Lecture 9: Random Walks, Markov Chains, Mixing Time

Rafael Oliveira

University of Waterloo
Cheriton School of Computer Science

rafael.oliveira.teaching@gmail.com

October 5, 2021

Overview

- Introduction
  - Why Random Walks & Markov Chains?
  - Basics on Theory of Finite Markov Chains
- Main Topics
  - Stationary Distributions and Mixing Time
  - Fundamental Theorem of Markov Chains
- Conclusion
- Acknowledgements

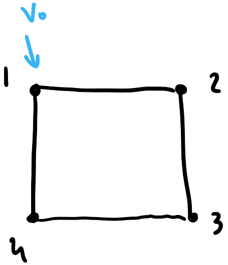
What is a Random Walk?

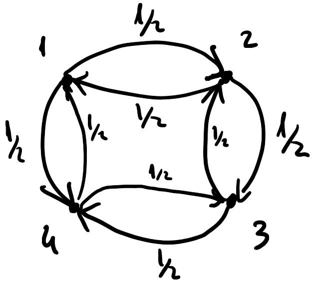
vertices of graph ↔ state of random walk

Given a graph $G(V, E)$

1. random walk starts from a vertex v_0
2. at each time step it moves *uniformly* to a *random neighbor* of the current vertex in the graph

$$v_{t+1} \leftarrow_R N_G(v_t)$$





1 → 2 → 1 → 2 → 3 → 4

What is a Random Walk?

Given a graph $G(V, E)$

1. random walk starts from a vertex v_0
2. at each time step it moves *uniformly* to a *random neighbor* of the current vertex in the graph

$$v_{t+1} \leftarrow_R N_G(v_t)$$

Basic questions involving random walks:

What is a Random Walk?

Given a graph $G(V, E)$

1. random walk starts from a vertex v_0
2. at each time step it moves *uniformly* to a *random neighbor* of the current vertex in the graph

$$v_{t+1} \leftarrow_R N_G(v_t)$$

Basic questions involving random walks:

- *Stationary distribution:* does the random walk converge to a "stable" distribution? If it does, what is this distribution?

if random walk has been happening for a while
what's is the probability of being in a given vertex?

What is a Random Walk?

Given a graph $G(V, E)$

1. random walk starts from a vertex v_0
2. at each time step it moves *uniformly* to a *random neighbor* of the <u>current vertex</u> in the graph

$$v_{t+1} \leftarrow_R N_G(v_t)$$

Basic questions involving random walks:

- *Stationary distribution:* does the random walk converge to a "stable" distribution? If it does, what is this distribution?
- *Mixing time:* how long does it take for the walk to converge to the stationary distribution?

What is a Random Walk?

Given a graph $G(V, E)$

1. random walk starts from a vertex v_0
2. at each time step it moves *uniformly* to a *random neighbor* of the current vertex in the graph

$$v_{t+1} \leftarrow_R N_G(v_t)$$

Basic questions involving random walks:

- *Stationary distribution:* does the random walk converge to a "stable" distribution? If it does, what is this distribution?
- *Mixing time:* how long does it take for the walk to converge to the stationary distribution?
- *Hitting time:* starting from a vertex v_0, what is expected number of steps until it reaches a vertex v_f?

What is a Random Walk?

Given a graph $G(V, E)$

1. random walk starts from a vertex v_0
2. at each time step it moves *uniformly* to a *random neighbor* of the <u>current vertex</u> in the graph

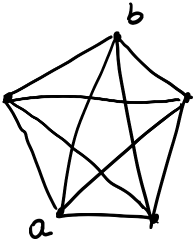
$$v_{t+1} \leftarrow_R N_G(v_t)$$

Basic questions involving random walks:

- *Stationary distribution:* does the random walk converge to a "stable" distribution? If it does, what is this distribution?
- *Mixing time:* how long does it take for the walk to converge to the stationary distribution?
- *Hitting time:* starting from a vertex v_0, what is expected number of steps until it reaches a vertex v_f?
- *Cover time:* how long does it take to reach every vertex of the graph at least once?

Random Walk: Example

- Suppose $G(V, E) = K_n$, the complete graph, $a, b \in V$ two vertices



K_5

Random Walk: Example

- Suppose $G(V,E) = K_n$, the complete graph, $a, b \in V$ two vertices
    1. What is expected number of steps to reach b in simple random walk starting at a? (i.e., hitting time)

$$a \neq b$$

$$\Pr[a \to b] = \frac{1}{n-1}$$

$$E[T_{a,b}] = \sum_{k \geq 1} k \cdot \underbrace{\Pr[T_{a,b} = k]}_{\left(\frac{n-2}{n-1}\right)^{k-1} \cdot \frac{1}{n-1}} = \sum_{k \geq 1} k \cdot \left(\frac{n-2}{n-1}\right)^{k-1} \cdot \frac{1}{n-1}$$

$$= \sum_{k \geq 1} \underbrace{\Pr[T_{a,b} \geq k]}_{= \sum_{j \geq n} \Pr[T_{a,b} = j]} = \sum_{n \geq 1} \left(\frac{n-2}{n-1}\right)^{k-1} = \frac{1}{1 - \frac{n-2}{n-1}}$$

$$= n-1$$

Random Walk: Example

- Suppose $G(V, E) = K_n$, the complete graph, $a, b \in V$ two vertices
  1. What is expected number of steps to reach b in simple random walk starting at a? (i.e., hitting time)
  2. Starting from a, what is expected number of steps to visit all vertices? (i.e, cover time)

Practice problem: (you have seen this before)

Random Walk: Example

- Suppose $G(V, E) = K_n$, the complete graph, $a, b \in V$ two vertices
  1. What is expected number of steps to reach b in simple random walk starting at a? (i.e., hitting time)
  2. Starting from a, what is expected number of steps to visit all vertices? (i.e, cover time)
  3. Stationary Distribution?

Yes: $\boxed{\pi = \left(\frac{1}{n}, \frac{1}{n}, \dots, \frac{1}{n}\right)}$ uniform distribution

$$\Pr\left[\text{being at } a \text{ at time } t \;\middle|\; \text{uniform at time } t-1\right] =$$

$$= \sum_{i=1}^{n} \Pr\left[X_t = a \mid X_{t-1} = i\right] \underbrace{\Pr\left[X_{t-1} = i\right]}_{1/n} = 0 + \underbrace{\frac{1}{n} \cdot \sum_{i \neq a} \frac{1}{n-1}}_{=1}$$

Random Walk: Example

- Suppose $G(V, E) = K_n$, the complete graph, $a, b \in V$ two vertices
  1. What is expected number of steps to reach b in simple random walk starting at a? (i.e., hitting time)
  2. Starting from a, what is expected number of steps to visit all vertices? (i.e, cover time)
  3. Stationary Distribution?
  4. Mixing time? (we'll do it later)

Random Walk: Example

- Suppose $G(V, E) = K_n$, the complete graph, $a, b \in V$ two vertices
  1. What is expected number of steps to reach b in simple random walk starting at a? (i.e., hitting time)
  2. Starting from a, what is expected number of steps to visit all vertices? (i.e, cover time)
  3. Stationary Distribution?
  4. Mixing time? (we'll do it later)
- **Practice question:** Compare question 2 to coupon collector problem!

What is a Markov Chain?

Random walk is a special kind of *stochastic process*:

$$\Pr[X_t = v_t \mid \underbrace{X_0 = v_0, \ldots, X_{t-1} = v_{t-1}}_{\text{history}}] = \Pr[X_t = v_t \mid \underbrace{X_{t-1} = v_{t-1}}_{\text{random walk only cares where we were last}}]$$

$X_t \leftarrow \underbrace{\text{state}}_{\text{vertex in graph}}$ of random walk at time t

What is a Markov Chain?

Random walk is a special kind of *stochastic process*:

$$\Pr[X_t = v_t \mid X_0 = v_0, \ldots, X_{t-1} = v_{t-1}] = \Pr[X_t = v_t \mid X_{t-1} = v_{t-1}]$$

Probability that we are at vertex v_t at time t only depends on the state of our process at time $t - 1$.

What is a Markov Chain?

Random walk is a special kind of *stochastic process*:

$$\boxed{\Pr[X_t = v_t \mid X_0 = v_0, \ldots, X_{t-1} = v_{t-1}] = \Pr[X_t = v_t \mid X_{t-1} = v_{t-1}]}$$

Probability that we are at vertex v_t at time t only depends on the state of our process at time $t-1$.

Process is "*forgetful*/*memoryless*"

Markov chain is characterized by this property.

Why study Markov Chains and Random Walks?

Markov Chains and Random Walks are ubiquitous in randomized algorithms.

Why study Markov Chains and Random Walks?

Markov Chains and Random Walks are ubiquitous in randomized algorithms.

- Page Rank algorithm (next lecture)

Why study Markov Chains and Random Walks?

Markov Chains and Random Walks are ubiquitous in randomized algorithms.

- Page Rank algorithm (next lecture)
- Approximation algorithms for counting problems [Karp, Luby & Madras]
    - Permanent of non-negative matrices [Jerrum, Vigoda & Sinclair] (*great final project topic!*)

Why study Markov Chains and Random Walks?

Markov Chains and Random Walks are ubiquitous in randomized algorithms.

- Page Rank algorithm (next lecture)
- Approximation algorithms for counting problems [Karp, Luby & Madras]
  - Permanent of non-negative matrices [Jerrum, Vigoda & Sinclair] (*great final project topic!*)
- Sampling Problems
  - Gibbs sampling in statistical physics (*great final project topic!*)
  - many more places

Why study Markov Chains and Random Walks?

Markov Chains and Random Walks are ubiquitous in randomized algorithms.

- Page Rank algorithm (next lecture)
- Approximation algorithms for counting problems [Karp, Luby & Madras]
  - Permanent of non-negative matrices [Jerrum, Vigoda & Sinclair] (*great final project topic!*)
- Sampling Problems
  - Gibbs sampling in statistical physics (*great final project topic!*)
  - many more places
- Probability amplification without too much randomness (efficient)
  - random walks on expander graphs (*great final project topic!*)

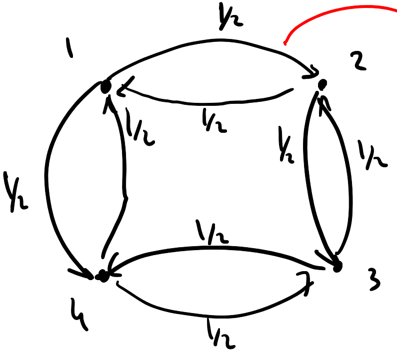
Why study Markov Chains and Random Walks?

Markov Chains and Random Walks are ubiquitous in randomized algorithms.

- Page Rank algorithm (next lecture)
- Approximation algorithms for counting problems [Karp, Luby & Madras]
  - Permanent of non-negative matrices [Jerrum, Vigoda & Sinclair] (*great final project topic!*)
- Sampling Problems
  - Gibbs sampling in statistical physics (*great final project topic!*)
  - many more places
- Probability amplification without too much randomness (efficient)
  - random walks on expander graphs (*great final project topic!*)
- many more

Representing Finite Markov Chains

Markov chain can be seen as weighted directed graph.

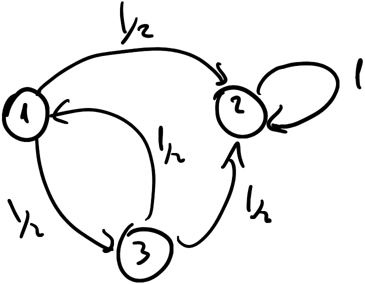


$\Pr[1 \rightsquigarrow 2] = 1/2$

Representing Finite Markov Chains

Markov chain can be seen as weighted directed graph.

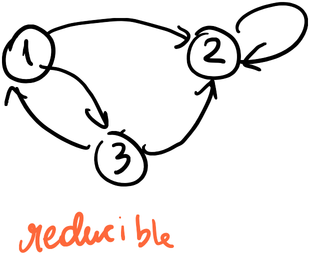
- Vertex is a state of Markov chain
- edge (i, j) corresponds to transition probability from i to j

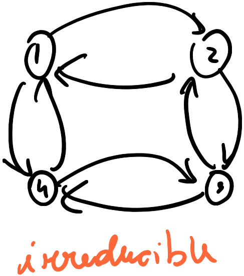


Representing Finite Markov Chains

Markov chain can be seen as weighted directed graph.

- Vertex is a state of Markov chain
- edge (i, j) corresponds to transition probability from i to j





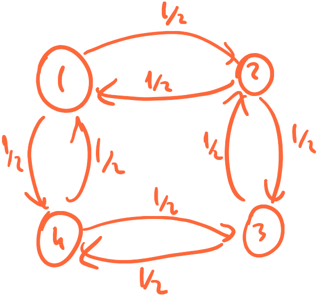
- Markov Chain *irreducible* if underlying directed graph is *strongly connected* (i.e. there is directed path from i to j for any pair $i, j \in V$)

Representing Finite Markov Chains

Markov chain can be seen in weighted adjacency matrix format.

Representing Finite Markov Chains

Markov chain can be seen in weighted adjacency matrix format.



$$P = \begin{array}{c} \\ 1 \\ 2 \\ 3 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} 1 & 2 & 3 & 4 \end{array} \\ \begin{pmatrix} 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \\ 0 & 1/2 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix} \end{array}$$

- $P \in \mathbb{R}^{n \times n}$ transition matrix

$$\boxed{P_{ij} = \Pr[i \to j]}$$

Representing Finite Markov Chains

Markov chain can be seen in weighted adjacency matrix format.

- $P \in \mathbb{R}^{n \times n}$ transition matrix
- entry $P_{i,j}$ corresponds to transition probability from i to j

Representing Finite Markov Chains

Markov chain can be seen in weighted adjacency matrix format.

if Markov chain starts at vertex/state 1

$p_0(1) = 1 \quad p_0(j) = 0 \quad \forall j > 1$

$p_0 = e_1$

$(1, 0, 0, \ldots, 0)$

- $P \in \mathbb{R}^{n \times n}$ transition matrix
- entry $P_{i,j}$ corresponds to transition probability from i to j
- $p_t \in \mathbb{R}^n$ probability vector: $p_t(i) :=$ Pr[being at *state i* at *time t*]

Representing Finite Markov Chains

Markov chain can be seen in weighted adjacency matrix format.

$$p_{t+1}(j) = \underbrace{\sum_{i=1}^{n} p_t(i) \cdot \underbrace{\Pr[i \to j]}_{P_{ij}}}_{(\vec{p}_t P)_j}$$

- $P \in \mathbb{R}^{n \times n}$ transition matrix
- entry $P_{i,j}$ corresponds to transition probability from i to j
- $p_t \in \mathbb{R}^n$ probability vector: $p_t(i) := \Pr[\text{being at } \textit{state } i \text{ at } \textit{time } t]$
- Transition given by

$$p_{t+1} = p_t \cdot P$$

Properties of Markov Chains

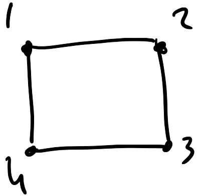
- *Period* of a state i is:

time

$$\gcd\{t \in \mathbb{N} \mid P^t_{i,i} > 0\}$$

That is, gcd of all times t such that the probability of starting at state i and being back at i at time t is positive

start at 1



even times: 1, 3

odd times: 2, 4

$\boxed{\text{Period for 1 is = 2}}$

Properties of Markov Chains

- *Period* of a state i is:

$$\gcd\{t \in \mathbb{N} \mid P_{i,i}^t > 0\}$$

  That is, gcd of all times t such that the probability of starting at state i and being back at i at time t is positive
- State i is *aperiodic* if its period is 1.

Properties of Markov Chains

- *Period* of a state i is:

$$\gcd\{t \in \mathbb{N} \mid P^t_{i,i} > 0\}$$

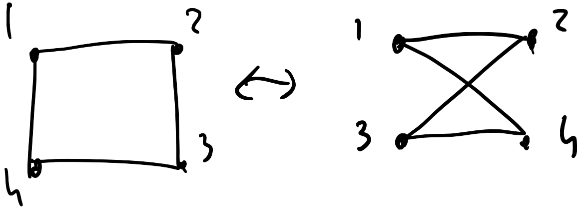
  That is, gcd of all times t such that the probability of starting at state i and being back at i at time t is positive
- State i is *aperiodic* if its period is 1.
- Markov Chain aperiodic if *all states* are aperiodic (otherwise periodic)

Properties of Markov Chains

- *Period* of a state i is:

$$\gcd\{t \in \mathbb{N} \mid P^t_{i,i} > 0\}$$

  That is, gcd of all times t such that the probability of starting at state i and being back at i at time t is positive
- State i is *aperiodic* if its period is 1.
- Markov Chain aperiodic if *all states* are aperiodic (otherwise periodic)
  - Bipartite graphs yield periodic Markov Chains



Properties of Markov Chains

- *Period* of a state i is:

$$\gcd\{t \in \mathbb{N} \mid P^t_{i,i} > 0\}$$

  That is, gcd of all times t such that the probability of starting at state i and being back at i at time t is positive
- State i is *aperiodic* if its period is 1.
- Markov Chain aperiodic if *all states* are aperiodic (otherwise periodic)
  - Bipartite graphs yield periodic Markov Chains

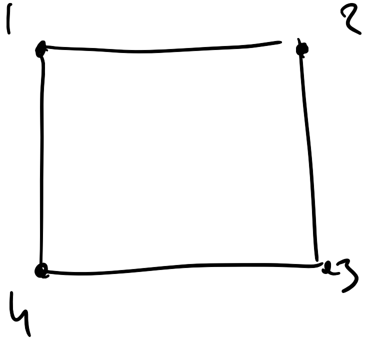
Lemma

For any finite, irreducible and aperiodic Markov Chain, there exists $\underbrace{T < \infty}_{\text{finite}}$ such that

$$P^t_{i,j} > 0 \quad \text{for any } i,j \in V \text{ and } t \geq T.$$

See proof in reference of [Häggström, Chapter 4].

Periodic Markov Chain



Periodic Markov Chain

Reducible Markov Chain

Stationary Distributions

Definition (Stationary Distribution)

A stationary distribution of a Markov Chain is a probability distribution $\pi \in \mathbb{R}^n$ such that

$$\underbrace{\pi}_{p_t} P = \underbrace{\pi}_{p_{t+1}}.$$

Stationary Distributions

Definition (Stationary Distribution)

A stationary distribution of a Markov Chain is a probability distribution $\pi \in \mathbb{R}^n$ such that

$$\pi P = \pi.$$

- Informally, π is an "equilibrium/fixed point" state, as we have $\pi = \pi P^t$ for any $t \geq 0$.

Stationary Distributions

Definition (Stationary Distribution)

A stationary distribution of a Markov Chain is a probability distribution $\pi \in \mathbb{R}^n$ such that

$$\pi P = \pi.$$

- Informally, π is an "equilibrium/fixed point" state, as we have $\pi = \pi P^t$ for any $t \geq 0$.
- **Intuition:** If we run finite, irreducible and aperiodic Markov Chain long enough, we will converge to a stationary distribution.

unique

reducible → could have more than 1 stationary dist.

periodic → have no stationary distribution

Stationary Distributions

Definition (Stationary Distribution)

A stationary distribution of a Markov Chain is a probability distribution $\pi \in \mathbb{R}^n$ such that

$$\pi P = \pi.$$

- Informally, π is an "equilibrium/fixed point" state, as we have $\pi = \pi P^t$ for any $t \geq 0$.
- **Intuition:** If we run finite, irreducible and aperiodic Markov Chain long enough, we will converge to a stationary distribution.
  - what do you mean by converge here?

Stationary Distributions

Definition (Stationary Distribution)

A stationary distribution of a Markov Chain is a probability distribution $\pi \in \mathbb{R}^n$ such that

$$\pi P = \pi.$$

- Informally, π is an "equilibrium/fixed point" state, as we have $\pi = \pi P^t$ for any $t \geq 0$.
- **Intuition:** If we run finite, irreducible and aperiodic Markov Chain long enough, we will converge to a stationary distribution.
  - what do you mean by converge here?
- Given two distributions $p, q \in \mathbb{R}^n$, their *total variational distance* is

$$\Delta_{TV}(p, q) = \frac{1}{2} \sum_{i=1}^{n} |p_i - q_i| = \frac{1}{2} \cdot \|p - q\|_1$$

Stationary Distributions

Definition (Stationary Distribution)

A stationary distribution of a Markov Chain is a probability distribution $\pi \in \mathbb{R}^n$ such that

$$\pi P = \pi.$$

- Informally, π is an "equilibrium/fixed point" state, as we have $\pi = \pi P^t$ for any $t \geq 0$.
- **Intuition:** If we run finite, irreducible and aperiodic Markov Chain long enough, we will converge to a stationary distribution.
  - what do you mean by converge here?
- Given two distributions $p, q \in \mathbb{R}^n$, their *total variational distance* is

$$\Delta_{TV}(p, q) = \frac{1}{2} \sum_{i=1}^{n} |p_i - q_i| = \frac{1}{2} \cdot \|p - q\|_1$$

- p_t *converges* to q iff $\lim_{t \to \infty} \Delta_{TV}(p_t, q) = 0$

Mixing Time of Markov Chains

Definition (Mixing Time)

The ε-mixing time of a Markov Chain is the smallest t such that

$$\Delta_{TV}(p_t, \pi) \leq \varepsilon$$

regardless of the initial starting distribution p_0.

Mixing Time of Markov Chains

Definition (Mixing Time)

The ε-mixing time of a Markov Chain is the smallest t such that

$$\Delta_{TV}(p_t, \pi) \leq \varepsilon$$

regardless of the initial starting distribution p_0.

- For complete graph, eigenvalues $\boxed{\lambda_1 = 1,}$ $\boxed{\lambda_2 = \cdots = \lambda_n = -1/(n-1),}$ corresponding eigenvectors $v_1, \ldots, v_n$ (orthonormal)

$n, 0, 0, 0, \ldots, 0$

$$P = \begin{pmatrix} 0 & & & 1/_{n-1} \\ & 0 & & \\ & & 0 & \\ 1/_{n-1} & & \ddots & \\ & & & 0 \end{pmatrix} = \frac{1}{n-1} J - \frac{1}{n-1} I$$

all ones matrix

$\frac{1}{n-1}$ $n-1, -1, -1, \ldots, -1$

$(1, 1, 1, \ldots, 1) \Leftrightarrow$ eigenvalue n

$(1, -1, 0, \ldots, 0)$ eigenvalue 0

Mixing Time for Complete Graph

$$\left(\tfrac{1}{n}, \tfrac{1}{n}, \ldots, \tfrac{1}{n}\right) P = \underbrace{\left(\tfrac{1}{n}, \ldots, \tfrac{1}{n}\right)}_{v_1}$$

eigenvalue 1 corresponds to stationary distribution

$$P_t = p_0 P^t = p_0 \left(\lambda_1 v_1 v_1^T + \sum_{i=2}^{n} \lambda_i v_i v_i^T \right)^t$$

P symmetric orthonormal basis of eigenvectors

$$\boxed{v_i^T v_j = 0} \quad \forall\, i \neq j$$

stationary distribution

$$P_t = p_0 \left(\underbrace{\lambda_1^t}_{=1} \cdot v_1 v_1^T + \sum_{i=2}^{n} \lambda_i^t v_i v_i^T \right) \rightarrow (p_0 v_1) \cdot v_1^T$$

$$\lambda_i^t = \left(-\frac{1}{n-1}\right)^t \rightarrow 0$$

Mixing Time for Complete Graph

$$\boxed{p_0 = e_i}$$

$$e_i J = n \underset{\sim}{\pi} = (1/n, 1/n, \dots, \tfrac{1}{n})$$

$$p_1 = p_0 P = \frac{1}{n-1} e_i (J - I) = \frac{n}{n-1}\pi - \frac{1}{n-1} e_i$$

$$p_2 = p_1 P = \left(\frac{n}{n-1}\pi - \frac{1}{n-1}e_i\right) P = \frac{n}{n-1}\pi - \frac{1}{n-1}\left(\frac{n}{n-1}\pi - \frac{1}{n-1}e_i\right)$$

$$= \frac{n}{n-1}\pi\left(1 - \frac{1}{n-1}\right) + \frac{1}{(n-1)^2} e_i$$

$$p_t = \frac{n}{n-1}\pi \cdot \sum_{k=0}^{t-1}\left(\frac{-1}{n-1}\right)^k + \underbrace{\left(\frac{-1}{n-1}\right)^t e_i}_{\text{norm} < \epsilon}$$

$$\|p_t\|_1 = 1$$

$$\|\pi\|_1 = 1$$

$$\|e_i\|_1 = 1$$

$$p_t - \alpha_t \pi = \left(\frac{-1}{n-1}\right)^t e_i$$

$$\boxed{t = \log_{n-1}(1/\epsilon)}$$

$$p_0 = \sum_{i=1}^{n} \alpha_i \underbrace{e_i}_{} \qquad \alpha_i \in [0,1] \quad \text{s.t.} \quad \sum_{i=1}^{n} \alpha_i = 1$$

ϵ mixing time

is $\log_{n-1}(1/\epsilon)$

$\Rightarrow$ p_0 ϵ-mixes in $\log_{n-1}(1/\epsilon)$ time

(for any p_0)

Hitting Time

- Given states i, j in a Markov chain, the *hitting time* from state i to state j is defined as

$$T_{i,j} := \min\{t \geq 1 \mid X_t = j, X_0 = i\}$$

We say $T_{i,j} = \infty$ if the Markov chain never visits j starting from i.

and $T_{ij} < \infty$ otherwise

Hitting Time

- Given states i, j in a Markov chain, the *hitting time* from state i to state j is defined as

$$T_{i,j} := \min\{t \geq 1 \mid X_t = j, X_0 = i\}$$

  We say $T_{i,j} = \infty$ if the Markov chain never visits j starting from i.
- The *mean hitting time* $\tau_{i,j} := \mathbb{E}[T_{i,j}]$

Hitting Time

- Given states i, j in a Markov chain, the *hitting time* from state i to state j is defined as

$$T_{i,j} := \min\{t \geq 1 \mid X_t = j, X_0 = i\}$$

  We say $T_{i,j} = \infty$ if the Markov chain never visits j starting from i.
- The *mean hitting time* $\tau_{i,j} := \mathbb{E}[T_{i,j}]$
- *Hitting time lemma*: For any *finite*, *irreducible*, *aperiodic* Markov chain, and for any two states i, j (not necessarily distinct) we have that:

$$\Pr[T_{i,j} < \infty] = 1 \quad \text{and} \quad \mathbb{E}[T_{i,j}] < \infty$$

Proof of Hitting Time Lemma

- We know that we can find $M < \infty$ such that $(P^M)_{i,j} > 0$ for all i, j, since our Markov chain is finite, irreducible and aperiodic.

(lemma from previous part)

Proof of Hitting Time Lemma

- We know that we can find $M < \infty$ such that $(P^M)_{i,j} > 0$ for all i, j, since our Markov chain is finite, irreducible and aperiodic.
- set $\alpha := \min_{i,j}(P^M)_{i,j}$ = the smallest entry of P^M

P^M

Proof of Hitting Time Lemma

- We know that we can find $M < \infty$ such that $(P^M)_{i,j} > 0$ for all i, j, since our Markov chain is finite, irreducible and aperiodic.
- set $\alpha := \min_{i,j}(P^M)_{i,j}$
- Note that

$$\Pr[T_{i,j} > M] \leq \Pr[X_M \neq j] \leq 1 - \alpha$$

state of MC at step M

$$1 - \underbrace{\Pr[X_M = j]}_{= (P^M)_{ij} \geq \alpha}$$

Proof of Hitting Time Lemma

- We know that we can find $M < \infty$ such that $(P^M)_{i,j} > 0$ for all i, j, since our Markov chain is finite, irreducible and aperiodic.
- set $\alpha := \min_{i,j}(P^M)_{i,j}$
- Note that

$$\Pr[T_{i,j} > M] \leq \Pr[X_M \neq s_j] \leq 1 - \alpha$$

- Moreover, we can prove:

$$\begin{aligned}\Pr[T_{i,j} > 2M] &= \Pr[T_{i,j} > M] \cdot \Pr[T_{i,j} > 2M \mid T_{i,j} > M] \\ &\leq (1-\alpha) \cdot \underbrace{\Pr[X_{2M} \neq j \mid T_{i,j} > M]} \\ &\leq (1-\alpha)^2\end{aligned}$$

same reasoning as previous slide

Proof of Hitting Time Lemma

- We know that we can find $M < \infty$ such that $(P^M)_{i,j} > 0$ for all i, j, since our Markov chain is finite, irreducible and aperiodic.
- set $\alpha := \min_{i,j}(P^M)_{i,j}$
- Note that

$$\Pr[T_{i,j} > M] \leq \Pr[X_M \neq s_j] \leq 1 - \alpha$$

- Moreover, we can prove:

$$\begin{aligned}\Pr[T_{i,j} > 2M] &= \Pr[T_{i,j} > M] \cdot \Pr[T_{i,j} > 2M \mid T_{I,j} > M] \\ &\leq (1-\alpha) \cdot \Pr[X_{2M} \neq j \mid T_{i,j} > M] \\ &\leq (1-\alpha)^2\end{aligned}$$

- Iterating, we have $\Pr[T_{i,j} > \ell M] \leq (1-\alpha)^\ell$

$\Rightarrow \Pr[T_{i,j} = \infty] = 0$

first part done

Proof of Hitting Time Lemma

$$\sum_{n\geq 0} \Pr[T_{i,j} > n] = \sum_{\ell\geq 0} \underbrace{\sum_{k=\ell M}^{(\ell+1)M-1}}_{M} \underbrace{\Pr[T_{i,j} > k]}_{\leq \Pr[T_{ij} > M\ell]}$$

$$\leq \sum_{\ell\leq 0} M\cdot \Pr[T_{ij} > \ell M] \leq \sum_{\ell\geq 0} M\cdot(1-\alpha)^{\ell}$$

$$= M\cdot\frac{1}{1-(1-\alpha)} = \frac{M}{\alpha}$$

- Iterating, we have $\Pr[T_{i,j} > \ell M] \leq (1-\alpha)^{\ell}$
- Thus, we have

$$\mathbb{E}[T_{i,j}] = \sum_{n\geq 1} \Pr[T_{i,j} \geq n] = \sum_{n\geq 0} \Pr[T_{i,j} > n] \leq M/\alpha < \infty \qquad \square$$

Fundamental Theorem of Markov Chains

- The *return time* from state i to itself is $T_{i,i}$
- *Expected return time:* defined as $\tau_{i,i} := \mathbb{E}[T_{i,i}]$.

Fundamental Theorem of Markov Chains

- The *return time* from state i to itself is $T_{i,i}$
- *Expected return time:* defined as $\tau_{i,i} := \mathbb{E}[T_{i,i}]$.

Theorem (Fundamental Theorem of Markov Chains)

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

1. *There exists a unique stationary distribution π, where $\pi_i > 0$ for all $i \in [n]$*
2. *The sequence of distributions $\{p_t\}_{t \geq 0}$ will converge to π, no matter what the initial distribution is*
3. $$\pi_i = \lim_{t \to \infty} P^t_{i,i} = \frac{1}{\tau_{i,i}}$$

Fundamental Theorem of Markov Chains

Theorem (Fundamental Theorem of Markov Chains)

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

1. *There is unique stationary distribution* π, *where* $\pi_i > 0$ *for all* $i \in [n]$
2. *For every distribution* $p_0 \in \mathbb{R}^n_{\geq 0}$, $\lim_{t \to \infty} p_0 \cdot P^t = \pi$
3. $$\pi_i = \lim_{t \to \infty} P^t_{i,i} = \frac{1}{\tau_{i,i}}$$

Intuition for proof of this theorem:

Fundamental Theorem of Markov Chains

Theorem (Fundamental Theorem of Markov Chains)

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

1. *There is unique stationary distribution π, where $\pi_i > 0$ for all $i \in [n]$*
2. *For every distribution $p_0 \in \mathbb{R}^n_{\geq 0}$,* $\lim_{t \to \infty} p_0 \cdot P^t = \pi$
3. $$\pi_i = \lim_{t \to \infty} P^t_{i,i} = \frac{1}{\tau_{i,i}}$$

Intuition for proof of this theorem:

- two random walks are "indistinguishable" after they "meet" at the *same vertex v* at a particular *time t*

Fundamental Theorem of Markov Chains

Theorem (Fundamental Theorem of Markov Chains)

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

1. *There is unique stationary distribution* π, *where* $\pi_i > 0$ *for all* $i \in [n]$
2. *For every distribution* $p_0 \in \mathbb{R}^n_{\geq 0}$, $\lim_{t \to \infty} p_0 \cdot P^t = \pi$
3. $$\pi_i = \lim_{t \to \infty} P^t_{i,i} = \frac{1}{\tau_{i,i}}$$

Intuition for proof of this theorem:

- two random walks are "indistinguishable" after they "meet" at the *same vertex v* at a particular *time t*
- By finiteness, irreducibility and aperiodicity, two walks will meet with positive probability (and thus by Markov property) become *same distribution*

Fundamental Theorem of Markov Chains

Theorem (Fundamental Theorem of Markov Chains)

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

1. *There is unique stationary distribution π, where $\pi_i > 0$ for all $i \in [n]$*
2. *For every distribution $p_0 \in \mathbb{R}^n_{\geq 0}$,* $\lim_{t \to \infty} p_0 \cdot P^t = \pi$
3. $$\pi_i = \lim_{t \to \infty} P^t_{i,i} = \frac{1}{\tau_{i,i}}$$

If our underlying graph is undirected:

Fundamental Theorem of Markov Chains

Theorem (Fundamental Theorem of Markov Chains)

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

1. *There is unique stationary distribution* π, *where* $\pi_i > 0$ *for all* $i \in [n]$
2. *For every distribution* $p_0 \in \mathbb{R}^n_{\geq 0}$, $\lim_{t\to\infty} p_0 \cdot P^t = \pi$
3. $$\pi_i = \lim_{t\to\infty} P^t_{i,i} = \frac{1}{\tau_{i,i}}$$

If our underlying graph is undirected:

- If A_G adjacency matrix of $G(V,E)$ and $D = diag(d_1, d_2, \ldots, d_n)$, transition matrix:

$$P = D^{-1} \cdot A_G$$

Fundamental Theorem of Markov Chains

Theorem (Fundamental Theorem of Markov Chains)

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

1. *There is unique stationary distribution π, where $\pi_i > 0$ for all $i \in [n]$*
2. *For every distribution $p_0 \in \mathbb{R}^n_{\geq 0}$,* $\lim_{t\to\infty} p_0 \cdot P^t = \pi$
3. $$\pi_i = \lim_{t\to\infty} P^t_{i,i} = \frac{1}{\tau_{i,i}}$$

If our underlying graph is undirected:

- If A_G adjacency matrix of $G(V,E)$ and $D = diag(d_1, d_2, \ldots, d_n)$, transition matrix:
$$P = D^{-1} \cdot A_G$$
- Note that in this case, easy to guess stationary distribution:
$$\pi_i = \frac{d_i}{2m}, \quad m = |E|$$

Fundamental Theorem of Markov Chains

If our underlying graph is undirected:

Fundamental Theorem of Markov Chains

If our underlying graph is undirected:

- In this case, easy to guess stationary distribution:

$$\pi_i = \frac{d_i}{2m}, \quad m = |E|$$

Fundamental Theorem of Markov Chains

If our underlying graph is undirected:

- In this case, easy to guess stationary distribution:

$$\pi_i = \frac{d_i}{2m}, \quad m = |E|$$

- If A_G adjacency matrix of $G(V, E)$ and $D = diag(d_1, d_2, \dots, d_n)$, transition matrix:

$$P = D^{-1} \cdot A_G$$

Fundamental Theorem of Markov Chains

If our underlying graph is undirected:

- In this case, easy to guess stationary distribution:

$$\pi_i = \frac{d_i}{2m}, \quad m = |E|$$

- If A_G adjacency matrix of $G(V, E)$ and $D = diag(d_1, d_2, \ldots, d_n)$, transition matrix:

$$P = D^{-1} \cdot A_G$$

  - P not symmetric, but *similar* to a symmetric matrix:

$$D^{1/2} P D^{-1/2} = D^{1/2} D^{-1} A_G D^{-1/2} = D^{-1/2} A_G D^{-1/2} = P'$$

Fundamental Theorem of Markov Chains

If our underlying graph is undirected:

- If A_G adjacency matrix of $G(V,E)$ and $D = diag(d_1, d_2, \ldots, d_n)$, transition matrix:

$$P = D^{-1} \cdot A_G$$

  - P not symmetric, but *similar* to a symmetric matrix:

$$D^{1/2} P D^{-1/2} = D^{1/2} D^{-1} A_G D^{-1/2} = D^{-1/2} A_G D^{-1/2} = P'$$

  - P and P' has *same eigenvalues*! And P' has only *real eigenvalues*!

Fundamental Theorem of Markov Chains

If our underlying graph is undirected:

- If A_G adjacency matrix of $G(V, E)$ and $D = diag(d_1, d_2, \dots, d_n)$, transition matrix:

$$P = D^{-1} \cdot A_G$$

  - P not symmetric, but *similar* to a symmetric matrix:

$$D^{1/2} P D^{-1/2} = D^{1/2} D^{-1} A_G D^{-1/2} = D^{-1/2} A_G D^{-1/2} = P'$$

  - P and P' has *same eigenvalues*! And P' has only *real eigenvalues*!
  - Eigenvectors of P are $D^{-1/2} v_i$ where v_i are eigenvectors of P'. And v_i can be taken to form *orthonormal basis*.

Fundamental Theorem of Markov Chains

- Stationary distribution: $\pi_i = \dfrac{d_i}{2m}, \quad m = |E|$
- Transition matrix: $P = D^{-1} \cdot A_G$
  - P *similar* to a symmetric matrix: $P' = D^{-1/2} A_G D^{-1/2}$
  - P and P' has *same eigenvalues*! And P' has only *real eigenvalues*!
  - Eigenvectors of P are $D^{-1/2} v_i$ where v_i are eigenvectors of P'. And v_i can be taken to form *orthonormal basis*.

Fundamental Theorem of Markov Chains

- Stationary distribution: $\pi_i = \dfrac{d_i}{2m}, \quad m = |E|$
- Transition matrix: $P = D^{-1} \cdot A_G$
  - P *similar* to a symmetric matrix: $P' = D^{-1/2} A_G D^{-1/2}$
  - P and P' has *same eigenvalues*! And P' has only *real eigenvalues*!
  - Eigenvectors of P are $D^{-1/2} v_i$ where v_i are eigenvectors of P'. And v_i can be taken to form *orthonormal basis*.
  - Graph strongly connected $\Rightarrow$ *Perron-Frobenius* for irreducible non-negative matrices

Fundamental Theorem of Markov Chains

- Stationary distribution: $\pi_i = \dfrac{d_i}{2m}, \quad m = |E|$
- Transition matrix: $P = D^{-1} \cdot A_G$
  - P *similar* to a symmetric matrix: $P' = D^{-1/2} A_G D^{-1/2}$
  - P and P' has *same eigenvalues*! And P' has only *real eigenvalues*!
  - Eigenvectors of P are $D^{-1/2} v_i$ where v_i are eigenvectors of P'. And v_i can be taken to form *orthonormal basis*.
  - Graph strongly connected $\Rightarrow$ *Perron-Frobenius* for irreducible non-negative matrices
    - *unique* eigenvector whose eigenvalue has *max absolute value*
    - eigenvector has *all positive coordinates*
    - eigenvalue is *positive*

Fundamental Theorem of Markov Chains

- Stationary distribution: $\pi_i = \dfrac{d_i}{2m}, \quad m = |E|$
- Transition matrix: $P = D^{-1} \cdot A_G$
  - P *similar* to a symmetric matrix: $P' = D^{-1/2} A_G D^{-1/2}$
  - P and P' has *same eigenvalues*! And P' has only *real eigenvalues*!
  - Eigenvectors of P are $D^{-1/2} v_i$ where v_i are eigenvectors of P'. And v_i can be taken to form *orthonormal basis*.
  - Graph strongly connected $\Rightarrow$ *Perron-Frobenius* for irreducible non-negative matrices
    - *unique* eigenvector whose eigenvalue has *max absolute value*
    - eigenvector has *all positive coordinates*
    - eigenvalue is *positive*
  - This eigenvector is π!

Fundamental Theorem of Markov Chains

- Stationary distribution: $\pi_i = \dfrac{d_i}{2m}, \quad m = |E|$
- Transition matrix: $P = D^{-1} \cdot A_G$
  - P *similar* to a symmetric matrix: $P' = D^{-1/2} A_G D^{-1/2}$
  - P and P' has *same eigenvalues*! And P' has only *real eigenvalues*!
  - Eigenvectors of P are $D^{-1/2} v_i$ where v_i are eigenvectors of P'. And v_i can be taken to form *orthonormal basis*.
  - Graph strongly connected $\Rightarrow$ *Perron-Frobenius* for irreducible non-negative matrices
    - *unique* eigenvector whose eigenvalue has *max absolute value*
    - eigenvector has *all positive coordinates*
    - eigenvalue is *positive*
  - This eigenvector is π!
  - All random walks converge to π, as we wanted to show.

Revisiting the complete graph

Conclusion

Markov Chains and Random Walks are ubiquitous in randomized algorithms.

- Page Rank algorithm (next lecture)
- Approximation algorithms for counting problems [Karp, Luby & Madras]
    - Permanent of non-negative matrices [Jerrum, Vigoda & Sinclair]
- Sampling Problems
    - Gibbs sampling in statistical physics
    - many more places
- Probability amplification without too much randomness (efficient)
    - Random walks on expander graphs
- many more

Acknowledgement

- Lecture based largely on:
  - Lap Chi's notes
  - [Motwani & Raghavan 2007, Chapter 6]
  - [Häggström]
- See Lap Chi's notes at `https://cs.uwaterloo.ca/~lapchi/cs466/notes/L11.pdf`
- Also see Lap Chi's notes `https://cs.uwaterloo.ca/~lapchi/cs466/notes/L14.pdf` for a proof of fundamental theorem of Markov chains for undirected graphs.

References I

Motwani, Rajeev and Raghavan, Prabhakar (2007)
Randomized Algorithms

Karp, R.M. and Luby, M. and Madras, N. (1989)
Monte-Carlo approximation algorithms for enumeration problems.
Journal of algorithms, 10(3), pp.429-448.

Jerrum, M. and Sinclair, A. and Vigoda, E. (2004)
A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries.
Journal of the ACM (JACM), 51(4), pp.671-697.

Häggström, Olle (2002)
Finite Markov Chains and Algorithmic Applications
Cambridge University Press