CMPSCI 690RA: Randomized Algorithms

Lecture 7 – Markov Chains and Random Walks

Andrew McGregor

Last Compiled: February 13, 2020

Outline

Markov Chains

Random Walks on Graphs

Motivating Example

- An algorithm for 2-SAT:
  1. Pick arbitrary assignment.
  2. Pick an unsatisfied clause: randomly flip the value assigned to one of the two variables.
  3. Repeat Step 2 until there are no unsatisfied clauses.

Motivating Example

- An algorithm for 2-SAT:
  1. Pick arbitrary assignment.
  2. Pick an unsatisfied clause: randomly flip the value assigned to one of the two variables.
  3. Repeat Step 2 until there are no unsatisfied clauses.
- Let $x^{(t)}$ be the assignment at time t.

Motivating Example

- An algorithm for 2-SAT:
  1. Pick arbitrary assignment.
  2. Pick an unsatisfied clause: randomly flip the value assigned to one of the two variables.
  3. Repeat Step 2 until there are no unsatisfied clauses.
- Let $x^{(t)}$ be the assignment at time t.
- Consider $X^{(t)} = n - \Delta(x^{(t)}, y)$ for a fixed satisfying assignment y, i.e., the number of variables that are set the same in both $x^{(t)}$ and y

Motivating Example

- An algorithm for 2-SAT:
  1. Pick arbitrary assignment.
  2. Pick an unsatisfied clause: randomly flip the value assigned to one of the two variables.
  3. Repeat Step 2 until there are no unsatisfied clauses.
- Let $x^{(t)}$ be the assignment at time t.
- Consider $X^{(t)} = n - \Delta(x^{(t)}, y)$ for a fixed satisfying assignment y, i.e., the number of variables that are set the same in both $x^{(t)}$ and y
- $X^{(t+1)} = X^{(t)} \pm 1$ and

$$\mathbb{P}\left[X^{(t+1)} = X^{(t)} + 1\right] \geq 1/2$$

  since either 1 or 2 of the variables in the clause we pick have a different assignment that in y.

Motivating Example

- An algorithm for 2-SAT:
  1. Pick arbitrary assignment.
  2. Pick an unsatisfied clause: randomly flip the value assigned to one of the two variables.
  3. Repeat Step 2 until there are no unsatisfied clauses.
- Let $x^{(t)}$ be the assignment at time t.
- Consider $X^{(t)} = n - \Delta(x^{(t)}, y)$ for a fixed satisfying assignment y, i.e., the number of variables that are set the same in both $x^{(t)}$ and y
- $X^{(t+1)} = X^{(t)} \pm 1$ and

$$\mathbb{P}\left[X^{(t+1)} = X^{(t)} + 1\right] \geq 1/2$$

  since either 1 or 2 of the variables in the clause we pick have a different assignment that in y.
- How long until we terminate?

Markov Chain

A Markov chain is a discrete-time stochastic process that defines a sequence of random variables $(X_0, X_1, X_2, \ldots)$ and is defined by:

- State space: e.g., $X_t \in \{1, \ldots, n\}$
- Transition probabilities: $P_{ij} = \mathbb{P}[X_t = j | X_{t-1} = i]$
- Initial distribution for X_0.

Markov Chain

A Markov chain is a discrete-time stochastic process that defines a sequence of random variables $(X_0, X_1, X_2, \ldots)$ and is defined by:

- State space: e.g., $X_t \in \{1, \ldots, n\}$
- Transition probabilities: $P_{ij} = \mathbb{P}[X_t = j | X_{t-1} = i]$
- Initial distribution for X_0.

Memoryless: X_t only depends on X_{t-1}:

$$\begin{aligned} \mathbb{P}[X_t = i_t | X_{t-1} = i_{t-1}, \ldots, X_1 = i_1, X_0 = i_0] &= \mathbb{P}[X_t = i_t | X_{t-1} = i_{t-1}] \\ &= P_{i_{t-1} i_t} \end{aligned}$$

State Probability Vector and Stationary Distribution

Definition

Let $q_i^{(t)} = \mathbb{P}[X_t = i]$. The distribution of the chain at time t is the row vector $(q_1^{(t)}, q_2^{(t)}, \ldots, q_n^{(t)})$.

Proof.

Follows since $q_j^{(t)} = \sum_i q_i^{(t-1)} P_{i,j}$. □

State Probability Vector and Stationary Distribution

Definition

Let $q_i^{(t)} = \mathbb{P}[X_t = i]$. The distribution of the chain at time t is the row vector $(q_1^{(t)}, q_2^{(t)}, \ldots, q_n^{(t)})$.

Lemma

$q^{(t)} = q^{(0)} P^t$ where P be the matrix whose (i,j)-th entry is P_{ij}.

Proof.

Follows since $q_j^{(t)} = \sum_i q_i^{(t-1)} P_{i,j}$. □

State Probability Vector and Stationary Distribution

Definition

Let $q_i^{(t)} = \mathbb{P}[X_t = i]$. The distribution of the chain at time t is the row vector $(q_1^{(t)}, q_2^{(t)}, \ldots, q_n^{(t)})$.

Lemma

$q^{(t)} = q^{(0)} P^t$ where P be the matrix whose (i,j)-th entry is P_{ij}.

Proof.

Follows since $q_j^{(t)} = \sum_i q_i^{(t-1)} P_{i,j}$. □

Definition

A stationary distribution for a Markov chain with transition matrix P is a probability distribution π such that $\pi = \pi P$.

Underlying Graphs and Irreducible Markov Chains

Definition

The underlying graph of a Markov chain is a directed graph $G = (V, E)$ where $V = [n]$ and $(i, j) \in E$ iff $P_{ij} > 0$.

Underlying Graphs and Irreducible Markov Chains

Definition

The underlying graph of a Markov chain is a directed graph $G = (V, E)$ where $V = [n]$ and $(i, j) \in E$ iff $P_{ij} > 0$.

Definition

A Markov chain is irreducible if its underlying graph consists of a single strong component. (Recall that a strong component of a directed graph is a maximal subgraph C of G such that for any vertices i and j in C, there is a directed path from i to j.)

Transient States and Persistent States

Probability that t is the first time that the chain visits state j if it starts at state i:

$$r_{ij}^{(t)} = \mathbb{P}\left[X_t = j \text{ and } X_s \neq j \text{ for all } 1 \leq s < t | X_0 = i\right]$$

Transient States and Persistent States

Probability that t is the first time that the chain visits state j if it starts at state i:

$$r_{ij}^{(t)} = \mathbb{P}\left[X_t = j \text{ and } X_s \neq j \text{ for all } 1 \leq s < t | X_0 = i\right]$$

Probability that the chain ever visits state j if is starts at state i:

$$f_{ij} = \sum_{t>0} r_{ij}^{(t)}$$

Transient States and Persistent States

Probability that t is the first time that the chain visits state j if it starts at state i:

$$r_{ij}^{(t)} = \mathbb{P}\left[X_t = j \text{ and } X_s \neq j \text{ for all } 1 \leq s < t | X_0 = i\right]$$

Probability that the chain ever visits state j if is starts at state i:

$$f_{ij} = \sum_{t>0} r_{ij}^{(t)}$$

Expected time to until a visit to state j if is starts at state i:

$$h_{ij} = \sum_{t>0} t r_{ij}^{(t)} \text{ if } f_{ij} = 1 \text{ and } h_{ij} = \infty \text{ otherwise}$$

Transient States and Persistent States

Probability that t is the first time that the chain visits state j if it starts at state i:

$$r_{ij}^{(t)} = \mathbb{P}\left[X_t = j \text{ and } X_s \neq j \text{ for all } 1 \leq s < t | X_0 = i\right]$$

Probability that the chain ever visits state j if is starts at state i:

$$f_{ij} = \sum_{t>0} r_{ij}^{(t)}$$

Expected time to until a visit to state j if is starts at state i:

$$h_{ij} = \sum_{t>0} t r_{ij}^{(t)} \text{ if } f_{ij} = 1 \text{ and } h_{ij} = \infty \text{ otherwise}$$

Definition

If $f_{ii} < 1$ then state i is transient. If $f_{ii} = 1$ then state i is persistent. Those persitent states i for which $h_{ii} = \infty$ are said to be null peristent and those for which $h_{ii} \neq \infty$ are said to be non-null persistent.

Periodicity and Ergodicity

Definition

The periodicity of state i is the maximum integer T for which there exists an initial distribution $q^{(0)}$ and positive integer a such that for all t we have: $q_i^{(t)} > 0$ implies $t \in \{a + Ti : i \geq 0\}$. If $T > 1$ then state is periodic and aperiodic otherwise. If every state is aperiodic then the Markov chain is aperiodic.

Periodicity and Ergodicity

Definition

The periodicity of state i is the maximum integer T for which there exists an initial distribution $q^{(0)}$ and positive integer a such that for all t we have: $q_i^{(t)} > 0$ implies $t \in \{a + Ti : i \geq 0\}$. If $T > 1$ then state is periodic and aperiodic otherwise. If every state is aperiodic then the Markov chain is aperiodic.

Definition

An ergodic state is one that is aperiodic and non-null persistent. A Markov chain is ergodic if *all* states are ergodic.

Fundamental Theorem of Markov Chains

Theorem

Any irreducible, finite, and aperiodic Markov chain has the following properties:

1. *All states are ergodic.*
2. *There is a unique stationary distribution* π *with* $\pi_i > 0$*.*
3. $f_{ii} = 1$ *and* $h_{ii} = 1/\pi_i$*.*
4. *Let* $N(i, t)$ *be the number of times the Markov chain visits state* i *in* t *steps. Then,*

$$\lim_{t \to \infty} \frac{N(i, t)}{t} = \pi_i$$

Outline

Markov Chains

Random Walks on Graphs

Random Walks on Graphs

A connected, non-bipartite, undirected graph $G = (V, E)$ defines Markov Chain M_G with states V and transition matrix:

$$P_{uv} = \begin{cases} 1/d(u) & \text{if } (u, v) \in E \\ 0 & \text{otherwise} \end{cases}$$

where $|V| = n$ and $|E| = m$.

Lemma

M_G is irreducible & aperiodic because it's connected & non-bipartite.

Random Walks on Graphs

A connected, non-bipartite, undirected graph $G = (V, E)$ defines Markov Chain M_G with states V and transition matrix:

$$P_{uv} = \begin{cases} 1/d(u) & \text{if } (u, v) \in E \\ 0 & \text{otherwise} \end{cases}$$

where $|V| = n$ and $|E| = m$.

Lemma

M_G is irreducible & aperiodic because it's connected & non-bipartite.

Lemma (Stationary Distribution of M_G)

For all $v \in V$, $\pi_v = d(v)/(2m)$.

Random Walks on Graphs

A connected, non-bipartite, undirected graph $G = (V, E)$ defines Markov Chain M_G with states V and transition matrix:

$$P_{uv} = \begin{cases} 1/d(u) & \text{if } (u,v) \in E \\ 0 & \text{otherwise} \end{cases}$$

where $|V| = n$ and $|E| = m$.

Lemma

M_G is irreducible & aperiodic because it's connected & non-bipartite.

Lemma (Stationary Distribution of M_G)

For all $v \in V$, $\pi_v = d(v)/(2m)$.

Proof.

Let $\Gamma(v)$ denote the graph neighborhood of v:

$$\sum_{u,v} \pi_u P_{u,v} = \sum_{u \in \Gamma(v)} \frac{d(u)}{2m} \cdot \frac{1}{d(u)} = \frac{d(v)}{2m} = \pi_v$$

□

Hitting Time

Definition

The hitting time h_{uv} is the expected time taken by a random walk starting at node u until it arrives at v.

Hitting Time

Definition

The hitting time h_{uv} is the expected time taken by a random walk starting at node u until it arrives at v.

Theorem

For any edge $(u, v) \in E$, $h_{vu} < 2m$.

Hitting Time

Definition

The hitting time h_{uv} is the expected time taken by a random walk starting at node u until it arrives at v.

Theorem

For any edge $(u, v) \in E$, $h_{vu} < 2m$.

Proof.

- By Fundamental Theorem: $h_{uu} = \pi_u^{-1} = 2m/d(u)$

□

Hitting Time

Definition

The hitting time h_{uv} is the expected time taken by a random walk starting at node u until it arrives at v.

Theorem

For any edge $(u, v) \in E$, $h_{vu} < 2m$.

Proof.

- By Fundamental Theorem: $h_{uu} = \pi_u^{-1} = 2m/d(u)$
- But we can also write h_{uu} as

$$h_{uu} = \frac{1}{d(u)} \sum_{w \in \Gamma(u)} (1 + h_{wu})$$

□

Hitting Time

Definition

The hitting time h_{uv} is the expected time taken by a random walk starting at node u until it arrives at v.

Theorem

For any edge $(u, v) \in E$, $h_{vu} < 2m$.

Proof.

- By Fundamental Theorem: $h_{uu} = \pi_u^{-1} = 2m/d(u)$
- But we can also write h_{uu} as

$$h_{uu} = \frac{1}{d(u)} \sum_{w \in \Gamma(u)} (1 + h_{wu})$$

- Hence,

$$2m = \sum_{w \in \Gamma(u)} (1 + h_{wu}) > h_{vu}$$

□

Cover Time (1/2)

Definition

$\mathcal{C}_u(G)$ denote the expected time to visit every node after starting at u.
The cover time of G is $\mathcal{C}(G) = \max_u \mathcal{C}_u(G)$.

Cover Time (1/2)

Definition

$\mathcal{C}_u(G)$ denote the expected time to visit every node after starting at u.
The cover time of G is $\mathcal{C}(G) = \max_u \mathcal{C}_u(G)$.

Theorem

$\mathcal{C}(G) < 4mn$ where $|V| = n$ and $|E| = m$.

Cover Time (1/2)

Definition

$\mathcal{C}_u(G)$ denote the expected time to visit every node after starting at u.
The cover time of G is $\mathcal{C}(G) = \max_u \mathcal{C}_u(G)$.

Theorem

$\mathcal{C}(G) < 4mn$ where $|V| = n$ and $|E| = m$.

Corollary

Expected time of at 2-SAT algorithm is $O(n^2)$.

Cover Time (2/2)

Proof.

□

Cover Time (2/2)

Proof.

- Let T be a spanning tree of G and $v_0 \in V$

$\square$

Cover Time (2/2)

Proof.

- Let T be a spanning tree of G and $v_0 \in V$
- Let $v_0, v_1, \ldots, v_{2n-2} = v_0$ be the vertices visited in a traversal of T that traverses each edge exactly twice.

□

Cover Time (2/2)

Proof.

- Let T be a spanning tree of G and $v_0 \in V$
- Let $v_0, v_1, \ldots, v_{2n-2} = v_0$ be the vertices visited in a traversal of T that traverses each edge exactly twice.
- Then

$$\mathcal{C}_{v_0}(G) \leq \sum_{j=0}^{2n-3} h_{v_j, v_{j+1}}$$

□

Cover Time (2/2)

Proof.

- Let T be a spanning tree of G and $v_0 \in V$
- Let $v_0, v_1, \ldots, v_{2n-2} = v_0$ be the vertices visited in a traversal of T that traverses each edge exactly twice.
- Then

$$\mathcal{C}_{v_0}(G) \leq \sum_{j=0}^{2n-3} h_{v_j, v_{j+1}}$$

- We know $h_{v_j, v_{j+1}} \leq 2m$ because (u, v) is an edge

□

Cover Time (2/2)

Proof.

- Let T be a spanning tree of G and $v_0 \in V$
- Let $v_0, v_1, \ldots, v_{2n-2} = v_0$ be the vertices visited in a traversal of T that traverses each edge exactly twice.
- Then

$$\mathcal{C}_{v_0}(G) \leq \sum_{j=0}^{2n-3} h_{v_j, v_{j+1}}$$

- We know $h_{v_j, v_{j+1}} \leq 2m$ because (u, v) is an edge
- Hence, $\mathcal{C}_{v_0}(G) \leq 2(n-1)2m$ for all v_0.

□