Fall 2014 Randomized Algorithms	Nov 5, 2014
Lecture 7	
Prof. Friedrich Eisenbrand	*Scribes: Szymon Dudycz*

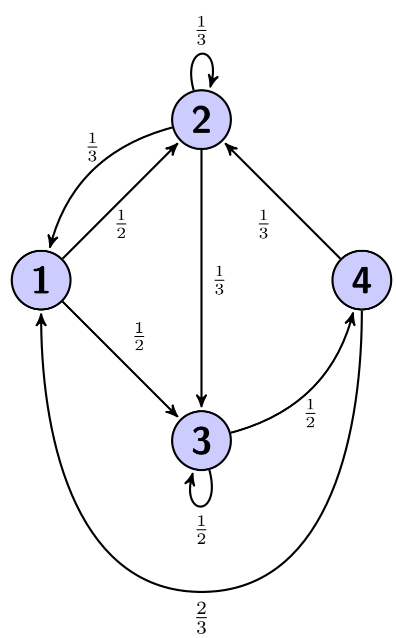
In this lecture we introduce Markov Chains and prove theirs property, mainly existence of unique stationary distribution. Then we present Metropolis-Hastings algorithm used for sampling vertices from graph using Markov Chains.

1 Markov Chains

A Markov Chain has a finite set of states and associated with them transition matrix P, such that P_{xy} is a probability of moving from x to y, where for each $x \sum_y P_{xy} = 1$. Markov Chains are often represented as a directed graph with each vertex representing a state and edge from x to y having weight P_{xy}.

Let $\pi \in [0,1]^{|V|}, \pi^T \mathbf{1} = 1$ be initial probability distribution. Then $\pi P^{(t)}$ is a probability distribution after t steps. It can be easily seen that $\pi P^{(t)}$ is indeed a distribution, because it is non-negative and using induction $\pi P^{(t)} \mathbf{1} = \pi P^{(t-1)} P \mathbf{1} = \pi P^{(t-1)} \mathbf{1} = \pi \mathbf{1} = \mathbf{1}$.

1.1 Example of Markov Chain



The transition matrix associated with this Markov chain is:

$$P = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{2}{3} & \frac{1}{3} & 0 & 0 \end{pmatrix}$$

1.2 Properties of Markov Chains

Definition 1 *Let $p^{(0)}$ be any initial distribution and let $p^{(t)}$ denote distribution after t steps, so $p^{(t)} = pP^{(t)}$. Then a* long-term probability distribution *is:* $a^{(t)} = \frac{1}{t}\left(p^{(0)} + \cdots + p^{(t-1)}\right)$.

Definition 2 *Markov Chain is* connected *if for each pair of states i, j there is a path from i to j.*

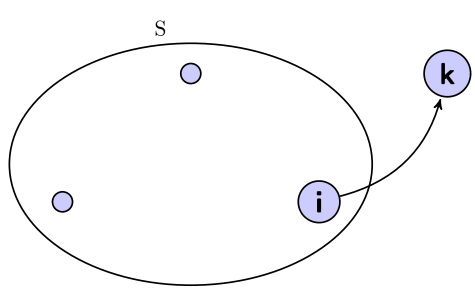
Definition 3 π *is a* stationary distribution *if* $\pi P = \pi$.

If π is a stationary distribution then π is left eigenvector of P with eigenvalue 1. P has no eigenvalue λ such that $|\lambda| > 1$, because otherwise let v be an eigenvector of P, so $vP^{(t)} = \lambda^{(t)}$. Then $\|\lambda^{(t)}v\|_\infty$ is large, whereas all entries in $P^{(t)}$ are at most 1.

1.3 Fundamental theorem of Markov Chains

Lemma 4 *Let P be a transition matrix of connected Markov Chain and let $A = [P - I, \mathbf{1}]$. Then rank(A)=n, where n is number of states in Markov Chain.*

Proof $Ax = 0$ has a solution $\begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix}$. Let's assume that $\text{rank}(A) < n$. Then there is another solution $\begin{pmatrix} x \\ \alpha \end{pmatrix}$ which is orthogonal to $\begin{pmatrix} \mathbf{1} \\ 0 \end{pmatrix}$, in particular $\sum_i x_i = 0, x \neq \mathbf{0}$. Let's consider set S of largest entries in x. Because Markov Chain is connected, then there is some $i \in S, k \notin S$ such that there is an edge from i to k.



From $Ax = 0$ follows that $x_i = \sum_j P_{ij} x_j + \alpha$, so x_i is a convex combination of neighbour states of i. Because $x_k < x_i$ and $P_{ik} > 0$, then α must be greater than 0.

From similar argument for smallest entries in x follows that $\alpha < 0$, which is a contradiction. ■

Corollary 5 $\{\pi : \pi P = \pi\}$ *has dimension at most 1.*

Proof π such that $\sum_i \pi_i = 1$ is an eigenvector of P with eigenvalue 1 if and only if it is a solution of $\pi A = (0, 0, \ldots, 0, 1)$. Because this system has $n + 1$ equations of rank n, then if there is any solution, it is unique. ■

Theorem 6 (Fundamental theorem of Markov Chains) *For a connected Markov Chain there exists a unique probability distribution π such that $\pi P = \pi$. Moreover for any initial distribution p^0 long-term probability distribution satisfies $\lim_{t\to\infty} a^{(t)} = \pi$.*

Proof Let $b^{(t)} = a^{(t)} P - a^{(t)} = \frac{1}{t}\left(p^{(1)} + \cdots + p^{(t)}\right) - \frac{1}{t}\left(p^{(0)} + \cdots + p^{(t-1)}\right) = \frac{1}{t}\left(p^{(t)} - p(0)\right)$. Because $|b^{(t)}| \leq \frac{2}{t}$, $a^{(t)} P - a^{(t)} = b^{(t)}$ converges to $\mathbf{0}$, so $a^{(t)}$ converges to such π that $\pi P = \pi$.

Uniqueness of this solution follows from Corollary 5. ■

1.4 Time-reversible Markov Chains

Definition 7 *Markov Chain is* time-reversible *if there exists $\pi, \sum \pi_i = 1$ such that $\forall_{i,j} \pi_i P_{ij} = \pi_j P_{ji}$.*

Theorem 8 *A connected time reversible Markov Chain has stationary distribution π, where π is a distribution from the defintion of time-reversible Markov Chains.*

Proof π is a stationary distribution if $\forall_i \pi P^i = \pi_i$. But $\pi_i = \sum_j \pi_i P_{ij} = \sum_j \pi_j P_{ji} = \pi P^i$, so π is indeed a stationary distribution. ■

2 Application of Markov Chains

2.1 Metropolis Hastings

Let G be a huge graph and vol some weight function on vertices of G. We want to sample vertices from G with respect to vol, so the probability of choosing vertex i is $\pi_i = \frac{\mathrm{vol}(i)}{\sum_j \mathrm{vol}(j)}$. Because G is to big to calculate probabilities and directly sample vertices, we need to create Markov Chain for walking through G.

Let d be largest degree of vertex in G. The transition matrix of Markov Chain will be as follows. The probability of going from i to neighbour j is $\frac{1}{d}\min\{1, \frac{\pi_j}{\pi_i}\} = \frac{1}{d}\min\{1, \frac{\mathrm{vol}(j)}{\mathrm{vol}(i)}\}$, and the probability of staying in i is $1 - \frac{1}{d}\sum_{j \in N(i)} \min\{1, \frac{\mathrm{vol}(j)}{\mathrm{vol}(i)}\}$.

This is a time-reversible Markov Chain, because $\pi_i P_{ij} = \frac{1}{d}\min\{\pi_i, \pi_j\} = \pi_j P_{ji}$. From Theorem 8 follows that π is a stationary distribution of Markov Chain, so random walk will converge to sampling vertices with respect to π.