COMPSCI 690RA: Randomized Algorithms and Probabilistic Data Analysis

Prof. Cameron Musco

University of Massachusetts Amherst. Spring 2022.

Lecture 10

Logistics

- Problem Set 3 is due this Friday 4/15 at 8pm.
- We have no class next Wednesday – it's a Monday at UMass.
- I will post a quiz due Tuesday 4/26 at 8pm.
- Remember that office hours are now Thursday at 4pm.

Summary

Last Week: Markov Chains.

- Finish spectral graph sparsification and physical interpretation
- Start on Markov chains and their analysis
- Markov chain based algorithms for satisfiability: $\approx n^2$ time for 2-SAT, and $\approx (4/3)^n$ for 3-SAT.

Today: Markov Chains Continued

- The gambler's ruin problem.
- Aperiodicity and stationary distribution of a Markov chain.
- Mixing time and its analysis via coupling.
- Markov Chain Monte Carlo (MCMC) methods.

Question 3

Not complete

Points out of 1.00

Flag question

Edit question

Consider a matrix $A \in \mathbb{R}^{5 \times 3}$ such that $x = [0, 2, 2, 1, 1]$ is in the column span of A.

What can we say about the leverage score of the second row of A, i.e., τ_2?

- a. $\tau_2 \leq 2$
- b. $\tau_2 \leq 0.4$
- c. $\tau_2 \geq 2$
- d. Nothing, without knowing A.
- e. $\tau_2 \geq 0.4$

Check

Consider a matrix $A \in \mathbb{R}^{n \times d}$ with full column rank. Let $U \in \mathbb{R}^{n \times d}$ be its left singular vector matrix. Let $Q \in \mathbb{R}^{n \times d}$ be an orthonormal basis for its column span, computed e.g., using Gram-Schmidt.

True of False: for all $i \in [n]$, $\|Q_{i,:}\|_2 = \|U_{i,:}\|_2$.

Select one:

- ◯ True
- ◯ False

Check

Question **4**

Not complete

Points out of 1.00

Flag question

Edit question

Let $E_0, E_1, \ldots$ be independent, identically distributed random variables. Which of the following are Markov chains? Select all that apply.

- [ ] a. $X_0, X_1, \ldots$ where $X_0 = E_0$ and $X_{i+1} = X_i + E_i$
- [ ] b. $E_0, E_1, \ldots$ themselves.
- [ ] c. $X_0, X_1, \ldots$ where $X_0 = E_0$ and $X_{i+1} = X_i \cdot E_i$
- [ ] d. $X_0, X_1, \ldots$ where $X_0 = E_0$, and $X_{i+1} = X_i + E_{i-1} + E_i$
- [ ] e. $X_0, X_1, \ldots$ where $X_0 = E_0$, $X_1 = E_1$, and $X_{i+1} = X_i + X_{i-1} + E_i$

Check

Question 5

Not complete

Points out of 1.00

Flag question

Edit question

Markov Chain Review

- A discrete time stochastic process is a Markov chain if is it memoryless:

$$\Pr(X_t = a_t | X_{t-1} = a_{t-1}, \ldots, X_0 = a_0) = \Pr(X_t = a_t | X_{t-1} = a_{t-1})$$

- If each X_t can take m possible values, the Markov chain is specified by the transition matrix $P \in [0,1]^{m \times m}$ with

$$P_{i,j} = \Pr(X_{t+1} = j | X_t = i).$$

- Let $q_t \in [0,1]^{1 \times m}$ be the distribution of X_i. Then $q_{t+1} = q_t P$.

q_1

.5	.5	0	0

P

.5	.5	0	0
.25	.25	.5	0
0	1	0	0
0	.5	.5	0

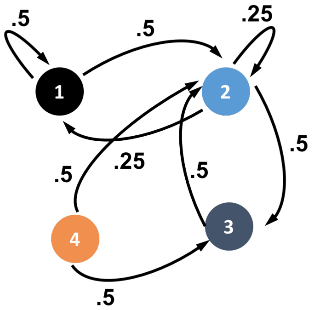
=

q_2

.375	.375	.25	0

Markov Chain Review

Often viewed as an underlying state transition graph. Nodes correspond to possible values that each X_t can take.



P

.5	.5	0	0
.25	.25	.5	0
0	1	0	0
0	.5	.5	0

The Markov chain is irreducible if the underlying graph consists of single strongly connected component.

Gambler’s Ruin

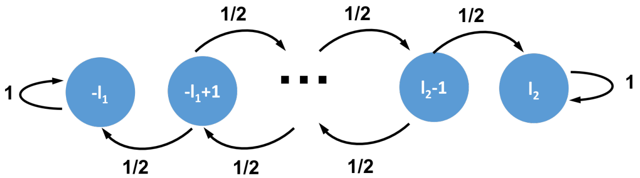
Gambler's Ruin

- You and 'a friend' repeatedly toss a fair coin. If it hits heads, you give your friend \$1. If it hits tails, they give you \$1.
- You start with $\$\ell_1$ and your friend starts with $\$\ell_2$. When either of you runs out of money the game terminates.
- What is the probability that you win $\$\ell_2$?

Let $X_0, X_1, \ldots$ be the Markov chain where X_t is your profit at step t. $X_0 = 0$ and:

$$P_{-\ell_1,-\ell_1} = P_{\ell_2,\ell_2} = 1$$

$$P_{i,i+1} = P_{i,i-1} = 1/2 \text{ for } -\ell_1 < i < \ell_2$$



- ℓ_1 and ℓ_2 are absorbing states.
- All i with $-\ell_1 < i < \ell_2$ are transient states. I.e., $\Pr[X_{t'} = i \text{ for some } t' > t \mid X_t = i] < 1$.

Observe that this Markov chain is also a Martingale since $\mathbb{E}[X_{t+1}|X_t] = X_t$.

Gambler's Ruin Analysis

Let $\mathsf{X}_0, \mathsf{X}_1, \ldots$ be the Markov chain where X_t is your profit at step t.
$\mathsf{X}_0 = 0$ and:

$$P_{-\ell_1,-\ell_1} = P_{\ell_2,\ell_2} = 1$$

$$P_{i,i+1} = P_{i,i-1} = 1/2 \text{ for } -\ell_1 < i < \ell_2$$

We want to compute $q = \lim_{t\to\infty} \Pr[\mathsf{X}_t = \ell_2]$.

By linearity of expectation, for any i, $\mathbb{E}[\mathsf{X}_i] = 0.$ Further, for $q = \lim_{t\to\infty} \Pr[\mathsf{X}_t = \ell_2]$, since $-\ell_1, \ell_2$ are the only non-transient states,

$$\lim_{t\to\infty} \mathbb{E}[\mathsf{X}_t] = \ell_2 q + -\ell_1(1-q) = 0.$$

Solving for q, we have $q = \frac{\ell_1}{\ell_1+\ell_2}$.

Gambler's Ruin Thought Exercise

What if you always walk away as soon as you win just $1. Then what is your probability of winning, and what are your expected winnings?

Stationary Distributions

Stationary Distribution

A stationary distribution of a Markov chain with transition matrix $P \in [0,1]^{m \times m}$ is a distribution $\pi \in [0,1]^m$ such that $\pi = \pi P$.

I.e. if $X_t \sim \pi$, then $X_{t+1} \sim \pi P = \pi$.

$$\overset{\boldsymbol{\pi}}{\begin{bmatrix} .4 & .4 & .2 & 0 \end{bmatrix}} \overset{\mathbf{P}}{\begin{bmatrix} .5 & .5 & 0 & 0 \\ .5 & 0 & .5 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & .5 & .5 & 0 \end{bmatrix}} = \overset{\boldsymbol{\pi}}{\begin{bmatrix} .4 & .4 & .2 & 0 \end{bmatrix}}$$

Think-pair-share: Do all Markov chains have a stationary distribution?

Claim (Existence of Stationary Distribution)

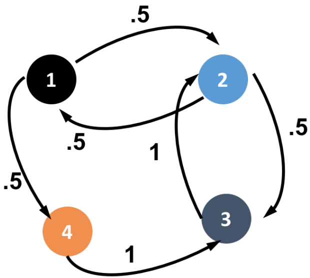
Any Markov chain with a finite state space, and transition matrix $P \in [0,1]^{m \times m}$ has a stationary distribution $\pi \in [0,1]^m$ with $\pi = \pi P$.

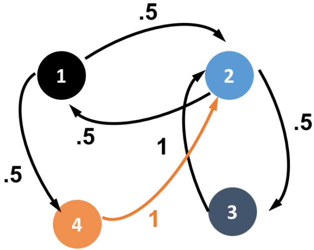
Follows from the Brouwer fixed point theorem: for any continuous function $f : \mathcal{S} \to \mathcal{S}$, where $\mathcal{S}$ is a compact convex set, there is some x such that $f(x) = x$.

Periodicity

The periodicity of a state i is defined as:

$$T = \gcd\{t > 0 : \Pr(X_t = i \mid X_0 = i) > 0\}.$$





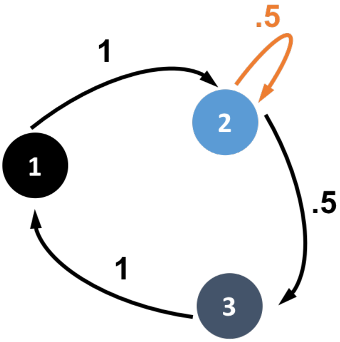
The state is aperiodic if it has periodicity $T = 1$.

A Markov chain is aperiodic if all states are aperiodic.

Periodicity

Claim

If a Markov chain is irreducible, and has at least one self-loop, then it is aperiodic.



Fundamental Theorem

Theorem (The Fundamental Theorem of Markov Chains)

Let $X_0, X_1, \ldots$ *be a Markov chain with a finite state space and transition matrix* $P \in [0,1]^{m \times m}$*. If the chain is both irreducible and aperiodic,*

1. *There exists a unique stationary distribution* $\pi \in [0,1]^m$ *with* $\pi = \pi P$*.*
2. *For any states* i, j*,* $\lim_{t \to \infty} \Pr[X_t = i | X_0 = j] = \pi(i)$*. I.e., for any initial distribution* q_0*,* $\lim_{t \to \infty} q_t = \lim_{t \to \infty} q_0 P^t = \pi$*.*
3. $\pi(i) = \frac{1}{\mathbb{E}[\min(t : X_t = i) | X_0 = i]}$*. I.e.,* $\pi(i)$ *is the inverse of the average expected return time from state* i *back to* i*.*

In the limit, the probability of being at any state i is independent of the starting state.

Stationary Distribution Example 1

Shuffling Markov Chain: Given a pack of c cards. At each step draw a random card, place it on top, and repeat.

- What is the state space of this chain?
- What is the transition probability $P_{i,j}$? How does it compare to $P_{j,i}$?
- This Markov chain is symmetric and thus its stationary distribution is uniform, $\pi(i) = \frac{1}{c!}$.

Letting $m = c!$ denote the size of the state space,

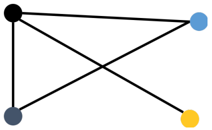
$$\pi P_{:,i} = \sum_j \pi(j) P_{j,i} = \sum_j \pi(j) P_{i,j} = \frac{1}{m} \sum_j P_{i,j} = \frac{1}{m} = \pi(i).$$

Once we have exhibited a stationary distribution, we know that it is unique and that the chain converges to it in the limit!

Stationary Distribution Example 2

Random Walk on an Undirected Graph: Consider a random walk on an undirected graph. If it is at node i at step t, then it moves to any of i's neighbors at step $t+1$ with probability $\frac{1}{d_i}$.

- What is the state space of this chain?
- What is the transition probability $P_{i,j}$?
- Is this chain aperiodic?
- If the graph is not bipartite, then there is at least one odd cycle, making the chain aperiodic.



Stationary Distribution Example 2

Random Walk on an Undirected Graph: Consider a random walk on an undirected graph. If it is at node i at step t, then it moves to any of i's neighbors at step $t+1$ with probability $\frac{1}{d_i}$.

Claim: When the graph is not bipartite, the unique stationary distribution of this Markov chain is given by $\pi(i) = \frac{d_i}{|E|}$.

$$\pi P_{:,i} = \sum_j \pi(j) P_{j,i} = \sum_j \frac{d_j}{|E|} \cdot \frac{1}{d_j} = \sum_j \frac{1}{|E|} = \frac{d_i}{|E|} = \pi(i).$$

I.e., the probability of being at a given node i is dependent only on the node's degree, not on the structure of the graph in any other way.

Mixing Times

Total Variation Distance

Definition (Total Variation (TV) Distance)

For two distributions $p, q \in [0,1]^m$ over state space $[m]$, the total variation distance is given by:

$$\|p - q\|_{TV} = \frac{1}{2}\sum_{i \in [m]} |p(i) - q(i)| = \max_{A \subseteq [m]} |p(A) - q(A)|.$$

Kontorovich-Rubinstein duality: Let P, Q be possibility correlated random variables with marginal distributions p, q. Then

$$\|p - q\|_{TV} \leq \Pr[\mathsf{P} \neq \mathsf{Q}].$$

Mixing Time

Definition (Mixing Time)

Consider a Markov chain $X_0, X_1, \ldots$ with unique stationary distribution π. Let $q_{i,t}$ be the distribution over states at time t assuming $X_0 = i$. The mixing time is defined as:

$$\tau(\epsilon) = \min\left\{t : \max_{i \in [m]} \|q_{i,t} - \pi\|_{TV} \leq \epsilon\right\}.$$

I.e., what is the maximum time it takes the Markov chain to converge to within ϵ in TV distance of the stationary distribution?

Claim: If $X_0, X_1, \ldots$ is finite, irreducible, and aperiodic, then $\tau(\epsilon) \leq \tau(1/2) \cdot c \log(1/\epsilon)$ for large enough constant c.

Coupling Motivation

Claim: $\max_{i\in[m]} \|q_{i,t} - \pi\|_{TV} \leq \max_{i,j\in[m]} \|q_{i,t} - q_{j,t}\|_{TV}$.

$$\begin{aligned}
\|q_{i,t} - \pi\|_{TV} &= \|q_{i,t} - \pi P^t\|_{TV} \\
&= \|q_{i,t} - \sum_j \pi(j) e_j P^t\|_{TV} \\
&= \|q_{i,t} - \sum_j \pi(j) q_{j,t}\|_{TV} \\
&\leq \sum_j \|\pi(j) q_{i,t} - \pi(j) q_{j,t}\|_{TV} \\
&\leq \max_{j\in[m]} \|q_{i,t} - q_{j,t}\|_{TV}.
\end{aligned}$$

Coupling: A common technique for bounding the mixing time by showing that $\max_{i,j\in[m]} \|q_{i,t} - q_{j,t}\|_{TV}$ is small.