G_a'(j + 1/m)$ for a such that $j + 1/m < u_0(a)$ and that when $j > u_0(a)$, $G_a'(j) < G_a'(j + 1/m)$, whereas when $j < u_0(a) < j+1/m$, then $G_a'(j) > G_a' j + 1/m$. Hence $b_{i,j+1/m} - b_{i,j}$ is zero for $u_0(i) > b + 1/m$, negative for $u_0(i) < b$ and positive for the i's such that $j < u_0(i) < j + 1/m$. Hence the sum $\sum_{i:j<u_0(i)<j+1/m}(b_{i,j+1/m} - b_{i,j})$ gives the total variation distance between the distributions of two cards making one move according to B^T and starting from j and $j + 1/m$ respectively. The number of i's in the sum equals $m(u_0^{-1}(j + 1/m) - u_0^{-1}(j))$ and

$$b_{i,j+1/m} \leq 1 \wedge \frac{1}{mG_i'(u_0(i))} < 1 \wedge \frac{2}{mi}.$$

Each of the i's correspond to a $j < u_0(i)$, i.e. $i > u_0^{-1}(j)$. Hence the total variation distance is bounded by $2(u_0^{-1}(j + 1/m) - u_0^{-1}(j))/u_0^{-1}(j)$. Now $u_0'(a) = ae^a$, so

$$u_0^{-1}(j + 1/m) - u_0^{-1}(j) < \frac{e^{-i}}{mi} \leq \frac{1}{mi} < \frac{1}{u_0^{-1}(j)}.$$

Hence the total variation bound becomes $1 \wedge 2/mu_0^{-1}(j)^2$. Since $u_0(a) = 1-(1-a)e^a \leq a^2$, we have $u_0^{-1}(j) \geq \sqrt{j}$, so the bound is no larger than $1 \wedge 2/mj$. Now, in analogy with the above, let $Y_a^*(m)$ be distributed as the position of a card after one amove according to B^T, started from a. Then one can construct a coupling such that $\mathbb{P}(Y_a^*(m) \neq Y_{a+k/m}^*(m)) < 1 \wedge 2k/ma$ and hence

$$|\mathbb{E}[\psi(Y_a^*(m))] - \mathbb{E}[\psi(Y_{a+k/m}^*(m))]| < \hat{\psi}\left(1 \wedge \frac{2k}{ma}\right) < 4.5\left(1 \wedge \frac{2k}{ma}\right). \quad (7)$$

Let us now compare $\mathbb{E}[\psi(Y_a(m)]$ with $\mathbb{E}[\psi(Y_a(n))]$. For convenience, assume that $n|m$ and set $m = nl$. For $a = k/n - r/m$, $0 \leq r \leq l-1$, we have that $\mathbb{E}[\psi(Y_a(n)] = \mathbb{E}[\psi(Y_{a_0}(n))]$, where $a_0 := n^{-1}\lceil na \rceil$. Using (6) shows that

$$|\mathbb{E}[\psi(Y_a(m))] - \mathbb{E}[\psi(Y_{a_0}(m))| < \frac{50}{n} = 0.0005.$$

From our Matlab calculations, we get $\max_x |\psi'(x)| < 100$. Then it is clear that

$$|\mathbb{E}[\psi(Y_{a_0}(m))] - \mathbb{E}[Y_{a_0}(n)]| \leq \frac{100}{n} = 0.001.$$

Hence

$$\|\mathbb{E}[\psi(Y_\cdot(m))] - \mathbb{E}[\psi(Y_\cdot(n))]\|_2 < 0.0015. \quad (8)$$

Analogously for comparing $\mathbb{E}[\psi(Y_a^*(m)]$ with $\mathbb{E}[\psi(Y_a^*(n))]$, use (7) to get

$$|\mathbb{E}[\psi(Y_a^*(m))] - \mathbb{E}[\psi(Y_{a_0}^*(m))]| < 4.5\left(1 \wedge \frac{2r}{(k-1)l}\right)$$

and hence some straightforward calculations give, using (7), that $\sum_1^l k^2 \leq l^3/3$ and that $\sum_1^\infty 1/k^2 = \pi^2/6$,

$$\begin{aligned} \|\mathbb{E}[\psi(Y_\cdot^*(m))] - \mathbb{E}[\psi(Y_\cdot^*(n))]\|_2 &< 0.001 + \frac{2 \cdot 4.5}{\sqrt{n}}\sqrt{1 + \frac{5}{24} + \frac{1}{3}\left(\frac{\pi^2}{6} - 1\right)} \\ &< 0.001 + \frac{10}{\sqrt{n}} < 0.033. \end{aligned}$$

Since X_1 is the convex combination of Y and Y^*, it follows from (8) that

$$\mathbb{E}[\psi(X_\cdot(m))] - \mathbb{E}[\psi(X_\cdot(n))]\|_2 < 0.018. \quad (9)$$

Combining (9) with (5), we find that

$$\|\mathbb{E}[\psi(X_\cdot(m)] - \psi\|^2 < 0.0192$$

for all $m \geq 10^5$. From this it follows that S has an eigenvalue λ with $\lambda > \kappa - 0.0192 > 0.21$ as desired.

Next we prove Lemma 2.5 in a completely analogous way. We have that $\lambda, \phi)$ is an eigenvalue/eigenvector pair for C if $(1/2)(\mathbb{E}[\phi(Y)|X_0 = i] - \mathbb{E}[\phi(Y^*)|X_0 = i]) = \lambda(i)$ for all i, where Y and Y^* are, as above, random variables distributed according one step of B and B^T respectively, starting from X_0. Again we take $n = 10^5$ and use Matlab to get κ and ψ close to an eigenvalue and eigenvector respectively. It turns out that $\kappa = 0.0793...i$, so $|\kappa| < 0.08$ and we get $\hat{\psi} < 5$. In terms of variability however, this case turns out to be less well behaved. We get $\max_x |\psi'(x)| < 400$ and $\|\frac{1}{2}(\mathbb{E}[\psi(Y(n)|X_0 = \cdot] - \mathbb{E}[\psi(Y^*(n))|X_0 = \cdot]) - \kappa\psi(\cdot)\|_2 < 0.017$. Then the above calculations now give

$$\left\| \frac{1}{2} \left(\mathbb{E}[\psi(Y(m)|X_0 = \cdot] - \mathbb{E}[\psi(Y^*(m))|X_0 = \cdot] \right) - \kappa\psi(\cdot) \right\|_2 < 0.047 < 0.05.$$

The desired result follows. □

For the remainder of the paper, in the light of Lemma 2.3, we fix λ to be the eigenvalue of B with the second largest modulus. Let ϕ be an eigenvector corresponding to λ with $\|\phi\|_2 = 1$. The next lemma, which we extract from the proof of Lemma 2.4, will be useful in order to show that $\phi(i)$ and $\phi(j)$ cannot differ much if i and j are close.

Lemma 2.6 *Let $f : [n] \to \mathbb{C}$ and for $i \in [n]$, let X_i be a random variable distributed according to the law the position of card i after one move according to B. Then for all $i \in [n-1]$,*

$$|\mathbb{E}[f(X_{i+1})] - \mathbb{E}[f(X_i)]| \leq \frac{22}{n}\|f\|_\infty.$$

Lemma 2.7 *For the eigenvector ϕ, of B, we have*

$$\|\phi\|_1 \geq c_1 n^{4/9},$$

$$\|\phi\|_\infty \leq c_2 n^{-4/9}$$

and

$$|\phi(i+1) - \phi(i)| \leq c_3 n^{-13/9}$$

for constants c_1, c_2 and c_3 independent of n.

Proof. Let, as in the previous proof, X_i be distributed as the position of card i after one move according to B. By definition of eigenvalue/eigenvector, $\mathbb{E}\phi(X_i) = \lambda\phi(i)$. Hence by Lemma 2.6,

$$|\phi(i) - \phi(i+1)| \leq \frac{22|\lambda|^{-1}}{n} < \frac{275}{n}$$

since $\|\phi\|_\infty \leq 1$. Write $\|\phi\|_\infty = C_1 n^{-a}$ for a large constant C_1. Since $|\phi(i) - \phi_{i+1}| < 275/n$, it follows that

$$1 \geq \|\phi\|_2^2 > \frac{275^2}{n^2} \sum_1^{C_1 n^{1-a}/10} \frac{1}{j^2} > n^{1-3a}$$

provided that C_1 is large enough, which entails that $a \geq 1/3$. This however means that $\|\phi\|_\infty \leq C_1 n^{-1/3}$ so that the conclusion from Lemma 2.6 above can be strengthened to

$$|\phi(i) - \phi(i+1)| < \frac{275 C_1}{n^{4/3}}.$$

Now writing $\|\phi\|_\infty = C_2 n^{-b}$ for some large constant C_2 gives that

$$1 \geq \|\phi\|_2^2 > \frac{275^2 C_1^2}{n^2} \sum_{j=1}^{C_2 n^{1-a}/10C_1} \frac{1}{j^2} > n^{4/3-3b}$$

for large enough C_2, so that $b \leq 4/9$. This shows that $\|\phi\|_\infty = O(n^{-4/9})$. Once again bootstrapping the bound on $|\phi(i) - \phi(i+1)|$ gives an upper bound of $O(n^{-13/9})$. Since $\|\phi\|_\infty \geq n^{-1/2}$, it follows that

$$\|\phi\|_1 = \Omega(n^{-13/9}) \sum_{j=1}^{\Omega(n^{17/18})} \frac{1}{j} = \Omega(n^{4/9}).$$

□

Let $S_t := \sum_{i:\phi(i)>0} \phi(X_i^t)$ where X_i^t is the position of card i after t rounds of CCRR. In accordance with the the above, we write for simplicity X_i for X_i^1. The random variable S_t is going to be the test statistic used to verify that order $\log n$ rounds are necessary for the deck to mix. Let X_∞ be the deck at stationarity (i.e. uniform on $\mathbb{S}_n$) and let $S_\infty = \sum_{\phi(i)>0} \phi(X_t^\infty)$. Note that $S_0 = \Omega(n^{4/9})$ by Lemma 2.7. Since the cards now move according to A and not B, ϕ is not quite an

eigenvector. However, letting Y_t^i be the position of a card after t steps according to B and coupling X_1^i and Y_1^i by using the same uniform random variable for updating, (i.e. we use $n^{-1}\lceil nU\rceil$ for X_t^i) (2) gives that $|\mathbb{E}[X_1^i] - \mathbb{E}[Y_1^i]| = O(n^{1/2})$. Hence by Lemma 2.7,

$$|\mathbb{E}[\phi(X_1^i)|X_0^i = a] - \mathbb{E}[\phi(Y_t^i)|Y_0^i = a]| = O(n^{-13/9}n^{1/2}) = O(n^{-17/18}).$$

Hence summing over i with $X_0^i > 0$ and using the triangle inequality gives

$$|\mathbb{E}[S_1|S_0] - \lambda S_0| < C_1 n^{1/18}.$$

for a sufficiently large constant C_1. A straightforward recursion gives, using (12),

$$\mathbb{E}[S_t] \geq \lambda^t S_0 - \left(\sum_{r=1}^{t-1} \lambda^r\right) C_1 n^{1/18} > C_2 \lambda^t n^{4/9} - C_3 n^{1/18}. \tag{10}$$

We also need to bound the variance of S_t. Let $f_i(U_i) = \mathbb{E}[X_i|U_i]$, where U_i is the position where card i is reinserted in round 1. Then we can write $X_i = f_i(U_i) + \epsilon_i$, where $\mathbb{E}[\epsilon_i^2] = O(n)$ by (2). Hence $\phi(X_i) = \phi(f_i(U_i)) + \delta_i$, where the variance of δ_i is order $n \cdot (n^{-13/9})^2 = n^{-17/9}$ by Lemma 2.7.

Now observe that $f_i(U_i)$ and $f_j(U_j)$ are independent and $\mathbb{C}\mathrm{ov}(\delta_i, \delta_j) = O(n^{-17/9})$. Also,

$$\mathbb{C}\mathrm{ov}(\phi(f_i(U_i)), \delta_j) = \mathbb{C}\mathrm{ov}(\phi(f_i(U_i)), \mathbb{E}[\delta_j|U_i]) = O(n^{-4/9}n^{-13/9}) = O(n^{-17/9}).$$

Summing up, we get

$$\mathbb{C}\mathrm{ov}(\phi(X_i), \phi(X_j)) = O(n^{-17/9}) \tag{11}$$

from which it follows that

$$\mathbb{V}\mathrm{ar}(S_1) = O(n^{1/9}). \tag{12}$$

From the considerations leading up to (10), we can write $\mathbb{E}[S_{t+1}|S_t] = \lambda S_t + Z$ for a random variable Z, which is function of S_t such that $|Z| \leq n^{1/18}$. Hence

$$\begin{aligned} \mathbb{V}\mathrm{ar}(\mathbb{E}[S_{t+1}|S_t]) &\leq \lambda^2 \mathbb{V}\mathrm{ar}(S_t) + 2\lambda \mathbb{E}[S_t] n^{1/18} + n^{1/9} \\ &< \lambda^2 \mathbb{V}\mathrm{ar}(S_t) + 2\lambda n^{11/18} + n^{1/9}. \end{aligned}$$

By (12), $R := \max_x \mathbb{E}[\mathbb{V}\mathrm{ar}(S_{t+1}|X_t = x)] = O(n^{1/9})$. Hence, with $v_t := \mathbb{V}\mathrm{ar}(S_t)$, we have the recursive inequality,

$$v_{t+1} \leq R + 2\lambda n^{11/18} + n^{1/9}\lambda^2 v_t$$

with $v_0 = 0$. It follows that

$$v_t \leq (R + O(n^{1/9})) \sum_{j=0}^{t} \lambda^{2j} + n^{11/18} \lambda \sum_{j=0}^{t} \lambda^{2j},$$

so for all t,

$$\mathbb{V}\mathrm{ar}(S_t) O(n^{11/18}).$$

By continuity we also get $\mathbb{V}\mathrm{ar}(S_\infty) = O(n^{11/18})$.

Finally let $\tau := \lfloor \log n / 9 \log \lambda^{-1} \rfloor$. Then by (10), $\mathbb{E}[S_\tau] \geq c_1 n^{1/3}$, so by Chebyshev's inequality,

$$\mathbb{P}(S_\tau \leq n^{23/72}) \to 0$$

as $n \to \infty$, whereas

$$\mathbb{P}(S_\infty \leq n^{23/72}) \to 1.$$

This proves the main theorem.

References

[1] D. Aldous and P. Diaconis (1986), Shuffling cards and stopping times, *Amer. Math. Monthly* **93** 333-348.

[2] Diaconis, P. and Shahshahani, M.; Generating a random permutation with random transpositions, *Z. Wahrsch. Verw. Gebeite* **57** (1981), 159-179.

[3] B. Morris, W. Ning and Y. Peres; Mixing time of the card-cyclic-to-random shuffle, To appear in Annals of Applied Probability. http://arxiv.org/abs/1207.3406.

[4] E. Mossel, Y. Peres and A. Sinclair; Shuffing by semi-random transpositions, *Proceedings of the 45th Annual IEEE Symposium on Foundations of Computer Science(FOCS'04)* October 17-19, 2004, Rome, Italy, 572-581, IEEE(2004).

[5] R. Pinsky, Probabilistic and Combinatorial Aspects of the Card-Cyclic to Random Shuffle, *Random Struct. Alg.* (2013), on-line.

[6] L. Saloff-Coste and J. Zuniga; Convergence of some time inhomogeneous Markov chains via spectral techniques, *Stochastic Proc. Appl.* **117** (2007) 961-979.

[7] Tao, T.; When are eigenvalues stable?, http://terrytao.wordpress.com/2008/10/28/when-are-eigenvalues-stable/.

[8] Wilson, D. B.; Mixing times of lozenge tiling and card shuffling Markov chains, *A.. Appl. Probab.* **14** (2004), 274-325.

[9] Wilson, D. B.; Mixing time of the Rudvalis shuffle, *Electron. Commun. Probab.* **8** (2003), 77-85.

3 Appendix

For the Matlab computations, we have used three functions, `rimatris`, `riprod` and `riprod2`. Recall from Lemmas 2.3, 2.4 and 2.5, the transition matrix $B(n)$ for which there was established that the second eigenvalue has modulus at least 0.08, via considerations of approximate eigenvalues and eigenvectors for the matrices $S(n) = (B(n)+B(n)^T)/2$ and $D(n) = (B(n)-B(n)^T)/2$. The command `rimatris(n)` produces $B(n)$. The two other functions take an n-dimensional vector $\mathbf{v}$ as input and return $S(n)\mathbf{v}$ and $D(n)\mathbf{v}$ respectively. Since we needed n to be as large as 10^5, computation time was an important issue. Therefore the code has been optimized for computational speed and it is not quite as straightforward as one would at first believe on knowing $B(n)$. Here is the code.

```
function A=rimatris(n)

A=zeros(n,n);
r=zeros(1,n+1);
e=exp(1);
ep=1/n;
a=0;
ea=1;
ema=1;
eema=exp(1);
eep=exp(ep);
emep=1/eep;

for i=1:n,
a=a+ep;
```

```
ea=ea*eep;
ema=ema*emep;
eema=eemaêmep;
u=0;

for j=0:n,
z=j*ep;
s=min(e*ema*u,eema^(1-u)-e*ema*(1-u))-z;
while abs(s)>1e-12,
I=(u <= 1-(1-a)*ea);
u=u-s/(I*e*ema + (1-I)*(e*ema-ema*eema^(1-u)));
s=min(e*ema*u,eema^(1-u)-e*ema*(1-u))-z;
end
r(j+1)=u;
end
A(i,:)=r(2:n+1)-r(1:n);
end
function y=riprod(x);

n=length(x);
y=zeros(1,n);
e=exp(1);
ep=1/n;
z=ep*(0:n);
u=z;
a=0;
ea=1;
ema=1;
eema=exp(1);
eep=exp(ep);
emep=1/eep;

for i=1:n,
a=a+ep;
ea=ea*eep;
ema=ema*emep;
eema=eemaêmep;
```

```
s=min(e*ema*u,eema.^(1-u)-e*ema*(1-u))-z;
while max(abs(s))>1e-12,
I=(u <= 1-(1-a)*ea);
u=u-s./(I*e*ema + (1-I).*(e*ema-ema*eema.^(1-u)));
s=min(e*ema*u,eema.^(1-u)-e*ema*(1-u))-z;
end
r=u(2:n+1)-u(1:n);
y(i)=y(i)+r*x;
y=y+x(i)*r;
end
y=0.5*y';
function y=riprod2(x);

n=length(x);
y=zeros(1,n);
e=exp(1);
ep=1/n;
z=ep*(0:n);
u=z;
a=0;
ea=1;
ema=1;
eema=exp(1);
eep=exp(ep);
emep=1/eep;

for i=1:n,
a=a+ep;
ea=ea*eep;
ema=ema*emep;
eema=eemaêmep;

s=min(e*ema*u,eema.^(1-u)-e*ema*(1-u))-z;
while max(abs(s))>1e-12,
I=(u <= 1-(1-a)*ea);
u=u-s./(I*e*ema + (1-I).*(e*ema-ema*eema.^(1-u)));
```

```
s=min(e*ema*u,eema.^(1-u)-e*ema*(1-u))-z;
end
r=u(2:n+1)-u(1:n);
y(i)=y(i)+r*x;
y=y-x(i)*r;
end
y=0.5*y';
```

Given these functions, they have been used with the following set of commands.

```
A=rimatris(10001);
B=(A+A')/2;
C=(A-A')/2;
[u,l]=eigs(B,2);
u=u(:,2);
u=100*u;
l=l(2,2);
[w,k]=eigs(C,1); w=100*w;
for i=2:25, u(26-i)=u(26)-i*(u(26)-u(25));, end
for i=2:75, w(76-i)=w(76)-i*(w(76)-w(75));, end
du=10000*(u(2:10001)-u(1:10000));
dw=10000*(u(2:10001)-w(1:10000));
x=0:10000;
xx=0:0.1:10000;
y=interp1(x,u,xx);
y=y';
z=interp1(x,w,xx);
z=conj(z');
r=riprod(y)-l*y;
s=riprod2(z)-k*z;
sqrt(r'*r/100000);
sqrt(s'*s/100000);
max(abs(du));
max(abs(dw));
```

Then u and w are first the normalized eigenvectors of $B(n)$ and $D(n)$ respectively for $n = 10^4$. These are then smoothed out, whereupon y and z are the linear interpolations of the smoothed-out vectors. The commands $max(abs(du))$ and

$max(abs(dw))$ give $\hat{\phi}$ in the respective cases.