Strong Stationary Times for Non-Uniform Markov Chains

Dinah Shender

May 18, 2010

Abstract

This thesis studies several approaches to bounding the total variation distance of a Markov chain, focusing primarily on the strong stationary time approach. While strong stationary times have been used successfully with uniform walks on groups, non-uniform walks have proven harder to analyze. This project applied strong stationary techniques to simple non-uniform walks in the hope of finding some more general tools.

1 Introduction

Markov Chains describe processes where the future depends only on the present and not on the past. A simple example is to imagine that if it is raining today, it will rain tomorrow with probability 60% and if it isn't raining today then it will rain tomorrow with probability 30%. Once we know the weather yesterday, knowing the weather two days ago does not give us extra information about whether it will rain tomorrow. More formally, we say that a Markov Chain is a series of random variables $X_0, X_1, ...$ with state space Ω and a transition matrix, P, fulfilling the following conditions:

- $P(x, y) \geq 0$ and $\displaystyle\sum_{y \in \Omega} P(x, y) = 1$ for all $x \in \Omega$
- $P(X_{t+1} = x_{t+1} | X_t = x_t, X_{t-1} = x_{t-1}, ..., X_0 = x_0) = P(X_{t+1} = x_{t+1} | X_t = x_t) = P(x_t, x_{t+1})$

The first property makes P a **stochastic matrix**, while the second is a called the **Markov Property**. In this thesis we will focus on Markov Chains that eventually converge to some limiting distribution. This distribution is called the **stationary distribution** and is defined by $\pi P = \pi$. We can write this as $\displaystyle\sum_{x \in \Omega} \pi(x) P(x, y) = \pi(y)$ which says that if we sample from the distribution π, and move one step in the process, the distribution at the next step will still be π. Thus once the chain reaches the stationary distribution it stays there for the rest of time. In the example above, if we let rain be state 1, and not rain be state 2, then we see that the transition matrix P would be

$$\begin{pmatrix} .6 & .4 \\ .3 & .7 \end{pmatrix}$$

We can also see that the row vector $\pi = \begin{pmatrix} \frac{3}{7} & \frac{4}{7} \end{pmatrix}$ will be the stationary distribution since $\pi P = \pi$.

The next reasonable question to ask is what sorts of Markov chains have stationary distributions and whether the chain will ever reach the stationary distribution. It turns out that under mild conditions a Markov chain will have a unique stationary distribution to which it converges. In order to discuss this we introduce some terminology. A chain is called **irreducible** if for any $x, y \in \Omega$ there exists an integer n such that $P^n(x, y) > 0$. This says that starting from any state it is possible to eventually get to any other state. Let $T(x) = \{t \geq 1 : P^t x, x > 0\}$. Then the **period** of a state x in Ω is $\gcd T(x)$. If we start at a state with period 2, for example, it is only possible to return at even times. The following proof, taken from [6, section 1.3], shows that for irreducible chains, the period is a property of the chain.

Lemma 1. *If a chain is irreducible, then all states have the same period.*

Proof. Fix two states, x and y. Since the chain is irreducible, there exist r and l such that $P^r(x, y) > 0$ and $P^l(y, x) > 0$. Then $P^{r+l}(x, x) \geq P^r(x, y)P^l(y, x) > 0$ so $r + l \in T(x)$. Similarly, $P^{r+l}(y, y) \geq P^l(y, x)P^r(x, y) > 0$ so $r + l \in T(y)$. Then $r + l \in T(x) \cap T(y)$. Moreover, if $k \in T(x)$ then $P^{k+r+l}(y, y) \geq P^l(y, x)P^k(x, x)P^r(x, y) > 0$ so $k + r + l \in T(y)$. In other words, $T(x) \subset T(y) - (r + l)$. Since $\gcd T(y)$ divides $r + l$, $\gcd T(y)$ must divide every element of $T(x)$. Thus $\gcd T(y) \leq \gcd T(x)$. However by an analogous argument, $\gcd T(x) \leq \gcd T(y)$ and so $\gcd T(x) = \gcd T(y)$. □

Since for irreducible chains the period is a property of the chain, we will call such chains **aperiodic** if all states have period 1 and **periodic** otherwise. Note that a sufficient condition for a chain to be aperiodic is for $P(x, x) > 0$ for some $x \in \Omega$.

Lemma 2. *Let P be the transition matrix of an irreducible Markov Chain. Then there exists a unique stationary distribution π.*

The proof is rather long and is omitted here. See [6, sections 1.5.3 and 1.5.4].

Most of the chains discussed here will be random walks on groups. These random walks will have an increment distribution μ defined on the group G. The chain will start with the identity and proceeds by multiplying on the left by an element h according to μ so that $P(g, hg) = \mu(h)$. Let S be the support of μ. Then the chain is irreducible if and only if S generates G. All walks on groups discussed here will be irreducible. Due to the group structure, all such random walks have the same stationary distribution.

Lemma 3. *Let P be the transition matrix for a random walk on a group G with increment distribution μ, which has support S. Suppose S generates G. Then the uniform probability distribution on G is the unique stationary distribution.*

Proof. We simply check the definition of a stationary distribution given above.

$$\sum_{h\in G} \frac{1}{|G|} P(h,g) = \frac{1}{|G|} \sum_{k\in G} P(k^{-1}g,g) = \frac{1}{|G|} \sum_{k\in G} \mu(k) = \frac{1}{|G|}$$

Uniqueness follows from irreducibility. □

The final piece needed to state the Convergence Theorem is a notion of distance. The most commonly used notion of distance is the **total variation distance** between two probability distributions. For two probability distributions μ and ν on Ω the total variation distance between them is defined to be

$$||\mu - \nu||_{TV} = \max_{A\subset\Omega} |\mu(A) - \nu(A)|$$

where of course $\mu(A) = \sum_{x\in A} \mu(x)$ and similarly for $\nu(A)$. We can now state the Convergence Theorem.

Theorem 4. *Suppose that P is irreducible and aperiodic with stationary distribution π. Then there exists constants $\alpha \in (0,1)$ and $C > 0$ such that*

$$\max_{x\in\Omega} ||P^t(x,\cdot) - \pi||_{TV} \leq C\alpha^t$$

Note that $P^t(x,\cdot) = P(X_t = \cdot|X_0 = x)$. This theorem can be traced back to Markov and there are several different proofs. A particularly accessible one can be found in [6, section 4.3]. It is common to abbreviate the left hand side as $d(t)$. We then refer to the **mixing time** as

$$t_{mix}(\epsilon) = \min\{t : d(t) \leq \epsilon\}$$

It is also common to in particular define $t_{mix} = t_{mix}(1/4)$. We say that a chain has "mixed" when $d(t)$ is small.

Another commonly used notion of distance is the separation distance. It is defined to be

$$s_x(t) = \max_{y\in\Omega}(1 - \frac{P^t(x,y)}{\pi(y)})$$

. It is convenient to define

$$s(t) = \max_{x\in\Omega} s_x(t)$$

We will show in the next section that $d(t) \leq s(t)$.

Before delving into stopping times as a technique to bound the mixing time of a walk, it is worth briefly mentioning a few other methods. The first is coupling. A coupling of Markov chains with transition matrix P is defined to be a process (X_t, Y_t) with the property that both X_t and Y_t are Markov chains with transition matrix P. We will usually let $X_0 = x$ and $Y_0 = y$. Then the first time that $X_t = Y_t$ turns out to provide a bound for the total variation distance of the chain.

Theorem 5. *Let* $\tau_{couple} = \min t : X_t = Y_t$*. Then*

$$\max_{x\in\Omega} \|P^t(x,\cdot) - \pi\| \leq P(\tau_{couple} > t)$$

.

The proof is omitted here but can be found in [6, section 5.2].

Another set of techniques used is that of analyzing the eigenvalues of the walk. Since the stationary distribution is a left eigenvector for P, it is natural that the eigenvalues would give information about how long until the stationary distribution is reached. One useful technique of this kind is that of comparison, which will be discussed more at the end of section 3.

The next section will focus on using strong stationary times to bound the total variation distance and provide some examples in both the weighted and unweighted cases. Section 3 makes an effort to bound the stopping time found for a variation of the top to random shuffle as well as exploring other techniques to bound this walk.

2 Stopping Times

Many Markov Chains have physical or geometric interpretations. In thinking about how long it takes a chain to mix, it is natural to think about this interpretation and look for an event that would signal that the chain must have mixed by that point. Stationary times are a way of formalizing this idea.

A **stopping time** for a chain X_t is a time τ such that the event $\{\tau = t\}$ doesn't depend on X_s for $s > t$. Intuitively, a stopping time is just a rule that says when to stop the chain as some function of the current and previous states of the chain. A simple example is a chain corresponding to the result of a series of coin flips and the stopping time "stop the first time heads comes up." There are also **randomized stopping times**, where each potential path $X_1, X_2, ..., X_i$ is assigned a probability of stopping, q_i, which depends only on the first i states of the walk. Intuitively, at each step we flip a coin with probability of heads equal to q_i. If the coin comes up heads then the chain is stopped, otherwise it continues. Thus we can look at a stopping time as a randomized stopping time where the probability of heads is always either 0 or 1. Extending this further, we may define a **stationary time** to be a time τ such that $P_x(X_\tau = y) = \pi(y)$. Note that we do not require that X_τ and τ be independent. Finally, a **strong stationary time** is a randomized stopping time where $P_x\{\tau = t, X_\tau = y\} = P_x\{\tau = t\}\pi(y)$. Thus a strong stationary time is a stopping time where the state X_τ is independent of τ and where the probability of being in any particular state is the stationary distribution. This last definition was first made in [1]. That paper first introduced a technique to find upper bounds on the mixing time that relies on the following connection between the probability of reaching the stopping time by time t and the total variation distance. While this theorem was first shown in [1], the following proof is taken from [6, section 6.4]. The results of each step of the proof are of independent interest.

Theorem 6. *If τ is a strong stationary time for a chain X with state space Ω then*

$$d(t) = \max_{x \in \Omega} \|P^t(x, \cdot) - \pi\|_{TV} \leq \max_{x \in \Omega} P_x(\tau > t)$$

Proof. The proof goes in two steps. The first step is to show that $P_x(\tau > t)$ is an upper bound for the separation distance starting from x. The second step is to show that the separation distance is an upper bound for the total variation distance.

The separation distance starting at x is $s_x(t) = \max_{y \in \Omega}(1 - \frac{P^t(x,y)}{\pi(y)})$. Notice that

$$1 - \frac{P^t(x, y)}{\pi(y)} = 1 - \frac{P_x(X_t = y)}{\pi(y)} \leq 1 - \frac{P_x(X_t = y, \tau \leq t)}{\pi(y)} = 1 - \frac{\pi(y) P_x(\tau \leq t)}{\pi(y)} = P_x(\tau > t)$$

from whence it follows that $s_x(t) \leq P_x(\tau > t)$. The second to last equality holds because

$$P_x(\tau \leq t)\pi(y) = \sum_{s \leq t} P_x(\tau = s)\pi(y) = \sum_{s \leq t} \sum_{z \in \Omega} P_x(\tau = s) P^{t-s}(z, y)\pi(z) =$$

$$\sum_{s \leq t} \sum_{z \in \Omega} P_x(\tau = s) P(X_\tau = z) P^{t-s}(z, y) = \sum_{s \leq t} \sum_{z \in \Omega} P_x(\tau = s, X_\tau = z) P^{t-s}(z, y) =$$

$$\sum_{s \leq t} P_x(\tau = s, X_{\tau + t - s} = y) = \sum_{s \leq t} P_x(\tau = s, X_t = y) = P_x(\tau \leq t, X_t = y)$$

For the second part of the proof consider that

$$\|P^t(x, \cdot) - \pi\|_{TV} = \sum_{y \in \Omega, P^t(x,y) < \pi(y)} (\pi(y) - P^t(x, y)) =$$

$$\sum_{y \in \Omega, P^t(x,y) < \pi(y)} \pi(y)(1 - \frac{P^t(x, y)}{\pi(y)}) \leq \max_y (1 - \frac{P^t(x, y)}{\pi(y)}) = s_x(t)$$

where the second to last step follows because $\pi(\cdot)$ is a probability distribution. Putting the two parts together gives the needed statement. □

It is worth emphasizing that this theorem is what makes the search for strong stationary times worthwhile. Since stationary times often have straightforward distributions, they make it simple to give a bound on total variation distance. What is more, [2] shows that a time giving a good bound exists.

Theorem 7. *For any Markov chain X_t there exists a strong stationary time τ such that $s_x(t) = P_x(\tau > t)$ for all t.*

Unfortunately, while we know that strong stationary times exist, finding such times has proven difficult. In particular, most known examples of strong stationary times involve random walks on groups with uniform increment distribution. The project of this thesis is to study some simple chains with non-uniform increment distributions in the hope of finding tools which can be applied more generally. However, we will start by going back and looking at one of the first examples [1] of a strong stationary time in order to get a feel for what we are trying to do.

Imagine a numbered deck of n cards from which we take the top card and place it in the deck uniformly at random. If we continue to do this at each step, we get what is known as the top to random shuffle. Letting the permutation map from the number of the card to the current position of the card be the state of the chain we see that this is a random walk on S_n and that the current state depends only on the previous state. We may define a stopping time by setting τ_{top} to be the step when the original bottom card first reaches the top and defining $\tau = \tau_{top} + 1$ to be one step after, ie the step when the original bottom card is being placed into the deck.

Theorem 8. *τ is a strong stationary time.*

Proof. Note first that because this is a random walk on S_n the stationary distribution is the uniform distribution. Consider the first time when a card is placed under the bottom card. This card is trivially randomly placed with respect to itself. Now consider the first time there are two cards under the initial bottom card. Given that a card was placed below the initial bottom card, it is equally likely to have been placed between the bottom card and the first card or below the first card (ie at the very bottom of the deck). Thus all permutations of the bottom two cards are equally likely and we may say that these cards are randomly arranged. Now proceed by induction and suppose that there are k cards below the original bottom card, which are randomly arranged. Consider the step when the $k + 1^{st}$ card is placed below the original. There are $k + 1$ spots where we may put this card, all with the same probability. Thus all permutations of the $k + 1$ cards are equally likely. Since the order of these cards does not change until a new card is placed below the original bottom card, all permutations are equally likely as long as there are $k + 1$ cards under the original bottom card and the induction holds.

Now consider the state of the deck at time τ_{top}. By the above argument, all permutations of the bottom $n-1$ cards are equally likely at this time. However, the bottom card is equally likely to be placed in any of the n possible positions, and so it follows that at time τ, all permutations of the deck are equally likely. Thus the state of the deck at time τ is equally likely to be any element of S_n. More formally, $P(X_\tau = y) = \pi(y)\ \forall y \in S_n$. Finally, throughout this argument, given the number of cards below the bottom card, the state of the chain has been independent of time. Thus it follows that X_τ is also independent of τ. ☐

To think about how long this time takes, note that if there are k cards below the bottom card, then the top card has a $\frac{k+1}{n}$ chance of being placed below the bottom card. In other words, the original bottom card has probability $\frac{k+1}{n}$ of rising one spot. This distribution of probabilities is the same as those in the coupon collector's problem and so we may use the bound for the coupon collector to bound τ. Thus we have that the variation distance for the walk is bounded by $d(n \log n + cn) \leq P(\tau > n \log n + cn) \leq e^{-c}$. It can be shown that this bound is tight and that this walk actually achieves cutoff [1].

While it is usually easier to find a bound for a strong stationary time than it is to bound the variation distance directly, finding such a time is often quite difficult. One technique, used multiple times in [7], that is useful when considering walks on S_n is to think of marking elements and letting the candidate time be the first time all cards are marked. As an example, consider the random walk on S_n generated by the transpositions $(1, i),\ i = 1, ..., n$ with each generator having probability $\frac{1}{n}$. We will think about this walk as moving the elements or cards $1, 2, ..., n$ to different positions, also labeled $1, 2, ..., n$. A step in the walk is equivalent to picking a position i and switching the cards in the first and i^{th} positions. We will let J_t be the set of marked cards at time t, I_t the set of their positions at time t, and $k_t = |I_t| = |J_t|$. Let $\pi_t : I_t \to J_t$ be the map from a position containing a marked card at time t to the actual marked card in that position. That is, if card 3 is marked and in position 6 at time t then $\pi_t(6) = 3$. The idea is to mark the card in position 1 if it is unmarked and staying or if it is unmarked and being switched with a marked card. The proof below, adapted from [7, section 5.1], shows that then π_t is always equally likely to be any bijection from I_t to J_t. Formally, the time is as follows:

Theorem 9. *Let $I_1 = J_1 = \{n\}, k_1 = 1, t = 1,$ and $\sigma_1 = id$. While $k_t < n$*

- *Pick a position i uniformly at random.*
- *If $1 \notin I_t$ and $i \in I_t \cup \{1\}$, then mark the element in the first position by*
  - *Letting $I'_t = I_t \cup \{1\}, J_{t+1} = J \cup \{\sigma_t(1)\}$ and $k_{t+1} = k_t + 1$*
  - *Let $I_{t+1} = (1, i)I'_t$*
- *Otherwise, let $J_{t+1} = J_t$, $I_{t+1} = (1, i)I_t,$ and $k_{t+1} = k_t$.*
- *In all cases, let $\sigma_{t+1} = (1, i)\sigma_t$.*

Let τ be the first time that $k_t = n$. Then τ is a strong stationary time.

Note that in the notation above $(1, i)I$ simply means apply the transposition $(1, i)$ to all elements of I. Also, we multiply on the left in calculating σ_t because we are interested in tracking the permutation map from positions to elements. This is exactly the inverse of the permutation map from elements to positions. Since the latter is also the state of the chain, multiplying on the left gives the wanted map.

Proof. As mentioned, the main claim of the proof is that given $I_t, J_t,$ and k_t, the correspondence π_t defined above is uniform over all possible correspondences. We proceed inductively. When $k_t = 1$ there is only one possible correspondence and so the claim is obvious.

Now assume the claim is true when $k_t = m$. For any $i = 1, 2, ..., n$ define $\tau_i = \inf\{t : k_t = i\}$. Then the claim is true until a new element is marked, ie until τ_{m+1}. For convenience, denote $\alpha = \tau_{m+1} - 1$. At this time a new card is about to be marked, which means that $1 \notin I_\alpha$and $i \in I_\alpha \cup 1$. In words, the card in position 1 is unmarked at time α but is about to either switch with an already marked card or stay in place. Each of these events have the same $\frac{1}{n}$ probability of occurring. Furthermore, since π_α is uniform, the already marked cards are in uniform random order relative to each other. Since there are m previously marked cards, then after the transposition is applied the card in position 1 is equally likely to be in one of $m+1$ positions while the previously marked cards move to the first position with probability $\frac{1}{m+1}$ and stay in place with probability $\frac{m}{m+1}$. Since at time α the marked cards are randomly arranged among the m positions in I_α, it follows that the cards in J_α are in any of the $m+1$ positions of $I_{\tau_{m+1}}$ with equal probability. Thus $\pi_{\tau_{m+1}}$ is in fact uniform over all possible correspondences. Until time τ_{m+2}, the only moves involving marked cards are those where either a marked card in the first position switches places with another marked card or a marked card in the first position switches places with an unmarked card. In the former case, since the cards are already in random order relative to each other, switching them does not change anything and so all maps remain equally likely. In the latter case, all that occurs is that some other position i replaces the place of 1 in I_t. Thus instead of $\pi_t(1)$ we have $\pi_t(i)$, however pi_t itself is still uniform. Thus at all times up until τ_{m+2} π_t is uniform as claimed.

The final points to note are that when $k_t = n$ then $\pi_t \in S_n$ and that the induction tells us that it is equally likely to be any permutation. Also π_t and σ_t are equal by definition whenever both are defined. Since σ_t is the inverse of the state of the chain at time t, showing that π_t is uniform on S_n when $k_t = n$ is equivalent to showing that σ_t is uniform which is equivalent to showing that the chain is uniform since they are inverses. Thus the time is certainly a stationary time. Finally, given k_t, π_t is independent of t throughout. Thus it follows that π_τ is independent of τ and so by the same argument the state of the chain at time τ is independent of the value of τ. Thus τ is indeed a strong stationary time. □

As an example of how the time works consider the case $n = 3$. Initially the cards are in the order $(1, 2, 3_m)$ with the 'm' indicating that the third card starts off being marked. The first and second cards continue to get switched with no change in the number of marked cards until one of two equally likely options occurs. Either the card in the first position stays and is marked, or it switches places with 3 and is marked. Whether the card in the second position is 1 or 2 depends on the parity of the time, for convenience call whatever that card is a. Then at the first time two cards are marked we have $(*_m, a, *_m)$. Note that the two elements that aren't a are equally likely to be in either order. Now the walk either stays in that order (if the card in the first position doesn't change) or it switches around the two marked cards, neither of which change the random relative order of the two marked cards. This continues until a is switched with the first card resulting in the state $(a, *_m, *_m)$. Now a gets marked in the next step no matter what happens. However since a is equally likely to be in any of the positions, and since the two cards are still in random order relative to each other, it follows that after this step all three cards will be marked and in uniform random order relative to each other, illustrating the time proven above.

We can bound the total variation distance by calculating $E(\tau)$ and using Markov's inequality as follows. The proof is taken from [7, section 5.1].

Lemma 10. *For the walk described above,*

$$d(t) = \max_{x\in\Omega} ||P^t(x,\cdot) - \pi||_{TV} \leq \frac{2n\log n - n - 2}{t}$$

implying that $t_{mix} \leq 8n\log n - 4n - 8$

Proof. We know that in general, $d(t) \leq \max_{x\in\Omega} P_x(\tau > t)$. Using Markov's inequality we see that $d(t) \leq \max_{x\in\Omega} P_x(\tau > t) \leq \frac{E(\tau)}{t}$ and so it is enough to find $E(\tau)$. To do this we define $\tau_2, \tau_3, .., \tau_n$. Let τ_2 be the first time that there are two marked cards. Then for $k > 2$ let τ_k be the number of steps after the $k-1^{st}$ card is marked until the k^{th} card is marked. Thus $\tau = \tau_2 + \tau_3 + ... + \tau_n$ and so $E(\tau) = E(\tau_2) + E(\tau_3) + ... + E(\tau_n)$. Let $\tau_k = \tau'_k + \tau''_k$ where τ'_k is the number of steps after the $k-1^{st}$ card is marked until $1 \notin I_t$ and τ''_k is the number of steps after τ'_k stops counting until $i \in I_t \cup \{1\}$. where i is the position picked at the beginning of each move. These events are chosen because in order for a new card to be marked we must have $1 \notin I_t$ and $i \in I_t \cup \{1\}$. Since after a card is marked the newly marked card will be in position 1, an unmarked card must first be moved into position 1 for the next card to be marked. Once an unmarked card is placed in position 1, it will remain there until it is marked so our choice of events makes sense.

For $k = 2$, $E(k'_2) = 0$ since we start out with an unmarked card in the first position. Now notice that for $k > 2$ after time $\tau_2 + \ldots + \tau_{k-1}$ there are $n-k+1$ unmarked card remaining so the probability that the first card is replaced with an unmarked card is $\frac{n-k+1}{n}$. This is a geometric random variable and so $E(\tau'_k) = \frac{n}{n-k+1}$. Similarly, $|I_t \cup \{1\}| = k$ for $\tau_2 + \ldots + \tau_{k-1} \leq t < \tau_2 + \ldots + \tau_k$ and so the probability that $i \in I_t \cup \{1\}$ is $\frac{k}{n}$ meaning that $E(\tau''_k) = \frac{n}{k}$. Thus $E(\tau_2) = \frac{n}{2}$ while for $k > 2$, $E(\tau_k) = E(\tau'_k) + E(\tau''_k) = \frac{n}{n-k-1} + \frac{n}{k}$. Then

$$E(\tau) = (\frac{n}{n-2} + \frac{n}{n-3} + \ldots + \frac{n}{1}) + (\frac{n}{3} + \frac{n}{4} + \ldots + \frac{n}{n}) + \frac{n}{2} = 2n\sum_{j=1}^{n}\frac{1}{j} - n - \frac{n}{n-1} - 1$$

For large n we can make the approximation $E(\tau) = 2n\log n - n - 2$. □

A bound of the same order is implied by [5].

While this idea of marking cards is useful, it needs to be adapted in walks where the generators are not all equally likely. For example, consider the walk on S_n generated by the transpositions $(1,i)$, $i = 1, 2, ..., n$, where $(1,i)$ has probability p_i of occurring. The time above is no longer valid since its proof relied on the fact that all the permutations were equally likely to occur. To fix this, we change the way the cards are marked. Instead of marking them whenever we can, we mark them with probability $\frac{p_{I_t}}{p_i}$, where $p_{I_t} = \min_{i\in I_t\cup\{1\}} p_i$. This new time comes from [7, section 5.6].

Theorem 11. *Let* $I_1 = J_1 = \{n\}, k_1 = 1, t = 1,$ *and* $\sigma_1 = id$. *While* $k_t < n$

- *Pick a position i uniformly at random.*

- *If* $1 \notin I_t$ *and* $i \in I_t \cup \{1\}$, *then with probability* $\frac{p_{I_t}}{p_i}$ *do the following:*
  - *Letting* $I'_t = I_t \cup \{1\}$, $J_{t+1} = J \cup \sigma_t(1)$ *and* $k_{t+1} = k_t + 1$
  - *Let* $I_{t+1} = (1, i)I'_t$
- *Otherwise, let* $J_{t+1} = J_t$, $I_{t+1} = (1, i)I_t$, *and* $k_{t+1} = k_t$.
- *In all cases, let* $\sigma_{t+1} = (1, i)\sigma_t$.

Let τ *be the first time that* $k_t = n$. *Then* τ *is a strong stationary time for this walk.*

Proof. The proof is almost identical as before. The inductive step of the previous proof relied on the fact that given that a card was marked, the card was equally likely to go to any of the $m + 1$ positions in $I_\alpha \cup \{1\}$. Here, the fact that the card is not always marked compensates for the fact that the permutations are not all equally likely. Thus, if the card in position i is marked, the probability that the card in position 1 card will be placed in position i and marked is $p_i \cdot \frac{p_{I_t}}{p_i} = p_{I_t}$, and is the same for all $i \in I_\alpha \cup \{1\}$. All other portions of the previous proof hold and so we are done. □

Theorem 12. *For the walk described above,*

$$d(t) \leq \frac{1 - (n-1)p}{p^2}\left[2\log n - 1 - \frac{1}{n-1} - \frac{1}{n}\right]$$

where $p = \min_{i=1,\ldots,n} p_i$.

Note that $p \leq \frac{1}{n}$ since otherwise the sum of the probabilities would be greater than 1. Also note that in the case where $p_i = \frac{1}{n}$ for all i and so $p = \frac{1}{n}$ we get essentially the same bound as before. Unfortunately, this bound is not always sharp. For example if p_i is proportional to $\frac{1}{i^r}$ then the bound is off by a factor of about $\frac{n^2}{\log^2(n)}$ [7, Example 5.6.4].

Proof. The idea of the proof is similar to the unweighted case. As before, let τ_2 be the first time that there are two marked cards. For $k > 2$, let τ_k be the number of steps after the $k-1^{st}$ card is marked until the k^{th} card is marked so that $\tau = \tau_2 + \tau_3 + \ldots + \tau_n$. Unlike the unweighted case, we can not figure out the exact expectation of these random variables because we do not know the probabilities of the marked or unmarked cards. However, we can look at τ_k more or less as the sum of geometric random variables and use those to bound $E(\tau_k)$ for $k > 2$ as

$$E(\tau_k) \leq 1 \cdot \left[\frac{1}{(n-k+1)p} + \frac{1}{kp}\right] + \frac{1-np}{1-(n-1)p}E(\tau_k)$$

and for k=2 we get the bound

$$E(\tau_k) \leq 1 \cdot \left[\frac{1}{2p}\right] + \frac{1-np}{1-(n-1)p}E(\tau_k)$$

The first thing to notice is that for a card to be marked, there must first be an unmarked card in the first position, it must then be switched with an unmarked card, and finally it must actually be marked. The probability that a card is marked is $\frac{p_{I_t}}{p_i}$ where i is the position of the chosen card. The most that this can be is 1, a value that is certainly achieved when there is only 1 marked card. The least this value can be is $\frac{p}{1-(n-1)p}$, where $p = \min_{i=1,\dots,n} p_i$. This value will in fact be achieved if we choose a position i such that $p_i = (n-1)p$. Then p_j must equal p for all $j \neq i$ and so if the i^{th} card is chosen after another card has been marked, the upper value will be achieved.

Now consider the probability of an unmarked card replacing the marked card in the first position when there are $k-1$ marked cards. This probability must be at least $(n-(k-1))p$. The lower bound for this probability provides an upper bound of $\frac{1}{(n-k+1)p}$ for the expected time for this event to occur since these are geometric random variables. Of course when $k = 2$ there is already an unmarked card in the first position so the expected time is 0. Similarly, the probability of choosing a position in $I_t \cup \{1\}$ after we have placed an unmarked card in the first position is at least kp. Again this lower bound for the probability provides an upper bound of $\frac{1}{kp}$ for the expected time for this event.

Finally, we have two possibilities. With maximal probability 1, these two events lead to a card being marked. With maximal probability $1 - \frac{p}{1-(n-1)p}$ these events will occur but a card will not be marked. Then there will be a marked card in the first position and the time will effectively reset. Putting all of these together results in the upper bound given above. Rearranging we see that

$$E(\tau_k) \leq \frac{1-(n-1)p}{p^2}\left[\frac{1}{n-k+1} + \frac{1}{k}\right]$$

while

$$E(\tau_2) \leq \frac{1-(n-1)p}{p^2} \cdot \frac{1}{2}$$

Using the fact that $E(\tau) = E(\tau_2) + E(\tau_3) + \ldots + E(\tau_n)$ we see that

$$E(\tau) \leq \frac{1-(n-1)p}{p^2}\left[\frac{1}{2} + \frac{1}{3} + \ldots + \frac{1}{n} + \frac{1}{n-2} + \ldots + \frac{1}{1}\right] \approx \frac{1-(n-1)p}{p^2}\left[2\log n - 1 - \frac{1}{n-1} - \frac{1}{n}\right]$$

where the last approximation is valid for large n. □

Returning to the first example, what if in the top to random shuffle, the top card were placed in position i with probability p_i where p_i need not equal $\frac{1}{n}$. Using the idea of marking cards, we can find a stopping time for this problem. The idea behind this time is that we mark a card with a certain probability if it moves either to the bottom of the deck or it moves in front of a previously marked card. This argument is novel to this thesis.

Theorem 13. *Let* $p_t = min\{p_i : i+1 \in I_t \cup \{n+1\}\}$. *Let* I_t *be the set of positions of the marked cards at time t and let* J_t *be their values. Let* k_t *be the number of marked cards at time t. Set* $I_1 = J_1 = \emptyset$, $k_1 = 0$, *and* $\sigma_1 = id$. *While* $k_t < n$ *do the following:*

- *Choose $i = 1, 2, ..., $ or n with probability p_i.*
- *If $i + 1 \in I_t \cup \{n + 1\}$ then with probability $\frac{p_t}{p_i}$*
  - *Let $I'_{t+1} = (i\,, i-1, \dots, 1)I_t \cup \{i\}$ and $J_{t+1} = J_t \cup \{\sigma_t(1)\}$. Set $k'_{t+1} = k_t + 1$.*
- *Otherwise set $k'_{t+1} = k_t$.*
- *If $1 \in I'_{t+1}$, let $I_{t+1} = I'_{t+1} \backslash \{1\}$ and $k_{t+1} = k'_{t+1} - 1$. Otherwise let $I_{t+1} = I'_{t+1}$ and $k_{t+1} = k'_{t+1}$.*
- *In all cases, let $\sigma_{t+1} = (i\, i-1 ... 1)\sigma_t$.*

Let τ be the first time that $k_t = n$. Then τ is a strong stationary time.

Note that as before, σ_t is actually the inverse of the current state of the chain and thus we multiply on the left.

Proof. The proof begins similarly to the previous two. Let $\pi_t : J_t \to I_t$ be the map from the value of the marked card to its position at time t. That is, if at time t card 3 is marked and in position 7, then $\pi_t(3) = 7$. Note that unlike the previous example, π_t is not equivalent to σ_t but is instead equivalent to the state of the chain. As before, the claim is that given I_t and J_t π_t is uniform over all possible correspondences. We use induction on k_t; the cases where $k_t = 0$ or $k_t = 1$ are obvious.

Assume that the claim holds when $k = m$. Let $\tau_{m+1} = inf\{t : k_t = m + 1\}$ and let $\alpha = \tau_{m+1} - 1$. Then the probability that $\sigma_\alpha(1)$ is marked and goes to some l such that $l + 1 \in I_\alpha$ is $\frac{p_t}{p_l} \cdot p_l = p_t$. Similarly, the probability of $\sigma_\alpha(1)$ being marked and moving to the bottom of the deck is $\frac{p_t}{p_n} \cdot p_n = p_t$. Thus $\sigma_\alpha(1)$ is equally likely to be in any of $m+1$ positions. Now consider $j \in J_\alpha$. The inductive hypothesis is equivalent to saying that the probability that j is the l^{th} marked card from the top of the deck is $\frac{1}{m}$. In order to be the l^{th} marked card from the top at time τ_{m+1}, j must either be the l^{th} or $l - 1^{st}$ marked card from the top at time α. The probability that j is the l^{th} card at times α and τ_{m+1} is

$$P(j \text{ is the } l^{th} \text{ card at time } \alpha) \cdot P(\sigma_\alpha(1) \text{ does not move in front of } j) = \frac{1}{m} \cdot \frac{m + 1 - l}{m + 1}$$

The probability that j is the $l - 1^{st}$ card at time α and the l^{th} card at time τ_{m+1} is

$$P(j \text{ is the } l^{th} \text{ card at time } \alpha) \cdot P((\sigma_\alpha(1) \text{ moves in front of } j) = \frac{1}{m} \cdot \frac{l - 1}{m + 1}$$

Summing these two probabilities gives the probability that j is the l^{st} marked card from the top. Since the sum is $\frac{1}{m+1}$, it follows that j is equally likely to be in any of $m + 1$ positions. These calculations are valid for all $l \neq 1$. For the case $l = 1$, the only way for j to be the first marked card from the top at time τ_{m+1} is if it is the first card at time α and if $\sigma_\alpha(1)$ does not move in front of j. The probability of this $\frac{1}{m} \cdot \frac{m}{m+1} = \frac{1}{m+1}$ and so the conclusion remains valid. Finally, since all $m + 1$ of the marked cards are equally likely to be in any of the positions, even if a card is unmarked (because it is in the first position after the permutation is applied) the rest of the marked cards are still distributed uniformly at random and so π_t is still uniformly random and the inductive hypothesis holds at time τ_{m+1}.

Now consider a step where the top card is unmarked, and is put in some position i and is not marked. Then this does not affect the relative order of the marked cards and so π_t is still uniform after this. If the top card was marked and is unmarked at the end of a step in which a new card is not marked, then the hypothesis still holds because the rest of the marked cards are still in random order relative to each other and so π_t is still uniform. Thus at any step where a new card is not marked the inductive hypothesis continues to hold and so the induction is complete. But when $k_t = n$, π_t is simply the state of the chain and so it follows that τ is a stationary time. Since, given k_t, π_t is independent of t throughout the argument, it follows that τ is indeed a stationary time.

□

3 Examples

Finding a general upper bound for the time discussed in the last example of the previous chapter has proven difficult. In the previous examples it was possible to look at the time it took to mark the k^{th} card. However, in this case when a card fails to be marked it eventually forces the marked cards in front of it to become "unmarked" when they come back to the top. So instead I have concentrated on a few special cases and have used computer simulations to get an idea for whether this time might be reasonable.

For each deck size, I ran the simulation 100 times. The tables show the average of these trials. For some deck sizes running the simulation wasn't possible, indicated by a '-'.

Deck s	$\frac{1}{i}$ (forward)	$\frac{1}{i}$ (backward)	$\frac{1}{i^2}$	One weight 100x	One weight 10x	$n\log(n)$
10	172.6	40955.2	10238.2	474473.4	136.9	23.0
20	5809.78	-	186249598.4	-	382.0	59.9
30	331693.1	-	-	-	832.7	102.0
40	23204429	-	-	-	1204.2	147.6

The column headings correspond to the following probability distributions, where p_i is the probability of putting a card in the i^{th} position:

- $\frac{1}{i}$ forward refers to a distribution where p_i is proportional to $\frac{1}{i}$. Thus for large n, p_i is about $\frac{1}{i\log(n)}$.
- $\frac{1}{i}$ backward refers to a distribution where p_i is proportional to $\frac{1}{n-i+1}$. Thus for large n, p_i is about $\frac{1}{(n-i+1)\log(n)}$.
- $\frac{1}{i^2}$ refers to a distribution where p_i is proportional to $\frac{1}{i^2}$.

- ”One weight 100x” refers to a distribution where all the weights are the same except for $p_{n/2}$, which 100 times larger than the rest. Thus $p_i = \frac{1}{n+99}$ for all $i \neq \frac{n}{2}$, while $p_{n/2} = \frac{100}{n+99}$.
- ”One weight 10x” refers to to a distribution where all the weights are the same except for $p_{n/2}$, which 10 times larger than the rest. Thus $p_i = \frac{1}{n+9}$ for all $i \neq \frac{n}{2}$, while $p_{n/2} = \frac{10}{n+9}$.
- The final column gives the value of $n\log(n)$ for each deck size. Since $n\log(n)$ is the amount of time it takes for the analogous stopping time for the unweighted top to random shuffle to complete, this is a relevant comparison.

Since most of the difficulty in analyzing this stopping time comes from not knowing when cards do not get marked, the single weight distribution is most amenable to analysis, since there is only one spot where marking fails. In order to generalize, I will consider the following distribution: Let i be fixed and let $i \neq n$. Let $p_j = q$ for $j \neq i$ and let $p_i = aq$ for some constant a. This means that a card placed in the i^{th} slot is marked with probability $\frac{q}{aq} = \frac{1}{q}$. Now let κ_l be the time from when there are marked cards in all of the last l slots to when all the cards are marked. Finding the expected time until the last $n-i$ cards are marked is straightforward. Since a card placed anywhere other than position i will always be marked if the card behind it is marked or if it is th last card in the deck, then the last $n-i$ cards will be marked exactly when there are $n-i$ cards behind the initial bottom card. This is simply the sum of the expectation of $n-i$ geometric variables and equals $\sum_{j=1}^{n-i} \frac{1}{qj}$. Then we are left to compute $E(\kappa_{n-i})$. Fortunately, it is possible to set up a system of equations for $E(\kappa_j)$ for $j = n-i, ..., n-1$. For convenience I will denote $E(\kappa_j)$ by x_j. The system is as follows:

$$x_{n-i} = 1 + (n-i+1)qx_{n-i+1} + (a+i-2)qx_{n-i}$$

$$\text{For } 1 < j < i: \; x_{n-j} = 1 + (n-j+1)qx_{n-j+1} + (j-1)qx_{n-j} + (a-1)qx_{n-i}$$

$$x_{n-1} = 1 + (a-1)x_{n-i}$$

Each of these equations is reached via first step analysis. For the first equation, imagine that the last $n-i$ cards are marked. There is at least 1 more step needed, hence we add 1. For the future, there are four possibilities. The first is that the top card goes into one of the first $i-1$ spots and cannot be marked. The probability of this is $(i-1)q$. In this case, the expected number of steps remains x_{n-i}. The second possibility is that the card is placed in the i^{th} position and not marked. Then again the expected number of extra steps needed would be x_{n-i} since only the last $n-i$ cards are marked. A card placed in the ith slot is not marked with probability $\frac{a-1}{a}$ so the probability of this possibility is $\frac{a-1}{a}aq = (a-1)q$. The third possibility is that the card is placed in the i^th slot and marked. The probability of this is $\frac{1}{a}aq = q$ and the expected time remaining would be x_{n-i+1} since the last $n-i+1$ cards would now be marked. The final possibility is that the card is placed in one of the last $n-i$ slots and marked. The probability of this is $(n-i)q$ since anything placed in one of these positions would bet marked. The expected time remaining, given that this occurs, would also be x_{n-i+1}

For the second equations, suppose that the last j cards are marked. There are four possibilities for the future here as well, similar to those in the first case. The first is that the top card goes into one of the first $j-1$ spots. This has probability $(j-1)q$. The expected number of steps remaining would be x_{n-j} since the number of marked cards hasn't changed. The second possibility is that the card is placed in the i^{th} spot and not marked. Since the time does not complete until all cards are marked, we will not finish before this card, which is now in the i^{th} position, comes back to the top and is eventually placed somewhere where it is marked. That means that all marked cards in front of the i^{th} position must eventually move to the top, at which point they will be "unmarked." Thus, the expected time remaining is the same as if only the last $n-i$ cards were marked, ie x_{n-i}. The probability of this is $(a-1)q$. The third possibility is that the card is placed in the i^{th} slot and marked. Then the the last $n-j+1$ cards are all marked and the expected time remaining is x_{n-j+1}. The probability of this is q. Finally, the card could be placed in one of the $n-j-1$ spots in front of a marked card which is not the i^{th} one or it could be placed in the last position. Then the newly placed card will definitely be marked. This occurs with probability $(n-j)q$ and since the last $n-j+1$ cards will now the marked the expected time remaining is x_{n-j+1}.

For the final equation, suppose that all the cards but the top one are marked. If the card is placed anywhere but i we are automatically done and the expected time remaining is 0. If the card is placed in the i^{th} position then we are done unless the card is not marked. This occurs with probability $(a-1)q$. As in the second possibility above, this means that all the cards in front of the i^{th} position might as well not be marked, since they must be pushed to the front and eventually remarked. Thus the expected time remaining is x_{n-i} since only the last $n-i$ cards remain marked.

After simplifying the equations we can represent this system by the following matrix equation:

$$\begin{pmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ \frac{1-a}{n+a-(i-1)} & 1 & -\frac{n-(i-1)+1}{n+a-(i-1)} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ \frac{1-a}{n+a-2} & 0 & 0 & \cdots & 1 & -\frac{n-2+1}{n+a-2} \\ 1-a & 0 & 0 & \cdots & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_{n-i} \\ x_{n-i+1} \\ \vdots \\ x_{n-2} \\ x_{n-1} \end{pmatrix} = \begin{pmatrix} \frac{1}{n-i+1} \\ \frac{1}{(n-(i-1)+a)q} \\ \vdots \\ \frac{1}{(n-2+a)q} \\ 1 \end{pmatrix}$$

Unfortunately there is no simple general solution for this equation. Another way of attacking it is by looking at it as a skip-free chain. A skip free Markov chain is a chain whose state space is a subset of the nonnegative integers and where upward jumps may only be of unit size. In our case, we define Y_t to be the number of marked cards beyond the last $n-i$. Since it is possible to mark at most 1 card per a step this chain is indeed skip-free. Then we have the following theorem, found in [4]:

Theorem 14. *Consider an irreducible discrete-time skip-free chain y on the nonnegative integers with* $Y(0) = 0$. *Given d, let X (with state space {0,...,d}) be obtained from Y by making d an absorbing state, and let P denote the transition matrix for X. Then the hitting time of state d (same for X and Y has probability generating function*

$$f(u) = \prod_{j=o}^{d-1} \frac{(1-\theta_j)u}{1-\theta_j u}$$

where $\theta_0, \dots, \theta_{d-1}$ *are the d non-unit eigenvalues of P. In particular, if every* θ_j *is real and nonnegative, then the hitting time distribution is the convolution of Geometric*$(1-\theta_j)$ *distribution.*

In out case $d = i$. The transition matrix for $(a-1+n)Y$ is

$$\begin{pmatrix} a+i-2 & n-i+1 & 0 & \dots & 0 & 0 & 0 \\ a-1 & i-2 & n-(i-2) & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ a-1 & 0 & 0 & \dots & 1 & n-1 & 0 \\ a-1 & 0 & 0 & \dots & 0 & 0 & n \\ a-1 & 0 & 0 & \dots & 0 & 0 & n \end{pmatrix}$$

where there are $i+1$ rows with the last row being the one step probability distribution if we start with all the cards marked. The proof that this is the correct transition matrix is very similar to the proof of the recursive equations for the expectation above and will not be repeated here. Then if we make staying with all the cards marked an absorbing state, we get the following transition matrix for $(a-1+n)X$:

$$\begin{pmatrix} a+i-2 & n-i+1 & 0 & \dots & 0 & 0 & 0 \\ a-1 & i-2 & n-(i-2) & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ a-1 & 0 & 0 & \dots & 1 & n-1 & 0 \\ a-1 & 0 & 0 & \dots & 0 & 0 & n \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \end{pmatrix}$$

Unfortunately the eigenvalues for this matrix are quite complicated and do not have a simple general form. Even in the $i = 2$ case the eigenvalues are

$$\lambda_1 = 1$$

$$\lambda_2 \approx \frac{-(a+n-1)\sqrt{a^2+4an-4a-4n+4}+a^2+an-a)}{2(a^2+2an-2a+n^2-2n+1)}$$

$$\lambda_3 \approx \frac{(a+n-1)\sqrt{a^2+4an-4a-4n+4}+a^2+an-a}{2(a^2+2an-2a+n^2-2n+1)}$$

making using the theorem above untenable.

A final approach I tried is actually unrelated to strong stationary time. It is called comparison and uses eigenvalue techniques to compare the mixing times of two chains. There are several theorems relating the eigenvalues of a Markov chain to its mixing time. Many of them require that the chain be **reversible**, that is for all $x, y \in \Omega$, $\pi(x)P(x, y) = \pi(y)P(y, x)$. This name is fitting since it implies that

$$\pi(x_0)P(x_0, x_1) \dots P(x_{n-1}, x_n) = \pi(x_n)P(x_n, x_{n-1}) \dots P(x_1, x_0)$$

or equivalently that

$$P_\pi(X_0 = x_0, \dots, X_n = x_n) = P_\pi(X_0 = x_n, \dots, X_n = x_0)$$

which says that if the initial distribution is π, $(X_0, \dots, X_n)$ and $(X_n, \dots, X_0)$ have the same distribution.

It is known that for a reversible Markov chain, the eigenvalues of the transition matrix all have absolute value less than or equal to 1 and that 1 is itself an eigenvalue [6, section 12.1]. Then label the eigenvalues in decreasing order $1 = \lambda_1 > \lambda_2 \geq \dots \geq \lambda_{|\Omega|} \geq -1$ and define the **spectral gap** $\gamma = 1 - \lambda_2$. Also let $\lambda_* = \max\{|\lambda| : \lambda \text{ is an eigenvalue of } P, \lambda \neq 1\}$ and define $\gamma_* = 1 - \lambda_*$. Then from [6, section 12.2] we have

Theorem 15. *Let P be the transition matrix of a reversible, irreducible Markov chain with state space Ω and let $\pi_{min} = \min_{x \in \Omega} \pi(x)$. Then*

$$t_{mix}(\epsilon) \leq \log\left(\frac{1}{\epsilon \pi_{min}}\right)\frac{1}{\gamma_*}.$$

Unfortunately the comparison theorem gives a bound in terms γ rather than γ_*. We resolve this by using lazy chains. For a given transition matrix P, define $\hat{P} = \frac{1}{2}P + \frac{1}{2}I$. Then $\hat{P}$ is called the **lazy version** of P because it always has at least $\frac{1}{2}$ probability of staying. For such a chain $\gamma_* = \gamma$ [6, section 12.2]. It is useful to think of a lazy chain as first flipping a coin and then staying if the coin comes up heads and moving according to P otherwise. From this it is clear that a bound on the mixing time for a lazy chain also provides a bound for the non-lazy version since while the lazy chain only moves once every two steps on average, when the it moves it moves according to P. Thus the mixing time of the non-lazy chain is half that of the lazy chain. We will use lazy chain in our analysis to avoid problems.

The Comparison Theorem takes two chains with transition matrices P and $\tilde{P}$ and gives a lower bound for γ in terms of $\tilde{\gamma}$ and some other parameters. The theorem is presented here for the special case when the Markov chain is a random walk on a group. The proof of this case as well as the more general case may be found in [6, section 13.5]. It was originally proven in [3].

Theorem 16. *Let μ and $\tilde{\mu}$ be the increment distributions of two irreducible and reversible random walks on a finite group G. Let S and $\tilde{S}$ be the support sets of μ and $\tilde{\mu}$ respectively. For each $a \in \tilde{S}$,* *fix an expansion $a = s_1 \ldots s_k$ where $s_i \in S$ for $1 \le i \le k$. Write $N(s,a)$ for the number of times $s \in S$ appears in the expansion of $a \in \tilde{S}$ and let $|a| = \sum_{s \in S} N(s,a)$ be the total number of factors in the expansion of a. Let γ and $\tilde{\gamma}$ be their spectral gaps, respectively. Then*

$$\tilde{\gamma} \le B\gamma$$

where

$$B = max_{s \in S} \frac{1}{\mu(s)} \sum_{a \in \tilde{S}} \tilde{\mu}(a) N(s,a)|a|$$

Unfortunately we cannot use this technique directly on the top to random shuffle since it is not a reversible walk. Instead, we look at it together with its reversal. We call the matrix defined by

$$\hat{P}(x,y) = \frac{\pi(y)P(y,x)}{\pi(x)}$$

the transition matrix for the time reversal for P. It is easy to show that $\hat{P}$ is in fact a stochastic matrix and that it has the same stationary distribution as P [6, section 1.6]. From this definition it follows that a reversible chain is its own reversal. In the case of the top to random shuffle, since the stationary distribution is uniform, the transition matrix for the time reversal is simply $P^\top$. We can think of the reversal as choosing a random card according to p_i and then placing it at the top.

Lemma 17. *Let P be the transition matrix for a weighted top to random shuffle and let π be its stationary distribution. Then $P^\top P$ is a stochastic matrix representing a reversible Markov chain with stationary distribution π.*

Proof. First notice that $P^\top P(x,y) = \sum_{z \in \Omega} P^\top(x,z)P(z,y) \ge 0$ since all the terms are greater than or equal to zero. Also,

$$\sum_{y \in \Omega} P^\top P(x,y) = \sum_{y \in \Omega} \sum_{z \in \Omega} P^\top(x,z)P(z,y) = \sum_{z \in \Omega} P^\top(x,z) \sum_{y \in \Omega} P(z,y) = \sum_{z \in \Omega} P^\top(x,z) = 1$$

so the matrix is stochastic. $\pi P^\top P = \pi P = \pi$ so π is stationary for $P^\top P$. Since π is the uniform distribution, it follows from the comments above that the reversal of $P^\top P$ is $(P^\top P)^\top = P^\top P$ and so $P^\top P$ is in fact reversible. □

In order to ensure that the eigenvalues are all positive, we will work with the lazy version, $Q = \frac{1}{2}P^\top P + \frac{1}{2}I$. Half the time the walk will not move. The other half we notice that $P^\top P(x,y) = \sum_{z \in \Omega} P^\top(x,z)P(z,y)$ and so we can interpret this walk as first moving according to $P^\top$ and then moving according to P. In other words, when the walk represented by Q moves, we can think of this as taking a random card in position i and placing it at the top and then taking this random card and placing it at a random position j. Then a step in the walk is $(1,2,\ldots,i)(j,j-1,\ldots,1)$. For $i < j$ this is the same as $(j,j-1,\ldots,i)$, while for $i > j$ this is the same as $(j,j+1,\ldots,i)$. When $i = j$ we get the identity. In particular, as long as the increment distribution for P gives non-zero probability to $(j,j-1,\ldots,1)$ for all $j = 1,2,\ldots,n$, any adjacent transposition will have nonzero probability of occurring. This interpretation of the walk will be the basis of the comparison.

Theorem 18. *Let $Q = P^\top P$ as above and let R be the following walk on S_n: with probability $\frac{1}{2}$ do nothing, otherwise pick an adjacent pair of cards uniformly at random and transpose them. Then the probability of any adjacent transposition is $\frac{1}{2(n-1)}$. Let γ be the spectral gap of the former and $\tilde{\gamma}$ be the spectral gap of the latter. Let μ and $\tilde{\mu}$ be their increment distributions, respectively. Let ν be the increment distribution for P and let $\hat{S}$ be its support. Assume that $\nu(j,j-1,\ldots,1) > 0$ for all $j = 1,2,\ldots,n$. Then*

$$\tilde{\gamma} \leq \frac{1}{2p_{min}^2}\gamma$$

where $p_{min} = \min_{x\in\hat{S}} \nu(x)$. This implies that for Q

$$t_{mix}(\epsilon) \leq \log\left(\frac{n!}{\epsilon}\right)\frac{n^3}{p_{min}^2\pi^2}$$

Proof. We first find an upper bound on

$$B = max_{s\in S}\frac{1}{\mu(s)}\sum_{a\in\tilde{S}}\tilde{\mu}(a)N(s,a)|a|$$

as defined in the Comparison Theorem. Note that since we are comparing Q against R, the quantities with a tilde correspond to R. Let $a \in \tilde{S}$ where $\tilde{S}$ is the support of $\tilde{\mu}$. Since all adjacent transpositions are in S, per the comments above, $|a| = 1$ and $N(s,a) \leq 1$ for all $a \in \tilde{(S)}$, $s \in S$. However, no element of S is in the expansion of more than element one of $\tilde{S}$ meaning that

$$\max_{s\in S}\sum_{a\in\tilde{S}}\tilde{\mu}(a)N(s,a)|a| = \max_{s\in S, a\in\tilde{S}}\tilde{\mu}(a)N(s,a)|a| \leq \max_{a\in\tilde{S}}\tilde{\mu}(a) = \frac{1}{2}$$

since $\tilde{\mu}(id) = \frac{1}{2}$ while for all other $a \in \tilde{(S)}$, $\tilde{\mu}(a) = \frac{1}{2(n-1)}$. Then

$$B \leq max_{s\in S}\frac{1}{\mu(s)}max_{s\in S}\sum_{a\in\tilde{S}}\tilde{\mu}(a)N(s,a)|a| \leq \frac{1}{2}max_{s\in S}\frac{1}{\mu(s)} \leq \frac{1}{2p_{min}^2}$$

That $p_{min}^2 \leq \mu(s)$ for all $s \in S$ follows directly from the matrix multiplication $P^\top P$. Then the first part of the theorem follows by applying the comparison theorem.

It is known[8, Table 1] that $\tilde{\gamma} = \frac{1-\cos(\frac{\pi}{n})}{n-1}$. This implies that

$$\frac{1}{\gamma} \leq \frac{n-1}{2p_{min}^2(1-\cos(\frac{\pi}{n}))}$$

Recall the previously shown bound

$$t_{mix}(\epsilon) \leq \log\left(\frac{1}{\epsilon \pi_{min}}\right)\frac{1}{\gamma_*}.$$

$\pi_{min} = \frac{1}{n!}$ since $P^\top P$ is a random walk on S_n with uniform stationary distribution. Now using the fact that for lazy chains $\gamma = \gamma_*$, we have that

$$t_{mix}(\epsilon) \leq \log\left(\frac{n!}{\epsilon}\right)\frac{n-1}{2p_{min}^2(1-\cos(\frac{\pi}{n}))}$$

. Using a first order approximation for cos and simplifying, we see that for n large

$$t_{mix}(\epsilon) \leq \log\left(\frac{n!}{\epsilon}\right)\frac{n^3}{p_{min}^2 \pi^2}$$

as claimed. □

As an example, consider the top to random shuffle with weights proportional to $\frac{1}{i}$. Then using comparison we would get that

$$t_{mix}(\epsilon) \leq \log\left(\frac{n!}{\epsilon}\right)\frac{n^5}{\pi^2}$$

While this bound is not very good, it still seems to beat the time found via simulation for the stopping time discussed in section 2. However, for the single weight case where one card has weight 10 times that of the others, $p_{min} = \frac{1}{n+9} \approx \frac{1}{n}$. Then comparison gives the same bound as for when p_i is proportional to $\frac{1}{i}$ but the simulation suggests that the stopping time bound is in fact much better.

4 Conclusion

While finding an explicit bound for the stopping time for the weighted top to random shuffle proved difficult, the results of the computer simulation suggest that at least in the case where there is only one big weight this time would provide a good bound. It is not clear, however, that the bound from this time would be tight in other cases, reinforcing the difficulty in applying strong stationary techniques to random walks on groups with non-uniform increment distributions.

References

[1] David Aldous and Persi Diaconis. Shuffling cards and stopping times. *American Mathematical Monthly*, 93(5):333–348, 1986.

[2] David Aldous and Persi Diaconis. Strong uniform times and finite random walks. *Advances in Applied Math*, 8(1):69–97, 1987.

[3] Persi Diaconis and Laurent Saloff-Coste. Comparison theorems for reversible markov chains. *Annals of Applied Probability*, 3(3):696=730, 1993.

[4] James Allen Fill. On hitting times and fastest strong stationary times for skip-free and more general chains. *Journal of Theoretical Probability*, 22:587–600, 2009.

[5] L. Flatto, A. M. Odlyzko, and D. B. Wales. Random shuffles and group representations. *Annals of Probability*, 13:154–178, 1985.

[6] David A. Levin, Yuval Peres, and Elizabeth L. Wilmer. *Markov Chains and Mixing Times*. American Mathematical Society, 2008.

[7] Igor Pak. *Random Walks on Groups: Strong Uniform Time Approach*. PhD thesis, Harvard University, 1997.

[8] David B. Wilson. Mixing times of lozenge tiling and card shuffling markov chains. *The Annals of Applied Probability*, 14(1):274–325, 2004.