CSE5401 Lec. 18

Path coupling analysis in MCMC

1. Counting independent sets
2. Path coupling analysis

Previously, we saw how to approximately count colorings using a "Markov Chain Monte Carlo" algorithm: in this algorithm, we 1) set up a Markov chain for which the stationary dist'n was uniform on the proper colorings. Then we

2) used a coupling analysis of the Markov chain to show that it is approximately uniform after a poly. # of steps. Finally

3) We invoked our reduction from approx. counting to approx. sampler

This framework can be used to solve many approximate counting problems. Today we will use it to approximately count the # of "independent sets" in graphs of degree ≤ 4. We'll use a slightly more elaborate analysis of the coupling in which we consider the change in distances on a path of states.

Recall, given an undirected graph $G=(V,E)$, an independent set in G is a set of vertices $S \subseteq V$ s.t. no pair $u,v \in S$ have an edge: $\forall u,v \in S$, $(u,v) \notin E$ (cf. clique: $\forall u,v \in S$, $(u,v) \in E$ — very similar problems, but today independent set will be more natural.)

We'll focus on graphs of degree ≤ 4. (Note that graphs with few edges may have many large independent sets.) We'll assume G is connected (sample components indep.)

1. The Markov chain

Given an undirected graph G, the states of our Markov chain

will be the independent sets of G. The transitions are sampled as follows: first, we choose $(u,v) \in E$ uniformly at random Then, if our current independent set is X_t with probability $\frac{1}{3}$ we put $X_{t+1} = X \setminus \{u,v\}$; with probability $\frac{1}{3}$ we check if $X_t \cup \{u\} \setminus \{v\}$ is an independent set, taking this as X_{t+1} if so and leaving $X_{t+1} = X_t$ otherwise. Similarly, with probability $\frac{1}{3}$, we put $X_{t+1} = X_t \cup \{v\} \setminus \{u\}$ if this is an independent set and leave $X_{t+1} = X_t$ otherwise.

How do we show that this Markov chain has a uniform stationary distribution? We know that it suffices to show that the Markov chain (a) has a symmetric transition matrix (b) is not bipartite and (c) is connected. For part a, we first observe that there are two types of adjacent independent sets: those that differ in one vertex (add/delete), and those that differ in two (swap the endpoints of an edge.) Note that we cannot have both endpoints of an edge in X_t, so the first move just drops whichever is in X_t Clearly, for pairs that differ in a swap, the probability of this transition is $\frac{1}{3|E|}$ in either direction. For pairs that differ in adding or removing a vertex v, we see that for each vertex u not in either with an edge to v, when we pick the (u,v) edge, there is one move (resp. delete both and add v / remove u) that transitions from one to the other. Since v cannot have edges to any u in the rest of the independent set, these are the only moves that add/delete v, so indeed the probabilities are symmetric. To see b, observe that if there is an edge with neither endpoint in X_t, dropping both gives

X_t again, and otherwise, there is a swap move for each edge that also yields X_t again. Thus, the graph on states has self-loops and is not bipartite. Finally, for part c, we notice that we can get from any independent set X to any other X' by first swap moves to add all of the vertices in $X' \setminus X$ and delete all of the neighbors of X' in X. Then moves that delete both endpoints can remove the vertices of $X \setminus X'$ that are not neighbors of vertices of X'. Therefore, we've shown a, b, & c, so the Markov chain has a uniform stationary distribution.

2. The coupling

We use the natural coupling, similar to our previous examples: we sample the same edge and attempt the same move in X_t and Y_t. Observe that this gives the original Markov chain when we consider X_t or Y_t alone, so it is a valid coupling. It now remains to bound the time for the coupling to meet. Again, we let $D_t = |X_t \Delta Y_t|$: the symmetric difference, i.e. $X_t \Delta Y_t = (X_t \setminus Y_t) \cup (Y_t \setminus X_t)$, so $|X_t \Delta Y_t| = |X_t \setminus Y_t| + |Y_t \setminus X_t|$ (Notice that $X_t \cap Y_t$ is removed from both, so these are disjoint.) We now wish to bound $\mathbb{E}[D_{t+1} \mid D_t]$ as before.

To bound this, we'll consider a sequence of independent sets that differ in exactly one vertex, $X_t = Z_t^{(0)}, Z_t^{(1)}, \ldots, Z_t^{(D_t)} = Y_t$ e.g., the first $k = |X_t \setminus Y_t|$ sets delete the vertices in $X_t \setminus Y_t$, $Z_t^{(k)} = X_t \cap Y_t$, and the next $|Y_t \setminus X_t|$ sets add the vertices of Y_t. Thus, $|X_t \Delta Y_t| = \sum_{i=1}^{D_t} |Z_t^{(i)} \Delta Z_t^{(i-1)}|$ by construction. Now, for the move drawn for the coupling of X_t and Y_t, we put $Z_{t+1}^{(i)}$ equal

to the result of the same move from $Z_t^{(i)}$. Now, notice,
For any three sets A, B, and C, $|A \triangle B| = |A \setminus B| + |B \setminus A|$
$|A \setminus B| = |(A \setminus C) \setminus B| + |(A \cap C) \setminus B| \leq |A \setminus C| + |C \setminus B|$ and
similarly, $|B \setminus A| \leq |B \setminus C| + |C \setminus B|$ so: $|A \triangle B| \leq |A \triangle C| + |C \triangle B|$.
Thus, $|X_{t+1} \triangle Y_{t+1}| \leq |X_{t+1} \triangle Z_{t+1}^{(i)}| + |Z_{t+1}^{(i)} \triangle Y_t| \leq |X_{t+1} \triangle Z_{t+1}^{(1)}| + |Z_{t+1}^{(1)} \triangle Z_{t+1}^{(2)}| + \cdots$
i.e., $|X_{t+1} \triangle Y_{t+1}| \leq \sum_{i=1}^{D_t} |Z_{t+1}^{(i)} \triangle Z_{t+1}^{(i-1)}|$. It therefore suffices to
bound $\mathbb{E}\left[|Z_{t+1}^{(i)} \triangle Z_{t+1}^{(i-1)}| \mid Z_t^{(i)}, Z_t^{(i-1)}\right]$ where $Z_t^{(i)}$ & $Z_t^{(i-1)}$ differ
in one vertex, since $\mathbb{E}[D_{t+1} \mid D_t] = \mathbb{E}_{\substack{X_t Y_t : \\ |X_t \triangle Y_t| = D_t}}\left[\mathbb{E}[|X_{t+1} \triangle Y_{t+1}| \mid X_t, Y_t]\right]$
$\leq \sum_{i=1}^{D_t} \mathbb{E}_{|X_t \triangle Y_t| = D_t}\left[\mathbb{E}[|Z_{t+1}^{(i)} \triangle Z_{t+1}^{(i-1)}| \mid Z_t^{(i)}, Z_t^{(i-1)}]\right]$.
Thus, finally, suppose WLOG $Z_t^{(i)} = Z_t^{(i-1)} \cup \{v\}$. (O.W. $Z_t^{(i)} \cup \{v\} = Z_t^{(i-1)}$,
and we just rename them) Observe that we can only have $|Z_{t+1}^{(i)} \triangle Z_{t+1}^{(i-1)}| \neq$
$|Z_t^{(i)} \triangle Z_t^{(i-1)}|$ if the move involves an edge (u, w) with $(u, v) \in E$;
as otherwise the move succeeds or fails in both, and adds/removes
the same vertex to both if it succeeds. Since $v \in Z_t^{(i)}$, $u \notin Z_t^{(i)}$ or $Z_t^{(i-1)}$.
We now invoke the assumption that G has degree ≤ 4:
u may have $0, 1, 2,$ or 3 neighbors in $Z_t^{(i-1)}$ (giving it
resp. $1, 2, 3,$ or 4 neighbors in $Z_t^{(i)}$ when we include v).
If u has 2 or 3 neighbors in $Z_t^{(i-1)}$, then it cannot be added
since its other neighbor always remains in $Z_t^{(i-1)}$ if we attempt to
swap it in. Moves that would add another neighbor $w \notin \{v\} \cup Z_t^{(i-1)}$
or delete both succeed in both states, so $|Z_{t+1}^{(i)} \triangle Z_{t+1}^{(i-1)}|$ cannot be
larger for edges containing such neighbors u.
If u has 1 neighbor in $Z_t^{(i-1)}$ then a move that adds u and
deletes $w \in Z_t^{(i-1)}$ succeeds in $Z_t^{(i-1)}$ but fails in $Z_t^{(i)}$ since v is a
neighbor of u present in $Z_t^{(i)}$. In this case, the difference increases

by 2. On the other hand, if it is the (u,v) edge that is chosen, either adding v and deleting u or deleting both each decrease the difference by 1. Notice, other moves on these edges do not change the difference: adding u and deleting v fails on account of $w \in Z_t^{(i-1)}$; adding u and deleting u or deleting both w and u resp. leave both unchanged or delete w from both. Similarly, moves with the non-adjacent neighbors of u fail to add u and otherwise make the same change to both. Thus, the expected difference is 0.

What about u with no neighbors in $Z_t^{(i-1)}$? In this case, adding u succeeds in $Z_t^{(i-1)}$ and fails in $Z_t^{(i)}$ with any of the ≤ 3 neighbors $w \neq v$. None of the other moves with a (u,w) edge change the difference. But, all three moves on the (u,v) edge decrease the difference by 1. Since the choice of edge and type of move is uniformly random, the expected change is 0.

Thus we see that for all pairs $\mathbb{E}\left[|Z_{t+1}^{(i-1)} \triangle Z_{t+1}^{(i)}| - |Z_t^{(i-1)} \triangle Z_t^{(i)}| \,\middle|\, Z_t^{(i)}, Z_t^{(i-1)}\right]$ is ≤ 0, hence $\mathbb{E}[D_{t+1} \mid D_t] \leq D_t$; and the probability that it changes by -1 for $D_t > 0$ is $\geq \frac{1}{3|E|}$ (if we delete both endpoints on one of the edges with a difference) Thus, essentially* as we saw on homework the coupling reaches the same state in $\text{poly}(|E|, |V|, \log \frac{1}{\varepsilon})$ steps with probability $1-\varepsilon$. The coupling lemma thus tells us that it is ε-close to uniform after a polynomial # of steps, and so we can use our reduction from approximate counting to approximate sampling to count independent sets. (step 3).

The Markov chain here may increase the distance by 2 on some moves, though the expected change is still non-positive: $\mathbb{E}[D_{t+1} - D_t \mid D_t] \le 0$. Thus $\Pr[D_{t+1} - D_t = -1 \mid D_t] \ge \Pr[D_{t+1} - D_t = 1 \mid D_t] + 2 \cdot \Pr[D_{t+1} - D_t = 2 \mid D_t]$

We solve a similar hitting time problem on D_t. Let h_{ij} denote the expected # of steps before reaching $D_t = j$ from $D_t = i$. Then for $i = 1, \dots, n-2$, we have the recurrence

$$\begin{aligned} h_{i,i-1} &= \Pr[D_{t+1} - D_t = -1 \mid D_t = i] + \Pr[D_{t+1} = D_t \mid D_t = i](1 + h_{i,i-1}) + \\ &\quad \Pr[D_{t+1} - D_t = 1 \mid D_t = i](1 + h_{i+1,i-1}) + \Pr[D_{t+1} - D_t = 2 \mid D_t = i](1 + h_{i+2,i-1}) \\ &= 1 + (1 - \Pr[\cdots = -1 \mid D_t = i])h_{i,i-1} + (\Pr[\cdots = 1 \mid D_t = i] + \Pr[\cdots = 2 \mid D_t = i])h_{i+1,i} \end{aligned}$$

$$h_{i,i-1} = \frac{1}{\Pr[\cdots = -1 \mid D_t = i]} + \frac{\Pr[\cdots = 1 \mid D_t = i] + \Pr[\cdots = 2 \mid D_t = i]}{\Pr[\cdots = -1 \mid D_t = i]} h_{i+1,i} + \frac{\Pr[\cdots = 2 \mid D_t = i]}{\Pr[\cdots = -1 \mid D_t = i]} h_{i+2,i+1}$$

where $h_{n,n-1} = \frac{1}{\Pr[\cdots = -1 \mid D_t = n]}$, $h_{n-1,n-2} = \frac{1}{\Pr[\cdots = -1 \mid D_t = n-1]} + \frac{\Pr[\cdots = 1 \mid D_t = n-1]}{\Pr[\cdots = -1 \mid D_t = n-1]} h_{n,n-1}$

$$\le \frac{1}{\Pr[\cdots = -1 \mid D_t = n-1]} + \frac{1}{\Pr[\cdots = -1 \mid D_t = n]}$$

Claim $h_{n-k,n-k-1} \le \sum_{i=0}^{k-1} \frac{1}{\Pr[\cdots = -1 \mid D_t = n-i]}$

Proof by induction on k; base cases $k = 1, 2$ above. In general, our recurrence gives

$$h_{n-k,n-k-1} \le \frac{1}{\Pr[\cdots = -1 \mid D_t = n-k]} + \frac{\Pr[\cdots = 1 \mid D_t = n-k] + \Pr[\cdots = 2 \mid D_t = n-k]}{\Pr[\cdots = -1 \mid D_t = n-k]} \sum_{i=0}^{k-2} \frac{1}{\Pr[\cdots = -1 \mid D_t = n-i]} + \frac{\Pr[\cdots = 2 \mid D_t = n-k]}{\Pr[\cdots = -1 \mid D_t = n-k]} \sum_{i=0}^{k-3} \frac{1}{\Pr[\cdots = -1 \mid D_t = n-i]}$$

$$\le \frac{1}{\Pr[\cdots = -1 \mid D_t = n-k]} + \sum_{i=0}^{k-2} \frac{1}{\Pr[\cdots = -1 \mid D_t = n-i]}$$ since

$$\frac{\Pr[\cdots = 1 \mid D_t = n-k] + 2 \cdot \Pr[\cdots = 2 \mid D_t = n-k]}{\Pr[\cdots = -1 \mid D_t = n-k]} \le 1.$$

Claim $h_{n,n-k} \le \sum_{i=0}^{k-1} (k-i) \frac{1}{\Pr[\cdots = -1 \mid D_t = n-i]}$

Proof by induction on k. Again, $k = 1$ is above. In general,

$$h_{n,n-k-1} \le h_{n,n-k} + h_{n-k,n-k-1} \le \sum_{i=0}^{k-1} (k-1) \frac{1}{\Pr[\cdots = -1 \mid D_t = n-i]} + \sum_{i=0}^{k-1} \frac{1}{\Pr[\cdots = -1 \mid D_t = n-i]}$$

$$\leq \sum_{i=0}^{k-1} (k+1-i) \frac{1}{\Pr[\dots=-1 \mid D_t = n-i]}$$

Thus, since $\Pr[\dots=-1 \mid D_t = i] \geq \frac{1}{3|E|}$, $h_{n,0} \leq O(n^2|E|)$.

Since we also have $h_{i,0} \leq h_{n,0}$ for all i, Markov's inequality gives that after any sequence of $O(n^2|E|)$ moves, $\Pr[D_t \neq 0] < 1/2$.

After $\log\frac{1}{\varepsilon}$ repetitions — i.e., $O(n^2|E|\log\frac{1}{\varepsilon})$ moves — $\Pr[D_t \neq 0] < \varepsilon$.