Randomized Algorithms

Lecture 14: Markov Chain Monte Carlo and Approximate Counting

Raul Garcia-Patron

School of Informatics
University of Edinburgh

Markov chain Monte Carlo (MCMC)

The Markov chain Monte Carlo (MCMC) method provides a very general approach to sampling from a desired probability distribution.

- The idea is to build a *Markov chain M* on the state space Ω that we want to sample from.
- We ensure the stationary distribution of the Markov chain is unique and corresponds to the target distribution.
- We can then run M to generate a sequence of $X_0, X_1, \ldots, X_k$ of states so X_k distribution is the stationary distribution: x_k is our output sample.
- How large k has to be to have a valid sample is called mixing-time.
- Knowing the mixing-time of a Markov chain is non-trivial and will be the core of the last section of the course.

MCMC for Independent Sets

- Given an input graph $G = (V, E)$, an IS is subsets $I \subseteq V$ which satisfy $|I \cap \{u, v\}| = 0$ for all u, v such that $e = (u, v) \in E$.
- Our interest is to sample from the uniform distribution over the state space Ω.

MCMC for Independent Sets: Algorithm

The IS Markov chain generates a random sequence of ISs:

Algorithm GENERATEIS$(n; G = (V, E))$

1. Start with an arbitrary IS X_0
2. **for** $i \leftarrow 0$ **to** "whenever"
3. Choose v uniformly at random from V.
4. **if** $v \in X_i$ **then**
5. $X_{i+1} \leftarrow X_i \setminus \{v\}$
6. **elseif** ($v \notin X_i$ **and** $X_i \cup \{v\}$ is also an IS in G) **then**
7. $X_{i+1} \leftarrow X_i \cup \{v\}$
8. **else** $X_{i+1} \leftarrow X_i$

Unique stationary distribution

- If a Markov chain is finite, irreducible, aperiodic:
  - The chain has an unique stationary distribution.
- Time-reversal or detailed balance: if $\sum_{i=0}^{\infty} \pi_i = 1$ and

$$\pi_i P_{i,j} = \pi_j P_{j,i} \tag{1}$$

  then π is the stationary distribution of P.
- Having unique stationary distribution does not give us a Fully Polynomial Almost Uniform Sampler (FPAUS) for Ω. We need to also show the chain is *rapidly mixing*.

MCMC for Independent Sets: convergence to stationary

Algorithm GENERATEIS$(n; G = (V, E))$

1. Start with an arbitrary IS X_0
2. **for** $i \leftarrow 0$ **to** "whenever"
3. Choose v uniformly at random from V.
4. **if** $v \in X_i$ **then**
5. $X_{i+1} \leftarrow X_i \setminus \{v\}$
6. **elseif** ($v \notin X_i$ **and** $X_i \cup \{v\}$ is also an IS in G) **then**
7. $X_{i+1} \leftarrow X_i \cup \{v\}$
8. **else** $X_{i+1} \leftarrow X_i$

- Finite: Yes.
- Irreducible: there is always a path between two configurations.
- Aperiodicity: $\exists$ self-loops.
- Detail balance?

MCMC for Independent Sets: Irreducible

- For a finite state space Ω. Let call the set of states reachable in one step from state x **the neighbors of** x**, denoted by** $N(x)$. We also have that if $y \in N(x)$ then also $x \in N(y)$.
- For any starting IS x and final IS y there is always a connecting path:
  - All vertices that belong to $X \cup y$ are divide into: $x \setminus y$ (in x but not in y), $x \cap y$ (in both) and $y \setminus x$ (in y but not in x).
  - To move from configuration x to y, remove all $x \setminus y$ one by one and then add all $y \setminus x$ one by one.
- The connecting path has non-zero probability:
  - Adjacent IS state neighbors differ in a single vertex of G. Probability of the jump is $1/|V|$, i.e., probability you select the right vertex v allowing the transition.

MCMC for Independent Sets: Detail balance

From the previous slide, we know M_{IS} has a unique stationary distribution π_{IS}, but not what it *is*. We now show it must be the uniform one.

- Detail balance:

$$\forall i, j : \pi_i P_{i,j} = \pi_j P_{j,i} \tag{2}$$

- Adjacent IS state neighbors differ in a single vertex of G. Assume $X = Y + v$.
  - $P_{x,y}$: we jump from Y to X only if vertex v is selected (probability $1/|V|$) followed by the algorithm deterministically adding v to X (line 7).
  - $P_{y,x}$: we jump from X to Y only if vertex v is selected (probability $1/|V|$) followed by the algorithm deterministically removing v to Y (line 5).
- Because $P_{x,y} = P_{y,x}$ detail balance $\Rightarrow \pi_x = \pi_y = 1/|\Omega|$.
- You extend the equality between any pair using the same paths as be defined for irreducibility.

From Sampling to Approximate Counting

Last lecture DNF was an example of *uniform sampling from the target set* that can be used to obtain an FPRAS to approximately count the elements:

$$\text{Estimation of } |S| = \frac{\#a \in S}{m}|U|$$

In what follows we are going to explore how to transform a sampling algorithm into a counting one.

- Won't always have an immediately-samplable "superset" like U whose cardinality is bigger by a low factor like T.
- Won't always be able to do *exact* uniform sampling from the bigger set, that may sometimes be *almost-uniform* instead.

ϵ-uniform sampler and FPAUS

Definition (Definition 11.3)

Let ω be the (random) output of a sampling algorithm for a finite sample space Ω. Then a sampling algorithm is said to generate an ϵ-uniform sample of Ω if for every $S \subset \Omega$,

$$\left| \Pr[\omega \in S] - \frac{|S|}{|\Omega|} \right| \leq \epsilon.$$

Definition (FPAUS)

A sampling algorithm is a **fully-polynomial almost uniform sampler (FPAUS)** for a problem if, given input x and a parameter $\epsilon > 0$, it generates a ϵ-uniform sample of $\Omega(x)$ after running in time polynomial in $\ln(\frac{1}{\epsilon})$ and the size of x.

Independent sets ordering

Imagine that we have an "off the shelf" fully polynomial approximation uniform sampler (FPAUS) for sampling independent sets of an input graph. We show how to create a **fully polynomial approximation scheme (FPRAS)** from this.

Definition (IS)

For a given undirected graph $G = (V, E)$, the subset $I \subseteq V$ is said to be an *independent set* if for every $e \in E$, $e = (u, v)$, at most one of u, v lie in I.

Definition (Ordering of IS)

For a given graph $G = (V, E)$ consider some ordering $e_1, e_2, \ldots, e_m$ of the edges of E.

- For every $i = 1, \ldots, m$, set $E_i = \cup_{j=1}^{i}\{e_j\}$, and $G_i = (V, E_i)$.
- Let $\Omega(G_i)$ be the number of Independent sets in G_i.

Observe that G_0 is an n-vertex graph with no edges, and G_m is G. Each G_{i+1} is G_i with an extra edge added.

Telescopic product

Now consider the following *telescoping product*:

$$|\Omega(G)| = \frac{|\Omega(G_m)|}{|\Omega(G_{m-1})|} \times \frac{|\Omega(G_{m-1})|}{|\Omega(G_{m-2})|} \times \frac{|\Omega(G_{m-2})|}{|\Omega(G_{m-3})|} \times \ldots \times \frac{|\Omega(G_1)|}{|\Omega(G_0)|} \times |\Omega(G_0)|.$$

- $|\Omega(G_0)| = 2^n$ as every subset of V is an I.S. for G_0 (G_0 has no edges!).
- We will show how to obtain close approximate values $\tilde{r}_i$ for each ratio $r_i = \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}$, for $i = 1, \ldots, m$.
- Our *estimate* for the number of I.S.s will be:

$$2^n \prod_{i=1}^{m} \tilde{r}_i.$$

Proof of FPRAS via telescopic product

It is possible to show the following lemma:

Lemma (Lemma 11.4)

When $m \geq 1$ and $0 < \epsilon \leq 1$, $\exists$ a $(\frac{\epsilon}{2m}, \frac{\delta}{m})$-approximation for the quantity r_i using Algorithm EstimRatio.

1. We run Algorithm EstimRatio for each $\frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|}$ to obtain estimates $\tilde{r}_m, \tilde{r}_{m-1}, \ldots, \tilde{r}_2, \tilde{r}_1$.
2. By Lemma 11.4, $\Pr[|\frac{\tilde{r}_i}{r_i} - 1| > \frac{\epsilon}{2m}] \leq \frac{\delta}{m}$, for every $1 \leq i \leq m$.
3. $\Pr[\cap_{i=1}^{m} |\frac{\tilde{r}_i}{r_i} - 1| < \frac{\epsilon}{2m}] = 1 - Pr[\cup_{i=1}^{m} |\frac{\tilde{r}_i}{r_i} - 1| > \frac{\epsilon}{2m}]$
4. Hence (Union Bound on bad events):
   $Pr[\cap_{i=1}^{m} |\frac{\tilde{r}_i}{r_i} - 1| < \frac{\epsilon}{2m}] \geq 1 - \sum_{i=1}^{m} Pr[|\frac{\tilde{r}_i}{r_i} - 1| > \frac{\epsilon}{2m}] \geq 1 - \delta$.
5. So with probability of at least $1 - \delta$, we have:

$$\left(1 - \frac{\epsilon}{2m}\right)^m \leq \prod_{i=1}^{m} \frac{\tilde{r}_i}{r_i} \leq \left(1 + \frac{\epsilon}{2m}\right)^m.$$

Proof of FPRAS via telescopic product II

1. So with probability of at least $1 - \delta$, we have:

$$\left(1 - \frac{\epsilon}{2m}\right)^m \le \prod_{i=1}^{m} \frac{\tilde{r}_i}{r_i} \le \left(1 + \frac{\epsilon}{2m}\right)^m .$$

2. Easy to show (for $\epsilon < 1$): $1 - \epsilon \le (1 - \frac{\epsilon}{2m})^m$
3. Easy to show (for $\epsilon < 1$): $(1 + \frac{\epsilon}{2m})^m \le 1 + \epsilon$
4. Hence we have

$$\begin{aligned} 1 - \epsilon \ \le & \quad \textstyle\prod_{i=1}^{m} \frac{\tilde{r}_i}{r_i} \quad \le 1 + \epsilon, \\ (1-\epsilon)2^n \prod_{i=1}^{m} r_i \le & \quad 2^n \textstyle\prod_{i=1}^{m} \tilde{r}_i \quad \le (1+\epsilon)2^n \prod_{i=1}^{m} r_i \end{aligned}$$

5. We have an FPRAS for counting IS on G, i.e, $|\Omega(G)|$ with ϵ relative error with probability of failure of δ.

Algorithm EstimRatio

- Key idea: sample from $\Omega(G_{i-1})$ and check if in $\Omega(G_i)$.
- Uses the assumed FPAUS as a subroutine in step 4.

Algorithm EstimRatio$(G_{i-1} = (V, E_{i-1}); e_i)$

1. $count \leftarrow 0$
2. $G_i \leftarrow (V, E_{i-1} \cup \{e_i\})$
3. **for** $k \leftarrow 1$ **to** $M = \lceil 1296 m^2 \epsilon^{-2} \ln(\frac{2m}{\delta}) \rceil$
4. Generate a $\frac{\epsilon}{6m}$-uniform sample from $\Omega(G_{i-1})$.
5. **if** (the sample generated is *also* an I.S for G_i) **then**
6. $count \leftarrow count + 1$
7. **return** $\tilde{r}_i \leftarrow \frac{count}{M}$

We will compute a $\tilde{r}_i$ that is within $\pm\frac{\epsilon}{2m}$ of the true value with probability at least $1 - \frac{\delta}{m}$, for each $i, 1 \leq i \leq m$.

Intuition of Lemma 11.4

- G_{i-1} and G_i differ in a single edge $\{u, v\}$.
- An IS of G_{i-1} is also IS of G_i: $\Omega(G_i) \subseteq \Omega(G_{i-1})$.
- An independent set in $\Omega(G_{i-1}) \setminus \Omega(G_i)$ contains both u and v:
  - If it contains only one or none it belongs to $\Omega(G_i)$ already.
- We can associate to each IS $I \in \Omega(G_{i-1}) \setminus \Omega(G_i)$ with an IS $I \setminus \{v\} \in \Omega(G_i)$ (remove v), therefore:

$$\Omega(G_{i-1}) \setminus \Omega(G_i) \subseteq \Omega(G_i)$$

- We finally obtain:

$$r_i = \frac{|\Omega(G_i)|}{|\Omega(G_{i-1})|} = \frac{|\Omega(G_i)|}{|\Omega(G_i)| + |\Omega(G_{i-1}) \setminus \Omega(G_i)|} \geq \frac{1}{2}$$

- Further technical details are needed due to the sampling from $\Omega(G_{i-1})$ nor being exact.