CSE 5401 Lec. 13

Fast mixing via couplings

1. Coupling and mixing time
2. Example coupling
3. Sampling/counting graph colorings (with many colors)

We know that in order to solve counting problems, it's enough to be able to sample solutions approximately uniformly at random. But in general, the set of solutions may be exponentially large. We need to reach a uniform distribution on this large set very quickly in order to get tractable algorithms. One technique that can obtain bounds that show such fast mixing occurs is to present a "coupling":

Defn: A coupling of a Markov chain $W_0, W_1, \ldots, W_t, \ldots$ on the state space S is a Markov chain $Z_0, Z_1, \ldots, Z_t, \ldots$ on the state space $S \times S$ such that if $Z_t = (X_t, Y_t)$, for all $(x,y), (x',y')$ in $S \times S$, $\Pr[X_{t+1} = x' \mid Z_t = (x,y)] = \Pr[W_{t+1} = x' \mid W_t = x]$ and

$\Pr[Y_{t+1} = y' \mid Z_t = (x,y)] = \Pr[W_{t+1} = y' \mid W_t = y]$

In other words, projecting $\{Z_t\}$ on to the first or second component gives a Markov chain with the same distribution as $\{W_t\}$.

Indeed, two independent runs of $\{W_t\}$ are an example of a coupling, but not a very useful one. Rather, a useful coupling will correlate the transitions in some way so that if $X_t = Y_t$ for some t, then $X_{t+1} = Y_{t+1}$, so the chains remain in identical states. The upshot is that if Y_0 is distributed according to the stationary distribution of the Markov chain, then so is $Y_1, Y_2, \ldots, Y_t$, and

hence then so is $X_t, X_{t+1}, \ldots$, regardless of what x_0 was. Then, the Markov chain will have mixed from x_0 within t steps. A good coupling makes this occur quickly. Let's see this formally

Coupling Lemma Let $\{(X_t, Y_t)\}$ be a coupling of the Markov chain $\{W_t\}$, and let π be the stationary distribution of $\{W_t\}$. Suppose that for some $\varepsilon > 0$ and $T \in \mathbb{N}$, for all $x_0, y_0 \in S$, $\Pr[X_T \neq Y_T \mid X_0 = x_0, Y_0 = y_0] \leq \varepsilon$. Then for all $A \subseteq S$, $w_0 \in S$, $|\Pr[W_T \in A \mid W_0 = w_0] - \pi(A)| \leq \varepsilon$

Proof: First observe that if we sample y_0 from π, since π is the stationary distribution of $\{W_t\}$ and $\Pr[Y_{t+1} = y' \mid Y_t = y] = \Pr[W_{t+1} = y' \mid W_t = y]$, $Y_1, Y_2, \ldots, Y_T$ are all also distributed according to π. Similarly, since $\Pr[X_{t+1} = x' \mid X_t = x] = \Pr[W_{t+1} = x' \mid W_t = x]$, If $X_0 = w_0$, $\Pr[X_T = x' \mid X_0 = w_0] = \Pr[W_T = x' \mid W_0 = w_0]$. Hence, for all $A \subseteq S$, $\Pr[X_T \in A \mid X_0 = w_0] = \Pr[W_T \in A \mid W_0 = w_0]$ and $\mathbb{E}_{Y_0 \sim \pi}[\Pr[Y_T \in A \mid Y_0 = y_0]] = \pi(A)$.

Now, observe

$$\Pr[X_T \in A \mid X_0 = w_0] \geq \mathbb{E}_{y_0 \sim \pi}[\Pr[X_T = Y_T \text{ and } Y_T \in A \mid X_0 = w_0, Y_0 = y_0]]$$

(union bound)

$$\geq \mathbb{E}_{Y_0 \sim \pi}[1 - \Pr[X_T \neq Y_T \mid X_0 = w_0, Y_0 = y_0] - \Pr[Y_T \notin A \mid X_0 = w_0, Y_0 = y_0]]$$

$$\geq \mathbb{E}_{Y_0 \sim \pi}[\Pr[Y_T \in A \mid X_0 = w_0, Y_0 = y_0]] - \varepsilon \quad \text{by assumption}$$

$$= \pi(A) - \varepsilon.$$

Furthermore, since A is arbitrary, $\Pr[X_T \in A^c \mid X_0 = w_0] \geq \pi(A^c) - \varepsilon = 1 - \pi(A) - \varepsilon$ so $\pi(A) + \varepsilon \geq 1 - \Pr[X_T \notin A^c \mid X_0 = w_0] = 1 - \Pr[X_T \in A \mid X_0 = w_0] = \Pr[X_T \in A \mid X_0 = w_0]$ ▨

Example Consider the Markov chain on $\{0,1\}^n$ in which we choose a coordinate i at random and set it equal to 0 or 1 w.p. $\frac{1}{2}$ ea.

What is the stationary distribution of this Markov chain?

Notice, we can transition from each $\vec{w} \in \{0,1\}^n$ to each of the n strings that differ in one coordinate, each with prob. $\frac{1}{2n}$, and remain at $\vec{w}$ w.p. $\frac{1}{2}$. Thus, this is (essentially) a random walk on an undirected graph where every vertex has n neighbors and n self-loops. It is not bipartite because of the self loops. Thus, we saw previously that it has a stationary distribution that is uniform on $\{0,1\}^n$. Now, consider the following coupling: we pick i uniformly at random as before, but we set x_i and y_i to the same value chosen uniformly at random. Since each was indeed chosen according to the original distribution, it is indeed a coupling and observe that once a coordinate is chosen, $x_i = y_i$ on all subsequent steps. Thus, $\vec{x}_t = \vec{y}_t$ once every coordinate has been chosen at least once. The number of steps before this occurs is given by the "coupon collector" problem. Consider any coordinate i. The probability i is not chosen in the first t steps is $(1-\frac{1}{n})^t$, so for $t \geq n \ln\frac{n}{\varepsilon}$, this is at most $(1-\frac{1}{n})^{n \cdot \ln\frac{n}{\varepsilon}} \leq e^{-\ln\frac{n}{\varepsilon}} = \frac{\varepsilon}{n}$. A union bound over the coordinates gives that with probability $1-\varepsilon$, all n coordinates are chosen within $n \ln\frac{n}{\varepsilon}$ steps. Thus, by the coupling lemma, the distribution is ε-close to uniform after $n \ln\frac{n}{\varepsilon}$ steps.
It is not hard to simply sample $\vec{w} \in \{0,1\}^n$ uniformly at random. The power of the technique is that we can use it when the structure of solutions makes them hard to sample uniformly.

Sampling colorings

Recall that a c-coloring of an undirected graph $G=(V,E)$ is an assignment of one of c colors to each vertex of G

such that vertices connected by an edge receive distinct colors. 3-coloring is NP-complete in general, but if every vertex in the graph has at most d neighbors, then d+1 coloring is trivial: however we color a vertex's neighbors, there is always a distinct color we can use for the vertex. So, we can greedily assign colors to the vertices. Counting the number of possible colorings, however, remains difficult. Let's suppose $c \geq 4d+1$. Consider the following Markov chain on colorings of the graph: we pick a vertex and a color uniformly at random. If assigning the chosen color to the vertex yields a proper coloring, we let the next coloring be given by changing the color of that vertex on the next step to the chosen color, and leaving all other vertices the same. Otherwise, we leave the coloring as-is.

What is the stationary distribution of this Markov chain?

We observe first that each coloring has all proper colorings that differ in the color of a single vertex as its neighbors, and the probability of transitioning to each is precisely $\frac{1}{|V| \cdot c}$. Since any two colorings that differ in a single vertex can reach each other (by choosing the respective colors) the transitions are symmetric, and with probability $\geq \frac{1}{c}$ we choose the same color we started with, so the transition graph is not bipartite. Now, we note that this graph on colorings is connected: Fix some order of the vertices, and given colorings $\vec{x}$ and $\vec{y}$, we consider the vertices in order: for $i = 1, \ldots, |V|$, if none of the i^{th} vertex's

neighbors in $\vec{X}$ are colored y_i's color, put $x_i \leftarrow y_i$. Otherwise, for the neighbors $j > i$ currently colored y_i's color in $\vec{X}$, we recolor x_j to some remaining color other than y_i's color (which is possible since $c > d+2$) and then put $x_i \leftarrow y_i$. Each step leaves a proper coloring in $\vec{X}$ s.t. for all $j \le i$ $X_i = Y_i$, so at the end, $\vec{X} = \vec{Y}$.

Now we'll use a coupling to bound the # steps before this Markov chain is close to the stationary distribution:

Theorem For any n-vertex $G = (V, E)$ in which every vertex has $\le d$ neighbors and any initial c-coloring for $c \ge 4d+1$, for any $\varepsilon > 0$, after $T \ge \frac{nc}{c-4d} \ln \frac{n}{\varepsilon}$ steps, the Markov chain is ε-close to a uniform random proper c-coloring of G.

Proof: Consider the coupling $\{(\vec{X}_t, \vec{Y}_t)\}$ that chooses the same vertex and color to consider for both $\vec{X}$ and $\vec{Y}$. Let D_t be the number of vertices that are colored differently in $\vec{X}_t$ & $\vec{Y}_t$

Claim $\mathbb{E}[D_t] \le n \cdot \left(1 - \frac{c-4d}{nc}\right)^t$

Proof of Claim by induction on t. Base, $t=0$: $D_0 \le |V| = n$.

For the induction step, we consider two cases: either step t chooses vertices with different colors in $\vec{X}_t$ & $\vec{Y}_t$, or vertices with the same colors. In the first case, which occurs w.p. $\frac{D_t}{n}$, either the color chosen is distinct from the colors of the $\le d$ neighbors in both $\vec{X}_t$ & $\vec{Y}_t$ ($\le 2d$ colors), w.p. $\ge \frac{c-2d}{c}$, in which case $D_{t+1} = D_t - 1$, or else the vertex remains different colors in $\vec{X}_{t+1}$ & $\vec{Y}_{t+1}$, so $D_t = D_{t+1}$. If vertices with the same colors in $\vec{X}_t$ & $\vec{Y}_t$ are chosen, these colors remain the same ($D_{t+1} = D_t$) unless we choose a neighbor of one of the vertices colored differently in $\vec{X}_t$ & $\vec{Y}_t$, w.p..

$\leq \frac{d \cdot D_t}{n}$, and choose one of the colors of that neighbor in $\vec{X}_t$ or $\vec{Y}_t$, w.p. $\frac{2}{c}$ — then the vertex is colored differently in $\vec{X}_{t+1}$ & $\vec{Y}_{t+1}$

i.e., $D_{t+1} = D_t + 1$. Notice, we have

$$\mathbb{E}[D_{t+1} \mid D_t] = D_t + \Pr[D_{t+1} = D_t + 1] - \Pr[D_{t+1} = D_t - 1]$$

$$\leq D_t + \frac{d \cdot D_t}{n} \cdot \frac{2}{c} - \frac{D_t}{n} \cdot \frac{c-2d}{c} = D_t \left(1 - \frac{c-4d}{nc}\right) \text{ so}$$

$$\mathbb{E}[D_{t+1}] = \mathbb{E}[\mathbb{E}[D_{t+1} \mid D_t]] \leq \mathbb{E}[D_t] \cdot \left(1 - \frac{c-4d}{nc}\right) \leq n \cdot \left(1 - \frac{c-4d}{nc}\right)^{t+1} \text{ by IH}$$

Now, we simply note that D_t is a nonnegative integer, so

$$\Pr[D_T \geq 1] \leq \mathbb{E}[D_T] \leq n\left(1 - \frac{c-4d}{nc}\right)^T \leq n \cdot e^{-\ln\left(\frac{n}{\varepsilon}\right)} = \varepsilon$$

by choice of T. When $D_t = 0$, $\vec{X}_t = \vec{Y}_t$, so by the coupling lemma, the Markov chain is thus indeed ε-close to its stationary distribution, which we just argued is uniform on proper colorings after $T \geq \frac{nc}{c-4d} \ln \frac{n}{\varepsilon}$ steps. ▨

Furthermore, by the reduction of counting to sampling, this gives us our first example of a nontrivial FPRAS for a counting problem.