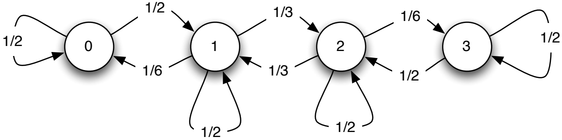
COMPSCI 240: Reasoning Under Uncertainty

Andrew Lan and Nic Herndon

University of Massachusetts at Amherst

Spring 2019

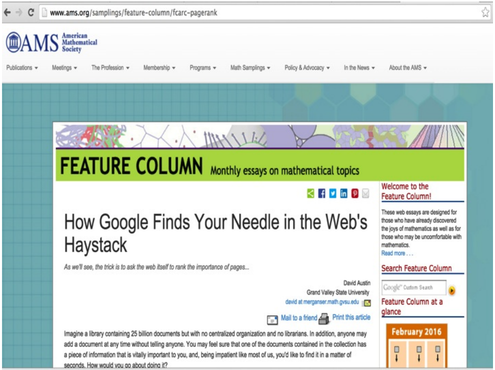
Lecture 24: Markov Chains



Life without Google

PageRank

How can one rank the Web pages according to its relevance to a search query?

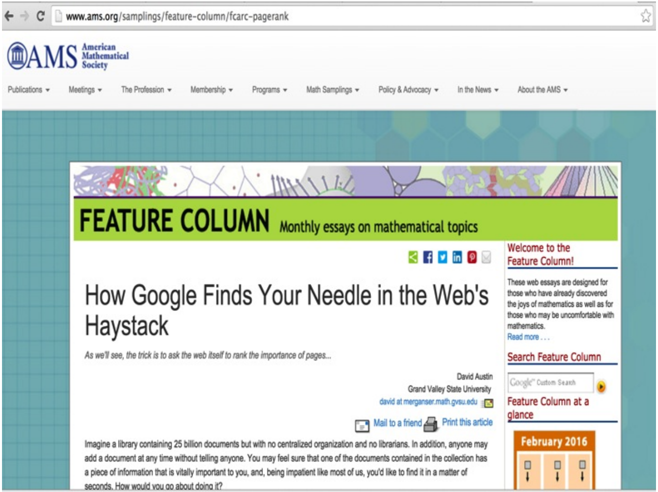


Page, Lawrence and Brin, Sergey and Motwani, Rajeev and Winograd, Terry (1999) The PageRank Citation Ranking: Bringing Order to the Web. Technical Report. Stanford InfoLab.

PageRank

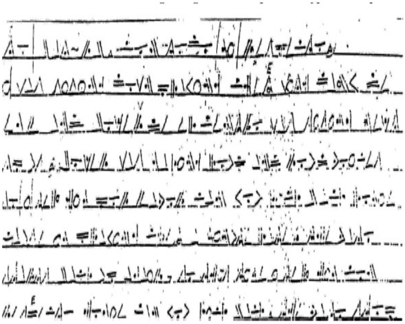
An important page has many links to it and can be reached easily.

PageRank, which is the stationary vector of an enormous Markov chain, is the driving force behind Google's success in ranking webpages.



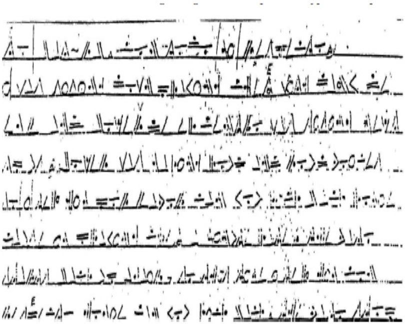
Cryptography

Stanford's Statistics Department has a drop-in consulting service. One day, a psychologist from the state prison system showed up with a collection of coded messages.



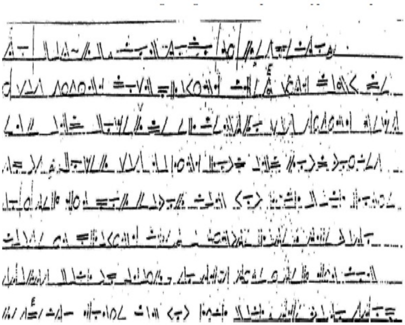
Cryptography

Stanford's Statistics Department has a drop-in consulting service. One day, a psychologist from the state prison system showed up with a collection of coded messages. What could be a feasible way to decode the passage?



Cryptography

Stanford's Statistics Department has a drop-in consulting service. One day, a psychologist from the state prison system showed up with a collection of coded messages. Suppose the first symbol corresponds to "A". What is the probability that the next symbol is A/B/C/..../Z?



Markov Chain

- We learned discrete random variables, such as the Bernoulli and Poisson, that are memoryless.
- That is, each event is identical and independent – current event is independent from the history (memory) of previous events.
- We now consider scenarios where the future depends on past only through present!
- The condition of the future is summarized by a state, which changes over time according to given probabilities.

Markov Chain

More Examples:

- Whether you would understand the content of the next class only depends on whether you understand the concept in today's class.
- Performance of a person's daily activity (e.g., driving, walking, cooking, eating, walking) at time t depends on the activity at $t - 1$.

Discrete Markov Chain

- We consider discrete-time Markov chain, in which the state changes at certain discrete time instances, indexed by an integer variable t.
- A discrete Markov chain defines a series of random variables X_t, e.g., $\{X_0, X_1, X_2, \ldots\}$.
- A Markov Chain consists:
  - **State space**: a set of states in which the chain can be described at time t:

$$\mathcal{S} = \{s_1, \ldots, s_k\}$$

  - **Transition probabilities** that describe the probability of transitioning from a state at $t-1$ to another state at t:

$$P\left(X_t = s_j | X_{t-1} = s_i\right) = p_{ij} \text{ for all } 1 \leq i, j \leq k$$

  - **An initial state** X_0, in which the chain is initiated.

Markov Property

- The key assumption is that the transition probabilities (p_{ij}) for the state at time $t+1$ (state j) only depends on the state at time t (state i).
  - The value of X_{t+1} only depends on the value of X_t.
- Mathematically, the **Markov property** defines that

$$P\left(X_{t+1}=j|X_t=i, X_{t-1}=x_{t-1}, \cdots, X_0=x_0\right)$$

$$=P\left(X_{t+1}=j|X_t=i\right)$$

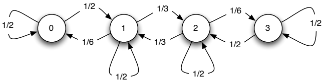
$$=p_{ij}$$

- The transition probability p_{ij} must be **non-negative** and **sum to 1**:

$$\sum_{j=1}^{k} p_{ij}=1, \text{ for all } i.$$

Transition Probability Graph

A Markov chain can be described using **transition probability graph**, whose nodes are the states and whose arrows are the possible transitions (with probabilities).



- **Example:** Ivan is a student who has three emotions: 1) neutral, 2) sad, and 3) happy.
- If he is neutral at a give time t, he will be neutral with a probability of 0.8, sad with a probability of 0.1, and happy with a probability of 0.1 at time $t + 1$.
- If he is sad at t, he will be sad with a probability of 0.5, neural with 0.4, and happy with 0.1.
- If he is happy at t, he will be happy with a probability of 0.5, neural with 0.4, and sad with 0.1.
- Draw the probability transition graph.

Transition Probability Graph

- Weights on arrows out of each state i sum to one: $\sum_j p_{ij} = 1$

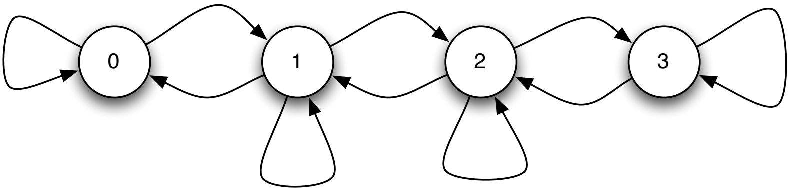
Analyzing the Queue at Amherst Coffee

- Consider a queue at Amherst Coffee
- Every minute, someone joins the queue...
  - With probability 1 if the queue has length 0
  - With probability 2/3 if the queue has length 1
  - With probability 1/3 if the queue has length 2
  - With probability 0 if the queue has length 3.
- Every minute, the server serves a customer with probability 1/2.

Suppose there is one person in line at noon. How many people might be in line at 12:10pm?

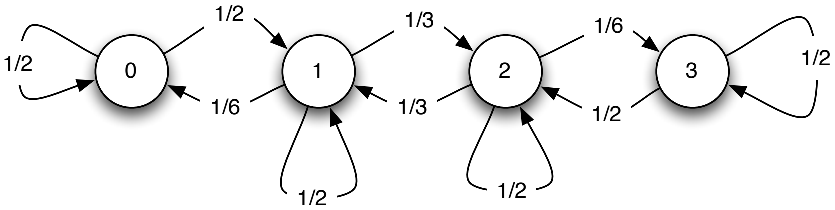
States

- Let X_t be the number of people in the queue at time t
- At any given time t, the queue is in one of four *states*: either there are 0, 1, 2, or 3 people in the queue.
- Arrows indicate that it is possible to move from one state to the next at each step.



States with Transition Probabilities

- Weight p_{ij} on arrow from state i to state j indicates the probability of transitioning to state j given we're in state i.



- **Question: If we're in state 2, what's the probability we're in state 3 after one step: A) 1, B) 1/2, C) 1/6, D) 0, E) 1/3.**
  Answer: C) 1/6.
- **Question: If we're in state 2, what's the probability we're in state 2 after two steps: A) 1/3, B) 4/9, C) 1/4, D) 1/12, E) 1/9.**
  Answer: B) 4/9.

States with Transition Probabilities

- Weight p_{ij} on arrow from state i to state j indicates the probability of transitioning to state j given we're in state i.

i	p_i(join)	p_i(not join)	p_i(served)	p_i(not served)
0	1	0	1/2	1/2
1	2/3	1/3	1/2	1/2
2	1/3	2/3	1/2	1/2
3	0	1	1/2	1/2

$$p_{ij} = \begin{cases} p_i(\text{not join}) \cdot p_i(\text{served}) & , j = i - 1 \\ p_i(\text{join}) \cdot p_i(\text{served}) + p_i(\text{not join}) \cdot p_i(\text{not served}) & , j = i \\ p_i(\text{join}) \cdot p_i(\text{not served}) & , j = i + 1 \end{cases}$$

$$p_{10} = p_1(\text{not join}) \cdot p_1(\text{served}) = 1/3 \cdot 1/2 = 1/6$$
$$p_{11} = p_1(\text{join}) \cdot p_1(\text{served}) + p_1(\text{not join}) \cdot p_1(\text{not served}) = 2/3 \cdot 1/2 + 1/3 \cdot 1/2 = 1/2$$
$$p_{12} = p_1(\text{join}) \cdot p_1(\text{not served}) = 2/3 \cdot 1/2 = 1/3$$

What if the current state is uncertain?

- What if we don't know X_{t-1}, but know $P(X_{t-1} = i)$ for each i, what's $P(X_t = j)$?
- Then, by the **Law of Total Probability**:

$$
\begin{aligned}
P(X_t = j) &= \sum_i P(X_t = j, X_{t-1} = i) \\
&= \sum_i P(X_t = j | X_{t-1} = i) P(X_{t-1} = i) \\
&= \sum_i p_{ij} P(X_{t-1} = i)
\end{aligned}
$$

- **Question**: If there's a $1/3$ probability we're in state 1 and a $2/3$ probability we're in state 3, what's the probability we're in state 2 after one step.
  A) $1/3$, B) $1/4$, C) $4/9$, D) $7/9$, E) $1/9$.
  Ans: C) $4/9$.

Markov Chain Theorem

Theorem

*We define the **distribution** of X_t as*

$$\begin{aligned} v_t &= \langle v_t[1], v_t[2], \cdots, v_t[k] \rangle \\ &= \langle P(X_t = 1), P(X_t = 2), \ldots, P(X_t = k) \rangle . \end{aligned}$$

where

$$\begin{aligned} v_t[j] &= P(X_t = j) \\ &= \sum_i P(X_t = j | X_{t-1} = i) P(X_{t-1} = i) \\ &= \sum_i p_{ij} v_{t-1}[i]. \end{aligned}$$

Thus,

$$v_t = \left\langle \sum_i p_{i1} v_{t-1}[i], \sum_i p_{i2} v_{t-1}[i], \cdots, \sum_i p_{ik} v_{t-1}[i] \right\rangle .$$

Markov Chain Theorem

This implies that if we know the distribution at $t = 0$ (i.e., v_0), then we can compute any v_t where $t > 0$:

$$v_1 = \left\langle \sum_i p_{i1} v_0[i], \sum_i p_{i2} v_0[i], \cdots, \sum_i p_{ik} v_0[i] \right\rangle .$$

$$v_2 = \left\langle \sum_i p_{i1} v_1[i], \sum_i p_{i2} v_1[i], \cdots, \sum_i p_{ik} v_1[i] \right\rangle .$$

$$v_3 = \left\langle \sum_i p_{i1} v_2[i], \sum_i p_{i2} v_2[i], \cdots, \sum_i p_{ik} v_2[i] \right\rangle .$$

$$\vdots$$

$$v_t = \left\langle \sum_i p_{i1} v_{t-1}[i], \sum_i p_{i2} v_{t-1}[i], \cdots, \sum_i p_{ik} v_{t-1}[i] \right\rangle .$$

Markov Chain Theorem

- This theorem can be **effectively** represented using matrices, but it requires knowledge about linear algebra.
- To give you a short overview, a Markov chain model can be encoded in a **transition probability matrix**. Make sure that you remember the following notation:

$$A = \begin{pmatrix} p_{1,1} & p_{1,2} & \cdots & p_{1,k} \\ p_{2,1} & p_{2,2} & \cdots & p_{2,k} \\ \vdots & \vdots & \ddots & \vdots \\ p_{k,1} & p_{k,2} & \cdots & p_{k,k} \end{pmatrix}$$

- **Markov Chain Theorem**: Given v_0, we can compute $v_1 = v_0 A$, and

$$v_t = v_{t-1}A = v_{t-2}AA = v_{t-3}AAA = \ldots = v_0 A^t$$