12

Markov chains

Summary. The chapter begins with an introduction to discrete-time Markov chains, and to the use of matrix products and linear algebra in their study. The concepts of recurrence and transience are introduced, and a necessary and sufficient criterion for recurrence is proved. This leads to Pólya's theorem: symmetric random walk is recurrent in one and two dimensions, and transient in higher dimensions. It is shown how to calculate hitting probabilities and hitting times. Stopping times are introduced, and the strong Markov property is stated and proved. After a section on the classification of states, there is a discussion of invariant distributions. The ergodic theorem is proved for positive recurrent chains. A criterion for time-reversibility is presented, and applied in the special case of random walk on a finite graph.

12.1 The Markov property

A stochastic process is said to have the 'Markov property' if, conditional on its present value, its future is independent of its past. This is a very restrictive assumption, but it has two benefits. First, many processes in nature may be thus modelled, and secondly, the mathematical theory of such processes is strikingly beautiful and complete.

Let S be a countable set called the *state space*, and let $\mathbf{X} = (X_n : n \geq 0)$ be a sequence of random variables taking values in S. The X_n are functions on some common probability space, but we shall not be specific about that. The following is an informal way of explaining what it means to be a Markov chain: the sequence $\mathbf{X}$ is a Markov chain if, conditional on the present value X_n, the future $(X_r : r > n)$ is independent of the past $(X_m : m < n)$.

Definition 12.1 *The sequence* $\mathbf{X}$ *is called a* ***Markov chain*** *if it satisfies the* ***Markov property***

$$\mathbb{P}\big(X_{n+1} = i_{n+1} \,\big|\, X_0 = i_0,\ X_1 = i_1, \dots,\ X_n = i_n\big) = \mathbb{P}\big(X_{n+1} = i_{n+1} \,\big|\, X_n = i_n\big) \tag{12.2}$$

for all $n \geq 0$ *and all* $i_0, i_1, \dots, i_{n+1} \in S$. *The chain is called* ***homogeneous*** *if, for all* $i, j \in S$, *the conditional probability* $\mathbb{P}(X_{n+1} = j \mid X_n = i)$ *does not depend on the value of* n.

Here are some examples of Markov chains. Each has a coherent theory relying on an assumption of independence tantamount to the Markov property.

(a) **(Branching processes)** The branching process of Chapter 9 is a simple model of the growth of a population. Each member of the nth generation has a number of offspring that is independent of the past.
(b) **(Random walk)** A particle performs a random walk on the line, as in Chapter 10. At each epoch of time, it jumps a random distance that is independent of previous jumps.
(c) **(Poisson process)** The Poisson process of Section 11.2 is a Markov chain but in *continuous* time. Arrivals after time t are independent of arrivals before t.
(d) **(An example from finance)** Here is an example of a Markov model in action. In an idealized financial model, a 'stock' price S_n is such that $\log S_n$ performs a type of random walk on $\mathbb{R}$, while a 'bond' accumulates value at a constant interest rate. A so-called 'European call option' permits a buyer to purchase one unit of stock at a given future time and price. The problem is to determine the correct value of this option at time 0. The answer is known as the Black–Scholes formula. A key element of the theory is that the stock satisfies the Markov property.

The basic theory of Markov chains is presented in this chapter. For simplicity, *all Markov chains here will be assumed to be homogeneous*. In order to calculate probabilities associated with such a chain, we need to know two quantities:

(a) the *transition matrix* $P = (p_{i,j} : i, j \in S)$ given by $p_{i,j} = \mathbb{P}(X_1 = j \mid X_0 = i)$, and
(b) the *initial distribution* $\lambda = (\lambda_i : i \in S)$ given by $\lambda_i = \mathbb{P}(X_0 = i)$.

By the assumption of homogeneity,

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = p_{i,j} \qquad \text{for } n \geq 0.$$

The pair (λ, P) is characterized as follows.

Proposition 12.3

(a) *The vector λ is a* **distribution** *in that $\lambda_i \geq 0$ for $i \in S$, and $\sum_{i \in S} \lambda_i = 1$.*
(b) *The matrix $P = (p_{i,j})$ is a* **stochastic matrix** *in that*
   (i) *$p_{i,j} \geq 0$ for $i, j \in S$, and*
   (ii) *$\sum_{j \in S} p_{i,j} = 1$ for $i \in S$, so that P has row sums 1.*

Proof (a) Since λ_i is a probability, it is non-negative. Also,

$$\sum_{i \in S} \lambda_i = \sum_{i \in S} \mathbb{P}(X_0 = i) = 1.$$

(b) Since $p_{i,j}$ is a probability, it is non-negative. Finally,

$$\begin{aligned}\sum_{j \in S} p_{i,j} &= \sum_{j \in S} \mathbb{P}(X_1 = j \mid X_0 = i) \\ &= \mathbb{P}(X_1 \in S \mid X_0 = i) = 1.\end{aligned}$$

□

Theorem 12.4 *Let λ be a distribution and P a stochastic matrix. The random sequence* $\mathbf{X}$ *is a Markov chain with initial distribution λ and transition matrix P if and only if*

$$\mathbb{P}\big(X_0 = i_0,\ X_1 = i_1,\ \dots,\ X_n = i_n\big) = \lambda_{i_0} p_{i_0,i_1} \cdots p_{i_{n-1},i_n} \tag{12.5}$$

for all $n \geq 0$ and $i_0, i_1, \dots, i_n \in S$.

Proof Write A_k for the event $\{X_k = i_k\}$, so that (12.5) may be written as

$$\mathbb{P}(A_0 \cap A_1 \cap \dots \cap A_n) = \lambda_{i_0} p_{i_0,i_1} \cdots p_{i_{n-1},i_n}. \tag{12.6}$$

Suppose $\mathbf{X}$ is a Markov chain with initial distribution λ and transition matrix P. We prove (12.6) by induction on n. It holds trivially when $n = 0$. Suppose N (≥ 1) is such that (12.6) holds for $n < N$. Then

$$\begin{aligned}\mathbb{P}(A_0 \cap A_1 \cap \dots \cap A_N) &= \mathbb{P}(A_0 \cap A_1 \cap \dots \cap A_{N-1})\mathbb{P}\big(A_N \,\big|\, A_0 \cap A_1 \cap \dots \cap A_{N-1}\big)\\ &= \mathbb{P}(A_0 \cap A_1 \cap \dots \cap A_{N-1})\mathbb{P}(A_N \mid A_{N-1})\end{aligned}$$

by the Markov property. Now $\mathbb{P}(A_N \mid A_{N-1}) = p_{i_{N-1},i_N}$, and the induction step is complete.

Suppose conversely that (12.6) holds for all n and sequences (i_m). Setting $n = 0$ we obtain the initial distribution $\mathbb{P}(X_0 = i_0) = \lambda_{i_0}$. Now,

$$\mathbb{P}\big(A_{n+1} \,\big|\, A_0 \cap A_1 \cap \dots \cap A_n\big) = \frac{\mathbb{P}(A_0 \cap A_1 \cap \dots \cap A_{n+1})}{\mathbb{P}(A_0 \cap A_1 \cap \dots \cap A_n)}$$

so that, by (12.6),

$$\mathbb{P}\big(A_{n+1} \,\big|\, A_0 \cap A_1 \cap \dots \cap A_n\big) = p_{i_n,i_{n+1}}.$$

Therefore, $\mathbf{X}$ is a homogeneous Markov chain with transition matrix P. □

The Markov property (12.2) asserts in essence that the past affects the future only via the present. This is made formal in the next theorem, in which X_n is the present value, F is a future event, and H is a historical event.

Theorem 12.7 (Extended Markov property) *Let* $\mathbf{X}$ *be a Markov chain. For $n \geq 0$, for any event H given in terms of the past history $X_0, X_1, \dots, X_{n-1}$, and any event F given in terms of the future $X_{n+1}, X_{n+2}, \dots$,*

$$\mathbb{P}(F \mid X_n = i,\ H) = \mathbb{P}(F \mid X_n = i) \qquad \text{for } i \in S. \tag{12.8}$$

Proof A slight complication arises from the fact that F may depend on the *infinite* future. There is a general argument of probability theory that allows us to restrict ourselves to the case when F depends on the values of the process at only *finitely* many times, and we do not explain this here.

By the definition of conditional probability and Theorem 12.4,

$$\begin{aligned}\mathbb{P}(F \mid X_n = i,\ H) &= \frac{\mathbb{P}(H,\ X_n = i,\ F)}{\mathbb{P}(H,\ X_n = i)}\\ &= \frac{\sum_{<n}\sum_{>n}\lambda_{i_0}p_{i_0,i_1}\cdots p_{i_{n-1},i}\,p_{i,i_{n+1}}\cdots}{\sum_{<n}\lambda_{i_0}p_{i_0,i_1}\cdots p_{i_{n-1},i}}\\ &= \sum_{>n} p_{i,i_{n+1}}p_{i_{n+1},i_{n+2}}\cdots\\ &= \mathbb{P}(F \mid X_n = i),\end{aligned}$$

where $\sum_{<n}$ sums over all sequences $(i_0, i_1, \dots, i_{n-1})$ corresponding to the event H, and $\sum_{>n}$ sums over all sequences $(i_{n+1}, i_{n+2}, \dots)$ corresponding to the event F. □

Exercise 12.9 Let X_n be the greatest number shown in the first n throws of a fair six-sided die. Show that $\mathbf{X} = (X_n : n \geq 1)$ is a homogeneous Markov chain, and write down its transition probabilities.

Exercise 12.10 Let $\mathbf{X}$ and $\mathbf{Y}$ be symmetric random walks on the line $\mathbb{Z}$. Is $\mathbf{X} + \mathbf{Y}$ necessarily a Markov chain? Explain.

Exercise 12.11 A square matrix with non-negative entries is called *doubly stochastic* if all its row-sums and column-sums equal 1. If P is doubly stochastic, show that P^n is doubly stochastic for $n \geq 1$.

12.2 Transition probabilities

Let $\mathbf{X}$ be a Markov chain with transition matrix $P = (p_{i,j})$. The elements $p_{i,j}$ are called the *one-step transition probabilities*. More generally, the *n-step transition probabilities* are given by

$$p_{i,j}(n) = \mathbb{P}(X_n = j \mid X_0 = i),$$

and they form a matrix called the *n-step transition matrix* $P(n) = (p_{i,j}(n) : i, j \in S)$. The matrices $P(n)$ satisfy a collection of equations named after Chapman and Kolmogorov.

Theorem 12.12 (Chapman–Kolmogorov equations) *We have that*

$$p_{i,j}(m+n) = \sum_{k \in S} p_{i,k}(m)p_{k,j}(n)$$

for $i, j \in S$ *and* $m, n \geq 0$. *That is to say,* $P(m+n) = P(m)P(n)$.

Proof By the definition of conditional probability,

$$\begin{aligned}p_{i,j}(m+n) &= \mathbb{P}(X_{m+n} = j \mid X_0 = i)\\ &= \sum_{k\in S}\mathbb{P}\big(X_{m+n} = j \,\big|\, X_m = k,\ X_0 = i\big)\mathbb{P}(X_m = k \mid X_0 = i).\end{aligned} \tag{12.13}$$

By the extended Markov property, Theorem 12.7,

$$\mathbb{P}\big(X_{m+n} = j \,\big|\, X_m = k,\ X_0 = i\big) = \mathbb{P}(X_{m+n} = j \mid X_m = k),$$

and the claim follows. □

By the Chapman–Kolmogorov equations, Theorem 12.12, the n-step transition probabilities form a matrix $P(n) = (p_{i,j}(n))$ that satisfies $P(n) = P(1)^n = P^n$. One way of calculating the probabilities $p_{i,j}(n)$ is therefore to find the nth power of the matrix P. When the state space is finite, then so is P, and this calculation is usually done best by diagonalizing P. We illustrate this with an example.

Example 12.14 (Two-state Markov chain) Suppose $S = \{1, 2\}$ and

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$$

where $\alpha, \beta \in (0, 1)$. Find the n-step transition probabilities.

Solution A (by diagonalization) In order to calculate the n-step transition matrix P^n, we shall diagonalize P. The eigenvalues κ of P are the roots of the equation $\det(P - \kappa I) = 0$, which is to say that $(1-\alpha-\kappa)(1-\beta-\kappa) - \alpha\beta = 0$, with solutions

$$\kappa_1 = 1, \qquad \kappa_2 = 1 - \alpha - \beta.$$

Therefore,

$$P = U^{-1} \begin{pmatrix} 1 & 0 \\ 0 & 1-\alpha-\beta \end{pmatrix} U$$

for some invertible matrix U. It follows that

$$P^n = U^{-1} \begin{pmatrix} 1 & 0 \\ 0 & (1-\alpha-\beta)^n \end{pmatrix} U,$$

and so

$$p_{1,1}(n) = A + B(1-\alpha-\beta)^n, \tag{12.15}$$

for some constants A, B which are found as follows. Since $p_{1,1}(0) = 1$ and $p_{1,1}(1) = 1 - \alpha$, we have that $A + B = 1$ and $A + B(1-\alpha-\beta) = 1 - \alpha$. Therefore,

$$A = \frac{\beta}{\alpha+\beta}, \qquad B = \frac{\alpha}{\alpha+\beta}.$$

Now, $p_{1,2}(n) = 1 - p_{1,1}(n)$, and $p_{2,2}(n)$ is found by interchanging α and β. In summary,

$$P^n = \frac{1}{\alpha+\beta} \begin{pmatrix} \beta + \alpha(1-\alpha-\beta)^n & \alpha - \alpha(1-\alpha-\beta)^n \\ \beta - \beta(1-\alpha-\beta)^n & \alpha + \beta(1-\alpha-\beta)^n \end{pmatrix}.$$

We note for future reference that

$$P^n \to \frac{1}{\alpha+\beta} \begin{pmatrix} \beta & \alpha \\ \beta & \alpha \end{pmatrix} \qquad \text{as } n \to \infty,$$

which is to say that

$$p_{i,1}(n) \to \frac{\beta}{\alpha+\beta}, \quad p_{i,2}(n) \to \frac{\alpha}{\alpha+\beta} \qquad \text{for } i = 1, 2.$$

This conclusion may be stated as follows. The distribution of X_n settles down to a limiting distribution $(\beta, \alpha)/(\alpha + \beta)$, which does not depend on the choice of initial state i. This hints at a general property of Markov chains to which we shall return in Sections 12.9–12.10.

Solution B (by difference equations) By conditioning on X_n (or, alternatively, by the Chapman–Kolmogorov equations),

$$\begin{aligned}p_{1,1}(n+1) &= \mathbb{P}(X_{n+1} = 1 \mid X_0 = 1)\\ &= \mathbb{P}(X_{n+1} = 1 \mid X_n = 1)p_{1,1}(n) + \mathbb{P}(X_{n+1} = 1 \mid X_n = 2)p_{1,2}(n)\\ &= (1-\alpha)p_{1,1}(n) + \beta p_{1,2}(n)\\ &= (1-\alpha)p_{1,1}(n) + \beta(1 - p_{1,1}(n)).\end{aligned}$$

This is a difference equation with boundary condition $p_{1,1}(0) = 1$. Solving it in the usual way, we obtain (12.15). △

Finally, we summarize the matrix method illustrated in Example 12.14. Suppose the state space is finite, $|S| = N$ say, so that P is an $N \times N$ matrix. It is a general result for stochastic matrices[1] that $\kappa_1 = 1$ is an eigenvalue of P, and no other eigenvalue has larger absolute value. We write $\kappa_1\ (= 1), \kappa_2, \dots, \kappa_N$ for the (possibly complex) eigenvalues of P, arranged in decreasing order of absolute value. *We assume for simplicity that the κ_i are distinct*, since the diagonalization of P is more complicated otherwise. There exists an invertible matrix U such that $P = U^{-1}KU$ where K is the diagonal matrix with entries $\kappa_1, \kappa_2, \dots, \kappa_N$. Then

$$P^n = (U^{-1}KU)^n = U^{-1}K^nU = U^{-1}\begin{pmatrix}\kappa_1^n & 0 & \cdots & 0\\ 0 & \kappa_2^n & \cdots & 0\\ \vdots & \vdots & \ddots & 0\\ 0 & 0 & \cdots & \kappa_N^n\end{pmatrix}U, \tag{12.16}$$

from which the individual probabilities $p_{i,j}(n)$ may in principle be found.

The situation is considerably simpler if the chain has two further properties that will be encountered soon, namely 'irreducibility' (see Section 12.3) and 'aperiodicity' (see Definition 12.72 and Theorem 12.73). Under these conditions, by the Perron–Frobenius theorem, $\kappa_1 = 1$ is the unique eigenvalue with absolute value 1, so that $\kappa_k^n \to 0$ as $n \to \infty$, for $k \geq 2$. By (12.16), the long-run transition probabilities of the chain satisfy

$$P^n \to U^{-1}\begin{pmatrix}1 & 0 & \cdots & 0\\ 0 & 0 & \cdots & 0\\ \vdots & \vdots & \ddots & 0\\ 0 & 0 & \cdots & 0\end{pmatrix}U \qquad \text{as } n \to \infty. \tag{12.17}$$

One may gain further information from (12.17) as follows. The rows of U are the normalized left eigenvectors of P, and the columns of U^{-1} are the normalized right eigenvectors. Since P is stochastic, $P\mathbf{1}' = \mathbf{1}'$, where $\mathbf{1}'$ is the column-vector of ones. Therefore, the first column of U^{-1} is constant. By examining the product in (12.17), we find that $p_{i,j}(n) \to \pi_j$ for some vector $\pi = (\pi_j : j \in S)$ that does not depend on the initial state i.

[1]This is part of the so-called Perron–Frobenius theorem, for which the reader is referred to Grimmett and Stirzaker (2001, Sect. 6.6).

Remark 12.18 (Markov chains and linear algebra) Much of the theory of Markov chains involves the manipulation of vectors and matrices. The equations are usually linear, and thus much of the subject can be phrased in the language of linear algebra. For example, if X_0 has distribution λ, interpreted as a row-vector $(\lambda_i : i \in S)$, then

$$\mathbb{P}(X_1 = j) = \sum_{i \in S} \lambda_i p_{i,j} \qquad \text{for } j \in S,$$

so that X_1 has as distribution the row-vector λP. By iteration, X_2 has distribution λP^2, and so on. We therefore adopt the convention that probability distributions are by default row-vectors, and they act on the left side of matrices. Thus, λ' denotes the transpose of the row-vector λ, and is itself a column-vector.

Exercise 12.19 Let $\mathbf{X}$ be a Markov chain with transition matrix P, and let $d \geq 1$. Show that $Y_n = X_{nd}$ defines a Markov chain with transition matrix P^d.

Exercise 12.20 A fair coin is tossed repeatedly. Show that the number H_n of heads after n tosses forms a Markov chain.

Exercise 12.21 A flea hops randomly between the vertices of a triangle. Find the probability that it is back at its starting point after n hops.

12.3 Class structure

An important element in the theory of Markov chains is the interaction between the state space S and the transition mechanism P.

Let $\mathbf{X}$ be a homogeneous Markov chain with state space S and transition matrix P. For $i, j \in S$, we say that i *leads to* j, written $i \to j$, if $p_{i,j}(n) > 0$ for some $n \geq 0$. By setting $n = 0$ we have that $i \to i$ for all $i \in S$. We write $i \leftrightarrow j$ if $i \to j$ and $j \to i$, and in this case we say that i and j *communicate*.

Proposition 12.22 *The relation $\leftrightarrow$ is an equivalence relation.*

Proof Since $i \to i$, we have that $i \leftrightarrow i$. Suppose that $i, j, k \in S$ satisfy $i \leftrightarrow j$ and $j \leftrightarrow k$. Since $i \to j$ and $j \to k$, there exist $m, n \geq 0$ such that $p_{i,j}(m) > 0$ and $p_{j,k}(n) > 0$. By the Chapman–Kolmogorov equations, Theorem 12.12,

$$\begin{aligned} p_{i,k}(m+n) &= \sum_{l \in S} p_{i,l}(m) p_{l,k}(n) \\ &\geq p_{i,j}(m) p_{j,k}(n) > 0, \end{aligned}$$

so that $i \to k$. Similarly, $k \to i$, and hence $i \leftrightarrow k$. □

Since $\leftrightarrow$ is an equivalence relation, it has *equivalence classes*, namely the subsets of S of the form $C_i = \{j \in S : i \leftrightarrow j\}$. These classes are called *communicating classes*. The chain $\mathbf{X}$ (or the state space S) is called *irreducible* if there is a single equivalence class, which is to say that $i \leftrightarrow j$ for all $i, j \in S$.

A subset $C \subseteq S$ is called *closed* if

$$i \in C, \ i \to j \quad \Rightarrow \quad j \in C. \tag{12.23}$$

If the chain ever hits a closed set C, then it remains in C forever afterwards. If a singleton set $\{i\}$ is closed, we call i an *absorbing* state.

Proposition 12.24 *A subset C of states is closed if and only if*

$$p_{i,j} = 0 \qquad \text{for } i \in C, \ j \notin C. \tag{12.25}$$

Proof Let $C \subseteq S$. If (12.25) fails then so does (12.23), and C is not closed.

Suppose conversely that (12.25) holds. Let $k \in C$, $l \in S$ be such that $k \to l$. Since $k \to l$, there exists $m \geq 0$ such that $\mathbb{P}(X_m = l \mid X_0 = k) > 0$, and so there exists a sequence $k_0\,(= k), k_1, \dots, k_m\,(= l)$ with $p_{k_r,k_{r+1}} > 0$ for $r = 0, 1, \dots, m-1$. By (12.25), $k_r \in C$ for all r, so that $l \in C$. Statement (12.23) follows. □

Example 12.26 Let $S = \{1, 2, 3, 4, 5, 6\}$ and

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

Possible transitions of the chain are illustrated in Figure 12.1. The equivalence classes are $C_1 = \{1, 2, 3\}$, $C_2 = \{4\}$, and $C_3 = \{5, 6\}$. The classes C_1 and C_2 are not closed, but C_3 is closed. △

Exercise 12.27 Find the communicating classes, and the closed communicating classes, when the transition matrix is

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

It may be useful to draw a diagram.

Exercise 12.28 If the state space is finite, show that there must exist at least one closed communicating class. Give an example of a transition matrix with no such class.

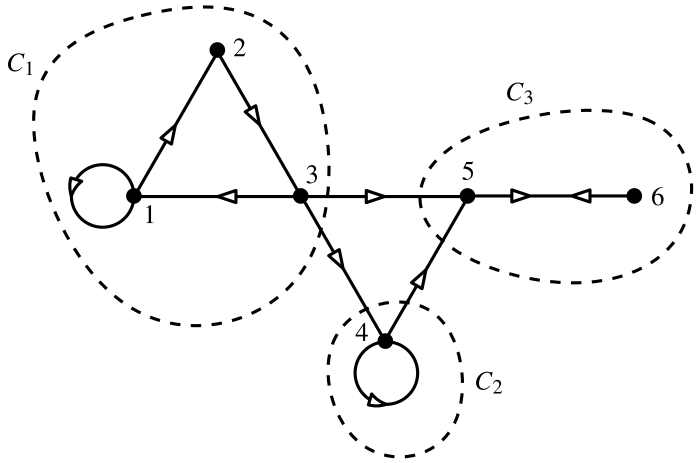

Fig. 12.1 The arrows indicate possible transitions of the chain of Example 12.26. The communicating classes are circled.

12.4 Recurrence and transience

Let $\mathbf{X}$ be a homogeneous Markov chain with state space S and transition matrix P. The better to economize on notation, we write henceforth $\mathbb{P}_i(A)$ for $\mathbb{P}(A \mid X_0 = i)$, and similarly $\mathbb{E}_i(Z)$ for the mean of a random variable Z conditional on the event $X_0 = i$.

The *passage-time* to state j is defined as

$$T_j = \min\{n \geq 1 : X_n = j\},$$

and the *first-passage probabilities* are given by

$$f_{i,j}(n) = \mathbb{P}_i(T_j = n).$$

If a chain starts in state i, is it bound to return to i at some later time?

Definition 12.29 *A state i is called* ***recurrent***[2] *if $\mathbb{P}_i(T_i < \infty) = 1$. A state is called* ***transient*** *if it is not recurrent.*

Here is a criterion for recurrence in terms of the transition matrix P and its powers.

Theorem 12.30 *The state i is recurrent if and only if*

$$\sum_{n=0}^{\infty} p_{i,i}(n) = \infty.$$

[2]The word 'persistent' is also used in this context.

In proving this, we shall make use of the following Theorem 12.33. It was proved in Theorem 10.12 that simple random walk on the line is recurrent if and only it is unbiased. The proof used generating functions, and this method may be extended to prove Theorem 12.30. We introduce next the generating functions in this more general context. For $i, j \in S$, let

$$P_{i,j}(s) = \sum_{n=0}^{\infty} p_{i,j}(n)s^n, \qquad F_{i,j}(s) = \sum_{n=0}^{\infty} f_{i,j}(n)s^n,$$

with the conventions that $f_{i,j}(0) = 0$ and $p_{i,j}(0) = \delta_{i,j}$, the Kronecker delta defined by

$$\delta_{i,j} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{otherwise.} \end{cases} \tag{12.31}$$

We let

$$f_{i,j} = F_{i,j}(1) = \mathbb{P}_i(T_j < \infty). \tag{12.32}$$

and note that i is recurrent if and only if $f_{i,i} = 1$.

Theorem 12.33 *For $i, j \in S$, we have that*

$$P_{i,j}(s) = \delta_{i,j} + F_{i,j}(s)P_{j,j}(s), \qquad s \in (-1, 1].$$

Proof Using conditional probability and the extended Markov property, Theorem 12.7, for $n \ge 1$,

$$p_{i,j}(n) = \sum_{m=1}^{\infty} \mathbb{P}_i(X_n = j \mid T_j = m)\mathbb{P}_i(T_j = m). \tag{12.34}$$

The summand is 0 for $m > n$, since in this case the first passage to j has not taken place by time n. For $m \le n$,

$$\mathbb{P}_i(X_n = j \mid T_j = m) = \mathbb{P}_i(X_n = j \mid X_m = j,\ H),$$

where $H = \{X_r \ne j \text{ for } 1 \le r < m\}$ is an event defined prior to time m. By the extended Markov property,

$$\mathbb{P}_i(X_n = j \mid T_j = m) = \mathbb{P}(X_n = j \mid X_m = j) = \mathbb{P}_j(X_{n-m} = j).$$

We substitute this into (12.34) to obtain

$$p_{i,j}(n) = \sum_{m=1}^{n} p_{j,j}(n-m) f_{i,j}(m).$$

Multiply through this equation by s^n and sum over $n \ge 1$ to obtain

$$P_{i,j}(s) - p_{i,j}(0) = P_{j,j}(s)F_{i,j}(s).$$

The claim follows since $p_{i,j}(0) = \delta_{i,j}$. □

Proof of Theorem 12.30 By Theorem 12.33 with $i = j$,

$$P_{i,i}(s) = \frac{1}{1 - F_{i,i}(s)} \qquad \text{for } |s| < 1. \tag{12.35}$$

In the limit as $s \uparrow 1$, we have by Abel's lemma that

$$F_{i,i}(s) \uparrow F_{i,i}(1) = f_{i,i}, \qquad P_{i,i}(s) \uparrow \sum_{n=0}^{\infty} p_{i,i}(n).$$

By (12.35),

$$\sum_{n=0}^{\infty} p_{i,i}(n) = \infty \quad \text{if and only if} \quad f_{i,i} = 1,$$

as claimed. □

The property of recurrence is called a *class property*, in that any pair of communicating states are either both recurrent or both transient.

Theorem 12.36 *Let C be a communicating class.*

(a) *Either every state in C is recurrent or every state is transient.*
(b) *Suppose C contains some recurrent state. Then C is closed.*

Proof (a) Let $i \leftrightarrow j$ and $i \neq j$. By Theorem 12.30, it suffices to show that

$$\sum_{n=0}^{\infty} p_{i,i}(n) = \infty \quad \text{if and only if} \quad \sum_{n=0}^{\infty} p_{j,j}(n) = \infty. \tag{12.37}$$

Since $i \leftrightarrow j$, there exist $m, n \geq 1$ such that

$$\alpha := p_{i,j}(m)p_{j,i}(n) > 0.$$

By the Chapman–Kolmogorov equations, Theorem 12.12,

$$p_{i,i}(m + r + n) \geq p_{i,j}(m)p_{j,j}(r)p_{j,i}(n) = \alpha p_{j,j}(r) \qquad \text{for } r \geq 0.$$

We sum over r to obtain

$$\sum_{r=0}^{\infty} p_{i,i}(m + r + n) \geq \alpha \sum_{r=0}^{\infty} p_{j,j}(r).$$

Therefore, $\sum_r p_{i,i}(r) = \infty$ whenever $\sum_r p_{j,j}(r) = \infty$. The converse holds similarly, and (12.37) is proved.

(b) Assume $i \in C$ is recurrent, and C is not closed. By Proposition 12.24, there exist $j \in C$, $k \notin C$ such that $p_{j,k} > 0$ and $k \nrightarrow j$. By part (a), j is recurrent. However,

$$\mathbb{P}_j(X_n \neq j \text{ for all } n \geq 1) \geq \mathbb{P}_j(X_1 = k) = p_{j,k} > 0,$$

a contradiction. Therefore, C is closed. □

Theorem 12.38 *Suppose that the state space S is finite.*

(a) *There exists at least one recurrent state.*
(b) *If the chain is irreducible, all states are recurrent.*

Here is a preliminary result which will be useful later.

Proposition 12.39 *Let $i, j \in S$. If j is transient, then $p_{i,j}(n) \to 0$ as $n \to \infty$.*

Proof of Proposition 12.39 Let j be transient. By Theorem 12.30 and Abel's lemma, $P_{j,j}(1) < \infty$. By Theorem 12.33, $P_{i,j}(1) < \infty$, and hence the nth term in this sum, $p_{i,j}(n)$, tends to zero as $n \to \infty$. □

Proof of Theorem 12.38 Suppose $|S| < \infty$.

(a) We have that

$$1 = \sum_{j \in S} \mathbb{P}_i(X_n = j) = \sum_{j \in S} p_{i,j}(n). \tag{12.40}$$

Assume every state is transient. By Proposition 12.39, for all $j \in S$, $p_{i,j}(n) \to 0$ as $n \to \infty$. This contradicts (12.40).

(b) Suppose the chain is irreducible. By Theorem 12.36, either every state is recurrent or every state is transient, and the claim follows by part (a). □

Exercise 12.41 A Markov chain $\mathbf{X}$ has an absorbing state s to which all other states lead. Show that all states except s are transient.

Exercise 12.42

(a) Let j be a recurrent state of a Markov chain. Show that $\sum_n p_{i,j}(n) = \infty$ for all states i such that $i \to j$.
(b) Let j be a transient state of a Markov chain. Show that $\sum_n p_{i,j}(n) < \infty$ for all states i.

12.5 Random walks in one, two, and three dimensions

One-dimensional random walks were explored in some detail in Chapter 10. The purpose of the current section is to extend the theory to higher dimensions within the context of Markov chains.

The graphs of this section are the d-dimensional lattices. Let $\mathbb{Z} = \{\dots, -1, 0, 1 \dots\}$ denote the integers, and let $\mathbb{Z}^d$ be the set of all d-vectors of integers, written $x = (x_1, x_2, \dots, x_d)$ with each $x_i \in \mathbb{Z}$. The set $\mathbb{Z}^d$ may be interpreted as a graph with vertex-set $\mathbb{Z}^d$, and with edges joining any two vectors x, y which are separated by Euclidean distance 1. Two such vertices are declared *adjacent*, and are said to be *neighbours*. We denote the ensuing graph by $\mathbb{Z}^d$ also, and note that each vertex has exactly $2d$ neighbours. The graphs $\mathbb{Z}$ and $\mathbb{Z}^2$ are drawn in Figure 12.2.

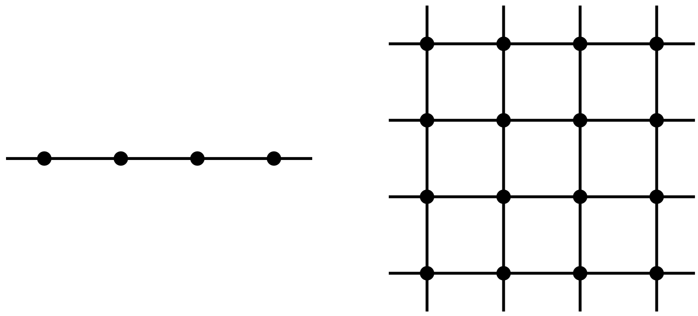

Fig. 12.2 The line $\mathbb{Z}$ and the square lattice $\mathbb{Z}^2$.

Let $d \geq 1$. The symmetric random walk on $\mathbb{Z}^d$ is the Markov chain on the state space $\mathbb{Z}^d$ which, at each step, jumps to a uniformly chosen neighbour. The transition matrix is given by

$$p_{x,y} = \begin{cases} \dfrac{1}{2d} & \text{if } y \text{ is a neighbour of } x, \\ 0 & \text{otherwise.} \end{cases}$$

It is clear that the chain is irreducible. By Theorem 12.36, either every state is recurrent or every state is transient.

Theorem 12.43 (Pólya's theorem) *The symmetric random walk on $\mathbb{Z}^d$ is recurrent if $d = 1, 2$ and transient if $d \geq 3$.*

The case $d = 1$ was proved at Theorem 10.12, and the cases $d = 2, 3$ featured in Exercise 10.11 and Problems 10.5.9 and 10.5.13.

Proof Let $d = 1$ and $X_0 = 0$. The walker can return to 0 only after an even number of steps. The probability of return after $2n$ steps is the probability that, of the first $2n$ steps, exactly n are to the right. Therefore,

$$p_{0,0}(2n) = \left(\frac{1}{2}\right)^{2n} \binom{2n}{n}. \tag{12.44}$$

By Stirling's formula, Theorem A.4,

$$p_{0,0}(2n) = \left(\frac{1}{2}\right)^{2n} \frac{(2n)!}{(n!)^2} \sim \frac{1}{\sqrt{\pi n}}. \tag{12.45}$$

In particular, $\sum_n p_{0,0}(2n) = \infty$. By Theorem 12.30, the state 0 is recurrent.

Suppose that $d = 2$. There is a clever but special way to handle this case, which we defer until after this proof. Instead we develop next a method that works also when $d \geq 3$. The walk is at the origin $\mathbf{0}$ at time $2n$ if and only if it has taken equal numbers of leftward and rightward steps, and also equal numbers of upward and downward steps. Therefore,

$$p_{\mathbf{0},\mathbf{0}}(2n) = \left(\frac{1}{4}\right)^{2n} \sum_{m=0}^{n} \frac{(2n)!}{[m!\,(n-m)!]^2}.$$

Now,

$$\sum_{m=0}^{n} \frac{(2n)!}{[m!\,(n-m)!]^2} = \binom{2n}{n} \sum_{m=0}^{n} \binom{n}{m}\binom{n}{n-m} = \binom{2n}{n}^2,$$

by (A.2). Therefore,

$$p_{\mathbf{0},\mathbf{0}}(2n) = \left(\frac{1}{2}\right)^{4n} \binom{2n}{n}^2. \tag{12.46}$$

This is simply the square of the one-dimensional answer (12.44) (this is no coincidence), so that

$$p_{\mathbf{0},\mathbf{0}}(2n) \sim \frac{1}{\pi n}. \tag{12.47}$$

Therefore, $\sum_n p_{\mathbf{0},\mathbf{0}}(2n) = \infty$, and hence $\mathbf{0}$ is recurrent.

Suppose finally that $d = 3$, the general case $d \geq 3$ is handled similarly. By the argument that led to (12.46), and a little reorganization,

$$\begin{aligned} p_{\mathbf{0},\mathbf{0}}(2n) &= \left(\frac{1}{6}\right)^{2n} \sum_{i+j+k=n} \frac{(2n)!}{(i!\,j!\,k!)^2} \\ &= \left(\frac{1}{2}\right)^{2n} \binom{2n}{n} \sum_{i+j+k=n} \left(\frac{n!}{3^n i!\,j!\,k!}\right)^2 \\ &\leq \left(\frac{1}{2}\right)^{2n} \binom{2n}{n} M \sum_{i+j+k=n} \frac{n!}{3^n i!\,j!\,k!}, \end{aligned} \tag{12.48}$$

where

$$M = \max\left\{ \frac{n!}{3^n i!\,j!\,k!} : i, j, k \geq 0,\ i + j + k = n \right\}.$$

It is not difficult to see that the maximum M is attained when i, j, and k are all closest to $\frac{1}{3}n$, so that

$$M \leq \frac{n!}{3^n \left(\lfloor \frac{1}{3}n \rfloor !\right)^3}.$$

Furthermore, the final summation in (12.48) equals 1, since the summand is the probability that, in allocating n balls randomly to three urns, the urns contain respectively i, j, and k balls. It follows that

$$p_{\mathbf{0},\mathbf{0}}(2n) \leq \frac{(2n)!}{12^n n! \left(\lfloor \frac{1}{3}n \rfloor !\right)^3}$$

which, by Stirling's formula, is no bigger than $Cn^{-\frac{3}{2}}$ for some constant C. Therefore,

$$\sum_{n=0}^{\infty} p_{\mathbf{0},\mathbf{0}}(2n) < \infty,$$

implying that the origin $\mathbf{0}$ is transient. □

This section closes with an account of the 'neat' way of studying the two-dimensional random walk (see also Problem 10.5.10). It is the precisely 'squared' form of (12.46) that suggests an explanation using independence. Write $X_n = (A_n, B_n)$ for the position of the walker at time n, and let $Y_n = (U_n, V_n)$ where

$$U_n = A_n - B_n, \quad V_n = A_n + B_n.$$

Thus, Y_n is derived from X_n by referring to a rotated and re-scaled coordinate system, as illustrated in Figure 12.3.

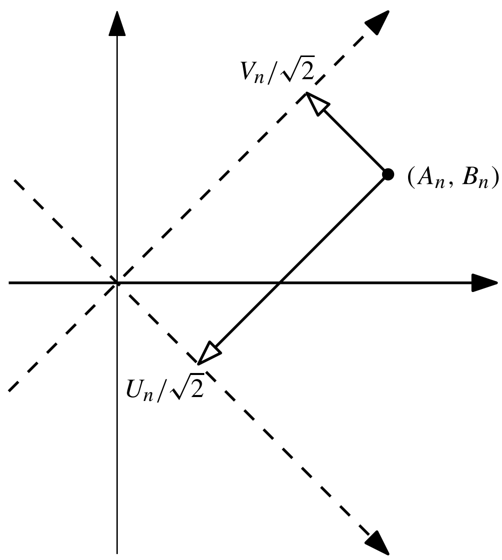

Fig. 12.3 The new coordinate system for the process $\mathbf{Y} = (Y_n)$.

The key fact is that $\mathbf{U} = (U_n)$ and $\mathbf{V} = (V_n)$ are independent, symmetric random walks on the line $\mathbb{Z}$. This is checked by a set of four calculations of the following type. First,

$$\begin{aligned}\mathbb{P}\big(Y_{n+1} - Y_n = (1, 1)\big) &= \mathbb{P}(A_{n+1} - A_n = 1)\\ &= \mathbb{P}\big(X_{n+1} - X_n = (1, 0)\big) = \tfrac{1}{4},\end{aligned}$$

and similarly for the other three possibilities for $Y_{n+1} - Y_n$, namely $(-1, 1)$, $(1, -1)$, and $(-1, -1)$. It follows that $\mathbf{U}$ and $\mathbf{V}$ are symmetric random walks. Furthermore, they are independent since

$$\mathbb{P}\big(Y_{n+1} - Y_n = (u, v)\big) = \mathbb{P}(U_{n+1} - U_n = u)\mathbb{P}(V_{n+1} - V_n = v) \qquad \text{for } u, v = \pm 1.$$

Finally, $X_n = \mathbf{0}$ if and only if $Y_n = \mathbf{0}$, and this occurs if and only if both $U_n = 0$ and $V_n = 0$. Therefore, in agreement with (12.46),

$$p_{\mathbf{0},\mathbf{0}}(2n) = \mathbb{P}_0(U_n = 0)\mathbb{P}_0(V_n = 0) = \left[\left(\frac{1}{2}\right)^{2n}\binom{2n}{n}\right]^2,$$

by (12.44). This argument is invalid in three or more dimensions.

Exercise 12.49 The infinite binary tree T is the tree-graph in which every vertex has exactly three neighbours. Show that a random walk on T is transient.

Exercise 12.50 Consider the asymmetric random walk on the line $\mathbb{Z}$ that moves one step rightwards with probability p, or one step leftwards with probability q $(= 1 - p)$. Show that the walk is recurrent if and only if $p = \frac{1}{2}$.

Exercise 12.51 In a variant of Exercise 12.50, the walker moves two steps rightwards with probability p, and otherwise one step leftwards. Show that the walk is recurrent if and only if $p = \frac{1}{3}$.

12.6 Hitting times and hitting probabilities

Let $A \subseteq S$. The *hitting time* of the subset A is the earliest epoch n of time at which $X_n \in A$:

$$H^A = \inf\{n \geq 0 : X_n \in A\}. \tag{12.52}$$

The infimum of an empty set is taken by convention to be ∞, so that H^A takes values in the extended integers $\{0, 1, 2, \dots\} \cup \{\infty\}$. Note that $H^A = 0$ if $X_0 \in A$.

In this section, we study the *hitting probability*

$$h_i^A = \mathbb{P}_i(H^A < \infty)$$

of ever hitting A starting from i, and also the mean value of H_A. If A is closed, then h_i^A is called an *absorption probability*.

Theorem 12.53 *The vector $h^A = (h_i^A : i \in S)$ is the minimal non-negative solution to the equations*

$$h_i^A = \begin{cases} 1 & \text{for } i \in A, \\ \displaystyle\sum_{j \in S} p_{i,j} h_j^A & \text{for } i \notin A. \end{cases} \tag{12.54}$$

The property of minimality is as follows. For any non-negative solution $(x_i : i \in S)$ of (12.54), we have that $h_i^A \leq x_i$ for all $i \in S$. Since the vector $h^A = (h_i^A)$ multiplies P on its right side, it is best considered as column vector.

Proof We show first that the hitting probabilities satisfy (12.54). Certainly $h_i^A = 1$ for $i \in A$, since $H^A = 0$ in this case. For $i \notin A$, we condition on the first step of the chain to obtain

$$h_i^A = \sum_{j \in S} p_{i,j} \mathbb{P}_i(H^A < \infty \mid X_1 = j) = \sum_{j \in S} p_{i,j} h_j^A$$

as required for (12.54).

We show next that the h_i^A are minimal. Let $x = (x_i : i \in S)$ be a non-negative solution to (12.54). In particular, $h_i^A = x_i = 1$ for $i \in A$. Let $i \notin A$. Since x satisfies (12.54),

$$x_i = \sum_{j \in S} p_{i,j} x_j = \sum_{j \in A} p_{i,j} x_j + \sum_{j \notin A} p_{i,j} x_j. \tag{12.55}$$

Since $x_j = 1$ for $j \in A$, and x is non-negative, we have that

$$\begin{aligned} x_i &\geq \sum_{j \in A} p_{i,j} \\ &= \mathbb{P}_i(X_1 \in A) = \mathbb{P}_i(H^A = 1). \end{aligned}$$

We iterate this as follows. By expanding the final summation in (12.55),

$$\begin{aligned} x_i &= \mathbb{P}_i(X_1 \in A) + \sum_{j \notin A} p_{i,j} \left(\sum_{k \in A} p_{j,k} x_k + \sum_{k \notin A} p_{j,k} x_k \right) \\ &\geq \mathbb{P}_i(X_1 \in A) + \mathbb{P}_i(X_1 \notin A,\ X_2 \in A) \\ &= \mathbb{P}_i(H^A \leq 2). \end{aligned}$$

By repeated substitution, we obtain $x_i \geq \mathbb{P}_i(H^A \leq n)$ for all $n \geq 0$. Take the limit as $n \to \infty$ to deduce as required that $x_i \geq \mathbb{P}_i(H^A < \infty) = h_i^A$. □

We turn now to the mean hitting times, and write

$$k_i^A = \mathbb{E}_i(H^A).$$

Theorem 12.56 *The vector $k^A = (k_i^A : i \in S)$ is the minimal non-negative solution to the equations*

$$k_i^A = \begin{cases} 0 & \text{for } i \in A, \\ 1 + \sum_{j \in S} p_{i,j} k_j^A & \text{for } i \notin A. \end{cases} \tag{12.57}$$

Proof This is very similar to the last proof. We show first that the k_i^A satisfy (12.57). Certainly $k_i^A = 0$ for $i \in A$, since $H^A = 0$ in this case. For $i \notin A$, we condition on the first step of the chain to obtain

$$k_i^A = \sum_{j \in S} p_{i,j} [1 + \mathbb{E}_j(H^A)] = 1 + \sum_{j \in S} p_{i,j} k_j^A$$

as required for (12.57).

We show next that the k_i^A are minimal. Let $y = (y_i : i \in S)$ be a non-negative solution to (12.57). In particular, $k_i^A = y_i = 0$ for $i \in A$. Let $i \notin A$. Since y satisfies (12.57),

$$\begin{aligned} y_i &= 1 + \sum_{j \in S} p_{i,j} y_j = 1 + \sum_{j \notin A} p_{i,j} y_j \\ &= 1 + \sum_{j \notin A} p_{i,j} \left(1 + \sum_{k \notin A} p_{j,k} y_k \right) \\ &\geq \mathbb{P}_i(H^A \geq 1) + \mathbb{P}_i(H^A \geq 2). \end{aligned}$$

By iteration,

$$y_i \geq \sum_{m=1}^{n} \mathbb{P}_i(H^A \geq m),$$

and we send $n \to \infty$ to obtain

$$y_i \geq \sum_{m=1}^{\infty} \mathbb{P}_i(H^A \geq m) = k_i^A,$$

as required. We have used the elementary fact (see Problem 2.6.6) that $\mathbb{E}M = \sum_{m=1}^{\infty} \mathbb{P}(M \geq m)$ for any random variable M taking non-negative integer values. □

Example 12.58 (Gambler's Ruin) Let S be the non-negative integers $\{0, 1, 2, \dots\}$, and $p \in (0, 1)$. A random walk on S moves one unit rightwards with probability p and one unit leftwards with probability q $(= 1 - p)$, and has an absorbing barrier at 0. Find the probability of ultimate absorption.

Solution Let h_i be the probability of absorption starting at i. By Theorem 12.53, (h_i) is the minimal non-negative solution to the equations

$$h_0 = 1, \qquad h_i = p h_{i+1} + q h_{i-1} \quad \text{for } i \geq 1.$$

Suppose $p \neq q$. The difference equation has general solution

$$h_i = A + B(q/p)^i \qquad \text{for } i \geq 0.$$

If $p < q$, the boundedness of the h_i forces $B = 0$, and the fact $h_0 = 1$ implies $A = 1$. Therefore, $h_i = 1$ for all $i \geq 0$.

Suppose $p > q$. Since $h_0 = 1$, we have $A + B = 1$, so that

$$h_i = (q/p)^i + A\big(1 - (q/p)^i\big).$$

Since $h_i \geq 0$, we have $A \geq 0$. By the minimality of the h_i, we have $A = 0$, and hence $h_i = (q/p)^i$, in agreement with Theorem 10.32.

Suppose finally that $p = q = \frac{1}{2}$. The difference equation has solution

$$h_i = A + Bi,$$

and the above arguments yield $B = 0$, $A = 1$, so that $h_i = 1$. △

Example 12.59 (Birth–death chain) Let $(p_i : i \geq 1)$ be a sequence of numbers satisfying $p_i = 1 - q_i \in (0, 1)$. The Gambler's Ruin example, above, may be extended as follows. Let $\mathbf{X}$ be a Markov chain on $\{0, 1, 2, \dots\}$ with transition probabilities

$$p_{i,i+1} = p_i, \quad p_{i,i-1} = q_i \qquad \text{for } i \geq 1,$$

and $p_{0,0} = 1$. What is the probability of ultimate absorption at 0, having started at i?

Solution As in Example 12.58, the required probabilities h_i are the minimal non-negative solutions of

$$h_0 = 1, \qquad h_i = p_i h_{i+1} + q_i h_{i-1} \quad \text{for } i \geq 1.$$

Set $u_i = h_{i-1} - h_i$ and reorganize this equation to find that

$$u_{i+1} = (q_i/p_i)u_i,$$

so that $u_{i+1} = \gamma_i u_1$ where

$$\gamma_i = \frac{q_1 q_2 \cdots q_i}{p_1 p_2 \cdots p_i}.$$

Now, $u_0 + u_1 + \cdots + u_i = h_0 - h_i$, so that

$$h_i = 1 - u_1(\gamma_0 + \gamma_1 + \cdots + \gamma_{i-1}),$$

where $\gamma_0 = 1$. It remains to determine the constant u_1.

There are two situations. Suppose first that $\sum_k \gamma_k = \infty$. Since $h_i \geq 0$ for all i, we have that $u_1 = 0$, and therefore $h_i = 1$. On the other hand, if $\sum_k \gamma_k < \infty$, the h_i are minimized when $1 - u_1 \sum_k \gamma_k = 0$, which is to say that $u_1 = \left(\sum_k \gamma_k\right)^{-1}$ and

$$h_i = \sum_{k=i}^{\infty} \gamma_k \Bigg/ \sum_{k=0}^{\infty} \gamma_k \qquad \text{for } i \geq 0.$$

Thus $h_i < 1$ for $i \geq 1$ if and only if $\sum_k \gamma_k < \infty$. △

Exercise 12.60 Let $\mathbf{X}$ be a Markov chain on the non-negative integers $\{0, 1, 2, \dots\}$ with transition probabilities satisfying

$$p_{0,1} = 1, \quad p_{i,i+1} + p_{i,i-1} = 1, \quad p_{i,i+1} = p_{i,i-1}\left(\frac{i+1}{i}\right)^2 \quad \text{for } i \geq 1.$$

Show that $\mathbb{P}_0(X_n \geq 1 \text{ for all } n \geq 1) = 6/\pi^2$. You may use the fact that $\sum_{k=1}^{\infty} k^{-2} = \frac{1}{6}\pi^2$.

Exercise 12.61 Consider Exercise 12.60 with the difference that

$$p_{i,i+1} = p_{i,i-1}\left(\frac{i+1}{i}\right)^{\alpha} \quad \text{for } i \geq 1,$$

where $\alpha > 0$. Find the probability $\mathbb{P}_0(X_n \geq 1 \text{ for all } n \geq 1)$ in terms of α.

12.7 Stopping times and the strong Markov property

The Markov property of Definition 12.2 requires that, conditional on the value of the chain at a given time n, the future evolution of the chain is independent of its past. We frequently require an extension of this property to a *random* time n. It is not hard to see that the Markov property cannot be true for *all* random times, and it turns out that the appropriate times are those satisfying the following definition.

Definition 12.62 *A random variable* $T : \Omega \to \{0, 1, 2, \dots\} \cup \{\infty\}$ *is called a* ***stopping time*** *for the chain* $\mathbf{X}$ *if, for all* $n \geq 0$*, the event* $\{T = n\}$ *is given in terms of* $X_0, X_1, \dots, X_n$ *only.*

That is to say, a random time T is a stopping time if you can tell whether it takes any given value by examining only the present and past of the chain. Times that 'look into the future' are not stopping times.

The principal examples of stopping times are the so-called hitting times. Let $A \subseteq S$, and consider the hitting time H^A of A given in (12.52). Note that

$$\{H^A = n\} = \{X_n \in A\} \cap \left(\bigcap_{0 \leq m < n} \{X_m \notin A\} \right),$$

so that H^A is indeed a stopping time: one can tell whether or not $H^A = n$ by examining $X_0, X_1, \dots, X_n$ only.

Two related examples: it is easily checked that $T = H^A + 1$ is a stopping time, and that $T = H^A - 1$ is not. See Exercise 12.69 for a further example.

Theorem 12.63 (Strong Markov property) *Let* $\mathbf{X}$ *be a Markov chain with transition matrix* P*, and let* T *be a stopping time. Given that* $T < \infty$ *and* $X_T = i$*, the sequence* $\mathbf{Y} = (Y_k : k \geq 0)$*, given by* $Y_k = X_{T+k}$*, is a Markov chain with transition matrix* P *and initial state* $Y_0 = i$*. Furthermore, given that* $T < \infty$ *and* $X_T = i$*,* $\mathbf{Y}$ *is independent of* $X_0, X_1, \dots, X_{T-1}$*.*

Proof Let H be an event given in terms of $X_0, X_1, \dots, X_{T-1}$. It is required to show that

$$\begin{aligned} &\mathbb{P}\big(X_{T+1} = i_1, X_{T+2} = i_2, \dots, X_{T+n} = i_n, H \,\big|\, T < \infty, X_T = i\big) \\ &\qquad = \mathbb{P}_i(X_1 = i_1, X_2 = i_2, \dots, X_n = i_n)\mathbb{P}(H \mid T < \infty, X_T = i). \end{aligned} \tag{12.64}$$

The event $H \cap \{T = m\}$ is given in terms of $X_1, X_2, \dots, X_m$ only. Furthermore, $X_T = X_m$ when $T = m$. We condition on the event $H \cap \{T = m\} \cap \{X_m = i\}$ and use the Markov property (12.8) at time m to deduce that

$$\begin{aligned} &\mathbb{P}(X_{T+1} = i_1, X_{T+2} = i_2, \dots, X_{T+n} = i_n, H, T = m, X_T = i) \\ &\qquad = \mathbb{P}_i(X_1 = i_1, X_2 = i_2, \dots, X_n = i_n)\mathbb{P}(H, T = m, X_T = i). \end{aligned}$$

Now sum over $m = 0, 1, 2, \dots$ and divide by $\mathbb{P}(T < \infty, X_T = i)$ to obtain (12.64). □

Example 12.65 Let S be the non-negative integers $\{0, 1, 2, \dots\}$, and $p \in (0, 1)$. Consider a random walk $\mathbf{X}$ on S which moves one step rightwards with probability p, one step leftwards with probability q $(= 1 - p)$, and with an absorbing barrier at 0. Let H be the time until absorption at 0. Find the distribution (and mean) of H given $X_1 = 1$.

Solution We shall work with the probability generating function of H. A problem arises since it may be the case that $\mathbb{P}_1(H = \infty) > 0$. One may either work with the conditional generating function $\mathbb{E}_1(s^H \mid H < \infty)$, or, equivalently, we can use the fact that, when $|s| < 1$, $s^n \to 0$ as $n \to \infty$. That is, we write

$$G(s) = \mathbb{E}_1(s^H) = \sum_{n=0}^{\infty} s^n \mathbb{P}_1(H = n) \qquad \text{for } |s| < 1,$$

valid regardless of whether or not $\mathbb{P}_1(H = \infty) = 0$. Henceforth, we assume that $|s| < 1$, and later we shall use Abel's lemma[3] to take the limit as $s \uparrow 1$.

By conditioning on the first step of the walk, we find that

$$G(s) = p\mathbb{E}_1(s^H \mid X_1 = 2) + q\mathbb{E}_1(s^H \mid X_1 = 0).$$

By the strong Markov property,

$$\mathbb{E}_1(s^H \mid X_1 = 2) = \mathbb{E}(s^{1+H'+H''}) = s\mathbb{E}(s^{H'+H''}),$$

where H' and H'' are independent copies of H. Therefore,

$$G(s) = psG(s)^2 + qs. \tag{12.66}$$

This is a quadratic in $G(s)$ with solutions

$$G(s) = \frac{1 \pm \sqrt{1 - 4pqs^2}}{2ps}. \tag{12.67}$$

Since G is continuous wherever it is finite, we must choose one of these solutions and stick with it for all $|s| < 1$. Since $G(0) \le 1$ and the positive root diverges as $s \downarrow 0$, we take the negative root in (12.67) for all $|s| < 1$.

The mass function of H is obtained from the coefficients in the expansion of $G(s)$ as a power series:

$$\mathbb{P}_1(H = 2k - 1) = \binom{1/2}{k}(-1)^{k-1}\frac{(4pq)^k}{2p} = \frac{(2k-2)!}{k!\,(k-1)!} \cdot \frac{(pq)^k}{p},$$

for $k = 1, 2, \dots$. This uses the extended binomial theorem, Theorem A.3.

[3] See the footnote on page 54 for a statement of Abel's lemma.

It is not certain that $H < \infty$. Since $\mathbb{P}_1(H < \infty) = \lim_{s \uparrow 1} G(s)$, we have by (12.67) and Abel's lemma that

$$\mathbb{P}_1(H < \infty) = \frac{1 - \sqrt{1 - 4pq}}{2p}.$$

It is convenient to write

$$1 - 4pq = 1 - 4p + 4p^2 = (1 - 2p)^2 = |p - q|^2,$$

so that

$$\mathbb{P}_1(H < \infty) = \frac{1 - |p - q|}{2p} = \begin{cases} 1 & \text{if } p \le q, \\ q/p & \text{if } q > p, \end{cases}$$

as in Theorem 10.32 and Example 12.58.

We turn to the mean value $\mathbb{E}_1(H)$. When $q > p$, $\mathbb{P}_1(H = \infty) > 0$, and so $\mathbb{E}_1(H) = \infty$. Suppose $p \le q$. By differentiating (12.66),

$$pG^2 + 2psGG' - G' + q = 0 \qquad \text{for } |s| < 1, \tag{12.68}$$

which we solve for G' to find that

$$G' = \frac{pG^2 + q}{1 - 2psG} \qquad \text{for } |s| < 1.$$

By Abel's lemma, $\mathbb{E}_1(H) = \lim_{s \uparrow 1} G'(s)$, so that

$$\mathbb{E}_1(H) = \lim_{s \uparrow 1} \left(\frac{pG^2 + q}{1 - 2psG} \right) = \begin{cases} \infty & \text{if } p = q, \\ \dfrac{1}{q - p} & \text{if } p < q. \end{cases} \qquad \triangle$$

Exercise 12.69

(a) Let H^A be the hitting time of the set A. Show that $T = H^A - 1$ is not generally a stopping time.

(b) Let L^A be the time of the last visit of a Markov chain to the set A, with the convention that $L^A = \infty$ if infinitely many visits are made. Show that L^A is not generally a stopping time.

12.8 Classification of states

We saw in Definition 12.29 that a state i is *recurrent* if, starting from i, the chain returns to i with probability 1. The state is *transient* if it is not recurrent. If the starting state i is recurrent, the chain is bound to return to it. Indeed, it is bound to return infinitely often.

Theorem 12.70 *Suppose $X_0 = i$, and let $V_i = |\{n \ge 1 : X_n = i\}|$ be the number of subsequent visits by the Markov chain to i. Then*

(a) $\mathbb{P}_i(V_i = \infty) = 1$ *if i is recurrent,*
(b) $\mathbb{P}_i(V_i < \infty) = 1$ *if i is transient.*

We return in Theorem 12.101 to the more detailed question of the rate of divergence of the number of visits to i in the recurrent case. The proof makes use of the recurrence time of a state i. Let

$$T_i = \inf\{n \geq 1 : X_n = i\} \tag{12.71}$$

be the first passage time to i. If $X_0 = i$, then T_i is the *recurrence time* of i, with mean $\mu_i = \mathbb{E}_i(T_i)$.

Proof Recall the first-passage probability $f_{i,i} = \mathbb{P}_i(T_i < \infty)$, so that i is recurrent if $f_{i,i} = 1$ and transient of $f_{i,i} < 1$. Let T_i^r be the epoch of the rth visit to i, with $T_i^r = \infty$ if $V_i < r$. Since the T_i^r are increasing,

$$\begin{aligned}
\mathbb{P}_i(V_i \geq r) &= \mathbb{P}_i(T_i^r < \infty) \\
&= \mathbb{P}_i(T_i^r < \infty \mid T_i^{r-1} < \infty)\mathbb{P}_i(T_i^{r-1} < \infty) \\
&= f_{i,i}\mathbb{P}_i(T_i^{r-1} < \infty) \qquad \text{for } r \geq 1,
\end{aligned}$$

by the strong Markov property, Theorem 12.63. By iteration, $\mathbb{P}_i(V_i \geq r) = f_{i,i}^r$. We send $r \to \infty$ to find that

$$\mathbb{P}_i(V_i = \infty) = \begin{cases} 1 & \text{if } f_{i,i} = 1, \\ 0 & \text{if } f_{i,i} < 1. \end{cases}$$

and the theorem is proved. When i is transient, we have proved the stronger fact that V_i has a geometric distribution. □

Definition 12.72

(a) *The **mean recurrence time** μ_i of the state i is defined by*

$$\mu_i = \mathbb{E}_i(T_i) = \begin{cases} \displaystyle\sum_{n=1}^{\infty} n f_{i,i}(n) & \text{if } i \text{ is recurrent}, \\ \infty & \text{if } i \text{ is transient}. \end{cases}$$

(b) *If i is recurrent, we call it **null** if $\mu_i = \infty$, and **positive** (or **non-null**) if $\mu_i < \infty$.*

(c) *The **period** d_i of the state i is given by*

$$d_i = \gcd\{n : p_{i,i}(n) > 0\}.$$

*The state i is called **aperiodic** if $d_i = 1$, and **periodic** if $d_i > 1$.*

(d) *State i is called **ergodic** if it is aperiodic and positive recurrent.*

It was proved in Theorem 12.36 that recurrence is a class property. This conclusion may be extended as follows.

Theorem 12.73 *If $i \leftrightarrow j$ then*

(a) *i and j have the same period,*

(b) *i is recurrent if and only if j is recurrent,*
(c) *i is positive recurrent if and only if j is positive recurrent,*
(d) *i is ergodic if and only if j is ergodic.*

We may therefore speak of an communicating class C as being recurrent, transient, ergodic, and so on. An irreducible chain has a single communicating class, and thus we may attribute these adjectives (when appropriate) to the chain itself.

Proof We may assume $i \neq j$.

(a) Since $i \leftrightarrow j$, there exist $m, n \geq 1$ such that

$$\alpha := p_{i,j}(m)p_{j,i}(n) > 0.$$

By the Chapman–Kolmogorov equations, Theorem 12.12,

$$p_{i,i}(m+r+n) \geq p_{i,j}(m)p_{j,j}(r)p_{j,i}(n) = \alpha p_{j,j}(r) \qquad \text{for } r \geq 0. \tag{12.74}$$

In particular, $p_{i,i}(m+n) \geq \alpha > 0$, so that $d_i \mid m+n$. Therefore, if $d_i \nmid r$, then $d_i \nmid m+r+n$, so that $p_{i,i}(m+n+r) = 0$. In this case, by (12.74), $p_{j,j}(r) = 0$, and hence $d_j \nmid r$. Therefore, $d_i \mid d_j$. By the reverse argument, $d_j \mid d_i$, and hence $d_i = d_j$.

(b) This was proved at Theorem 12.36.

(c) For this proof we look forward slightly to Theorem 12.81. Suppose that i is positive recurrent, and let C be the communicating class of states containing i. Since i is recurrent, by Theorem 12.36(b), C is closed. If $X_0 \in C$, then $X_n \in C$ for all n, and the chain is irreducible on the state space C. By part (a) of Theorem 12.81, it possesses an invariant distribution, and by part (b) every state (of C) is positive recurrent. If $i \leftrightarrow j$ then $j \in C$, so j is positive recurrent.

(d) This follows from (a), (b), and (c). □

Finally in this section, we note that recurrent states will be visited regardless of the initial distribution. This will be useful later.

Proposition 12.75 *If the chain is irreducible and $j \in S$ is recurrent, then*

$$\mathbb{P}(X_n = j \textit{ for some } n \geq 1) = 1,$$

regardless of the distribution of X_0.

Proof Let i, j be distinct states. Since the chain is assumed irreducible, there exists a least (finite) integer m such that $p_{j,i}(m) > 0$. Since m is least, it is the case that

$$p_{j,i}(m) = \mathbb{P}_j(X_m = i,\ X_r \neq j \text{ for } 1 \leq r < m). \tag{12.76}$$

Suppose $X_0 = j$, $X_m = i$, and no return to j takes place after time m. By (12.76), with probability 1 no return to j ever takes place. By the Markov property at time m,

$$p_{j,i}(m)(1 - f_{i,j}) \leq 1 - f_{j,j}.$$

If j is recurrent, then $f_{j,j} = 1$, so that $f_{i,j} = 1$ for all $i \in S$.

Let $\lambda_i = \mathbb{P}(X_0 = i)$. With $T_j = \inf\{n \geq 1 : X_n = j\}$ as usual,

$$\mathbb{P}(T_j < \infty) = \sum_{i \in S} \lambda_i f_{i,j} = 1,$$

by conditioning on X_0. □

Exercise 12.77 Let **X** be an irreducible Markov chain with period d. Show that $Y_n = X_{nd}$ defines an aperiodic Markov chain.

Exercise 12.78 Let $0 < p < 1$. Classify the states of the Markov chains with transition matrices

$$\begin{pmatrix} 0 & p & 0 & 1-p \\ 1-p & 0 & p & 0 \\ 0 & 1-p & 0 & p \\ p & 0 & 1-p & 0 \end{pmatrix}, \qquad \begin{pmatrix} 1-2p & 2p & 0 \\ p & 1-2p & p \\ 0 & 2p & 1-2p \end{pmatrix}.$$

Exercise 12.79 Let i be an aperiodic state of a Markov chain. Show that there exists $N \geq 1$ such that $p_{i,i}(n) > 0$ for all $n \geq N$.

12.9 Invariant distributions

We turn now towards the study of the long-term behaviour of a Markov chain: what can be said about X_n in the limit as $n \to \infty$? Since the sequence $(X_n : n \geq 0)$ is subject to random fluctuations, it does not (typically) converge to any given state. On the other hand, we will see in the next section that its distribution settles into an equilibrium. In advance of stating this limit theorem, we first explore the possible limits. Any distributional limit is necessarily invariant under the evolution of the chain, and we led to the following definition.

Definition 12.80 *Let* **X** *be a Markov chain with transition matrix P. The vector $\pi = (\pi_i : i \in S)$ is called an* ***invariant distribution***[4] *of the chain if:*

(a) $\pi_i \geq 0$ *for all $i \in S$, and* $\sum_{i \in S} \pi_i = 1$,
(b) $\pi = \pi P$.

An invariant distribution is invariant under the passage of time: if X_0 has distribution π, then X_n has distribution πP^n, and $\pi P^n = \pi P \cdot P^{n-1} = \pi P^{n-1} = \cdots = \pi$.

Theorem 12.81 *Consider an irreducible Markov chain.*

(a) *There exists an invariant distribution π if and only if some state is positive recurrent.*
(b) *If there exists an invariant distribution π, then every state is positive recurrent, and $\pi_i = 1/\mu_i$ where μ_i is the mean recurrence time of state i. In particular, π is the unique invariant distribution.*

[4] Also known as a *stationary* or *equilibrium* or *steady-state* distribution. An invariant distribution is sometimes referred to as an *invariant measure*, but it is more normal to reserve this expression for a non-negative solution π of the equation $\pi = \pi P$ with no assumption of having sum 1, or indeed of even having finite sum.

We shall prove Theorem 12.81 by exhibiting an explicit solution of the vector equation $\rho = \rho P$. In looking for a solution, it is natural to consider a vector ρ with entries indicative of the proportions of time spent in the various states. Towards this end, we fix a state $k \in S$ and start the chain from this state. Let W_i be the number of subsequent visits to state i before the first return to the initial state k. Thus W_i may be expressed in either of the forms

$$W_i = \sum_{m=1}^{\infty} 1(X_m = i,\ T_k \geq m) = \sum_{m=1}^{T_k} 1(X_m = i), \qquad i \in S, \tag{12.82}$$

where $T_k = \inf\{n \geq 1 : X_n = k\}$ is the first return time to the starting state k, and $1(A) = 1_A$ is the indicator function of A. Note that $W_k = 1$ if $T_k < \infty$. Our candidate for the vector ρ is given by

$$\rho_i = \mathbb{E}_k(W_i), \qquad i \in S. \tag{12.83}$$

Recall that $\mathbb{E}_k(Z)$ denotes the mean of Z given that $X_0 = k$.

Proposition 12.84 *For an irreducible, recurrent chain, and any given $k \in S$, the vector $\rho = (\rho_i : i \in S)$ satisfies:*

(a) $\rho_k = 1$,
(b) $\sum_{i \in S} \rho_i = \mu_k$, *whether or not* $\mu_k < \infty$,
(c) $\rho = \rho P$,
(d) $0 < \rho_i < \infty$ *for* $i \in S$.

One useful consequence of the above is the following. Consider an irreducible, positive recurrent Markov chain, and fix a state k. By Theorem 12.81, there exists a unique invariant distribution π, and $\pi_k = 1/\mu_k$. By Proposition 12.84(b), $\nu := \pi_k \rho$ satisfies $\nu = \nu P$ and $\sum_{i \in S} \nu_i = 1$. It follows that $\pi = \nu$. Therefore, $\rho_i = \nu_i/\pi_k = \pi_i/\pi_k$. In summary, for given $k \in S$, the mean number of visits to state i between two consecutive visits to k is π_i/π_k.

Proof of Proposition 12.84 (a) Since the chain is assumed recurrent, $\mathbb{P}_k(T_k < \infty) = 1$. By (12.82), $W_k = 1$, so that $\rho_k = \mathbb{E}_k(1) = 1$.

(b) Since the time between two visits to state k must be spent somewhere, we have that

$$T_k = \sum_{i \in S} W_i,$$

so that, by an interchange of expectation and summation,[5]

$$\mu_k = \mathbb{E}_k(T_k) = \sum_{i \in S} \mathbb{E}_k(W_i) = \sum_{i \in S} \rho_i.$$

(c) By (12.82) and a further interchange, for $j \in S$,

[5] Care is necessary when interchanging limits. This interchange is justified by the footnote on page 39. The forthcoming interchange at (12.87) holds since the order of summation is irrelevant to the value of a double sum of non-negative reals.

$$\rho_j = \sum_{m=1}^{\infty} \mathbb{P}_k(X_m = j,\ T_k \geq m). \tag{12.85}$$

The event $\{T_k \geq m\}$ depends only on $X_0, X_1, \dots, X_{m-1}$. By the extended Markov property, Theorem 12.7, for $m \geq 1$,

$$\begin{aligned}\mathbb{P}_k(X_m = j,\ T_k \geq m) &= \sum_{i \in S} \mathbb{P}_k(X_{m-1} = i,\ X_m = j,\ T_k \geq m) \\ &= \sum_{i \in S} \mathbb{P}_k\big(X_m = j \,\big|\, X_{m-1} = i,\ T_k \geq m\big)\mathbb{P}_k(X_{m-1} = i,\ T_k \geq m) \\ &= \sum_{i \in S} p_{i,j}\mathbb{P}_k(X_{m-1} = i,\ T_k \geq m). \end{aligned} \tag{12.86}$$

By (12.85)–(12.86) and another interchange of limits,

$$\rho_j = \sum_{i \in S} \sum_{m=1}^{\infty} p_{i,j}\mathbb{P}_k(X_{m-1} = i,\ T_k \geq m). \tag{12.87}$$

We rewrite this with $r = m - 1$ to find that

$$\rho_j = \sum_{i \in S} p_{i,j} \sum_{r=0}^{\infty} \mathbb{P}_k(X_r = i,\ T_k \geq r + 1) = \sum_{i \in S} p_{i,j}\rho_i,$$

where the last equality holds by separate consideration of the two cases $i = k$ and $i \neq k$. In summary, $\rho = \rho P$.

(d) Since the chain is irreducible, there exist $m, n \geq 0$ such that $p_{i,k}(m), p_{k,i}(n) > 0$. Since $\rho = \rho P$ and hence $\rho = \rho P^k$ for $k \geq 1$, we have that

$$\rho_k \geq \rho_i p_{i,k}(m), \qquad \rho_i \geq \rho_k p_{k,i}(n).$$

Since $\rho_k = 1$,

$$p_{k,i}(m) \leq \rho_i \leq \frac{1}{p_{i,k}(n)},$$

and the proof is complete. □

Proof of Theorem 12.81 (a) Suppose $k \in S$ is positive recurrent, so that $\mu_k < \infty$. Let ρ be given by (12.83). By Proposition 12.84, $\pi := (1/\mu_k)\rho$ is an invariant distribution.

(b) Suppose that π is an invariant distribution of the chain. We show first that

$$\pi_i > 0 \qquad \text{for } i \in S. \tag{12.88}$$

Since $\sum_{i \in S} \pi_i = 1$, there exists $k \in S$ with $\pi_k > 0$. Let $i \in S$. By irreducibility, there exists $m \geq 1$ such that $p_{k,i}(m) > 0$. We have that $\pi = \pi P$, so that $\pi = \pi P^m$. Therefore,

$$\pi_i = \sum_{j \in S} \pi_j p_{j,i}(m) \geq \pi_k p_{k,i}(m) > 0,$$

and (12.88) follows.

By irreducibility and Theorem 12.36, either all states are transient or all are recurrent. If all states are transient then $p_{i,j}(n) \to 0$ as $n \to \infty$ by Proposition 12.39. Since $\pi = \pi P^n$,

$$\pi_j = \sum_i \pi_i p_{i,j}(n) \to 0 \quad \text{as} \quad n \to \infty, \qquad \text{for } i, j \in S, \tag{12.89}$$

which contradicts (12.88). Therefore, all states are recurrent. A small argument is needed to justify the limit in (12.89), and this is deferred to Lemma 12.91.

We show next that the existence of π implies that all states are positive, and that $\pi_i = \mu_i^{-1}$ for $i \in S$. Suppose that X_0 has distribution π. Then

$$\pi_j \mu_j = \mathbb{P}(X_0 = j) \sum_{n=1}^{\infty} \mathbb{P}_j(T_j \geq n) = \sum_{n=1}^{\infty} \mathbb{P}(T_j \geq n,\ X_0 = j).$$

However, $\mathbb{P}(T_j \geq 1,\ X_0 = j) = \mathbb{P}(X_0 = j)$, and for $n \geq 2$,

$$\begin{aligned}
\mathbb{P}(T_j \geq n,\ X_0 = j) &= \mathbb{P}(X_0 = j,\ X_m \neq j \text{ for } 1 \leq m \leq n-1) \\
&= \mathbb{P}(X_m \neq j \text{ for } 1 \leq m \leq n-1) - \mathbb{P}(X_m \neq j \text{ for } 0 \leq m \leq n-1) \\
&= \mathbb{P}(X_m \neq j \text{ for } 0 \leq m \leq n-2) - \mathbb{P}(X_m \neq j \text{ for } 0 \leq m \leq n-1) \\
&= a_{n-2} - a_{n-1}
\end{aligned}$$

where

$$a_r = \mathbb{P}(X_m \neq j \text{ for } 0 \leq m \leq r),$$

and we have used the invariance of π. We sum over n to obtain

$$\pi_j \mu_j = \mathbb{P}(X_0 = j) + a_0 - \lim_{n \to \infty} a_n = 1 - \lim_{n \to \infty} a_n.$$

However, $a_n \to \mathbb{P}(X_m \neq j \text{ for all } m) = 0$ as $n \to \infty$, by the recurrence of j and Proposition 12.75.

We have shown that

$$\pi_j \mu_j = 1, \tag{12.90}$$

so that $\mu_j = \pi_j^{-1} < \infty$ by (12.88). Hence $\mu_j < \infty$ and all states of the chain are positive. Furthermore, (12.90) specifies π_j uniquely as μ_j^{-1}. □

Here is the little lemma used to establish the limit in (12.89). It is a form of the so-called bounded convergence theorem.

Lemma 12.91 *Let $\lambda = (\lambda_i : i \in S)$ be a distribution on the countable set S, and let $\alpha_i(n)$ satisfy*

$$|\alpha_i(n)| \leq M, \qquad \lim_{n \to \infty} \alpha_i(n) = 0 \qquad \text{for } i \in S,$$

where $M < \infty$. Then

$$\sum_{i \in S} \lambda_i \alpha_i(n) \to 0 \qquad \text{as } n \to \infty.$$

Proof Let F be a finite subset of S, and write

$$\begin{aligned}\sum_{i\in S}|\lambda_i\alpha_i(n)| &\le \sum_{i\in F}\lambda_i|\alpha_i(n)| + M\sum_{i\notin F}\lambda_i\\ &\to M\sum_{i\notin F}\lambda_i \qquad \text{as } n\to\infty, \text{ since } F \text{ is finite}\\ &\to 0 \qquad \text{as } F\uparrow S, \text{ since } \sum_i \lambda_i < \infty.\end{aligned}$$

□

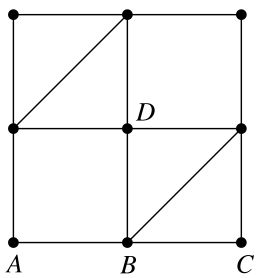

Fig. 12.4 Find the mean number of visits to B before returning to the starting state A.

Exercise 12.92 A particle starts at A and executes a symmetric random walk on the graph of Figure 12.4. Find the invariant distribution of the chain. Using the remark after Proposition 12.84 or otherwise, find the expected number of visits to B before the particle returns to A.

Exercise 12.93 Consider the symmetric random walk on the line $\mathbb{Z}$. Show that any invariant distribution π satisfies $\pi_n = \frac{1}{2}(\pi_{n-1} + \pi_{n+1})$, and deduce that the walk is null recurrent.

12.10 Convergence to equilibrium

The principal result for discrete-time Markov chains is that, subject to weak conditions, its distribution converges to the unique invariant distribution.

Theorem 12.94 (Ergodic theorem for Markov chains) *Consider an aperiodic, irreducible, positive recurrent Markov chain. For $i, j \in S$,*

$$p_{i,j}(n) \to \pi_j \qquad \textit{as } n\to\infty,$$

where π is the unique invariant distribution of the chain.

Proof The proof uses an important technique known as 'coupling'. Construct an ordered pair $\mathbf{Z} = (\mathbf{X}, \mathbf{Y})$ of *independent* Markov chains $\mathbf{X} = (X_n : n \ge 0)$, $\mathbf{Y} = (Y_n : n \ge 0)$ each of which has state space S and transition matrix P. Then $\mathbf{Z} = (Z_n : n \ge 0)$ is given by

$Z_n = (X_n, Y_n)$, and it is easy to check that $\mathbf{Z}$ is a Markov chain with state space $S \times S$ and transition probabilities

$$\begin{aligned} p_{ij,kl} &= \mathbb{P}\big(Z_{n+1} = (k,l) \,\big|\, Z_n = (i,j)\big) \\ &= \mathbb{P}(X_{n+1} = k \mid X_n = i)\mathbb{P}(Y_{n+1} = l \mid Y_n = j) \quad \text{by independence} \\ &= p_{i,k}p_{j,l}. \end{aligned}$$

Since $\mathbf{X}$ is irreducible and aperiodic, for $i, j, k, l \in S$ there exists $N = N(i,j,k,l)$ such that $p_{i,k}(n)p_{j,l}(n) > 0$ for all $n \geq N$ (see Exercise 12.79). Therefore, $\mathbf{Z}$ is irreducible. Only here is the aperiodicity used.

Suppose that $\mathbf{X}$ is positive recurrent. By Theorem 12.81, $\mathbf{X}$ has a unique stationary distribution π, and it is follows that $\mathbf{Z}$ has the stationary distribution $\nu = (\nu_{i,j} : i, j \in S)$ given by $\nu_{i,j} = \pi_i \pi_j$. Therefore, $\mathbf{Z}$ is also positive recurrent, by Theorem 12.81. Let $X_0 = i$ and $Y_0 = j$, so that $Z_0 = (i, j)$. Fix $s \in S$ and let

$$T = \min\{n \geq 1 : Z_n = (s,s)\}$$

be the first passage time of $\mathbf{Z}$ to (s,s). By the recurrence of Z and Proposition 12.75,

$$\mathbb{P}_{ij}(T < \infty) = 1, \tag{12.95}$$

where $\mathbb{P}_{ij}$ denotes the probablity measure conditional on $Z_0 = (i, j)$.

The central idea of the proof is the following observation. Suppose $X_m = Y_m = s$. Since T is a stopping time, by the strong Markov property X_n and Y_n have the same conditional distributions given the event $\{T \leq n\}$. We shall use this fact, together with the finiteness of T, to show that the limiting distributions of X and Y are independent of their starting points.

More precisely,

$$\begin{aligned} p_{i,k}(n) &= \mathbb{P}_{ij}(X_n = k) \\ &= \mathbb{P}_{ij}(X_n = k,\ T \leq n) + \mathbb{P}_{ij}(X_n = k,\ T > n) \\ &= \mathbb{P}_{ij}(Y_n = k,\ T \leq n) + \mathbb{P}_{ij}(X_n = k,\ T > n) \\ &\qquad \text{since, given that } T \leq n,\ X_n \text{ and } Y_n \text{ are identically distributed} \\ &\leq \mathbb{P}_{ij}(Y_n = k) + \mathbb{P}_{ij}(T > n) \\ &= p_{j,k}(n) + \mathbb{P}_{ij}(T > n). \end{aligned}$$

This, and the related inequality with i and j interchanged, yields

$$\big|p_{i,k}(n) - p_{j,k}(n)\big| \leq \mathbb{P}_{ij}(T > n) \to 0 \qquad \text{as } n \to \infty$$

by (12.95). Therefore,

$$p_{i,k}(n) - p_{j,k}(n) \to 0 \quad \text{as } n \to \infty, \qquad \text{for } i, j, k \in S. \tag{12.96}$$

Thus, if the limit $\lim_{n\to\infty} p_{ik}(n)$ exists, then it does not depend on i. To show that it exists, write

$$\pi_k - p_{j,k}(n) = \sum_{i \in S} \pi_i \big[p_{i,k}(n) - p_{j,k}(n)\big] \to 0 \qquad \text{as } n \to \infty, \tag{12.97}$$

by Lemma 12.91. The proof is complete. □

Example 12.98 Here is an elementary example which highlights the necessity of aperiodicity in the ergodic theorem, Theorem 12.94. Let $\mathbf{X}$ be a Markov chain with state space $S = \{1, 2\}$ and transition matrix

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

Thus, $\mathbf{X}$ alternates deterministically between the two states. It is immediate that $P^{2n} = I$ and $P^{2n+1} = P$ for $n \geq 0$, and, in particular, the limit $\lim_{k\to\infty} p_{i,j}(k)$ exists for no $i, j \in S$.

The proof of Theorem 12.94 fails since the paired chain $\mathbf{Z}$ is not irreducible: for example, if $Z_0 = (0, 1)$ then $Z_n \neq (0, 0)$ for all n. △

Example 12.99 (Coupling game) A pack of cards is shuffled, and the cards dealt (face up) one by one. A friend is asked to select some card, secretly, from amongst the first six or seven cards, say. If the face value of this card is m (aces count 1 and court cards count 10), the next $m-1$ cards are allowed to pass, and your friend is asked to note the face value of the mth card. Continuing according to this rule, there arrives a last card in this sequence, with face value X say, and with fewer than X cards remaining. We call X your friend's 'score'.

With high probability, you are able to guess accurately your friend's score, as follows. You follow the same rules as the friend, starting for simplicity at the first card. You obtain thereby a score Y, say. There is a high probability that $X = Y$.

Why is this the case? Suppose your friend picks the m_1th card, m_2th card, and so on, and you pick the n_1 $(= 1)$th, n_2th, ... If $m_i = n_j$ for some i, j, the two of you are 'stuck together' forever after. When this occurs first, we say that 'coupling' has occurred. Prior to coupling, each time you read the value of a card, there is a positive probability that you will arrive at the next stage on exactly the same card as the other person. If the pack of cards were infinitely large, then coupling would take place sooner or later. It turns out that there is a reasonable chance that coupling takes place before the last card of a regular pack has been dealt. △

A criterion for transience or recurrence was presented at Theorem 12.30. We now have a criterion for null recurrence.

Theorem 12.100 *Let* $\mathbf{X}$ *be an irreducible, recurrent Markov chain. The following are equivalent.*

(a) *There exists a state i such that $p_{i,i}(n) \to 0$ as $n \to \infty$.*
(b) *Every state is null recurrent.*

As an application, consider symmetric random walk on the graphs $\mathbb{Z}$ or $\mathbb{Z}^2$ of Section 12.5. By (12.45) or (12.47) as appropriate, $p_{\mathbf{0},\mathbf{0}}(n) \to 0$ as $n \to \infty$, from which we deduce that the one- and two-dimensional random walks are null recurrent. This may be compared with the method of Exercise 12.93.

Proof We shall prove only that (a) implies (b). See Grimmett and Stirzaker (2001, Thm (6.2.9)) for the other part. If the chain $\mathbf{X}$ is positive recurrent and in addition aperiodic, then

$$p_{i,i}(n) \to \frac{1}{\mu_i} > 0,$$

by Theorems 12.81 and 12.94. Therefore, (a) does not hold. The same argument may be applied in the periodic case by considering the chain $Y_n = X_{nd}$ where d is the period of the chain. Thus (a) implies (b). □

This section closes with a discussion of the rate at which a Markov chain visits a given state. Let $i \in S$ and let

$$V_i(n) = \sum_{k=1}^{n} 1(X_k = i)$$

denote the number of visits to i up to time n.

Theorem 12.101 *Let $i \in S$. If the chain is irreducible and positive recurrent,*

$$\frac{1}{n} V_i(n) \Rightarrow \frac{1}{\mu_i} \qquad \text{as } n \to \infty,$$

irrespective of the initial distribution of the chain.

There are various modes of convergence of random variables, of which we have chosen convergence in distribution for the sake of simplicity. (It is equivalent to convergence in probability in this case, see Theorem 8.47.) A more powerful result is valid, but it relies on the so-called strong law of large numbers which is beyond the range of this volume.

Proof The law of large numbers tells us about the asymptotic behaviour of the sum of independent, identically distributed random variables, and the key to the current proof is to write $V_i(n)$ in terms of such a sum. Let

$$U_1 = \inf\{n \geq 1 : X_n = i\},$$

be the time until the first visit to i, and for $m \geq 1$ let U_m be the time between the mth and $(m+1)$th visits. Since the chain is assumed positive recurrent, we have that $\mathbb{P}(U_m < \infty) = 1$ and $\mu_i = \mathbb{E}_i(U_1) < \infty$. The *first* passage time U_1 may have a different distribution from the remaining U_m if $X_0 \neq i$.

By the strong Markov property, the random variables $U_1, U_2, \ldots$ are independent, and $U_2, U_3, \ldots$ are identically distributed. Moreover,

$$V_i(n) \geq x \quad \text{if and only if} \quad S_{\lceil x \rceil} \leq n,$$

where $\lceil x \rceil$ is the least integer not less than x, and

$$S_m = \sum_{r=1}^{m} U_r.$$

is the time of the mth visit to i. Therefore,

$$\mathbb{P}\left(\frac{1}{n} V_i(n) \geq \frac{1+\epsilon}{\mu_i}\right) = \mathbb{P}\left(S_N \leq n\right), \tag{12.102}$$

where $N = \lceil (1+\epsilon)n/\mu_i \rceil$. By the weak law of large numbers, Theorem 8.17,

$$\frac{1}{N}S_N = \frac{1}{N}U_1 + \frac{1}{N}\sum_{r=2}^{N} U_r$$
$$\Rightarrow \mu_i \qquad \text{as } n \to \infty, \tag{12.103}$$

where we have used the fact that $U_1/N \to 0$ in probability (see Theorem 8.47 and Problem 8.6.9). By (12.102)–(12.103),

$$\mathbb{P}\left(\frac{1}{n}V_i(n) \geq \frac{1+\epsilon}{\mu_i}\right) \to \begin{cases} 0 & \text{if } \epsilon > 0, \\ 1 & \text{if } \epsilon < 0. \end{cases}$$

There is a gap in this proof, since Theorem 8.17 assumed that a typical summand, U_2 say, has finite variance. If that is not known, then it is necessary to appeal to the more powerful conclusion of Example 8.52 whose proof uses the method of characteristic functions. □

Exercise 12.104 Let π be the unique invariant distribution of an aperiodic, irreducible Markov chain $\mathbf{X}$. Show that $\mathbb{P}(X_n = j) \to \pi_j$ as $n \to \infty$, regardless of the initial distribution of X_0.

12.11 Time reversal

An important observation of physics is that many equations are valid irrespective of whether time flows forwards or backwards. Invariance under time-reversal is an important property of certain Markov chains.

Let $\mathbf{X} = (X_n : 0 \leq n \leq N)$ be an irreducible, positive recurrent Markov chain, with transition matrix P and invariant distribution π. Suppose further that X_0 has distribution π, so that X_n has distribution π for every n. The 'reversed chain' $\mathbf{Y} = (Y_n : 0 \leq n \leq N)$ is given by reversing time: $Y_n = X_{N-n}$ for $0 \leq n \leq N$. Recall from Theorem 12.81(b) that $\pi_i > 0$ for $i \in S$.

Theorem 12.105 *The sequence* $\mathbf{Y}$ *is an irreducible Markov chain with transition matrix* $\widehat{P} = (\widehat{p}_{i,j} : i, j \in S)$ *given by*

$$\widehat{p}_{i,j} = (\pi_j/\pi_i)p_{j,i} \qquad \text{for } i, j \in S, \tag{12.106}$$

and with invariant distribution π.

Proof We check first that $\widehat{P}$ is a stochastic matrix. Certainly its entries are non-negative, and also

$$\sum_{j \in S} \widehat{p}_{i,j} = \frac{1}{\pi_i}\sum_{j \in S} \pi_j p_{j,i} = \frac{1}{\pi_i}\pi_i = 1,$$

since π is invariant for P.

Next we show that π is invariant for $\widehat{P}$. By (12.106),

$$\sum_{i \in S} \pi_i \widehat{p}_{i,j} = \sum_{i \in S} \pi_j p_{j,i} = \pi_j,$$

since P has row-sums 1.

Finally, by Theorem 12.4,

$$\begin{aligned}\mathbb{P}\big(Y_0 = i_0,\ Y_1 = i_1,\ \ldots,\ Y_n = i_n\big) &= \mathbb{P}\big(X_{N-n} = i_n,\ X_{N-n+1} = i_{n-1},\ \ldots,\ X_N = i_0\big)\\ &= \pi_{i_n} p_{i_n,i_{n-1}} \cdots p_{i_1,i_0}\\ &= \pi_{i_0} \widehat{p}_{i_0,i_1} \cdots \widehat{p}_{i_{n-1},i_n} \qquad \text{by (12.106).}\end{aligned}$$

By Theorem 12.4 again, $\mathbf{Y}$ has transition matrix $\widehat{P}$ and initial distribution π. □

We call the chain $\mathbf{Y}$ the *time-reversal* of the chain $\mathbf{X}$, and we say that $\mathbf{X}$ is *reversible* if $\mathbf{X}$ and its time-reversal have the same transition probabilities.

Definition 12.107 *Let $\mathbf{X} = (X_n : 0 \le n \le N)$ be an irreducible Markov chain such that X_0 has the invariant distribution π. The chain is* **reversible** *if $\mathbf{X}$ and its time-reversal $\mathbf{Y}$ have the same transition matrices, which is to say that*

$$\pi_i p_{i,j} = \pi_j p_{j,i} \qquad \textit{for } i, j \in S. \tag{12.108}$$

Equations (12.108) are called the *detailed balance equations*, and they are pivotal to the study of reversible chains. More generally we say that a transition matrix P and a distribution λ are *in detailed balance* if

$$\lambda_i p_{i,j} = \lambda_j p_{j,i} \qquad \text{for } i, j \in S.$$

An irreducible chain $\mathbf{X}$ with invariant distribution π is said to be *reversible in equilibrium* if its transition matrix P is in detailed balance with π.

It turns out that, for an irreducible chain, P is in detailed balance with a distribution λ if and only if λ is the unique invariant distribution. This provides a good way of finding the invariant distribution of a reversible chain.

Theorem 12.109 *Let P be the transition matrix of an irreducible chain $\mathbf{X}$, and suppose that π is a distribution satisfying*

$$\pi_i p_{i,j} = \pi_j p_{j,i} \qquad \textit{for } i, j \in S. \tag{12.110}$$

Then π is the unique invariant distribution of the chain. Furthermore, $\mathbf{X}$ is reversible in equilibrium.

Proof Suppose that π is a distribution that satisfies (12.110). Then

$$\sum_{i \in S} \pi_i p_{i,j} = \sum_{i \in S} \pi_j p_{j,i} = \pi_j \sum_{i \in S} p_{j,i} = \pi_j,$$

since P has row-sums 1. Therefore, $\pi = \pi P$, whence π is invariant. The reversibility in equilibrium of $\mathbf{X}$ follows by checking Definition 12.107. □

The above discussion of reversibility is restricted to a Markov chain with only finitely many time-points $0, 1, 2, \ldots, N$. It is easily extended to the infinite time set $0, 1, 2, \ldots$. It may even be extended to the doubly-infinite time set $\dots, -2, -1, 0, 1, 2, \dots$, subject to the assumption that X_n has the invariant distribution π for all n.

Time-reversibility is a very useful concept in the theory of random networks. There is a valuable analogy using the language of flows. Let $\mathbf{X}$ be a Markov chain with state space S and invariant distribution π. To this chain there corresponds the following directed network (or graph). The vertices of the network are the states of the chain, and an arrow is placed from vertex i to vertex j if $p_{i,j} > 0$. One unit of a notional material ('probability') is distributed about the vertices and allowed to flow along the arrows. A proportion π_i of the material is placed initially at vertex i. At each epoch of time, a proportion $p_{i,j}$ of the amount of material at each vertex i is transported to each vertex j.

It is immediate that the amount of material at vertex i after one epoch is $\sum_j \pi_j p_{j,i}$, which equals π_i since $\pi = \pi P$. That is to say, the deterministic flow of probability is in equilibrium: there is 'global balance' in the sense that the total quantity leaving each vertex is balanced by an equal quantity arriving there. There may or may not be 'local balance', in the sense that, for each $i, j \in S$, the amount flowing from i to j equals the amount flowing from j to i. Local balance occurs if and only if $\pi_i p_{i,j} = \pi_j p_{j,i}$ for $i, j \in S$, which is to say that P and π are in detailed balance.

Example 12.111 (Birth–death chain with retaining barrier) Consider a random walk $\mathbf{X} = (X_n : n \geq 0)$ on the non-negative integers $\{0, 1, 2, \dots\}$ which, when at $i \geq 1$, moves one step rightwards with probability p_i, or one step leftwards with probability q_i $(= 1 - p_i)$. When at $i = 0$, it stays at 0 with probability q_0 and otherwise moves to 1. We assume for simplicity that $0 < p_i < 1$ for all i. This process differs from the birth–death chain of Example 12.59 in its behaviour at 0.

Under what conditions on the p_i is the Markov chain $\mathbf{X}$ reversible in equilibrium? If this holds, find the invariant distribution.

Solution We look for a solution to the detailed balance equations (12.110), which may be written as

$$\pi_{i-1} p_{i-1} = \pi_i q_i \qquad \text{for } i \geq 1.$$

By iteration, the solution is

$$\pi_i = \rho_i \pi_0 \qquad \text{for } i \geq 1, \tag{12.112}$$

where

$$\rho_i = \frac{p_{i-1} p_{i-2} \cdots p_0}{q_i q_{i-1} \cdots q_1}.$$

The vector π is a distribution if and only if $\sum_i \pi_i = 1$. By (12.112),

$$\sum_{i \in S} \pi_i = \pi_0 \sum_{i \in S} \rho_i.$$

We may choose π_0 appropriately if and only if $S = \sum_i \rho_i$ satisfies $S < \infty$, in which case we set $\pi_0 = 1/S$.

By Theorem 12.109, $\mathbf{X}$ is reversible in equilibrium if and only if $S < \infty$, in which case the invariant distribution is given by $\pi_i = \rho_i / S$. △

Example 12.113 (Ehrenfest dog–flea model) Two dogs, Albert and Beatrice, are infested by a total of m fleas that jump from one dog to the other at random. We assume that at each epoch of time one flea, picked uniformly at random from the m available, passes from its current host to the other dog. Let X_n be the number of fleas on Albert after n units of time has passed. Thus, $\mathbf{X} = (X_n : n \geq 0)$ is an irreducible Markov chain with transition matrix

$$p_{i,i+1} = 1 - \frac{i}{m}, \quad p_{i,i-1} = \frac{i}{m} \qquad \text{for } 0 \leq i \leq m.$$

Rather than solve the equation $\pi = \pi P$ to find the invariant distribution, we look for solutions of the detailed balance equations $\pi_i p_{i,j} = \pi_j p_{j,i}$. These equations amount to

$$\pi_{i-1}\left(\frac{m-i+1}{m}\right) = \pi_i \cdot \frac{i}{m} \qquad \text{for } 1 \leq i \leq m.$$

By iteration,

$$\pi_i = \binom{m}{i}\pi_0,$$

and we choose $\pi_0 = 2^{-m}$ so that π is a distribution. By Theorem 12.109, π is the unique invariant distribution. △

Exercise 12.114 Consider a random walk on a triangle, illustrated in Figure 12.5. The state space is $S = \{1, 2, 3\}$, and the transition matrix is

$$P = \begin{pmatrix} 0 & \alpha & 1-\alpha \\ 1-\alpha & 0 & \alpha \\ \alpha & 1-\alpha & 0 \end{pmatrix},$$

where $0 < \alpha < 1$. Show that the detailed balance equations possess a solution if and only if $\alpha = \frac{1}{2}$.

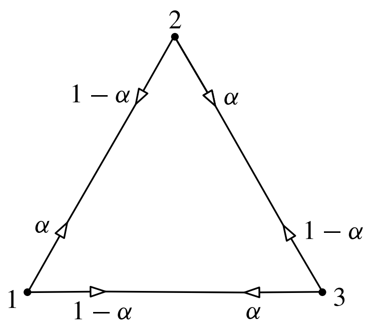

Fig. 12.5 Transition probabilities for a random walk on a triangle.

Exercise 12.115 Can a reversible Markov chain be periodic? Explain.

Exercise 12.116 A random walk moves on the finite set $\{0, 1, 2, \dots, N\}$. When in the interior of the interval, it moves one step rightwards with probability p, or one step leftwards with probability q $(= 1 - p)$. When it is at either endpoint, 0 or N, and tries to leave the interval, it is retained at its current position. Assume $0 < p < 1$, and use the detailed balance equations to find the invariant distribution.

12.12 Random walk on a graph

A *graph* $G = (V, E)$ is a set V of vertices, pairs of which are joined by edges. That is, the edge-set E is a set of distinct unordered pairs $\langle u, v \rangle$ of distinct elements of V. A graph is usually represented in the manner illustrated in Figure 12.6. The lattice-graphs $\mathbb{Z}^d$ of Section 12.5 are examples of infinite graphs.

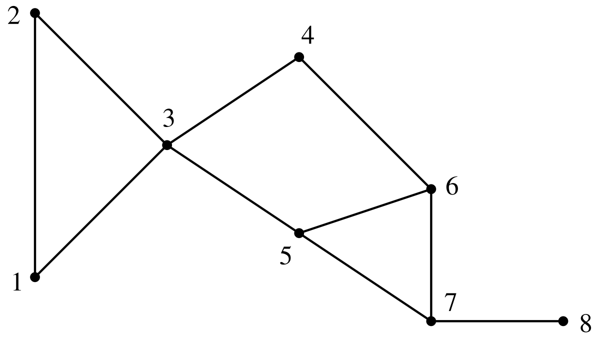

Fig. 12.6 A graph G with 8 vertices. A random walk on G moves around the vertex-set. At each step, it moves to a uniformly random neighbour of its current position.

Here is some language and notation concerning graphs. A graph is connected if, for distinct $u, v \in V$, there exists a path of edges from u to v. We write $u \sim v$ if $\langle u, v \rangle \in E$, in which case we say that u and v are *neighbours*. The *degree* $d(v)$ of vertex v is the number of edges containing v, that is, $d(v) = |\{u \in V : v \sim u\}|$.

There is a rich theory of random walks on finite and infinite graphs. Let $G = (V, E)$ be a connected graph with $d(v) < \infty$ for all $v \in V$. A particle moves about the vertices of G, taking steps along the edges. Let X_n be the position of the particle at time n. At time $n + 1$, it moves to a uniformly random neighbour of X_n. More precisely, a random walk is the Markov chain $\mathbf{X} = (X_n : n \geq 0)$ with state space V and transition matrix

$$p_{u,v} = \begin{cases} \dfrac{1}{d(u)} & \text{if } v \sim u, \\ 0 & \text{otherwise.} \end{cases} \tag{12.117}$$

When G is infinite, the main question is to understand the long-run behaviour of the walk, such as whether or not it is transient or recurrent. This was the question addressed in Section 12.5 for the lattice-graphs $\mathbb{Z}^d$. In this section, we consider a *finite* connected graph G, and we prove the following main result. It will be useful to note that

$$\sum_{v \in V} d(v) = 2|E|, \tag{12.118}$$

since each edge contributes 2 to the summation.

Theorem 12.119 *Random walk on the finite connected graph $G = (V, E)$ is an irreducible, Markov chain with unique invariant distribution*

$$\pi_v = \frac{d(v)}{2|E|} \qquad \text{for } v \in V.$$

The chain is reversible in equilibrium.

Proof Since G is connected, the chain is irreducible. The vector π is certainly a distribution since $\pi_v \geq 0$ for $v \in V$, and $\sum_{v \in V} \pi_v = 1$ by (12.118). By Theorem 12.109, it suffices to check the detailed balance equations (12.110), namely

$$\frac{d(u)}{2|E|} p_{u,v} = \frac{d(v)}{2|E|} p_{v,u}, \qquad \text{for } u, v \in V.$$

This holds by the definition (12.117) of the transition probabilities. □

Example 12.120 (Erratic knights) A knight is the sole inhabitant of a chess board, and it performs random moves. Each move is chosen at random from the set of currently permissible moves, as illustrated in Figure 12.7. What is the invariant distribution of the Markov chain describing the knight's motion?

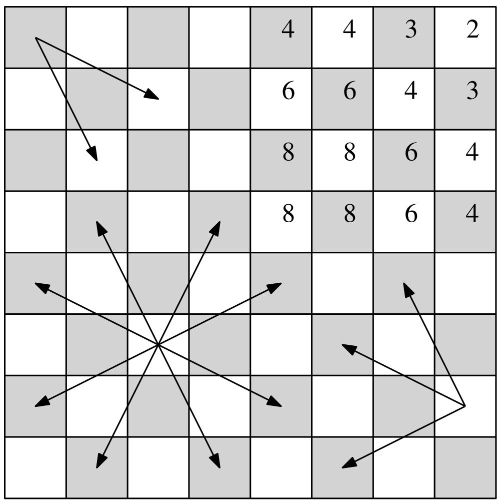

Fig. 12.7 A map for the erratic knight. The arrows indicate permissible moves. If the knight is at a square from which there are m permissible moves, then it selects one of these with equal probability $1/m$. The numbers are the degrees of the corresponding graph vertices.

Solution Let $G = (V, E)$ be the graph given as follows. The vertex-set V is the set of squares of the chess board, and the edge-set E is given by: two vertices u, v are joined by an edge if and only if the move between u and v is a legal knight-move. The knight performs a random walk on G. In order to find the invariant distribution we must count the vertex-degrees. The four corners have degree 2, and so on, as indicated in Figure 12.7. The sum of the vertex-degrees is

$$\sum_{v \in V} d(v) = 4 \cdot 2 + 8 \cdot 3 + 20 \cdot 4 + 16 \cdot 6 + 16 \cdot 8 = 336,$$

and the invariant distribution is given by Theorem 12.119 as $\pi_v = d(v)/336$. △

Exercise 12.121 An erratic king performs random (but legal) moves on a chess board. Find his invariant distribution.

12.13 Problems

1. A transition matrix is called *doubly stochastic* if its column-sums equal 1, that is, if $\sum_{i \in S} p_{i,j} = 1$ for $j \in S$.

   Suppose an irreducible chain with N $(< \infty)$ states has a doubly stochastic transition matrix. Find its invariant distribution. Deduce that all states are positive recurrent and that, if the chain is aperiodic, then $p_{i,j}(n) \to 1/N$ as $n \to \infty$.

2. Let $\mathbf{X}$ be a discrete-time Markov chain with state space $S = \{1, 2\}$, and transition matrix

   $$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}.$$

   Classify the states of the chain. Suppose that $0 < \alpha\beta < 1$. Find the n-step transition probabilities and show directly that they converge to the unique invariant distribution. For what values of α and β is the chain reversible in equilibrium?

3. We distribute N black balls and N white balls in two urns in such a way that each contains N balls. At each epoch of time, one ball is selected at random from each urn, and these two balls are interchanged. Let X_n be the number of black balls in the first urn after time n. Write down the transition matrix of this Markov chain, and find the unique invariant distribution. Is the chain reversible in equilibrium?

4. Consider a Markov chain on the set $S = \{0, 1, 2, \dots\}$ with transition probabilities

   $$p_{i,i+1} = a_i, \quad p_{i,0} = 1 - a_i,$$

   where $(a_i : i \geq 0)$ is a sequence of constants satisfying $0 < a_i < 1$ for all i. Let $b_0 = 1$ and $b_i = a_0 a_1 \cdots a_{i-1}$ for $i \geq 1$. Show that the chain is

   (a) recurrent if and only if $b_i \to 0$ as $i \to \infty$,
   (b) positive recurrent if and only if $\sum_i b_i < \infty$,

   and write down the invariant distribution when the last condition holds.

5. At each time n, a random number S_n of students enter the lecture room, where $(S_n : n \geq 0)$ are independent and Poisson distributed with parameter λ. Each student remains in the room for a geometrically distributed time with parameter p, different times being independent. Let X_n be the number of students present at time n. Show that $\mathbf{X}$ is a Markov chain, and find its invariant distribution.

6. Each morning, a student takes one of three books (labelled 1, 2, and 3) from her shelf. She chooses book i with probability α_i, and choices on successive days are independent. In the evening, she replaces the book at the left-hand end of the shelf. If p_n denotes the probability that on day n she finds the books in the order 1, 2, 3 from the left to right, show that p_n converges as $n \to \infty$, and find the limit.

7. Let **X** be an irreducible, positive recurrent, aperiodic Markov chain. Show that **X** is reversible in equilibrium if and only if

$$p_{i_1,i_2} p_{i_2,i_3} \cdots p_{i_{n-1},i_n} p_{i_n,i_1} = p_{i_1,i_n} p_{i_n,i_{n-1}} \cdots p_{i_2,i_1},$$

for all finite sequences $i_1, i_2, \dots, i_n \in S$.

8. A special die is thrown repeatedly. Its special property is that, at each throw, the outcome is equally likely to be any of the five numbers that are different from the immediately previous number. If the first score is 1, find the probability that the $(n+1)$th score is 1.

9. A particle performs a random walk about the eight vertices of a cube. Find
   (a) the mean number of steps before it returns to its starting vertex S,
   (b) the mean number of visits to the opposite vertex T to S before its first return to S,
   (c) the mean number of steps before its first visit to T.

10. *Markov chain Monte Carlo.* We wish to simulate a discrete random variable Z with mass function satisfying $\mathbb{P}(Z = i) \propto \pi_i$, for $i \in S$ and S countable. Let **X** be an irreducible Markov chain with state space S and transition matrix $P = (p_{i,j})$. Let $Q = (q_{i,j})$ be given by

$$q_{i,j} = \begin{cases} \min\{p_{i,j}, (\pi_j/\pi_i) p_{j,i}\} & \text{if } i \neq j, \\ 1 - \sum_{j:\, j \neq i} q_{i,j} & \text{if } i = j. \end{cases}$$

Show that Q is the transition matrix of a Markov chain which is reversible in equilibrium, and has invariant distribution equal to the mass function of Z.

11. Let i be a state of an irreducible, positive recurrent Markov chain **X**, and let V_n be the number of visits to i between times 1 and n. Let $\mu = \mathbb{E}_i(T_i)$ and $\sigma^2 = \mathbb{E}_i([T_i - \mu]^2)$ be the mean and variance of the first return time to the starting state i, and assume $0 < \sigma^2 < \infty$.

    Suppose $X_0 = i$. Show that

$$U_n = \frac{V_n - (n/\mu)}{\sqrt{n\sigma^2/\mu^3}}$$

converges in distribution to the normal distribution N(0, 1) as $n \to \infty$.

12. Consider a pack of cards labelled $1, 2, \dots, 52$. We repeatedly take the top card and insert it uniformly at random in one of the 52 possible places, that is, either on the top or on the bottom or in one of the 50 places inside the pack. How long on average will it take for the bottom card to reach the top?

    Let p_n denote the probability that after n iterations the cards are found to be in increasing order from the top. Show that, irrespective of the initial ordering, p_n converges as $n \to \infty$, and determine the limit p. You should give precise statements of any general results to which you appeal.

    Show that, at least until the bottom card reaches the top, the ordering of the cards inserted beneath it is uniformly random. Hence or otherwise show that, for all n,

$$|p_n - p| \leq \frac{52(1 + \log 51)}{n}.$$

(Cambridge 2003)

13. Consider a collection of N books arranged in a line along a bookshelf. At successive units of time, a book is selected randomly from the collection. After the book has been consulted, it is replaced on the shelf one position to the left of its original position, with the book in that position moved to the right by one. That is, the selected book and its neighbour to the left swap positions. If the selected book is already in the leftmost position it is returned there. All but one of the books have plain covers and are equally likely to be selected. The other book has a red cover. At each time unit, the red book will be selected with probability p, where $0 < p < 1$. Each other book will be selected with probability $(1-p)/(N-1)$. Successive choices of book are independent.

   Number the positions on the shelf from 1 (at the left) to N (at the right). Write X_n for the position of the red book after n units of time. Show that $\mathbf{X}$ is a Markov chain, with non-zero transition probabilities given by:

$$\begin{aligned} p_{i,i-1} &= p && \text{for } i = 2, 3, \dots, N, \\ p_{i,i+1} &= \frac{1-p}{N-1} && \text{for } i = 1, 2, \dots, N-1, \\ p_{i,i} &= 1 - p - \frac{1-p}{N-1} && \text{for } i = 2, 3, \dots, N-1, \\ p_{1,1} &= 1 - \frac{1-p}{N-1}, \\ p_{N,N} &= 1 - p. \end{aligned}$$

   If $(\pi_i : i = 1, 2, \dots, N)$ is the invariant distribution of the Markov chain $\mathbf{X}$, show that

$$\pi_2 = \frac{1-p}{p(N-1)}\pi_1, \qquad \pi_3 = \frac{1-p}{p(N-1)}\pi_2.$$

   Deduce the invariant distribution. (Oxford 2005)

* 14. Consider a Markov chain with state space $S = \{0, 1, 2, \dots\}$ and transition matrix given by

$$p_{i,j} = \begin{cases} qp^{j-i+1} & \text{for } i \geq 1 \text{ and } j \geq i-1, \\ qp^j & \text{for } i = 0 \text{ and } j \geq 0, \end{cases}$$

   and $p_{i,j} = 0$ otherwise, where $0 < p = 1 - q < 1$.

   For each $p \in (0, 1)$, determine whether the chain is transient, null recurrent, or positive recurrent, and in the last case find the invariant distribution. (Cambridge 2007)

15. Let $(X_n : n \geq 0)$ be a simple random walk on the integers: the random variables $\xi_n := X_n - X_{n-1}$ are independent, with distribution

$$\mathbb{P}(\xi = 1) = p, \quad \mathbb{P}(\xi = -1) = q,$$

   where $0 < p < 1$ and $q = 1 - p$. Consider the hitting time $\tau = \inf\{n : X_n = 0 \text{ or } X_n = N\}$, where $N > 1$ is a given integer. For fixed $s \in (0, 1)$ define

$$H_k = \mathbb{E}\big(s^\tau 1(X_\tau = 0) \,\big|\, X_0 = k\big) \qquad \text{for } k = 0, 1, \dots, N.$$

   Show that the H_k satisfy a second-order difference equation, and hence find them. (Cambridge 2009)

16. An erratic bishop starts at the bottom left of a chess board and performs random moves. At each stage, she picks one of the available legal moves with equal probability, independently of

earlier moves. Let X_n be her position after n moves. Show that $(X_n : n \geq 0)$ is a reversible Markov chain, and find its invariant distribution.

What is the mean number of moves before she returns to her starting square?

17. A frog inhabits a pond with an infinite number of lily pads, numbered $1, 2, 3 \dots$. She hops from pad to pad in the following manner: if she happen to be on pad i at a given time, she hops to one of the pads $(1, 2, \dots, i, i + 1)$ with equal probability.

    (a) Find the equilibrium distribution of the corresponding Markov chain.
    (b) Suppose the frog starts on pad k and stops when she returns to it. Show that the expected number of times the frog hops is $e(k - 1)!$ where $e = 2.718 \dots$. What is the expected number of times she will visit the lily pad $k + 1$?

    (Cambridge 2010)

18. Let $(X_n : n \geq 0)$ be a simple, symmetric random walk on the integers $\{\dots, -1, 0, 1, \dots\}$, with $X_0 = 0$ and
$$\mathbb{P}\big(X_{n+1} = i \pm 1 \,\big|\, X_n = i\big) = \tfrac{1}{2}.$$
    For each integer $a \geq 1$, let $T_a = \inf\{n \geq 0 : X_n = a\}$. Show that T_a is a stopping time.

    Define a random variable Y_n by the rule
$$Y_n = \begin{cases} X_n & \text{if } n < T_a, \\ 2a - X_n & \text{if } n \geq T_a. \end{cases}$$
    Show that $(Y_n : n \geq 0)$ is also a simple, symmetric random walk.

    Let $M_n = \max\{X_i : 0 \leq i \leq n\}$. Explain why $\{M_n \geq a\} = \{T_a \leq n\}$ for $a \geq 1$. By using the process $(Y_n : n \geq 0)$ constructed above, show that, for $a \geq 1$,
$$\mathbb{P}\big(M_n \geq a,\ X_n \leq a - 1\big) = \mathbb{P}(X_n \geq a + 1),$$
    and thus
$$\mathbb{P}(M_n \geq a) = \mathbb{P}(X_n \geq a) + \mathbb{P}(X_n \geq a + 1).$$
    Hence compute $\mathbb{P}(M_n = a)$ where a and n are positive integers with $n \geq a$. [Hint: if n is even, then X_n must be even, and if n is odd, then X_n must be odd.] (Cambridge 2010)