Chapter 7

Finite Markov Chains

Markov chains that we will be studying in this chapter is a stochastic process, which we have yet to define:

Definition 7.1. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability triple. A* $\mathbb{R}$*-valued* **stochastic process** *is a parametrized set of RVs. That is, we denote the collection*

$$(X_\alpha)_{\alpha \in \mathbb{A}}$$

for a parameter set $\mathbb{A}$*, a* $\mathbb{R}$*-valued stochastic process.*

If the index set $\mathbb{A} = \mathbb{N}$ *we call it a* **discrete (or discrete-time) stochatic process**.

That is, previously we have used the concept of an i.i.d. sequence of random variables, that is, $X_1, X_2, \ldots \overset{\text{IID}}{\sim} F$. This sequence is a simple case of a discrete stochastic process. To study this sequence as a process is quite uninteresting as all X_i are independent, we will in this chapter introduce some dependency an analyze the resulting structure. These discrete stochastic processes are termed finite Markov chains. We will cover their properties and simulation methods.

7.1 Introduction

A finite Markov chain is a stochastic process that moves among elements in a finite set $\mathbb{X}$ as follows: when at $x \in \mathbb{X}$ the next position is chosen at random according to a fixed probability distribution $P(\cdot|x)$. We define such a process more formally below.

7.1.1 Advanced intro*

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability triple and let $\mathcal{G}$ be a sigma algebra on Ω. Define for a $\mathbb{R}$ valued R.V. X the conditional expectation

$$\mathbb{E}[X \mid \mathcal{G}]$$

which is any $\mathcal{G}$ measurable function (is it unique?) $\Omega \to \mathbb{R}$ which satisfies for any $G \in \mathcal{G}$

$$\int_G \mathbb{E}[X \mid \mathcal{G}] dP = \int_G X dP$$

This can be thought of the best possible guess of X given the knowledge contained in $\mathcal{G}$. The conditional probability can be defined as

$$\mathbb{P}(X \in A \mid \mathcal{G}) := \mathbb{E}[\mathbb{1}_A(X) \mid \mathcal{G}]$$

which constitutes the single best guess for the if the event $X \in A$ happened given the information contained in $\mathcal{G}$.

Remark 7.2. *If we think of a random variable $X \in L^2(\mathbb{P})$, then for any σ-algebra $\mathcal{G}$ we see that*

$$\mathbb{E}[(X - \mathbb{E}[X \mid \mathcal{G}])\mathbb{1}_G] = 0$$

for all $G \in \mathcal{G}$, which says that the $L^2(\mathbb{P})$ random variable $X - \mathbb{E}[X \mid \mathcal{G}]$ is orthogonal to all indicators $\mathbb{1}_G$, $G \in \mathcal{G}$. In this case the conditional expecation is unique and can be thought of as a projection of X onto $\mathcal{G}$.

Remark 7.3. *Properties:*

- $\mathbb{E}[\mathbb{E}[X \mid \mathcal{G}]] = \mathbb{E}[X]$*, so the tower property still holds.*
- *If X is $\mathcal{G}$ measurable (i.e. we know X), then $\mathbb{E}[X \mid \mathcal{G}] = X$, i.e. we are allowed to guess X itself since we know it. There is no better guess.*
- *If X is independent of $\mathcal{G}$, i.e. the information in $\mathcal{G}$ is irrelevant for X, then $\mathbb{E}[X \mid \mathcal{G}] = \mathbb{E}[X]$, i.e. we gained nothing.*

Consider now a stochastic process $X_1, \ldots, X_n$ with index set $\mathbb{N}$. Define $\mathcal{F}_n$ as the smallest σ-algebra on Ω such $(X_1, \ldots, X_n)$ is an $\mathbb{R}^n$ valued RV.

We can evaluate this on a particular realization of $(X_1, \ldots, X_n)$ as follows

$$\mathbb{P}(X_t \in A \mid \mathcal{F}_n) = \mathbb{P}(X_t \in A \mid x_1, \ldots, x_n)((X_1, \ldots, X_n)).$$

From the above it is clear that $\mathbb{P}(X_t \in A \mid \mathcal{F}_n)$ is a random variable that depends on $(X_1, \ldots, X_n)$. Note that $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdot$ is an increasing family of σ-algebras, all of which are subsets of $\mathcal{F}$, such a sequence is denoted a filtration.

Remark 7.4. *This can be seen as over-complicating it. Why do we even need this level of formalism? Since the statespace $\mathbb{X}$ is finite we could equally well say*

$$\mathbb{P}(A \mid X_1 = x, X_2 = x_2, \ldots, X_n = x_n)$$

which we already know what it means. The reason is

1. *It simplifies notation.*
2. *If you take a course in continuous stochastic processes then you would have to use this notation.*
3. *If the statespace is not enumerable, then you would also have to use this formalism.*
4. *The object $\mathcal{F}_n$ has a natural interpretation as the "history", or specifically when mapped by $(X_1, \ldots, X_n)$ as the set of trajectories of the stochastic process up to and including time n.*

Definition 7.5 (Finite Markov Chain)**.** *A stochastic process,*

$$\{X_n : n \in \mathbb{N}\}$$

is a **Markov chain** *with* **state space** $\mathbb{X}$*, if for any $t \in \mathbb{N}$ the following holds*

$$\mathbb{P}(X_{t+1} = x|\mathcal{F}_t) = \mathbb{P}(X_{t+1} = x|X_t).$$

We say that the Markov chain is **homogeneous** *if*

$$\mathbb{P}(X_{t+1} = x|X_t) = \mathbb{P}(X_{s+1} = x|X_s)$$

for all $t, s \in \mathbb{N}$.

From the above and with the intuition that $\mathbb{P}(X_{t+1} = x \mid \mathcal{F}_t)$ constitutes our best guess for the event $X_{t+1} = x$ being true given the history of the process up to time t, we can interpret the Markov chain condition as saying that the only information from the history that we need is the value at time t.

7.1.2 Non advanced introduction

Definition 7.6 (Finite Markov Chain)**.** *A stochastic process,*

$$\{X_n : n \in \mathbb{N}\}$$

is a **Markov chain** *with* **state space** $\mathbb{X}$*, if for any* $t \in \mathbb{N}$ *the following holds*

$$\mathbb{P}(X_{t+1} = x | X_0, X_1, \ldots, X_t) = \mathbb{P}(X_{t+1} = x | X_t).$$

We say that the Markov chain is **homogeneous** *if*

$$\mathbb{P}(X_{t+1} = x | X_t) = \mathbb{P}(X_{s+1} = x | X_s)$$

for all $t, s \in \mathbb{N}$.

Note that if the Markov chain is homogeneous then it is enough to know

$$\mathbb{P}(X_1 = x_1 | X_0 = x_0) : \mathbb{X} \times \mathbb{X} \to [0, 1]$$

i.e. we can identify it with a $|\mathbb{X}| \times |\mathbb{X}|$ matrix

$$P_{xy} = \mathbb{P}(X_1 = y | X_0 = x).$$

Usually we identify $x \in \mathbb{X}$ with the enumeration of elements in $\mathbb{X}$ and thus we can write for $N = |\mathbb{X}|$ an $N \times N$ matrix P_{ij}. This matrix is denoted the **transition matrix**.

Lemma 7.7. *Let* $\{X_n, n \in \mathbb{N}\}$ *be a homogeneous Markov chain. Let the statespace* $\mathbb{X} = \{s_1, \ldots, s_N\}$ *be enumerated and let* μ_0 *be the PMF of* X_0*. Then the PMF* μ_n *for* X_n *is*

$$\mu_n = \mu_0 P^n$$

Proof. Let us start with applying the law of total probability

$$\begin{aligned}\mathbb{P}(X_n = x_n) &= \sum_{x_{n-1}} \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}) \, \mathbb{P}(X_{n-1} = x_{n-1}) \\ &= \sum_{x_{n-1}} P_{x_{n-1} x_n} \, \mathbb{P}(X_{n-1} = x_{n-1})\end{aligned}$$

Since n was arbitrary we can apply it again until we reach X_0, namely

$$\mathbb{P}(X_n = x_n) = \sum_{x_0, \ldots, x_{n-1}} P_{x_{n-1} x_n} \ldots P_{x_0 x_1} \, \mathbb{P}(X_0 = x_0)$$

The above is just a sequence of matrix multiplications, we can write

$$\mu_n = \mu_0 P^n.$$

□

Since we will be interested in Markov chains on $\mathbb{X} = \{s_1, s_2, \ldots, s_k\}$ with the same transition matrix P but different initial distributions, we introduce $\mathbb{P}_\mu$ and $\mathbb{E}_\mu$ for probabilities and expectations given that the initial distribution is μ, respectively. When the initial distribution is concentrated at a single initial state x given by:

$$\mathbb{1}_{\{x\}}(y) := \begin{cases} 1 & \text{if } y = x \\ 0 & \text{if } y \neq x \end{cases}$$

we represent it by e_x, the $1 \times k$ ortho-normal basis row vector with a 1 in the x-th entry and a 0 elsewhere. We simply write $\mathbb{P}_x$ for $\mathbb{P}_{\mathbb{1}_{\{x\}}}$ or $\mathbb{P}_{e_x}$ and $\mathbb{E}_x$ for $\mathbb{E}_{\mathbb{1}_{\{x\}}}$ or $\mathbb{E}_{e_x}$. Thus, Lemma 7.7 along with our new notations means that:

$$\mathbb{P}_x(X_t = y) = (e_x P^t)(y) = P^t(x, y) \ .$$

In words, the probability of going to y from x in t steps is given by the (x, y)-th entry of P^t, the **t-step transition matrix**. We refer to the x-th row and the x-th column of P by $P(x, \cdot)$ and $P(\cdot, x)$, respectively.

Let a function $f(x) : \mathbb{X} \to \mathbb{R}$ be given, then we can define

$$(P^t f)(x) := \sum_y P^t(x, y) f(y) = \sum_y f(y)\, \mathbb{P}_x(X_t = y) = \mathbb{E}_x(f(X_t)) \ . \quad (7.1)$$

This is the expected value of f under the distribution of states in t steps given that we start at state x.

Identified in the above way, we see that $P^t : (\mathbb{X} \to \mathbb{R}) \to (\mathbb{X} \to \mathbb{R})$ i.e. it maps a $\mathbb{R}$ valued function to an $\mathbb{R}$ valued function. Let us look at some properties of P^t

Lemma 7.8. *Let $f : \mathbb{X} \to \mathbb{R}$, then the mapping P^t defined in (7.1) satisfies:*

- *Let $g : \mathbb{X} \to \mathbb{R}$, then $P^t(f + g) = P^t f + P^t g$, i.e. it is a linear functional.*
- *Let $t > s > 0$ be positive integers, then $P^t f = P^{t-s}(P^s f)$.*

Proof. Let us denote X_t, Y_t two homogeneous and independent Markov processes with the same transition Matrix. First, define $g(x) = (P^s f)(x) =$

$\mathbb{E}_x[f(X_s)]$, then consider

$$\begin{aligned}(P^{t-s}(P^s f))(x) &= \mathbb{E}_x[g(Y_{t-s})] = \sum_y g(y)\,\mathbb{P}(Y_{t-s} = y \mid Y_0 = x) \\ &= \sum_y \sum_z f(z)\,\mathbb{P}(X_s = z \mid X_0 = y)\,\mathbb{P}(Y_{t-s} = y \mid Y_0 = x) \\ &= \sum_y \sum_z f(z)\,\mathbb{P}(X_t = z \mid X_{t-s} = y)\,\mathbb{P}(X_{t-s} = y \mid X_0 = x) \\ &= \sum_z f(z)\,\mathbb{P}(X_t = z \mid X_0 = x) \\ &= (P^t f)(x)\end{aligned}$$

□

This is the prime example of a so called **Semigroup**,

Definition 7.9. *A semigroup is a set S together with a binary operator $\odot$, i.e. a function $\odot : S \times S \to S$ that satisfies the associative property*

$$(a \odot b) \odot c = a \odot (b \odot c)$$

for all $a, b, c \in S$.

Let $S = \{P^t, t > 0\}$ and define the operator $\odot$ as

$$P^t \odot P^s = P^{t+s}$$

then we see from the above lemma Lemma 7.8 that $\{P_t, t > 0\}$ forms a semigroup, specifically a one parameter semigroup.

Remark 7.10. *The semigroup property is retained when moving over to continuous time Markov processes. If the semigroup is what is called strongly continuous there is also a time dependent (parabolic) partial differential equation which $P^t f$ solves.*

Until now our Markov chains have been **homogeneous** in time according to Definition 7.5, i.e., the transition matrix P does not change with time. We define inhomogeneous Markov chains as Markov chains that are not homogeneous. Such Markov chains are more realistic as models in some situations and more flexible as algorithms in the sequel.

Lemma 7.11. *For a finite inhomogeneous Markov chain $(X_t)_{t\in\mathbb{Z}_+}$ with state space $\mathbb{X} = \{s_1, s_2, \ldots, s_k\}$, initial distribution*

$$\mu_0 := (\mu_0(s_1), \mu_0(s_2), \ldots, \mu_0(s_k)),$$

where $\mu_0(s_i) = \mathbb{P}(X_0 = s_i)$*, and transition matrices*

$$(P_1, P_2, \ldots), \quad P_t := (P_t(s_i, s_j))_{(s_i,s_j)\in\mathbb{X}\times\mathbb{X}}, \ t \in \{1, 2, \ldots\}$$

we have for any $t \in \mathbb{Z}_+$ *that the distribution at time* t *given by:*

$$\mu_t := (\mu_t(s_1), \mu_t(s_2), \ldots, \mu_t(s_k)),$$

where $\mu_t(s_i) = \mathbb{P}(X_t = s_i)$*, satisfies:*

$$\mu_t = \mu_0 P_1 P_2 \cdots P_t \ . \tag{7.2}$$

Proof. Left as Exercise 7.12. □

Exercise 7.12. *Prove Lemma 7.11 in a similar way to Lemma 7.7.*

7.2 Random Mapping Representation and Simulation

In order to simulate $(x_0, x_1, \ldots, x_n)$, a sequential realisation or sequence of states visited by a Markov chain, we need a random mapping representation of a Markov chain.

Definition 7.13 (Random mapping representation (RMR))**.** *A* **random mapping representation** *(RMR) of a transition matrix* $P := (P(x, y))_{(x,y)\in\mathbb{X}^2}$ *is a function*

$$\rho(x, w) : \mathbb{X} \times \mathbb{W} \to \mathbb{X} \ , \tag{7.3}$$

along with a $\mathbb{W}$*-valued random variable* W*, satisfying*

$$\mathbb{P}(\{\rho(x, W) = y\}) = P(x, y), \quad \text{for each } (x, y) \in \mathbb{X}^2 \ . \tag{7.4}$$

Theorem 7.14. *Every Markov chain on* $\mathbb{X}$ *has a random mapping representation.*

Proof. Let X_t be a Markov chain with transition matrix P_t at t, for simplicity assume that $\mathbb{X} \subset \mathbb{N}$. Let t be an arbitrary time. Let $Z_t \sim \text{Uniform}([0, 1])$. For any $i, j \in \mathbb{X}$, set

$$F_{i,j} = \sum_{m=1}^{j} P_t(i, m).$$

Define

$$f_t(i, z) := j \quad \text{when} \quad F_{i,j-1} < z \leq F_{i,j}.$$

We have

$$\begin{aligned}\mathbb{P}(f_t(i,Z)=j) &= \mathbb{P}(F_{i,j-1} < Z \leq F_{i,j}) = F_{i,j} - F_{i,j-1} \\ &= \sum_{m=1}^{j} P_t(i,m) - \sum_{m=1}^{j-1} P_t(i,m) = P_t(i,j).\end{aligned}$$

We see that (f_t, Z_t) is a RMR for X_t at t. □

Exercise 7.15. *Think about what $F_{i,j}$ actually is and what is $f_t(i,z)$? Hint: Recall the concept of inversion sampling Theorem 5.38 and relate it to that.*

Theorem 7.16. *Let $W_1, \ldots, \overset{\text{IID}}{\sim} F$ such that (ρ_t, W_t) is a RMR for a transition matrix P_t, for all $t \in \mathbb{N}$. Then if $X_0 \sim \mu_0$,*

$$X_t := \rho_t(X_{t-1}, W_t), t \in \mathbb{N},$$

is a Markov chain with initial distribution μ_0 and transition matrix P_t at time t.

Exercise 7.17. *Do the proof of Theorem 7.16 by using the necessary Definitions.*

Exercise 7.18. *Show that the RMR for a Markov chain is not necessarily unique.*

7.3 Irreducibility and Aperiodicity

The utility of our mathematical constructions with Markov chains depends on a delicate balance between generality and specificity. We introduce two specific conditions called irreducibility and aperiodicity that make Markov chains more useful to model real-word phenomena.

Definition 7.19. *Let X_t be a homogeneous Markov chain on state space $\mathbb{X} = \{s_1, \ldots, s_N\}$. We say that $s_i \to s_j$ (* **communicates***) if there exists a $t \in \mathbb{N}$ such that*

$$\mathbb{P}(X_t = s_j \mid X_0 = s_i) > 0.$$

We say that s_i, s_j **intercommunicates** *if $s_i \to s_j$ and $s_j \to s_i$, we write this as $s_i \leftrightarrow s_j$.*

Definition 7.20 (Irreducible)**.** *A homogeneous Markov chain $(X_t)_{t\in\mathbb{Z}_+}$ on state space $\mathbb{X} := \{s_1, s_2, \ldots, s_k\}$ is said to be* **irreducible** *if $s_i \leftrightarrow s_j$ for each $(s_i, s_j) \in \mathbb{X}^2$. Otherwise the chain is said to be* **reducible***.*

Definition 7.21 (Return times and period)**.** *Let* $\mathbb{T}(x) := \{t \in \mathbb{N} : P^t(x,x) > 0\}$ *be the set of* **possible return times** *to the starting state* x*. The* **period** *of state* x *is defined to be* $\gcd(\mathbb{T}(x))$*, the greatest common divisor of* $\mathbb{T}(x)$*. When the period of a state* x *is* 1*, i.e.,* $\gcd(\mathbb{T}(x)) = 1$*, then* x *is said to be an* **aperiodic state***.*

Proposition 7.22. *If the Markov chain* $(X_t)_{t\in\mathbb{Z}_+}$ *with transition matrix* P *on state space* $\mathbb{X}$ *is irreducible then* $\gcd(\mathbb{T}(x)) = \gcd(\mathbb{T}(y))$ *for any* $(x,y) \in \mathbb{X}^2$*.*

Proof. Fix any pair of states $(x,y) \in \mathbb{X}^2$. Since, P is irreducible, $x \leftrightarrow y$ and therefore there exists natural numbers $\eta(x,y)$ and $\eta(y,x)$ such that $P^{\eta(x,y)}(x,y) > 0$ and $P^{\eta(y,x)}(y,x) > 0$. Let $\eta' = \eta(x,y) + \eta(y,x)$ and observe that $\eta' \in \mathbb{T}(x) \cap \mathbb{T}(y)$, $\mathbb{T}(x) \subset \mathbb{T}(y) - \eta' := \{t - \eta' : t \in \mathbb{T}(y)\}$ and $\gcd(\mathbb{T}(y))$ divides all elements in $\mathbb{T}(x)$. Thus, $\gcd(\mathbb{T}(y)) \leq \gcd(\mathbb{T}(x))$. By a similar argument we can also conclude that $\gcd(\mathbb{T}(x)) \leq \gcd(\mathbb{T}(y))$. Therefore $\gcd(\mathbb{T}(x)) = \gcd(\mathbb{T}(y))$. □

Definition 7.23 (Aperiodic)**.** *A Markov chain* $(X_t)_{t\in\mathbb{Z}_+}$ *with transition matrix* P *on state space* $\mathbb{X}$ *is said to be aperiodic if all of its states are aperiodic, i.e.,* $\gcd(\mathbb{T}(x)) = 1$ *for every* $x \in \mathbb{X}$*. If a chain is not aperiodic, we call it* **periodic***.*

7.4 Stationarity

We are interested in statements about a Markov chain that has been running for a long time. For any nontrivial Markov chain $(X_0, X_1, \ldots)$ the value of X_t will keep fluctuating in the state space $\mathbb{X}$ as $t \to \infty$ and we cannot hope for convergence to a fixed point state $x^* \in \mathbb{X}$ or to a k-cycle of states $\{x_1, x_2, \ldots, x_k\} \subset \mathbb{X}$. However, we can look one level up into the space of probability distributions over $\mathbb{X}$ that give the probability of the Markov chain visiting each state $x \in \mathbb{X}$ at time t, and hope that the distribution of X_t over $\mathbb{X}$ settles down as $t \to \infty$. The Markov chain convergence theorem indeed sattes that the distribution of X_t over $\mathbb{X}$ settles down as $t \to \infty$, provided the Markov chain is irreducible and aperiodic.

Definition 7.24 (Stationary distribution)**.** *Let* $(X_t)_{t\in\mathbb{Z}_+}$ *be a Markov chain with state space* $\mathbb{X} = \{s_1, s_2, \ldots, s_k\}$ *and transition matrix* $P = (P(x,y))_{(x,y)\in\mathbb{X}^2}$*. A row vector*

$$\pi = (\pi(s_1), \pi(s_2), \ldots, \pi(s_k)) \in \mathbb{R}^{1\times k}$$

is said to be a **stationary distribution** *for the Markov chain, if it satisfies the conditions of being:*

1. a probability distribution*:* $\pi(x) \geq 0$ *for each* $x \in \mathbb{X}$ *and* $\sum_{x\in\mathbb{X}} \pi(x) = 1$*, and*

2. a fixed point*:* $\pi P = \pi$*, i.e.,* $\sum_{x\in\mathbb{X}} \pi(x)P(x,y) = \pi(y)$ *for each* $y \in \mathbb{X}$*.*

Proposition 7.25 (Existence of Stationary distribution)**.** *For any irreducible and aperiodic Markov chain there exists at least one stationary distribution.*

Proof. See the Perron Frobenius theorem, Wikipedia. □

7.5 Reversibility

We introduce another specific property called reversibility. This property will assist in conjuring Markov chains with a desired stationary distibution.

Definition 7.26 (Reversible)**.** *A probability distribution* π *on* $\mathbb{X} = \{s_1, s_2, \ldots, s_k\}$ *is said to be a* **reversible distribution** *for a Markov chain* $(X_t)_{t\in\mathbb{Z}}$ *on* $\mathbb{X}$ *with transition matrix* P *if for every pair of states* $(x, y) \in \mathbb{X}^2$*:*

$$\pi(x)P(x,y) = \pi(y)P(y,x) \ . \tag{7.5}$$

A Markov chain that has a reversible distribution is said to be a reversible Markov chain.

In words, $\pi(x)P(x,y) = \pi(y)P(y,x)$ says that if you start the chain at the reversible distribution π, i.e., $\mu_0 = \pi$, then the probability of going from x to y is the same as that of going from y to x.

Proposition 7.27 (A reversible π is a stationary π)**.** *Let* $(X_t)_{t\in\mathbb{Z}_+}$ *be a Markov chain on* $\mathbb{X} = \{s_1, s_2, \ldots, s_k\}$ *with transition matrix* P*. If* π *is a reversible distribution for* $(X_t)_{t\in\mathbb{Z}_+}$ *then* π *is a stationary distribution for* $(X_t)_{t\in\mathbb{Z}_+}$*.*

Exercise 7.28. *Prove Proposition 7.27.*

7.5.1 Random Walks on Graphs

Random walks on graphs is one of the most useful applications of Markov chains. In this section, we will see some basic definitions from graph theory and define simple random walks on graphs as finite Markov chains to shed light on the random surfer model of Google.

Definition 7.29 (Definitions in Graph Theory)**.** *Here we take a brief tour of the most basic defintions in graph theory. A* **Graph** $\mathbb{G} := (\mathbb{V}, \mathbb{E})$ *consists of a* **vertex set** $\mathbb{V} := \{v_1, v_2, \ldots, v_k\}$ *together with an* **edge set** $\mathbb{E} := \{e_1, e_2, \ldots, e_l\}$*. Each* **edge** *in* $\mathbb{E}$ *connects two of the* **vertices** *in* $\mathbb{V}$*. A* **directed edge** e_h *connecting vertex* v_i *to* v_j *is denoted by the ordered pair* (v_i, v_j)*. An* **undirected edge** *simply connects two vertices without regard to order and is denoted by* $\{v_i, v_j\}$ *to represent both of the directed edges* (v_i, v_j) *and* (v_j, v_i)*. Thus,* $\mathbb{E} \subset \mathbb{V}^2$ *and a graph* $\mathbb{G}$ *with directed eges in* $\mathbb{E}$ *is said to be an* **directed graph** *and a graph* $\mathbb{G}$ *with undirected edges is said to be an* **undirected graph***. Two vertices are* **neighbours** *if they share an edge, i.e., if they are connected by an edge. The* **neighbourhood** *of a vertex* v_i *denoted by* $\mathrm{nbhd}(v_i) := \{v_j : (v_i, v_j) \in \mathbb{E}\}$ *is the set of neighbouring vertices of* v_i*. The number of neighbours of a vertex* v_i *in an undirected graph is called its* **degree** *and is denoted by* $\deg(v_i)$*. Note that* $\deg(v_i) = \#\,\mathrm{nbhd}(v_i)$*. If there is a sequences of edges or a path from every vertex to every other vertex then the undirected graph is said to be connected. In a graph we only allow one edge per pair of vertices but in a* **multigraph** *we allow more than one edge per pair of vertices. An edge can be* **weighted** *by being associated with a real number called its weight. More generally, vertices and edges can be augmented with various properties, including addresses, names, etc., and weights, relation types, etc. Graphs whose vertices and edges are further augmented by various properties are called* **property graphs***, an extremely useful and versatile representation of data from different domains. We can represent a directed graph by its* **adjacency matrix** *given by:*

$$A := (A(v_i, v_j))_{(v_i, v_j) \in \mathbb{V} \times \mathbb{V}}, \quad A(v_i, v_j) = \begin{cases} 1 & \text{if } (v_i, v_j) \in \mathbb{E} \\ 0 & \text{otherwise .} \end{cases}$$

Thus the adjacency matrix of an undirected graph is symmetric. In a directed graph, each vertex v_i *has* **in-edges** *that come into it and* **out-edges** *that go out of it. The number of in-edges and out-edges of* v_i *is denoted by* $\mathrm{ideg}(v_i)$ *and* $\mathrm{odeg}(v_i)$ *respectively. Note that a transition diagram of a Markov chain is a weighted directed graph and is represented by the transition probability matrix.*

Model 7.30 (Random Walk on a Connected Undirected Graph)**.** *A random walk on a connected undirected graph* $\mathbb{G} = (\mathbb{V}, \mathbb{E})$ *is a Markov chain with state space* $\mathbb{V} := \{v_1, v_2, \ldots, v_k\}$ *and the following transition rules: if the chain is at vertex* v_i *at time* t *then it moves uniformly at random to one of the neighbours of* v_i *at time* $t + 1$*. If* $\deg(v_i)$ *is the degree of* v_i *then the*

transition probabilities of this Markov chain is

$$P(v_i, v_j) = \begin{cases} \frac{1}{\deg(v_i)} & \text{if } (v_i, v_j) \in \mathbb{E} \\ 0 & \text{otherwise,} \end{cases}$$

Proposition 7.31. *The random walk on a connected undirected graph* $\mathbb{G} = (\mathbb{V}, \mathbb{E})$*, with vertex set* $\mathbb{V} := \{v_1, v_2, \ldots, v_k\}$ *and degree sum* $d = \sum_{i=1}^{k} \deg(v_i)$ *is a reversible Markov chain with the reversible distribution* π *given by:*

$$\pi = \left(\frac{\deg(v_1)}{d}, \frac{\deg(v_2)}{d}, \ldots, \frac{\deg(v_k)}{d} \right) .$$

Exercise 7.32. *Prove Proposition 7.31 by directly showing that* π *is reversible.*

Example 7.33 (Google's random surfer on the word wide web)**.** *Consider the huge graph with vertices as webpages and hyper-links as undirected edges. Then Model 7.30 gives a random walk on this graph. However if a page has no links to other pages, it becomes a sink and therefore terminates the random walk. Let us modify this random walk into a* **random surf** *to avoid getting stuck. If the random surfer arrives at a sink page, she picks another page at random and continues surfing at random again. Google's PageRank formula uses a random surfer model who gets bored after several clicks and switches to a random page. The PageRank value of a page reflects the chance that the random surfer will land on that page by clicking on a link. The stationary distribution of the random surfer on the world wide web is a very successful model for ranking pages and has grown into a Trillion dollar company.*

7.6 Computer exercises

Exercise 7.34. *Take the text file* `pride_and_prejudice.txt` *and define a Markov chain where each state is a word. You could either model everything as an observation of a single Markov chain, i.e. the entire book is one long chain. Otherwise you could consider each sentence to be an observation, but now you have different starting points. You need to assume homogeneity to estimate the transition matrix.*

- *Estimate the transition matrix.*
- *Calculate the probability of going from the word 'the' to the word 'her'.*
- *Use Theorem 7.14 or Exercise 7.15 to generate a sentence using the estimated transition matrix starting from the word 'Lady'.*

Exercise 7.35. *Consider a set of bacteria, each bacteria either splits into* 2*, stays the same or dies. I.e. each bacteria can create either* $\{0, 1, 2\}$ *successors. The branching process. Let* Z_n *denote the number of bacteria at time* n*, let* $X_{n,i}$ *be a random variable denoting the number of direct successors for member* i *(can only be* $\{0, 1, 2\}$ *in period* n*, where* $X_{n,i}$ *are i.i.d., here* $n \in \{0, 1, \ldots\}$ *and* $i \in \{1, \ldots, Z_n\}$*. The recurrence relation is*

$$Z_{n+1} = \sum_{i=1}^{Z_n} X_{n,i}$$

with $Z_0 = 1$*.*

- *Is this a Markov chain?*
- *Let* $p = (1/3, 1/3, 1/3)$ *be the probabilities of* $\{0, 1, 2\}$ *offspring and simulate the chain.*
- *We say that the population dies out if* $Z_n = 0$ *for some* n *and* n *denotes the life-time of the population, by simulation, calculate the expected life-time of the population.*
- *If you want to read more, check out the Galton-Watson process.*