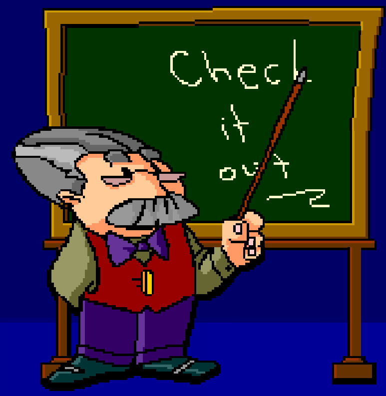
Randomized Algorithms

Markov Chains and Random Walks

Speaker: Chuang-Chieh Lin
Advisor: Professor Maw-Shang Chang
National Chung Cheng University

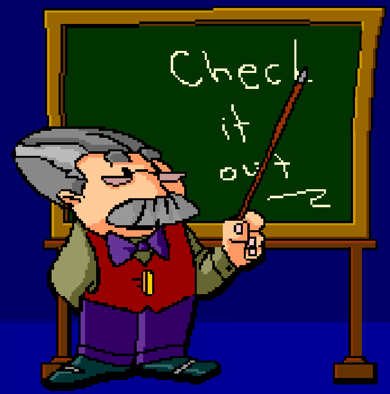
2007/4/25

References



- Professor S. C. Tsai's slides.
- *Randomized Algorithms*, Rajeev Motwani and Prabhakar Raghavan.
- *Probability and Computing - Randomized Algorithms and Probabilistic Analysis*, Michael Mitzenmacher and Eli Upfal.
- Wikipedia: Markov Chains

Outline



- Introduction to Markov chains
- Classification of states
- Stationary distribution
- Random walks on undirected graphs
- Connectivity problem

Introduction to Markov Chains

- Markov chains provide a simple but powerful framework for modeling random processes.

- Markov chains can be used to analyze simple randomized algorithms applying random walks.

Introduction to Markov Chains (cont'd)

Definition:

- A stochastic process $X = \{X(t), t \in T\}$ is a collection of random variables.
- If T is a countable set, say $T = \{0, 1, 2, \ldots\}$, we say that X is a *discrete time stochastic process*.
- Otherwise it is called *continuous time stochastic process*.
- Here we consider a discrete time stochastic process X_n, for $n = 0, 1, 2, \ldots$.

Introduction to Markov Chains (cont'd)

- If $X_n = i$, then the process is said to be in state i at time n.
- Denote $\mathbf{Pr}[X_{n+1} = j \mid X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0] = P_{i,j}$ for all states $i_0, i_1, \ldots, i_{n-1}, i, j$ and all $n \geq 0$.
- X_{n+1} depends only on X_n.

Introduction to Markov Chains (cont'd)

That is,

$$P_{i,j} = \mathbf{Pr}[X_{n+1} = j \mid X_n = i],$$

for all states i, j and all $n \geq 0$

Such a stochastic process is known as a **Markov chain**.

- Formal definitions.

Markov property

- In probability theory, a stochastic process has the Markov property if the conditional probability distribution of future states of the process, given the present state and all past states, depends only upon the current state and not on any past states.
- Mathematically, if $X(t)$, $t > 0$, is a stochastic process, the Markov property states that

$$
\begin{aligned}
& \Pr[X(t+h) = y \mid X(s) = x(s), \forall s \leq t] \\
= \quad & \Pr[X(t+h) = y \mid X(s) = x(t)], \quad \forall h > 0
\end{aligned}
$$

Markov chain

- In mathematics, a Markov chain, named after Andrey Markov, is a discrete-time stochastic process with the Markov property.

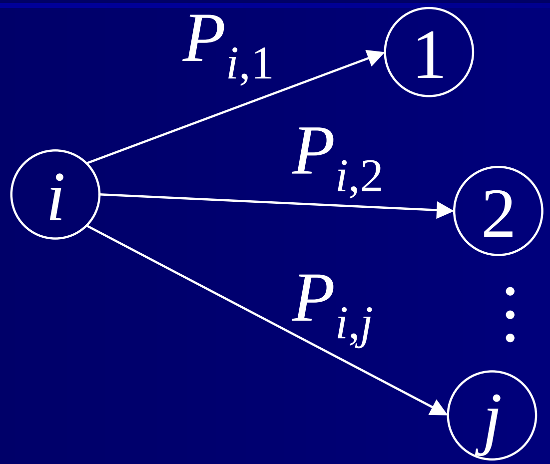
June 14, 1856 – July 20, 1922

Homogeneous

- Markov processes are typically termed (*time*-) *homogeneous* if

$$
\begin{aligned}
& \Pr[X(t+h) = y \mid X(t) = x(t)] \\
= \; & \Pr[X(h) = y \mid X(0) = x(t)], \quad \forall t, h > 0
\end{aligned}
$$

Transition matrix



- $P_{i,j} \geq 0.$
- $\sum_j P_{i,j} = 1.$

$$\mathbf{P} = \begin{bmatrix} P_{0,0} & P_{0,1} & \cdots & P_{0,j} & \cdots \\ P_{1,0} & P_{1,1} & \cdots & P_{1,j} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots \\ P_{i,0} & P_{i,1} & \cdots & P_{i,j} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots \end{bmatrix}$$

transition matrix

Transition probability

The m-step transition probability $P_{i,j}^m$ of the Markov chain is defined as the conditional probability, given that the chain is currently in state i, that will be in state j after m additional transitions. That is,

$$P_{i,j}^m = \mathbf{Pr}[X_{n+m} = j \mid X_n = i], \text{ for } m \geq 0, \ i, j \geq 0.$$

Conditioning on the first transition from i, we have the following equation:

$$P_{i,j}^m = \sum_{k \geq 0} P_{i,k} P_{k,j}^{m-1}.$$

Chapman-Kolmogorov equation

- Generalize the previous result, we have *Chapman-Kolmogorov equation* as follows.

$$P_{i,j}^{n+m} = \sum_{k \geq 0} P_{i,k}^{n} P_{k,j}^{m}.$$

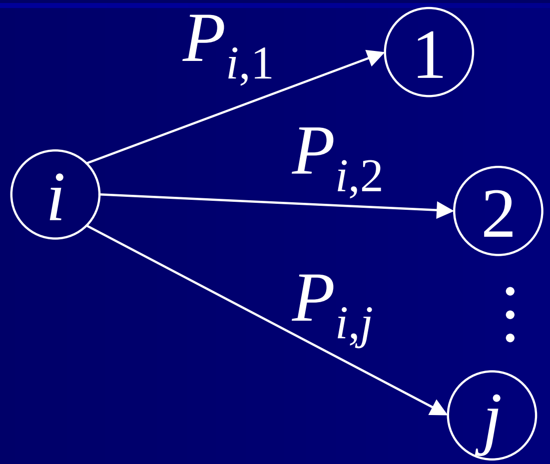
Chapman-Kolmogorov equation (cont'd)

Proof: By the definition of the n-step transition probability,

$$
\begin{aligned}
P_{i,j}^{n+m} &= \sum_{k \geq 0} \Pr[X_{n+m} = j, X_n = k \mid X_0 = i] \\
&= \sum_{k \geq 0} \Pr[X_n = k \mid X_0 = i] \cdot \\
&\quad \Pr[X_{n+m} = j \mid X_n = k, X_0 = i]
\end{aligned}
$$

By the Markov property, $\Pr[X_{n+m} = j \mid X_n = k, X_0 = i] = \Pr[X_{n+m} = j \mid X_n = k] = P_{k,j}^m$. With the additional observation that $\mathbf{Pr}[X_n = k \mid X_0 = i] = P_{i,k}^n$, the theorem immediately follows. ■

Recall: Transition matrix



- ★ $P_{i,j} \geq 0.$
- ★ $\sum_j P_{i,j} = 1.$

$$\mathbf{P} = \begin{bmatrix} P_{0,0} & P_{0,1} & \cdots & P_{0,j} & \cdots \\ P_{1,0} & P_{1,1} & \cdots & P_{1,j} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots \\ P_{i,0} & P_{i,1} & \cdots & P_{i,j} & \cdots \\ \vdots & \vdots & \ddots & \vdots & \ddots \end{bmatrix}$$

transition matrix

Recall: Homogeneous

- Markov processes are typically termed (*time*-) *homogeneous* if

$$\begin{aligned} & \Pr[X(t+h) = y \mid X(t) = x(t)] \\ = \; & \Pr[X(h) = y \mid X(0) = x(t)], \quad \forall t, h > 0 \end{aligned}$$

★ Let $\mathbf{P}^{(n)}$ denote the matrix of n-step transition probabilities $P^n_{i,j}$, then the Chapman-Kolmogorov equations implies that $\mathbf{P}^{(n)} = \mathbf{P}^n$.

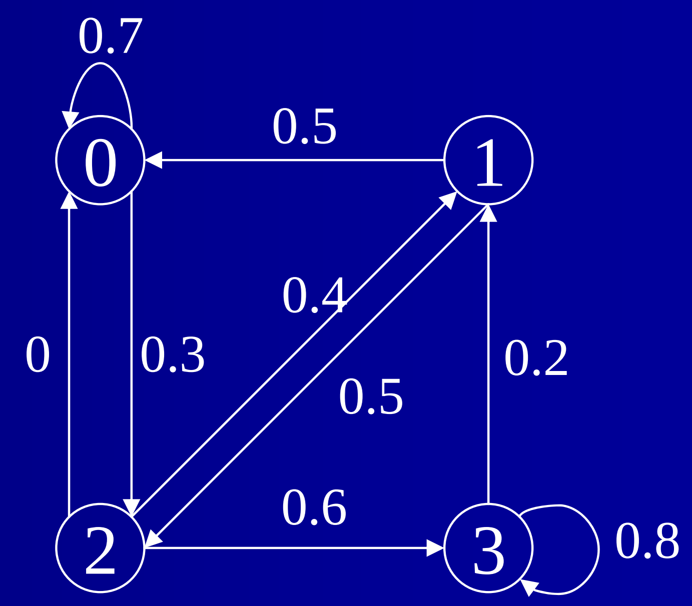
- For example,

$\mathbf{P} =$

	0	1	2	3
0	0.7	0	0.3	0
1	0.5	0	0.5	0
2	0	0.4	0	0.6
3	0	0.2	0	0.8

$\mathbf{P}^2 =$

	0	1	2	3
0	0.49	0.12	0.21	0.18
1	0.35	0.2	0.15	0.3
2	0.2	0.12	0.2	0.48
3	0.1	0.16	0.1	0.64



Classification of states

- A first step in analyzing the long-term behavior of a Markov chain is to classify its states.
- In the case of a finite Markov chain, this is equivalent to analyzing the connected connectivity structure of the directed graph representing the Markov chain.

Basic definitions

- State j is said to be accessible from state i if $P_{i,j}^{n} > 0$ for some $n \geq 0$.
- We say states i and j communicate if they are both accessible from each other. $(i \leftrightarrow j)$
- The Markov chain is said to be irreducible if all states communicate with each other.

Basic definitions (cont'd)

- Let $r_{i,j}^t$ denote the probability that starting at state i, the first transition to state j occurs at time t. That is,

  $r_{i,j}^t = \mathbf{Pr}[X_t = j$ and, for $1 \leq s \leq t-1, X_s \neq j \mid X_0 = i]$.

- State i is said to be recurrent if $\sum_{t\geq 1} r_{i,i}^t = 1$, and transient if $\sum_{t \geq 1} r_{i,i}^t < 1$.

Basic definitions (cont'd)

- A state j in a discrete time Markov chain is periodic if there exists an integer $\Delta > 1$ such that $\mathbf{Pr}[X_{t+s} = j \mid X_t = j] = 0$ for some integer $t \geq 0$ unless s is divisible by Δ.
- A state i has period d if $d = \gcd\{n \mid P^n_{i,i} > 0\}$, where *gcd* means the *greatest common divisor*.
- A discrete time Markov chain is periodic if there exists *at least one periodic state* in the chain.

Basic definitions (cont'd)

- A state with period 1 is said to be aperiodic.
- We denote by $h_{i,j}$ the expected time from state i to state j. So we have $h_{i,j} = \sum_{t \geq 1} t \cdot r_{i,j}^{t}$.
- A recurrent state i is said to be positive recurrent, if $h_{i,i} < \infty$. Otherwise it is null recurrent.
- Positive recurrent, aperiodic states are called ergodic.

Null recurrent?

- For example, consider a Markov chain whose states are the positive integers.
- From state i, the probability of going to state i+1 is $i/(i+1)$.
- With probability $1/(i+1)$, the chain returns to state 1.

Null recurrent? (cont'd)

- Starting at state 1, the probability of not having returned to state 1 within the first *t* steps is thus

$$\prod_{j=1}^{t} \frac{j}{j+1} = \frac{1}{t+1}.$$

- Hence the probability of never returning to state 1 from state 1 is 0, then we have state 1 is recurrent. It follows that

$$r_{1,1}^{t} = \frac{1}{t} \cdot \frac{1}{t+1} = \frac{1}{t(t+1)}.$$

Null recurrent? (cont'd)

- However, the expected number of steps until the first return to state 1 from state 1 is

$$h_{1,1} = \sum_{t=1}^{\infty} t \cdot r_{1,1}^{t} = \sum_{t=1}^{\infty} \frac{1}{t+1},$$

which is unbounded.

- Thus this Markov chain has null recurrent states.

- In the foregoing example, the Markov chain had an infinite number of states.
- This is necessary for null recurrent states to exist.
- Yet for a *finite* Markov chain, we have the following lemma.

Lemma 1

- In a *finite* Markov chain:
  - At least one state is recurrent; and
  - All recurrent states are positive recurrent.

- We omit the proof here, though it is not hard.

Recall that...

- State i is said to be recurrent if $\sum_{t\geq 1} r_{i,i}^t = 1$, and transient if $\sum_{t\geq 1} r_{i,i}^t < 1$.

Proposition 1

- State i is recurrent if $\sum_{n \geq 0} 1 \cdot P^n_{i,i} = \infty$.
  - That is, the expected number of visits to state i over all time is infinite.
- State i is transient if $\sum_{n \geq 0} 1 \cdot P^n_{i,i} < \infty$.
  - That is, the expected number of visits to state i over all time is finite.

Proof of this proposition is a little bit complicated, so we omit it here.

Proof of the second statement

- Let N_i be the number of visits to state i over all time, then $\mathbf{E}[N_i] = \sum_{n\geq 0} P_{i,i} = 0$.
- Given an initial state distribution, let V_i denote the event that the system eventually goes to state i.
  - Obviously, $\Pr[V_i] \leq 1$.
- If V_i does not occur, then $N_i = 0$.
  - This implies that $\mathbf{E}[N_i \mid V_i^c] = 0$.

Proof of the second statement (cont'd)

- Otherwise (i.e., V_i occurs), there exists a time t when the system first enters state i.

- In this case, given that the state is i, let V_{ii} denote the event that the system *eventually returns to state* i.

  - Thus V_{ii}^c is the event that the system never returns to state i.

Proof of the second statement (cont'd)

- Since i is transient, there exists a state, say j, such that for some t', $P_{i,j}^{t'} > 0$ but i is not accessible from j.
  - Thus if we enter state j at time t', the event V_{ii}^c will occur.
- Since this is one possible way that V_{ii}^c can occur, $\Pr[V_{ii}^c] \geq P_{i,j}^{t'} > 0$.

Proof of the second statement (cont'd)

- After each return to i, there is a probability $\Pr[V_{ii}^c] > 0$ that state i will never be reentered.
- Hence, given V_i, the expected number of visits to i is geometric with conditional expected value $\mathbf{E}[N_i \mid V_i] = 1/\Pr[V_{ii}^c] \leq 1/P_{i,j}^{t'}$.
- Finally we have

$$\begin{aligned} \mathbf{E}[N_i] &= \mathbf{E}[N_i \mid V_i^c] \cdot \Pr[V_i^c] + \mathbf{E}[N_i \mid V_i] \cdot \Pr[V_i] \\ &= \mathbf{E}[N_i \mid V_i] \cdot \Pr[V_i] < \infty. \end{aligned}$$

□

Corollary 1

- If state *i* is recurrent, and state *i* communicates with state *j*, then state *j* is recurrent.

- Proof:
  – Exercise.

Stationary Distribution

Definition: A stationary distribution (also called an equilibrium distribution) of a Markov chain is a probability distribution $\bar{\pi}$ such that

$$\bar{\pi} = \bar{\pi}\mathbf{P}.$$

Recall: P is the one-step transition probability matrix of a Markov chain.

Computing the stationary distribution of a finite Markov chain

- One way to compute the stationary distribution of a finite Markov chain is to solve the system of linear equations

$$\bar{\pi} = \bar{\pi}\mathbf{P}.$$

- This is particularly useful if one is given a specific chain.

- For example, given the transition matrix

$$\mathbf{P}=\begin{bmatrix} 0 & 1/4 & 0 & 3/4 \\ 1/2 & 0 & 1/3 & 1/6 \\ 1/4 & 1/4 & 1/2 & 0 \\ 0 & 1/2 & 1/4 & 1/4 \end{bmatrix},$$

we have five equations for the four unknowns π_0, π_1, π_2, and π_3 given by $\bar{\pi} = \bar{\pi}\mathbf{P}$ and $\sum_{i=0}^{3} \pi_i = 1$.

Another technique

- Another useful technique is to study the cut-sets of the Markov chain.

- For any state *i* of the chain,

$$\sum_{j=0}^{n} \pi_j P_{j,i} = \pi_i = \pi_i \sum_{j=0}^{n} P_{i,j}$$

or

$$\sum_{j \neq i}^{n} \pi_j P_{j,i} = \sum_{j=0}^{n} \pi_i P_{i,j}.$$

- That is, in the stationary distribution the probability that a chain leaves a state equals the probability that it enters the state.
- This observation can be generalized to sets of states as follows.

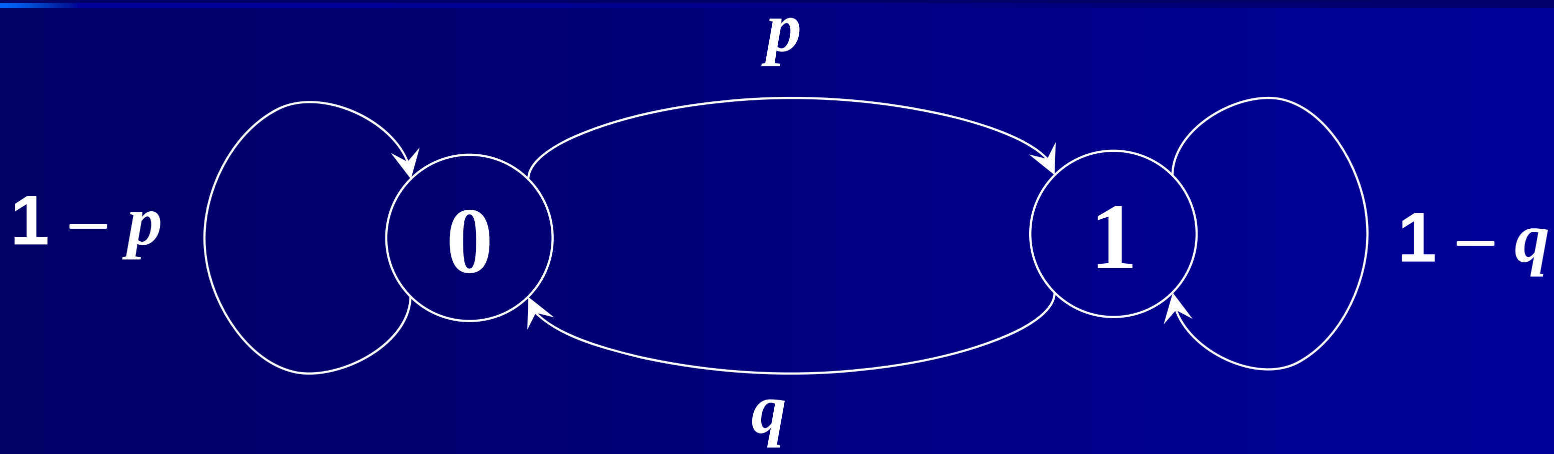
Theorem

- Let *S* be a set of states of a finite, irreducible, aperiodic Markov chain. In the stationary distribution, the probability that the chain leaves the set *S* equals the probability that it enters *S*.

- In other words, if *C* is a cut-set in the graph representation of the Markov chain, then in the stationary distribution the probability of crossing the cut-set in one direction is equal to the probability of crossing the cut-set in the other direction.

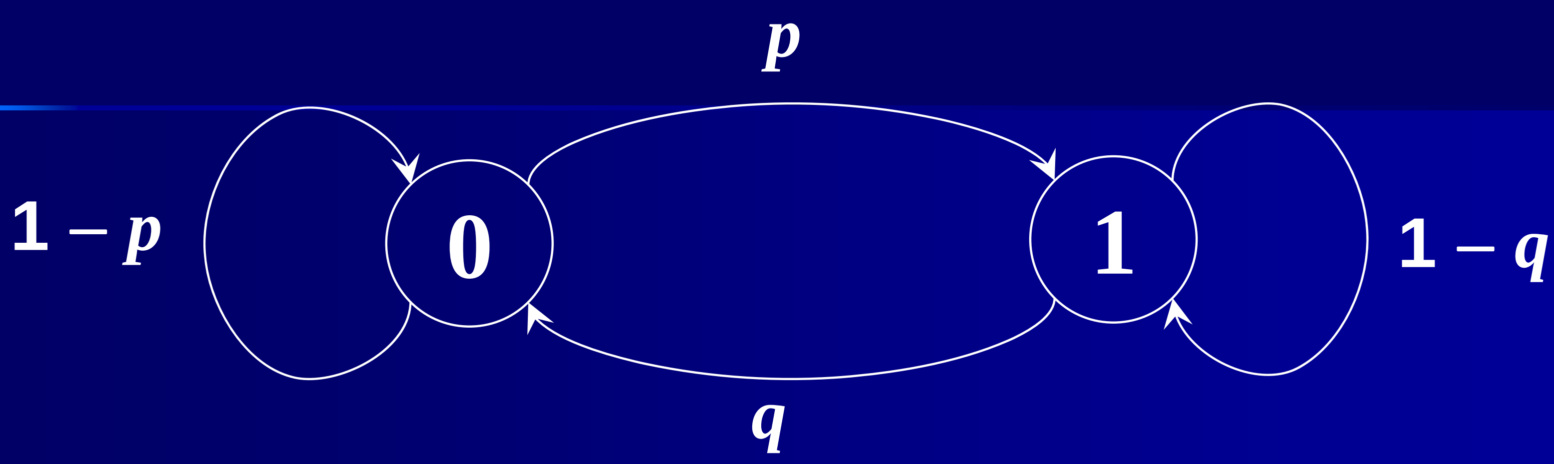
- That is, in the stationary distribution the probability that a chain leaves a state equals the probability that it enters the state.

For example,

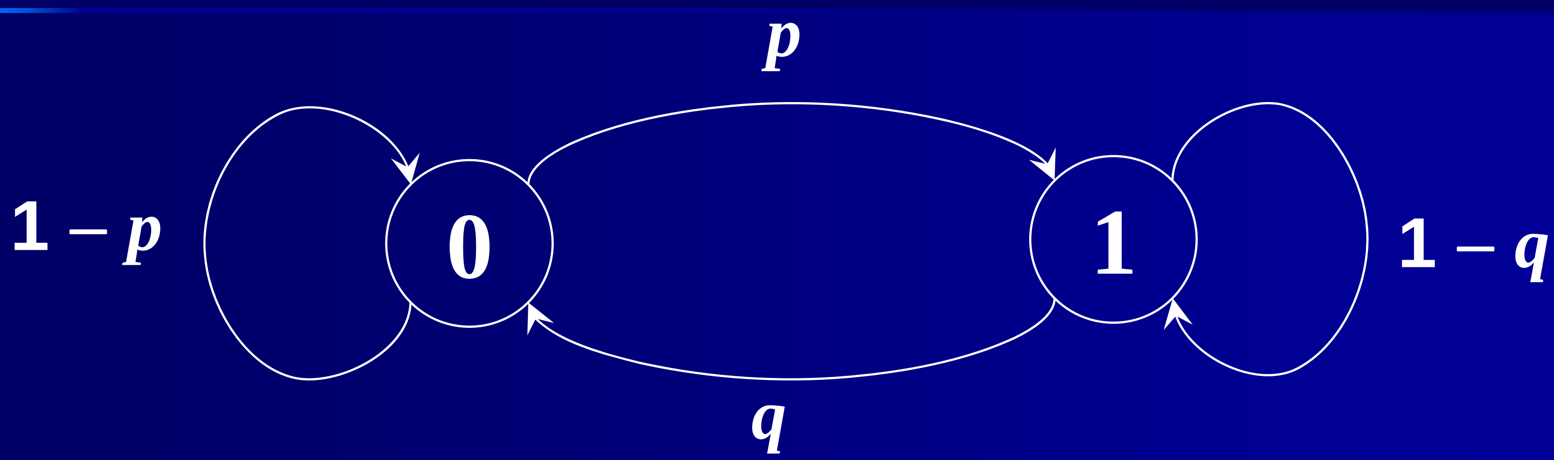


- The transition matrix is

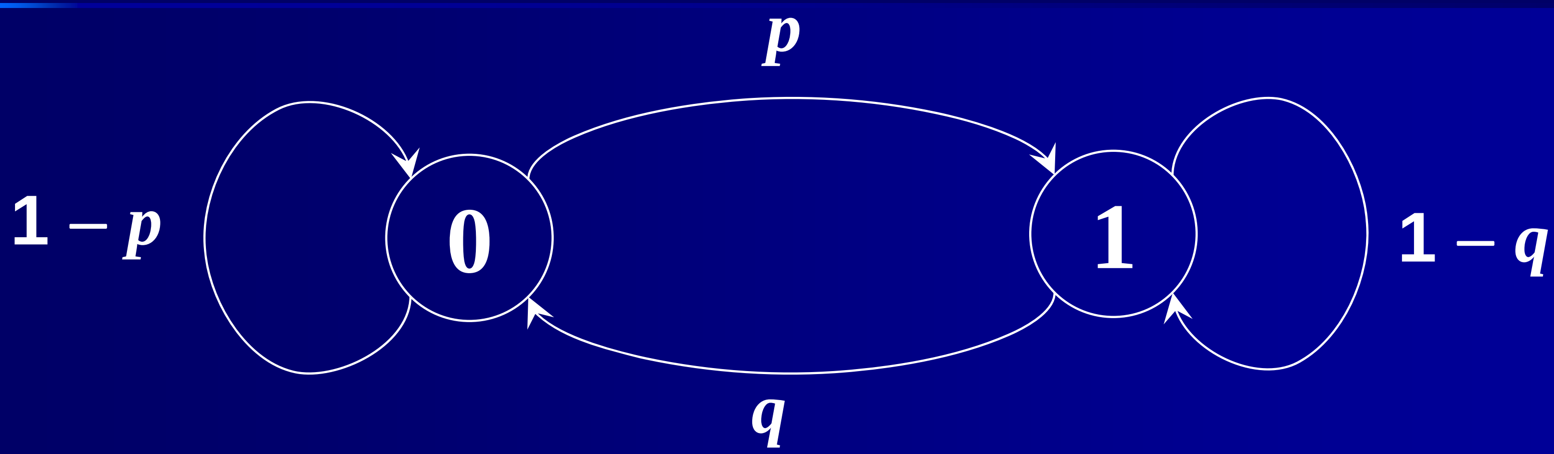
$$\mathbf{P} = \begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix}.$$



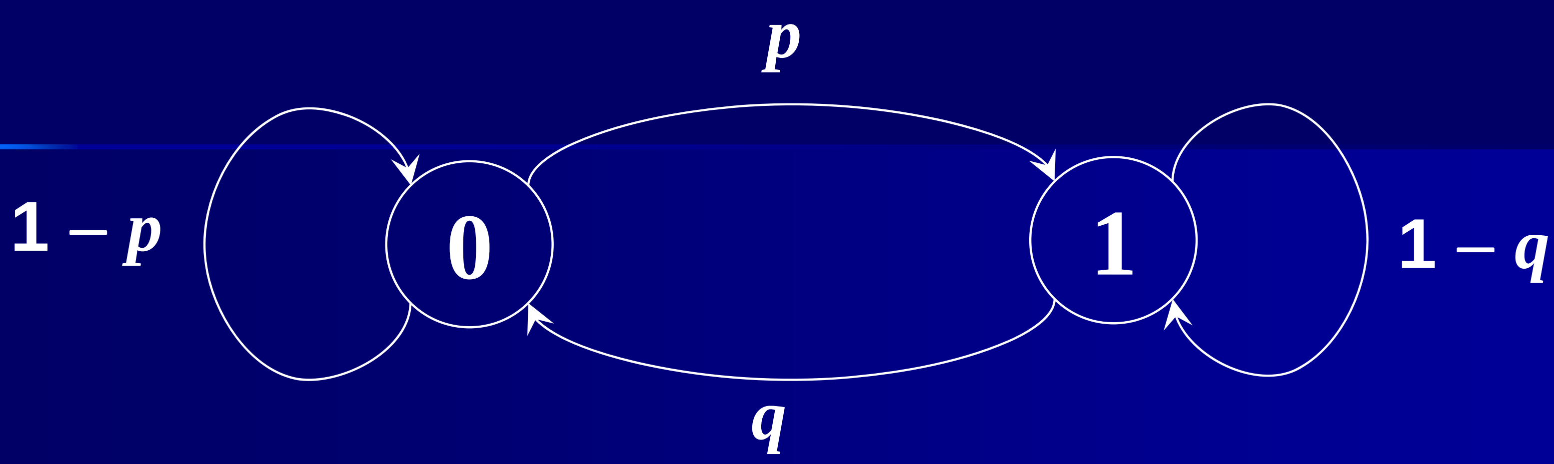
- This Markov chain is often used to represent bursty behavior.
- For example, when bits are corrupted in transmissions they are often corrupted in *large blocks*, since errors are often caused by an external phenomenon of *some duration*.



- In this setting, begin in state 0 after *t* steps represents that the *t*th bit was sent successfully, while being in state 1 represents that the bit was corrupted.



- Blocks of successfully sent bits and corrupted bits both have lengths that follow a geometric distribution.
- When p and q are small, state changes are rare, and the bursty behavior is modeled.



- Solving $\bar{\pi} = \bar{\pi}\mathrm{P}$ corresponds to solving the following system of three equations:

$$\begin{aligned} \pi_0(1-p) + \pi_1 q &= \pi_0; \\ \pi_0 p + \pi_1(1-q) &= \pi_1; \\ \pi_0 + \pi_1 &= 1. \end{aligned}$$

- Using the cut-set formulation,

$$\pi_0 p = \pi_1 q.$$

we have that in the stationary distribution the probability of leaving state 0 must equal the probability of entering state 0.

- Again, now using $\pi_0 + \pi_1 = 1$ yields $\pi_0 = q/(p+q)$ and $\pi_1 = p/(p+q)$.

- For example, with the natural parameters $p = 0.005$ and $q = 0.1$, in the stationary distribution more than 95% of the bits are received successfully.

Theorem 1

positive recurrent or aperiodic

- Any *finite*, *irreducible*, and *ergodic* Markov chain has the following properties:

★ The chain has a unique stationary distribution $\bar{\pi} = (\pi_0, \pi_1, \ldots, \pi_n)$;

★ For all j and i, $\lim_{t\to\infty} P^t_{j,i}$ exists and it is independent of j;

★ $\pi_i = \lim_{t\to\infty} P^t_{j,i} = \frac{1}{h_{i,i}}$.

Proof of Theorem 1

- Please refer to the textbook for the details.
  - page 168–170 in [MU05]
- We omit the proof here.
- Yet we can still eye on the following lemma.

Lemma 2

- For any *irreducible*, *ergodic* Markov chain and for any state *i*,

$$\lim_{t\to\infty} P^t_{i,i} \text{ exists and } \lim_{t\to\infty} P^t_{i,i} = \frac{1}{h_{i,i}}.$$

- We explain instead of proving the lemma as follows.

Informal justifications…

- The expected time between visits to i is $h_{i,i}$ and therefore state i is visited $1/h_{i,i}$ of the time.
- Thus the limit of $P^t_{i,i}$, which represents the probability that, a state chosen far in the future is at state i when the chain starts at state i, must be $1/h_{i,i}$.
- Since the limit exists, we can show that

$$\underbrace{\lim_{t\to\infty} P^t_{j,i} = \lim_{t\to\infty}}_{} P^t_{i,i} = \frac{1}{h_{i,i}}.$$

We omit the detail here for simplicity.

Markov Chains and Random Walks on Undirected Graphs…

Random walks on undirected graphs

- A *random walk* on G is a Markov chain defined by the sequence of moves of a particle between vertices of G.
- In this process, the *place* of the particle at a given time step is the *state* of the system.
- If the particle is at vertex i and if i has $d(i)$ outgoing edges, then the probability that the particle follows edge (i, j) and moves to neighbor j is $1/d(i)$.

Lemma 3

- A random walk on an undirected graph G is aperiodic iff G is not bipartite.

Proof of Lemma 3

- A graph is bipartite iff it does not have any odd cycle.
- In an undirected graph, there is always a path of length 2 from a vertex to itself.
- $\Rightarrow$ Thus, if the graph is bipartite then the random walk is periodic (d = 2).
- $\Leftarrow$ If not bipartite, then it has an odd cycle and gcd(2,odd-number) = 1. Thus, the Markov chain is aperiodic.

Some technical restrictions

- Due to some technical reasons, for the remainder of our discussion, we assume that the graph *G* that we will discuss is not bipartite.

Something needs to be clarified…

ergodic

aperiodic

- A random walk on a finite, undirected, connected, and non-bipartite graph *G* satisfies the conditions of Theorem 1.
  - A Markov chain that is *finite*, *irreducible*, and *ergodic*.
- Hence the random walk converges to a stationary distribution. (Due to Theorem 1.)
- The following Theorem shows that this distribution depends only on the degree sequence of the graph.

Theorem 2

A random walk on G converges to a stationary distribution $\bar{\pi}$, where

$$\pi_v = \frac{d(v)}{2|E|}.$$

Proof of Theorem 2

Since $\sum_{v \in V} d(v) = 2|E|$, $\sum_{v \in V} \pi_v = \sum_{v \in V} \frac{d(v)}{2|E|} = 1$.
That is, $\bar{\pi}$ is indeed a distribution.

Let $\mathbf{P}$ be the transition probability matrix and $N(v)$ be the neighbors of v.

$$(\bar{\pi}\mathbf{P})_v = \sum_{u \in N(v)} \frac{d(u)}{2|E|} \cdot \frac{1}{d(u)} = \frac{d(v)}{2|E|}.$$

thus the theorem follows. ■

Corollary 2

For any vertex u in G, $h_{u,u} = \frac{2|E|}{d(u)}$.

Recall that $h_{v,u}$ is the expected number of steps to reach u from v. So the corollary immediately follows.

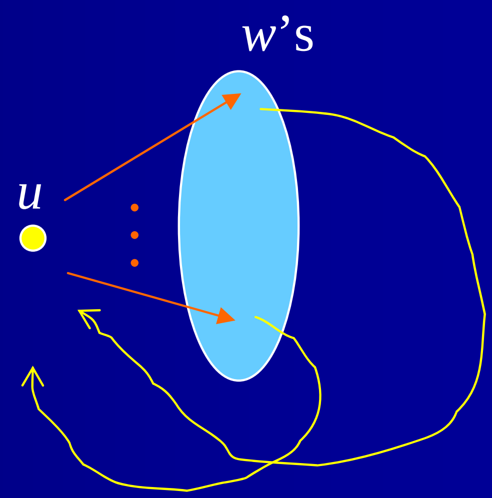
Lemma 4

If $(u, v) \in E$, then $h_{v,u} \leq 2|E|$.

Proof of Lemma 4

Recall that $N(u)$ denotes the neighbors of u in the given graph G.

Since $\frac{2|E|}{d(u)} = h_{u,u} = \frac{1}{d(u)} \cdot \sum_{w \in N(u)} (1 + h_{w,u})$, we have $2|E| = \sum_{w \in N(u)} (1 + h_{w,u})$.



Therefore $h_{v,u} < 2|E|$ (since $(u, v) \in E$.) $\blacksquare$

The Covering Time

- Definition:
  - The cover time of a graph $G = (V, E)$ is the maximum over all $v \in V$ of the expected time to visit all of the nodes in the graph by a random walk starting from v.

Lemma 5

The cover time of $G = (V, E)$ is bounded by $4|V| \cdot |E|$.

Proof of Lemma 5

- Choose a spanning tree of *G*. Then there exists a cyclic (Eulerian) tour on this tree, where each edge is traversed once in each direction, which can be found by doing a DFS.

- Let $v_0, v_1, \ldots, v_{2|V|-2} = v_0$ be the sequence of vertices in the tour, starting from v_0.

Proof of Lemma 5 (cont'd)

- Clearly the expected time to go through the vertices in the tour is an upper bound on the cover time.

- Hence the cover time is bounded above by

$$\sum_{i=0}^{2|V|-3} h_{v_i,v_{i+1}} < (2|V|-2)\boxed{(2|E|)} < 4|V|\cdot|E|. \quad \blacksquare$$

from Lemma 4

Application: *s* - *t* connectivity

- Suppose we are given an undirected graph $G(V, E)$ and two vertices *s* and *t* in G.
  - Let $n = |V|$ and $m = |E|$.

- We want to determine if there is a path connecting *s* and *t*.

Application: *s*-*t* connectivity (cont'd)

- This is easily done in linear time using a standard breadth-first search or depth-first search.
- However, such algorithms require $\Omega(n)$ space.
- The following randomized algorithm works only $O(\log n)$ bits of memory.

Application: *s* - *t* connectivity (cont'd)

s-t Connectivity Algorithm:

Input: G and two vertices s and t

Output: **Yes** or **No**

- ★ Start a random walk from s.
- ★ If the walk reaches t within $4n^3$ steps, return **YES**. Otherwise, return **No**.

Application: *s* - *t* connectivity (cont'd)

- We have assumed that *G* has no bipartite connected component.
  - The results can be made to apply to bipartite graphs with some additional technical work.
- Let us consider the following theorem.

Theorem 3

The s-t Connectivity Algorithm returns the correct answer with probability $\frac{1}{2}$ and it only errs by returning that there is no path from s to t when there is such a path.

Proof of Theorem 3

- The algorithm gives correct answer, when G has no s-t path.
- If G has an s-t path, the algorithm errs if it does not find the path in $4n^3$ steps.
- Now we consider the case that G has an s-t path.

Proof of Theorem 3 (cont'd)

- The expected time to reach t from s is bounded by the cover time, which is at most $4|V| \cdot |E| < 2n^3$.
  - $|E| \leq n(n-1)/2$
- Let a random variable X denote the number of steps needed for the s-t Connectivity Algorithm.
- By Markov's inequality,

$$\mathbf{Pr}[X > 4n^3] \leq \mathbf{Pr}[X \geq 4n^3] \leq \frac{\mathbf{E}[X]}{4n^3} < \frac{2n^3}{4n^3} = \frac{1}{2}. \quad \blacksquare$$

Proof of Theorem 3 (cont'd)

- The algorithm must keep track of its current position, which takes $O(\log n)$ bits, as well as the number of steps taken in the random walk, which also takes only $O(\log n)$ bits.
  - Since we count up only $4n^3$, which requires $\log(4n^3) = O(\log n)$ bits

Thank you.