Lecture 5

Lecturer: Constantinos Daskalakis *Scribe: Alan Deckelbaum*

NOTE: The content of these notes has not been formally reviewed by the lecturer. It is recommended that they are read critically.

In today's lecture we discuss bounding mixing times in Markov chains via coupling and via strong stationary times.

1 Top-in-at-random Shuffle

Recall that in the top-in-at-random shuffle, we take the top card from the deck, place the card in a random position, and repeat. We will bound the mixing time using coupling. As in the last lecture, we have the bound

$$\Delta(t) \leq Pr_{\max x,y}[T_{xy} > t]$$

where T_{xy} is the time when the coupled chains starting from x and y first meet. (Recall that $\Delta(t) = \max_x ||P_x^{(t)} - \pi||_{TV}$.)

1.1 Inverse Top-in-at-random Shuffle

To bound the mixing time, we will look at the *inverse top-in-at-random shuffle.* Define this shuffle by picking a card c uniformly at random, placing the card on the top of the deck, and repeating. Consider the following coupling of the inverse shuffle:

- Set $X_0 = x$, $Y_0 = y$, where x, y are arbitrary.
- At each time t, we pick the same card c to use in both chains.

We notice that the time it takes for the two chains to meet is equivalent to the coupon collector problem, where each card value c is a coupon. (Once a card c has been chosen, that card will always be aligned in X_t and Y_t for all later times t.) Therefore, we have

$$Pr[T_{xy} > t] = Pr[\text{coupon collector with } n \text{ coupons takes} > t \text{ days to get all coupons}] \leq e^{-c}$$

where $t = n \log n + cn$.

By taking $c = \ln 2 + 1$, the right-hand-side is less than $\frac{1}{2e}$, and hence $\Delta^{inv}(t) \leq e^{-c} < \frac{1}{2e}$. Therefore, the mixing time of the inverse shuffle is at most $n \log n + O(n)$.

1.2 Relating the inverse and standard shuffles

We now need to relate $\Delta(t)$ to $\Delta^{inv}(t)$. We will do this by taking a more general view of shuffles. We will think of a shuffle as a random walk on the symmetric group S_n. The walk is defined by a set of generators $G = \{g_1, \ldots, g_k\}$ and a probability distribution P over G. (At each step of the walk, we pick a random g_i according to P and apply g_i to the current state.) Since each generator in G has an inverse in S_n, we see that the random walk process is doubly-stochastic. Therefore, the stationary distribution of the walk is uniform over S_n. We define the inverse shuffle by the generator set $G' = \{g_1^{-1}, \ldots, g_k^{-1}\}$ with the probability distribution $P'(g_i^{-1}) = P(g_i)$.

Claim 1. $\Delta(t) = \Delta^{inv}(t)$.

Proof: We can define a function f mapping paths in the original shuffle to paths in the inverse shuffle. In particular, we map a path $x \circ \sigma_1 \circ \sigma_2 \circ \cdots \circ \sigma_t$ (where the notation means that we start at x and then apply the permutations σ_i) to the path

$$f(x \circ \sigma_1 \circ \sigma_2 \circ \cdots \circ \sigma_t) = x \circ \sigma_t^{-1} \circ \sigma_{t-1}^{-1} \circ \cdots \circ \sigma_1^{-1}.$$

We notice that the probability of the original path is equal to the probability of the new path in the inverse shuffle (given that we start at the appropriate start state.) Thus, we observe that f is a 1-to-1 mapping from paths in the original shuffle to paths in the inverse shuffle and f is probability-preserving.

Furthermore, suppose that $x \circ \sigma_1 \circ \sigma_2 \circ \cdots \circ \sigma_t = x \circ \tau_1 \circ \tau_2 \circ \cdots \circ \tau_t$. It then follows (from the group laws of the form $(\sigma_1\sigma_2)^{-1} = \sigma_2^{-1}\sigma_1^{-1}$) that $x \circ \sigma_t^{-1} \circ \sigma_{t-1}^{-1} \circ \cdots \circ \sigma_1^{-1} = x \circ \tau_t^{-1} \circ \tau_{t-1}^{-1} \circ \cdots \circ \tau_1^{-1}$. Therefore, we see that f induces a 1-to-1 mapping $\hat{f}$ between the states reachable from x in the original and inverse chains. We have

$$P_x^{(t)}(y) = P_x^{inv,(t)}(\hat{f}(y)) \text{ for all } y.$$

Therefore, letting π be the uniform distribution, and noting again that $\hat{f}$ is a bijection, we have

$$||P_x^{inv,(t)} - \pi||_{TV} = ||P_x^t - \pi||_{TV}$$

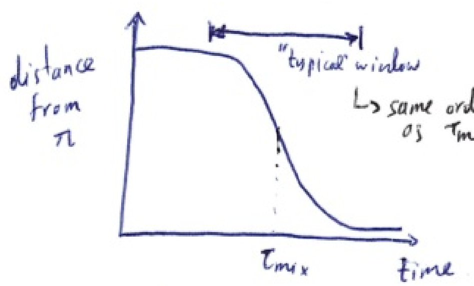
and hence

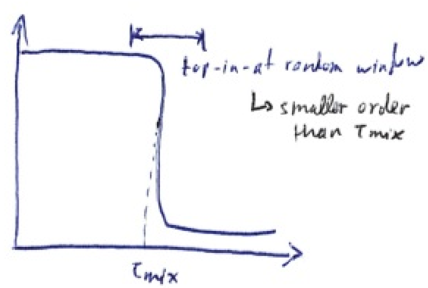
$$\Delta^{inv}(t) = \Delta(t).$$

$\square$ Therefore, we conclude that for the top-in-at-random shuffle, we have $\Delta(t) \leq e^{-c}$ for $t = n\log n + cn$, and hence $\tau_{mix} \leq n\log n + (\log 2 + 1)n$. Furthermore, to get the TV distance from the stationary distribution to be at most ϵ, we have $\tau(\epsilon) \leq n\log n + \log(1/\epsilon)n$.

We recall from last class the inequality $\Delta(t) \leq e^{-\lfloor \frac{t}{\tau_{mix}} \rfloor}$. This inequality implies, in our example, only that $\tau(\epsilon) \leq \log(1/\epsilon)\tau_{mix}$, which is a weaker bound than the upper-bound on $\tau(\epsilon)$ computed above.

This is an example of *sharp cutoff*, where the distance from stationarity drops off faster than we'd expect from the equation $\Delta(t) \leq e^{-\lfloor \frac{t}{\tau_{mix}} \rfloor}$ alone.





We will show that the following two properties hold for any function $\alpha(n)$ which goes to ∞ as $n \to \infty$:

- $\Delta(n\log n + \alpha(n)n) \to 0$ as $n \to \infty$. This follows immediately from $\Delta(t) \leq e^{-c}$ for $t = n\log n + cn$.
- $\Delta(n\log n - \alpha(n)n) \to 1$ as $n \to \infty$. We will show a certificate for this inequality. Look at the bottom k cards $C_1, \ldots, C_k$ in X_0. We will denote by A_k the event "$C_1 < C_2 < \cdots < C_k$", where $<$ denotes "below." Observe that in the stationary distribution, $\pi(A_k) = \frac{1}{k!}$, while we have $P_{X_0}^{(0)}(A_k) = 1$. We now see that

$$Pr[A_k \text{ at time } t] \geq Pr[C_k \text{ hasn't been reinserted by time } t]$$
$$= Pr[T_k + T_{k+1} + \cdots + T_{n-1} + 1 > T]$$

where T_i is the time for card k to go form position i to position $i-1$, where the final $+1$ corresponds to the one step to reinsert the card once its reached the top of the deck. Notice that T_i is a geometric random variable with parameter i/n. Therefore, we observe that the expression "$T_k + T_{k+1} + \cdots + T_{n-1} + 1 > t$" is equivalent to the event that a coupon collector is missing k coupons after t days.

Exercise (0.5pts): Show that the probability that, after $n\log n - \alpha(n)n$ days, the coupon collector still needs k coupons approaches 1 as $n \to \infty$ (for fixed k.)

This implies that $||P_x^{n\log n - \alpha(n)n} - \pi||_{TV} \geq 1 - \frac{1}{k!} - o(1)$.

2 Mixing of Riffle Shuffle via Strong Stationary Times

In this section we define the concept of a stopping time and a strong stationary time, and then we use strong stationary times to bound the mixing times of Markov chains.

Definition 1. *A stopping time is a random variable $T \in \mathbb{N}$ such that the event $\{T = t\}$ depends only on $X_0, \ldots, X_t$.*

Definition 2. *A stopping time is called a strong stationary time if, for all x, z:*

$$Pr[X_t = z | T = t; X_0 = x] = \pi(z).$$

Intuitively, a strong stationary time is a time such that, conditioned on the time happing, we are in the stationary distribution. We can use strong stationary times to bound mixing times by the following simple lemma.

Lemma 1. *If T is a strong stationary time, then for all x*

$$\Delta_x(t) \leq Pr[T > t | X_0 = x].$$

Furthermore,

$$\Delta(t) \leq \max_x Pr[T > t | X_0 = x].$$

The proof of the above lemma follows immediately from the fact that, once T occurs, we are in the stationary distribution.

Recall that in the riffle shuffle we split the deck into left and right stacks (denoted L and R), where the size of the left stack is chosen from the binomial distribution with parameters $(n, 1/2)$. (The right stack consists of the remaining cards.) We then take a random interleaving of the two stacks, and repeat the process.

From the properties of the binomial distribution, we know that $Pr[|L| = k] = \frac{\binom{n}{k}}{2^k}$. When $|L| = k$ we know that there are $\binom{n}{k}$ possible ways of interleaving of L and R. Therefore, each pair of a cut and an interleaving of that cut has probability of $\frac{1}{2^n}$ of being selected.

We now define the *inverse riffle shuffle*. In this shuffle, we label the cards with 0 or 1 independently with probability 1/2. We then pull all 0-labeled cards out of the deck (preserving their relative order) and place these cards on top of the deck. We observe that, indeed, each move[1] in the standard riffle shuffle has a corresponding inverse move with the same probability of $\frac{1}{2^n}$.

We will now use strong stationary times to analyze the mixing time of the inverse riffle shuffle. Consider running the inverse riffle shuffle for t steps, keeping track of the label that each card receives in each step. After t steps, each card will have been assigned a t-bit number. We define a stopping time by

$$T = \min\{t | \text{ all labels are distinct}\}.$$

We claim that T is a strong stationary time. When T occurs, the relative order of any two cards is random: the cards must have labels differing in some bit, and the value of the rightmost differing bit determines the relative order of the cards. More precisely, as long as the labels are distinct, the final order of the cards depends only on the labels and not on the original arrangement. Since the labels are all chosen randomly, we conclude that, conditioned on T, the deck is in a random permutation regardless of the initial configuration.

We will now bound the mixing time of the riffle shuffle, using the fact that

$$\Delta^{inv}(t) = \Delta(t) \leq Pr[T > t].$$

Analyzing the probability that $T > t$ is equivalent to the standard "birthday problem." In particular, after t steps, there are n "people" (cards) with 2^t possible "birthdays" (labels). We want to find the

[1]By "move" we mean the pair consisting of the way the deck was divided into L and R as well as the method of interleaving the two halves.

probability that all of the birthdays are distinct. It is a standard result that, with k people and ck^2 possible birthdays:

$$Pr[\text{some pair has the same birthday}] \rightsquigarrow 1 - e^{-\frac{1}{2c}} \quad \text{as } k \to \infty.$$

Therefore, we want to choose c such that $1 - e^{-\frac{1}{2c}} < \frac{1}{2e}$ and pick a t such that $2^t > cn^2$. We can achieve this with $t = 2\log_2 n + O(1)$. Therefore, the mixing time of the riffle shuffle is $\tau_{mix} \leq 2\log_2 n + O(1)$. This indeed has the correct asymptotic behavior, as a result of Aldous proves that $\tau_{mix} \approx \frac{3}{2}\log_2 n$.

Bayer and Diaconis have performed a numerical computation of the mixing behavior for any t and n. (Note that this is not a trivial computation, since the number of arrangements of the deck can be very large.) For a standard size deck, their results are as follows:

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