CS5120: Probability & Computing **Jan-May 2020**

Markov Chains 5

1 A method for approximate counting

Let $S \subseteq U$. To find $|S|$, we saw the idea of sampling uniformly at random from U. The number of samples needed for an (ε, δ)-approximation is $\Theta\left(\frac{|U|}{|S|}\frac{1}{\varepsilon^2}\log(\frac{1}{\delta})\right)$.

Approximately uniform sampling suffices. Suppose that X is a random sample and let Y be an indicator variable so that $Y = 1$ if $X \in S$ and $Y = 0$ otherwise. Then the estimate on sufficient number of samples is $\Theta(\frac{1}{E[Y]})$ and we have $E[Y] = \frac{|S|}{|U|}$. We can thus observe that if we are able to sample approximately u.a.r. then the number of samples doesn't increase significantly. That is, suppose that X is a random sample such that $Pr[X \in S] \in \frac{|S|}{|U|}\left(\frac{1}{2}, \frac{3}{2}\right)$. Then $E[Y]$ in this case, compared to the exactly-uniform sample case, is at most halved, so that the required number of samples at most doubles.

We now consider a new idea. As before, we will assume that $\frac{|U|}{|S|}$ is large. The idea is to find a sequence of sets $S_0, S_1, \ldots, S_m$ such that the following are satisfied.

- $S = S_0 \subset S_1 \subset \ldots \subset S_m \subset U$.
- $|S_m|$ is known or easy to compute.
- Sampling from each S_i possible.
- $\frac{|S_i|}{|S_{i-1}|}$ is small for every i.

Then we can estimate successively, $|S_{m-1}|$, $|S_{m-2}|$, ..., $|S_0|$; to estimate $|S_i|$, we sample from S_{i+1}.

Let $r_1 = \frac{|S_1|}{|S_0|}$, $r_2 = \frac{|S_2|}{|S_1|}$, $\ldots$, $r_m = \frac{|S_m|}{|S_{m-1}|}$. Then we have: $\frac{|S_m|}{|S_0|} = \prod_i r_i$.

Suppose that we have an $(\bar{\varepsilon}, \bar{\delta})$ approximation for each of the sets, that is:

$$Pr\left[\frac{\bar{r_i}}{r_i} \in (1-\bar{\varepsilon}, 1+\bar{\varepsilon})\right] \geq 1-\bar{\delta}.$$

Claim 1 *If $\bar{\varepsilon} \leq \frac{\varepsilon}{2m}$ and $\bar{\delta} \leq \frac{\delta}{m}$, then we have an (ε, δ)-approximation for S, that is:*

$$Pr\left[\frac{\bar{r_1}}{r_1}\frac{\bar{r_2}}{r_2}\cdots\frac{\bar{r_m}}{r_m} \in (1-\varepsilon, 1+\varepsilon)\right] \geq 1-\delta.$$

Proof The probability that $\frac{\bar{r_i}}{r_i} \notin (1-\bar{\varepsilon}, 1+\bar{\varepsilon})$ for some i is, by the union bound at most $m\bar{\delta} \leq \delta$.

Thus, with probability at least $1-\delta$, all the ratios $\frac{\bar{r_i}}{r_i}$ are in $(1-\bar{\varepsilon}, 1+\bar{\varepsilon})$. Now, it suffices to show that

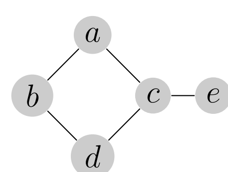
$$\left(1-\frac{\varepsilon}{2m}\right)^m \geq 1-\varepsilon \text{ and } \left(1+\frac{\varepsilon}{2m}\right)^m \leq 1+\varepsilon.$$

The first inequality follows from

If $\frac{|S_{i+1}|}{|S_i|} \leq r$ for every i, then the total number of samples needed is thus $mO\left(t\frac{m^2}{\varepsilon^2}\log(\frac{m}{\delta})\right) = O\left(\frac{tm^3}{\varepsilon^2}\log(\frac{m}{\delta})\right)$.

2 Counting Independent Sets

Given an undirected graph $G = (V, E)$, we denote by $\Omega(G)$ the set of independent sets in G. We wish to compute $|\Omega(G)|$ approximately. For example, in the graph shown below, $\Omega(G)$ consists of $\emptyset$,$\{a\}$, $\{b\}$, $\{c\}$, $\{d\}$, $\{e\}$, $\{a,d\}$, $\{a,e\}$, $\{b,c\}$, $\{b,e\}$, $\{d,e\}$, and $\{a,d,e\}$. Thus, $|\Omega(G)| = 12$.



To use the idea in the previous section, we need to define a suitable sequence $S_0, S_1, \ldots, S_m$ such that $S_0 = \Omega(G)$. Let the edges of G be $e_1, e_2, \ldots, e_m$. We define $G_0 = G$ and for $i \geq 1$, $G_i = G \setminus \{e_1, \ldots, e_i\}$. Thus G_m is the empty graph. We now let $S_i = \Omega(G_i)$. Clearly $S_i \subset S_{i+1}$ and $|S_m| = 2^n$.

We thus need to bound the ratios $\dfrac{|S_{i+1}|}{|S_i|}$ and also find a way to uniformly sample from $\Omega(G)$ for any graph G.

We first show a bound on the ratios: $\dfrac{|\Omega(G_{i+1})|}{|\Omega(G_i)|} \leq 2$. That is, if G is any graph and $\{u, v\} \in E(G)$, then $|\Omega(G \setminus \{u, v\})| \leq 2|\Omega(G)|$. To see this, consider an independent set $I \in \Omega(G \setminus \{u, v\}) \setminus \Omega(G)$. Then I must contain both u and v. Clearly, $I \setminus \{v\}$ is an independent set in G; the map from $\Omega(G \setminus \{u, v\}) \setminus \Omega(G)$ to $\Omega(G)$ that takes I to $I \setminus \{v\}$ is injective; thus $|\Omega(G \setminus \{u, v\})| \leq |\Omega(G)|$ and the desired claim follows.

2.1 Sampling via random walks

We now turn to the problem of sampling uniformly from $\Omega(G)$. For this, we define a biased random walk on $\Omega(G)$ and then show that the stationary distribution of this random walk is the uniform distribution. When walk is executed for sufficiently large number of time-steps, the resulting state will be close to uniform; however we shall be unable to prove any bound on this; indeed only for graphs of maximum degree at most 5 does this walk converge fast enough.

The random walk is defined as follows.

- Pick an arbitrary vertex v and set $X_0 = v$.
- For $i = 1$ to t
- Pick a vertex w u.a.r from V.
- If $X_i \cup \{w\}$ is an ind-set, $X_{i+1} = X_i \cup \{w\}$.
- Else if $w \in X_i$, $X_{i+1} = X_i \setminus \{w\}$.
- Else $X_{i+1} = X_i$.

Note that if I, J are adjacent, then I, J differ in exactly one vertex, and the transition from I to J is made precisely when that vertex is chosen in step

3. Thus $M_{I,J} = \dfrac{1}{|V|}$. Also, if $deg(I) = d$, then $M_{I,I} = 1 - \dfrac{d}{|V|}$.

To show that the stationary distribution is the uniform distribution, we observe and use the fact that $\sum_{J \in N(I)} M_{J,I} = 1$, with the help of the following result.

Theorem 1 *For an irreducible, aperiodic Markov chain, if*

$$\sum_{u \in N(v)} M_{u,v} = 1 \forall v \in V,$$

then π^* is the uniform distribution.

Proof We have:

$$\pi^*(v) = \sum_{u \in N(v)} M_{u,v} \pi^*(u). \tag{1}$$

Since $\sum_{u \in N(v)} M_{u,v} = 1$, it follows that the value of $\pi^*(v)$ is a weighted average of the value of the neighbors of v, and is thus constant. ■

3 Reversible chains and the Metropolis Algorithm

We now focus on the problem of finding transition probabilities to achieve a given target distribution as the stationary distribution. For this, we need the concept of reversible Markov chains. Given a Markov chain $X_1, \ldots, X_n, \ldots$, the reverse sequence $X_m, X_{m-1}, \ldots, X_0$ for any given m, also satisfies the Markovian (memoryless) property.

The transition probability is given by:

$$Pr[X_k = v | X_{k+1} = u] = \frac{Pr[X_k = v \wedge X_{k+1} = u]}{Pr[X_{k+1} = u]} = \frac{Pr[X_k = v] M_{v,u}}{Pr[X_{k+1} = u]}.$$

Taking limits as $m \to \infty$, we obtain: $Q_{u,v} = \dfrac{\pi^*(v)}{\pi^*(v)} M_{v,u}$.

We say that a Markov chain is reversible if the reverse Markov chain has the same transition probabilities, that is if $Q_{u,v} = M_{u,v}$. Note that this means $M_{u,v} = \dfrac{\pi^*(v)}{\pi^*(v)} M_{v,u}$ for all u, v. The following result shows that whenever the above condition holds for a distribution π, it must be the stationary distribution.

Theorem 2 *For an irreducible, aperiodic Markov chain, if*

$$\pi(u)M_{u,v} = \pi(v)M_{v,u} \forall u, v,$$

and $\sum_v \pi(v) = 1$, *then* π *is the stationary undistribution.*

Proof Can be seen from substitution in Equation 1. ■

We now use the above result to design transition probabilities for which a given distribution π is the stationary distribution.

Metropolis algorithm:

- Choose a constant c such that $c \leq \dfrac{1}{deg(v)}$ for all v.
- Set $M_{u,v} = c\dfrac{\pi(v)}{\pi(u)}$ if $\pi(v) \leq \pi(u)$. In this case, $M_{v,u}$ is set to c.
- Else set $M_{u,v} = c$. [This is the reverse of the previous situation.]
- Set $M_{v,v} = 1 - \sum_{u \in N(v)} M_{v,u}$.

The last condition ensures that the sum of the outgoing transition probabilities sums to one, for every vertex. The chosen values satisfy the equation in Theorem 2, thus the stationary distribution for the Markov chain is π, as desired.