Lecture 25: November 29

Lecturer: Alistair Sinclair — *Based on scribe notes by: Omar Khan, Grant Schoenebeck*

25.1 Markov Chain Review

We begin by briefly reviewing background on Markov chains from the last lecture.

Let Ω be a (finite) state space. A Markov chain $(X_t)_{t=0}^{\infty}$ on Ω is specified by a stochastic matrix P such that $P(x,y) = \Pr[X_t = y | X_{t-1} = x]$. We write $p_x^{(t)}(y) = \Pr[X_t = y | X_0 = x]$, so that $p_x^{(t)} = p_x^{(0)} P^t$.

Theorem 25.1 (Fundamental Theorem of Markov Chains) *Provided that P is irreducible and aperiodic,*

$$p_x^{(t)}(y) \to \pi(y) \quad \text{as } t \to \infty,$$

where the stationary distribution $\pi(y)$ is the unique (normalized) vector such that $\pi P = \pi$.

Definition 25.2 *The* mixing time, *$\tau_{mix} = \min\{t : \Delta(t) \leq \frac{1}{2e}\}$, where $\Delta(t) = \max_x \|p_x^{(t)} - \pi\|$ is the variation distance from π at time t, maximized over initial states x.*

Recall that variation distance is defined as $\|\mu - \xi\| = \frac{1}{2}\sum_{x\in\Omega} |\mu(x) - \xi(x)| = \max_{A\subset\Omega} |\mu(A) - \xi(A)|$.

Recall also that for any $t > \tau_{mix} \cdot \lceil \ln \epsilon^{-1} \rceil$ we have $\Delta(t) \leq \epsilon$.

In general, our goal was to randomly sample elements of a large set Ω from a distribution defined implicitly by assigning a positive weight $w(x)$ to each $x \in \Omega$ and then normalizing. So, $\Pr[x \text{ is choosen}] = \frac{w(x)}{Z}$ where $Z = \sum_{x\in\Omega} w(x)$. However, Z is in general not known; in fact, often the goal is to compute Z (see the next lecture for examples).

Last time we looked at the example of shuffling cards by two different methods: the top-in-at-random shuffle and the riffle shuffle. (In these examples w is uniform, so π was also.) We used the idea of a strong stationary time to analyze the mixing time in both cases. However, in general strong stationary times are not available so we need more widely applicable techniques. In this lecture we examine the simplest of these, known as "coupling."

25.2 Coupling

Definition 25.3 *Let (X_t), (Y_t) be two copies of a Markov chain. A coupling of (X_t) and (Y_t) is a joint process (X_t, Y_t) such that*

1. *Marginally (i.e., viewed in isolation), (X_t) and (Y_t) are both copies of the original chain.*

2. $X_t = Y_t \Rightarrow X_{t+1} = Y_{t+1}$.

We shall see in a moment that, for any coupling, the number of steps until the two copies "meet" (i.e., $X_t = Y_t$) provides an upper bound on the mixing time. The simplest example of a coupling is to make (X_t) and (Y_t) evolve independently. However, such a coupling will not tend to make the two copies meet rapidly. The art in applications is to find a coupling that causes the two copies to meet as quickly as possible.

Definition 25.4 $T_{xy} = \min\{t : X_t = Y_t | X_0 = x, Y_0 = y\}$. *I.e., T_{xy} is the (random) time until two copies meet, starting in states x, y.*

The following result formalizes the idea that T_{xy} provides a bound on the mixing time.

Claim 25.5 *For any coupling, $\Delta(t) \leq \max_{x,y} \Pr[T_{xy} \geq t]$.*

Observation 25.6 *For any two random variables X, Y on a common probability space with distributions μ and ξ respectively, any joint distribution satisfies*

$$\Pr[X \neq Y] \geq \|\mu - \xi\|.$$

We "prove" Observation 25.6 with a proof-by-picture:

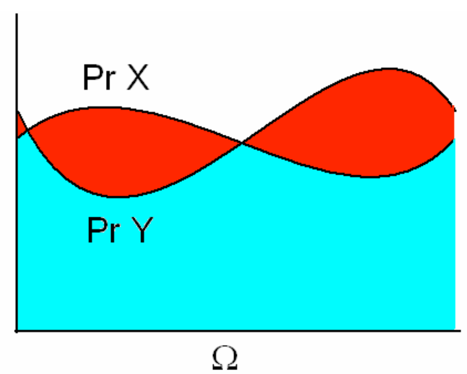

Figure 25.1: Example distributions for X and Y

According to the picture, the best we can do is to make the random variables equal on the overlapping region (below both curves) in the picture. I.e., for each z, we can make $X = Y = z$ with probability $\min\{\Pr[X = z], \Pr[Y = z]\}$. With the remaining probability, X and Y must be unequal. This probability is exactly half of the non-overlapping region (between the curves), which in turn is exactly the variation distance.

Note that this bound is tight, in the sense that there exists a coupling of X and Y s.t. equality is attained.

Proof:[of Claim 25.5]

$$\Delta(t) \equiv \max_x \|P_x^{(t)} - \pi\| \leq \max_{x,y} \|P_x^{(t)} - P_y^{(t)}\| \leq \max_{x,y} \Pr[X_t \neq Y_t | X_0 = x, Y_0 = y] \leq \max_{x,y} \Pr[T_{xy} \geq t].$$

In the second inequality here, we have used Observation 25.6. ■

Corollary 25.7 $\tau_{mix} \le 2e \max_{x,y} \mathbb{E}[T_{xy}]$

Proof: The proof is a simple application of Markov's inequality. ■

This means that, in order to devise a good coupling, we can focus on making $\mathbb{E}[T_{xy}]$ small.

25.3 Random Transposition Shuffle

Recall from last lecture the card shuffling process based on random transpositions. The following are two equivalent ways of describing this shuffle:

1. pick positions i, j u.a.r, and switch the cards at i, j.
2. pick position i and card c u.a.r, and switch c with the card at position i

(The first is the original definition of the shuffle; the second is more convenient for describing our coupling.)

Since we want to coordinate the MCs X_t and Y_t to make progress towards meeting, we consider the following coupling:

Coupling: both X_t and Y_t choose the same position i and the same card c at every step.

Let D_t denote the number of positions where X_t and Y_t disagree. Our goal is to determine how long it takes until $D_t = 0$. To study the evolution of D_t, there are two cases to consider based on the common choice of position i and card c:

1. Card c is already matched. In this case, D_t does not change.
2. Card c not matched. In this case, D_t does not increase, and decreases by at least 1 if in addition the cards at position i don't match. If the current distance is $D_t = d$, then $\Pr[D_t \text{ decreases}] \ge (d/n)^2$.

Thus we see that D_t is non-increasing, and the time until $D_t = 0$ is dominated by the sum

$$T_1 + T_2 + \ldots + T_n,$$

where T_d is a geometric random variable with expectation $\mathbb{E}[T_d] = (n/d)^2$ (an upper bound on the expected time taken for D_t to go from d to $d-1$). Hence the expected time to meet is $\mathbb{E}[T_{xy}] \le \sum_{d=1}^{n} (n/d)^2 = O(n^2)$. By Corollary 25.7 the mixing time is thus $O(n^2)$.

Note: The exact mixing time for this process (obtained by more sophisticated methods based on group representations [DS]) is $\Theta(n \log n)$.

Thus the card shuffling process mixes in a number of steps that is a low-degree polynomial in n. As for the other shuffling processes we analyzed in the last lecture, this is fairly remarkable in light of the fact that the size of the state space is $n!$, which is exponential in n.

25.4 Graph Colorings

Input: $G = (V, E)$ an undirected graph, max degree Δ; k colors.
Goal: Pick a random (proper) k-coloring of G.

Recall our Markov Chain from the last lecture:

1. pick a vertex v and color c u.a.r.
2. recolor v with c if this is legal (i.e., no neighbor of v has color c).

Recall that provided $k \geq \Delta+1$, G is k-colorable. (Actually, $k \geq \Delta$ is enough if G doesn't contain a Δ-clique [Brooks' Theorem]). Also, recall that in order to guarantee that the Markov Chain is connected (irreducible) we require that $k \geq \Delta + 2$.

Conjecture 25.8 *For all $k \geq \Delta + 2$ this MC has small mixing time ($O(n \log n)$, or at least polynomial in n).*

This remains an important conjecture, with applications not only in combinatorics but also in statistical physics (through the connection with the so-called "anti-ferromagnetic Potts model"). Much effort has gone into determining the smallest value of k for which rapid mixing can be proved. The current state-of-the-art uses fancy coupling arguments and extensions to show that $k \geq 1.76\Delta$ is sufficient [H03], and this has even been pushed down to $k \geq (1+\epsilon)\Delta$ for any $\epsilon > 0$ with (somewhat severe) additional assumptions on G [HV03].

Here we prove a weaker version, which says that about 2Δ colors are enough to ensure rapid mixing.

Theorem 25.9 [J95,SS97] *If $k \geq 2\Delta + 1$ then the MC has mixing time $O(n \log n)$.*

Proof: We'll first prove a slightly looser version of Theorem 25.9 by showing that the statement is true for $k \geq 4\Delta + 1$.

For the coupling, let X_t and Y_t choose the same v and c at every step. Let $D_t = \{v : X_t, Y_t$ disagree on color of vertex $v\}$, and $A_t = V - D_t$. We denote by $d_t = |D_t|$ the distance between X_t and Y_t. Our goal is to determine how long it takes before $d_t = 0$.

Unlike the random transposition shuffle, here it is not the case that the distance d_t never increases. Accordingly, we look at two classes of moves: *Good Moves*, which decrease d_t, and *Bad Moves* which increase d_t.

Good moves:

Suppose vertex v and color c are chosen. We get a good move if vertex $v \in D_t$ and color c does not appear in the neighborhood of v in X_t or Y_t (see Fig. 25.2). For in this case, v will be recolored to c in both processes, and they will now agree on v. Let g_t be the number of good moves available at step t. Then $g_t \geq d_t(k-2\Delta)$, because there are d_t possible vertices to choose from, and at most 2Δ colors are present in the neighborhood of v in either process. Therefore, of the k-colors, at least $k - 2\Delta$ will provide good moves.

Bad moves:

Suppose vertex $v \in A_t$ is chosen, along with a color c that is in the neighborhood of v in one of X_t, Y_t but not the other. Then the recoloring with c will take place in one process and not the other, thus creating a new disagreement at v. As an example, in Fig. 25.3, choosing vertex v and color R (or B) will result in a bad move, because v already agrees on its color (Y) but v can be recolored with R in the right-hand process but not in the left-hand process.

Let b_t be the number of bad moves. Then we claim that $b_t \leq 2d_t\Delta$. To see this, note that the color c chosen must be the color of some neighbor of v in one process and not in the other. Thus v itself must be a neighbor of a vertex in D_t, and c must be the color of that vertex in one of the two processes. Thus the number of choices for v is at most $d_t\Delta$, and the number of choices for c is 2, resulting in the value claimed for b_t.

All other moves (neither bad nor good) cause no change in d_t.

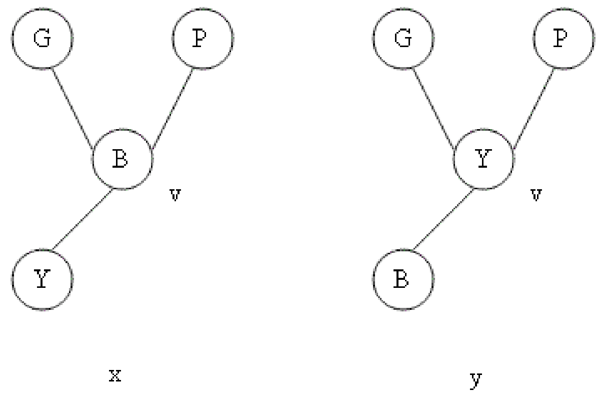

Figure 25.2: Good move: v is chosen, and the color choice is R (which is not in the neighborhood of v in either X or Y)

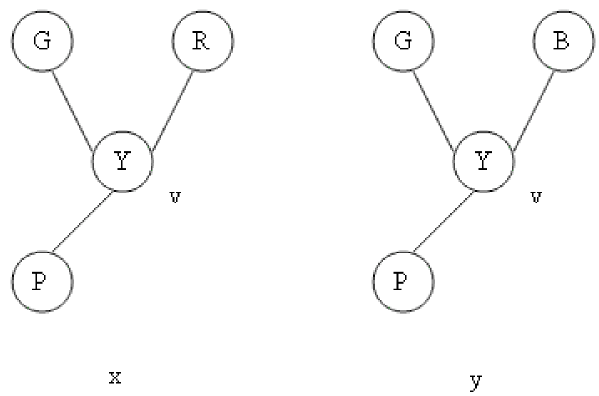

Figure 25.3: Bad move: v, colored Y at step t, is chosen to be colored R at step $t+1$

Clearly the probability of making any particular move is $1/kn$. Thus,

$$\begin{aligned} \mathbb{E}[d_{t+1}|d_t] &= d_t + \frac{b_t - g_t}{kn} \qquad \text{(bad moves contribute $+1$ to d_t, good moves -1)} \\ &\le d_t + d_t \frac{4\Delta - k}{kn} \\ &\le d_t(1 - \frac{1}{kn}) \qquad \text{(assuming $k \ge 4\Delta + 1$)} \end{aligned}$$

and thus

$$\mathbb{E}[d_t|d_0] \le d_0(1 - \frac{1}{kn})^t \le 1/2e \qquad \text{for } t = Ckn\log n \text{ and recalling } d_0 \le n.$$

Finally, note by Markov's inequality that with this value of t

$$\Pr[T_{xy} > t] = \Pr[d_t > 0 \mid X_0 = x, Y_0 = y] = \Pr[d_t \ge 1 \mid X_0 = x, Y_0 = y] \le \mathbb{E}[d_t \mid X_0, Y_0] \le 1/2e.$$

Hence by Claim 25.5 we deduce that the mixing time is $O(n \log n)$.

To strengthen the above argument from $k \ge 4\Delta + 1$ to $k \ge 2\Delta + 1$, we can change the coupling so that it pairs off colors in the neighborhood of v in X_t and not in Y_t with those in the neighborhood of v in Y_t and not in X_t. To illustrate this idea, suppose the colors in the neighborhood of v in X_t are red, green, yellow, blue, and in Y_t they are red, orange, white. (So the 'bad' colors in the neighborhoods are {green, yellow, blue} and {orange, white} respectively.) Then we couple colors as follows: (red,red), (green,orange), (yellow,white), (blue,blue) (and all other colors with themselves; note that we have to couple blue with itself because there are more bad colors in the first set). The advantage of this is the following: when (X_t, Y_t) choose the color pair (green,orange), *neither* of them will move, so the move will not be bad; a bad move will occur, of course, when they choose the complementary pair (orange,green). Contrast this with our previous coupling, when *both* color pairs (green,green) and (orange,orange) caused a bad move. Under this scheme, the number of color choices that cause a bad move at v is the *maximum* number of bad colors for v in X_t and Y_t.

The above improvement effectively decreases the number of bad moves by a factor of 2. Taking a bit more care with the analysis, one can show (Exercise!) that $g_t - b_t \geq (k - 2\Delta)d_t$. This is exactly the same as in the analysis above, but with 2Δ replacing 4Δ. Thus we get mixing time $O(n \log n)$ for $k \geq 2\Delta + 1$ also. ■

Exercise: In the boundary case $q = 2\Delta$, prove that the mixing time is $O(n^3)$. [Hint: Compare the evolution of d_t with a symmetric random walk, with a holding probability of $1 - \frac{1}{n}$ at each step.]

References

[HV03] T.P. Hayes and E. Vigoda, "A Non-Markovian Coupling for Randomly Sampling Colorings," *Proceedings of FOCS 2003*, pp. 618–627.

[H03] T.P. Hayes, "Randomly coloring graphs of girth at least five," *Proceedings of STOC 2003*, pp. 269–278.

[J95] M. Jerrum, "A very simple algorithm for estimating the number of k-colorings of a low-degree graph," *Random Structures and Algorithms*, 1995, pp. 157–165.

[SS97] J. Salas and A. D. Sokal, "Absence of phase transition for antiferromagnetic Potts models via the Dobrushin uniqueness theorem," *Journal of Statistical Physics* **86** (1997), pp. 551–579.