15-252
More Great Ideas in Theoretical Computer Science
Markov Chains

April 27th, 2018
Markov Chain

Andrey Markov (1856 - 1922)
Russian mathematician.
Famous for his work on random processes.

\[ \Pr[X \geq c \cdot E[X]] \leq 1/c \text{ is Markov’s Inequality.} \]

A model for the evolution of a random system.

The future is independent of the past, given the present.
Cool things about Markov Chains

- It is a very general and natural model.

Applications in:
  - computer science, mathematics, biology, physics,
  - chemistry, economics, psychology, music, baseball,
...

- The model is simple and neat.

- Cilantro
The plan

Motivating examples and applications

Basic mathematical representation and properties

A bit more on applications
The future is independent of the past, given the present.
Some Examples of Markov Chains
Example: Drunkard Walk

Salvador Dali (1922)
The Drunkard

Home
Example: Diffusion Process
Example: Weather

A very(!!) simplified model for the weather.

Probabilities on a daily basis:

\[
\begin{align*}
\text{Pr}[\text{sunny to rainy}] &= 0.1 \\
\text{Pr}[\text{sunny to sunny}] &= 0.9 \\
\text{Pr}[\text{rainy to rainy}] &= 0.5 \\
\text{Pr}[\text{rainy to sunny}] &= 0.5
\end{align*}
\]

Encode more information about current state for a more accurate model.
Example: Life Insurance

Goal of life insurance company:
   figure out how much to charge the clients.

Find a model for how long a client will live.

Probabilistic model of health on a monthly basis:

\[
\begin{align*}
\Pr[\text{healthy to sick}] &= 0.3 \\
\Pr[\text{sick to healthy}] &= 0.8 \\
\Pr[\text{sick to death}] &= 0.1 \\
\Pr[\text{healthy to death}] &= 0.01 \\
\Pr[\text{healthy to healthy}] &= 0.69 \\
\Pr[\text{sick to sick}] &= 0.1 \\
\Pr[\text{death to death}] &= 1
\end{align*}
\]
Example: Life Insurance

Goal of life insurance company:
figure out how much to charge the clients.

Find a model for how long a client will live.

Probabilistic model of health on a monthly basis:
Some Applications of Markov Models
Application: Algorithmic Music Composition

Nicholas Vasallo

Megalithic Copier #2: Markov Chains (2011)

written in Pure Data
Application: Image Segmentation
Random text generated by a computer (putting random words together):

“While at a conference a few weeks back, I spent an interesting evening with a grain of salt.”

Google: Mark V Shaney
Speech recognition software programs use Markov models to listen to the sound of your voice and convert it into text.
Application: Google PageRank

1997: Web search was horrible

Sorts webpages by number of occurrences of keyword(s).
Application: Google PageRank

Founders of Google

Larry Page    Sergey Brin

$40Billionaires
Jon Kleinberg

Nevanlinna Prize
How does Google order the webpages displayed after a search?

2 important factors:

- Relevance of the page.

- Reputation of the page.

  The number and reputation of links pointing to that page.

Reputation is measured using PageRank.

PageRank is calculated using a Markov Chain.
The plan

Motivating examples and applications

Basic mathematical representation and properties

A bit more on applications
The Setting

There is a system with \( n \) possible states/values \{1, 2, \ldots, n\}. At each time step, the state changes probabilistically.

**Memoryless**

The next state only depends on the current state.

*Evolution of the system:* random walk on the graph.
A **Markov Chain** is a digraph with $V = \{1, 2, \ldots, n\}$ such that:

- Each edge is labeled with a value in $(0, 1]$ (a probability).
  - self-loops allowed

- At each vertex, the probabilities on outgoing edges sum to 1.

(We usually assume the graph is strongly connected.
  i.e. there is a directed path from $i$ to $j$ for any $i$ and $j$.)

The vertices of the graph are called **states**.

The edges are called **transitions**.

The label of an edge is a **transition probability**.
Given some Markov Chain with $n$ states:

Define

$$\pi_t[i] = \text{probability of being in state } i \text{ after exactly } t \text{ steps.}$$

$$\pi_t = [p_1 \ p_2 \ \cdots \ p_n] \quad \quad \sum_{i} p_i = 1$$

Note that someone has to provide $\pi_0$.

Once this is known, we get the distributions $\pi_1, \pi_2, \ldots$
A Markov Chain with \( n \) states can be characterized by the \( n \times n \) transition matrix \( K \)

\[
\forall i, j \in \{1, 2, \ldots, n\} \quad K[i, j] = \Pr[i \to j \text{ in one step}]
\]

**Note:** rows of \( K \) sum to 1.
Some Fundamental and Natural Questions

What is the probability of being in state $i$ after $t$ steps (given some initial state)?

$$\pi_t[i] = ?$$

What is the expected time of reaching state $i$ when starting at state $j$?

What is the expected time of having visited every state (given some initial state)?

How do you answer such questions?
Mathematical representation of the evolution

Suppose we start at state 1 and let the system evolve.

How can we mathematically represent the evolution?

\[
\pi_0 = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 1 & 0 & 0 & 0 \end{bmatrix}
\]

What is \( \pi_1 \)?

By inspection, \( \pi_1 = \begin{bmatrix} 0 & 1/2 & 0 & 1/2 \end{bmatrix} \).
Mathematical representation of the evolution

The probability of states after 1 step:

\[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
\end{bmatrix}
\begin{bmatrix}
0 & \frac{1}{2} & 0 & \frac{1}{2} \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
0 & \frac{1}{4} & \frac{3}{4} & 0 \\
\end{bmatrix}
\begin{bmatrix}
\pi_0 \\
\end{bmatrix}
= 
\begin{bmatrix}
0 & \frac{1}{2} & 0 & \frac{1}{2} \\
\end{bmatrix}
\begin{bmatrix}
\pi_1 \\
\end{bmatrix}
\]

the new state (probabilistic)
Mathematical representation of the evolution

The probability of states after 2 steps:

\[
\begin{bmatrix}
0 & \frac{1}{2} & 0 & \frac{1}{2}
\end{bmatrix}
\begin{bmatrix}
0 & \frac{1}{2} & 0 & \frac{1}{2} \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
0 & \frac{1}{4} & \frac{3}{4} & 0
\end{bmatrix}
= \begin{bmatrix}
0 & \frac{1}{8} & \frac{7}{8} & 0
\end{bmatrix}
\]

the new state (probabilistic)
Mathematical representation of the evolution

\[ \pi_1 = \pi_0 \cdot K \]

\[ \pi_2 = \pi_1 \cdot K \]

So \( \pi_2 = (\pi_0 \cdot K) \cdot K \]

\[ = \pi_0 \cdot K^2 \]
Mathematical representation of the evolution

In general:
If the initial probabilistic state is \( \begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix} = \pi_0 \)

\( p_i = \) probability of being in state \( i \),

\( p_1 + p_2 + \cdots + p_n = 1 \),

after \( t \) steps, the probabilistic state is:

\[
\begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix} \begin{bmatrix} \text{Transition Matrix} \end{bmatrix}^t = \pi_t
\]
Remarkable Property of Markov Chains

What happens in the long run?

i.e., can we say anything about \( \pi_t \) for large \( t \) ?

Suppose the Markov chain is “aperiodic”.

Then, as the system evolves, the probabilistic state **converges** to a limiting probabilistic state.

As \( t \to \infty \), for any \( \pi_0 = [p_1 \ p_2 \ \cdots \ p_n] \):

\[
[p_1 \ p_2 \ \cdots \ p_n] \begin{bmatrix}
\text{Transition Matrix} \\
\end{bmatrix}^t \to \pi
\]
Remarkable Property of Markov Chains

In other words:

$$\pi_t \rightarrow \pi \quad \text{as} \quad t \rightarrow \infty.$$ 

Note:

$$\pi \begin{bmatrix} \text{Transition Matrix} \end{bmatrix} = \pi$$  

stationary/invariant distribution

This $\pi$ is unique.
Remarkable Property of Markov Chains

Stationary distribution is \[
\begin{bmatrix}
\frac{5}{6} & \frac{1}{6}
\end{bmatrix}.
\]

\[
\begin{bmatrix}
\frac{5}{6} & \frac{1}{6}
\end{bmatrix} \begin{bmatrix}
0.9 & 0.1 \\
0.5 & 0.5
\end{bmatrix} = \begin{bmatrix}
\frac{5}{6} & \frac{1}{6}
\end{bmatrix}
\]

**In the long run, it is Sunny \(5/6\) of the time, it is Rainy \(1/6\) of the time.**
Remarkable Property of Markov Chains

How did I find the stationary distribution?

\[
\begin{bmatrix}
0.9 & 0.1 \\
0.5 & 0.5
\end{bmatrix}^2 = \begin{bmatrix}
0.86 & 0.14 \\
0.7 & 0.3
\end{bmatrix}
\]

\[
\begin{bmatrix}
0.9 & 0.1 \\
0.5 & 0.5
\end{bmatrix}^4 = \begin{bmatrix}
0.8376 & 0.1624 \\
0.812 & 0.188
\end{bmatrix}
\]

\[
\begin{bmatrix}
0.9 & 0.1 \\
0.5 & 0.5
\end{bmatrix}^8 = \begin{bmatrix}
0.833443 & 0.166557 \\
0.832787 & 0.167213
\end{bmatrix}
\]

**Exercise:** Why do the rows converge to \( \pi \)?
Markov Chains can be characterized by the transition matrix $K$.

$$K[i, j] = \Pr[i \rightarrow j \text{ in one step}]$$

What is the probability of being in state $i$ after $t$ steps?

$$\pi_t = \pi_0 \cdot K^t \quad \pi_t[i] = (\pi_0 \cdot K^t)[i]$$
**Theorem** (Fundamental Theorem of Markov Chains):

Consider a Markov chain that is strongly connected and aperiodic.

- There is a unique invariant/stationary distribution $\pi$ such that
  $$\pi = \pi K.$$ 

- For any initial distribution $\pi_0$,
  $$\lim_{t \to \infty} \pi_0 K^t = \pi$$

- Let $T_{ij}$ be the number of steps it takes to reach state $j$ provided we start at state $i$. Then,
  $$\mathbb{E}[T_{ii}] = \frac{1}{\pi[i]}.$$
The plan

Motivating examples and applications

Basic mathematical representation and properties

A bit more on applications
How are Markov Chains applied?

2 common types of applications:

1. Build a Markov chain as a statistical model of a real-world process.
   
   Use the Markov chain to simulate the process.
   
   e.g. text generation, music composition.

2. Use a measure associated with a Markov chain to approximate a quantity of interest.
   
   e.g. Google PageRank, image segmentation
Generate a superficially real-looking text given a sample document.

**Idea:**
- From the sample document, create a Markov chain.
- Use a random walk on the Markov chain to generate text.

**Example:**
- Collect speeches of Obama, create a Markov chain.
- Use a random walk to generate new speeches.
1. For each word in the document, create a node/state.

2. Put an edge \texttt{word1} ---\texttt{word2} if there is a sentence in which \texttt{word2} comes after \texttt{word1}.

3. Edge probabilities reflect frequency of the pair of words.

The Markov Chain:

- \texttt{like} ---\texttt{a} (3/9)  
  \texttt{like} ---\texttt{the} (4/9)  
  \texttt{like} ---\texttt{to} (2/9)

- Edge probabilities:
  - \texttt{like a} 3 times
  - \texttt{like the} 4 times
  - \texttt{like to} 2 times
“I jumped up. I don't know what's going on so I am coming down with a road to opportunity. I believe we can agree on or do about the major challenges facing our country.”
Another use:

Build a Markov chain based on speeches of Obama.
Build a Markov chain based on speeches of Bush.

Given a new quote, can predict if it is by Obama or Bush.

(by testing which Markov model the quote fits best)
PageRank is a measure of reputation:

The number and reputation of links pointing to you.

**The Markov Chain:**
Google PageRank

PageRank is a measure of reputation:

The number and reputation of links pointing to you.

The Markov Chain:

1. Every webpage is a node/state.

2. Each hyperlink is an edge:

   if webpage A has a link to webpage B,  \( A \rightarrow B \)

3a. If A has \( m \) outgoing edges, each gets label \( \frac{1}{m} \).

3b. If A has no outgoing edges, put edge \( A \rightarrow B \) \( \forall B \)

(jump to a random page)
A little tweak:

Random surfer jumps to a random page with 15% prob.

Stationary distribution:

probability of being at webpage $A$ in the long run

PageRank of webpage $A$

= 

The stationary probability of $A$
Google PageRank

B: 38.4%

C: 34.3%

A: 3.3%

D: 3.9%

E: 8.1%

F: 3.9%

1.6% 1.6% 1.6% 1.6%
Google: 

“PageRank continues to be the heart of our software.”
The plan

Motivating examples and applications

Basic mathematical representation and properties

A bit more on applications