Lecture 5: Random Walks and Markov Chain

Lecturer: Thomas Sauerwald & He Sun

1 Introduction to Markov Chains

Definition 5.1. *A sequence of random variables $X_0, X_1, \ldots \subseteq \Omega$ is a Markov chain with state space Ω if for all possible states $x_{t+1}, x_t, \ldots, x_1 \in \Omega$ we have*

$$\mathbf{Pr}\left[\, X_{t+1} = x_{t+1} \mid X_t = x_t \wedge X_{t-1} = x_{t-1} \wedge \cdots \wedge X_0 = x_0 \,\right] = \mathbf{Pr}\left[\, X_{t+1} = x_{t+1} \mid X_t = x_t \,\right].$$

The Markov chain is called time-homogenous if the latter probability is independent of t.

In this course, we will only consider finite Markov chains which are also time-homogenous. In this case, it is convenient to store all information about the transition in a Markov chain in the so-called transition matrix $\mathbf{P}$ which is a $|\Omega| \times |\Omega|$ matrix defined by:

$$\mathbf{P}_{x,y} := \mathbf{Pr}\left[\, X_1 = y \mid X_0 = x \,\right].$$

Properties of the transition matrix $\mathbf{P}$:

- all entries in $\mathbf{P}$ are in $[0, 1]$,
- every row sum equals 1.

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

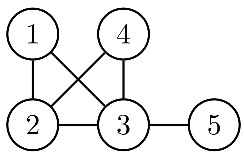

Figure 1: Example of a Markov chain corresponding to a random walk on a graph G with 5 vertices.

A very important special case is the Markov chain that corresponds to a random walk on an undirected, unweighted graph. Here, the random walk picks each step a neighbor chosen uniformly at random and moves to that neighbor. Hence, the transition matrix is $\mathbf{P} = \mathbf{D}^{-1} \cdot \mathbf{A}$, where $\mathbf{D}$ is the diagonal matrix with $\mathbf{D}_{i,i} = \frac{1}{\deg(i)}$.

The next lemma shows how to compute the probability for multiple steps.

Lemma 5.2. *For any two states $x, y \in \Omega$ and any round $t \in \mathbb{N}$,*

$$\mathbf{Pr}\left[X_t = y \mid X_0 = x \right] = \mathbf{P}^t_{x,y},$$

where $\mathbf{P}^0$ is the identity matrix.

Proof. Let μ^0 be the (row-)unit-vector which has 1 at component x and 0 otherwise (the starting distribution). Then let μ^t be the distribution of the random walk at step t. Clearly,

$$\mu^1_y = 1 \cdot \mathbf{P}_{x,y} = \sum_{z \in \Omega} \mu^0_z \cdot \mathbf{P}_{z,y},$$

and therefore,

$$\mu^1 = \mu^0 \cdot \mathbf{P},$$

and more generally,

$$\mu^t = \mu^{t-1} \cdot \mathbf{P} = \mu^0 \cdot \mathbf{P}^t.$$

□

Let us now return to the above example and see how the powers of $\mathbf{P}$ look like.

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

$$\mathbf{P}^{10} = \begin{pmatrix} 0.17 & 0.24 & 0.32 & 0.17 & 0.09 \\ 0.16 & 0.25 & 0.34 & 0.16 & 0.08 \\ 0.16 & 0.26 & 0.35 & 0.16 & 0.08 \\ 0.17 & 0.24 & 0.32 & 0.17 & 0.09 \\ 0.18 & 0.24 & 0.31 & 0.18 & 0.09 \end{pmatrix}$$

$$\mathbf{P}^{\infty} = \begin{pmatrix} 1/6 & 1/4 & 1/3 & 1/6 & 1/12 \\ 1/6 & 1/4 & 1/3 & 1/6 & 1/12 \\ 1/6 & 1/4 & 1/3 & 1/6 & 1/12 \\ 1/6 & 1/4 & 1/3 & 1/6 & 1/12 \\ 1/6 & 1/4 & 1/3 & 1/6 & 1/12 \end{pmatrix}$$

Another important special case of a Markov chain is the random walk on edge-weighted graphs. Here, every (undirected) edge $\{x, y\} \in E$ has a non-negative weight $w_{\{x,y\}}$. The transition matrix $\mathbf{P}$ is defined as

$$\mathbf{P}_{\{x,y\}} = \frac{w_{\{x,y\}}}{w_x},$$

where

$$w_x = \sum_{z \in N(x)} w_{\{x,z\}}.$$

Definition 5.3. *A distribution π is called* stationary, *if $\pi \mathbf{P} = \pi$.*

A distribution π is (time)-reversible w.r.t. $\mathbf{P}$, if it satisfies for all $x, y \in \Omega$,

$$\pi_x \mathbf{P}_{x,y} = \pi_y \mathbf{P}_{y,x}.$$

Lemma 5.4. *Let π be a probability distribution that is time-reversible w.r.t. $\mathbf{P}$. Then π is a stationary distribution for $\mathbf{P}$.*

Proof. Summing over all states $y \in \Omega$ yields

$$\sum_{x \in \Omega} \pi_x \mathbf{P}_{x,y} = \sum_{x \in \Omega} \pi_y \mathbf{P}_{y,x} = \pi_y.$$

□

Hence, if π is time-reversible w.r.t. $\mathbf{P}$, then once the distribution π is attained, the chain moves with the same frequency from x to y than from y to x.

Random walks on graphs and random walks on edge-weighted graphs are always reversible. (A simple example for a non-reversible Markov chain is a Markov chain for which there are two states with $\mathbf{P}_{x,y} > 0$ but $\mathbf{P}_{x,y} = 0$.)

Let us now compute the stationary distribution for three important examples of Markov chains:

- For a random walk on an (unweighted) graph, $\pi_x = \frac{\deg(x)}{2|E|}$:

$$\pi_x = \sum_{y \in N(x)} \frac{\deg(y)}{2|E|} \cdot \mathbf{P}_{y,x} = \sum_{y \in N(x)} \frac{\deg(y)}{2|E|} \cdot \frac{1}{\deg(y)} = \frac{\deg(x)}{2|E|}.$$

  An even more easier way to verify this is to use Lemma 5.4.

- If the transition matrix $\mathbf{P}$ is symmetric, then $\pi = (1/n, \ldots, 1/n)$ is uniform.

- For a random walk with non-negative edge-weights $w_{\{u,v\}}$, $\{u,v\} \in E$, let

$$w_G = \sum_{x \in V} w_x.$$

  Then the Markov chain is reversible with $\pi_x = w_x / w_G$ (and hence π is the stationary distribution):

$$\pi_x \mathbf{P}_{x,y} = \frac{w_x}{w_G} \cdot \frac{w_{\{x,y\}}}{w_x} = \frac{w_{\{x,y\}}}{w_G} = \frac{w_{\{y,x\}}}{w_G} = \frac{w_y}{w_G} \cdot \frac{w_{\{y,x\}}}{w_y} = \pi_y \mathbf{P}_{y,x}.$$

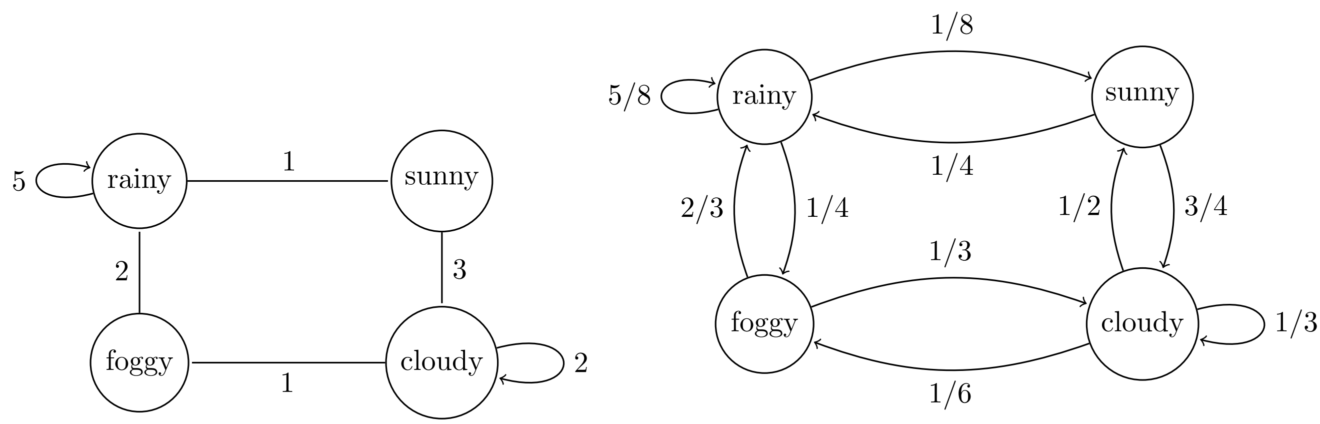

Figure 2: The Markov chain representing the weather in Saarbrücken.

The transition matrix for the weather chain (the states are ordered R, F, C, S).

$$\mathbf{P} = \begin{pmatrix} 5/8 & 1/4 & 0 & 1/8 \\ 2/3 & 0 & 1/3 & 0 \\ 0 & 1/6 & 1/3 & 1/2 \\ 1/4 & 0 & 3/4 & 0 \end{pmatrix}$$

By using the above formula for edge-weighted random walks, we can compute the stationary distribution: $\pi = (8/21, 3/21, 6/21, 4/21)$ (note that 21 is the sum of all weights of the vertices, where every edge weight is counted twice, but every loop weight is counted once)

The naive approach to compute $\mathbf{P}^t$ numerically involves $\approx \log_2(t)$ matrix multiplications (fastest algorithm $\mathcal{O}(n^{2.376})$). Using the reversibility, we now show how to do it using only a constant number of matrix multiplications (assuming that we can efficiently compute the eigenvectors – for very small t it might be actually more efficient to compute $\mathbf{P}^t$ directly). Moreover, the approach will give us a precise formula for any entry $\mathbf{P}^t_{x,y}$ as a function of t.

Since the matrix $\mathbf{P}$ is reversible w.r.t. to π, it follows that the matrix

$$\mathbf{N}_{x,y} := \pi_x^{1/2} \mathbf{P}_{x,y} \pi_y^{-1/2},$$

is symmetric. Let us rewrite this using diagonal matrices:

$$\mathbf{N} = \mathbf{D}_\pi^{1/2} \mathbf{P} \mathbf{D}_\pi^{-1/2}.$$

Hence the matrix $\mathbf{N}$ can be diagonalized, i.e., there is an orthogonal matrix $\mathbf{R}$ consisting of the eigenvalues of $\mathbf{N}$ so that $\mathbf{R}\mathbf{N}\mathbf{R}^{-1} = \mathbf{D}$, where $\mathbf{D}$ is the diagonal matrix with eigenvalues $\lambda_1 \geqslant \lambda_2 \geqslant \cdots \geqslant \lambda_n$ of $\mathbf{N}$. Therefore, we can compute powers of the matrix $\mathbf{P}$ in the following way:

$$\begin{aligned} \mathbf{P}^t &= \left(\mathbf{D}_\pi^{-1/2} \mathbf{N} \mathbf{D}_\pi^{+1/2}\right)^t \\ &= \left(\mathbf{D}_\pi^{-1/2} \mathbf{R}^{-1} \mathbf{D} \mathbf{R} \mathbf{D}_\pi^{+1/2}\right)^t \\ &= \mathbf{D}_\pi^{-1/2} \mathbf{R}^{-1} \cdot \mathbf{D}^t \cdot \mathbf{D}_\pi^{+1/2} \mathbf{R}^1 \end{aligned}$$

Definition 5.5. *A Markov chain is called aperiodic if for all states $x \in \Omega$:*

$$gcd(\{t \in \mathbb{N} \colon \mathbf{P}^t_{x,x} > 0\}) = 1.$$

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$\mathbf{P}^2 = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

$$\mathbf{P}^3 = \begin{pmatrix} 1/2 & 0 & 0 & 1/2 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1/2 & 0 & 0 & 1/2 \end{pmatrix}$$

$$\mathbf{P}^4 = \mathbf{P}^1 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1/2 & 0 & 0 & 1/2 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

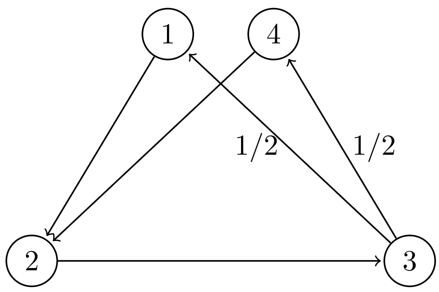

Figure 3: Example of a Markov chain with period 3. The set $\{t \in \mathbb{N} \colon \mathbf{P}^t_{1,1} > 0\}$ contains only multiples of 3.

Definition 5.6. *A Markov chain is irreducible if for any two states $x, y \in \Omega$ there exists an integer t such that $\mathbf{P}^t_{x,y} > 0$.*

Intuitively, a Markov chain is called irreducible if it is connected viewed as a graph.

The following Markov chain corresponds to a random walk on the (unconnected) graph which consists of two disjoint cycles of length 2. Hence the transition matrix $\mathbf{P}$ below is not irreducible.

$$\mathbf{P} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

Note that $\mathbf{P}$ has two stationary distributions: $\pi = (1/2, 1/2, 0, 0)$ and $\pi = (0, 0, 1/2, 1/2)$.

Theorem 5.7 (Fundamental Theorem for Markov Chains)**.** *If a Markov chain is aperiodic and ireducible, then it holds for all states $x, y \in \Omega$ that:*

- *there exists an integer $T = T(x, y) \in \mathbb{N}$ so that for all $t \geqslant T$, $\mathbf{P}^t_{x,y} > 0$.*

- *in the limit,*

$$\lim_{t\to\infty} \mathbf{P}^t_{x,y} = \pi_y = \frac{1}{\mathbf{E}\left[\,\tau_{y,y}\,\right]},$$

where $\tau_{y,y} = \min\{t \in \mathbb{N} \setminus \{0\} \colon X_t = y, X_0 = y\}$ *is the first return to* y*.*

Hence for large t, the matrix $\mathbf{P}^t$ has value π_y in column y. Moreover, note that $\pi_y = \frac{1}{\mathbf{E}[\,\tau_{y,y}\,]}$ means that the proportion of steps spent in state y is equal to the expected waiting times between two consecutive visits.

2 Speed of Convergence

We now focus on random walks on graphs. We assume $G = (V, E)$ is an undirected, unweighted and connected graph. For simplicity, we also assume that G is d-regular. Recall that $\mathbf{M} = \mathbf{D}^{-1}\mathbf{A}$ is the normalized adjacency matrix which will be transition matrix of the random walk on G, so $\mathbf{P} = \mathbf{M}$. Since $\mathbf{P}$ is a symmetric matrix, $\pi = (1/n, \ldots, 1/n)$ is the stationary distribution.

Lemma 5.8. *Let* $\mathbf{P}$ *be any symmetric transition matrix. Then for any probability vector* x*:*

$$\left\| x\mathbf{P}^t - \pi \right\|_2 \leqslant \lambda^t,$$

where $\pi = (1/n, \ldots, 1/n)$ *is the uniform vector and* $\lambda = \max\{|\lambda_2|, |\lambda_n|\}$*.*