Randomised Algorithms: Markov Chains

February 2023

1 Properties of Markov Chains

Exercise 1 [Chapman-Kolmogorov Equations] Consider a Markov chain (ρ^0, P). Let $\rho^t \in \mathbb{R}$ be such that $\rho_i^t = \mathbf{Pr}\left[\, X^t = i \,\right]$. Then, show that for any $t \geq 0$,

$$\rho^t = \rho^0 \cdot P^t.$$

(*Answer*) We will prove the claim using induction over $t \geq 0$. For $t = 0$, it is follows trivially from the definition of ρ^0. For $t \geq 1$, assuming it holds for $t-1$, we have that

$$\begin{aligned}
\mathbf{Pr}\left[\, X^t = i \,\right] &= \sum_{j \in \Omega} \mathbf{Pr}\left[\, X^t = i, X^{t-1} = j \,\right] \quad \text{(law of total probability)} \\
&= \sum_{j \in \Omega} \mathbf{Pr}\left[\, X^t = i \mid X^{t-1} = j \,\right] \cdot \mathbf{Pr}\left[\, X^{t-1} = j \,\right] \quad \text{(properties of cond. probability)} \\
&= \sum_{j \in \Omega} P_{ji} \cdot \rho_j^{t-1} \quad \text{(definitions of } P \text{ and } \rho^{t-1}\text{)} \\
&= \rho^{t-1} \cdot P \quad \text{(definition of matrix multiplication)} \\
&= (\rho^0 \cdot P^{t-1}) \cdot P \quad \text{(by the induction hypothesis} \\
&= \rho^0 \cdot P^t.
\end{aligned}$$

Exercise 2 [Two-state MC] Consider the Markov Chain with (μ, P) with state space $\Omega = \{0, 1\}$ and transition matrix

$$\begin{bmatrix} 1-p & p \\ q & 1-q \end{bmatrix},$$

for some $p, q \in (0, 1)$. By induction or otherwise, prove that

$$P^n = \frac{1}{p+q} \cdot \begin{bmatrix} q + p(1-p-q)^n & p - p(1-p-q)^n \\ q - q(1-p-q)^n & p + q(1-p-q)^n \end{bmatrix}.$$

Show that

$$\lim_{n \to \infty} P^n = \frac{1}{p+q} \cdot \begin{bmatrix} q & p \\ q & p \end{bmatrix}.$$

1.1 Communicating states

Exercise 3 Consider a Markov chain with transition matrix P and state space Ω. Let $G = (\Omega, E)$ be the graph where there is an edge $(x, y) \in E$ iff $P(x, y) > 0$. State x is *accessible* from state y if there is a path from x to y in G. We say that i and j *communicate* (and write $i \leftrightsquigarrow j$) iff i is accessible from j and the other way round. Prove that for any $i, j, k \in \Omega$

(a) $i \leftrightsquigarrow i$ (**reflexive**).

(b) If $i \leftrightsquigarrow j$, then also $j \leftrightsquigarrow i$ (**symmetry**).

(c) If $i \leftrightsquigarrow j$ and $j \leftrightsquigarrow k$, then $i \leftrightsquigarrow k$ (**transitive**).

Exercise 4 [Connection to Strongly Connected Components] Consider a finite state Markov chain. Argue that the communicating classes are exactly the *strongly connected components* of the graph with vertices Ω and edges $uv \in E$ iff P^t_{uv}.

Exercise 5 (optional) How do the communicating classes of the Markov chain (or equivalently the induced strongly connected components of the graph) relate to *recurrent states*, i.e. states that will be visited infinitely often?

1.2 Periodic/Aperiodic Markov Chains

The following exercises explore some core properties of periodicity of Markov chains, but they are not very easy to prove (relying on number theory) and can be skipped.

Exercise 6 (++) Consider an irreducible finite Markov chain with state space Ω and transition matrix P. Let $T(x) = \{t \geq 1 : P^t(x, x) > 0\}$ be the set of all time steps that a Markov chain can return to x, having started at x. Then, the period of x is defined as $\gcd T(x)$.
Show that for any $x, y \in \Omega$, it holds that $\gcd T(x) = \gcd T(y)$.

(*Answer*) See Lemma 4.2 in this paper.

Exercise 7 (++) For any aperiodic and irreducible Markov chain with state space Ω and transition matrix P, there exists $r \geq 0$ such that for any $x, y \in \Omega$,

$$P^r(x, y) > 0.$$

(*Answer*) See Lemma 4.4 and 4.5 in this paper.

Exercise 8 Prove that in a periodic graph with period k it is possible to split the vertices into k disjoint sets $C_0, \ldots, C_{k-1}$ such that each edge $(u, v) \in E$ satisfies $u \in C_i$ and $v \in C_{i+1 \mod \pmod{k}}$.

1.3 Stationary distributions

Exercise 9 [Non-unique stationary distribution] Show that there is a non-irreducible Markov chain with more than one stationary distributions.

(*Answer*) Consider the Markov chain with states $\Omega = \{0, 1\}$ and transition matrix $P = I$. Then, the following distributions are stationary:

$$\pi_A = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

and

$$\pi_B = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

More generally, if the graph consists of k disjoint irreducible components, then each of these could have a different stationary distribution.

Exercise 10 [Exact convergence] Most Markov chains covered in this course never reach a stationary distribution exactly, but only get arbitrarily close. Can you find an irreducible Markov chain with n states such that for any starting state x there is an integer t such that $P^t_x = \pi$?

(*Answer*) Consider the simple random walk on the complete graph (with self-loops) with n vertices. The transition matrix satisfies $P_{ij} = \frac{1}{n}$ for any vertices $i, j \in [n]$.

Then, in this graph the stationary distribution is the uniform distribution $\pi_i = \frac{1}{n}$ for any $i \in [n]$. Then, this distribution satisfies

$$\pi_i = \sum_{i=1}^{n} \pi_j \cdot P_{ji} = \sum_{i=1}^{n} \frac{1}{n} \cdot \frac{1}{n} = \frac{1}{n}.$$

But starting from any vertex x after one step we could end up in any vertex y with probability $1/n$. Hence, this MC reaches the stationary distribution after exactly one step.
Question: Do there exist any other examples?

Exercise 11 [Singleton cut-set] Show that for any finite state Markov chain with stationary distribution π, we have that for any state $x \in \Omega$

$$\sum_{y \in \Omega : y \neq x} \pi_y P_{y,x} = \sum_{y \in \Omega : x \neq y} \pi_x P_{x,y}.$$

(*Answer*)

$$\sum_{y \in \Omega : y \neq x} \pi_y P_{y,x} = \sum_{y \in \Omega} \pi_y P_{y,x} - \pi_x P_{x,x} = \pi_x - \pi_x P_{x,x} = \pi_x \cdot (1 - P_{x,x}) = \pi_x \cdot \sum_{y \in \Omega : x \neq y} P_{x,y} = \sum_{y \in \Omega : x \neq y} \pi_x P_{x,y}.$$

Exercise 12 [Cut-set of Markov chain] (+) Consider a finite, irreducible and aperiodic Markov chain. Then, for any $S \subseteq \Omega$, we have that,

$$\sum_{x \in S} \sum_{y \in S^c} \pi_x P_{xy} = \sum_{y \in S} \sum_{x \in S^c} \pi_y P_{yx}.$$

This means that the probability that the chain leaves the set S is equal to the probability that the chain enters S.

Exercise 13 [Time-reversible Markov chains] This property is very useful for determining the stationary distribution of a Markov chain (see Exercises 23, 20, 42, 31).
Consider a finite irreducible Markov chain with transition matrix P. Assume that for some $\tilde{\pi}$ with $\sum_{x \in \Omega} \tilde{\pi}_x = 1$ and $\tilde{\pi}_x \geq 0$, we have that for any $x, y \in \Omega$

$$\tilde{\pi}_x P_{xy} = \tilde{\pi}_y P_{yx}.$$

Then, $\tilde{\pi}$ is the stationary distribution of the Markov chain.

(*Answer*) We just need to verify that

$$\tilde{\pi} P = \tilde{\pi},$$

and so it will follow by the uniqueness that $\tilde{\pi}$ is the stationary distribution. For any state $y \in \Omega$,

$$(\tilde{\pi} P)_y = \sum_{x \in \Omega} \tilde{\pi}_x P_{xy} = \sum_{x \in \Omega} \tilde{\pi}_y P_{yx} = \tilde{\pi}_y \sum_{x \in \Omega} P_{yx} = \tilde{\pi}_y.$$

Extended Note 1 [Verifying the stationary distribution] Exercise 13 gives us a way to quickly verify that a probability vector is the stationary distribution of an irreducible Markov chain: we just have to verify that it satisfies $\pi_x P_{xy} = \pi_y P_{yx}$.

Exercise 14 [Time-reversal] This exercise explains where the name time-reversal comes from (Probability & Computing Ex. 7.13).
Consider a finite Markov chain with stationary distribution π and transition matrix P. Imagine starting the chain at time 0 and running it for m steps, obtaining the sequence $X_0, \ldots, X_m$. Consider the states in reverse order, $X_m, X_{m-1}, \ldots, X_0$.

(a) Argue that given X_{k+1}, the state X_k is independent of $X_{k+2}, \ldots, X_m$. Thus the reverse sequence is Markovian.

(b) Argue that for the reverse sequence, the transition probabilities $Q_{x,y}$ are given by

$$Q_{x,y} = \frac{\pi_y P_{y,x}}{\pi_x}.$$

(c) Prove that if the original Markov chain is time reversible, so that $\pi_x P_{x,y} = \pi_y P_{y,x}.$, then $Q_{x,y} = P_{x,y}$.

(d) Based on your answer in (c), argue about the name *time-reversible*.

(*Answer*)

(a) Consider any states $x, y, z \in \Omega$. Then,

$$\begin{aligned}
\mathbf{Pr}[X_{k+2} = z, X_k = x \mid X_{k+1} = y] &= \frac{\mathbf{Pr}[X_{k+2} = z, X_{k+1} = y, X_k = x]}{\mathbf{Pr}[X_{k+1} = y]} \\
&= \frac{\mathbf{Pr}[X_{k+2} = z, X_{k+1} = y, X_k = x]}{\mathbf{Pr}[X_{k+1} = y, X_k = x]} \cdot \frac{\mathbf{Pr}[X_{k+1} = y, X_k = x]}{\mathbf{Pr}[X_{k+1} = y]} \\
&= \mathbf{Pr}[X_{k+2} = z \mid X_{k+1} = y, X_k = x] \cdot \mathbf{Pr}[X_k = x \mid X_{k+1} = y] \\
&= \mathbf{Pr}[X_{k+2} = z \mid X_{k+1} = y] \cdot \mathbf{Pr}[X_k = x \mid X_{k+1} = y],
\end{aligned}$$

which shows that the X_k is independent of X_{k+2} given X_{k+1}.

(b) By Bayes' rule, for any two states $x, y \in \Omega$ we have that

$$\mathbf{Pr}[X_{k+1} = y \mid X_k = x] \cdot \mathbf{Pr}[X_k = x] = \mathbf{Pr}[X_k = x \mid X_{k+1} = y] \cdot \mathbf{Pr}[X_{k+1} = y].$$

Writing in terms of P_{xy} and Q_{yx} we have that

$$P_{xy} \cdot \mathbf{Pr}[X_k = x] = Q_{yx} \cdot \mathbf{Pr}[X_{k+1} = y]. \tag{1}$$

Since the Markov chain has a stationary distribution we have that

$$\pi_x = \lim_{k \to \infty} \mathbf{Pr}[X_k = x] = \lim_{k \to \infty} \mathbf{Pr}[X_{k+1} = x].$$

Therefore, by taking limits on both sides in eq. (1), we have that

$$\lim_{k \to \infty} P_{x,y} \cdot \mathbf{Pr}[X_k = x] = \lim_{k \to \infty} Q_{y,x} \cdot \mathbf{Pr}[X_{k+1} = y] \Leftrightarrow P_{x,y}\pi_x = Q_{yx}\pi_y \Leftrightarrow Q_{y,x} = \frac{\pi_x P_{x,y}}{\pi_y}.$$

Exercise 15 [Ehrenfest Chain] Recalling the Ehrenfest Chain defined on **Lecture 4 slide 3**.

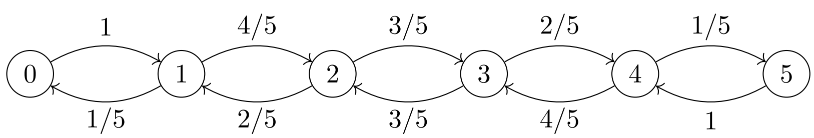
(a) Draw the Ehrenfest Chain for $d = 5$.

(b) What is the stationary distribution of the Ehrenfest Chain?

(c) Does the Ehrenfest Chain converge to the stationary distribution?

(*Answer*)

(a) Here is a drawing for the chain with $d = 5$.



(b) (**Method 1**) The first method is to try to find a distribution π that satisfies the time-reversibility condition, i.e., for any two states $x, y \in \Omega$,

$$\pi_x P_{x,y} = \pi_y P_{y,x}.$$

For the Ehrenfest Chain, we only need to verify this for $y = x + 1$. So we need to find π such that

$$\pi_x P_{x,x+1} = \pi_{x+1} P_{x+1,x},$$

which is equivalent to

$$\pi_x \cdot \frac{d-x}{d} = \pi_{x+1} \cdot \frac{x+1}{d},$$

and to

$$\pi_{x+1} = \pi_x \cdot \frac{d-x}{x+1}.$$

By applying this recurrence x times, we get that

$$\pi_{x+1} = \pi_0 \cdot \prod_{k=0}^{x} \frac{d-k}{k+1} = \pi_0 \cdot \frac{(d-x) \cdot (d-x-1) \cdot \ldots \cdot d}{x \cdot (x-1) \cdot \ldots \cdot 1} = \pi_0 \cdot \frac{d!}{k!(d-k)!} = \pi_0 \cdot \binom{d}{k}.$$

To find out π_0, we just need to use that π is a distribution so $\sum_{x \in \Omega} \pi_x = 1$ and so

$$1 = \pi_0 \cdot \sum_{x=0}^{d} \binom{d}{x} = \pi_0 \cdot 2^d.$$

Hence,

$$\pi_x = 2^{-d} \cdot \binom{d}{x}.$$

(**Method 2**) The second method is to "guess" the answer and verify that it satisfies the condition $\pi = \pi P$. It is "reasonable" to assume that this will converge to the binomial distribution and so by normalising, we get want to try

$$\pi_x = 2^{-d} \cdot \binom{d}{x}.$$

Now we verify the condition $\pi = \pi P$ by considering the following cases:

- **Case 1 [$1 \leq x \leq d-1$]:**

$$\begin{aligned}
(\pi P)_x &= \pi_{x+1} P_{x+1,x} + \pi_{x-1} P_{x-1,x} \\
&= \frac{1}{2^d} \cdot \binom{d}{x+1} \frac{x+1}{d} + \frac{1}{2^d} \cdot \binom{d}{x-1} \frac{d-x+1}{d} \\
&= \frac{1}{2^d} \cdot \frac{d!}{(x+1)!(d-x-1)!} \cdot \frac{x+1}{d} + \frac{1}{2^d} \cdot \frac{d!}{(x-1)!(d-x+1)!} \frac{d-x+1}{d} \\
&= \frac{1}{2^d} \cdot \frac{(d-1)!}{x!(d-x-1)!} \cdot \left(\frac{1}{x} + \frac{1}{d-x} \right) \\
&= \frac{1}{2^d} \cdot \frac{(d-1)!}{x!(d-x-1)!} \cdot \frac{d}{x \cdot (d-x)} \\
&= \frac{1}{2^d} \cdot \frac{d!}{x!(d-x)!} = \frac{1}{2^d} \cdot \binom{d}{x}.
\end{aligned}$$

- **Case 2 [$x = 0$]:**

$$(\pi P)_0 = \pi_1 \cdot P_{1,0} = \frac{1}{2^d} \cdot d \cdot \frac{1}{d} = \frac{1}{2^d} = \pi_0.$$

- **Case 2 [$x = d$]:**

$$(\pi P)_d = \pi_{d-1} \cdot P_{d-1,d} = \frac{1}{2^d} \cdot d \cdot \frac{1}{d} = \frac{1}{2^d} = \pi_d.$$

(**Method 3**) In the lectures you showed that for the d-dimensional hypercube, the stationary distribution is the uniform distribution $\pi_x = \frac{1}{2^d}$, where $x \in \{0,1\}^d$.

We can interpret $x \in \{0,1\}^d$ the vector representing the colours of each of the particles and then there are d possible transitions between these as in the hypercube. The stationary distribution for having k particles of one colour is just the sum over all states with that number of particles. Since there are $\binom{d}{k}$ of these, we conclude that in the Ehrenfest chain

$$\pi_k = 2^{-d} \cdot \binom{d}{k}.$$

Exercise 16 Let X_n be the sum of n independent rolls of a fair die. Show that for any $k \geq 2$,

$$\lim_{n\to\infty} \mathbf{Pr}\left[\, X_n \text{ is divisible by } k \,\right] = \frac{1}{k}.$$

(*Answer*) We construct a Markov chain $(X_n)_{n\geq 0}$, where X_n gives us the remainder of the sum of the first n die when divided by k. So, the state space $\Omega = \{0, \ldots, k-1\}$ and initially $X_0 = 0$. From each state $x \in \Omega$, there are 6 possible transitions to $x + 1 \pmod{k}$ (if the outcome of the die is 1), to $x + 2 \pmod{k}$ (if the outcome of the die is 2), ... , to $x + 6 \pmod{k}$ (if the outcome of the die is 6).
The graph of the Markov chain is symmetric to the states and hence the stationary distribution satisfies $\pi_x = \frac{1}{k}$ for all states $x \in \Omega$, and one can verify that this satisfies

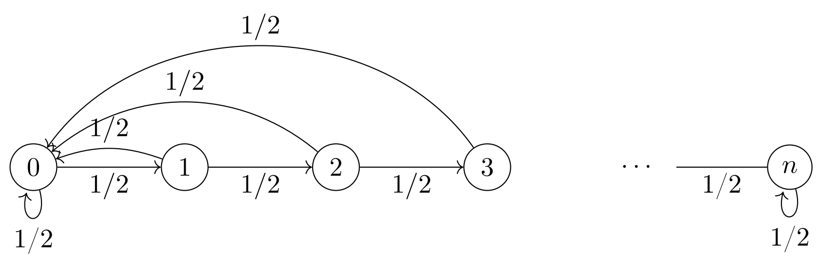
$$\pi_x = \sum_{i=1}^{6} \pi_{(x+i) \pmod 6} \cdot \frac{1}{6}.$$

Exercise 17 Consider a Markov chain on the states $\Omega = \{0, 1, \ldots, n\}$, where for $i < n$ we have $P_{i,i+1} = 1/2$ and $P_{i,0} = 1/2$. Also, $P_{n,n} = 1/2$ and $P_{n,0} = 1/2$. This process can be viewed as a random walk on a directed graph with vertices $\{0, 1, \ldots, n\}$, where each vertex has two directed edges: one that returns to 0 and one that moves to the vertex with the next higher number (with a self-loop at vertex n).

(a) Find the stationary distribution of this chain.

(b) (optional) What does this say about random walks in directed graphs?

(*Answer*) This chain looks as follows:



We need to find a probability distribution π satisfying the condition $\pi P = \pi$. For any $1 \leq i \leq n-1$, we have that

$$\pi_i = \frac{1}{2}\pi_{i-1}.$$

By applying the recurrence equation for i times, we get

$$\pi_i = \frac{1}{2^i} \cdot \pi_0.$$

For $i = n$, we get that

$$\pi_n = \frac{1}{2}\pi_n + \frac{1}{2}\pi_{n-1} \Rightarrow \pi_n = \pi_{n-1}.$$

Now, since π is a probability distribution,

$$\sum_{i=0}^{n} \pi_i = 1 \Rightarrow \pi_0 \cdot \left(1 + \frac{1}{2} + \ldots + \frac{1}{2^{n-1}} + \frac{1}{2^{n-1}}\right) = 1 \Rightarrow \pi_0 = \frac{1}{2}.$$

Therefore,

$$\pi_i = \begin{cases} \frac{1}{2^{i+1}} & \text{if } 1 \leq i \leq n-1, \\ \frac{1}{2^n} & \text{otherwise.} \end{cases}$$

Exercise 18 Let $(X_n)_{n\geq 1}$ be a Markov chain on the states $\{0,1,2\}$ with transition matrix

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1-p & p \\ 1-p & p & 0 \end{bmatrix},$$

where $p \in (0,1)$.

(a) Draw the state space diagram for the Markov Chain X_n.

(b) Explain why X_n is an irreducible and aperiodic Markov Chain.

(c) Determine the stationary distribution of the Markov Chain.

[Source: [2017P6Q8]]

Exercise 19 Attempt **[2022P8Q12 (a),(b)(i)]**.

Exercise 20 [Top-to-Random] Consider the Top-to-Random shuffle defined on **Lecture 4 slide 23**.

(a) Define the Markov chain corresponding to the shuffle.

(b) Argue that the Markov chain is irreducible.

(c) Argue that the Markov chain is aperiodic.

(d) Find the stationary distribution of the chain.

(*Answer*)

(a) The Markov chain has as states Ω all possible $n!$ permutations of the n cards. For any permutation $x = (x_1, \ldots, x_n)$, there are n connections to permutations $(x_1, \ldots, x_n)$ (a self-loop), $(x_2, x_1, \ldots, x_n)$, $(x_2, x_3, x_1 \ldots, x_n)$, and so on. Each such connection carries a $1/n$ probability.

(b) The Markov chain is irreducible since it is possible to sort the permutation using insertion sort and also reach any other permutation from the sorted one (using the reverse moves). Hence, we can reach any permutation x from permutation y via the sorted permutation.

(c) The chain is aperiodic since each permutation has a self-loop.

(d) We will show that the uniform distribution π with $\pi_x = 1/n!$ satisfies the stationary distribution condition,

$$(\pi P)_x = \sum_{y \in N(x)} \pi_y P_{y,x} = \sum_{y \in N(x)} \frac{1}{n!} \cdot \frac{1}{n} = \frac{1}{n!} = \pi_x.$$

Exercise 21 [Riffle shuffle] Consider the riffle shuffle defined on **Lecture 4 slide 23**.

(a) Define the state space of the Markov chain.

(b) (+) Consider the riffle operation. Given two decks of cards A and B with a and b cards, at each step, the next card is chosen from A with probability $\frac{a}{a+b}$ and otherwise from B. Prove that when starting with n cards in total, drawing n cards using the above operation results into a uniform distribution over all permutations such that the subsequences of cards in A (and in B, respectively) are ordered increasingly.

(c) Deduce the transition probabilities of the Markov chain.

(d) Argue that the Markov chain is irreducible.

(e) Find the stationary distribution of this chain.

(*Answer*)

(a) The state space consists of all $n!$ possible permutations and there is a transition

(b) It all boils down to computing the probability of encountering a sequence of the form $\alpha\beta\beta\alpha\ldots$ (telling us from which of the two piles to sample). The probability that the i-th character is the k-th α contributes a factor of

$$\frac{a-k}{i},$$

and similarly the i-th character being the k-th β contributes a factor of

$$\frac{b-k}{i}.$$

Hence, the probability of encountering any such sequence consisting of a $\alpha's$ and b β's is given by

$$\frac{1}{(a+b)\cdot(a+b-1)\cdot\ldots\cdot 2\cdot 1}\cdot(a\cdot(a-1)\cdot\ldots 2\cdot 1)\cdot(b\cdot(b-1)\cdot\ldots\cdot 2\cdot 1)=\frac{a!b!}{(a+b)!}=\frac{1}{\binom{a+b}{a}}.$$

(c) Therefore, transition probabilities are all equal to

$$q=\sum_{i=0}^{n}2^{-n}\cdot\binom{n}{i}\cdot\frac{1}{\binom{n}{i}}=\frac{n}{2^{-n}}.$$

(d) By cutting at one card from A we can move that card (with some non-zero probability) to an arbitrary position. Hence, by following the steps of insertion sort, we can reach the sorted permutation and then from the sorted permutation we can go to any other permutation following the steps in reverse. Hence, the chain is irreducible.

Question: What would change if we deterministically split the deck in half?

(e) Again, we will verify that the uniform distribution distribution π with $\pi_x = 1/n!$ satisfies the stationary distribution condition

$$(\pi P)_x=\sum_{y\in N(x)}\pi_y\cdot P_{y,x}=\sum_{y\in N(x)}\frac{1}{n!}\cdot\frac{n}{2^{-n}}=\frac{1}{n!}.$$

1.4 Omitted proofs

Further Reading 1 [Theorems for finite MCs] This paper contains relatively short proofs for the major theorems for finite Markov chains: (i) existence, (ii) uniqueness and (iii) convergence.

Further Reading 2 [2D and 3D Random walks] See Lecture 6 in Part IB Markov Chains course from the Maths department, to find out why a drunk person finds his way home, but not a drunk bird.

2 Classical Markov Chains

Exercise 22 [Birth-Death chains] The *birth-death* Markov chains have state space $\Omega = \mathbb{N}$ and transition probability matrix P satisfying

$$\begin{aligned} P_{i,i-1} &= p \quad \text{for } i = 1,2,\ldots \\ P_{i,i+1} &= 1-p \quad \text{for } i = 0,1,\ldots \\ P_{0,0} &= p, \end{aligned}$$

for some parameter $p \in (0,1)$.

(a) Draw the first 3 states of the Markov chain. Why is it called a *birth-death* MC?

(b) Show that the chain is irreducible.

(c) Show that the chain is aperiodic.

(d) For any $p > 1/2$, find the stationary distribution.

[Source: [2009P4Q10 (a)]]

Exercise 23 [Bernoulli-Laplace Chain] Consider two urns one with j balls and the other with k balls, having in total r red balls (and the other being red). In each step, we select uniformly at random one ball from each urn and swap them.

(a) Design a Markov chain for X the number of red balls in urn 1, giving the state space and the transition matrix.

(b) Prove that the chain is irreducible.

(c) Prove that the chain is aperiodic.

(d) (+) Using Exercise 13 or otherwise, prove that the stationary distribution π is given by

$$\pi_x = \frac{\binom{r}{x} \cdot \binom{j+k-r}{k-x}}{\binom{j+k}{k}}.$$

Exercise 24 [Gambler's ruin] In the Gambler's ruin game there are two players A and B, starting with fortunes a and b respectively. In each step player A wins 1 with probability p and looses 1 with probability $1-p$. Let X^t be the amount that player A has lost after t rounds (so $X^0 = 0$). Player A will get broke when $X^t = -a$, while player B will get broke when $X^t = b$.
In this exercise, we are interested in the expected length $\ell(x)$ of a game starting with player A having lost x games.

(a) Define the Markov chain for this problem.

(b) Show that for $p = 1/2$ and $\ell(x)$ satisfying for any $-a < x < b$,

$$\ell(x) = \frac{1}{2}\ell(x-1) + \frac{1}{2}\ell(x+1) + 1.$$

(c) Show that for $p = 1/2$,

$$\ell(x) = (x+a) \cdot (b-x).$$

(d) (+ optional) Find a similar recurrence relation for $p \neq 1/2$, and deduce the expected length.

3 Applications

3.1 Algorithms

Exercise 25 [2-SAT] Prove rigorously the claim made in lecture that the expected time for RAND 2-SAT to find a given solution is at most the hitting time $h(0, n)$ of the random walk on a path.

Exercise 26 [3-SAT]

(a) Design an algorithm for the 3-SAT problem similar to the one for 2-SAT described in the lectures.

(b) Let h_j be the number of expected steps to output a solution given that j of the literals are correct. Then, show that h_j satisfies

$$\begin{aligned} h_n &= 0, \\ h_j &= \frac{2}{3}h_{j-1} + \frac{1}{3}h_{j+1} + 1, \quad 1 \le j \le n-1, \\ h_0 &= h_1 + 1. \end{aligned}$$

(c) Prove by induction or otherwise that $h_j = 2^{n+2} - 2^{j+2} - 3(n-j)$.

(d) Does the resulting algorithm have a good running time?

Exercise 27 [k-SAT]

(a) Generalise the algorithm for the 2-SAT problem to k-SAT.

(b) (+) What is the expected running time as a function of k? How does this compare to the brute force algorithm?

Exercise 28 [Faster 3-SAT] (+) Read 176-178 in "Probability and Computing" to see how to improve the expected running time for the 3-SAT algorithm (of Exercise 26) to $\mathcal{O}(n^{3/2} \cdot (4/3)^n)$ time.

Exercise 29 [3-Colourability] A *colouring* of a graph is an assignment of a colour to each of its vertices. A graph is k-colourable if there is a colouring of the graph with k colours such that no two adjacent vertices have the same colour. Let G be a 3-colourable graph.

(a) Show that there exists a colouring of the graph with two colours such that no triangle is monochromatic. (A triangle of a graph G is a subgraph of G with three vertices which are all adjacent to one another).

(b) Consider the following algorithm for colouring the vertices of G with two colours so that no triangle is monochromatic. The algorithm begins with an arbitrary 2-colouring of G. While there are any monochromatic triangles in G, the algorithm chooses one such triangle and changes the colour of a randomly chosen vertex of that triangle. Derive an upper bound on the expected number of such recolouring steps before the algorithm finds a 2-colouring with the desired property.

Exercise 30 [Reachability in log-space] The reachability problem is the following: Given an undirected graph $G = (V, E)$ and a pair of vertices $s, t \in V$, we want to determine if there is a path from s to t in G.

(a) Argue that the standard DFS/BFS solution requires $\Theta(n)$ space.

(b) Argue that running a random walk with at most n^k (with constant k) steps, requires $\Theta(\log n)$ space.

(c) Design an algorithm based on random walks which is able to answer the reachability problem with high probability.

3.2 Sampling

Exercise 31 [MCMC] In this exercise, you will show that modifying the random walk by adding self-loops with appropriate probability, we obtain a Markov chain with the uniform stationary distribution. In particular, given a graph $G = (V, E)$ and letting $N(u)$ be the set of neighbours of $u \in V$, show that the Markov chain with transition matrix

$$P_{u,v} = \begin{cases} 1/M & \text{if } u \neq v \text{ and } v \in N(u), \\ 0 & \text{if } u \neq v \text{ and } v \notin N(u), \\ 1 - N(u)/M & \text{if } u = v, \end{cases}$$

for any $M \geq |V|$ has the uniform stationary distribution. *Hint: You may want to use Exercise 13.*

(*Answer*) By Exercise 13, we just need to verify that for any vertices $u, v \in V$, $\frac{1}{|V|} P_{u,v} = \frac{1}{|V|} P_{v,u}$. Which is trivial, because $P_{u,v} = P_{v,u} = 1/M$.

Exercise 32 [Sampling an independent set] In this exercise, you will analyse the Markov chain algorithm for sampling an independent set (**Lecture 4 slide 25**).

(a) Define the Markov chain for this setting.

(b) Argue that the chain is irreducible. *Hint: Is there a state that can be reached from all states?*

(c) Argue that the chain is aperiodic.

(d) Using Exercise 31, argue that the uniform distribution is the stationary distribution of the Markov chain.

(*Answer*)

(a) The states are the independent sets $\mathcal{I}$ of the graph. For each non-empty (independent set) set $S \in \mathcal{I}$ there is $\frac{1}{|S|}$ probability to transition to any

(b) Every state can reach the empty state by removing the vertices one by one. And by the reverse process every independent set can be reached from the empty set. Hence, every independent set can reach every other, by going through the empty set.

(c) If the graph has at least one edge (u, v), then the state $\{u\}$ has a self-loop and since the chain is irreducible it is also aperiodic.

Exercise 33 [Metropolis Algorithm] In this exercise, we will generalise Exercise 31, so that given a connected graph $G = (V, E)$ and a distribution π over V, we create a Markov chain over V to have π as its stationary distribution.
Show that the chain with transition matrix

$$P_{u,v} = \begin{cases} (1/M) \cdot \min(1, \pi_v/\pi_u), & \text{if } u \neq v \text{ and } v \in N(u), \\ 0, & \text{if } u \neq v \text{ and } v \notin N(u), \\ 1 - \sum_{v \neq u} P_{u,v}, & \text{if } u = v. \end{cases}$$

Why would we want to do this?

(*Answer*) Again, we will use Exercise 13 to prove this. Consider u and $v \in N(u)$ with $\pi_u \geq \pi_v$, then

$$\pi_u \cdot P_{u,v} = \pi_u \cdot \frac{1}{M} \cdot \frac{\pi_v}{\pi_u} = \pi_v \cdot \frac{1}{M} = \pi_v \cdot P_{v,u}.$$

The case $\pi_u < \pi_v$ follows similarly.

Exercise 34 Combining Exercises 32 and 33, design a Markov chain whose stationary distribution is the exponential distribution over the independent sets $\mathcal{I}$ of a graph $G = (V, E)$, where for any $I \in \mathcal{I}$

$$\pi_I = \frac{e^{\lambda|I|}}{\sum_{I' \in \mathcal{I}} e^{\lambda|I'|}},$$

for any $\lambda \in \mathbb{R}$. What is the main advantage of this algorithm?

(*Answer*) We use the Markov chain in Exercise 32 but modify the transition probabilities between $s = X_i$ and $t = X_i \setminus \{v\}$ such that

$$P_{s,t} = \min(1, e^{|t|}/e^{|s|}) = \min(1, 1/e^{\lambda}).$$

Similarly, we modify the transition probability from $s = X_i$ and $t = X_i \cup \{v\}$ such that

$$P_{s,t} = \min(1, e^{|s|}/e^{|t|}) = \min(1, e^{\lambda}).$$

This algorithm does not require computing the term in the denominator

$$\sum_{I' \in \mathcal{I}} e^{\lambda|I'|},$$

which could be very expensive as $\mathcal{I}$ could be of exponential size. The algorithm relies on just computing the ratio of the value for two adjacent terms, so the normalising factor cancels out.

4 Random walks

4.1 Properties

Exercise 35 [Periodicity]

(a) Prove that a simple random walk on a graph is periodic if the graph G is bipartite.

(b) (+) Can you also prove that the random walk is aperiodic if G is not bipartite?

Exercise 36 [Stationary distribution]

(a) Verify that $\pi(u) = \frac{2\deg(u)}{2|E|}$ is a stationary distribution of a simple random walk.

(b) Verify that the lazy random walk has the same stationary distribution.

(c) Does this hold for any graph?

(*Answer*)

(a) We need to verify that the π is a distribution and that it satisfies $\pi P = \pi$. For the first property,

$$\sum_{u \in V} \pi_u = \sum_{u \in V} \frac{\deg(u)}{2|E|} = \frac{2|E|}{2|E|} = 1,$$

using that $\sum_{u \in V} \deg(u) = 2|E|$, given that each edge is counted twice.

For the second property, let $N(u)$ denote the neighbours of u, then

$$\pi_u = \sum_{v \in N(u)} \pi(v) \cdot P_{v,u} = \sum_{v \in N(u)} \frac{\deg(v)}{2|E|} \cdot \frac{1}{\deg(v)} = \sum_{v \in N(u)} \frac{1}{2|E|} = \frac{\deg(u)}{2|E|}.$$

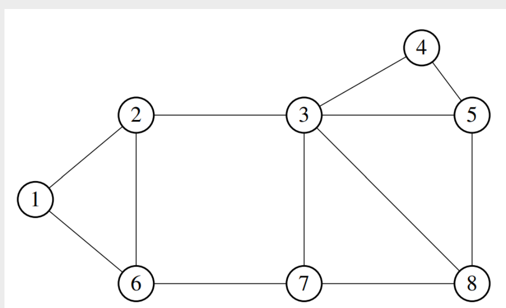
(b) For the lazy random walk, the first property holds for the same reason. For the second property,

$$\pi_u = \frac{1}{2}\pi_u + \frac{1}{2} \sum_{v \in N(u)} \pi(v) \cdot P_{v,u} \Leftrightarrow \pi_u = \sum_{v \in N(u)} \pi(v) \cdot P_{v,u},$$

which we proved in the first part of the question.

(c) π is the stationary distribution for any graph. The connectedness property (which implies that the graph is irreducible) is only needed to argue that the stationary distribution is unique.

Exercise 37 [Concrete Example] Suppose that G has eight vertices and undirected edges as shown in the figure below.



(a) Find the stationary distribution for the Markov Chain on G.

(b) Determine the relative proportions of time spent at each of the eight vertices.

[Source: [2007P4Q5 (c)]]

4.2 Variation distance

Exercise 38 Prove that the total variation distance between two distributions over a finite probability space Ω is a value in $[0, 1]$.

(*Answer*) The lower bound follows trivially from the fact that the total variation distance $\|\cdot\|_{TV}$ is a sum of absolute values, hence non-negative.

For the upper bound, consider two distributions μ and ν. Then, we write

$$\|\mu - \nu\|_{TV} = \frac{1}{2} \sum_{\omega \in \Omega} |\mu(\omega) - \nu(\omega)|$$

$$
\begin{aligned}
&= \frac{1}{2} \sum_{\substack{\omega\in\Omega \\ \mu(\omega)\geq\nu(\omega)}} (\mu(\omega)-\nu(\omega)) + \sum_{\substack{\omega\in\Omega \\ \mu(\omega)<\nu(\omega)}} (\nu(\omega)-\mu(\omega)) \\
&\overset{(a)}{\leq} \frac{1}{2}\cdot\left(\sum_{\omega\in\Omega}\mu(\omega)+\sum_{\omega\in\Omega}\nu(\omega)\right) \\
&\overset{(b)}{=} \frac{1}{2}\cdot(1+1)=1.
\end{aligned}
$$

using in (a) that $\mu(\omega)\geq 0$ and $\nu(\omega)\geq 0$ for all $\omega\in\Omega$ and in (b) that μ and ν are probability distributions.

Exercise 39 Consider a finite Markov chain (μ, P) with state space Ω. Show that

$$
\|P^t_\mu - \pi\|_{TV} \leq \max_{x\in\Omega}\|P^t_x - \pi\|_{TV}
$$

(*Answer*)

$$
\begin{aligned}
\|P^t_\mu - \pi\|_{TV} &= \sum_{\omega\in\Omega}\left|P^t_\mu(\omega)-\pi(\omega)\right| \\
&= \sum_{\omega\in\Omega}\left|(\mu P^t)(\omega)-\pi(\omega)\right| \\
&= \sum_{\omega\in\Omega}\left|\sum_{\omega'\in\Omega}\mu(\omega')P^t_{\omega'\omega}-\pi(\omega)\right| \\
&\overset{(a)}{=} \sum_{\omega\in\Omega}\left|\sum_{\omega'\in\Omega}\left(\mu(\omega')P^t_{\omega'\omega}-\mu(\omega')\pi(\omega)\right)\right| \\
&= \sum_{\omega\in\Omega}\left|\sum_{\omega'\in\Omega}\mu(\omega')\left(P^t_{\omega'\omega}-\pi(\omega)\right)\right| \\
&\overset{(b)}{\leq} \sum_{\omega'\in\Omega}\mu(\omega')\sum_{\omega\in\Omega}\left|P^t_{\omega'\omega}-\pi(\omega)\right| \\
&= \sum_{\omega'\in\Omega}\mu(\omega')\|P^t_{\omega',\omega}-\pi(\omega)\|_{TV} \\
&\leq \sum_{\omega'\in\Omega}\mu(\omega')\max_{x\in\Omega}\|P^t_{x,\omega}-\pi(\omega)\|_{TV} \\
&\leq \max_{x\in\Omega}\|P^t_{x,\omega}-\pi(\omega)\|_{TV},
\end{aligned}
$$

using in (a) that μ is a distribution over Ω and in (b) that $|a+b|\leq|a|+|b|$.

Exercise 40 Let P be a transition matrix of a Markov chain with state space Ω and μ and ν be two probability distributions on Ω. Prove that

$$
\|\mu P - \nu P\|_{TV} \leq \|\mu-\nu\|_{TV}
$$

(*Answer*)

$$
\begin{aligned}
\|\mu P - \nu P\|_{TV} &= \sum_{\omega\in\Omega}|(\mu P)(\omega)-(\nu P)(\omega)| \\
&= \sum_{\omega\in\Omega}\left|\sum_{\omega'\in\Omega}\mu(\omega')P_{\omega',\omega}-\sum_{\omega'\in\Omega}\nu(\omega')P_{\omega',\omega}\right| \\
&= \sum_{\omega\in\Omega}\left|\sum_{\omega'\in\Omega}(\mu(\omega')-\nu(\omega'))\cdot P_{\omega',\omega}\right| \\
&\overset{(a)}{\leq} \sum_{\omega'\in\Omega}\sum_{\omega\in\Omega}|\mu(\omega')-\nu(\omega')|\cdot P_{\omega',\omega}
\end{aligned}
$$

$$
\begin{aligned}
&= \sum_{\omega' \in \Omega} |\mu(\omega') - \nu(\omega')| \cdot \sum_{\omega \in \Omega} P_{\omega',\omega} \\
&= \|\mu - \nu\|_{TV},
\end{aligned}
$$

using in (a) that $|a+b| \leq |a| + |b|$ and in (b) that $\sum_{\omega \in \Omega} P_{\omega',\omega} = 1$ for any $\omega' \in \Omega$.

Exercise 41 Consider a state space Ω and for any probability p distribution over Ω, define for any set $A \subseteq \Omega$, $p(A) = \sum_{\omega \in A} p(\omega)$. Then, for probability distributions μ and ν

$$\|\mu - \nu\|_{TV} = \max_{A \subseteq \Omega} |\mu(A) - \nu(A)|.$$

(*Answer*) Let $A^+ := \{\omega \in \Omega : \mu(\omega) \geq \nu(\omega)\}$ and $A^- := \{\omega \in \Omega : \mu(\omega) < \nu(\omega)\}$. Then, since μ and ν are distributions

$$\mu(A) = \mu(A^+) + \mu(A^-) = \nu(A) = \nu(A^+) + \nu(A^-) = 1.$$

Therefore,

$$
\begin{aligned}
\|\mu - \nu\|_{TV} &= \frac{1}{2} \sum_{\omega \in A^+} (\mu(\omega) - \nu(\omega)) + \frac{1}{2} \sum_{\omega \in A^-} (\nu(\omega) - \mu(\omega)) \\
&= \frac{1}{2} \cdot \big(\mu(A^+) - \nu(A^+) + \nu(A^-) - \mu(A^-)\big) \\
&= \frac{1}{2} \cdot 2 \cdot (\mu(A^+) - \nu(A^+)) \\
&= \mu(A^+) - \nu(A^+).
\end{aligned}
$$

TODO: Argue that A^+ has the maximum by a greedy argument. Perhaps draw a figure here.

Exercise 42 Consider the Ehrenfest Markov Chain P with state space $\Omega = \{0, 1, \ldots, d\}$, and assume that the chain starts from state 0.

(a) Define a Markov chain Q with state space $\Omega' = \{0,1\}^d$, where $x \in \Omega'$ gives the colours of the particles.

(b) Relate the stationary distribution π_P to π_Q.

(c) In the Markov chain of (a), show that starting from state 0, then for any $x_1, x_2 \in \Omega'$ with the same number of 1s, satisfy

$$Q^t_{0,x_1} = Q^t_{0,x_2}.$$

(d) Can you relate the variation distance $\|P^t_0 - \pi_P\|$ in the Ehrenfest chain to the variation distance $\|Q^t_{0^d} - \pi_Q\|$ in the hypercube? *Hint: Use some symmetry argument.*

4.3 Mixing times

Exercise 43 Attempt **[2022P8Q12 (b)(ii)(iii)]**.

Exercise 44 This question asks you to prove lower bounds on the mixing time of some lazy random walks on graphs.

(a) Let $G = (V_1 \cup V_2, E)$ be a graph made of two disjoint complete graphs of n vertices, supported respectively on V_1 and V_2, connected by a single edge. This is called the *Barbell graph*. Consider a lazy random walk on G. Prove that $t_{mix}(1/4) = \Omega(n^2)$.

(b) Suppose now we add $s < n$ edges to the Barbell graph, where each edge has one endpoint in V_1 and the other endpoint in V_2. What happens to $t_{mix}(G)$?

(c) Consider now a version of the Barbell graph where $|V_1| = n$, $|V_2| = \lfloor \log n \rceil$ and there exists only an edge between V_1 and V_2. What is the mixing time of this graph?

(*Answer*) See solution notes here.

4.4 Cover times

Extended Note 2 [Cover times] The *cover time* of a random walk is defined as the expected time for a random walk to visit all vertices of a graph, starting from the worst possible vertex.

Exercise 45 [Cover time of clique] Analyse the cover time of a simple random walk on the complete graph (clique), i.e., the graph where each pair of vertices is connected by an undirected edge. *Hint: Use coupon collector.*

Exercise 46 [Cover time of path] Consider a path P_n with vertex set $\{0, 1, \ldots, n\}$ for even n. Can you determine the cover time? *Hint: use Exercise 24.*

Exercise 47 [Cover time of cycle] What is the cover time of a cycle C_n? *Hint: use Exercise 24.*

Exercise 48 [Cover time using mixing time] Consider any regular graph $G = (V, E)$. In this exercise, we will derive an upper bound on the cover time based on the mixing time $t := t_{mix}(1/(2n))$ which is $\mathcal{O}(t \cdot n \log n)$.

(a) Let $T_u := \min\{t \geq 1 : X_t = u\}$. Show that starting from any vertex $X_0 = u$, we have that

$$\mathbf{Pr}\left[\, T_v \leq t \mid X_0 = u \,\right] \geq \frac{1}{2n}.$$

(b) Using probability amplification and the union bound, show that

$$\mathbf{Pr}\left[\max_{u \in V} T_u \leq 4nt \log n\right] \geq 1 - n^{-2}.$$

(c) Bound the expectation of the cover time.