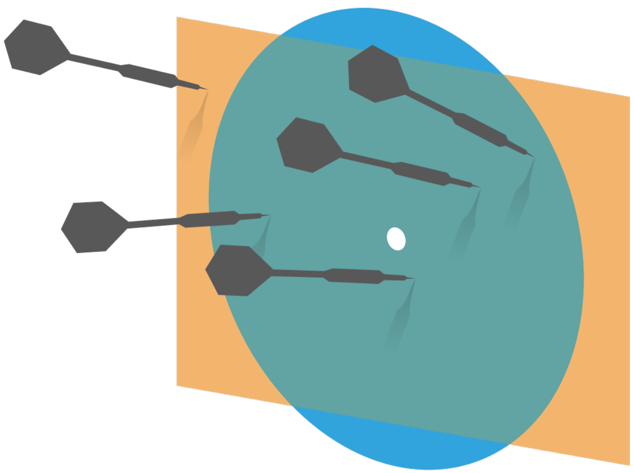
Probability and Statistics for Computer Science



Credit: wikipedia

Conditional probability comes back in matrix!

Hongye Liu, Teaching Assistant Prof, CS361, UIUC, 12.03.2021

Which of the following matrices

is your favorite?

A) Covariance Matrix

B) Confusion Matrix

C) Data matrix X & $X^T X$

D) None

E) All

F) Laplacian matrix

Last Time

- Application of Clustering

  Cluster Center Histogram

- Spectral Clustering

Objectives

Markov chain

1 2 3 4 3 2

An example of dependent events in a sequence

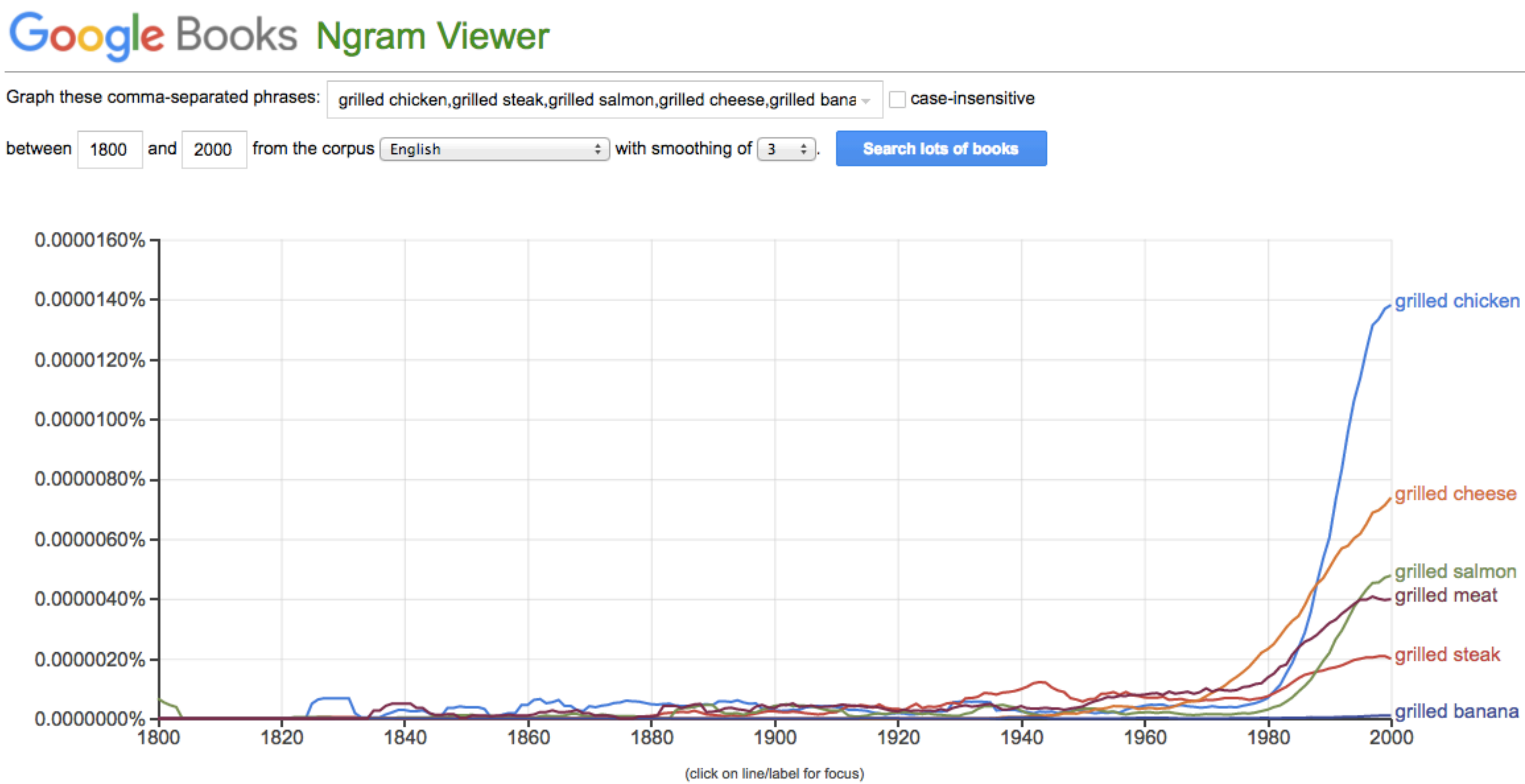
chicken

I had a glass of wine with my grilled ___________ steak

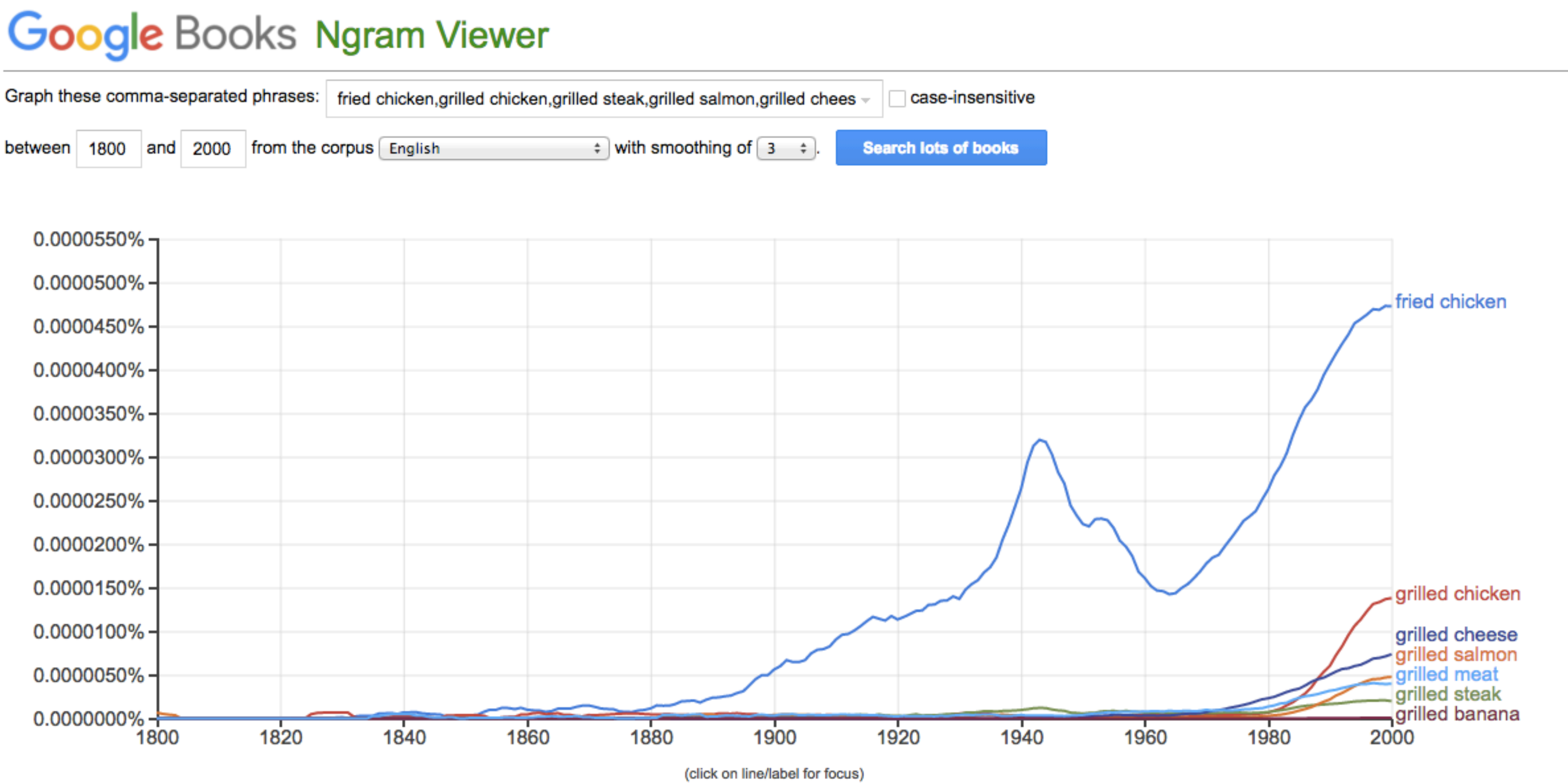
- - - -

paper

An example of dependent events in a sequence



An example of dependent events in a sequence



Markov chain

- Markov chain is a process in which outcome of any trial in a sequence is **conditioned by the outcome of the trial immediately preceding, but not by earlier ones.**
- Such dependence is called **chain dependence**

$P(X_{n+1} | X_n)$

$X_{n-1} \dots X_0$

Andrey Markov (1856-1922)

$= f(n)$

Markov chain in terms of probability

- Let $X_0, X_1, \ldots$ be a sequence of discrete finite-valued random variables

- The sequence is a Markov chain if the probability distribution X_t only depends on the distribution of the immediately preceding random variable X_{t-1}

$$\boxed{P(X_t | X_0 \ldots, X_{t-1}) = P(X_t | X_{t-1})}$$

Markov property

- If the conditional probabilities (transition probabilities) do **NOT change with time**, it's called **constant Markov chain.**

$$P(X_t | X_{t-1}) = P(X_{t-1} | X_{t-2}) = \ldots = \underbrace{P(X_1 | X_0)}_{= f(t) = C}$$

Coin example

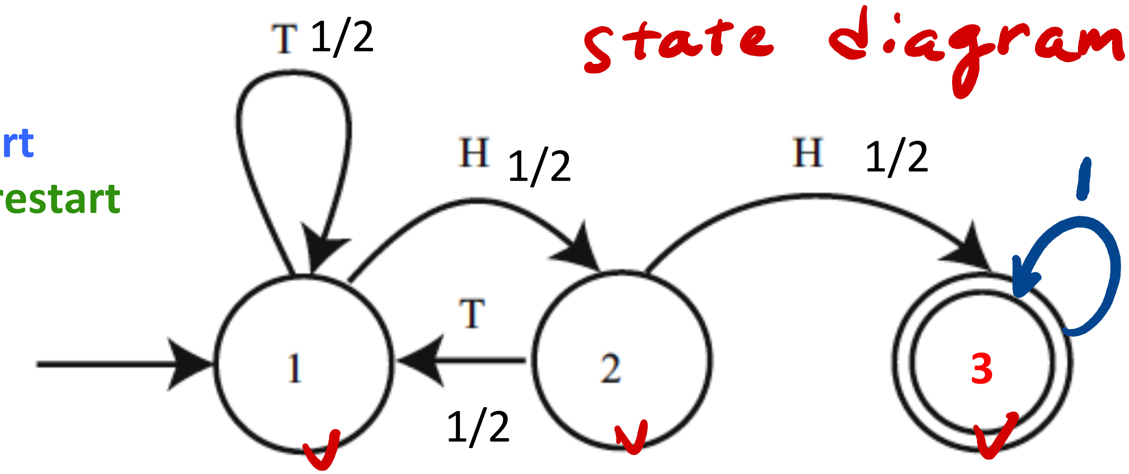
- Toss a fair coin until you see two heads in a row and then stop, what is the probability of stopping after exactly **n** flips?

$$\underbrace{* * * HH}_{n}$$

$$P(n = n_0) = ?$$

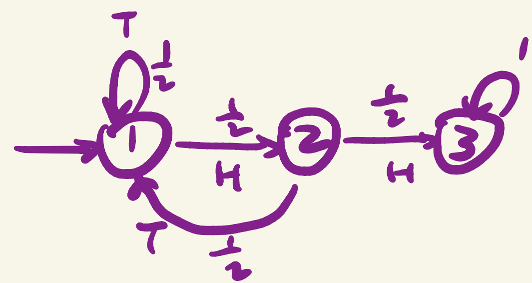
1 -> Start or just had tail/restart
2 -> had one head after start/restart
3 -> 2heads in a row/Stop

state diagram



$N =$	#1	#2	#3	#4	#5	#6	
Trials	T	T	H	T	H	H	θ
$X_N =$	X_1	X_2	X_3	X_4	X_5	X_6	
State	1	1	2	1	2	3	

Markov Property:



$$p_{ij} = P(X_{n+1}=j \mid X_n = i)$$

$$= P(X_{n+1}=j \mid X_n=i, \boxed{X_{n-1}=? \; - \; \ldots \; X_0=?})$$

This part can be any!!

The model helps form recurrence formula

- Let p_n be the probability of stopping after **n** flips

$$p_1 = 0 \quad p_2 = 1/4 \quad p_3 = 1/8 \quad p_4 = 1/8 \quad \dots$$

$P(n = n_0) = ?$

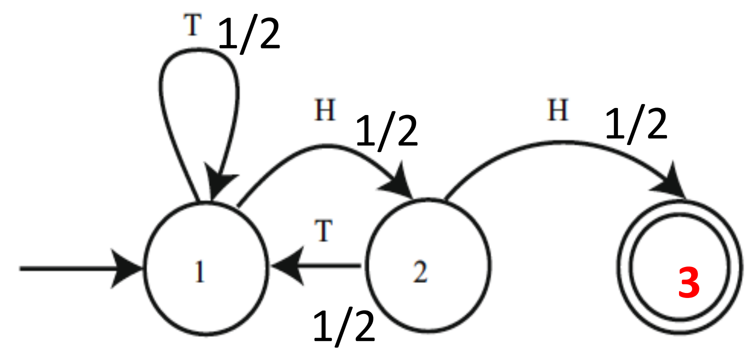
THH

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

HTHH

TTHH $= \frac{1}{8}$

$\frac{1}{16}$



The model helps form recurrence formula

- Let p_n be the probability of stopping after **n** flips

$$p_1 = 0 \quad p_2 = 1/4 \quad p_3 = 1/8 \quad p_4 = 1/8 \quad \ldots$$

T (n-1)

n

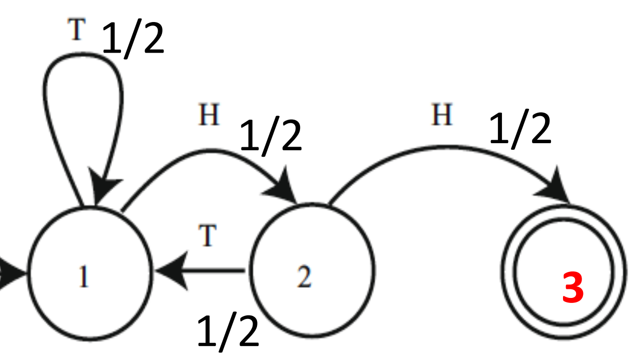
- If $n > 2$, there are two ways the sequence starts
  - Toss T and finish in n-1 tosses
  - Or toss HT and finish in n-2 tosses

HT n-2

- So we can derive a recurrence relation

$$p_n = \frac{1}{2} p_{n-1} + \frac{1}{4} p_{n-2}$$

P(T)· P(n-1|T) P(HT)· P(n-2|HT)



$$r^n = \frac{1}{2} r^{n-1} + \frac{1}{4} r^{n-2}$$

$$r^2 = \frac{1}{2} r + \frac{1}{4}$$

$$4r^2 = 2r + 1$$

$$4r^2 - 2r - 1 = 0$$

$$r = \frac{2 \pm \sqrt{20}}{2 \times 4} = \frac{2 \pm 2\sqrt{5}}{8}$$

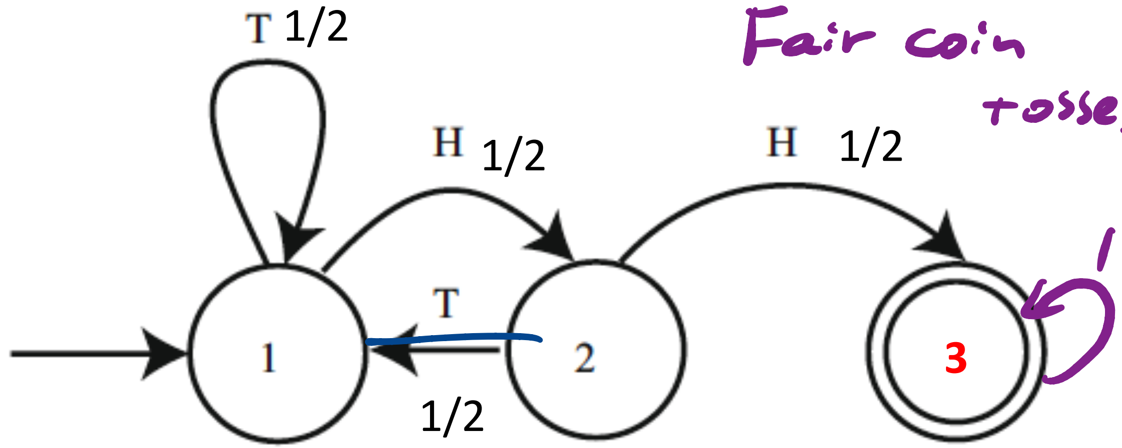
$$p_n = a r_1^n + b r_2^n$$

solve for a, b
using
$P_1 = 0$
$P_2 = \frac{1}{4}$

Transition probability btw states

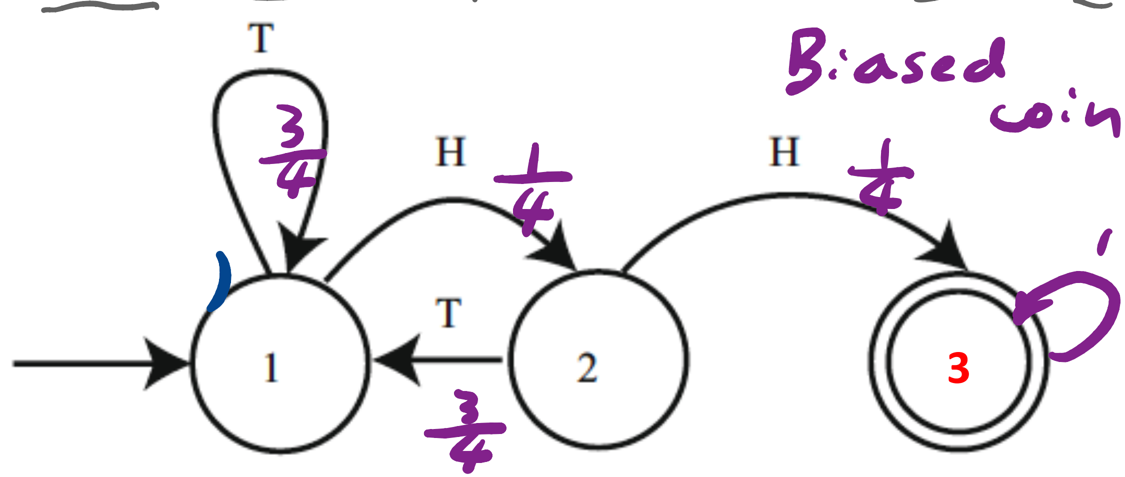
State (t ↓, $t+1$ →)	1	2	3
1	$\frac{1}{2}$	$\frac{1}{2}$	0
2	$\frac{1}{2}$	0	$\frac{1}{2}$
3	0	0	1

Fair coin tosses



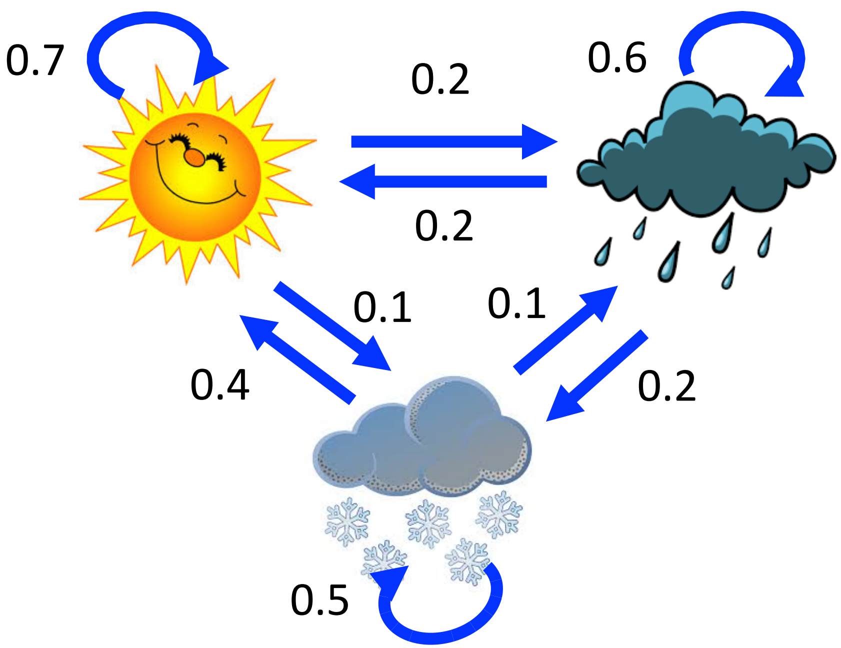
	1	2	3
1	$\frac{3}{4}$	$\frac{1}{4}$	0
2	$\frac{3}{4}$	0	$\frac{1}{4}$
3	0	0	1

Biased coin



Transition probability matrix: weather model

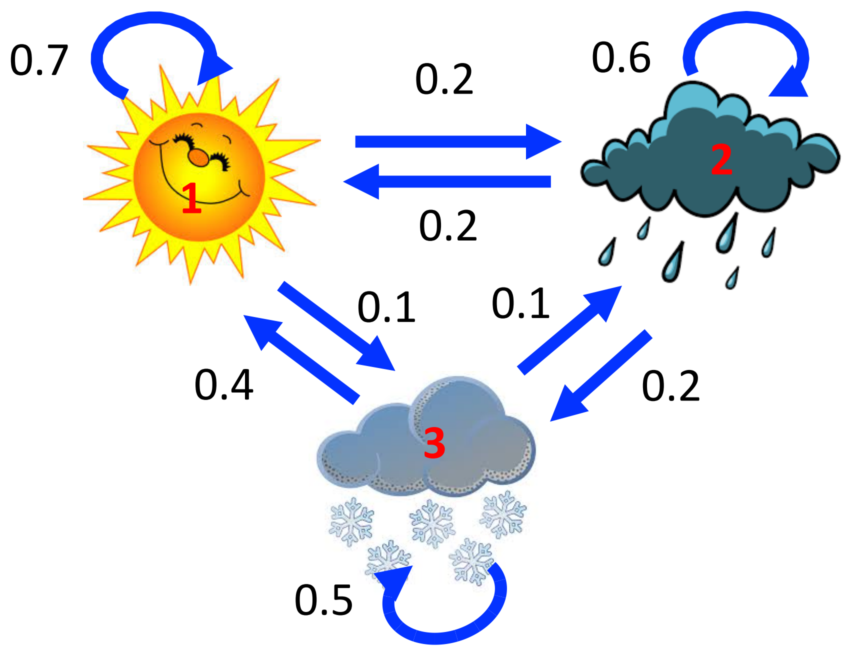
- Let's model daily weather as one of the three states (Sunny, Rainy, and Snowy) with Markov chain that has the transition probabilities as shown here.



	Sunny	Rainy	Snowy
Sun	0.7	0.2	0.1
R	0.2	0.6	0.2
Snow	0.4	0.1	0.5

Transition probability matrix: weather model

- Let's model daily weather as one of the three states (Sunny, Rainy, and Snowy) with Markov chain that has the transition probabilities as shown here.



i, the current state at time point t

j, the next state at time point t+1

$P(X_{t+1}=Sun \mid X_t=Sun)$

$$
\begin{array}{cc}
 & \begin{array}{ccc} \text{Sunny} & \text{Rainy} & \text{Snowy} \end{array} \\
\boldsymbol{P} = & \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.4 & 0.1 & 0.5 \end{bmatrix} \begin{array}{l} \text{Sunny} \\ \text{Rainy} \\ \text{Snowy} \end{array}
\end{array}
$$

The transition probability matrix

Q: Is this TRUE?

For a constant Markov Chain, at any step **t**, the probability distribution among the states remain the same.

$$P(X_{t+1} = s_a \mid X_t = s_b) = C$$

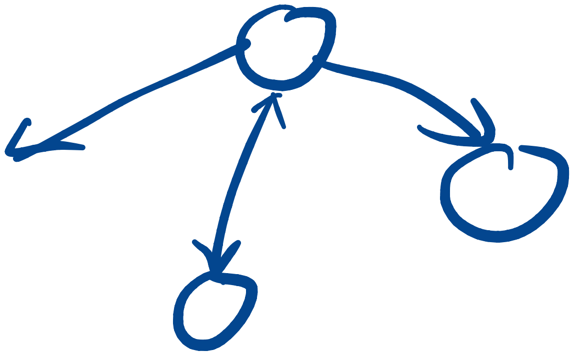
A. Yes.

$\boxed{\text{B.}}$ No.

Q: The transition probabilities for a node sum to 1

A. Yes.

B. No.



Transition probability matrix properties

- The transition probability matrix $\boldsymbol{P}$ is a square matrix with entries p_{ij}
- Since $p_{ij} = P(X_t = j | X_{t-1} = i)$

① $p_{ij} \geq 0$ and ② $\sum_j p_{ij} = 1$

Stochastic Matrix

$$\boldsymbol{P} = \begin{array}{c} \\ \\ \\ \end{array}\begin{array}{cc} \begin{array}{ccc} \text{Sunny} & \text{Rainy} & \text{Snowy} \end{array} & \\ \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.4 & 0.1 & 0.5 \end{bmatrix} & \begin{array}{c} \text{Sunny} \\ \text{Rainy} \\ \text{Snowy} \end{array} \end{array}$$

The transition probability matrix

Probability distributions over states

- Let $\boldsymbol{\pi}$ be a row vector containing the probability distribution over all the finite discrete states at t=0

$$\pi_i = P(X_0 = i)$$

- For example: if it is rainy today, and today is t=0, then

$$\boldsymbol{\pi} = \begin{array}{ccc} \text{Sunny} & \text{Rainy} & \text{Snowy} \\ \end{array}$$
$$\boldsymbol{\pi} = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$$

$P(A \cap B)$
$= P(A) P(B|A)$

- Let $\mathbf{P}^{(t)}$ be a row vector containing the probability distribution over states at time point t

$$\mathrm{p}_i^{(t)} = P(X_t = i)$$

$P(X_1 = 1)$
$\sum_j P(X_0 = S_j) P(X_1 | X_0)$

Propagating the probability distribution

- Propagating from t=0 to t=1,

$$\begin{aligned} P_j^{(1)} &= P(X_1 = j) \\ &= \sum_i P(X_1 = j, X_0 = i) \\ &= \sum_i P(X_1 = j | X_0 = i) P(X_0 = i) \\ &= \sum_i p_{ij} \pi_i \end{aligned}$$

$\pi_0 = [0, 1, 0]$

- In matrix notation,

$$\boxed{\boldsymbol{p}^{(1)} = \boldsymbol{\pi} P}$$

Probability distributions:

- Suppose that it is rainy, we have the initial probability distribution. $\boldsymbol{\pi} = \begin{bmatrix} 0 & 1 & 0 \end{bmatrix}$

- What are the probability distributions for tomorrow and the day after tomorrow?

$$\boldsymbol{p}^{(1)} = \boldsymbol{\pi} P$$

new prior

$$\boldsymbol{p}^{(2)} = \boldsymbol{p}^{(1)} P$$

Propagating to t= ∞

- We have just seen that

$$\boldsymbol{p}^{(2)} = \boldsymbol{p}^{(1)}P = (\boldsymbol{\pi}P)P = \boldsymbol{\pi}P^2$$

- So in general

$$\underset{1\times n\#s}{\boldsymbol{p}^{(t)}} = \underset{1\times n\#s}{\boldsymbol{\pi}} \underset{\#s\times\#s}{P^t}$$

- If one state can be reached from any other state in the graph, the Markov chain is called **irreducible** (single chain).
- Furthermore , if it satisfies:

$$\lim_{t\to\infty} \boldsymbol{\pi}P^t = \boldsymbol{S}$$

then the Markov chain is stationary and **S** is the stationary distribution.

$$\underset{t\to\infty}{P}(X_t)$$

Stationary distribution

- The stationary distribution $\boldsymbol{s}$ has the following property: $\underset{S_{t+1}}{\boldsymbol{s}P} = \underset{S_t}{\boldsymbol{s}}$

$$S_{\text{at } t} \cdot P = S_{\text{at } t+1}$$

- $\boldsymbol{s}$ is a row eigenvector of **P** with eigenvalue 1

- In the example of the weather model, regardless of the initial distribution,

Sunny Rainy Snowy

$$\boldsymbol{S} = \lim_{t\to\infty} \boldsymbol{\pi} \begin{bmatrix} 0.7 & 0.2 & 0.1 \\ 0.2 & 0.6 & 0.2 \\ 0.4 & 0.1 & 0.5 \end{bmatrix}^{\mathbf{t}} = \begin{bmatrix} \frac{18}{37} & \frac{11}{37} & \frac{8}{37} \end{bmatrix}$$

$$sp = s$$

$$(sp)^T = s^T$$

$$p^T s^T = s^T$$

$$X = s^T$$

$$AX = X$$

$$\uparrow$$

$$\lambda = 1$$

Chance of being up-to-date

In a class, students are either up-to-date or behind regarding progress. If a student is up-to-date, the student has 0.8 probability remaining up-to-date, if a student is behind, the student has 0.6 probability becoming up-to-date. Suppose the course is so long that it runs life long, what is the probability any student eventually gets up-to-date?

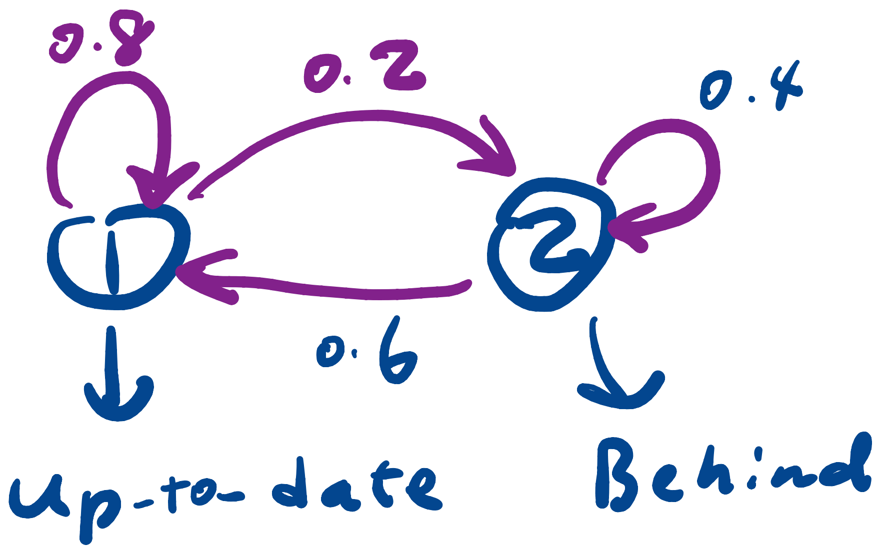
A) 25%

B) 50%

C) 75%

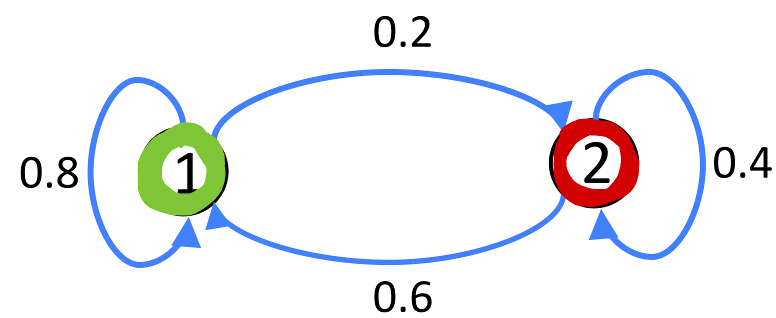
D) 95%

The Markov Model



Example: Up-to-date or behind model

State 1: Up-to-date
State 2: Behind



What's the transition matrix?
If I start with π = [0, 1], what is my probability of being up-to-date eventually? 3/4

$$P = \begin{array}{cc} & \begin{array}{cc} 1 & 2 \end{array} \\ \begin{array}{c} (U)\ 1 \\ (B)\ 2 \end{array} & \begin{bmatrix} 0.8 & 0.2 \\ 0.6 & 0.4 \end{bmatrix} \end{array}$$

$S = ?$

Solving the stationary Markov

Given $sP = s$ What is s ? u^T

$s = [\frac{3}{4} \quad \frac{1}{4}]$

$(sP)^T = s^T$

$P^T s^T = s^T$

$Au = u \quad (A = P^T, \ u = s^T)$

$Au = 1 \times u \Rightarrow Au = \lambda u \quad (\lambda = 1)$

$[A - I]u = 0$

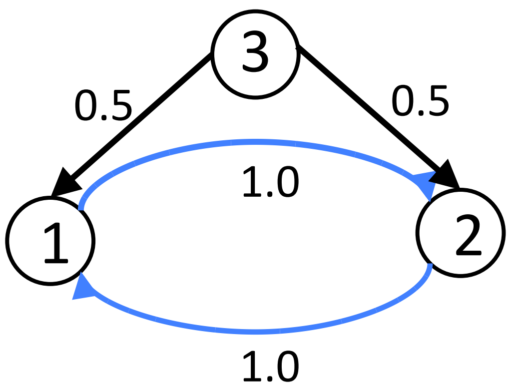
$\begin{bmatrix} 0.8-1 & 0.2-0 \\ 0.6-0 & 0.4-1 \end{bmatrix} u = 0$

$u = ?$

$u = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \quad u_1 + u_2 = 1$

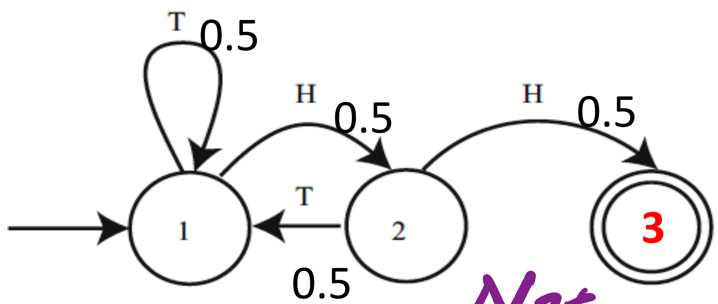
$u_1 = \frac{3}{4} \quad u_2 = \frac{1}{4}$

Examples of non-stationary Markov chains



Periodic

Not irreducible



Not stationary

Absorbing

$SP = S$

Additional References

- Robert V. Hogg, Elliot A. Tanis and Dale L. Zimmerman. “Probability and Statistical Inference”
- Kelvin Murphy, “Machine learning, A Probabilistic perspective”

Acknowledgement

See You!