Library Lunchtime Lecture Series

Markov and His Chains

Tuğrul Dayar
İhsan Doğramacı Bilkent University
Department of Computer Engineering

20 November 2014
Who Is A. A. Markov?

- Andrei Andreevich Markov  
  (Ryazan 1856 - St. Petersburg 1922)

- Son of:  
  Andrei Grigorievich Markov  
  Nadezhda Petrovna Fedorova

- Doctoral (1880 - 1885) Advisor: 
  Pafnuty Lvovich Chebyshev  
  Thesis on number theory

- Academic life associated with 
  St. Petersburg University and 
  Imperial Academy of Sciences

http://logic.pdmi.ras.ru/Markov
On A. A. Markov (continued)

- One of the best chess players in St. Petersburg
- Paid great attention to teaching mathematics, emphasized problem solving

A. A. Markov excelled in:
  - Theory of Numbers
  - Mathematical Analysis
  - Probability Theory

Taught course on probability theory yearly, starting in 1883 after Chebyshev’s departure from university, even after retirement as Emeritus Professor in 1905 until 1913

- His son, Andrei Andreevich Markov, Jr. (1903-1979), also a prominent mathematician
On A. A. Markov (continued)

1846: Chebyshev’s PhD thesis at Moscow University
   “An Experience in an Elementary Analysis of the Probability Theory” (1st Russian dis. on probability theory)

1880: Markov’s MS thesis at St. Petersburg University
   “On the Binary Square Forms with Positive Determinants”

1884: Markov’s PhD thesis at St. Petersburg University
   “On Certain Applications of the Algebraic Continuous Fractions”

1890: Publication of Markov’s book “Calculus of Probabilities”
   (published 4 times in Russian, translated to German)
Earlier Ideas Related to Markov chains

- Some *urn problems* studied by
  - Jacob Bernoulli (1654-1705)
  - Pierre-Simon Laplace (1749-1827)
  - Paul Ehrenfest (1880-1933)
  
  are special cases of Markov chains (MCs)

- *Brownian motion* due to
  - Robert Brown (1773-1858)

- *Random walks*, such as gambler’s ruin problem, studied by
  - Viktor Yakovlevich Bunyakovsky (1804-1889)

- Study of the stock exchange by
  - Louis Jean-Baptiste Alphonse Bachelier (1870-1946)
First Appearance of Markov Chains

Earlier studies were either unknown or unrecognized by Markov.

Gely P. Basharin, Amy N. Langville, Valeriy A. Naumov,
The life and work of A. A. Markov,
*Linear Algebra and Its Applications* 386 (2004), 3-26

First appearance of *Markov chains* in:


as a sequence of dependent random variables for which the weak law of large numbers hold
Some Other Contributors

- Oskar Perron (1880-1975),
  Ferdinand Georg Frobenius (1849-1917)
  Theory of finite nonnegative matrices (1907, 1909, 1912)

- Richard Edler von Mises (1883-1953)
  Connection between MCs and matrices (1931)

- Yakov Viktorovich Uspensky (1883-1947)
  Book bringing analytical probability in St. Petersburg tradition to States (1937)

- Maurice René Fréchet (1878-1973)
  First book on finite MCs from a matrix standpoint (1938)
Some Other Contributors (continued)

- **Vsevolod Ivanovich Romanovsky** (1879-1954)
  Book on discrete MCs having intricate algebraic treatment (1949)

- **William Feller** (1906-1970)
  Book on probability theory including denumerable MCs (1951)

- **John George Kemeny** (1926-1992),
  **James Laurie Snell** (1925-2011)
  Matrix-theoretic treatment of finite MCs avoiding spectral theory (1960)


Tuğrul Dayar,
20 November 2014
Dept. of Computer Engineering
Uses of **Markov Chains**

In 1926, Sergei Natanovich Bernstein (1880-1968) uses the phrase “Markov Chain” in a paper

*Application areas*:

- Chemistry (Michaelis-Menten kinetics)
- Economics and Finance (Leontief models)
- Internet Applications (PageRank)
- Mathematical biology (interacting species)
- Physics (thermodynamics and statistical mechanics)
- Queueing Theory (performance of computer networks)
- Social Sciences (demographic studies)
- Statistics (Markov Chain Monte Carlo)

and others
Umbrella Problem
(from book by Harchol-Balter, 2013)

- Absent-minded professor has 2 umbrellas she uses when commuting from home to office and back
- If it rains and an umbrella is available in her location, she takes it
  If it is not raining, she always forgets to take an umbrella
- Suppose it rains with probability $p$ each time she commutes, independently of prior commutes

What is the percentage of commutes during which she gets wet?
We need 3 states to model what goes on

States track *the number of available umbrellas at current location* (regardless of what current location is)

State space is \{0,1,2\}
Umbrella Problem (continued)

Next state depends *only* on current state.

Possible sequences of 10-step states:

- 0, 2, 0, 2, 1, 1, 1, 2, 0, 2, 1
  0 -> 2 -> 0 -> 2 -> 1 -> 1 -> 2 -> 0 -> 2 -> 1
  with probability
  \[ 1 \cdot (1-p) \cdot 1 \cdot p \cdot (1-p) \cdot (1-p) \cdot p \cdot (1-p) \cdot 1 \cdot p = (1-p)^4 p^3 \]

- 2, 1, 2, 0, 2, 1, 1, 2, 0, 2, 1 with probability \((1-p)^3 p^5\)

What is the probability that after \(n\) steps the professor ends up in state \(j\) if she starts in state \(i\)?
## Markov Chain Model as a Matrix

Row indices denote current state, column indices denote next state.

<table>
<thead>
<tr>
<th></th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>1-p</td>
<td>p</td>
</tr>
<tr>
<td>2</td>
<td>1-p</td>
<td>p</td>
<td>0</td>
</tr>
</tbody>
</table>

\[ P = \begin{pmatrix}
0 & 0 & 1 \\
1 & 0 & 1-p & p \\
2 & 1-p & p & 0
\end{pmatrix} \]

\*\*P is a \textit{stochastic} matrix of 1-step transition probabilities\*\*

\( P^{2} = P \cdot P = \begin{pmatrix}
0 & 1 & 2 \\
0 & 1-p & p & 0 \\
1-p & p & 0 & 1-p & p & (1-p) & p & (1-p) \\
2 & 0 & p (1-p) & 1-p + p^2 & 0 & p (1-p) & 0 & p (1-p) & 1-p + p^2 & 0
\end{pmatrix} \)

\( P^{2} \) is a \textit{probability matrix},

\( P^{2} = P \cdot P \) is also \textit{stochastic}.

\( P^{2} \) shows the 2-step transition probabilities.
Discrete-time Markov Chains

What is the probability that the professor ends up after $n$ steps in state $j$ if she starts in state $i$?

$p_{ij}^n$ : $ij$ th element of $P^n$ ($n$-step probability transition matrix)

Model of a discrete-time Markov chain (DTMC):
- sequence of random variables forming the chain are dependent (with dependency of next state only on current state, order 1)
- # of steps to exit current state $i$ is geometrically distributed (with parameter equal to $p_{ii}$)

If the professor starts with initial probability distribution $\pi^{(0)}$, whose $i$th element is the probability of being in state $i$ initially, what is the probability of being in state $j$ after $n$ steps?
Discrete-time Markov Chains (continued)

Probability distribution at step $n$
(also known as transient probability distribution at step $n$):

$$\pi^{(n)} = \pi^{(n-1)} \cdot P \quad \text{for} \quad n > 0$$
$$\pi^{(n)} = \pi^{(0)} \cdot P^n$$

Here, $\pi^{(n)}$ is a $(1 \times n)$ probability vector (taken as a row vector), meaning $\pi^{(n)} \geq 0$ and $\sum_i \pi_i^{(n)} = 1$

Ex: If $\pi^{(0)} = (1 \ 0 \ 0)$, then $\pi^{(1)} = \pi^{(0)} \cdot P = (0 \ 0 \ 1)$
$$\pi^{(2)} = \pi^{(1)} \cdot P = (1-p \ p \ 0)$$

As $n$ becomes larger (that is, in the long-run), how do the elements of $P^n$ behave?
Discrete-time Markov Chains (continued)

If $P$ corresponds to an ergodic DTMC, that is:

- irreducible (i.e., each state is reachable from every other state)
- aperiodic (i.e., current state is not tied to time step)
- positive recurrent (i.e., each state is revisited infinitely often and mean time between visits is finite)

then

$$\lim_{n \to \infty} P^n = \begin{pmatrix} \pi \\ \pi \\ \pi \end{pmatrix}$$

$\pi$: steady-state probability distribution
(satisfies $\pi.P = \pi$, $\sum_i \pi_i = 1$; hence, also a stationary distr.)
Umbrella Problem (continued)

\[ \pi_0 = \frac{1-\rho}{3-\rho}, \quad \pi_1 = \frac{1}{3-\rho}, \quad \pi_2 = \frac{1}{3-\rho} \]

Note that \( \pi_0 + \pi_1 + \pi_2 = 1 \)

What is the percentage of commutes during which the absent-minded professor gets wet?

The professor gets wet if she has no umbrellas \textit{and} it rains:

\[ \pi_0 \cdot \rho \]

If \( \rho = 0.6 \), then 10% of the time she gets wet!!!
A Queueing Station with Unbounded Waiting Line

\[ \lambda \rightarrow \text{waiting line} \rightarrow \mu \rightarrow \text{server} \rightarrow \text{departure} \]

- \( \lambda \): arrival rate
- \( \mu \): service rate
- \( \rho = \lambda / \mu \) (load)
States track total # of customers in the server and waiting line

State space is \{0, 1, 2, \ldots\}

State 0: Server is *idle*
State 1: Server is *busy* serving one customer and nobody in waiting line
State 2: Server is *busy* serving one customer and one customer in waiting line

Possible sequences of states within 10 transitions:
- 0, 1, 0, 1, 2, 3, 2, 1, 2, 1 (transitions to states with values different by 1)
Markov Chain Model as a Matrix (continued)

\[
\begin{pmatrix}
0 & 1 & 2 & \ldots & i-1 & i & i+1 & \ldots & \ldots \\
0 & -\lambda & \lambda & & & & & & \\
1 & \mu & -(\lambda+\mu) & \lambda & & & & & \\
& \ddots & & \ddots & \ddots & \ddots & \ddots & \ddots & \\
& & & & & & & \ddots & \ddots \\
Q = & i & & & & \mu & -(\lambda+\mu) & \lambda & & \\
& & & & & & & \ddots & \ddots \\
& & & & & & & \ddots & \ddots \\
& & & & & & & \ddots & \ddots \\
& & & & & & & \ddots & \ddots \\
& & & & & & & \ddots & \ddots \\
\end{pmatrix}
\]

*Q* is the *generator matrix* associated with the MC (this particular example is a *birth-and-death process*)
Continuous-time Markov Chains

Model of a continuous-time Markov chain (CTMC):
- sequence of random variables forming the chain are dependent (with dependency of next state only on current state, order 1)
- time $T$ to exit current state $i$ (negative) exponentially distributed (with parameter equal to $q_{ii}$):

$$\text{Prob}(T \leq t) = 1 - e^{q_{ii}t}$$

$$\text{Prob}(X \leq x) = 1 - e^{-\lambda x}$$

($e$: base of natural logarithm)

Tuğrul Dayar,
Dept. of Computer Engineering
Continuous-time Markov Chains (continued)

Steady-state probability distribution \( \pi \) of \( Q \) satisfies:

\[
\pi \cdot Q = 0, \quad \sum_{i} \pi_i = 1
\]

if \( Q \) corresponds to an irreducible and positive recurrent CTMC

Transient probability distribution of \( Q \) at time \( t \):

\[
\pi^{(t)} = \pi^{(0)} \cdot e^{Qt}
\]

\( e^{Qt} \): matrix exponential

What is the steady-state probability of having an empty system?
A Queueing Station with Unbounded Waiting Line (continued)

\( \rho < 1 \) (i.e., \( \lambda < \mu \)) for steady-state probability distribution to exist

Then \( \pi = (1-\rho, \rho(1-\rho), \rho^2(1-\rho), \ldots) \) Note that \( \sum_i \pi_i = 1 \Rightarrow \pi_0 = 1-\rho \)

What is the expected number of customers in the station?
\( E[N] = \sum_i i \cdot \pi_i = \rho/(1-\rho) \)

What is the expected time spent in the station?
\( E[T] = E[N] / \lambda = 1/(\mu-\lambda) \) (using Little’s law for open systems)

What is the expected time spent in the waiting line?
\( E[W] = E[T] - 1/\mu = \rho/(\mu-\lambda) \)
Web Graph and the Google Matrix


Figure 1. Directed graph representing web of six pages

\[
\begin{pmatrix}
1 & 2 & 3 & 4 & 5 & 6 \\
1 & 0 & 1/2 & 1/2 & 0 & 0 & 0 \\
2 & 0 & 0 & 0 & 0 & 0 & 0 \\
3 & 1/3 & 1/3 & 0 & 0 & 1/3 & 0 \\
4 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \\
5 & 0 & 0 & 0 & 1/2 & 0 & 1/2 \\
6 & 0 & 0 & 0 & 1 & 0 & 0 \\
\end{pmatrix}
\]
PageRank Algorithm

- Google matrix, $P$, is an extremely large, reducible and sparse substochastic matrix
- Some of its rows are zero (dangling nodes)
- In order to find the ranking of web pages, first, rows with zero sums are made to have row sums of 1:

$$P' = P + a.v$$

$u$: Column vector with all 1s

$a$: Column vector with missing row probabilities, $a = u - P.u$

$v$: Personalization probability (row) vector; generally, $v = (1/n \ 1/n \ \ldots \ 1/n)$ uniform distribution
PageRank Algorithm (continued)

- Then $P'$ is modified:
  \[ P'' = \beta P' + (1 - \beta) u.v \]

  with the convex combination ($0 < \beta < 1$) so that it becomes full, therefore:
  - irreducible
  - aperiodic
  
  $P''$ is positive recurrent, because it is irreducible and finite

- Generally, $\beta = 0.85$

- $\pi$ of $P''$ yields the ranks of web pages once it is sorted descendingly
Challenge Today

Multi-dimensional discrete-event dynamic systems are normally composed of interacting subsystems:

- Specify them as multi-dimensional MCs
- Compactly store the specification on computer
- Analyze the MC on computer

For instance, we can analyze:

- A system of stochastic chemical kinetics modeling the biological process associated with a gene expression, toggle switch, exclusive switch, metabolite synthesis with repressilator, or multiple metabolites and multiple enzymes
- Multi-class multi-server retrial queues with after-call work under different control policies
Level-dependent Quasi-birth-and-death Processes (LDQBDs)

\[
\begin{array}{ccccccc}
S_0 & S_1 & S_2 & \ldots & S_{/1} & S_l & S_{/1+1} & \ldots & \ldots \\
S_0 & Q_{0,0} & Q_{0,1} & \ & \ & \ & \ & \\
S_1 & Q_{1,0} & Q_{1,1} & Q_{1,2} & \ & \ & \ & \\
\vdots & & & & \ & \ & \ & \\
\vdots & & & & \ & \ & \ & \\
\vdots & & & & \ & \ & \ & \\
Q = S_l & & & & \ & \ & \ & \\
\vdots & & & & \ & Q_{l,-1} & Q_{l,l} & Q_{l,l+1} & \\
\vdots & & & & \ & \ & \ & \\
\vdots & & & & \ & \ & \ & \\
S_l : \text{states in level } l
\end{array}
\]

Tuğrul Dayar,
Dept. of Computer Engineering
20 November 2014
Kronecker Representations

An example
Consider the following matrices for a 3-dimensional problem (with 2, 3, and 2 states, respectively) having 4 terms of Kronecker products:

\[ Q_2^{(1)} = Q_3^{(1)} = I_2 , \quad Q_1^{(1)} = \begin{pmatrix} \mu_1 & \lambda_1 \end{pmatrix} , \quad Q_4^{(1)} = \begin{pmatrix} \mu \lambda_1 \end{pmatrix} , \]

\[ Q_1^{(2)} = Q_3^{(2)} = I_3 , \quad Q_2^{(2)} = \begin{pmatrix} \mu_2 & \lambda_2 \\ \mu_2 & \lambda_2 \end{pmatrix} , \quad Q_4^{(2)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} , \]

\[ Q_1^{(3)} = Q_2^{(3)} = I_2 , \quad Q_3^{(3)} = \begin{pmatrix} \mu_3 & \lambda_3 \end{pmatrix} , \quad Q_4^{(3)} = \begin{pmatrix} 1 \\ 1 \end{pmatrix} . \]

Then,

\[ Q = \sum_{k=1}^{4} \otimes_{k=1}^{3} Q_{k}^{(h)} + Q_D. \]
Kronecker Representations (continued)

An example (continued)

\[
Q = \begin{pmatrix}
0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
0 & 0 & 1 & 1 & 2 & 2 & 0 & 0 & 1 & 1 & 2 & 2 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & \mu_3 & \lambda_2 & \lambda_1 \\
0 & 0 & 1 & \mu_2 & \lambda_1 & \lambda_1 \\
0 & 1 & 0 & \mu_2 & \lambda_2 & \lambda_1 & \lambda_1 \\
0 & 1 & 1 & \mu_2 & \lambda_2 & \lambda_1 \\
0 & 2 & 0 & \mu_2 & \lambda_2 & \lambda_2 \\
0 & 2 & 1 & \mu_2 & \lambda_2 & \lambda_2 \\
0 & 2 & 1 & \mu_2 & \lambda_2 & \lambda_2 \\
1 & 0 & 0 & \mu_1 & \lambda_3 & \lambda_2 & \lambda_2 \\
1 & 0 & 1 & \mu_1 & \lambda_3 & \lambda_2 & \lambda_2 \\
1 & 1 & 0 & \mu_1 & \lambda_3 & \lambda_2 & \lambda_2 \\
1 & 1 & 1 & \mu_1 & \lambda_3 & \lambda_2 & \lambda_2 \\
1 & 2 & 0 & \mu_1 & \lambda_3 & \lambda_2 & \lambda_2 \\
1 & 2 & 1 & \mu_1 & \lambda_3 & \lambda_2 & \lambda_2 \\
\end{pmatrix}
\]

- # of floating-point values stored in Kronecker representation is 11 for matrices and 12 for diagonal, thus totaling 23; whereas, it is 53 for flat representation.
- Discrepancy between Kronecker and flat representations becomes substantial for larger values of the state space size, \( n \).
Thank You

Questions