Lecture 10: Markov Chains

Ziyu Shao

School of Information Science and Technology
ShanghaiTech University

January 02, 2024

Outline

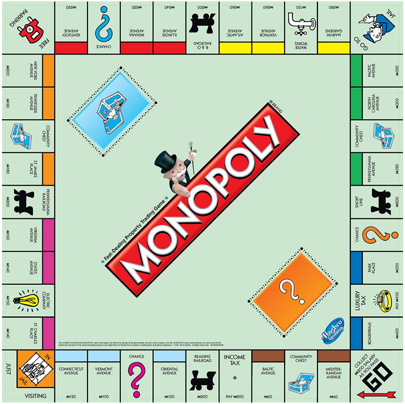
Outline

Definition

- A stochastic process is a collection of random variables $\{X_t, t \in I\}$. The set I is the *index set* of the process. The random variables are defined on a common state space $\mathcal{S}$.
- I is discrete: discrete-time stochastic processes (sequences of random variables)
- I is continuous: continuous-time stochastic processes (uncountable collections of random variables)

Example: Discrete Time & Discrete State Space

- State Space: $\{1, \ldots, 40\}$
- X_k: the player's board position after k dice rollings.
- Stochastic Process for Monopoly: $X_0, X_1, \ldots$



Example: Discrete Time & Continuous State Space

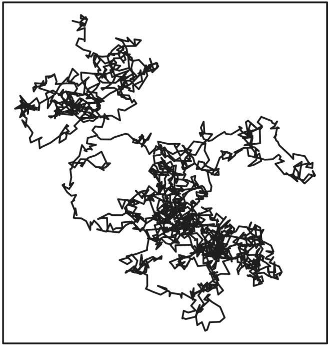
- Air-monitoring with PM2.5 measurements every hour
- State Space:$(0, 2000)$
- X_k: the PM2.5 measurement at the kth hour.
- Stochastic Process for Air-monitoring: $X_0, X_1, \ldots$

Example: Continuous Time & Discrete State Space

- We receive emails at random times day and night.
- State Space:$\{0, 1, 2, \dots\}$
- $X_t, t \in [0, \infty)$: the number of emails we receive up to time t
- Stochastic Process for Email: $\{X_t\}$

Example: Continuous Time & Continuous State Space

- Two-dimensional Brownian Motion
- State Space: $\mathbb{R}^2$
- $X_t, t \in [0, \infty)$: position of the particle at time t
- Stochastic Process for random motion of particles: $\{X_t\}$



Outline

Model Selection in Stochastic Modeling

- Enough complexity to capture the complexity of the phenomena in question
- Enough structure and simplicity to allow one to compute things of interest

Motivation

- Introduced by Andrey Markov in 1906
- IID sequence of random variables: too restrictive assumption
- Completely dependent among random variables: hard to analysis
- Markov chain: happy medium between complete independence & complete dependence.

Markov Model

Three basic components of Markov model

- A sequence of random variables $\{X_t, t \in \mathcal{T}\}$, where $\mathcal{T}$ is an index set, usually called "**time**".
- All possible sample values of $\{X_t, t \in \mathcal{T}\}$ are called "**states**", which are elements of a sate space $\mathcal{S}$.
- "**Markov property**": given the present value(information) of the process, the future evolution of the process is independent of the past evolution of the process.

Classification of Markov Model

- *Discrete-Time Markov Chain*: Discrete $\mathcal{S}$ & Discrete $\mathcal{T}$
- *Continuous-Time Markov Chain*: Discrete $\mathcal{S}$ & Continuous $\mathcal{T}$
- *Discrete Markov Process*: Continuous $\mathcal{S}$ & Discrete $\mathcal{T}$
- *Continuous Markov Process*: Continuous $\mathcal{S}$ & Continuous $\mathcal{T}$

Our focus: Discrete-Time Markov Chain with finite state space

Outline

Markov Chain

Definition

A sequence of random variables $X_0, X_1, X_2, \ldots$ taking values in the *state space* $\{1, 2, \ldots, M\}$ is called a *Markov chain* if for all $n \geq 0$,

$$P\left(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\right) = P\left(X_{n+1} = j | X_n = i\right).$$

Time-homogeneous Markov Chains

Definition

Given a Markov chain $X_0, X_1, X_2, \ldots$ It is called time-homogeneous Markov chain if for all $n \geq 0$,

$$P(X_{n+1} = j | X_n = i) = q_{i,j}.$$

where $q_{i,j}$ is a constant independent of n.

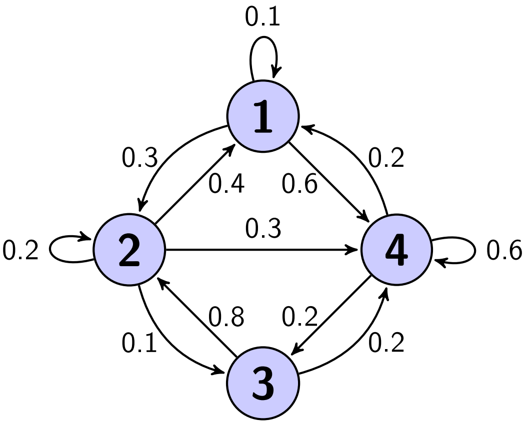
From now on, we focus on time-homogeneous Markov Chains, and we call it Markov chain in brevity.

Transition Matrix

Definition

Let $X_0, X_1, X_2, \ldots$ be a Markov chain with state space $\{1, 2, \ldots, M\}$, and let $q_{i,j} = P(X_{n+1} = j | X_n = i)$ be the transition probability from state i to state j. The $M \times M$ matrix $Q = (q_{i,j})$ is called the *transition matrix* of the chain.

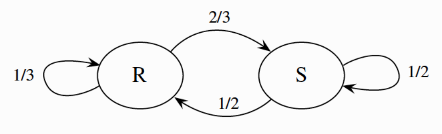
Graphical and Matrix Form of Markov Chain



$$Q = \begin{bmatrix} 0.1 & 0.3 & 0 & 0.6 \\ 0.4 & 0.2 & 0.1 & 0.3 \\ 0 & 0.8 & 0 & 0.2 \\ 0.2 & 0 & 0.2 & 0.6 \end{bmatrix}$$

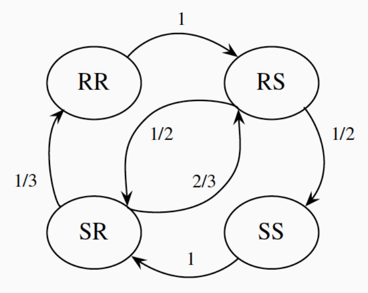
Row sum $= 1$ in Transition Matrix.

Example: Rainy-sunny Markov Chain



$$\begin{array}{cc} & \begin{array}{cc} R & S \end{array} \\ \begin{array}{c} R \\ S \end{array} & \begin{bmatrix} \frac{1}{3} & \frac{2}{3} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \end{array}$$

Example: Rainy-sunny Markov Chain

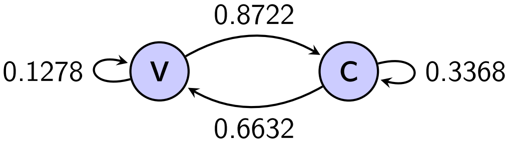


$$
\begin{array}{c}
 \\ RR \\ RS \\ SR \\ SS
\end{array}
\begin{array}{c}
\begin{array}{cccc} RR & RS & SR & SS \end{array} \\
\begin{bmatrix}
0 & 1 & 0 & 0 \\
0 & 0 & 1/2 & 1/2 \\
1/3 & 2/3 & 0 & 0 \\
0 & 0 & 1 & 0
\end{bmatrix}
\end{array}
$$

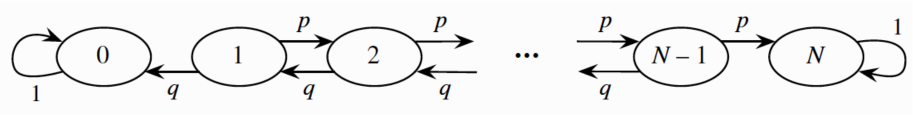
Example: The First Markov Chain in History

- Andrey Andreyevich Markov was interested in investigating the way the vowels and consonants alternate in Russian literature, e.g., "Eugene Onegin" by Pushkin
- He classified 20,000 consecutive characters: 8638 vowels & 11362 consonants

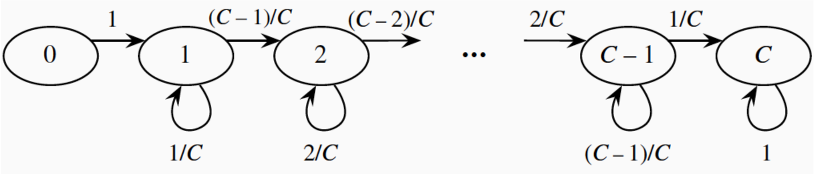
$$
\begin{array}{cc}
 & \begin{array}{cc} \text{vowel} & \text{consonant} \end{array} \\
\begin{array}{c} \text{vowel} \\ \text{consonant} \end{array} & \begin{bmatrix} 1104/8638 & 7534/8638 \\ 7535/11362 & 3827/11362 \end{bmatrix}
\end{array}
= \begin{bmatrix} 0.1278 & 0.8722 \\ 0.6632 & 0.3368 \end{bmatrix}
$$



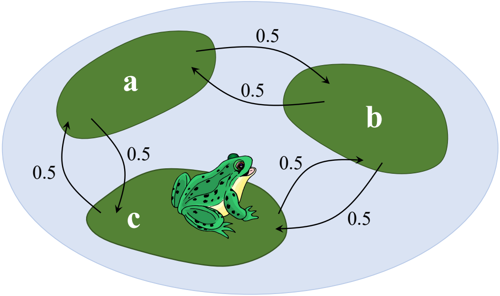
Gambler's Ruin As A Markov Chain



Coupon Collector As A Markov Chain



Example: Random Walk on A Graph



$$
\begin{array}{c c}
 & \begin{array}{ccc} a & b & c \end{array} \\
\begin{array}{c} a \\ b \\ c \end{array} & \begin{bmatrix} 0 & 0.5 & 0.5 \\ 0.5 & 0 & 0.5 \\ 0.5 & 0.5 & 0 \end{bmatrix}
\end{array}
$$

Outline

n-step Transition Probability

Definition

Let $X_0, X_1, X_2, \dots$ be a Markov chain with transition matrix Q. The n-step transition probability from i to j is the probability of being at j exactly n steps after being at i. We denote this by $q_{i,j}^{(n)}$:

$$q_{i,j}^{(n)} = P\left(X_n = j | X_0 = i\right).$$

Example: 2-step Transition Probability

$$q_{i,j}^{(2)} = P(X_2 = j | X_0 = i) = \sum_k q_{i,k} q_{k,j} = \text{(i,j) entry of } Q^2.$$

Chapter-Kolmogorov Relationship

$$q_{i,j}^{(m+n)} = P(X_{m+n} = j | X_0 = i) = \sum_k q_{i,k}^{(m)} q_{k,j}^{(n)} = \text{(i,j) entry of } Q^{m+n}.$$

Proof

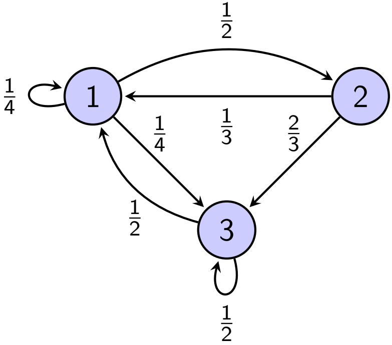
Distribution of X_n

Let X_0, X_1,... be a Markov chain with transition matrix Q and initial distribution $\boldsymbol{\alpha}$, where $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \cdots, \alpha_M)$, $\alpha_i = P(X_0 = i), i = 1, \ldots, M$. For all n ≥ 0, the distribution of X_n is $\boldsymbol{\alpha} Q^n$. That is, the jth component of $\boldsymbol{\alpha} Q^n$ is $P(X_n = j)$, denoted as:

$$P(X_n = j) = (\alpha Q^n)_j, \ \textit{for all } j.$$

Example

Given a Markov chain $X_0, X_1, X_2, \ldots$ with state space $\mathcal{S} = \{1, 2, 3\}$.

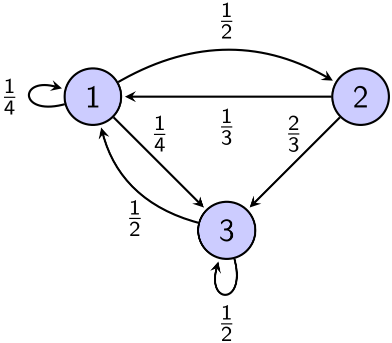


$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \end{array}$$

- Find $P(X_3 = 1 | X_2 = 1)$ and $P(X_4 = 3 | X_3 = 2)$.

Example

Given a Markov chain $X_0, X_1, X_2, \dots$ with state space $\mathcal{S} = \{1, 2, 3\}$.

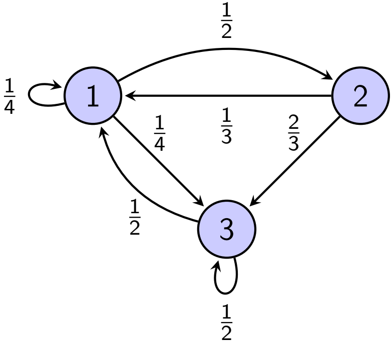


$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \end{array}$$

- If $P(X_0 = 1) = \frac{1}{3}$, find $P(X_0 = 1, X_1 = 2, X_2 = 3)$.

Example

Given a Markov chain $X_0, X_1, X_2, \ldots$ with state space $\mathcal{S} = \{1, 2, 3\}$.

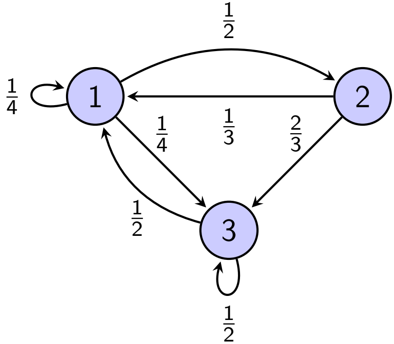


$$
\begin{array}{c c}
 & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\
\begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix}
\end{array}
$$

- Find $P(X_2 = 1 | X_0 = 1)$, $P(X_2 = 2 | X_0 = 1)$, and $P(X_2 = 3 | X_0 = 1)$.

Example

Given a Markov chain $X_0, X_1, X_2, \ldots$ with state space $\mathcal{S} = \{1, 2, 3\}$.



$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \end{array}$$

- Find $E(X_2 | X_0 = 1)$.

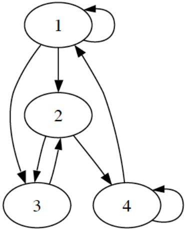
Outline

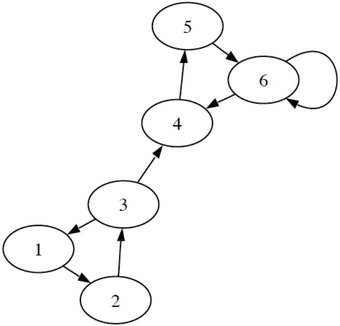
Recurrent and Transient States

Definition

State i of a Markov chain is **recurrent** if starting from i, the probability is 1 that the chain will eventually return to i. Otherwise, the state is **transient**, which means that if the chain starts from i, there is a positive probability of never returning to i.

Example





Irreducible and Reducible Chain

Definition

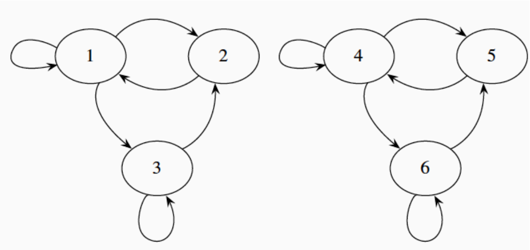
A Markov chain with transition matrix Q is *irreducible* if for any two states i and j, it is possible to go from i to j in a finite number of steps (with positive probability). That is, for any states i,j there is some positive integer n such that the (i, j) entry of Q^n is positive. A Markov chain that is not irreducible is called *reducible*.

Irreducible Implies All States Recurrent

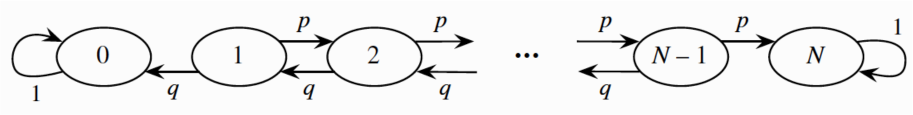
Theorem

In an irreducible Markov chain with a finite state space, all states are recurrent.

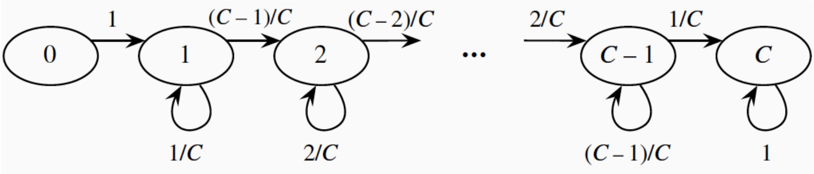
A Reducible Markov Chain with Recurrent States



Gambler's Ruin As A Markov Chain



Coupon Collector As A Markov Chain



Period

Definition

For a Markov chain with transition matrix Q, the *period* of state i, denoted *$d(i)$*, is the greatest common divisor of the set of possible return times to i. That is,

$$d(i) = gcd\{n > 0 : Q^n_{i,i} > 0\}.$$

If $d(i) = 1$, state i is said to be *aperiodic*. If the set of return times is empty, set $d(i) = +\infty$.

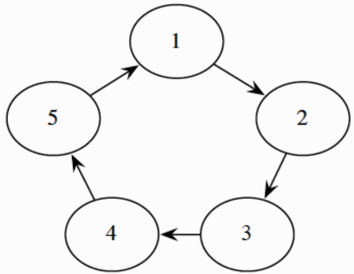
Periodic, Aperiodic Markov Chain

Definition

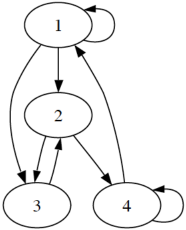
A Markov chain is called *periodic* if it is irreducible and all states have period greater than 1.

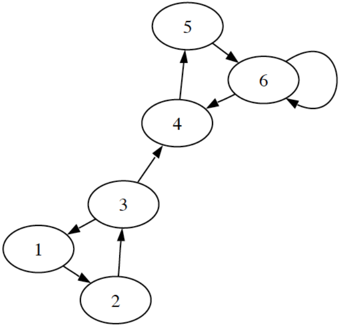
A Markov chain is called *aperiodic* if it is irreducible and all states have period equal to 1.

Example: Periodic Chain



Example





Outline

Definition

Definition

A row vector $\mathbf{s} = (s_1, \ldots, s_M)$ such that $s_i \geq 0$ and $\sum_i s_i = 1$ is a *stationary distribution* for a Markov chain with transition matrix Q if

$$\sum_i s_i q_{i,j} = s_j.$$

for all j, or equivalently,

$$\mathbf{s}Q = \mathbf{s}.$$

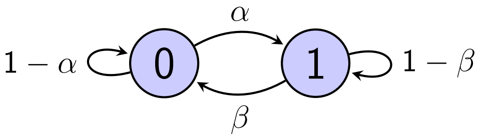
Example: Double Stochastic Matrix

$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & 0 & \frac{1}{2} \end{bmatrix} \end{array}$$

Theorem

If each column of the transition matrix Q sums to 1, then the uniform distribution over all states, $(1/M, 1/M, \ldots, 1/M)$, is a stationary distribution. (A nonnegative matrix such that the row sums and the column sums are all equal to 1 is called a doubly stochastic matrix.)

Example: Two-State Markov Chain



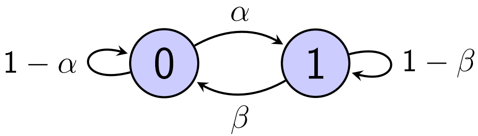
Theorem on Stationary Distribution

Theorem

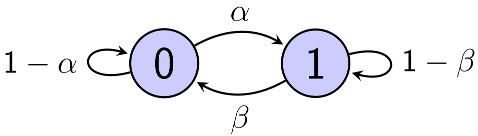
Given a Markov chain with finite state space.

- *If such Markov chain is irreducible, then it has a unique stationary distribution. In this distribution, every state has positive probability.*
- *If such Markov chain is irreducible and aperiodic, with stationary distribution* $\mathbf{s}$ *and transition matrix* Q*, then* $P(X_n = i)$ *converges to* s_i *as* $n \to \infty$*. In terms of the transition matrix,* Q^n *converges to a matrix in which each row is* $\mathbf{s}$*.*

Example



Example



Outline

Reversibility

Definition

Let $Q = (q_{i,j})$ be the transition matrix of a Markov chain. Suppose there is $\mathbf{s} = (s_1, \dots, s_M)$ with $s_i \geq 0$, $\sum_i s_i = 1$, such that

$$s_i q_{i,j} = s_j q_{j,i}$$

for all states i and j. This equation is called the *reversibility* or *detailed balance* condition, and we say that the chain is *reversible* with respect to $\mathbf{s}$ if it holds.

Reversible implies Stationary

Theorem

Suppose that $Q = (q_{i,j})$ *is a transition matrix of a Markov chain that is reversible with respect to a nonnegative vector* $\mathbf{s} = (s_1, \ldots, s_M)$ *whose components sum to* 1*. Then* $\mathbf{s}$ *is a stationary distribution of the chain.*

Check the Detailed Balance Equation

Theorem

If for an irreducible Markov chain with transition matrix $Q = (q_{i,j})$, there exists a probability solution π to the detailed balance equations

$$\pi_i q_{i,j} = \pi_j q_{j,i}$$

for all pairs of states i, j, then this Markov chain is reversible and the solution π is the unique stationary distribution.

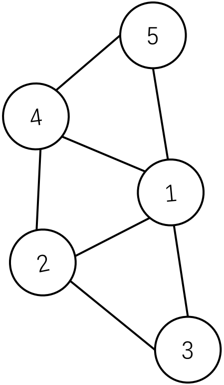
Example: Symmetric Transition Matrix

$$\begin{array}{cc} & \begin{array}{ccc} 1 & 2 & 3 \end{array} \\ \begin{array}{c} 1 \\ 2 \\ 3 \end{array} & \begin{bmatrix} \frac{1}{4} & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{4} & \frac{1}{4} & \frac{1}{2} \end{bmatrix} \end{array}$$

Theorem

If the transition matrix Q for an irreducible Markov chain is symmetric, then the uniform distribution over all states, $(1/M, 1/M, \ldots, 1/M)$, is the unique stationary distribution.

Example: Random Walk on Undirected Graph



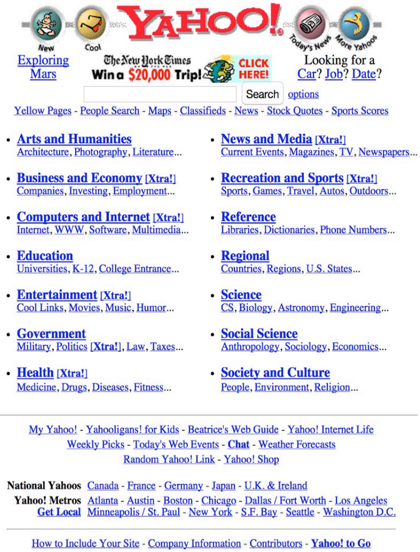
Example: Random Walk on Undirected Graph

Outline

How to Organize the Web?

- First Try: Web Directories
- Yahoo, DMOZ, LookSmart

Yahoo



YAHOO!
New
Cool
Today's News
More Yahoos
Exploring Mars
The New York Times
Win a $20,000 Trip!
CLICK HERE!
Looking for a Car? Job? Date?
Search
options
Yellow Pages - People Search - Maps - Classifieds - News - Stock Quotes - Sports Scores
Arts and Humanities
Architecture, Photography, Literature...
News and Media [Xtra!]
Current Events, Magazines, TV, Newspapers...
Business and Economy [Xtra!]
Companies, Investing, Employment...
Recreation and Sports [Xtra!]
Sports, Games, Travel, Autos, Outdoors...
Computers and Internet [Xtra!]
Internet, WWW, Software, Multimedia...
Reference
Libraries, Dictionaries, Phone Numbers...
Education
Universities, K-12, College Entrance...
Regional
Countries, Regions, U.S. States...
Entertainment [Xtra!]
Cool Links, Movies, Music, Humor...
Science
CS, Biology, Astronomy, Engineering...
Government
Military, Politics [Xtra!], Law, Taxes...
Social Science
Anthropology, Sociology, Economics...
Health [Xtra!]
Medicine, Drugs, Diseases, Fitness...
Society and Culture
People, Environment, Religion...
My Yahoo! - Yahooligans! for Kids - Beatrice's Web Guide - Yahoo! Internet Life
Weekly Picks - Today's Web Events - Chat - Weather Forecasts
Random Yahoo! Link - Yahoo! Shop
National Yahoos Canada - France - Germany - Japan - U.K. & Ireland
Yahoo! Metros Atlanta - Austin - Boston - Chicago - Dallas / Fort Worth - Los Angeles
Get Local Minneapolis / St. Paul - New York - S.F. Bay - Seattle - Washington D.C.
How to Include Your Site - Company Information - Contributors - Yahoo! to Go

How to Organize the Web?

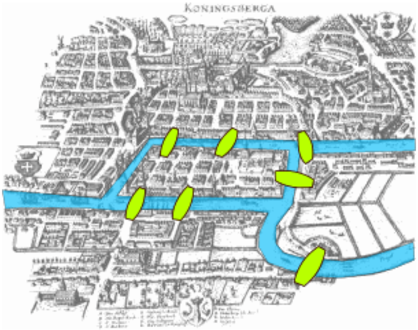
- Second Try: Web Search
- Information Retrieval:
  - find relevant docs in a small and trusted set
  - newspaper articles, patents, etc
- Hardness: web is huge, full of untrusted documents, random things, web spam, etc.

Challenges for Web Search

- Web contains many sources of information. Who to trust?
  - **Trick**: Trustworthy pages may point to each other!
- What is the best answer to query keywords?
  - Webpages are not equally important (www.nothing.com vs.www.stanford.edu)
  - **Trick**: rank pages containing keywords according to their importances (popularity)
  - Find the page with the highest rank
  - How to rank?

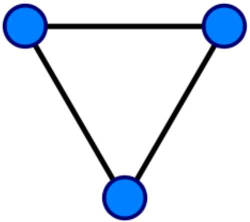
Modeling Language: Graph Theory

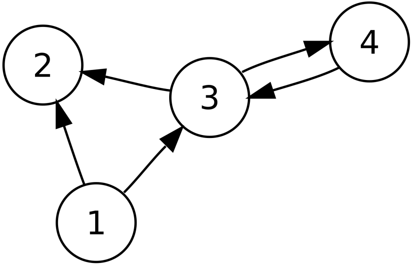
- Origin: 1735 Euler for Seven Bridges of Königsberg



Key Elements of A Graph

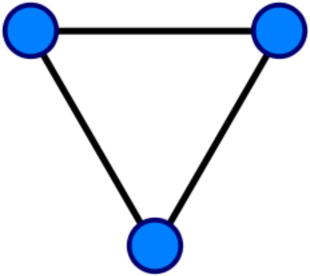
- A graph is an ordered pair $G = (V, E)$
- V: a set of vertices or nodes
- E: a set of edges or links between nodes
- Edge: undirected/directed





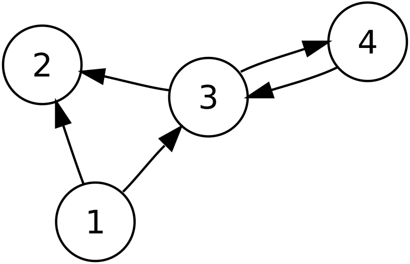
Undirected Graph

- Degree of vertex v: metric for connectivity of vertex v.
- $deg(v)$: the number of edges with v as an end vertex



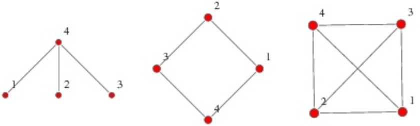
Directed Graph

- Indegree of vertex v: the number of incoming edges ends at v.
- Outdegree of vertex v: the number of outgoing edges starting from v.
- $I(v)$: indegree of v
- $O(v)$: outdegree of v



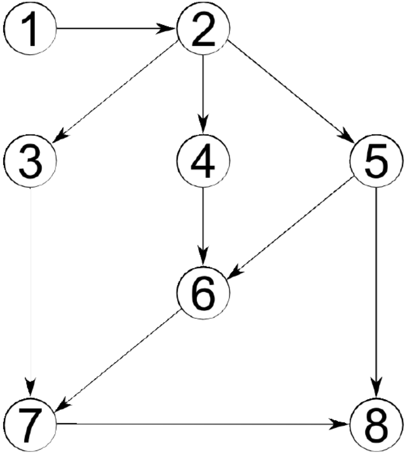
Adjacency Matrix

- A square $(0,1)-$matrix to represent a finite graph
- Matrix elements: pairs of vertices are adjacent or not
- Symmetric matrix: for undirected graph



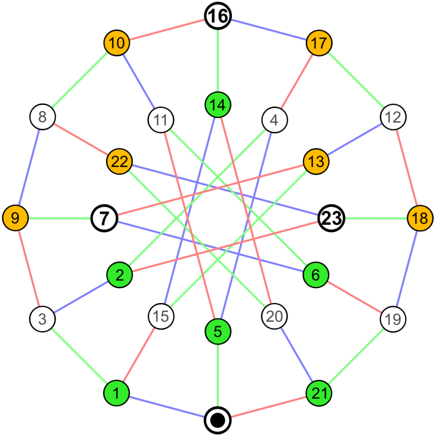
$$\begin{pmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \end{pmatrix} \quad \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

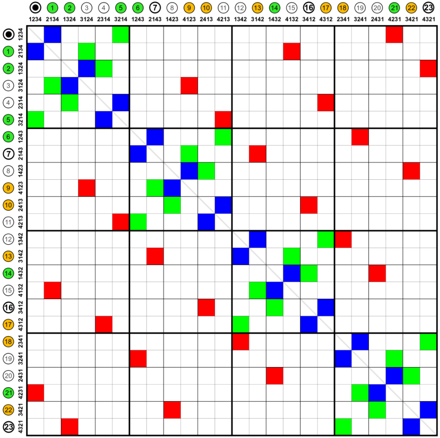
Adjacency Matrix: Directed Graph



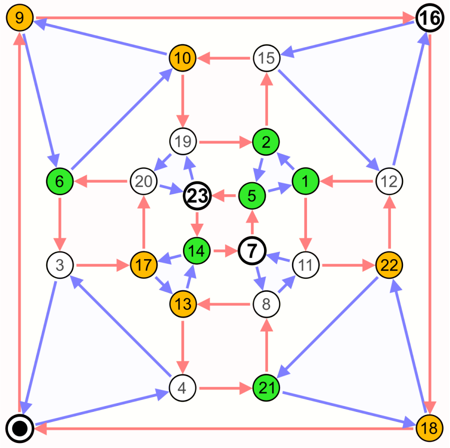
	1	2	3	4	5	6	7	8
1		1						
2			1	1	1			
3							1	
4						1		
5						1		1
6							1	
7								1
8								

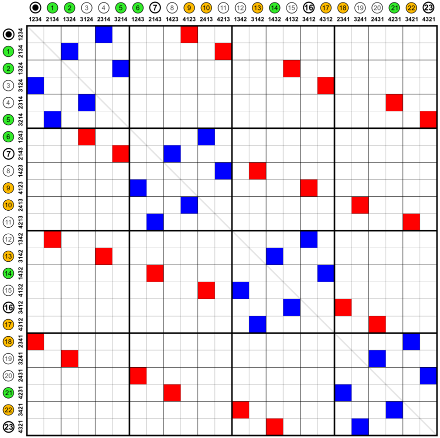
Adjacency Matrix of Nauru Graph





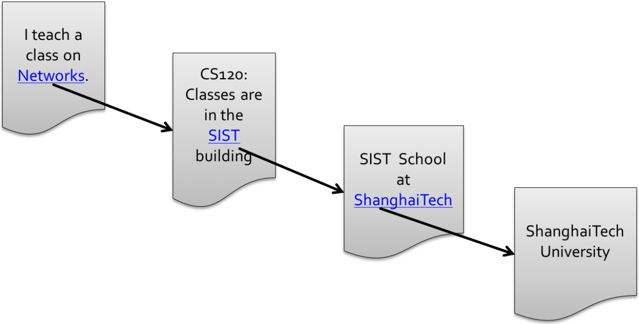
Adjacency Matrix of Cayley Graph



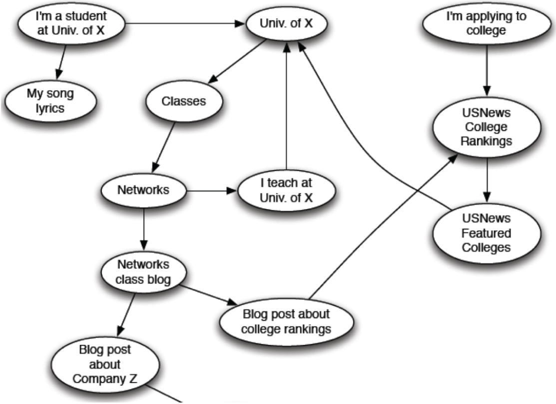


World Wide Web as A Graph

- Web as a directed graph
- Nodes: webpages
- Edges: hyperlinks

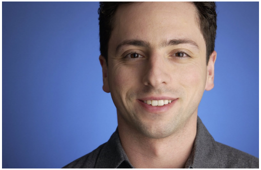


Web as A Directed Graph



Milestones in Networking

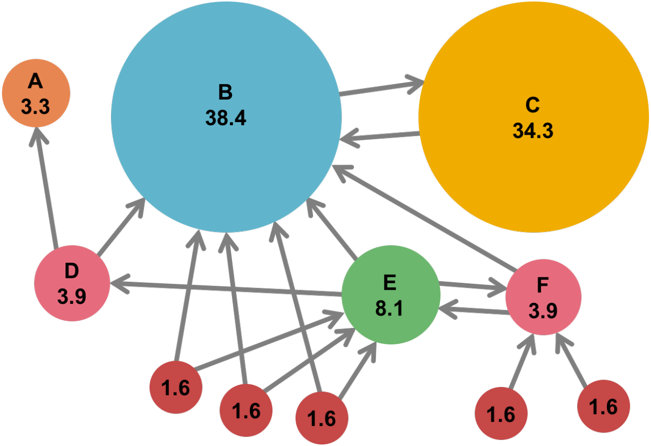
- 1998: Larry Page (1973-) and Sergey Brin (1973-) invented PageRank algorithm and then founded Google.
- 1998: Jon Kleinberg (1971-) invented Hyperlink-Induced Topic Search (HITS) algorithm.



Links as Votes

- Page is more important if it has more links
- Incoming links or outgoing links?
- Think of incoming links as votes:
  - www.stanford.edu has 23400 incoming links
  - www.nothing.com has 1 incoming link
- Are all in-links are equal?
  - Links from important pages count more
  - Recursive question!

Example: PageRank Scores

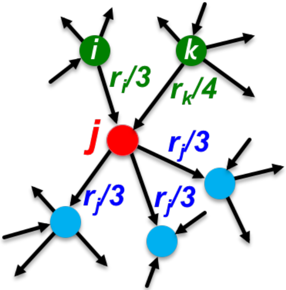


Simple Recursive Formulation

- Each link's vote is proportional to the importance of its source page.
- If page j with importance r_j has n out-links, each link gets r_j/n votes
- Page j's own importance is the sum of the votes on its in-links

Example

- $r_j = r_i/3 + r_k/4.$



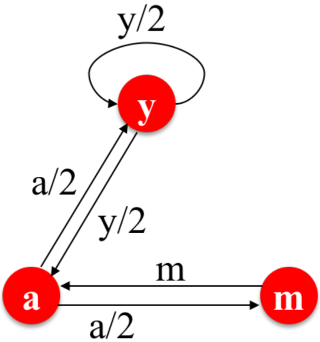
PageRank: The “Flow” Model

- A “vote” from an important page is worth more
- A page is important if it is pointed to by other important pages
- Define a rank r_j for page j:

$$r_j = \sum_{i \to j} \frac{r_i}{O_i}$$

where O_i is the outdegree of i.

Example: Flow Equation



$$r_y = r_y/2 + r_a/2$$
$$r_a = r_y/2 + r_m$$
$$r_m = r_a/2$$

Solving the Flow Equations

- Additional constraint forces uniqueness:
  - $r_y + r_a + r_m = 1$.
  - solution: $r_y = 2/5, r_a = 2/5, r_m = 1/5$.
- Gaussian elimination method works for small examples
- We need a better method for large web-size graphs

PageRank: Matrix Formulation

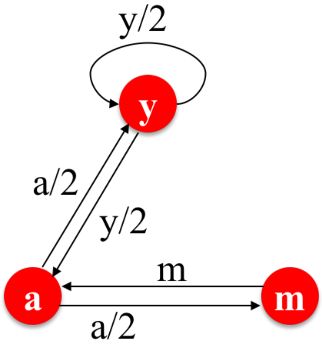
- Adjacency matrix Q
  - Each page i has O_i out-links
  - If $i \to j$, then $Q_{i,j} = \frac{1}{O_i}$, else $Q_{i,j} = 0$.
- Q is a stochastic matrix
- Row sum to 1.

PageRank: Matrix Formulation

- Rank vector $\mathbf{r}$
  - Vector with an entry per page
  - r_i is the importance score of page i
  - $\sum_i r_i = 1$
- The flow equations can be written

$$\mathbf{r} = \mathbf{r} \cdot Q$$

Example



$$r_y = r_y/2 + r_a/2$$
$$r_a = r_y/2 + r_m$$
$$r_m = r_a/2$$

$$Q = \begin{array}{c c} & \begin{array}{ccc} y & a & m \end{array} \\ \begin{array}{c} y \\ a \\ m \end{array} & \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 1 & 0 \end{pmatrix} \end{array}$$

Random Walk Interpretation

- Random Walk on Directed Graphs
- $\mathbf{r} \cdot Q = \mathbf{r}$
- $\mathbf{r}$: stationary distribution

Power Iteration Method

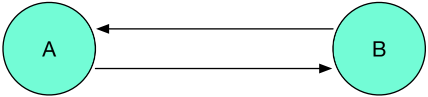
- Given a web graph with n nodes, where the nodes are pages and edges are hyperlinks.
- Power iteration: a simple iterative scheme
  - Suppose there are N web pages
  - Initialize: $\mathbf{r}(0) = [\frac{1}{N}, \ldots, \frac{1}{N}]$.
  - Iterate: $\mathbf{r}(t+1) = \mathbf{r}(t) \cdot Q$.
  - Stop when $|\mathbf{r}(t+1) - \mathbf{r}(t)|_1 < \epsilon$

The Google Formulation

$$r_j^{(t+1)} = \sum_{i \to j} \frac{r_i^{(t)}}{O_i} \implies \mathbf{r}^{(t+1)} = \mathbf{r}^{(t)} \cdot Q$$

$$\begin{cases} \text{Does this converge?} \\ \text{Converge to what we want?} \\ \text{Result reasonable?} \end{cases}$$

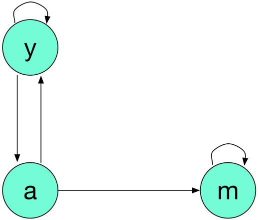
Example: Spider Traps with Two Nodes



$$r_A^{(t+1)} = r_B^{(t)}$$
$$r_B^{(t+1)} = r_A^{(t)}$$

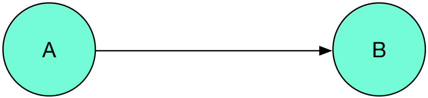
$$\begin{array}{cccccccc} r_A & \rightarrow & 1 & 0 & 1 & 0 & \ldots \\ r_B & \rightarrow & 0 & 1 & 0 & 1 & \ldots \end{array}$$

Example: Spider Traps with One Node



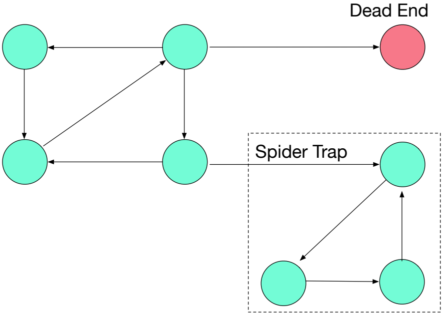
$$
\begin{array}{cccc}
 & y & a & m \\
y & 1/2 & 1/2 & 0 \\
a & 1/2 & 0 & 1/2 \\
m & 0 & 0 & 1
\end{array}
\implies
\begin{array}{cccccc}
r_y & \rightarrow & 1/3 & 2/6 & 3/12 & \ldots & 0 \\
r_a & \rightarrow & 1/3 & 1/6 & 2/12 & \ldots & 0 \\
r_m & \rightarrow & 1/3 & 3/6 & 7/12 & \ldots & 1
\end{array}
$$

Example: Dead End



$$\begin{array}{lcccccc} r_A & \rightarrow & 1 & 0 & 0 & 0 & \ldots \\ r_B & \rightarrow & 0 & 1 & 0 & 0 & \ldots \end{array}$$

Observations



- Some pages are dead ends (no out-links)
- Spider traps (all out-links are within the group)

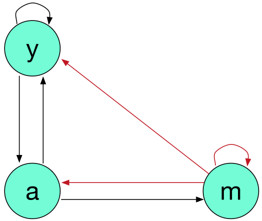
Google's Solution

- Idea: irreducibility leads to unique stationary distribution
- Random Teleports: create virtual links between any two pages
- Given the WWW graph $G = (V, E)$ and $N = |V|$.
- At each time, walker at page i has the following operations:
- If $O_i = 0$ (dead-end), then select any page j with equal probability $1/N$.
- Otherwise, walker has two options

$$\begin{cases} \text{w.p. } \beta & \text{Follow an out-link at random } \frac{1}{O_i} \\ \text{w.p. } 1-\beta & \text{Jump to some random pages} \end{cases} \tag{1}$$

- $\beta = (0.8, 0.9)$

Example: Google's Solution for Spider Traps



$$\begin{array}{c|ccc} & y & a & m \\ y & 1/2 & 1/2 & \mathbf{0} \\ a & 1/2 & 0 & \mathbf{1/2} \\ m & 0 & 0 & \mathbf{1} \end{array} \Longrightarrow \begin{array}{c|ccc} & y & a & m \\ y & 1/2 & 1/2 & 0 \\ a & 1/2 & 0 & 1/2 \\ m & \color{red}{1/3} & \color{red}{1/3} & \color{red}{1/3} \end{array}$$

General Solution

- Assume no Dead Ends
- Pagerank Equation:

$$r_j = \sum_{i \to j} \beta \cdot \frac{r_i}{O_i} + (1 - \beta) \cdot \frac{1}{N}$$

- Google Matrix

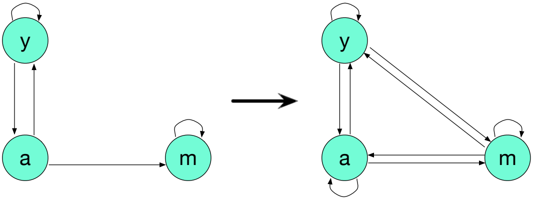
$$G = \beta \cdot Q + (1 - \beta) \left[\frac{1}{N} \right]_{N \times N}$$

$$\mathbf{r} = \mathbf{r} \cdot G$$

Random Teleports

- Google Matrix

$$Q \to G = \beta \cdot Q + (1-\beta)\left[\frac{1}{N}\right]_{N\times N}$$



Do Some Calculation

Let $\beta = 0.8$.

$$Q = 0.8 \cdot \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 0 & 1/2 \\ 0 & 0 & 1 \end{pmatrix} + 0.2 \cdot \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix}$$

As a result,

$$G = \begin{pmatrix} 7/15 & 7/15 & 1/15 \\ 7/15 & 1/15 & 7/15 \\ 1/15 & 1/15 & 13/15 \end{pmatrix}$$

Implementation of PageRank in Practice

- BigTable: distributed storage system
- GFS (Google File System): distributed file system
- Mapreduce: distributed computing system (followed by Hadoop & Spark)

Outline

Markov Chain Monte Carlo (MCMC)

- Revolutionized statistics and scientific computation
- Expanding the range of possible distributions that we can simulate from, including joint distributions in high dimensions
- Basic idea: build your own Markov chain $(X_0, X_1, \ldots)$ so that the desired distribution π is the stationary distribution of the chain.
- Sampling from distribution π (running the chain for a long time and then sampling)
- Further do sample mean & sample variance & other sample functions

MCMC Method

- Forward engineering: given the transition matrix Q, find the stationary distribution of Markov chain.
- Reverse engineering: given a distribution π that we want to simulate, we will engineer a Markov chain whose stationary distribution is π. Then run this engineered Markov chain for a long time, the distribution of the chain will approach π.

MCMC Method

- Markov Chain Monte Carlo (MCMC) is a remarkable methodology, which utilizes Markov sequences to effectively simulate from what would otherwise be intractable distributions.
- All MCMC algorithms construct reversible (time-reversible) Markov chain: detailed balance equations help us.
- Two most widely used algorithms: Metropolis-Hastings & Gibbs Sampling

Theory Justification: Strong Law of Large Numbers for Markov Chains

Theorem

Assume that $X_0, X_1, \cdots$ is an irreducible and aperiodic Markov chain with stationary distribution π. Let g be a bounded, real-valued function. Let X be a random variable with distribution π. Then, with probability 1,

$$\lim_{n\to\infty} \frac{g(X_1) + \cdots + g(X_n)}{n} = E(g(X)) = \sum_j \pi_j g(j).$$

Basic Idea of Metropolis–Hastings Algorithm

- Proposed by Nicholas Metropolis in 1953 & further developed by Wilfred Keith Hastings in 1970.
- Start with any irreducible Markov chain on the state space of interest
- Then modify it into a new Markov chain with desired stationary distribution
- Modification: moves are proposed according to the original chain, but the proposal may or may not be accepted.
- Art: choice of the probability of accepting the proposal

Algorithm 1 Metropolis-Hastings

Require:

Stationary distribution $\pi = (\pi_1, ..., \pi_M)$;
Original transition matrix $P = (p_{i,j})$;
State X_0 (chosen randomly or deterministically);

Ensure:

Modified transition matrix $P' = (p'_{i,j})$;

```
repeat
    If X_n = i, propose a new state j using the transition probabilities in the ith row of the original transition matrix P;
    Compute the acceptance probability a_{i,j} = min(π_j p_{j,i} / (π_i p_{i,j}), 1);
    Flip a coin that lands Heads with probability a_{i,j};
    If the coin lands Heads, accept the proposal (i.e., go to j), setting X_{n+1} = j. Otherwise, reject the proposal (i.e., stay at i), setting X_{n+1} = i;
until Convergence;
return P';
```

Basic Idea of Gibbs Sampling

- At each stage, one variable is updated (keeping all the other variables fixed) by drawing from the conditional distribution of that variable given all the other variables.
- Two major kinds of Gibbs sampler:
  - systematic scan: the updates sweep through the components in a deterministic order.
  - random scan: a randomly chosen component is updated at each stage.

Algorithm: Systematic Scan Gibbs Sampler

Let X and Y be discrete r.v.s with joint PMF $p_{X,Y}(x, y) = P(X = x, Y = y)$. We wish to construct a two-dimensional Markov chain (X_n, Y_n) whose stationary distribution is $p_{X,Y}$. The systematic scan Gibbs sampler proceeds by updating the X-component and the Y-component in alternation. If the current state is $(X_n, Y_n) = (x_n, y_n)$, then we update the X-component while holding the Y-component fixed, and then update the Y-component while holding the X-component fixed.

Algorithm 2 Systematic Scan Gibbs Sampler

Require:

Joint PMF $p_{X,Y}$;

Initial state (X_0, Y_0);

Ensure:

Two-dimensional Markov chain (X_n, Y_n);

```
repeat
    Draw a value x_{n+1} from the conditional distribution of X given
    Y = y_n, i.e. P(X|Y = y_n), and set X_{n+1} = x_{n+1};
    Draw a value y_{n+1} from the conditional distribution of Y given
    X = x_{n+1}, i.e. P(Y|X = x_{n+1}), and set Y_{n+1} = y_{n+1};
    return (X_{n+1}, Y_{n+1});
until n ≥ N;
```

Algorithm: Random Scan Gibbs Sampler

As above, let X and Y be discrete r.v.s with joint PMF $p_{X,Y}(x, y)$. We wish to construct a two-dimensional Markov chain (X_n, Y_n) whose stationary distribution is $p_{X,Y}$. Each move of the random scan Gibbs sampler picks a uniformly random component and updates it, according to the conditional distribution given the other component.

Algorithm 3 Random scan Gibbs sampler

Require:

Joint PMF $p_{X,Y}$;

Initial state (X_0, Y_0);

Ensure:

Two-dimensional Markov chain (X_n, Y_n);

repeat

  Choose which component to update, with equal probabilities;

  If the X-component was chosen, draw a value x_{n+1} from the conditional distribution of X given $Y = y_n$, and set $X_{n+1} = x_{n+1}, Y_{n+1} = y_n$. Similarly, if the Y-component was chosen, draw a value y_{n+1} from the conditional distribution of Y given $X = x_n$, and set $X_{n+1} = x_n$, $Y_{n+1} = y_{n+1}$;

  return (X_{n+1}, Y_{n+1});

until $n \geq N$;

References

- Chapter 11 of **BH**
- Chapter 7 of **BT**
- Reading: Chapter 12 of **BH**