

# STAT211: Random Processes

Non-examinable Supplement for Researchers

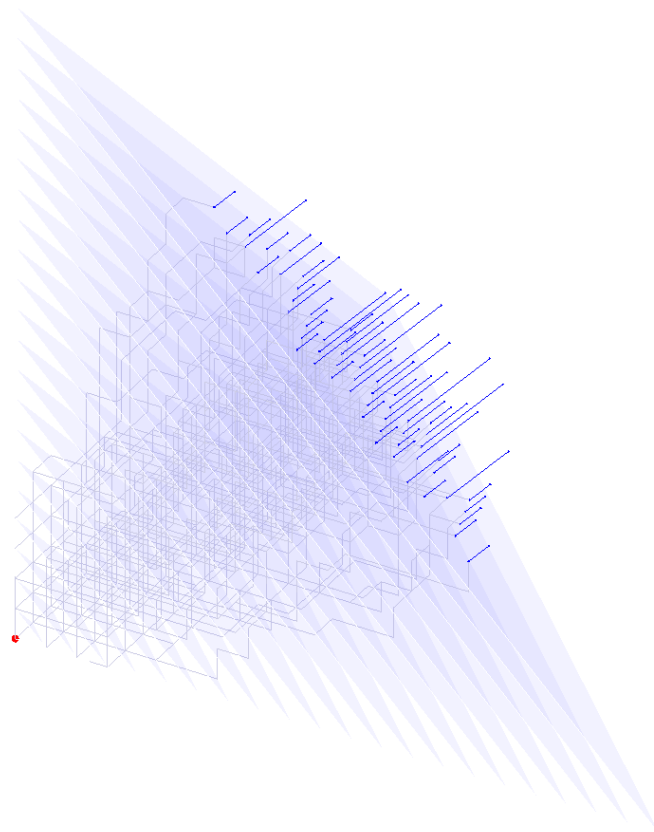
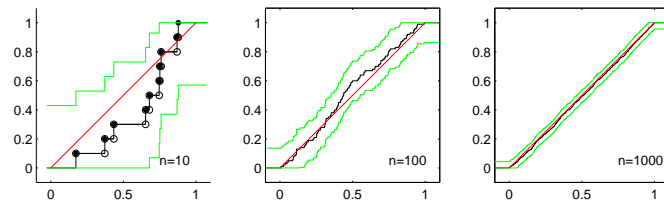
Raazesh Sainudiin and Dominic Lee  
School of Mathematics and Statistics, University of Canterbury,  
Christchurch, New Zealand

©2007–2016 Raazesh Sainudiin. ©2008–2013 Dominic Lee.

This work is licensed under the Creative Commons Attribution-Noncommercial-Share Alike 3.0 New Zealand License. To view a copy of this license, visit

<http://creativecommons.org/licenses/by-nc-sa/3.0/nz/>.

This work was partially supported by NSF grant DMS-03-06497 and NSF/NIGMS grant DMS-02-01037.



# Contents

<b>1</b>	<b>Finite Markov Chains</b>	<b>3</b>
1.1	Introduction . . . . .	3
1.2	Irreducibility and Aperiodicity . . . . .	11
1.3	Stationarity . . . . .	13
1.4	Reversibility . . . . .	15

# Chapter 1

## Finite Markov Chains

When a stochastic process  $(X_\alpha)_{\alpha \in \mathbb{A}}$  is not independent it is said to be dependent. So far we have mostly concerned ourselves with independent processes. In this chapter we introduce finite Markov chains and their simulation methods. Finite Markov chains are among the simplest stochastic processes with a ‘first-order’ dependence called Markov dependence.

### 1.1 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.

**Definition 1 (Finite Markov Chain)** A stochastic sequence,

$$(X_n)_{n \in \mathbb{Z}_+} := (X_0, X_1, \dots),$$

is a homogeneous **Markov chain** with **state space**  $\mathbb{X}$  and **transition matrix**  $P := (P(x, y))_{(x, y) \in \mathbb{X}^2}$  if for all pair of **states**  $(x, y) \in \mathbb{X}^2 := \mathbb{X} \times \mathbb{X}$ , all integers  $t \geq 1$ , and all probable historical events  $H_{t-1} := \bigcap_{n=0}^{t-1} \{X_n = x_n\}$  with  $\mathbf{P}(H_{t-1} \cap \{X_t = x\}) > 0$ , the following **Markov property** is satisfied:

$$\mathbf{P}(X_{t+1} = y | H_{t-1} \cap \{X_t = x\}) = \mathbf{P}(X_{t+1} = y | X_t = x) =: P(x, y) . \quad (1.1)$$

The Markov property means that the conditional probability of going to state  $y$  at time  $t+1$  from state  $x$  at current time  $t$  is always given by the  $(x, y)$ -th entry  $P(x, y)$  of the transition matrix  $P$ , no matter what sequence of states  $(x_0, x_1, \dots, x_{t-1})$  preceded the current state  $x$ . Thus, the  $|\mathbb{X}| \times |\mathbb{X}|$  matrix  $P$  is enough to obtain the state transitions since the  $x$ -th row of  $P$  is the probability distribution  $P(x, \cdot) := (P(x, y))_{y \in \mathbb{X}}$ . For this reason  $P$  is called a **stochastic matrix**, i.e.,

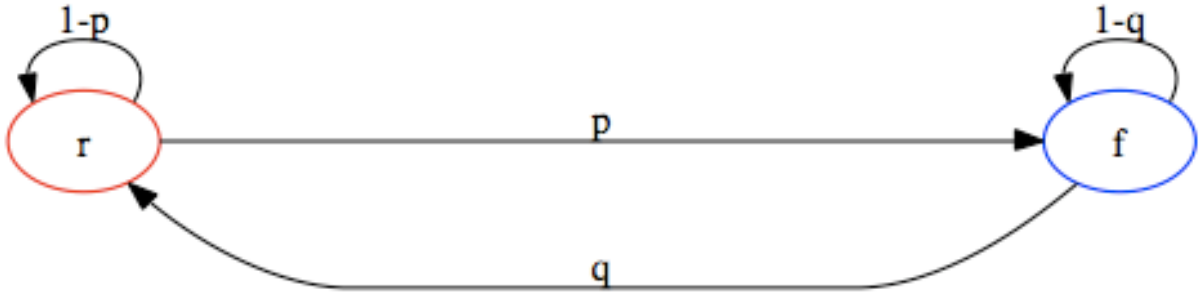
$$P(x, y) \geq 0 \quad \text{for all } (x, y) \in \mathbb{X}^2 \quad \text{and} \quad \sum_{y \in \mathbb{X}} P(x, y) = 1 \quad \text{for all } x \in \mathbb{X} . \quad (1.2)$$

Thus, for a Markov chain  $(X_n)_{n \in \mathbb{Z}_+}$ , the distribution of  $X_{t+1}$  given  $X_0, \dots, X_t$  depends on  $X_t$  alone. Because of this dependence on the previous state, the stochastic sequence,  $(X_0, X_1, \dots)$ , are *not* independent. We introduce the most important concepts using a simple example.

**Example 1 (Flippant Freddy)** Freddy the flippant frog lives in an enchanted pond with only two lily pads, *rollopia* and *flipopia*. A wizard gave a die and a silver coin to help flippant Freddy decide where to jump next. Freddy left the die on rollopia and the coin on flipopia. When Freddy got restless in rollopia he would roll the die and if the die landed odd he would leave the die behind and jump to flipopia, otherwise he would stay put. When Freddy got restless in flipopia he would flip the coin and if it landed Heads he would leave the coin behind and jump to rollopia, otherwise he would stay put.

Let the state space  $\mathbb{X} = \{r, f\}$ , and let  $(X_0, X_1, \dots)$  be the sequence of lily pads occupied by Freddy after his restless moments. Say the die on rollopia  $r$  has probability  $p$  of turning up odd and the coin on flipopia  $f$  has probability  $q$  of turning up heads. We can visualise the rules of Freddy's jumps by the following **transition diagram**:

Figure 1.1: Transition Diagram of Flippant Freddy's Jumps.



Then Freddy's sequence of jumps  $(X_0, X_1, \dots)$  is a Markov chain on  $\mathbb{X}$  with transition matrix:

$$P = \begin{matrix} & \begin{matrix} r & f \end{matrix} \\ \begin{matrix} r \\ f \end{matrix} & \begin{pmatrix} P(r, r) & P(r, f) \\ P(f, r) & P(f, f) \end{pmatrix} \end{matrix} = \begin{matrix} & \begin{matrix} r & f \end{matrix} \\ \begin{matrix} r \\ f \end{matrix} & \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix} \end{matrix}. \quad (1.3)$$

Suppose we first see Freddy in rollopia, i.e.,  $X_0 = r$ . When he gets restless for the first time we know from the first row of  $P$  that he will leave to flipopia with probability  $p$  and stay with probability  $1 - p$ , i.e.,

$$\mathbf{P}(X_1 = f | X_0 = r) = p, \quad \mathbf{P}(X_1 = r | X_0 = r) = 1 - p. \quad (1.4)$$

What happens when he is restless for the second time? By considering the two possibilities for  $X_1$ ,

Definition of conditional probability and the Markov property, we see that,

$$\begin{aligned}
\mathbf{P}(X_2 = f|X_0 = r) &= \mathbf{P}(X_2 = f, X_1 = f|X_0 = r) + \mathbf{P}(X_2 = f, X_1 = r|X_0 = r) \\
&= \frac{\mathbf{P}(X_2 = f, X_1 = f, X_0 = r)}{\mathbf{P}(X_0 = r)} + \frac{\mathbf{P}(X_2 = f, X_1 = r, X_0 = r)}{\mathbf{P}(X_0 = r)} \\
&= \mathbf{P}(X_2 = f|X_1 = f, X_0 = r) \frac{\mathbf{P}(X_1 = f, X_0 = r)}{\mathbf{P}(X_0 = r)} \\
&\quad + \mathbf{P}(X_2 = f|X_1 = r, X_0 = r) \frac{\mathbf{P}(X_1 = r, X_0 = r)}{\mathbf{P}(X_0 = r)} \\
&= \mathbf{P}(X_2 = f|X_1 = f, X_0 = r) \mathbf{P}(X_1 = f|X_0 = r) \\
&\quad + \mathbf{P}(X_2 = f|X_1 = r, X_0 = r) \mathbf{P}(X_1 = r|X_0 = r) \\
&= \mathbf{P}(X_2 = f|X_1 = f) \mathbf{P}(X_1 = f|X_0 = r) \\
&\quad + \mathbf{P}(X_2 = f|X_1 = r) \mathbf{P}(X_1 = r|X_0 = r) \\
&= P(f, f)P(r, f) + P(r, f)P(r, r) \\
&= (1 - q)p + p(1 - p)
\end{aligned} \tag{1.5}$$

Similarly,

$$\mathbf{P}(X_2 = r|X_0 = r) = P(f, r)P(r, f) + P(r, r)P(r, r) = qp + (1 - p)(1 - p) \tag{1.6}$$

Instead of elaborate computations of the probabilities of being in a given state after Freddy's  $t$ -th restless moment, we can store the state probabilities at time  $t$  in a row vector:

$$\mu_t := (\mathbf{P}(X_t = r|X_0 = r), \mathbf{P}(X_t = f|X_0 = r)) \quad ,$$

Now, we can conveniently represent Freddy starting in rollopiia by the **initial distribution**  $\mu_0 = (1, 0)$  and obtain the 1-step **state probability vector** in (1.4) from  $\mu_1 = \mu_0 P$  and the 2-step state probabilities in (1.5) and (1.6) by  $\mu_2 = \mu_1 P = \mu_0 P P = \mu_0 P^2$ . In general, multiplying  $\mu_t$ , the state probability vector at time  $t$ , by the transition matrix  $P$  on the right updates the state probabilities by another step:

$$\mu_t = \mu_{t-1} P \quad \text{for all } t \geq 1 \quad .$$

And for any initial distribution  $\mu_0$ ,

$$\mu_t = \mu_0 P^t \quad \text{for all } t \geq 0 \quad .$$

This can be easily implemented in MATLAB as follows:

```

>> p=0.85; q=0.35; P = [1-p p; q 1-q] % assume an unfair coin and an unfair die
P =
    0.1500    0.8500
    0.3500    0.6500
>> mu0 = [1, 0] % initial state vector since Freddy started in rollopiia
mu0 =
     1     0
>> mu0*P^0 % initial state distribution at t=0 is just mu0
ans =
     1     0
>> mu0*P^1 % state distribution at t=1
ans =
    0.1500    0.8500
>> mu0*P^2 % state distribution at t=2
ans =
    0.3200    0.6800
>> mu0*P^3 % state distribution at t=3
ans =
    0.2860    0.7140

```

Now, let us compute and look at the probability of being in rollopia after having started there for three values of  $p$  and  $q$  according to the following script:

---

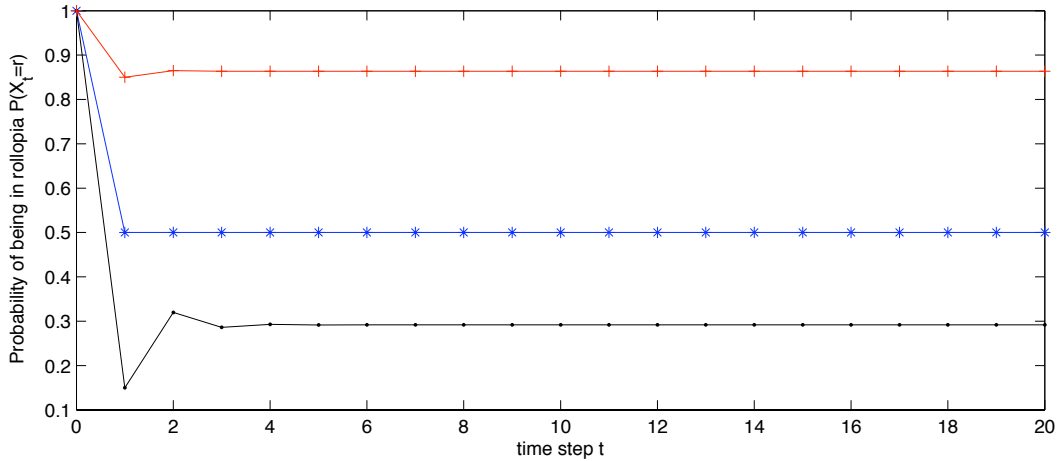
```

p=0.5; q=0.5; P = [1-p p; q 1-q]; % assume a fair coin and a fair die
mu0 = [1, 0]; % initial state vector since Freddy started in rollopia
for t = 1: 1: 21, mut(t,:) = mu0*P^(t-1); end
t=0:1:20; % vector of time steps t
plot(t,mut(:,1)','','b*--')
hold on;
p=0.85; q=0.35; P = [1-p p; q 1-q]; % assume an unfair coin and an unfair die
for t = 1: 1: 21, mut(t,:) = mu0*P^(t-1); end
t=0:1:20; % vector of time steps t
plot(t,mut(:,1)','','k.-')
p=0.15; q=0.95; P = [1-p p; q 1-q]; % assume another unfair coin and another unfair die
for t = 1: 1: 21, mut(t,:) = mu0*P^(t-1); end
t=0:1:20; % vector of time steps t
plot(t,mut(:,1)','','r+-')
xlabel('time step t'); ylabel('Probability of being in rollopia $P(X_t=r)$')
xlabel('time step t'); ylabel('Probability of being in rollopia $P(X_t=r)$')

```

---

Figure 1.2: The probability of being back in rollopia in  $t$  time steps after having started there under transition matrix  $P$  with (i)  $p = q = 0.5$  (blue line with asterisks), (ii)  $p = 0.85$ ,  $q = 0.35$  (black line with dots) and (iii)  $p = 0.15$ ,  $q = 0.95$  (red line with pluses).



It is evident from Figure 1.2 that as  $t \rightarrow \infty$ ,  $\mu_t$  approaches a distribution, say  $\pi$ , that depends on  $p$  and  $q$  in  $P$ . Such a limit distribution is called the **stationary distribution** and must satisfy the fixed point condition:

$$\pi P = \pi ,$$

that gives the solution:

$$\pi(r) = \frac{q}{p+q}, \quad \pi(f) = \frac{p}{p+q} .$$

In Figure 1.2 we see that  $\mathbf{P}(X_t = r)$  approaches  $\pi(r) = \frac{q}{p+q}$  for the three cases of  $p$  and  $q$ :

$$\begin{aligned}
\text{(i) } p = 0.50, q = 0.50, \quad & \mathbf{P}(X_t = r) \rightarrow \pi(r) = \frac{q}{p+q} = \frac{0.50}{0.50 + 0.50} = 0.5000, \\
\text{(ii) } p = 0.85, q = 0.35, \quad & \mathbf{P}(X_t = r) \rightarrow \pi(r) = \frac{q}{p+q} = \frac{0.35}{0.85 + 0.35} = 0.2917, \\
\text{(iii) } p = 0.15, q = 0.95, \quad & \mathbf{P}(X_t = r) \rightarrow \pi(r) = \frac{q}{p+q} = \frac{0.95}{0.15 + 0.95} = 0.8636.
\end{aligned}$$

Now let us generalise the lessons learned from Example 1.

**Proposition 2** For a finite Markov chain  $(X_t)_{t \in \mathbb{Z}_+}$  with state space  $\mathbb{X} = \{s_1, s_2, \dots, s_k\}$ , initial distribution

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

where  $\mu_0(s_i) = \mathbf{P}(X_0 = s_i)$ , and transition matrix

$$P := (P(s_i, s_j))_{(s_i, s_j) \in \mathbb{X}^2},$$

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

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

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

$$\mu_t = \mu_0 P^t. \quad (1.7)$$

**Proof:** We will prove this by induction on  $\mathbb{Z}_+ := \{0, 1, 2, \dots\}$ . First consider the case when  $t = 0$ . Since  $P^0$  is the identity matrix  $I$ , we get the desired equality:

$$\mu_0 P^0 = \mu_0 I = \mu_0.$$

Next consider the case when  $t = 1$ . We get for each  $j \in \{1, 2, \dots, k\}$ , that

$$\begin{aligned} \mu_1(s_j) &= \mathbf{P}(X_1 = s_j) = \sum_{i=1}^k \mathbf{P}(X_1 = s_j, X_0 = s_i) \\ &= \sum_{i=1}^k \mathbf{P}(X_1 = s_j | X_0 = s_i) \mathbf{P}(X_0 = s_i) \\ &= \sum_{i=1}^k P(s_i, s_j) \mu_0(s_i) \\ &= (\mu_0 P)(s_j), \quad \text{the } j\text{-th entry of the row vector } (\mu_0 P). \end{aligned}$$

Hence,  $\mu_1 = \mu_0 P$ . Now, we will fix  $m$  and suppose that (1.7) holds for  $t = m$  and prove that (1.7) also holds for  $t = m + 1$ . For each  $j \in \{1, 2, \dots, k\}$ , we get

$$\begin{aligned} \mu_{m+1}(s_j) &= \mathbf{P}(X_{m+1} = s_j) = \sum_{i=1}^k \mathbf{P}(X_{m+1} = s_j, X_m = s_i) \\ &= \sum_{i=1}^k \mathbf{P}(X_{m+1} = s_j | X_m = s_i) \mathbf{P}(X_m = s_i) \\ &= \sum_{i=1}^k P(s_i, s_j) \mu_m(s_i) \\ &= (\mu_m P)(s_j), \quad \text{the } j\text{-th entry of the row vector } (\mu_m P). \end{aligned}$$

Hence,  $\mu_{m+1} = \mu_m P$ . But  $\mu_m = \mu_0 P^m$  by the induction hypothesis, and therefore:

$$\mu_{m+1} = \mu_m P = \mu_0 P^m P = \mu_0 P^{m+1}.$$

Thus by the principle of mathematical induction we have proved the proposition.

Thus, multiplying a row vector  $\mu_0$  by  $P^t$  on the right takes you from current distribution over the state space to the distribution in  $t$  steps of the chain.

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

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

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  $\mathbf{P}_x$  for  $\mathbf{P}_{\mathbf{1}_{\{x\}}}$  or  $\mathbf{P}_{e_x}$  and  $\mathbf{E}_x$  for  $\mathbf{E}_{\mathbf{1}_{\{x\}}}$  or  $\mathbf{E}_{e_x}$ . Thus, Proposition 2 along with our new notations means that:

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

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.

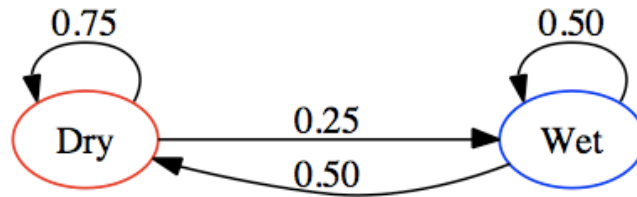
Let the function  $f(x) : \mathbb{X} \rightarrow \mathbb{R}$  be represented by the column vector  $f := (f(s_1), f(s_2), \dots, f(s_k)) \in \mathbb{R}^{k \times 1}$ . Then the  $x$ -th entry of  $P^t f$  is:

$$P^t f(x) = \sum_y P^t(x, y) f(y) = \sum_y f(y) \mathbf{P}_x(X_t = y) = \mathbf{E}_x(f(X_t)) .$$

This is the expected value of  $f$  under the distribution of states in  $t$  steps given that we start at state  $x$ . Thus multiplying a column vector  $f$  by  $P^t$  from the left takes you from a function on the state space to its expected value in  $t$  steps of the chain.

**Example 2 (Dry-Wet Christchurch Weather)** Consider a toy weather model for dry or wet days in Christchurch using a Markov chain with state space  $\{d, w\}$ . Let the transition diagram in Figure 1.3 give the transition matrix  $P$  for our dry-wet Markov chain. Using (1.7) we can find that

Figure 1.3: Transition Diagram of Dry and Wet Days in Christchurch.



the probability of being dry on the day after tomorrow is 0.625 given that it is wet today as follows:

```

>> P=[0.75 0.25; 0.5 0.5] % Transition Probability Matrix
P =
    0.7500    0.2500
    0.5000    0.5000
>> mu0=[0 1] % it is wet today gives the initial distribution
mu0 =
     0     1
>> mu0 * P^2 % the distribution in 2 days from today
ans =
    0.6250    0.3750
  
```



Suppose you sell \$100 of lemonade at a road-side stand on a hot day but only \$50 on a cold day. Then we can compute your expected sales tomorrow if today is dry as follows:

```
>> P=[0.75 0.25; 0.5 0.5] % Transition Probability Matrix
P =
    0.7500    0.2500
    0.5000    0.5000
>> f = [100; 50] % sales of lemonade in dollars on a dry and wet day
f =
    100
     50
>> P*f % expected sales tomorrow
ans =
    87.5000
    75.0000
>> mu0 = [1 0] % today is dry
mu0 =
     1     0
>> mu0*P*f % expected sales tomorrow if today is dry
ans =
    87.5000
```

**Exercise 3 (Freddy discovers a gold coin)** Flippant Freddy of Example 1 found a gold coin at the bottom of the pond. Since this discovery he jumps around differently in the enchanted pond. He can be found now in one of three states: flipopia, rolloppia and hydroppia (when he dives into the pond). His state space is  $\mathbb{X} = \{r, f, h\}$  now and his transition mechanism is as follows: If he rolls an odd number with his fair die in rolloppia he will jump to flipopia but if he rolls an even number then he will stay in rolloppia only if the outcome is 2 otherwise he will dive into hydroppia. If the fair gold coin toss at the bottom of hydroppia is Heads then Freddy will swim to flipopia otherwise he will remain in hydroppia. Finally, if he is in flipopia he will remain there if the silver coin lands Heads otherwise he will jump to rolloppia.

Make a Markov chain model of the new jumping mechanism adopted by Freddy. Draw the transition diagram, produce the transition matrix  $P$  and compute using MATLAB the probability that Freddy will be in hydroppia after one, two, three, four and five jumps given that he starts in hydroppia.

**Exercise 4** Let  $(X_t)_{t \in \mathbb{Z}_+}$  be a Markov chain with state space  $\{a, b, c\}$ , initial distribution  $\mu_0 = (1/3, 1/3, 1/3)$  and transition matrix

$$P = \begin{matrix} & \begin{matrix} a & b & c \end{matrix} \\ \begin{matrix} a \\ b \\ c \end{matrix} & \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \end{matrix}.$$

For each  $t$ , define  $Y_t = \mathbf{1}_{\{b,c\}}(X_t)$ . Show that  $(Y_t)_{t \in \mathbb{Z}_+}$  is not a Markov chain.

**Exercise 5** Let  $(X_t)_{t \in \mathbb{Z}_+}$  be a (homogeneous) Markov chain on  $\mathbb{X} = \{s_1, s_2, \dots, s_k\}$  with transition matrix  $P$  and initial distribution  $\mu_0$ . For a given  $m \in \mathbb{N}$ , let  $(Y_t)_{t \in \mathbb{Z}_+}$  be a stochastic sequence with  $Y_t = X_{mt}$ . Show that  $(Y_t)_{t \in \mathbb{Z}_+}$  is a Markov chain with transition matrix  $P^m$ . This establishes that Markov chains that are sampled at regular time steps are also Markov chains.

Until now our Markov chains have been **homogeneous** in time according to Definition 1, i.e., the transition matrix  $P$  does not change with time. We define inhomogeneous Markov chains that allow their transition matrices to possibly change with time. Such Markov chains are more realistic as models in some situations and more flexible as algorithms in the sequel.

**Definition 3 (Inhomogeneous finite Markov chain)** Let  $P_1, P_2, \dots$  be a sequence of  $k \times k$  stochastic matrices satisfying the conditions in Equation 1.2. Then, the stochastic sequence  $(X_t)_{t \in \mathbb{Z}_+} := (X_0, X_1, \dots)$  with finite state space  $\mathbb{X} := \{s_1, s_2, \dots, s_k\}$  is called an inhomogeneous Markov chain with transition matrices  $P_1, P_2, \dots$ , if for all pairs of states  $(x, y) \in \mathbb{X} \times \mathbb{X}$ , all integers  $t \geq 1$ , and all probable historical events  $H_{t-1} := \bigcap_{n=0}^{t-1} \{X_n = x_n\}$  with  $\mathbf{P}(H_{t-1} \cap \{X_t = x\}) > 0$ , the following **Markov property** is satisfied:

$$\mathbf{P}(X_{t+1} = y | H_{t-1} \cap \{X_t = x\}) = \mathbf{P}(X_{t+1} = y | X_t = x) =: P_{t+1}(x, y) . \quad (1.8)$$

**Proposition 4** For a finite inhomogeneous Markov chain  $(X_t)_{t \in \mathbb{Z}_+}$  with state space  $\mathbb{X} = \{s_1, s_2, \dots, s_k\}$ , initial distribution

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

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

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

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

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

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

$$\mu_t = \mu_0 P_1 P_2 \cdots P_t . \quad (1.9)$$

**Proof:** Left as Exercise 6.

**Exercise 6** Prove Proposition 4 using induction as done for Proposition 2.

**Example 7 (a more sophisticated dry-wet chain)** Let us make a more sophisticated version of the dry-wet chain of Example 2 with state space  $\{d, w\}$ . In order to take some seasonality into account in our weather model for dry and wet days in Christchurch, let us have two transition matrices for hot and cold days:

$$P_{\text{hot}} = \begin{matrix} & \begin{matrix} d & w \end{matrix} \\ \begin{matrix} d \\ w \end{matrix} & \begin{pmatrix} 0.95 & 0.05 \\ 0.75 & 0.25 \end{pmatrix} \end{matrix}, \quad P_{\text{cold}} = \begin{matrix} & \begin{matrix} d & w \end{matrix} \\ \begin{matrix} d \\ w \end{matrix} & \begin{pmatrix} 0.65 & 0.35 \\ 0.45 & 0.55 \end{pmatrix} \end{matrix} .$$

We say that a day is hot if its maximum temperature is more than 20° Celsius, otherwise it is cold. We use the transition matrix for today to obtain the state probabilities for tomorrow. If today is dry and hot and tomorrow is supposed to be cold then what is the probability that the day after tomorrow will be wet? We can use (1.9) to obtain the answer as 0.36:

```
>> Phot = [0.95 0.05; 0.75 0.25] % Transition Probability Matrix for hot day
Phot =
    0.9500    0.0500
    0.7500    0.2500
>> Pcold = [0.65 0.35; 0.45 0.55] % Transition Probability Matrix for cold day
Pcold =
    0.6500    0.3500
    0.4500    0.5500
>> mu0 = [1 0] % today is dry
mu0 =
    1    0
```

```

>> mu1 = mu0 * Phot % distribution for tomorrow since today is hot
mu1 =      0.9500      0.0500
>> mu2 = mu1 * Pcold % distribution for day after tomorrow since tomorrow is supposed to be cold
mu2 =      0.6400      0.3600
>> mu2 = mu0 * Phot * Pcold % we can also get the distribution for day after tomorrow directly
mu2 =      0.6400      0.3600

```

**Exercise 8** For the Markov chain in Example 7 compute the probability that the day after tomorrow is wet if today is dry and hot but tomorrow is supposed to be cold.

## 1.2 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-world phenomena.

**Definition 5 (Communication between states)** Let  $(X_t)_{t \in \mathbb{Z}_+}$  be a homogeneous Markov chain with transition matrix  $P$  on state space  $\mathbb{X} := \{s_1, s_2, \dots, s_k\}$ . We say that a state  $s_i$  **communicates** with a state  $s_j$  and write  $s_i \rightarrow s_j$  or  $s_j \leftarrow s_i$  if there exists an  $\eta(s_i, s_j) \in \mathbb{N}$  such that:

$$\mathbf{P} \left( X_{t+\eta(s_i, s_j)} = s_j | X_t = s_i \right) = P^{\eta(s_i, s_j)}(s_i, s_j) > 0 .$$

In words,  $s_i$  communicates with  $s_j$  if you can eventually reach  $s_j$  from  $s_i$ . If  $P^\eta(s_i, s_j) = 0$  for every  $\eta \in \mathbb{N}$  then we say that  $s_i$  **does not communicate** with  $s_j$  and write  $s_i \nrightarrow s_j$  or  $s_j \nleftarrow s_i$ .

We say that two states  $s_i$  and  $s_j$  **intercommunicate** and write  $s_i \leftrightarrow s_j$  if  $s_i \rightarrow s_j$  and  $s_j \rightarrow s_i$ . In words, two states intercommunicate if you can eventually reach one from another and vice versa. When  $s_i$  and  $s_j$  do not intercommunicate we write  $s_i \nleftrightarrow s_j$ .

**Definition 6 (Irreducible)** A homogeneous Markov chain  $(X_t)_{t \in \mathbb{Z}_+}$  with transition matrix  $P$  on state space  $\mathbb{X} := \{s_1, s_2, \dots, s_k\}$  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**.

We have already seen examples of reducible and irreducible Markov chains. For example, Flippant Freddy's family of Markov chains with the  $(p, q)$ -parametric family of transition matrices,  $\{P_{(p, q)} : (p, q) \in [0, 1]^2\}$ , where each  $P_{(p, q)}$  is given by Equation 1.3. If  $(p, q) \in (0, 1)^2$ , then the corresponding Markov chain is irreducible because we can go from rolloppia to flippopia or vice versa in just one step with a positive probability. Thus, the Markov chains with transition matrices in  $\{P_{(p, q)} : (p, q) \in (0, 1)^2\}$  are irreducible. But if  $p$  or  $q$  take probability values at the boundary of  $[0, 1]$ , i.e.,  $p \in \{0, 1\}$  or  $q \in \{0, 1\}$  then we have to be more careful because we may never get from at least one state to the other and the corresponding Markov chains may be reducible. For instance, if  $p = 0$  or  $q = 0$  then we will be stuck in either rolloppia or flippopia, respectively. However, if  $p = 1$  and  $q \neq 0$  or  $q = 1$  and  $p \neq 0$  then we can get from each state to the other. Therefore, only the transition matrices in  $\{P_{(p, q)} : p \in \{0\} \text{ or } q \in \{0\}\}$  are reducible.

The simplest way to verify whether a Markov chain is irreducible is by looking at its transition diagram (without the positive edge probabilities) and checking that from each state there is a sequence of arrows leading to any other state.

**Exercise 9** Revisit all the Markov chains we have considered up to now and determine whether they are reducible or irreducible by checking that from each state there is a sequence of arrows leading to any other state in their transition graphs.

**Definition 7 (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 8** 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 9 (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**.

We have already seen example of irreducible Markov chains that were either periodic or aperiodic. For instance, Freddy's Markov chain with  $(p, q) \in (0, 1)^2$  is aperiodic since the period of either of its two states is given by  $\gcd(\{1, 2, 3, \dots\}) = 1$ . However, the Markov chain model for a drunkard's walk around a block over the state space  $\{0, 1, 2, 3\}$  is periodic because you can only return to the starting state in an even number of time steps and

$$\gcd(\mathbb{T}(0)) = \gcd(\mathbb{T}(1)) = \gcd(\mathbb{T}(2)) = \gcd(\mathbb{T}(3)) = \gcd(\{2, 4, 6, \dots\}) = 2 \neq 1.$$

**Exercise 10** Show that the Markov chain corresponding to a drunkard's walk around a polygonal block with  $k$  corners is irreducible for any integer  $k > 1$ . Show that it is aperiodic only when  $k$  is odd and has period 2 when  $k$  is even.

**Proposition 10** Let  $A = \{a_1, a_2, \dots\} \subset \mathbb{N}$  that satisfies the following two conditions:

1.  $A$  is a **nonlattice**, meaning that  $\gcd(A) = 1$  and
2.  $A$  is closed under addition, meaning that if  $(a, a') \in A^2$  then  $a + a' \in A$ .

Then there exists a positive integer  $\eta < \infty$  such that  $n \in A$  for all  $n \geq \eta$ .

**Proof:** See Proofs of Lemma 1.1, Lemma 1.2 and Theorem 1.1 in Appendix of *Pierre Brémaud, Markov Chains, Gibbs Fields, Monte Carlo Simulation, and Queues, Springer, 1999*.

**Proposition 11** If the Markov chain  $(X_t)_{t \in \mathbb{Z}_+}$  with transition matrix  $P$  on state space  $\mathbb{X}$  is irreducible and aperiodic then there is an integer  $\tau$  such that  $P^t(x, x) > 0$  for all  $t \geq \tau$  and all  $x \in \mathbb{X}$ .

**Proof:** TBD

**Proposition 12** If the Markov chain  $(X_t)_{t \in \mathbb{Z}_+}$  with transition matrix  $P$  on state space  $\mathbb{X}$  is irreducible and aperiodic then there is an integer  $\tau$  such that  $P^t(x, y) > 0$  for all  $t \geq \tau$  and all  $(x, y) \in \mathbb{X}^2$ .

**Proof:** TBD

**Exercise 11 (King's random walk on a chessboard)** Consider the squares in the chessboard as the state space  $\mathbb{X} = \{0, 1, 2, \dots, 7\}^2$  with a randomly walking black king, i.e., for each move from current state  $(u, v) \in \mathbb{X}$  the king chooses one of his  $k(u, v)$  possible moves uniformly at random. Is the Markov chain corresponding to the randomly walking black king on the chessboard irreducible and/or aperiodic?

**Exercise 12 (King's random walk on a chesstorus)** We can obtain a chesstorus from a pliable chessboard by identifying the eastern edge with the western edge (roll the chessboard into a cylinder) and then identifying the northern edge with the southern edge (gluing the top and bottom end of the cylinder together by turning into a doughnut or torus). Consider the squares in the chesstorus as the state space  $\mathbb{X} = \{0, 1, 2, \dots, 7\}^2$  with a randomly walking black king, i.e., for each move from current state  $(x, y) \in \mathbb{X}$  the king chooses one of his 8 possible moves uniformly at random according to the scheme:  $X_t \leftarrow X_{t-1} + W_t$ , where  $W_t$  is independent and identically distributed as follows:

$$\mathbf{P}(W_t = w) = \begin{cases} \frac{1}{8} & \text{if } w \in \{(1, 1), (1, 0), (1, -1), (0, -1), (-1, -1), (-1, 0), (-1, 1), (0, 1)\}, \\ 0 & \text{otherwise.} \end{cases}$$

Is the Markov chain corresponding to the randomly walking black king on the chesstorus irreducible and/or aperiodic? Write a MATLAB script to simulate a sequence of  $n$  states visited by the king if he started from  $(0, 0)$  on the chesstorus.

### 1.3 Stationarity

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

**Definition 13 (Stationary distribution)** Let  $(X_t)_{t \in \mathbb{Z}_+}$  be a Markov chain with state space  $\mathbb{X} = \{s_1, s_2, \dots, s_k\}$  and transition matrix  $P = (P(x, y))_{(x, y) \in \mathbb{X}^2}$ . A row vector

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

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  $y \in \mathbb{X}$ .

**Definition 14 (Hitting times)** If a Markov chain  $(X_t)_{t \in \mathbb{Z}_+}$  with state space  $\mathbb{X} = \{s_1, s_2, \dots, s_k\}$  and transition matrix  $P = (P(x, y))_{(x, y) \in \mathbb{X}^2}$  starts at state  $x$ , then we can define the **hitting time**

$$T(x, y) = \min\{t \geq 1 : X_t = y\} .$$

and let  $T(x, y) = \min\{\} = \infty$  if the Markov chain never visits  $y$  after having started from  $x$ . Let the **mean hitting time**

$$\tau(x, y) := \mathbf{E}(T(x, y)),$$

be the expected time taken to reach  $y$  after having started at  $x$ . Note that  $\tau(x, x)$  is the **mean return time** to state  $x$ .

**Proposition 15 (Hitting times of irreducible aperiodic Markov chains)** If  $(X_t)_{t \in \mathbb{Z}_+}$  is an irreducible aperiodic Markov chain with state space  $\mathbb{X} = \{s_1, s_2, \dots, s_k\}$ , transition matrix  $P = (P(x, y))_{(x, y) \in \mathbb{X}^2}$  then for any pair of states  $(x, y) \in \mathbb{X}^2$ ,

$$\mathbf{P}(T(x, y) < \infty) = 1 ,$$

and the mean hitting time is finite, i.e.,

$$\tau(x, y) < \infty .$$

**Proposition 16 (Existence of Stationary distribution)** For any irreducible and aperiodic Markov chain there exists at least one stationary distribution.

**Proof:** TBD

**Definition 17 (Total variation distance)** If  $\nu_1 := (\nu_1(x))_{x \in \mathbb{X}}$  and  $\nu_2 := (\nu_2(x))_{x \in \mathbb{X}}$  are elements of  $\mathcal{P}(\mathbb{X})$ , the set of all probability distributions on  $\mathbb{X} := \{s_1, s_2, \dots, s_k\}$ , then we define the **total variation distance** between  $\nu_1$  and  $\nu_2$  as

$$d_{\text{TV}}(\nu_1, \nu_2) := \frac{1}{2} \sum_{x \in \mathbb{X}} \text{abs}(\nu_1(x) - \nu_2(x)), \quad d_{\text{TV}} : \mathcal{P}(\mathbb{X}) \times \mathcal{P}(\mathbb{X}) \rightarrow [0, 1] . \quad (1.10)$$

If  $\nu_1, \nu_2, \dots$  and  $\nu$  are probability distributions on  $\mathbb{X}$ , then we say that  $\nu_t$  **converges in total variation** to  $\nu$  as  $n \rightarrow \infty$  and write  $\nu_t \xrightarrow{\text{TV}} \nu$ , if

$$\lim_{t \rightarrow \infty} d_{\text{TV}}(\nu_t, \nu) = 0 .$$

Observe that if  $d_{\text{TV}}(\nu_1, \nu_2) = 0$  then  $\nu_1 = \nu_2$ . The constant 1/2 in Equation 1.10 ensures that the range of  $d_{\text{TV}}$  is in  $[0, 1]$ . If  $d_{\text{TV}}(\nu_1, \nu_2) = 1$  then  $\nu_1$  and  $\nu_2$  have disjoint supports, i.e., we can partition  $\mathbb{X}$  into  $\mathbb{X}_1$  and  $\mathbb{X}_2$ , i.e.,  $\mathbb{X} = \mathbb{X}_1 \cup \mathbb{X}_2$  and  $\mathbb{X}_1 \cap \mathbb{X}_2 = \emptyset$ , such that  $\sum_{x \in \mathbb{X}_1} \nu_1(x) = 1$  and  $\sum_{x \in \mathbb{X}_2} \nu_2(x) = 1$ . The total variation distance gets its name from the following natural interpretation:

$$d_{\text{TV}}(\nu_1, \nu_2) = \max_{A \subset \mathbb{X}} \text{abs}(\nu_1(A) - \nu_2(A)) .$$

This interpretation means that the total variation distance between  $\nu_1$  and  $\nu_2$  is the maximal difference in probabilities that the two distributions assign to any one event  $A \in \sigma(\mathbb{X}) = 2^{\mathbb{X}}$ .

In words, Proposition 18 says that if you run the chain for a sufficiently long enough time  $t$ , then, regardless of the initial distribution  $\mu_0$ , the distribution at time  $t$  will be close to the stationary distribution  $\pi$ . This is referred to as the Markov chain **approaching equilibrium** or **stationarity** as  $t \rightarrow \infty$ .

**Proposition 18 (Markov chain convergence theorem)** Let  $(X_t)_{t \in \mathbb{Z}_+}$  be an irreducible aperiodic Markov chain with state space  $\mathbb{X} = \{s_1, s_2, \dots, s_k\}$ , transition matrix  $P = (P(x, y))_{(x, y) \in \mathbb{X}^2}$  and initial distribution  $\mu_0$ . Then for any distribution  $\pi$  which is stationary for the transition matrix  $P$ , we have

$$\mu_t \xrightarrow{\text{TV}} \pi . \quad (1.11)$$

**Proof:** TBD

**Proposition 19 (Uniqueness of stationary distribution)** Any irreducible aperiodic Markov chain has a unique stationary distribution.

**Proof:** TBD

**Exercise 13** Consider the Markov chain on  $\{1, 2, 3, 4, 5, 6\}$  with the following transition matrix:

$$P = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 & 5 & 6 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{matrix} & \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ 0 & 0 & \frac{1}{4} & \frac{3}{4} & 0 & 0 \\ 0 & 0 & \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ 0 & 0 & 0 & 0 & \frac{3}{4} & \frac{1}{4} \\ 0 & 0 & 0 & 0 & \frac{1}{4} & \frac{3}{4} \end{pmatrix} \end{matrix} .$$

Show that this chain is reducible and it has three stationary distributions:

$$(1/2, 1/2, 0, 0, 0, 0), \quad (0, 0, 1/2, 1/2, 0, 0), \quad (0, 0, 0, 0, 1/2, 1/2) .$$

**Exercise 14** If there are two stationary distributions  $\pi$  and  $\pi'$  then show that there is a infinite family of stationary distributions  $\{\pi_p : p \in [0, 1]\}$ , called the convex combinations of  $\pi$  and  $\pi'$ .

**Exercise 15** Show that for a drunkard's walk chain started at state 0 around a polygonal block with  $k$  corners labelled  $\{0, 1, 2, \dots, k-1\}$ , the state probability vector at time step  $t$

$$\mu_t \xrightarrow{\text{TV}} \pi$$

if and only if  $k$  is odd. Explain what happens to  $\mu_t$  when  $k$  is even.

## 1.4 Reversibility

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

**Definition 20 (Reversible)** A probability distribution  $\pi$  on  $\mathbb{X} = \{s_1, s_2, \dots, s_k\}$  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  $(x, y) \in \mathbb{X}^2$ :

$$\pi(x)P(x, y) = \pi(y)P(y, x) . \quad (1.12)$$

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  $\pi$ , 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 21 (A reversible  $\pi$  is a stationary  $\pi$ )** Let  $(X_t)_{t \in \mathbb{Z}_+}$  be a Markov chain on  $\mathbb{X} = \{s_1, s_2, \dots, s_k\}$  with transition matrix  $P$ . If  $\pi$  is a reversible distribution for  $(X_t)_{t \in \mathbb{Z}_+}$  then  $\pi$  is a stationary distribution for  $(X_t)_{t \in \mathbb{Z}_+}$ .

**Proof:** Suppose  $\pi$  is a reversible distribution for  $(X_t)_{t \in \mathbb{Z}_+}$  then  $\pi$  is a probability distribution on  $\mathbb{X}$  and  $\pi(x)P(x, y) = \pi(y)P(y, x)$  for each  $(x, y) \in \mathbb{X}^2$ . We need to show that for any  $y \in \mathbb{X}$  we have

$$\pi(y) = \sum_{x \in \mathbb{X}} \pi(x)P(x, y) .$$

Fix a  $y \in \mathbb{X}$ ,

$$\begin{aligned} LHS &= \pi(y) = \pi(y) 1 = \pi(y) \sum_{x \in \mathbb{X}} P(y, x), \text{ since } P \text{ is a stochastic matrix} \\ &= \sum_{x \in \mathbb{X}} \pi(y)P(y, x) = \sum_{x \in \mathbb{X}} \pi(x)P(x, y), \text{ by reversibility} \\ &= RHS . \end{aligned}$$

**Definition 22 (Graph)** A **Graph**  $\mathbb{G} := (\mathbb{V}, \mathbb{E})$  consists of a **vertex set**  $\mathbb{V} := \{v_1, v_2, \dots, v_k\}$  together with an **edge set**  $\mathbb{E} := \{e_1, e_2, \dots, e_l\}$ . Each edge connects two of the vertices in  $\mathbb{V}$ . An edge  $e_h$  connecting vertices  $v_i$  and  $v_j$  is denoted by  $\langle v_i, v_j \rangle$ . Two vertices are **neighbours** if they share an edge. The **neighbourhood** of a vertex  $v_i$  denoted by  $\text{nbhd}(v_i) := \{v_j : \langle v_i, v_j \rangle \in \mathbb{E}\}$  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  $\deg(v_i) = \#\text{nbhd}(v_i)$ . 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 **directed** to preserve the order of the pair of vertices they connect or they can be **undirected**. An edge can be **weighted** by being associated with a real number called its weight. We can represent a directed graph by its **adjacency matrix** given by:

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

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  $\text{ideg}(v_i)$  and  $\text{odeg}(v_i)$  respectively. Note that a transition diagram of a Markov chain is a weighted directed graph and is represented by the transition probability matrix.

**Model 1 (Random Walk on an Undirected Graph)** A random walk on an undirected graph  $\mathbb{G} = (\mathbb{V}, \mathbb{E})$  is a Markov chain with state space  $\mathbb{V} := \{v_1, v_2, \dots, v_k\}$  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

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



**Proposition 23** The random walk on an undirected graph  $\mathbb{G} = (\mathbb{V}, \mathbb{E})$ , with vertex set  $\mathbb{V} := \{v_1, v_2, \dots, v_k\}$  and degree sum  $d = \sum_{i=1}^k \deg(v_i)$  is a reversible Markov chain with the reversible distribution  $\pi$  given by:

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

**Proof:** First note that  $\pi$  is a probability distribution provided that  $d > 0$ . To show that  $\pi$  is reversible we need to verify Equation 1.12 for each  $(v_i, v_j) \in \mathbb{V}^2$ . Fix a pair of states  $(v_i, v_j) \in \mathbb{V}^2$ , then

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

By Proposition 21  $\pi$  is also the stationary distribution.

**Exercise 16** Prove Proposition 23 by directly showing that  $\pi P = \pi$ , i.e., for each  $v_i \in \mathbb{V}$ ,  $\sum_{i=1}^k \pi(v_i)P(v_i, v_j) = \pi(v_j)$ .

**Example 17 (Random Walk on a regular graph)** A graph  $\mathbb{G} = (\mathbb{V}, \mathbb{E})$  is called regular if every vertex in  $\mathbb{V} = \{v_1, v_2, \dots, v_k\}$  has the same degree  $\delta$ , i.e.,  $\deg(v_i) = \delta$  for every  $v_i \in \mathbb{V}$ . Consider the random walk on a regular graph with symmetric transition matrix

$$Q(v_i, v_j) = \begin{cases} \frac{1}{\delta} & \text{if } \langle v_i, v_j \rangle \in \mathbb{E} \\ 0 & \text{otherwise} \end{cases} .$$

By Proposition 23, the stationary distribution of the random walk on  $\mathbb{G}$  is the uniform distribution on  $\mathbb{V}$  given by

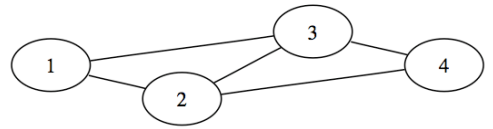
$$\pi = \left( \frac{\delta}{\delta \# \mathbb{V}}, \dots, \frac{\delta}{\delta \# \mathbb{V}} \right) = \left( \frac{1}{\# \mathbb{V}}, \dots, \frac{1}{\# \mathbb{V}} \right) .$$

**Example 18 (Triangulated Quadrangle)** The random walk on the undirected graph

$$\mathbb{G} = (\{1, 2, 3, 4\}, \{\langle 1, 2 \rangle, \langle 3, 1 \rangle, \langle 2, 3 \rangle, \langle 2, 4 \rangle, \langle 4, 3 \rangle\})$$

depicted below with adjacency matrix  $A$  is a Markov chain on  $\{1, 2, 3, 4\}$  with transition matrix  $P$ :

$$A = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 0 & 1 & 1 & 0 \end{pmatrix} \end{matrix}, \quad P = \begin{matrix} & \begin{matrix} 1 & 2 & 3 & 4 \end{matrix} \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \end{matrix} & \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix} \end{matrix},$$



By Proposition 23, the stationary distribution of the random walk on  $\mathbb{G}$  is

$$\pi = \left( \frac{\deg(v_1)}{d}, \frac{\deg(v_2)}{d}, \frac{\deg(v_3)}{d}, \frac{\deg(v_4)}{d} \right) = \left( \frac{2}{10}, \frac{3}{10}, \frac{3}{10}, \frac{2}{10} \right) .$$

**Example 19 (Drunkard's biased walk around the block)** Consider the Markov chain  $(X_t)_{t \in \mathbb{Z}_+}$  on  $\mathbb{X} = \{0, 1, 2, 3\}$  with initial distribution  $\mathbf{1}_{\{3\}}(x)$  and transition matrix

$$P = \begin{array}{c} \begin{array}{cccc} & 0 & 1 & 2 & 3 \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ 3 \end{array} & \begin{pmatrix} 0 & 1/3 & 0 & 2/3 \\ 1/3 & 0 & 2/3 & 0 \\ 0 & 1/3 & 0 & 2/3 \\ 1/3 & 0 & 2/3 & 0 \end{pmatrix} \end{array} \end{array}.$$

Draw the transition diagram for this Markov chain that corresponds to a drunkard who flips a biased coin to make his next move at each corner. The stationary distribution is  $\pi = (1/4, 1/4, 1/4, 1/4)$  (verify  $\pi P = \pi$ ).

We will show that  $(X_t)_{t \in \mathbb{Z}_+}$  is not a reversible Markov chain. Since  $(X_t)_{t \in \mathbb{Z}_+}$  is irreducible (aperiodicity is not necessary for uniqueness of  $\pi$ )  $\pi$  is the unique stationary distribution. Due to Proposition 21,  $\pi$  has to be a reversible distribution in order for  $(X_t)_{t \in \mathbb{Z}_+}$  to be a reversible Markov chain. But reversibility fails for  $\pi$  since,

$$\pi(0)P(0, 1) = \frac{1}{4} \times \frac{1}{3} = \frac{1}{12} < \frac{1}{6} = \frac{1}{4} \times \frac{2}{3} = \pi(1)P(1, 0).$$

**Exercise 20** Find the stationary distribution of the Markov chain in Exercise 12.

**Model 2 (Random Walk on a Directed Graph)** A random walk on a directed graph  $\mathbb{G} = (\mathbb{V}, \mathbb{E})$  is a Markov chain with state space  $\mathbb{V} := \{v_1, v_2, \dots, v_k\}$  and transition matrix given by:

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

**Example 21 (Directed Triangulated Quadrangle)** The random walk on the directed graph

$$\mathbb{G} = (\{1, 2, 3, 4\}, \{\langle 1, 2 \rangle, \langle 3, 1 \rangle, \langle 2, 3 \rangle, \langle 2, 4 \rangle, \langle 4, 3 \rangle\})$$

depicted below with adjacency matrix  $A$  is a Markov chain on  $\{1, 2, 3, 4\}$  with transition matrix  $P$ :

$$A = \begin{array}{c} \begin{array}{cccc} & 1 & 2 & 3 & 4 \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{array}, \quad P = \begin{array}{c} \begin{array}{cccc} & 1 & 2 & 3 & 4 \\ \begin{array}{c} 1 \\ 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & \frac{1}{2} \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \end{array}, \quad \begin{array}{c} \text{Diagram:} \\ \text{Nodes 1, 2, 3, 4 in a cycle with edges } 1 \rightarrow 2, 2 \rightarrow 3, 3 \rightarrow 1, 2 \rightarrow 4, 4 \rightarrow 3. \end{array}$$

**Exercise 22** Show that there is no reversible distribution for the Markov chain in Example 21.

**Example 23 (Random surf on the world wide web)** Consider the huge graph with vertices as webpages and hyper-links as undirected edges. Then Model 1 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.