# Random-to-Random Shuffles

Camille Briat
Mathematics Honors Thesis
Advisor: Todd Kemp
UCSD

May 16, 2013

**Abstract**

Considering random-to-random shuffles as Markov chains, it has already been proven that the transition matrices of $k$-random-to-random and $l$-random-to-random commute. We set out to provide a simpler, probabilistic proof of the same result, but show that the problem boils down to a tricky combinatorics question.

# Contents

# 1 Set Up

## 1.1 Terminology

In this paper, we will consider shuffling a deck of $n$ cards as a Markov chain. Its states are permutations of $n$ cards as in arrangements, or final orderings of the cards, while its steps are permutations as in bijections. Then, if $X_i$ represents the $i$th state of the Markov chain, and $\sigma$ is an element of the symmetric group $S_n$, the probability of transitioning from an ordered deck of card to a permuted deck of card in one step can be written as $\mathbb{P}(X_i = \sigma | X_{i-1} = \text{id})$ where $\sigma$ is the said permutation. We will think of this as $P(\sigma)$, where $P$ represents the probability distribution on transitions (steps). In fact, $P$ is a probability distribution on $S_n$. We will see later that, in addition to saving notation, this is sufficient since we are only interested in the first row of the transition matrices that we will study. Now a few definitions.

Definition 1. Denote by $X_i$ the $i$th state of a given shuffle, and let $\sigma$ be a possible state of the said shuffle. Define $P(\sigma) \equiv \mathbb{P}(X_i = \sigma | X_{i-1} = \text{id})$. Then the support of probability distribution $P$ is called the **generating set** of the shuffle and is denoted $G$. In other words, if $\sigma \in G$, then $P(\sigma) \neq 0$.

Thus, the generating set of a shuffle is the set of permutations (orderings) of the deck that are reachable from an ordered deck.

Definition 2. A **forward description** of a shuffle is a random walk on $S_n$ characterized by a generating set $G = \{\sigma_1, \sigma_2, \ldots, \sigma_m\}$ and a probability distribution $P$ on this generating set.

This means that a forward description is simply a description of how to perform a shuffle. Note however that forward descriptions are not unique; several descriptions may induce the same generating sets and probability distributions. Definition 2 is only interesting by contrast with definition 3:

Definition 3. Continuing definition 2, let $G' = \{\sigma_1^{-1}, \sigma_2^{-1}, \ldots, \sigma_m^{-1}\}$, and let $P'$ be a probability distribution on $G'$. Then an **inverse description** of a shuffle is a random walk on $S_n$ characterized by the generating set $G'$ and with the property that $P(\sigma_i) = P'(\sigma_i^{-1})$.

Note that $G'$ is well defined since $S_n$ is a group, so $\sigma_i^{-1}$ exists in $S_n$ and is unique if written in two-line notation. Once again, we use indefinite articles, as inverse descriptions need not be unique.

## 1.2 Example: Top-to-Random

Here is a basic example. Consider the following forward description of a card shuffle called top-to-random (TTR): take the first card on top of a deck of $n$

cards and insert it uniformly randomly back inside the deck. The elements of the generating set $G$ are permutations of the form

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & \cdots & p-1 & p & p+1 & \cdots & n \\ 2 & 3 & 4 & \cdots & p & 1 & p+1 & \cdots & n \end{pmatrix}$$

where $p$ indicates the position in which we have inserted the top card. The probability distribution on this generating set is uniform: there are $n$ possible positions where we could place the top card, each yielding a different final ordering, hence each element of the generating set has probability $1/n$, which we write $P(\sigma) = 1/n$ for any $\sigma \in G$.

Now consider the following inverse description: choose a card from the deck uniformly randomly and place it on top of the deck. Let the generating set of this random walk be $G'$ and let the probability distribution be $P'$. To verify that we have indeed described an inverse description, we need to check that if $\sigma \in G$, then $\sigma^{-1} \in G'$, and that $P'(\sigma^{-1}) = P(\sigma)$. Inverting $\sigma$ from above, we get that

$$\begin{aligned} \sigma^{-1} &= \begin{pmatrix} 2 & 3 & 4 & \cdots & p & 1 & p+1 & \cdots & n \\ 1 & 2 & 3 & \cdots & p-1 & p & p+1 & \cdots & n \end{pmatrix} \\ &= \begin{pmatrix} 1 & 2 & 3 & 4 & \cdots & p & p+1 & \cdots & n \\ p & 1 & 2 & 3 & \cdots & p-1 & p+1 & \cdots & n \end{pmatrix} \end{aligned}$$

which is exactly the type of permutations that we would get by picking a card from the deck and putting it back on top. To compute $P'(\sigma^{-1})$, note that there are $n$ possible cards that we can pick out of the deck, and since we do this uniformly randomly, each card is equally likely to be picked. We must put the card back on top, so we get a different final ordering for each card we pick. This means that $P'(\sigma^{-1}) = 1/n = P(\sigma)$. Thus the second description is indeed an inverse description of TTR.

# 2 $k$-Top-to-Random Shuffle

We now wish to generalize top-to-random shuffles to $k$-top-to-random shuffles ($k$TTR).

## 2.1 Forward Descriptions

Forward Description 1. Separate the top $k$ cards from the rest of the deck and successively insert them uniformly randomly inside the remaining deck. This is different from performing a top-to-random shuffle $k$ times since we pick the top $k$ cards without replacement.

Let $\{p_1, p_2, \ldots, p_k\}$ be the $k$ positions, in increasing order, in which we have inserted the $k$ top cards, and $\{a_1, a_2, \cdots, a_k\} = \{1, 2, \ldots, k\}$ be the top $k$ cards in a permuted order. That is, let $\gamma \in S_k$ be permutation of $k$ elements and define $a_i = \gamma(i)$ for $1 \leq i \leq k$. Then $a_i$ gets moved to the $p_i$th spot $(1 \leq i \leq k)$,

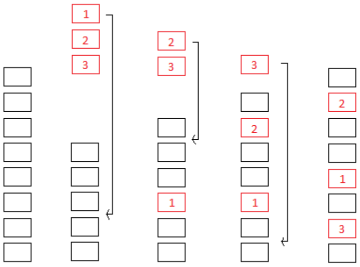


Figure 1: 3TTR on 8 cards, description 1

and all unmoved cards remain in their relative order. The $j$th unmoved card is the card labelled $j+k$, and it gets moved to position $j+m$ where $m$ is the number of moved cards placed above it (for $1 \leq j \leq n-k$). Now the card before the $i$th moved card (which is in spot $p_i$) is the $(p_i-1-(i-1))$th unmoved card (there are $p_i-1$ cards, $(i-1)$ of which were moved), hence the card in position $p_i-1$ is the card labelled $k+p_i-1-(i-1)=k+p_i-i$ (so $j=p_i-i$ and $m=i-1$). The card after the $i$th moved card has label one more than that, i.e. $k+p_i-i+1$. Therefore, the elements $\sigma$ of the generating set $G$ have the form

$$\begin{pmatrix} 1 & 2 & \cdots & p_1 & p_1+1 & \cdots & p_i-1 & p_i & p_i+1 & \cdots & n \\ k+1 & k+2 & \cdots & a_1 & k+p_1 & \cdots & k+p_i-i & a_i & k+p_i+1-i & \cdots & n \end{pmatrix}.$$

Note that if we place two or more cards next to each other, we may get $\sigma$s of the form

$$\begin{pmatrix} 1 & \cdots & p_1 & p_1+1 & \cdots & p_i-1 & p_i & p_{i+1} & p_i+2 & \cdots & n \\ k+1 & \cdots & a_1 & k+p_1 & \cdots & k+p_i-i & a_i & a_{i+1} & k+p_i+1-i & \cdots & n \end{pmatrix}$$

and so on. Since $p_{i+1}=p_i+1$, the next position is $p_i+2$. We work with the former form for simplicity.

Now let $P$ be the probability distribution over $G$. To compute the probability of a given generating element, observe that the number of choices for the position of the $i$th card is the number of unmoved cards plus the number of moved cards already placed plus one, i.e. there are $(n-k)+(i-1)+1=n-k+i$ choices. Since we insert the cards uniformly, each position has probability $1/(n-k+i)$. It follows from the independence of each insertion that

$$P(\sigma)=\frac{1}{n-k+1}\cdot\frac{1}{n-k+2}\cdot\dots\cdot\frac{1}{n}=\frac{1}{\binom{n}{k}k!}.$$

Forward Description 2. We now introduce a different method of reinserting the cards into the deck. Imagine a new "empty deck:" $n$ empty slots that are to be filled. Then proceed as follows: separate the top $k$ cards from the rest of the deck, permute them uniformly randomly, choose $k$ empty slots from the empty deck uniformly randomly, place the $k$ top cards into the slots in their permuted order, and finally place the remaining $n-k$ cards in the remaining slots in their same relative order.

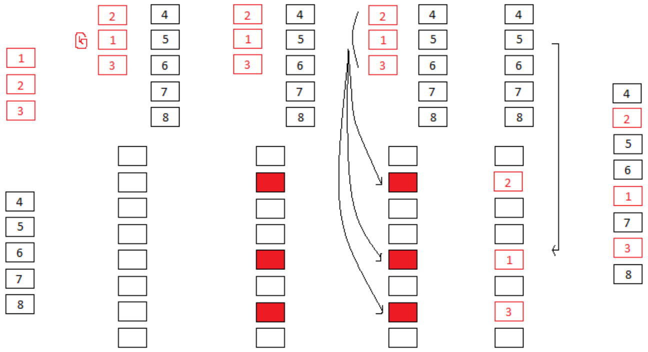


Figure 2: 3TTR on 8 cards, description 2

The only difference here is that we specify that the $k$ chosen cards are permuted uniformly randomly. The derivation of elements $\sigma$ of the generating set $G$ is exactly the same than in forward description 1 with the addition that $\gamma$ is chosen from $S_k$ uniformly randomly (recall that $\gamma$ is defined as $\gamma(i) = a_i$ where $\{a_1, a_2, \ldots, a_k\} = \{1, 2, \ldots, k\}$ are the top $k$ cards in their permuted order). A simple explanation of why $\gamma$ is uniformly random lies in calculating $P(\sigma)$. We choose $k$ cards independently of choosing a permutation of these $k$ cards, so

$$P(\sigma) = \frac{1}{\binom{n}{k}} \cdot \mathbb{P}(\gamma).$$

If we want $P(\sigma)$ to match with forward description 1, then we must have $\mathbb{P}(\gamma) = \frac{1}{k!}$. There are $k!$ permutations of $k$ cards, so this implies that $\gamma$ is chosen uniformly randomly from $S_k$.

## 2.2 Inverse Descriptions

This first description is closely related to forward description 2.

Inverse Description 1. Choose $k$ cards uniformly randomly from the deck, permute them uniformly randomly, and put them on top.

Let $\{p_1, p_2, \ldots, p_k\}$ be the original positions of the $k$ chosen cards in increasing order. Now let $\pi \in S_k$ be a uniformly randomly chosen permutation from the symmetric group on $k$ elements. Define $a_i = p_{\pi(i)}$ for $1 \leq i \leq k$ so

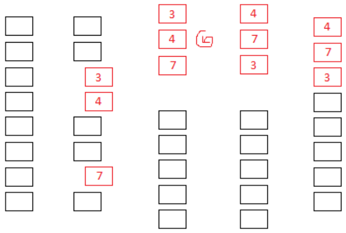


Figure 3: 3TTR$^{-1}$ on 8 cards, description 1

that $\{a_1, a_2, \cdots, a_k\}$ designates the permuted order of the $k$ chosen cards, i.e. $\{a_1, a_2, \cdots, a_k\} = \{p_1, p_2, \ldots, p_k\}$. Then $a_i$ gets moved to the $i$th spot and all unmoved cards remain in relative order, that is the $j$th unmoved cards is in spot $k + j$ (here $1 \leq j \leq n - k$). If $p_{i-1} < p_i - 1$ and $p_i + 1 < p_{i+1}$ (the $i$th chosen card is surrounded by unmoved cards), the $(p_i - 1)$th card is the $(p_i - 1) - (i - 1)$th unmoved card, so it is in position $k + p_i - i$. Similarly, the $(p_i + 1)$th card is the $(p_i + 1 - i)$th unmoved card so it is in position $k + p_i - i + 1$. If this is not the case and the adjacent previous card has been picked, then the $(p_i - 1)$th card is the $(p_{i-1})$th moved card, so the $(p_i - 2)$th card is in position $k + p_i - i$. This generalizes to any number of adjacent cards, preceding or succeeding $p_i$. However we will not list these scenarios for simplicity (it is easy to apply the same method as what follows to these cases). Then the elements $\tau$ of this description's generating set $H$ have the form

$$\begin{pmatrix} 1 & \cdots & k & k+1 & \cdots & k+p_1 & \cdots & k+p_i-i & k+p_i-i+1 & \cdots & n \\ a_1 & \cdots & a_k & 1 & \cdots & p_1+1 & \cdots & p_i-1 & p_i+1 & \cdots & n \end{pmatrix}.$$

To justify calling this description the inverse of the forward descriptions which have generating set $G$, we need to verify that $H = G'$. In other words, we need to check that $\tau = \sigma^{-1}$ were $\sigma \in G$. $\tau^{-1}$ has the form

$$\begin{pmatrix} a_1 & \cdots & a_k & 1 & \cdots & p_1+1 & \cdots & p_i-1 & p_i+1 & \cdots & n \\ 1 & \cdots & k & k+1 & \cdots & k+p_1 & \cdots & k+p_i-i & k+p_i-i+1 & \cdots & n \end{pmatrix}$$

which is the same as

$$\begin{pmatrix} 1 & \cdots & p_1 & p_1+1 & \cdots & p_i-1 & p_i & p_i+1 & \cdots & n \\ k+1 & \cdots & x_1 & k+p_1 & \cdots & k+p_i-i & x_i & k+p_i-i+1 & \cdots & n \end{pmatrix}$$

where $x_i$ is the card in the $p_i$th spot, i.e. $\pi^{-1}(x_i) = i$ and $\{x_1, \ldots, x_k\} = \{1, \ldots, k\}$. We can identify $\pi$ with $\gamma$ from the forward description, thus $x_i = \pi(i) = \gamma(i) = a_i$, from which it follows that $\tau^{-1}$ has the same form as $\sigma$. Thus $H = \{\sigma^{-1} | \sigma \in G\}$ as desired.

Let $P'$ be the probability distribution on $H$. We also need to verify that $P'(\sigma^{-1}) = P(\sigma)$. To compute $P'(\sigma^{-1})$, observe that uniformly picking $k$ cards out of $n$ is done with probability $1/\binom{n}{k}$ and that picking a permutation of $k$ cards uniformly randomly is done with probability $\frac{1}{k!}$. These two steps are applied independently, so:

$$P'(\sigma^{-1}) = \frac{1}{\binom{n}{k}} \cdot \frac{1}{k!} = \frac{1}{\binom{n}{k}k!}.$$

Therefore $P'(\sigma^{-1}) = P(\sigma)$ and we may call these descriptions inverses.

This second (equivalent) description is more closely related to forward description 1.

Inverse Description 2. Pick a card uniformly randomly from the deck and put it on top. Do this $k$ times, each time not picking one of the previously moved cards.

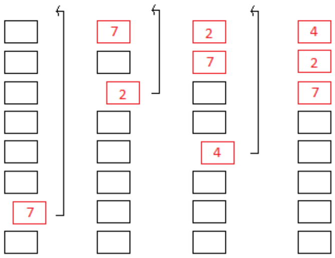


Figure 4: 3TTR$^{-1}$ on 8 cards, description 2

Let $p_i$ be the original position of the $i$th card we picked (for $1 \leq i \leq k$). Then $p_i$ is moved to spot $k-i+1$ and the unmoved cards retain their original relative order, as in inverse description 1. Therefore the elements $\tau$ of this description's generating set $H$ have the form

$$\begin{pmatrix} 1 & \cdots & k & k+1 & \cdots & k+p_1 & \cdots & k+p_i-i & k+p_i-i+1 & \cdots & n \\ p_k & \cdots & p_1 & 1 & \cdots & p_1+1 & \cdots & p_i-1 & p_i+1 & \cdots & n \end{pmatrix}$$

where once again, we may get slightly different forms if adjacent cards are moved.

To compare this with inverse description 1, we first observe that in this case, the $p_i$'s are not indexed in increasing order. In fact, if $\{q_1, q_2, \ldots, q_k\}$ are the positions in increasing order, then $p_i = q_{\pi(i)}$ where $\pi \in S_k$ is a uniformly randomly chosen permutation on $k$ elements. This is due to the fact that each card in $\{p_1, p_2, \ldots, p_k\}$ is equally likely to have been the $i$th card chosen. To keep notation consistent, let $a_i = p_{k-i+1}$. Then $a_i = q_{\pi'(i)}$ where $\pi' \in S_k$ is uniformly random as well since it is the composition of $\pi$ and the permutation that reverses order. Now we can rewrite $\tau$ as

$$\begin{pmatrix} 1 & \cdots & k & k+1 & \cdots & k+p_1 & \cdots & k+p_i-i & k+p_i-i+1 & \cdots & n \\ a_1 & \cdots & a_k & 1 & \cdots & p_1+1 & \cdots & p_i-1 & p_i+1 & \cdots & n \end{pmatrix}$$

and we see that the generating sets of inverse descriptions 1 and 2 are the same.

To compute probabilities, observe that the first card can be chosen among $n$ possibilities, the second among $n-1$, and so on: the $i$th card (recall that $1 \leq i \leq k$) can be chosen among $n-i+1$ possibilities. Thus, if $P'$ is the probability distribution on $H$, then:

$$P'(\tau) = \frac{1}{n} \cdot \frac{1}{n-1} \cdot \dots \cdot \frac{1}{n-k+1} = \frac{1}{\binom{n}{k}k!}.$$

Both inverse descriptions have the same probability distribution on the same generating sets, so they are equivalent.

## 2.3 Properties of $\mathbf{T_k}$

First, note that the following are equivalent notations:

1. $\mathbb{P}(X_i = \sigma_j | X_{i-1} = \sigma_i)$ where $X_i$ is state $i$ of the Markov chain for $k$TTR
2. $\mathbf{T_k}(\sigma_i, \sigma_j)$ where $\mathbf{T_k}$ is the transition matrix of $k$TTR
3. $[\mathbf{T_k}]_{i,j}$

Now recall that, at the very beginning of this paper, we mentioned that $P(\sigma) = \mathbb{P}(X_i = \sigma | X_{i-1} = \mathrm{id})$ was sufficient notation (here, $P$ is the probability distribution on the generating set of $k$TTR). We will now see why.

Property 1. $\mathbf{T_k}(\sigma_i, \sigma_j) = P(\sigma_i^{-1}\sigma_j)$. In other words, the first row of the transition matrix determines the entire matrix.

*Proof.* Consider $[\mathbf{T_k}]_{i,j} = \mathbf{T_k}(\sigma_i, \sigma_j)$. The trick is to relabel the cards according to $\tau$, which we pick to get $\tau\sigma_i = \mathrm{id}$ (there is only one such $\tau$ since permutations are bijections). That means that $\tau = \sigma_i^{-1}$. Thus, $\sigma_j$ is relabelled as $\tau\sigma_j = \sigma_i^{-1}\sigma_j$. Then $\mathbf{T_k}(\sigma_i, \sigma_j) = \mathbf{T_k}(\mathrm{id}, \sigma_i^{-1}\sigma_j) = P(\sigma_i^{-1}\sigma_j)$. Since $P(\sigma_i^{-1}\sigma_j)$ is an entry on the first row of $\mathbf{T_k}$, computing the $i$th row of $\mathbf{T_k}$ is only a matter of relabelling the cards according to $\sigma_i^{-1}$. □

Property 2. $\mathbf{T_k}$ is doubly stochastic.

*Proof.* Let $\sigma_i, \sigma_j, \tau \in S_n$. Then, by property 1, the column sums of $\mathbf{T_k}$ are:

$$\begin{aligned}
\sum_i \mathbf{T_k}(\sigma_i, \sigma_j) &= \sum_i P(\sigma_i^{-1}\sigma_j) \\
&= \sum_i P'((\sigma_i^{-1}\sigma_j)^{-1}) \\
&= \sum_i P'(\sigma_j^{-1}\sigma_i) \\
&= \sum_i \mathbf{T'_k}(\sigma_j, \sigma_i) \\
&= 1.
\end{aligned}$$

the second equality holds by the definition of an inverse description and the last equality holds because it is the row sum of an inverse description's transition matrix. □

Property 3. $\mathbf{T'_k} = \mathbf{T_k}^\top$.

*Proof.* $\mathbf{T_k}^\top$ makes sense as a transition matrix since $\mathbf{T_k}$ is doubly stochastic. Then,

$$\begin{aligned}\mathbf{T'_k}(\sigma_i, \sigma_j) &= P'(\sigma_i^{-1}\sigma_j)\\ &= P((\sigma_i^{-1}\sigma_j)^{-1})\\ &= P(\sigma_j^{-1}\sigma_i)\\ &= \mathbf{T_k}(\sigma_j, \sigma_i)\\ &= \mathbf{T_k}^\top(\sigma_i, \sigma_j). \quad \square\end{aligned}$$

## 2.4 Paths

Before we look at compositions of $k$TTR shuffles, we will introduce the notion of paths.

Definition 4. Consider a composition of shuffles. That is, let $\mathbf{P}$ and $\mathbf{P_i}$ for $1 \leq i \leq m$ be the transition matrices of shuffles with respective generating sets $G$ and $G_i$ such that $\mathbf{P} = \prod_1^m \mathbf{P_i}$. A **path** to a given generating element $\sigma$ of $G$ is a $k$-tuple $(\sigma_1, \ldots, \sigma_m)$ where $\sigma_i \in G_i$ such that $\prod_m^1 \sigma_i = \sigma$.

In other words, a path is a way of getting to a generating element. Note that so far, for $k$TTR and $k\text{TTR}^{-1}$ shuffles, paths have been indistinguishable from generating elements since there is only one path leading to each generating element. Indeed, in the forward description of $k$TTR, we are constrained to choosing the top $k$ cards. Thus, for any given generating element of $k$TTR, there is only one permutation of these $k$ cards and only one set of positions where we may put them to end up with the given generating element. Similarly, given a generating element of $k\text{TTR}^{-1}$, we know which cards must be selected from the deck to be placed on top. Then, there is a unique permutation that will arrange them in the correct order before we place them on top. We see that in both cases, paths correspond one-to-one with generating elements. For that reason, we will always take $k$TTR and $k\text{TTR}^{-1}$ shuffles as the elementary building blocks of paths.

## 2.5 Commutativity of $k$TTR and $l$TTR

Let $k$TTR have transition matrix $\mathbf{T_k}$ and generating set $T_k$, and $l$TTR have transition matrix $\mathbf{T_l}$ and generating set $T_l$. We are interested to see whether performing $k$TTR followed by $l$TTR is different from performing $l$TTR first and then $k$TTR. First, a couple of remarks.

Remark 1. $k$TTR followed by $l$TTR, which we will denote $(k,l)$TTR, is not the same as $(k+l)$TTR since the former allows replacement between the first set of $k$ cards and the second set of $l$ cards while the latter does not. Indeed, with $(k,l)$TTR we may end up moving less than $k+l$ distinct cards.

Remark 2. $(k,l)$TTR is an example of a process where paths and generating elements are distinguishable. For example, take $(1,2)$TTR on $n=4$ cards. Here, a path is a pair $(\sigma,\tau)$ with $\sigma \in T_1$ and $\tau \in T_2$. Now consider $\sigma_1 = \tau_1 = \mathrm{id}$,

$$\sigma_2 = \tau_2 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 1 & 3 & 4 \end{pmatrix}$$

and

$$\sigma_3 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 1 & 4 \end{pmatrix}, \quad \tau_3 = \begin{pmatrix} 1 & 2 & 3 & 4 \\ 3 & 1 & 2 & 4 \end{pmatrix}.$$

It turns out that $\tau_1\sigma_1 = \tau_2\sigma_2 = \tau_3\sigma_3 = \mathrm{id}$, i.e. several paths result in the same generating element (see figure 5 on next page).

Now let $l\mathrm{TTR} \circ k\mathrm{TTR}$ (abbreviated by $(k,l)$TTR) have transition matrix $\mathbf{T_{k,l}}$, i.e. $\mathbf{T_{k,l}} = \mathbf{T_k T_l}$, and have probability distribution $P_{k,l}$ on generating set $T_{k,l}$. Then $\mathbf{T_{k,l}} = \mathbf{T_{l,k}}$ if the following to conditions hold:

1. $T_{k,l} = T_{l,k}$.
2. If $\theta \in T_{k,l}$ (hence $\theta \in T_{l,k}$), then $P_{k,l}(\theta) = P_{l,k}(\theta)$.

Fact 1. $T_{k,l} = \bigcup_{m=0}^{\min\{k,l\}} T_{k+l-m}$.

*Proof.* Consider $\theta \in T_{k,l}$. A path representation of $\theta$ will indicate which cards were moved in each $k$TTR and $l$TTR shuffle, thus we may condition on the total number of cards moved. If $m$ denotes the number of overlapping cards (cards that were moved by $k$TTR and then again by $l$TTR), then $k+l-m$ distinct cards were moved by the path representation of $\theta$, so $\theta \in T_{k+l-m}$ has the form

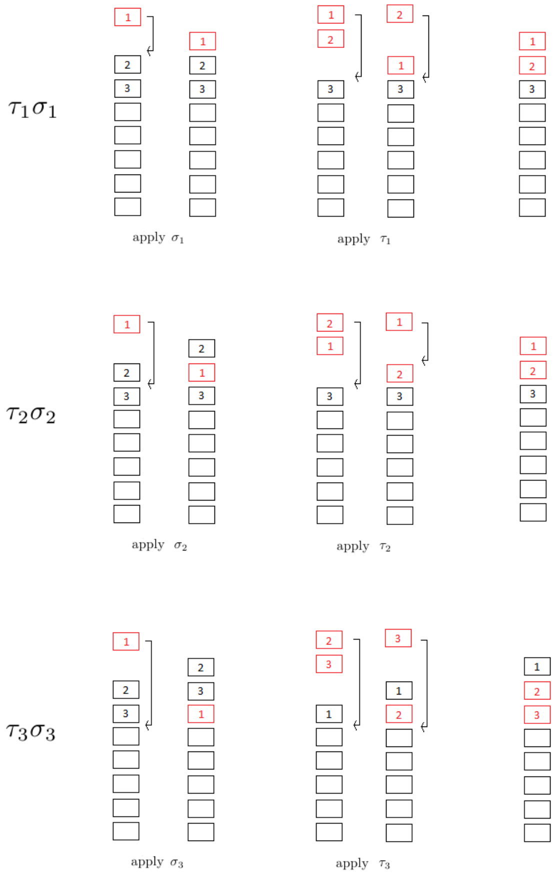
$$\begin{pmatrix} 1 & 2 & \cdots & p_1 & p_1+1 & \cdots & p_i-1 & p_i & p_i+1 & \cdots & n \\ k+1 & k+2 & \cdots & a_1 & k+p_1 & \cdots & k+p_i-i & a_i & k+p_i+1-i & \cdots & n \end{pmatrix}$$

where $\{a_1, a_2, \cdots, a_k\} = \{1, 2, \ldots, k+l-m\}$. Note that path representations are not unique, but that it is not a problem since the union need not be disjoint. □

Fact 1 implies condition 1. Condition 2 requires more work. Let's begin with the following 4 steps. Let $\sigma \in T_k$ and $\tau \in T_l$:

(1) $P_{k,l}(\theta) = \sum_{\text{paths}} P_{k,l}(\sigma\tau)$ where the sum extends over the set of paths $(\sigma,\tau)$ to the generating element $\theta$.

Figure 5: paths and generating elements are distinguishable for $(k, l)$TTR



(2) $P_{k,l}(\sigma\tau) = \sum\limits_m P_{k,l}(\sigma\tau \mid m) P_{k,l}(m)$ where $m$ is the overlap as described in Fact 1.

(3) $P_{k,l}(\sigma\tau \mid m)$ is symmetric in $k$ and $l$. The probability is calculated as follows: choose $m$ of the top $k$ cards and insert them among the top $l$ cards (so that they will be picked again with $l$TTR), then place the other $k - m$ cards of the top $k$ cards among the bottom $n - l$ cards of the deck. Next, perform the $l$TTR shuffle by placing the new $l$ top cards among the

entire deck:

$$\begin{aligned} & P_{k,l}(\sigma\tau \mid m) \\ = & \frac{1}{\binom{k}{m}} \cdot \frac{1}{\binom{l}{m} m!} \cdot \frac{1}{\binom{n-l}{k-m}(k-m)!} \cdot \frac{1}{\binom{n}{l} l!} \\ = & \frac{(k-m)!m!(l-m)!m!(n-l-k+m)!(k-m)!(n-l)!l!}{k!l!m!(n-l)!(k-m)!n!l!} \\ = & \frac{(k-m)!(l-m)!m!(n-l-k+m)!}{k!l!n!}. \end{aligned}$$

(4) $P_{k,l}(m)$ is hypergeometric, which is also symmetric in $k$ and $l$. This probability is calculated as such: choose $m$ cards out of the $k$ that will be moved twice, choosing the other $l-m$ cards out of the remaining $n-k$, this over the total possible choices of picking $l$ cards out of $n$.

$$\begin{aligned} P_{k,l}(m) &= \frac{\binom{k}{m}\binom{n-k}{l-m}}{\binom{n}{l}} \\ &= \frac{k!(n-k)!l!(n-l)!}{m!(k-m)!(l-m)!(n-k-l+m)!n!}. \end{aligned}$$

From (1) and (2), we get that $P_{k,l}(\theta) = \sum\limits_{\text{paths}} \sum\limits_{m} P_{k,l}(\sigma\tau \mid m) P_{k,l}(m)$. Combined with (3) and (4), this tells us that $P_{k,l}(\theta) = P_{l,k}(\theta)$ holds if the following conjecture is true:

Conjecture 1. $\#\{\sigma\tau \mid \sigma\tau \in G_{k,l}\} = \#\{\tau\sigma \mid \tau\sigma \in G_{l,k}\}$. That is, the number of paths to each generating element is the same in both shuffles.

# 3 $k$-Random-to-Random Shuffle

Consider the process with transition matrix $\mathbf{T_k}^\top \mathbf{T_k}$, which we will denote $\mathbf{R_k}$. For this process, we perform $k\text{TTR}^{-1}$ first, followed by $k$TTR. We call this a $k$-random-to-random shuffle, or $k$RTR. Loosely speaking, this process takes $k$ cards from anywhere inside a deck of $n$ cards and replaces them (also anywhere) inside the deck. Here are two possible descriptions of this process.

## 3.1 Descriptions

Referring back to inverse description 1 and forward description 1, we see that $k$RTR picks $k$ cards uniformly randomly from the deck, permutes them, puts them on top, and then successively inserts them back into the deck uniformly randomly. However, permuting the cards before inserting them back is redundant since we have already seen in forward description 2 of $k$TTR that inserting the cards back this way gives rise to a uniformly random permutation. Thus, our first description is as follows:

Description 1. Pick $k$ cards uniformly randomly, take them out of the deck, and successively insert them back uniformly randomly one by one.

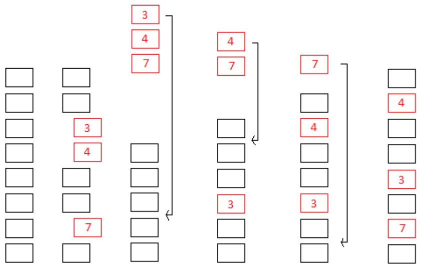


Figure 6: 3-RTR on 8 cards, description 1

If we refer back to inverse description 1 and forward description 2 instead, and once again omit the redundant step of permutating the $k$ cards a second time, we get the following description:

Description 2. Pick $k$ cards uniformly randomly, take them out of the deck and permute them, choose $k$ empty slots, place the $k$ cards in the empty slots in their permuted order, and finally place the remaining $n-k$ cards in the remaining slots in their same relative order.

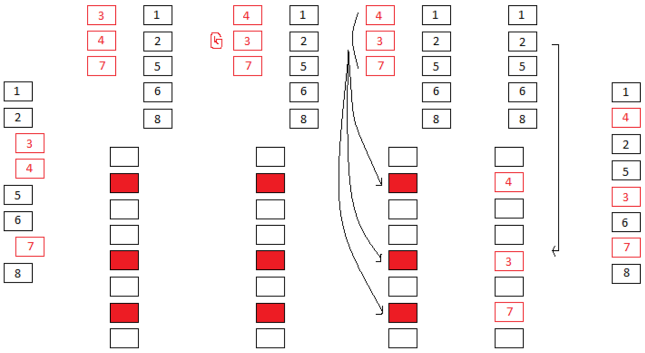


Figure 7: 3-RTR on 8 cards, description 2

Since we use equivalent forward and inverse $k$TTR descriptions in each $k$RTR description, these two descriptions are equivalent.

### 3.2 Properties of $\mathbf{R_k}$

PROPERTY 4. $\mathbf{R_k}(\sigma_i, \sigma_j) = P(\sigma_i^{-1}\sigma_j)$ i.e. the first row of $\mathbf{R_k}$ determines the entire matrix.

The proof is the same as for $k$TTR.

PROPERTY 5. $\mathbf{R_k}$ is a doubly stochastic matrix.

*Proof.* The product of doubly stochastic matrices is doubly stochastic: let $A$ and $B$ be doubly stochastic matrices, then

$$\begin{aligned}\sum_i [AB]_{i,j} &= \sum_i \sum_x A_{i,x}B_{x,j} \\ &= \sum_x \sum_i A_{i,x}B_{x,j} \\ &= \sum_x B_{x,j} \\ &= 1\end{aligned}$$

and similarly with column sums. Now recall that $\mathbf{T_k}$ is doubly stochastic, so $\mathbf{T_k}^\top$ is doubly stochastic. Then $\mathbf{R_k} = \mathbf{T'_k}\mathbf{T_k} = \mathbf{T_k}^\top\mathbf{T_k}$ is a product of doubly stochastic matrices, hence $\mathbf{R_k}$ is doubly stochastic. □

PROPERTY 6. $\mathbf{R_k}$ is symmetric.

*Proof.* Recall that $\mathbf{T'_k} = \mathbf{T_k}^\top$. Then,

$$\begin{aligned}\mathbf{R_k}^\top &= (\mathbf{T'_k}\mathbf{T_k})^\top \\ &= (\mathbf{T_k}^\top\mathbf{T_k})^\top \\ &= \mathbf{T_k}^\top(\mathbf{T_k}^\top)^\top \\ &= \mathbf{T_k}^\top\mathbf{T_k} \\ &= \mathbf{T'_k}\mathbf{T_k} \\ &= \mathbf{R_k}. \quad \square\end{aligned}$$

PROPERTY 7. $\mathbf{R'_k} = \mathbf{R_k}$.

*Proof.* Intuitively, it follows from the fact that $k$RTR $= k$TTR$^{-1} \circ k$TTR that $k$RTR$^{-1} = k$RTR. More rigorously, all we need to show is that $k$RTR and $k$RTR$^{-1}$ have the same paths. We do this by finding a bijection between the paths of both shuffles. Let $\sigma_1$ and $\sigma_2$ be any two generating elements of $k$TTR. Then $(\sigma_1^{-1}, \sigma_2)$ is an arbitrary path of $k$RTR. Now the inverse function gives us the desired correspondence: $(\sigma_2^{-1}, (\sigma_1^{-1})^{-1}) = (\sigma_2^{-1}, \sigma_1)$ is a valid path of $k$RTR$^{-1}$. □

Property 7 implies that $\mathbf{R'_k} = \mathbf{R_k}^\top$, as for $\mathbf{T_k}$, since $\mathbf{R_k}^\top = \mathbf{R_k}$. It is also the reason that we do not distinguish between forward and inverse descriptions of $k$RTR.

### 3.3 Commutativity of $k$RTR and $l$RTR

Let's recall and define some notation. Let $i$TTR have generating set $T_i$ and transition matrix $\mathbf{T_i}$. Let $i$RTR have generating set $G_i$ and transition matrix $\mathbf{R_i}$. Let $j\text{RTR} \circ i\text{RTR}$ be abbreviated by $(i,j)$RTR, and let its generating set be $G_{i,j}$ and transition matrix be $\mathbf{R_{i,j}}$. Thus, $\mathbf{R_{i,j}} = \mathbf{R_i}\mathbf{R_j} = \mathbf{T_i}^\top\mathbf{T_i}\mathbf{T_j}^\top\mathbf{T_j}$. Finally, let $P_i$ be the probability distribution on $G_i$, and $P_{i,j}$ the probability distribution on $G_{i,j}$.

Mirroring the analysis on the commutativity of $k$TTR and $l$TTR, we want to verify the following two conditions:

1. $G_{k,l} = G_{l,k}$.
2. If $\theta \in G_{k,l}$ (hence $\theta \in G_{l,k}$), then $P_{k,l}(\theta) = P_{l,k}(\theta)$.

Fact 2. $G_{k,l} = \bigcup_{m=0}^{\min\{k,l\}} G_{k+l-m}$.

*Proof.* We use the same conditioning argument as for $(k,l)$TTR, reproduced here: Consider $\theta \in G_{k,l}$. A path representation of $\theta$ will indicate which cards were moved in each $k$RTR and $l$RTR shuffle, thus we may condition on the total number of cards moved. If $m$ denotes the number of overlapping cards (cards that were moved by $k$RTR and then again by $l$RTR), then $k+l-m$ distinct cards were moved by the path representation of $\theta$, so $\theta \in G_{k+l-m}$. Note that path representations are not unique, but that it is not a problem since the union need not be disjoint. $\square$

Fact 2 implies condition 1. Now let $\sigma_1, \sigma_2 \in T_k$ and $\tau_1, \tau_2 \in T_l$,

(1) $P_{k,l}(\theta) = \sum_{\text{paths}} P_{k,l}(\sigma_1^{-1}\sigma_2\tau_1^{-1}\tau_2)$ where the sum extends over the set of paths $(\sigma_1^{-1}, \sigma_2, \tau_1^{-1}, \tau_2)$ to the generating element $\theta$.

(2) $P_{k,l}(\sigma_1^{-1}\sigma_2\tau_1^{-1}\tau_2) = \sum_m P_{k,l}(\sigma_1^{-1}\sigma_2\tau_1^{-1}\tau_2 \mid m)P_{k,l}(m)$ where $m$ is the overlap as described in Fact 2.

(3) $P_{k,l}(\sigma_1^{-1}\sigma_2\tau_1^{-1}\tau_2 \mid m)$ is symmetric in $k$ and $l$. The probability is calculated as follows: choose $k$ cards from the deck and reinsert them, then choose $m$ cards from the previous $k$ and $l-m$ cards from the other $n-k$

and reinsert those $l$ cards.

$$\begin{aligned}
&P_{k,l}(\sigma_1^{-1}\sigma_2\tau_1^{-1}\tau_2 \mid m) \\
&= \frac{1}{\binom{n}{k}} \cdot \frac{1}{\binom{n}{k}k!} \cdot \frac{1}{\binom{k}{m}\binom{n-k}{l-m}} \cdot \frac{1}{\binom{n}{l}l!} \\
&= \frac{k!(n-k)!k!(n-k)!m!(k-m)!(l-m)!(n-k-l+m)!l!(n-l)!}{n!n!k!k!(n-k)!n!l!} \\
&= \frac{(n-k)!(n-l)!m!(k-m)!(l-m)!(n-k-l+m)!}{n!n!n!}.
\end{aligned}$$

(4) $P_{k,l}(m)$ is still hypergeometric, which is also symmetric in $k$ and $l$. This probability is calculated in exactly the same way as for TTR shuffles.

(1) and (2) imply that $P_{k,l}(\theta) = \sum\limits_{\text{paths}} \sum\limits_{m} P_{k,l}(\sigma_1^{-1}\sigma_2\tau_1^{-1}\tau_2 \mid m)P_{k,l}(m)$. Thus, $P_{k,l}(\theta) = P_{l,k}(\theta)$ follows from the symmetries described in (3) and (4) and the following conjecture:

Conjecture 2.
$\#\{\sigma_1^{-1}\sigma_2\tau_1^{-1}\tau_2 \mid \sigma_1^{-1}\sigma_2\tau_1^{-1}\tau_2 \in G_{k,l}\} = \#\{\tau_1^{-1}\tau_2\sigma_1^{-1}\sigma_2 \mid \tau_1^{-1}\tau_2\sigma_1^{-1}\sigma_2 \in G_{l,k}\}$. That is, the number of paths to each generating element is the same in both shuffles.

With random-to-random shuffles, we may break this down further by using the next two facts:

Fact 3. If $\theta \in G_{k,l}$, then $\theta^{-1} \in G_{k,l}$.

*Proof.* By fact 1, $\theta \in G_{k+l-m}$ for some $m \leq \min\{k, l\}$. Then $\theta = \sigma_1^{-1}\sigma_2$ where $\sigma_1, \sigma_2 \in T_{k+l-m}$, therefore $\theta^{-1} = \sigma_2^{-1}\sigma_1 \in G_{k+l-m} \subset G_{k,l}$. □

Fact 4. If $\theta \in G_{k,l}$, then $\theta^{-1} \in G_{l,k}$.

*Proof.* Recall that a path in $l$RTR ∘ $k$RTR can be expressed as $\theta = \sigma_1^{-1}\sigma_2\tau_1^{-1}\tau_2$ for some $\sigma_1, \sigma_2 \in T_k$ and $\tau_1, \tau_2 \in T_l$. It follows that $\theta^{-1} = \tau_2^{-1}\tau_1\sigma_1^{-1}\sigma_2$ is an element of $G_{l,k}$. □

Thus, proving condition 2 can be reduced to finding a bijection between path inverses in $(k, l)$RTR. Indeed, fact 3 gives us a bijection between the paths of $\theta$ in $G_{k,l}$ and $\theta^{-1}$ in $G_{l,k}$. Thus, if we could find a bijection between the paths of $\theta$ in $G_{k,l}$ and $\theta^{-1}$ in $G_{k,l}$, we would have a bijection between paths of $\theta$ in $G_{k,l}$ and $\theta$ in $G_{l,k}$, proving conjecture 2.

Remark 3. As a side note, "$\mathbf{T_k}$, $\mathbf{T_l}^\top$ commute ⇒ $\mathbf{T_k}^\top\mathbf{T_k}$, $\mathbf{T_l}^\top\mathbf{T_l}$ commute" is only a one way implication. For a counter example of the other direction, take $k = l = 1$ so that $\mathbf{T_k}$ and $\mathbf{T_l}$ are transition matrices for 1TTR. Then

$\mathbf{T_k}^\top \mathbf{T_k} = \mathbf{T_l}^\top \mathbf{T_l}$ obviously commute, but $\mathbf{T_1}^\top \mathbf{T_1}$ moves only one card while $\mathbf{T_1} \mathbf{T_1}^\top$ can move two cards.

### 3.4 Another Perspective

Fact 5. Suppose $\mathbf{A}$ and $\mathbf{B}$ are symmetric matrices. Then $\mathbf{A}$ and $\mathbf{B}$ commute if and only if $\mathbf{AB}$ is symmetric.

*Proof.* Since $\mathbf{A}$ and $\mathbf{B}$ are symmetric, $\mathbf{A}^\top = \mathbf{A}$ and $\mathbf{B}^\top = \mathbf{B}$. Then $(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top = \mathbf{BA}$. Thus, $\mathbf{AB}$ is symmetric if $\mathbf{A}$ and $\mathbf{B}$ commute, and $\mathbf{A}$ and $\mathbf{B}$ commute if $\mathbf{AB}$ is symmetric. □

This means that we could show that $k$RTR and $l$RTR commute by showing that $\mathbf{R_{k,l}}$ is symmetric. From the same relabelling trick we used to show that only the first row of transition matrices matters, we can get that

$$\begin{aligned} &[\mathbf{R_{k,l}}]_{i,j} = [\mathbf{R_{k,l}}]_{j,i} \\ \Rightarrow\ &\mathbf{R_{k,l}}(\sigma_i, \sigma_j) = \mathbf{R_{k,l}}(\sigma_j, \sigma_i) \\ \Rightarrow\ &\mathbf{R_{k,l}}(\mathrm{id}, \sigma_i^{-1}\sigma_j) = \mathbf{R_{k,l}}(\mathrm{id}, \sigma_j^{-1}\sigma_i). \end{aligned}$$

Let $\theta = \sigma_i^{-1}\sigma_j$. Then $\theta^{-1} = (\sigma_i^{-1}\sigma_j)^{-1} = \sigma_j^{-1}\sigma_i$. Thus, finding a bijection between inverses in the generating set of $(k,l)$RTR would show that $\mathbf{R_{k,l}}$ is symmetric, hence that $k$RTR and $l$RTR commute. This is what we concluded in the previous section.

## References

[1] D. Aldous and P. Diaconis, Shuffling Cards and Stopping Times, *The American Mathematical Monthly*, Vol. 93, No. 5 (May, 1986), pp. 333-348.

[2] Daskalakis, Probability and Computation, *Lecture 5*, (Feb 16 2010), http://people.csail.mit.edu/costis/6896sp11/lec5s.pdf.