Disclaimer: *These notes have not been subjected to the usual scrutiny reserved for formal publications.*

In this lecture we will give one more example of coupling by analyzing Top-to-Random shuffling using the coupling method. Then, we talk about a new technique called strong stationary time and we will use it to analyze the riffle shuffle.

5.1 Top-to-Random Shuffling

Recall that in the Top-to-Random shuffling: Each time we pick up the top card in a deck and place it at uniformly random location in the deck. We want to design a coupling strategy to study the mixing time of this chain.

It turns out it is much easier to design a coupling for the following *inverse* chain: Each time choose a uniformly random card and place it at the top of the deck.

We claim that both of these walks have the same mixing time. This fact is in fact true for any Markov chain on groups. In this case note that we are considering a Markov chain on the symmetric group S_n. In general consider any group G and suppose we have a set of generators $\{g_1, \ldots, g_k\}$. In each time step we choose a generator from this set according to a probability distribution μ and we apply it to the current state. The inverse chain is defined as follows: Consider the set of generators $\{g_1^{-1}, \ldots, g_k^{-1}\}$. At any state x, choose a generator g_i^{-1} from the same distribution μ and apply it to x. Observe that both of these walks are doubly stochastic, so have a uniform stationary distribution.

Now for any path in the original walk $x \circ \sigma_1 \circ \cdots \circ \sigma_t$ we can construct a path in the inverse walk

$$f(x \circ \sigma_1 \circ \cdots \circ \sigma_t) = x \circ \sigma_1^{-1} \circ \cdots \circ \sigma_t^{-1}.$$

Observe that these two paths occur with exactly the same probability. Furthermore, if two paths $x \circ \sigma$ and $x \circ \tau$ reach the same state, so does the inverse path, $f(x \circ \sigma) = f(x \circ \tau)$. This means that f defines a bijection between the elements of the group, i.e., the states, in the original walk and the states of the inverse walk. Therefore, the distribution of the original walk at time t, $K^t(x, .)$, and the distribution of the inverse walk at time t, $\tilde{K}^t(x, .)$, are the same up to relabelling the states. But since the stationary distribution of the chain is uniform,

$$\|K^t(x, .) - \pi\|_{TV} = \|\tilde{K}^t(x, .) - \pi\|_{TV}.$$

Having this tool all we need to do to study the Top-to-Random shuffling is to study the mixing time of the Random-to-Top shuffling. Consider the following coupling: At any time both chains If X_t chooses the card labelled i to move to the top, Y_t also chooses the card labelled i and moves it to the top.

Note that this is a valid coupling. In particular, the chain X_t is exactly running the original Markov chain. Y_t is also running the original chain because it chooses each of the n card uniformly at random.

Now, let us study the time T_{XY} at which we get $X_t = Y_t$. Observe that whenever we choose the card labelled i, from now on this card will be in exactly the same location in both chains. So, we have $X_t = Y_t$ the first

time by which we have chosen each card at least once. This is again the coupon collector problem. So, this chain mixes in time $n \ln n$.

5.2   Strong Stationary Time

Definition 5.1 (Strong Stationary Time). *A strong stationary time for a Markov chain is a stopping time T such that X_T is stationary and independent of T, i.e.,*

$$\mathbb{P}\left[T = k, X_k = x\right] = \pi(x).$$

Note that strong stationary time is very desirable in practice because it can guarantee that the sample is chosen from the exact stationary distribution.

Strong Stationary time was introduced by Aldous and Diaconis [AD86].

Lemma 5.2. *If T is a strong stationary time, then for any starting state x,*

$$\Delta_x(t) \leq \mathbb{P}\left[T > t | X_0 = x\right]$$

Proof. First, observe that by definition of total variation distance,

$$\Delta_x(t) = \max_{A \subseteq \Omega} |K^t(x, A) - \pi(A)|.$$

Let T_x be the strong stationary time for the walk started at x. So, to prove the lemma it is enough to show that for any A, $|K^t(x, A) - \pi(A)| \leq \mathbb{P}\left[T_x > t\right]$. We can write

$$
\begin{aligned}
K^t(x, A) &= \mathbb{P}\left[X_t \in A\right] \\
&= \mathbb{P}\left[X_t \in A, T_x > t\right] + \mathbb{P}\left[X_t \in A, T_x \leq t\right] \\
&= \mathbb{P}\left[X_t \in A | T_x > t\right] \mathbb{P}\left[T_x > t\right] + \pi(A)\mathbb{P}\left[T_x \leq t\right] \\
&= \mathbb{P}\left[X_t \in A | T_x > t\right] \mathbb{P}\left[T_x > t\right] + \pi(A)(1 - \mathbb{P}\left[T_x > t\right]).
\end{aligned}
$$

Therefore,

$$|K^t(x, A) - \pi(A)| = \mathbb{P}\left[T_x > t\right] |\mathbb{P}\left[X_t \in A | T_x > t\right] - \pi(A)|.$$

But the quantity inside the aboslute value in the RHS is the difference of two probabilities so it is at most 1. This concludes the proof. $\square$

Let us start with a simple example of using the strong stationary time to bound the mixing time of an unbiased random walk on a cycle of length n. Consider an unbiased random walk on a cycle of length n with vertices $\{0, 1, 2, \ldots, n-1\}$ where at each vertex we stay with probability $1/2$ and we go to left/right with probability $1/4$. For simplicity, assume that $n = 2^k$. Suppose we start the walk at 0.

Let T_0 be the first time that we get distance 2^{k-2} away from zero. It follows that X_{T_0} is uniformly distributed on $2^{k-2} \cdot \{1, 3\}$. Following T_0, let T_1 be the first time that the walk goes distance 2^{k-3} away; so it is uniform on $2^{k-3} \cdot \{1, 3, 5, 7\}$ and so on. For example, if $n = 16$, at time T_0 the walker is uniformly distributed at $4, 12$ at time T_1 it is uniformly distributed at $2, 6, 10, 14$, and at T_2 it will be uniformly distributed on all odd vertices. Now, taking 1 more step the walker will be uniformly distributed on the cycle.

Now, let's bound the total variation distance. Recall that for $0 \leq i \leq k-2$, $T_i - T_{i-1}$ is the time that the walker moves 2^{k-2-i} steps away from its location at T_{i-1}. So, to compute $\mathbb{E}\left[T_i - T_{i-1}\right]$ we use the

following fact: A random walk on $\mathbb{Z}$ started at zero that stays with probability $1 - \theta$ and moves left/right with probability $\theta/2$ heats $\pm b$ in expected time b^2/θ. Therefore,

$$\mathbb{E}\left[T_i - T_{i-1}\right] = \frac{2^{2(k-2-i)}}{1/2}.$$

So, the final stopping time $1 + T_{k-2}$ has expectation

$$\mathbb{E}\left[1 + T_{k-2}\right] = 1 + \sum_{i=0}^{k-2} \frac{2^{2(k-2-i)}}{1/2} = \frac{2^{2k}}{8}(1 + 1/4 + 1/16 + \ldots) \le \frac{1}{6} \cdot 2^{2k} = \frac{1}{6}n^2.$$

Therefore, by Markov inequality,

$$\Delta_0(t) \le \frac{n^2/6}{t}$$

So, for $t = n^2$, the walk has total variation distance at most $1/6$.

5.3 Top-to-Random Shuffle

First, let us give another proof of the mixing time of the Top-to-Random shuffle. Fix the bottom cards of the deck. Without loss of generality say it is card n. At the beginning there is no card below n. Observe that whenever we put a card below n we put in a completely random position with respect to the rest of the card below n. So, we can card below n are in a uniformly random permutation. Let T be the first time that n becomes the top card of the deck. At this point we have a uniformly random permutation of all other cards. So, at time $T + 1$ we obtain a uniformly random permutation.

Let us study the distribution of T. We argue that T follows a coupon collector distribution. In particular, if we have k cards below n, the probability that the next card goes below n is exactly $(k + 1)/n$. So, $\mathbb{E}[T] = n \ln n$. Furthermore, $\mathbb{P}[T > n \ln n + cn] \le e^{-c}$. For this particular walk we can show that the mixing occurs precisely at time T, and indeed there is a *cutoff*. This is because the coupon collector process is strongly concentrated around its expected value. In particular, at any time $t < n \ln n - 10n$, the n-th card has not yet reach the top of the deck with high probability. Therefore, the set of possible

5.4 Riffle Shuffle

Suppose we have a deck of n cards and we want to generate a uniformly random permutation. First, let us recall the riffle shuffle. In each step we divide the deck of cards into two stacks, the top k cards and the bottom $n - k$ cards where k is chosen according to binomial distribution, $\frac{\binom{n}{k}}{2^n}$. Then, we choose an interleaving of the two stacks. That is we drop the cards at random from the two stacks one card at a time and the card is chosen with probability proportional to the size of the size of the stack at that time. Observe that the relative order of the cards from each stack are preserved under this operation. Note that there are $\binom{n}{k}$ possible interleaving of the two stacks. Furthermore, a simple calculation shows that each of these interleavings have the same probability. Putting the two steps together the probability that the cut followed by the interleaving generates a permutation is 2^n. It is not hard to see that any permutation can be generate in at most one way.

After a little bit of thinking one can construct the inverse of the riffle shuffle as follows:

- Label all cards with 0 and 1 independently and uniformly at random.

- place the 0 cards at the top while preserving their relative order.

- place the 1 cards at the bottom while preserving their relative order.

Note that the probability that we have k zeros follows exactly the binomial distribution.

Now, we study the mixing time of the inverse riffle shuffle. We design a strong stationary time. The idea is due to Aldous and Diaconis [AD86]. Suppose we have performed the inverse riffle shuffle t times. We define a label for each card by concatenating its 0/1 label in all t shuffles. Note that at any point in time the permutation can be obtained by sorting all cards with respect to their t bit strings. In particular, all cards with the same label will be ordered in the same order as the one we started with, but cards with different orders will be in a uniformly random order. The latter is because each time we assign a uniformly random bit to every card. Let T be the time where all cards have distinct labels. It follows that T is a strong stationary time.

So, it remains to bound the expected value of T. Fix any two cards. Observe that if we have t labels, the probability that these two cards have exactly the same label is 2^{-t}. Now, we can use a birthday paradox to analyze the expected value of T. Roughly speaking T should be big enough such that th eprobability that any two cards have the same label is less than $1/n^2$. Then, we can use union bound to prove the claim. In particular, we show
$$\mathbb{P}\left[T > 2\log_2(n) + c\right] \leq 2^{-c}.$$

If we run $2\log_2(n) + c$ inverse shuffles then any two cards will have the same label with probability $n^2 \cdot 2^{-c}$. Therefore, by union all cards will have different labels with probability 2^{-c}.

Bayer and Diaconis [BD92] found a way to explicitly calculate $\Delta_x(t)$ for the riffle shuffle for any value of t. They show that for $n = 52$, the total variation distance drops below $1/3$ in 7 interations. Note that they use

t	≤ 4	5	6	7	8	9
$\Delta(t)$	1.00	0.92	0.61	0.33	0.17	0.09

a spectral approach by studying all eigenvalues of the random walk matrix.

References

[AD86] David Aldous and Persi Diaconis. Shuffling cards and stopping times. *The American Mathematical Monthly*, 93(5):333–348, 1986. 5-2, 5-4

[BD92] Dave Bayer and Persi Diaconis. Trailing the dovetail shuffle to its lair. *The Annals of Applied Probability*, 2(2):294–313, 1992. 5-4