CMPUT 675: Randomized Algorithms **Fall 2005**

Lecture 14: Oct 25

Lecturer: Mohammad R. Salavatipour *Scribe: Zac Friggstad*

14.1 Example 1: Card Shuffling

Recall the Fundamental Theorem for Markov Chains.

Theorem 14.1 *Any finite, irreducible, ergodic Markov chain M with n states has the following properties.*

1. *M has a unique stationary distribution π.*
2. *For all $1 \leq i, j \leq n$,*
$$\lim_{t\to\infty} P^t_{j,i} = \pi_i = \frac{1}{h_{i,i}}$$
where $P^t_{j,i}$ is the probability of being in state i after t steps from state j and $h_{i,i}$ is the hitting time from state i to itself.

Consider the following process. Given a deck of n cards, a step is taken by choosing any card uniformly at random, removing it from the deck and placing it on top. This defines a Markov chain with $n!$ states where a state is some permuation of the cards. By choosing the top card, the state does not change in a step so the Markov chain is ergodic. Given permuations Π_1 and Π_2, Π_1 can reach Π_2 in at most n steps. This can be accomplished by selecting the bottom card of Π_2, finding it in Π_1 and moving that card to the top of Π_1. Remove the bottom card from Π_2 and repeat the process until all the cards from Π_2 are exhausted. This produces the permutation Π_2 from Π_1. Since each state can reach any state, then this Markov chain is irreducible. Then by the Fundamental Theorem for Markov Chains, this Markov chain has a unique stationary distribution. Let π_x be the probability associated with state x in the stationary distribution π. Since each state has n neighbors (one for each card choice) and the probability of a neighbor going to state x is the same for all neighbors (namely $\frac{1}{n}$), then the following system of equations is satisfied.

$$\sum_{y\in N(x)} \frac{1}{n}\pi_y = \pi_x \tag{14.1}$$
$$\sum_{x\in V} \pi_x = 1 \tag{14.2}$$

where $N(x)$ denotes the neighboring states of x. Notice that $\pi_x = \frac{1}{n!}$ satisfies this since

$$\sum_{y\in N(x)} \frac{1}{n}\pi_y = |N(x)|\,\frac{1}{n}\cdot\frac{1}{n!} = \frac{n}{n\cdot n!} = \frac{1}{n!} = \pi_x$$

and

$$\sum_{x\in V} \pi_x = |V|\,\frac{1}{n!} = \frac{n!}{n!} = 1$$

Definition 14.2 *The* variant distance *of distributions* D_1 *and* D_2 *on a countable set* S *is*

$$||D_1 - D_2|| = \frac{1}{2}\sum_{x \in S} |D_1(x) - D_2(x)|$$

So if we say variant distance between two distributions is at most ε it means any permunation of the cards has probability at most $\frac{1}{n!} + \varepsilon$ of occuring. Our goal is to find $t = t(\varepsilon)$ steps needed to be ε-close to π for a given $\varepsilon > 0$.

Definition 14.3 *Say* P_x^t *is the distribution of states after* t *steps if we start at state* x *in a Markov chain* M*. Define*

$$\begin{aligned} \Delta_x(t) &= ||P_x^t - \pi|| \\ \Delta(t) &= \max_x \Delta_x(t) \\ \tau_x(\varepsilon) &= \min\{t \mid \Delta_x(t) \leq \varepsilon\} \\ \tau(\varepsilon) &= \max_x \tau_x(\varepsilon) \end{aligned}$$

So $\tau(\varepsilon)$ *defines the minimum number of steps required to be within distance* ε *from* π *from any start state* x*.* $\tau(\varepsilon)$ *is also called the* mixing time *of* M*. We say* M *is* rapidly mixing *if* $\tau(\varepsilon)$ *is polynomial in the input size* n *and in* $\ln(\frac{1}{\varepsilon})$*.*

Typically we want to bound the mixing time (or say that a chain is rapidly mixing). A common technique for this is copuling. The idea is to start with two copies of the Markov chain and run the simultaneously. We make the moves at each step of these chain in such a way that these two chains get closer and closer to each other but still follow a Markov chain. If after some time they become identical it must be a stationary distribution.

Definition 14.4 *A* coupling *of a Markov chain* M *with states* S *is a Markov chain* $Z_t = (X_t, Y_t)$ *on* $S \times S$ *such that*

$$\Pr(X_{t+1} = x' \mid Z_t = (x, y)) = \Pr(M_{t+1} = x' \,|M_t = x)$$

and

$$\Pr(Y_{t+1} = y' \mid Z_t = (x, y)) = \Pr(M_{t+1} = y' \,|M_t = y)$$

Lemma 14.5 *(from* [1]*) For any* $A \subseteq S$*, let* $D_i(A) = \sum_{x \in A} D_i(x)$ *for* $i = 1, 2$*. Then*

$$||D_1 - D_2|| = \max_{A \subseteq S} |D_1(A) - D_2(A)|\,.$$

Proof: Let $S^+ \subseteq S$ be the set of states such that $D_1(x) \geq D_2(x)$, and let $S^- \subseteq S$ be the set of states such that $D_2(x) > D_1(x)$. Then,

$$\max_{A \subseteq S} D_1(A) - D_2(A) = D_1(S^+) - D_2(S^+)$$

and

$$\max_{A \subseteq S} D_2(A) - D_1(A) = D_2(S^-) - D_1(S^-).$$

Since $D_1(S) = D_2(S) = 1$, we have

$$D_1(S^+) + D_1(S^-) = D_2(S^+) + D_2(S^-) = 1,$$

which yields

$$D_1(S^+) - D_2(S^+) = D_2(S^-) - D_1(S^-).$$

Therefore

$$\max_{A \subseteq S} |D_1(A) - D_2(A)| = \left|D_1(S^+) - D_2(S^+)\right| = \left|D_1(S^-) - D_2(S^-)\right|.$$

Finally, since

$$\left|D_1(S^+) - D_2(S^+)\right| + \left|D_1(S^-) - D_2(S^-)\right| = \sum_{x \in S} |D_1(x) - D_2(x)| = 2\,||D_1 - D_2||,$$

we have

$$\max_{A \subseteq S} |D_1(A) - D_2(A)| = ||D_1 - D_2||.$$

■

Lemma 14.6 (Coupling Lemma) *Let $Z_t = (X_t, Y_t)$ be a coupling for a Markov chain M with state space S. If for a time T and for all $x, y \in S$,*

$$\Pr(X_T \neq Y_T \mid X_0 = x \text{ and } Y_0 = y) \leq \varepsilon$$

then $\tau(\varepsilon) \leq T$.

Proof: (from [1]) Consider the coupling when Y_0 is chosen according to the stationary distribution and X_0 takes on any arbitrary value. For the given T and ε and for any $A \subseteq S$,

$$\begin{aligned}
\Pr(X_T \in A) &\geq \Pr((X_T = Y_T) \cap (Y_T \in A)) \\
&= 1 - \Pr((X_T \neq Y_T) \cup (Y_T \notin A)) \\
&\geq (1 - \Pr(Y_T \notin A)) - \Pr(X_T \neq Y_T) \\
&\geq \Pr(Y_t \in A) - \varepsilon \\
&= \pi(A) - \varepsilon.
\end{aligned}$$

Here the second line follows from the union bound. For the third line, the fact that $\Pr(X_T \neq Y_T) \leq \varepsilon$ for any initial states X_0 and Y_0 is used. In particular, this will hold when Y_0 is chosen according to the stationary distribution. The last line follows from $\Pr(Y_T \in A) = \pi(A)$ since Y_T is distributed according to the stationary distribution if Y_0 is. Repeating the same argument for the set $S - A$ shows that $\Pr(X_t \in A) \geq \pi(S - A) - \varepsilon$ or $\Pr(X_t \in A) \leq \pi(A) + \varepsilon$. From this it follows that

$$\max_{x,A} \left|p_x^T(A) - \pi(A)\right| \leq \varepsilon,$$

so by Lemma 14.5 the variation distance from the stationary distribution after the chain runs for T steps is bounded above by ε ■

Consider the following coupling of the Markov chain for the cards example. Choose a j between 1 and n uniformly at random. Move the j'th card in X_t to the top of the deck to get X_{t+1}. Find the same card in the other deck and move that card to the top to get Y_{t+1}. This is a valid coupling of the Markov chain because in both chains, each card is placed on top with probability $\frac{1}{n}$. It is easy to see that the two chains get closer. Once a card moves to the top then it will always have the same position in both decks. The decks are guaranteed to be identical when every card has been drawn at least once. So we have to bound the number of time we need to draw cards such that every card is drawn at least once; this is just the coupon collector's problem. If $n \ln n + cn$ steps are taken, then the probability that a fixed card is not drawn is

$$\left(1 - \frac{1}{n}\right)^{n \ln n + cn} \leq e^{-(\ln n + c)} = \frac{e^{-c}}{n}.$$

By the union bound, the probability that any card is not drawn is upper bound by e^{-c}. So

$$e^{-c} \leq \varepsilon \rightarrow -c \leq \ln \varepsilon \rightarrow c \geq \ln\left(\frac{1}{\varepsilon}\right).$$

The expression $n \ln n + n \ln(\frac{1}{\varepsilon})$ is polynomial in n and $\ln(\frac{1}{\varepsilon})$ so the Markov chain is rapidly mixing.

14.2 Example 2: Independent Sets of Size k

Given a graph $G = G(V, E)$ of n vertices with maximum degree Δ, the goal is to select, uniformly at random, a sample among all independent sets of size $k \leq \frac{n}{3(\Delta+1)}$. Notice that we can always find one of size at least $\frac{n}{\Delta+1}$ by picking an arbitrary vertex, removing all adjacent vertices (at most Δ) and repeating until no more vertices can be found.

Consider the following process. Start with an arbitrary independent set of size k by performing a greedy selection strategy as stated above for k steps. At each step X_t, choose $v \in X_t$ and $w \in V$ uniformly at random and perform the operation $exchange(v, w, X_t)$ described as

$exchange(v, w, X_t)$
  if $w \notin X_t$ and $(X_t - v) \bigcup \{w\}$ is an independent set
    then $X_{t+1} \leftarrow (X_t - v) \bigcup \{w\}$
  else $X_{t+1} \leftarrow X_t$.

Define the states of the Markov chain for this problem to be independent sets of size k. Since there are $\binom{n}{k}$ subsets of V of size k then there can be no more than this number of independent sets of size k, so this Markov chain is finite. If $w \in X_t$, which can happen with probability $\frac{k}{n}$, then $X_{t+1} = X_t$ so this Markov chain is ergodic.

To show that this is irreducible, consider states S_1 and S_2. Fix a $v \in S_1 - S_2$ and a $w \in S_2 - S_1$. Our goal is to move our selection of v to w using legitimate moves and not moving any $u \in S_1 \bigcap S_2$. For each $u \in S_1$ that is adjacent to w ($u \notin S_2$ necessarily), move all vertices in S_1 adjacent to w to an arbitrary part of the graph such that the independent set property is retained and this new location is not adjacent to w. This can be done since there are k vertices selected and 1 vertex that we are after so there are at most

$$\left(\frac{n}{3(\Delta+1)} + 1\right)(\Delta + 1) = \frac{n}{3} + \Delta + 1$$

vertices that cannot be used for this temporary move. Thus there are at least

$$n - \left(\frac{n}{3} + \Delta + 1\right) = \left(\frac{2n}{3} - (\Delta + 1)\right)$$

vertices that can be used. Since $\Delta + 1 \leq \frac{n}{3}$ (otherwise $k = 0$) then there is a vertex that can be used for this intermediate move. Once all of the vertices in S_1 adjacent to w are moved, then v can be moved to w. Repeat this process for all vertices $v \in S_1 - S_2$ until $S_1 = S_2$. Therefore, this Markov chain is irreducible and has a stationary distribution π.

The goal is now to show that this Markov chain is rapidly mixing. Consider the coupling $Z_t = (X_t, Y_t)$ for the Markov chain where

$$\begin{aligned} X_t &= \{v_t^1, \ldots, v_t^i, v_t^{i+1}, \ldots, v_t^k\} \\ Y_t &= \{v_t^1, \ldots, v_t^i, u_t^{i+1}, \ldots, u_t^k\} \end{aligned}$$

for some $0 \le i \le k$ where $v_t^j \neq u_t^l$ for any $i < j, l \le k$ (order the elements so the same ones are "on the left"). Consider the following selection strategy for this coupled Markov chain.

step(X_t, Y_t)
  choose $v \in X_t$ and $w \in V$ uniformly at random
  exchange(v, w, X_t)
  if $v \in Y_t$ **then** *exchange*(v, w, Y_t)
  else say $v = u_t^{i+\alpha}$, **then** *exchange*$(u_t^{i+\alpha}, w, Y_t)$.

Let $d_t = |X_t - Y_t|$ be the difference in sets after t steps. Some facts about d_t are as follows.

- $d_t = 0 \to d_{t+1} = 0$.
- $|d_t - d_{t+1}| \le 1$.
- If $d_{t+1} = d_t + 1$ then $v \in X_t \bigcap Y_t$ and w is a neighbor of a vertex in $(Y_t - X_t) \bigcup (X_t - Y_t)$. The first happens with probability $\frac{k-d_t}{k}$ and the second with probability at most $\frac{2d_t(\Delta+1)}{n}$. Therefore

$$\Pr(d_{t+1} = d_t + 1 \mid d_t > 0) \le \frac{k - d_t}{k} \cdot \frac{2d_t(\Delta + 1)}{n}.$$

- If $d_{t+1} = d_t - 1$ then $v \notin Y_t$ and w is not a vertex or neighbor of $(X_t \bigcup Y_t) - \{v_t^{i+\alpha}, u_t^{i+\alpha}\}$. The first happens with probability $\frac{d_t}{k}$ and the second with probability at least $\frac{n-(k+d_t-2)(\Delta+1)}{n}$. Therefore

$$\Pr(d_{t+1} = d_t - 1 \mid d_t > 0) \ge \frac{d_t}{k} \cdot \frac{n - (k + d_t - 2)(\Delta + 1)}{n}.$$

Thus,

$$\begin{aligned}
\mathrm{E}[d_{t+1} \mid d_t] &= d_t + \Pr(d_{t+1} = d_t + 1) - \Pr(d_{t+1} = d_t - 1) \\
&\le d_t + \frac{k - d_t}{k} \cdot \frac{2d_t(\Delta + 1)}{n} - \frac{d_t}{k} \cdot \frac{n - (k + d_t - 2)(\Delta + 1)}{n} \\
&\le d_t \left(1 - \frac{n - (3k - d_t - 2)(\Delta + 1)}{kn}\right).
\end{aligned}$$

Using the conditional expectation equality, we have

$$\mathrm{E}[d_{t+1}] = \mathrm{E}[\mathrm{E}[d_{t+1} \,|d_t]] \le \mathrm{E}[d_t] \left(1 - \frac{n - (3k - d_t - 2)(\Delta + 1)}{kn}\right).$$

So by induction we get

$$\mathrm{E}[d_t] \le d_0 \left(1 - \frac{n - (3k - d_t - 2)(\Delta + 1)}{kn}\right)^t \le e^{-t(n-(3k-3)(\Delta+1))/kn}.$$

Since $k \le \frac{n}{3(\Delta+1)}$ then

$$\lim_{t\to\infty} \mathrm{E}[d_t] = 0.$$

Finally, it can be verified algebraically that

$$\tau(\varepsilon) \le \frac{kn \ln(\frac{1}{\varepsilon})}{n - (3k - 3)(\Delta + 1)}$$

which is bound by a polynomial in n and in $\ln(\frac{1}{\varepsilon})$. Therefore, this Markov chain is rapidly mixing.

References

1 M. Mitzenmacher and E. Upfal *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*, Cambridge University Press, 2005