Intro to Markov Chains

CS 70, Summer 2019

Lecture 26, 8/7/19

Applications of Markov Chains

- Models systems of **states** and **transitions**
- PageRank – Google's search algorithm. States are webpages, transitions are links.
- Tons of applications outside of CS: statistical physics, speech recognition, bioinformatics, sabermetrics...
  ↳ baseball statistics!

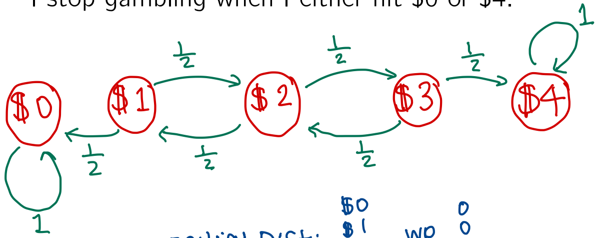
Markov Chain Definition

Three key components (and one assumption):

- Set $\mathcal{S}$ of states
  Think of these as vertices of a graph
- Transition probabilities. $\mathbb{P}[i \to j]$
  Think of these as directed edges in a graph
  Transitions **out of a node** should sum to 1
- Initial **distribution** $\mu^{(0)}$.
  Gives the probability that we start at a state.
- Memorylessness (aka **Markov property**)

Example: Gambling

I start with \$2. If I guess a coin flip correctly, I get \$1, and if I am incorrect, I lose \$1.
I stop gambling when I either hit \$0 or \$4.



Initial Dist: \$0 \$1 \$2 \$3 \$4 wp 0 0 1 0 0

Traversing the Chain

$t=0$ $X_0 = i$ state i $\xrightarrow{1/3} j$, $\xrightarrow{2/3} k$ $t=1$ $\mathbb{P}[X_1 = j] = \frac{1}{3}$ $\mathbb{P}[X_1 = k] = \frac{2}{3}$

X_0 is the initial state.
Choose transitions according to its probability.

X_i is the state you're on at time i. X_i is a **RV**.

Markov Property: ("Memoryless")
Only the **current state** matters for the next.

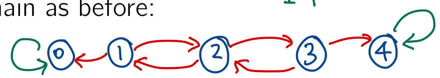
"Knowing the entire history of the chain is equivalent to just knowing the current state."

$$\mathbb{P}[X_{n+1} = S_{n+1} \mid X_0 = S_0, X_1 = S_1, \ldots, X_n = S_n]$$
$$= \mathbb{P}[X_{n+1} = S_{n+1} \mid X_n = S_n]$$

Gambling II

— $\frac{1}{2}$ prob
— 1 prob

Same chain as before:



What is $\mathbb{P}[X_1 = 3 | X_0 = 2]$? $\frac{1}{2}$

What is $\mathbb{P}[X_{100} = 3 | X_{99} = 2, X_0 = 2]$? $\to \mathbb{P}[X_{100} = 3 | X_{99} = 2]$ $= \mathbb{P}[X_1 = 3 | X_0 = 2]$ $= \frac{1}{2}$

$X_0 = 2$ What is $\mathbb{P}[X_1 = 3, X_2 = 2, X_3 = 3, X_4 = 4]$?

same initial setup

$\frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} \times \frac{1}{2} = \frac{1}{16}$

$\mathbb{P}[X_1 = 3 | X_0 = 2]$

$\mathbb{P}[X_2 = 2 | X_1 = 3]$

Gambling II

What is $\mathbb{P}[X_4 = 4]$? Same setup $X_0 = 2$

	X_0	X_1	X_2	X_3	$X_4 = 4$
Path 1:	2	3	4	4	4
Path 2:	2	3	2	3	4
Path 3:	2	1	2	3	4

$\mathbb{P}[X_4 = 4] = \mathbb{P}[\text{Path } 1] + \mathbb{P}[\text{Path } 2] + \mathbb{P}[\text{Path } 3]$

Good idea, but it doesn't scale.

The Transition Matrix

Calculations are easier to do when we stick the transition probabilities in a matrix.

Transition matrix P. $\mathbb{P}[i \to j]$

The (i, j) entry is $\mathbb{P}[X_1 = j | X_0 = i]$, or the **transition from** i **to** j.

row column

col index = states

$$P = \begin{array}{c|c} & \begin{matrix} 1 & 2 & 3 & 4 \cdots \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ \vdots \end{matrix} & \begin{bmatrix} \mathbb{P}[1\to 1] & \mathbb{P}[1\to 2] & \cdots \\ \mathbb{P}[2\to 1] & & \\ \vdots & \ddots & \\ & & \end{bmatrix} \end{array}$$

row index = states

Sanity check:
row sum: 1
col sum: non-neg, otherwise no restriction
entries: non-neg ≤ 1.

The Distribution Vectors

So far: saw initial distribution $\mu^{(0)}$.
Can represent it as a **row vector:**

$$\begin{array}{r} \text{index} = \text{states} \to \end{array} \begin{array}{c} \begin{bmatrix} \mathbb{P}[X_0=1] & \mathbb{P}[X_0=2] & \ldots & \end{bmatrix} \\ \begin{matrix} 1 & \quad 2 & \quad 3 & \ldots \end{matrix} \end{array}$$

entries sum to 1.

We can also define a **distribution at time** n:

$$\begin{array}{r} \text{index} = \text{states} \to \end{array} \begin{array}{c} \begin{bmatrix} \mathbb{P}[X_n=1] & \mathbb{P}[X_n=2] & \ldots & \end{bmatrix} \\ \begin{matrix} 1 & \quad 2 & \quad 3 & \ldots \end{matrix} \end{array}$$

call this $\mu^{(n)}$

Distribution at Time 1

$[\mathbb{P}[X_0=1] \;\; \mathbb{P}[X_0=2] \;\; \ldots \;]$

$$\begin{bmatrix} \mathbb{P}[1\to i] \\ \mathbb{P}[2\to i] \\ \vdots \end{bmatrix}$$

We'll prove that $\underset{1\times n}{\mu^{(0)}} \underset{n\times n}{P} = \underset{1\times n}{\mu^{(1)}}$.

$\mathbb{P}[X_1 = i] = i^{th}$ entry of $\mu^{(1)}$

$= \mu^{(0)} \times i^{th}$ column of P

$= \mathbb{P}[X_0=1] \times \mathbb{P}[X_1 = i | X_0 = 1] + \mathbb{P}[X_0 = 2] \times \mathbb{P}[X_1 = i | X_0 = 2] + \ldots$

casework on X_0

$= \mathbb{P}[X_1 = i]$ ← Total Prob. Rule.

If we know $\mu^{(1)}$, how do we get $\mu^{(2)}$?

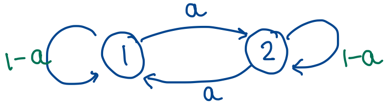
$\mu^{(2)} = \mu^{(1)} P = \mu^{(0)} P^2$

Distribution at Time n

$\mu^{(0)}$ row vector

In general: $\mu^{(n)} = P^n \mu^{(0)}$. (Proof optional.)

Example: Two State Markov Chain



$a \in (0,1)$

$P = \begin{bmatrix} 1-a & a \\ a & 1-a \end{bmatrix}$

(Notes:) $P^n = \begin{bmatrix} \overbrace{\tfrac{1}{2} + \tfrac{1}{2}(1-2a)^n}^{c} & \overbrace{\tfrac{1}{2} - \tfrac{1}{2}(1-2a)^n}^{c'} \\ c' & c \end{bmatrix}$

prove using induction

Aside: $n \to \infty$

For the two state Markov chain, as $n \to \infty$,

$$P^n \to \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

No matter what $\mu^{(0)}$ is:

$= [p \;\; 1-p]$

$$[p \;\; 1-p] \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} = [\tfrac{1}{2} \;\; \tfrac{1}{2}]$$

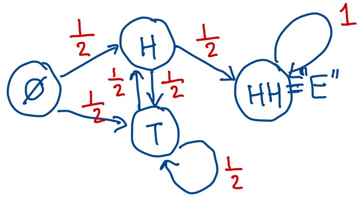
Tomorrow: we'll study this in greater detail!

Break

What's the weirdest thing you've ever eaten?

First Step Analysis: Two Heads

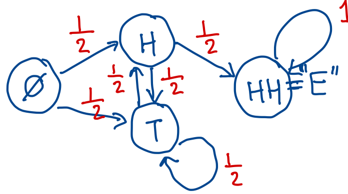
I repeatedly flip a coin, and stop when I get two heads in a row. What is the expected number of flips I need before stopping?



First Step Analysis: Two Heads

For state S, let $\tau(S)$ be the expected time to two heads, starting from state S. 4 variables: $\tau(\emptyset)$, $\tau(H)$, $\tau(T)$, $\tau(E)$

Analyze a **single transition** out of each state to get the **first step equations**:



$\emptyset$: $\tau(\emptyset) = 1 + \frac{1}{2}\tau(H) + \frac{1}{2}\tau(T)$

T: $\tau(T) = 1 + \frac{1}{2}\tau(H) + \frac{1}{2}\tau(T)$

H: $\tau(H) = 1 + \frac{1}{2}\tau(T) + \frac{1}{2}\tau(E)$

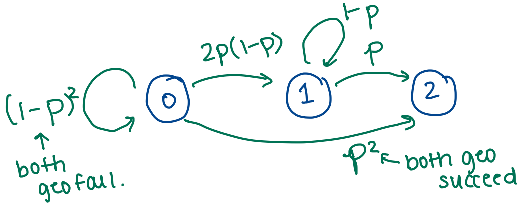
HH: $\tau(E) = 0$

Goal: $\tau(\emptyset)$. Notes.

Max of Two Geometrics

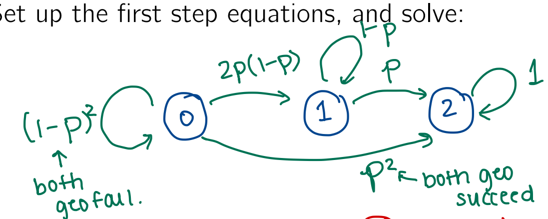
Let $X, Y \sim \text{Geometric}(p)$. X, Y are **independent**. Say X, Y model time until a success. $\max(X, Y)$ is the first time that both X, Y have succeeded at least once. What is $\mathbb{E}[\max(X, Y)]$?

states = # successes



Max of Two Geometrics

Set up the first step equations, and solve:



$\tau(i)$ = expected time until 2 from i

0: $\tau(0) = 1 + (1-p)^2\tau(0) + 2p(1-p)\tau(1) + p^2\tau(2)$

1: $\tau(1) = 1 + (1-p)\tau(1) + p\tau(2)$

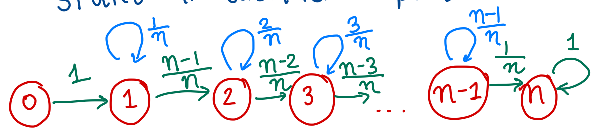
2: $\tau(2) = 0$ Exercise: Finish up.

Coupon Collector: A Markov Chain?

Can we reformulate Coupon Collector (with n distinct coupons) as a Markov chain?

How do we recover the expected number of coupons needed to get all n distinct ones?

states = # distinct coupons



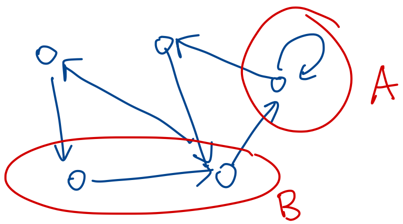
Let $\tau(i)$ = expected time to "n" from "i"

Goal: $\tau(0)$.

Probability of $\mathcal{A}$ Before $\mathcal{B}$

Let $\mathcal{A}$ and $\mathcal{B}$ be two **disjoint** subsets of the states $\mathcal{S}$ of a Markov chain.

Let $\alpha(i)$ be the probability that we enter $\mathcal{A}$ before entering $\mathcal{B}$, if we start at state i.



Probability of $\mathcal{A}$ Before $\mathcal{B}$

Can also run first step analysis!

If $i \in A$: $\alpha(i) = 1$ Already in A!

$i \in B$: $\alpha(i) = 0$ Impossible to get to A before B.

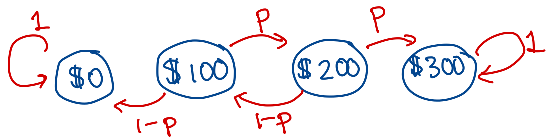
else: $\alpha(i) = \sum_{\substack{\text{neighbors } j \\ \text{of } i}} \mathbb{P}[i \to j]\,\alpha(j)$

Casework based on taking 1 step from i.

Gambling III

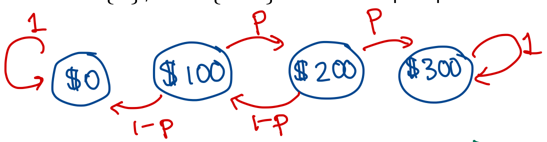
I start with \$100. In each round, I win \$100 with probability p and lose \$100 with probability $(1 - p)$. I end when I either have \$0 or \$300.

What is the probability I end the game with \$300?



Gambling III

Let $\mathcal{A} = \{0\}, \mathcal{B} = \{300\}$. First step equations:



$\alpha(i) = \mathbb{P}[B \text{ before } A \mid \text{at state } i]$

\$0: $\alpha(0) = 0$

\$100: $\alpha(100) = p\alpha(200) + (1-p)\alpha(0)$

\$200: $\alpha(200) = p\alpha(300) + (1-p)\alpha(100)$

\$300: $\alpha(300) = 1$

$\Rightarrow \alpha(100) = p\alpha(200)$

$\alpha(200) = (1-p)\alpha(100) + p$

$\Rightarrow \alpha(100) = \dfrac{p^2}{1-p(1-p)}$

Summary

- Markov chains let you model real world problems with **states** and **transition probabilities**
- The **Markov property** tells you that where you go next only depends on the **current state**, not on any previous history.
- The first step analysis is a simple way of analyzing expected hitting times and probabilities of hitting certain states before others.