6.842 Randomness and Computation February 24, 2014

Lecture 6

Lecturer: Ronitt Rubinfeld *Scribe: Mutaamba Maasha*

Outline

- Random Walks
- Markov Chains
- Stationary Distributions
- Hitting, Cover, Commute times

Markov Chains

A Markov chain is a random process specified in part by Ω = set of all "states" (or use V and call them nodes in a graph). In this class the set of states are always finite. We then want to consider $X_o ... X_t \in \Omega$ the sequence of visited states with the following property.

Markov Property

Next state depends only on the current state and not on the history of states. These state transitions are said to be **memoryless**.

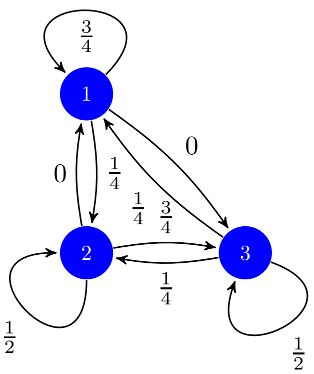
$$\Pr[X_{t+1} = Y | X_0 = x_0, X_1 = x_1, X_2 = x_2, ..., X_t = x_t] = \Pr[X_{t+1} = Y | X_t = x_t]$$

We can assume wlog transitions are independent of time as we can create a new set of states $\Omega \times T$ where T is the set of times.

$$P[i, j] = \Pr[X_{t+1} = i | X_t = j]$$

Markov chains with this property are called **time-homogeneous** or **stationary** Markov chains. This allows the transition probabilities $P[i, j]$ to be represented by a **transition matrix** $P_{i,j}$.

Example 1



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

For this Markov chain : $\Omega = \{1, 2, 3\}$ and the for the corresponding graph the edges are weighted the transition probability $P_{i,j}$.

Random walk

Random walks are an important special case where the state transitions are uniform on the subset corresponding to neighbors of node.

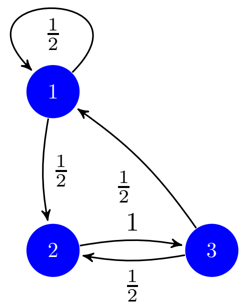
Definition 2 *Random Walk*

A random walk on $G = (V, E)$ is a sequence $s_0, s_1 \ldots$ of nodes where s_0 is a start node. At each step i, S_i is picked uniformly from $out(s_i)$, where $out(s_i)$ is the set of out edges of the state s_i.

The transition probability for the random walk is defined as :

$$P(i,j) = \begin{cases} (i,j) \in E & \frac{1}{|deg_{out}(i)|} \\ \text{otherwise} & 0 \end{cases}$$

Example 3



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

In this random walk on a digraph the probability of getting to adjacent vertices is uniform.

Properties of Markov Chains

The following are properties of Markov Chains.

Normalization

$$\forall i \in \Omega : \sum_{j \in \Omega} P(i,j) = 1$$

This condition gives a normalization of out going state transitions.

Transitions

Transitions probabilities for t steps

$$P^t(i,j) = \begin{cases} t = 1 & P(i,j) \\ t > 1 & \sum_k P(i,k) P^{t-1}(k,j) \end{cases}$$

This is just matrix multiplication

$$P^t = \underbrace{P \ldots P}_{t \text{ times}}$$

The initial distribution is given by

$$\pi^0 = (\pi_1^0 \ldots \pi_n^0)$$

where $\pi_i^0 = \Pr[\text{start at node } i]$
The after on step the intial distribution becomes the following distirubtion :

$$\pi^i = \pi^0 P$$

at step t the evolution of the inital distribution is given by the following :

$$\pi^i = \pi^0 P^t$$

Special Cases of Markov Chains

Definition 4 *Stocastic matrix*

A **(left) stocastic matrix** P is a square matrix where all rows sum to 1. All Markov chains are stocastic matrices.

$$\text{stocastic matrix} \equiv \forall i \in \Omega : \sum_{j \in \Omega} P_{i,j} = 1$$

Definition 5 *Doubly stocastic matrix*

A **doubly stocastic matrix** P is a square matrix where all rows and colums sum to one.

$$\text{doubly stocastic matrix} \equiv \forall i \in \Omega : \sum_{j \in \Omega} P_{i,j} = \sum_{j \in \Omega} P_{j,i} = 1$$

An example of a doubly stocastic matrice is a random walk on a regular (indegree matches out degree) directed graph.

$$P(i,j) = \frac{1}{|deg_{out}(i)|} = \frac{1}{|deg_{in}(j)|}$$

Definition 6 *Irreducible Markov chain*

An **irreducible Markov chain** has the property that for any two states i, j there is a t such that after t-steps it is always possible to get from i to j (with a non-zero probability).

$$\forall i, j \in \Omega : \exists t = t(i,j) : P^t > 0$$

This given a directed weighted graph $G = (V, E)$ of a markov chain the following graph can be constructed

$$G' = (V, \{(i,j) \in E : P(i,j) > 0\})$$

The markov chain represented by G is irreducible iff the directed graph G' is *strongly connected.*

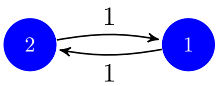
Definition 7 *Ergodic Markov chain*

An **ergodic Markov chain** is a markov chain if there is an t after which there is a positive probability to getting to any state.

$$\exists t \in \mathbb{N} : \forall i, j \in \Omega : \forall t' > t : P^{t'} > 0$$

This is a stronger property than irreducible. An example of non-ergodic markov chain is a bipartite graph for a given $i \in \Omega$ reachability of (i,i) is affected determined by the parity of t.

Example 8



$$P = \begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \end{array}$$

This markov chain is not erogdic as $P^t(i,i) = t \mod 2$.

Definition 9 *Aperiodic Markov chain*

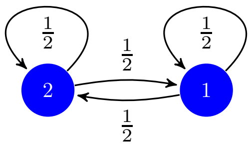
A **aperiodic Markov chain** is morkov chain such that for any state the returns to that state occurs at irregular times.

$$\forall i \in \Omega : \gcd\{t : P^t(i,i) > 0\} = 1$$

A way of making a graph aperiodic is to add loops.

Example 10

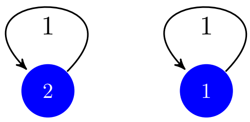
The following graph is aperiodic



$$P = \begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \end{array}$$

The following graph is aperiodic but not erogdic.

Example 11



$$P = \begin{array}{c} \\ 1 \\ 2 \end{array} \begin{array}{c} \begin{array}{cc} 1 & 2 \end{array} \\ \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \end{array}$$

Theorem 12 *Erodic iff Irreducible and aperiodic.*

Stationary Distribution

A stationary distribution Π is a state distribution where

$$\forall y \in \Omega : \Pi(y) = \sum_{x \in \Omega} \Pi(x) P(x,y)$$

or equivalently $\Pi = \Pi P$.

In example 8 there is no such Π as the state oscillates between the two states. In the example 11 there are many stationary distributions $\{(i, 1-j) : i \in [0,1]\}$. The set of Markov chains where there exists Π such that Π is unique is given by the following theorem.

Theorem 13 *Every erodoic Markov chain has a unique stationary distribution.*

Stationary distribution of undirected non-biparite graph is :

$$\Pi(i) = \frac{deg(i)}{2|E|}$$

So d-regular graphs have Π-uniform distributions.

Hitting times

The **hitting time** is the expect number of steps it takes to hit a given state j starting from state i.

$$h_{i,j} = E[\text{time starting at } i \text{ return to } j]$$

Theorem 14 $h_{i,i} = \frac{1}{\Pi_i}$ *for erogdic Markov Chains.*

Cover time

Cover time of an undirected graph is the expected number of steps for a random walk to hit every node in G starting from a give node.

$$C_i(G) = E[\text{ size of set start at u and hit every node in G}]$$

If no start node is given the maximum time over all nodes is taken.

$$C(G) = max_{i\in\Omega} C_i(G)$$

Example 15

$C(K_n^*) = \theta(n \ln n)$ where K_n^* is complete graph with n vertices. Each vertice has a self loops at each node making K_n^* aperiodic.
Each step is uniformly random and covering the graph becomes picking a vertex radomly with replacement until all vertices are sampled. This is equivalent to the coupon collector's problem which states it takes $n \log n$ steps to sample n items with replacement. Thus it takes $n \log n$ time to cover $C(K_n^*)$

Example 16

$C(L_n) = \theta(n^2)$ where L_n is a line of n vertices.
The expected distance a random walk on the line travels after n steps is $\theta(\sqrt{n})$. Thus to cover n vertices it takes $\theta(n^2)$ to cover n vertices.

Example 17

$C(lollipop(n)) = \theta(n^3)$ where $lollipop(n)$ is a complete graph $K_{n/2}$ attached to the line $L_{n/2}$.
If a random walk starts on the line then it takes $\theta(n^2)$ to cover the line and end up in the cluster which takes $\theta(n \log n)$ to cover giving a total complexity of $\theta(n \log n)$.
If a random walk starts in the cluster it enters the line with probability $\frac{2}{n-2}$. Thus it takes $O(n)$ steps to exit the cluster and even then it is likely to wander back for each of the $O(n^2)$ steps to cover $L_{n/2}$. Thus we get $O(n^3)$ expected time to cover $lollipop(n)$

Theorem 18 $C(G) \leq \theta(m * n)$ *actually* $< 8m(n-1)$

Definition 19 *Commute time*

The "commute time" is the the number of steps starting from i to j and back to i.

$$C_{i,j} = h_{i,j} + h_{j,i} = \text{Exp}[\text{ the number of steps to start at } i \text{ hit } j \text{ and return to } i]$$

Lemma 20 $\forall (i,j) \in E : C_{i,j} \leq O(m)$

Proof next time

Lemma 21 $C(G) = \theta(n^3)$

Proof Start a a vertex v_0 let T be any spanning tree rooted at v_0. The number of edges $= n-1$. Take a walk in this tree starting at v_0. (a dfs walk).
$v_0, v_1 ... v_{2n-2} = v_0$ is a depth first traversal each edge v, v_{i+1} appears twice (when you come down and when you return). What is the cover time of the graph.

$$C(G) \leq \sum_{i=0}^{2n-3} h_{v_j v_{j+1}}$$

using $C_(i,j) = h_{i,j} + hj, i$

$$= \sum_{(i,j) \in T} C_{i,j}$$

using lemma 20

$$= O(\sum_{(i,j) \in T} m)$$

whic reduces to

$$= O(n, m)$$

for $C(lolipop(n)) = O(n, m) = O(n^3)$ ■