1

Discrete-time Markov chains

This chapter is the foundation for all that follows. Discrete-time Markov chains are defined and their behaviour is investigated. For better orientation we now list the key theorems: these are Theorems 1.3.2 and 1.3.5 on hitting times, Theorem 1.4.2 on the strong Markov property, Theorem 1.5.3 characterizing recurrence and transience, Theorem 1.7.7 on invariant distributions and positive recurrence. Theorem 1.8.3 on convergence to equilibrium, Theorem 1.9.3 on reversibility, and Theorem 1.10.2 on long-run averages. Once you understand these you will understand the basic theory. Part of that understanding will come from familiarity with examples, so a large number are worked out in the text. Exercises at the end of each section are an important part of the exposition.

1.1 Definition and basic properties

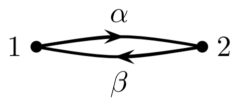
Let I be a countable set. Each $i \in I$ is called a *state* and I is called the *state-space*. We say that $\lambda = (\lambda_i : i \in I)$ is a *measure* on I if $0 \leq \lambda_i < \infty$ for all $i \in I$. If in addition the *total mass* $\sum_{i \in I} \lambda_i$ equals 1, then we call λ a *distribution*. We work throughout with a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Recall that a *random variable* X with values in I is a function $X : \Omega \to I$. Suppose we set

$$\lambda_i = \mathbb{P}(X = i) = \mathbb{P}(\{\omega : X(\omega) = i\}).$$

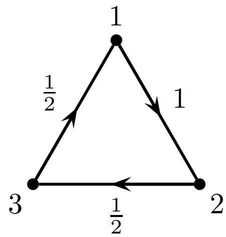
Then λ defines a distribution, the *distribution of* X. We think of X as modelling a random state which takes the value i with probability λ_i. There is a brief review of some basic facts about countable sets and probability spaces in Chapter 6.

We say that a matrix $P = (p_{ij} : i, j \in I)$ is *stochastic* if every row $(p_{ij} : j \in I)$ is a distribution. There is a one-to-one correspondence between stochastic matrices P and the sort of diagrams described in the Introduction. Here are two examples:

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



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



We shall now formalize the rules for a Markov chain by a definition in terms of the corresponding matrix P. We say that $(X_n)_{n\geq 0}$ is a *Markov chain* with *initial distribution* λ and *transition matrix* P if

(i) X_0 has distribution λ;
(ii) for $n \geq 0$, conditional on $X_n = i$, X_{n+1} has distribution $(p_{ij} : j \in I)$ and is independent of $X_0, \ldots, X_{n-1}$.

More explicitly, these conditions state that, for $n \geq 0$ and $i_1, \ldots, i_{n+1} \in I$,

(i) $\mathbb{P}(X_0 = i_1) = \lambda_{i_1}$;
(ii) $\mathbb{P}(X_{n+1} = i_{n+1} \mid X_0 = i_1, \ldots, X_n = i_n) = p_{i_n i_{n+1}}$.

We say that $(X_n)_{n\geq 0}$ is *Markov*(λ, P) for short. If $(X_n)_{0\leq n\leq N}$ is a finite sequence of random variables satisfying (i) and (ii) for $n = 0, \ldots, N-1$, then we again say $(X_n)_{0\leq n\leq N}$ is *Markov*(λ, P).

It is in terms of properties (i) and (ii) that most real-world examples are seen to be Markov chains. But mathematically the following result appears to give a more comprehensive description, and it is the key to some later calculations.

Theorem 1.1.1. *A discrete-time random process $(X_n)_{0\leq n\leq N}$ is Markov(λ, P) if and only if for all $i_1, \ldots, i_N \in I$*

$$\mathbb{P}(X_0 = i_1, X_1 = i_2, \ldots, X_N = i_N) = \lambda_{i_1} p_{i_1 i_2} p_{i_2 i_2} \ldots p_{i_{N-1} i_N}. \tag{1.1}$$

Proof. Suppose $(X_n)_{0\le n\le N}$ is Markov(λ, P), then

$$\begin{aligned}
&\mathbb{P}(X_0 = i_1, X_1 = i_2, \dots , X_N = i_N)\\
&\quad= \mathbb{P}(X_0 = i_1)\mathbb{P}(X_1 = i_2 \mid X_0 = i_1)\\
&\qquad\dots \mathbb{P}(X_N = i_N \mid X_0 = i_1, \dots , X_{N-1} = i_{N-1})\\
&\quad= \lambda_{i_1} p_{i_1 i_2} \dots p_{i_{N-1} i_N}.
\end{aligned}$$

On the other hand, if (1.1) holds for N, then by summing both sides over $i_N \in I$ and using $\sum_{j\in I} p_{ij} = 1$ we see that (1.1) holds for $N-1$ and, by induction

$$\mathbb{P}(X_0 = i_1, X_1 = i_2, \dots , X_n = i_n) = \lambda_{i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n}$$

for all $n = 0, 1, \dots , N$. In particular, $\mathbb{P}(X_0 = i_1) = \lambda_{i_1}$ and, for $n = 0, 1, \dots , N-1$,

$$\begin{aligned}
&\mathbb{P}(X_{n+1} = i_{n+1} \mid X_0 = i_1, \dots , X_n = i_n)\\
&\quad= \mathbb{P}(X_0 = i_1, \dots , X_n = i_n, X_{n+1} = i_{n+1})/\mathbb{P}(X_0 = i_1, \dots , X_n = i_n)\\
&\quad= p_{i_n i_{n+1}}.
\end{aligned}$$

So $(X_n)_{0\le n\le N}$ is Markov(λ, P). $\square$

The next result reinforces the idea that Markov chains have no memory. We write $\delta_i = (\delta_{ij} : j \in I)$ for the *unit mass* at i, where

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

Theorem 1.1.2 (Markov property). *Let $(X_n)_{n\ge0}$ be Markov(λ, P). Then, conditional on $X_m = i$, $(X_{m+n})_{n\ge0}$ is Markov(δ_i, P) and is independent of the random variables $X_0, \dots , X_m$.*

Proof. We have to show that for any event A determined by $X_0, \dots , X_m$ we have

$$\begin{aligned}
&\mathbb{P}(\{X_m = i_m, \dots , X_{m+n} = i_{m+n}\} \cap A \mid X_m = i)\\
&\qquad\qquad= \delta_{i i_m} p_{i_m i_{m+1}} \dots p_{i_{m+n-1} i_{m+n}} \mathbb{P}(A \mid X_m = i)
\end{aligned} \tag{1.2}$$

then the result follows by Theorem 1.1.1. First consider the case of elementary events

$$A = \{X_0 = i_1, \dots , X_m = i_m\}.$$

In that case we have to show

$$\begin{aligned}&\mathbb{P}(X_0 = i_1, \dots, X_{m+n} = i_{m+n} \text{ and } i = i_m)/\mathbb{P}(X_m = i)\\ &\quad = \delta_{ii_m} p_{i_m i_{m+1}} \dots p_{i_{m+n-1} i_{m+n}}\\ &\qquad \times \mathbb{P}(X_0 = i_1, \dots, X_m = i_m \text{ and } i = i_m)/\mathbb{P}(X_m = i)\end{aligned}$$

which is true by Theorem 1.1.1. In general, any event A determined by $X_0, \dots, X_m$ may be written as a countable disjoint union of elementary events

$$A = \bigcup_{k=1}^{\infty} A_k.$$

Then the desired identity (1.2) for A follows by summing up the corresponding identities for A_k. $\square$

The remainder of this section addresses the following problem: *what is the probability that after n steps our Markov chain is in a given state?* First we shall see how the problem reduces to calculating entries in the nth power of the transition matrix. Then we shall look at some examples where this may be done explicitly.

We regard distributions and measures λ as row vectors whose components are indexed by I, just as P is a matrix whose entries are indexed by $I \times I$. When I is finite we will often label the states $1, 2, \dots, N$; then λ will be an N-vector and P an $N \times N$-matrix. For these objects, matrix multiplication is a familiar operation. We extend matrix multiplication to the general case in the obvious way, defining a new measure λP and a new matrix P^2 by

$$(\lambda P)_j = \sum_{i \in I} \lambda_i p_{ij}, \quad (P^2)_{ik} = \sum_{j \in I} p_{ij} p_{jk}.$$

We define P^n similarly for any n. We agree that P^0 is the identity matrix I, where $(I)_{ij} = \delta_{ij}$. The context will make it clear when I refers to the state-space and when to the identity matrix. We write $p_{ij}^{(n)} = (P^n)_{ij}$ for the (i, j) entry in P^n.

In the case where $\lambda_i > 0$ we shall write $\mathbb{P}_i(A)$ for the conditional probability $\mathbb{P}(A \mid X_0 = i)$. By the Markov property at time $m = 0$, under $\mathbb{P}_i$, $(X_n)_{n\geq 0}$ is Markov(δ_i, P). So the behaviour of $(X_n)_{n\geq 0}$ under $\mathbb{P}_i$ does not depend on λ.

Theorem 1.1.3. *Let $(X_n)_{n\geq 0}$ be Markov(λ, P). Then, for all $n, m \geq 0$,*

(i) $\mathbb{P}(X_n = j) = (\lambda P^n)_j$;

(ii) $\mathbb{P}_i(X_n = j) = \mathbb{P}(X_{n+m} = j \mid X_m = i) = p_{ij}^{(n)}$.

Proof. (i) By Theorem 1.1.1

$$\begin{aligned}\mathbb{P}(X_n = j) &= \sum_{i_1 \in I} \dots \sum_{i_{n-1} \in I} \mathbb{P}(X_0 = i_1, \dots, X_{n-1} = i_{n-1}, X_n = j) \\ &= \sum_{i_1 \in I} \dots \sum_{i_{n-1} \in I} \lambda_{i_1} p_{i_1 i_2} \dots p_{i_{n-1} j} = (\lambda P^n)_j.\end{aligned}$$

(ii) By the Markov property, conditional on $X_m = i$, $(X_{m+n})_{n \geq 0}$ is Markov (δ_i, P), so we just take $\lambda = \delta_i$ in (i). □

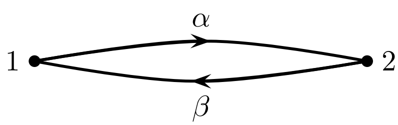
In light of this theorem we call $p_{ij}^{(n)}$ the *n-step transition probability from i to j*. The following examples give some methods for calculating $p_{ij}^{(n)}$.

Example 1.1.4

The most general two-state chain has transition matrix of the form

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

and is represented by the following diagram:



We exploit the relation $P^{n+1} = P^n P$ to write

$$p_{11}^{(n+1)} = p_{12}^{(n)} \beta + p_{11}^{(n)} (1 - \alpha).$$

We also know that $p_{11}^{(n)} + p_{12}^{(n)} = \mathbb{P}_1(X_n = 1 \text{ or } 2) = 1$, so by eliminating $p_{12}^{(n)}$ we get a recurrence relation for $p_{11}^{(n)}$:

$$p_{11}^{(n+1)} = (1 - \alpha - \beta) p_{11}^{(n)} + \beta, \quad p_{11}^{(0)} = 1.$$

This has a unique solution (see Section 1.11):

$$p_{11}^{(n)} = \begin{cases} \dfrac{\beta}{\alpha + \beta} + \dfrac{\alpha}{\alpha + \beta}(1 - \alpha - \beta)^n & \text{for } \alpha + \beta > 0 \\ 1 & \text{for } \alpha + \beta = 0. \end{cases}$$

Example 1.1.5 (Virus mutation)

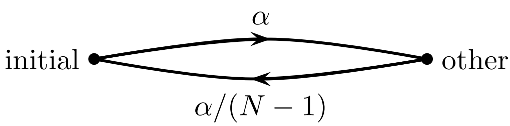
Suppose a virus can exist in N different strains and in each generation either stays the same, or with probability α mutates to another strain, which is chosen at random. What is the probability that the strain in the nth generation is the same as that in the 0th?

We could model this process as an N-state chain, with $N \times N$ transition matrix P given by

$$p_{ii} = 1 - \alpha, \quad p_{ij} = \alpha/(N-1) \quad \text{for } i \neq j.$$

Then the answer we want would be found by computing $p_{11}^{(n)}$. In fact, in this example there is a much simpler approach, which relies on exploiting the symmetry present in the mutation rules.

At any time a transition is made from the initial state to another with probability α, and a transition from another state to the initial state with probability $\alpha/(N-1)$. Thus we have a two-state chain with diagram



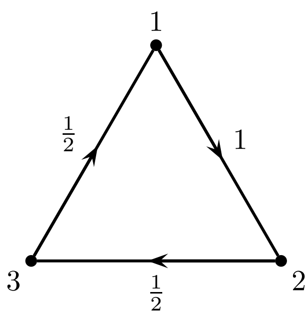
and by putting $\beta = \alpha/(N-1)$ in Example 1.1.4 we find that the desired probability is

$$\frac{1}{N} + \left(1 - \frac{1}{N}\right)\left(1 - \frac{\alpha N}{N-1}\right)^n.$$

Beware that in examples having less symmetry, this sort of lumping together of states may not produce a Markov chain.

Example 1.1.6

Consider the three-state chain with diagram



and transition matrix

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

The problem is to find a general formula for $p_{11}^{(n)}$.

First we compute the eigenvalues of P by writing down its characteristic equation

$$0 = \det(x - P) = x(x - \tfrac{1}{2})^2 - \tfrac{1}{4} = \tfrac{1}{4}(x-1)(4x^2+1).$$

The eigenvalues are $1, i/2, -i/2$ and from this we deduce that $p_{11}^{(n)}$ has the form

$$p_{11}^{(n)} = a + b\left(\frac{i}{2}\right)^n + c\left(-\frac{i}{2}\right)^n$$

for some constants a, b and c. (The justification comes from linear algebra: having distinct eigenvalues, P is diagonalizable, that is, for some invertible matrix U we have

$$P = U \begin{pmatrix} 1 & 0 & 0 \\ 0 & i/2 & 0 \\ 0 & 0 & -i/2 \end{pmatrix} U^{-1}$$

and hence

$$P^n = U \begin{pmatrix} 1 & 0 & 0 \\ 0 & (i/2)^n & 0 \\ 0 & 0 & (-i/2)^n \end{pmatrix} U^{-1}$$

which forces $p_{11}^{(n)}$ to have the form claimed.) The answer we want is real and

$$\left(\pm\frac{i}{2}\right)^n = \left(\frac{1}{2}\right)^n e^{\pm in\pi/2} = \left(\frac{1}{2}\right)^n \left(\cos\frac{n\pi}{2} \pm i\sin\frac{n\pi}{2}\right)$$

so it makes sense to rewrite $p_{11}^{(n)}$ in the form

$$p_{11}^{(n)} = \alpha + \left(\frac{1}{2}\right)^n \left\{\beta\cos\frac{n\pi}{2} + \gamma\sin\frac{n\pi}{2}\right\}$$

for constants α, β and γ. The first few values of $p_{11}^{(n)}$ are easy to write down, so we get equations to solve for α, β and γ:

$$\begin{aligned} 1 &= p_{11}^{(0)} = \alpha + \beta \\ 0 &= p_{11}^{(1)} = \alpha + \tfrac{1}{2}\gamma \\ 0 &= p_{11}^{(2)} = \alpha - \tfrac{1}{4}\beta \end{aligned}$$

so $\alpha = 1/5$, $\beta = 4/5$, $\gamma = -2/5$ and

$$p_{11}^{(n)} = \frac{1}{5} + \left(\frac{1}{2}\right)^n \left\{\frac{4}{5}\cos\frac{n\pi}{2} - \frac{2}{5}\sin\frac{n\pi}{2}\right\}.$$

More generally, the following method may in principle be used to find a formula for $p_{ij}^{(n)}$ for any M-state chain and any states i and j.

(i) Compute the eigenvalues $\lambda_1, \dots, \lambda_M$ of P by solving the characteristic equation.

(ii) If the eigenvalues are distinct then $p_{ij}^{(n)}$ has the form

$$p_{ij}^{(n)} = a_1\lambda_1^n + \ldots + a_M\lambda_M^n$$

for some constants $a_1, \dots, a_M$ (depending on i and j). If an eigenvalue λ is repeated (once, say) then the general form includes the term $(an + b)\lambda^n$.

(iii) As roots of a polynomial with real coefficients, complex eigenvalues will come in conjugate pairs and these are best written using sine and cosine, as in the example.

Exercises

1.1.1 Let $B_1, B_2, \dots$ be disjoint events with $\bigcup_{n=1}^\infty B_n = \Omega$. Show that if A is another event and $\mathbb{P}(A|B_n) = p$ for all n then $\mathbb{P}(A) = p$.

Deduce that if X and Y are discrete random variables then the following are equivalent:

(a) X and Y are independent;

(b) the conditional distribution of X given $Y = y$ is independent of y.

1.1.2 Suppose that $(X_n)_{n\geq 0}$ is Markov (λ, P). If $Y_n = X_{kn}$, show that $(Y_n)_{n\geq 0}$ is Markov (λ, P^k).

1.1.3 Let X_0 be a random variable with values in a countable set I. Let $Y_1, Y_2, \dots$ be a sequence of independent random variables, uniformly distributed on $[0, 1]$. Suppose we are given a function

$$G : I \times [0, 1] \to I$$

and define inductively

$$X_{n+1} = G(X_n, Y_{n+1}).$$

Show that $(X_n)_{n\geq 0}$ is a Markov chain and express its transition matrix P in terms of G. Can all Markov chains be realized in this way? How would you simulate a Markov chain using a computer?

Suppose now that $Z_0, Z_1, \ldots$ are independent, identically distributed random variables such that $Z_i = 1$ with probability p and $Z_i = 0$ with probability $1 - p$. Set $S_0 = 0$, $S_n = Z_1 + \ldots + Z_n$. In each of the following cases determine whether $(X_n)_{n \geq 0}$ is a Markov chain:

(a) $X_n = Z_n$, (b) $X_n = S_n$,
(c) $X_n = S_0 + \ldots + S_n$, (d)$X_n = (S_n, S_0 + \ldots + S_n)$.

In the cases where $(X_n)_{n \geq 0}$ is a Markov chain find its state-space and transition matrix, and in the cases where it is not a Markov chain give an example where $P(X_{n+1} = i | X_n = j, X_{n-1} = k)$ is not independent of k.

1.1.4 A flea hops about at random on the vertices of a triangle, with all jumps equally likely. Find the probability that after n hops the flea is back where it started.

A second flea also hops about on the vertices of a triangle, but this flea is twice as likely to jump clockwise as anticlockwise. What is the probability that after n hops this second flea is back where it started? [*Recall that* $e^{\pm i\pi/6} = \sqrt{3}/2 \pm i/2$.]

1.1.5 A die is 'fixed' so that each time it is rolled the score cannot be the same as the preceding score, all other scores having probability 1/5. If the first score is 6, what is the probability p that the nth score is 6? What is the probability that the nth score is 1?

Suppose now that a new die is produced which cannot score one greater (mod 6) than the preceding score, all other scores having equal probability. By considering the relationship between the two dice find the value of p for the new die.

1.1.6 An octopus is trained to choose object A from a pair of objects A, B by being given repeated trials in which it is shown both and is rewarded with food if it chooses A. The octopus may be in one of three states of mind: in state 1 it cannot remember which object is rewarded and is equally likely to choose either; in state 2 it remembers and chooses A but may forget again; in state 3 it remembers and chooses A and never forgets. After each trial it may change its state of mind according to the transition matrix

$$\begin{array}{cccc} \text{State 1} & \frac{1}{2} & \frac{1}{2} & 0 \\ \text{State 2} & \frac{1}{2} & \frac{1}{12} & \frac{5}{12} \\ \text{State 3} & 0 & 0 & 1. \end{array}$$

It is in state 1 before the first trial. What is the probablity that it is in state 1 just before the $(n+1)$th trial ? What is the probability $P_{n+1}(A)$ that it chooses A on the $(n+1)$th trial ?

Someone suggests that the record of successive choices (a sequence of As and Bs) might arise from a two-state Markov chain with constant transition probabilities. Discuss, with reference to the value of $P_{n+1}(A)$ that you have found, whether this is possible.

1.1.7 Let $(X_n)_{n\geq 0}$ be a Markov chain on $\{1,2,3\}$ with transition matrix

$$P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 2/3 & 1/3 \\ p & 1-p & 0 \end{pmatrix}.$$

Calculate $\mathbb{P}(X_n = 1 | X_0 = 1)$ in each of the following cases: (a) $p = 1/16$, (b) $p = 1/6$, (c) $p = 1/12$.

1.2 Class structure

It is sometimes possible to break a Markov chain into smaller pieces, each of which is relatively easy to understand, and which together give an understanding of the whole. This is done by identifying the communicating classes of the chain.

We say that i *leads to* j and write $i \to j$ if

$$\mathbb{P}_i(X_n = j \text{ for some } n \geq 0) > 0.$$

We say i *communicates with* j and write $i \leftrightarrow j$ if both $i \to j$ and $j \to i$.

Theorem 1.2.1. *For distinct states i and j the following are equivalent:*

(i) $i \to j$;

(ii) $p_{i_1 i_2} p_{i_2 i_2} \dots p_{i_{n-1} i_n} > 0$ *for some states* $i_1, i_2, \dots, i_n$ *with* $i_1 = i$ *and* $i_n = j$;

(iii) $p_{ij}^{(n)} > 0$ *for some* $n \geq 0$.

Proof. Observe that

$$p_{ij}^{(n)} \leq \mathbb{P}_i(X_n = j \text{ for some } n \geq 0) \leq \sum_{n=0}^{\infty} p_{ij}^{(n)}$$

which proves the equivalence of (i) and (iii). Also

$$p_{ij}^{(n)} = \sum_{i_2, \dots, i_{n-1}} p_{i i_2} p_{i_2 i_2} \dots p_{i_{n-1} j}$$

so that (ii) and (iii) are equivalent. $\square$

It is clear from (ii) that $i \to j$ and $j \to k$ imply $i \to k$. Also $i \to i$ for any state i. So $\leftrightarrow$ satisfies the conditions for an equivalence relation on I, and thus partitions I into *communicating classes*. We say that a class C is *closed* if

$$i \in C, i \to j \quad \text{imply } j \in C.$$

Thus a closed class is one from which there is no escape. A state i is *absorbing* if $\{i\}$ is a closed class. The smaller pieces referred to above are these communicating classes. A chain or transition matrix P where I is a single class is called *irreducible*.

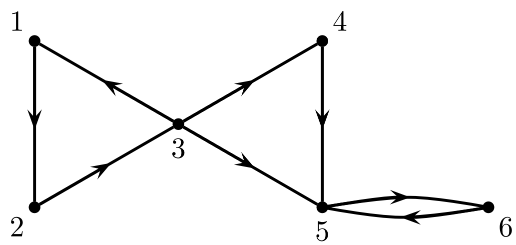
As the following example makes clear, when one can draw the diagram, the class structure of a chain is very easy to find.

Example 1.2.2

Find the communicating classes associated to the stochastic matrix

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

The solution is obvious from the diagram



the classes being $\{1, 2, 3\}$, $\{4\}$ and $\{5, 6\}$, with only $\{5, 6\}$ being closed.

Exercises

1.2.1 Identify the communicating classes of the following transition matrix:

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

Which classes are closed?

1.2.2 Show that every transition matrix on a finite state-space has at least one closed communicating class. Find an example of a transition matrix with no closed communicating class.

1.3 Hitting times and absorption probabilities

Let $(X_n)_{n\geq 0}$ be a Markov chain with transition matrix P. The *hitting time* of a subset A of I is the random variable $H^A : \Omega \to \{0, 1, 2, \ldots\} \cup \{\infty\}$ given by

$$H^A(\omega) = \inf\{n \geq 0 : X_n(\omega) \in A\}$$

where we agree that the infimum of the empty set $\emptyset$ is ∞. The probability starting from i that $(X_n)_{n\geq 0}$ ever hits A is then

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

When A is a closed class, h_i^A is called the *absorption probability*. The mean time taken for $(X_n)_{n\geq 0}$ to reach A is given by

$$k_i^A = \mathbb{E}_i(H^A) = \sum_{n<\infty} n\mathbb{P}(H^A = n) + \infty\mathbb{P}(H^A = \infty).$$

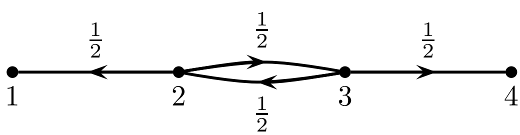
We shall often write less formally

$$h_i^A = \mathbb{P}_i(\text{hit } A), \quad k_i^A = \mathbb{E}_i(\text{time to hit } A).$$

Remarkably, these quantities can be calculated explicitly by means of certain linear equations associated with the transition matrix P. Before we give the general theory, here is a simple example.

Example 1.3.1

Consider the chain with the following diagram:



Starting from 2, what is the probability of absorption in 4? How long does it take until the chain is absorbed in 1 or 4?

Introduce

$$h_i = \mathbb{P}_i(\text{hit } 4), \quad k_i = \mathbb{E}_i(\text{time to hit } \{1, 4\}).$$

Clearly, $h_1 = 0$, $h_4 = 1$ and $k_1 = k_4 = 0$. Suppose now that we start at 2, and consider the situation after making one step. With probability 1/2 we jump to 1 and with probability 1/2 we jump to 3. So

$$h_2 = \tfrac{1}{2}h_1 + \tfrac{1}{2}h_3, \quad k_2 = 1 + \tfrac{1}{2}k_1 + \tfrac{1}{2}k_3.$$

The 1 appears in the second formula because we count the time for the first step. Similarly,

$$h_3 = \tfrac{1}{2}h_2 + \tfrac{1}{2}h_4, \quad k_3 = 1 + \tfrac{1}{2}k_2 + \tfrac{1}{2}k_4.$$

Hence

$$\begin{aligned} h_2 &= \tfrac{1}{2}h_3 = \tfrac{1}{2}(\tfrac{1}{2}h_2 + \tfrac{1}{2}), \\ k_2 &= 1 + \tfrac{1}{2}k_3 = 1 + \tfrac{1}{2}(1 + \tfrac{1}{2}k_2). \end{aligned}$$

So, starting from 2, the probability of hitting 4 is 1/3 and the mean time to absorption is 2. Note that in writing down the first equations for h_2 and k_2 we made implicit use of the Markov property, in assuming that the chain begins afresh from its new position after the first jump. Here is a general result for hitting probabilities.

Theorem 1.3.2. *The vector of hitting probabilities $h^A = (h_i^A : i \in I)$ is the minimal non-negative solution to the system of linear equations*

$$\begin{cases} h_i^A = 1 & \text{for } i \in A \\ h_i^A = \sum_{j \in I} p_{ij} h_j^A & \text{for } i \notin A. \end{cases} \tag{1.3}$$

(Minimality means that if $x = (x_i : i \in I)$ is another solution with $x_i \geq 0$ for all i, then $x_i \geq h_i$ for all i.)

Proof. First we show that h^A satisfies (1.3). If $X_0 = i \in A$, then $H^A = 0$, so $h_i^A = 1$. If $X_0 = i \notin A$, then $H^A \geq 1$, so by the Markov property

$$\mathbb{P}_i(H^A < \infty \mid X_1 = j) = \mathbb{P}_j(H^A < \infty) = h_j^A$$

and

$$\begin{aligned} h_i^A = \mathbb{P}_i(H^A < \infty) &= \sum_{j \in I} \mathbb{P}_i(H^A < \infty, X_1 = j) \\ &= \sum_{j \in I} \mathbb{P}_i(H^A < \infty \mid X_1 = j)\mathbb{P}_i(X_1 = j) = \sum_{j \in I} p_{ij} h_j^A. \end{aligned}$$

Suppose now that $x = (x_i : i \in I)$ is any solution to (1.3). Then $h_i^A = x_i = 1$ for $i \in A$. Suppose $i \notin A$, then

$$x_i = \sum_{j \in I} p_{ij} x_j = \sum_{j \in A} p_{ij} + \sum_{j \notin A} p_{ij} x_j.$$

Substitute for x_j to obtain

$$\begin{aligned} x_i &= \sum_{j \in A} p_{ij} + \sum_{j \notin A} p_{ij} \Big(\sum_{k \in A} p_{jk} + \sum_{k \notin A} p_{jk} x_k \Big) \\ &= \mathbb{P}_i(X_1 \in A) + \mathbb{P}_i(X_1 \notin A, X_2 \in A) + \sum_{j \notin A} \sum_{k \notin A} p_{ij} p_{jk} x_k. \end{aligned}$$

By repeated substitution for x in the final term we obtain after n steps

$$\begin{aligned} x_i = \mathbb{P}_i(X_1 \in A) + \ldots + \mathbb{P}_i(X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A) \\ + \sum_{j_1 \notin A} \cdots \sum_{j_n \notin A} p_{ij_1} p_{j_1 j_2} \ldots p_{j_{n-1} j_n} x_{j_n}. \end{aligned}$$

Now if x is non-negative, so is the last term on the right, and the remaining terms sum to $\mathbb{P}_i(H^A \le n)$. So $x_i \ge \mathbb{P}_i(H^A \le n)$ for all n and then

$$x_i \ge \lim_{n \to \infty} \mathbb{P}_i(H^A \le n) = \mathbb{P}_i(H^A < \infty) = h_i. \qquad \square$$

Example 1.3.1 (continued)

The system of linear equations (1.3) for $h = h^{\{4\}}$ are given here by

$$\begin{aligned} &h_4 = 1, \\ &h_2 = \tfrac{1}{2} h_1 + \tfrac{1}{2} h_3, \ h_3 = \tfrac{1}{2} h_2 + \tfrac{1}{2} h_4 \end{aligned}$$

so that

$$h_2 = \tfrac{1}{2} h_1 + \tfrac{1}{2}(\tfrac{1}{2} h_2 + \tfrac{1}{2})$$

and

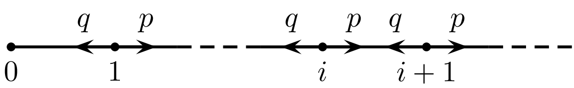
$$h_2 = \tfrac{1}{3} + \tfrac{2}{3} h_1, \ h_3 = \tfrac{2}{3} + \tfrac{1}{3} h_1.$$

The value of h_1 is not determined by the system (1.3), but the minimality condition now makes us take $h_1 = 0$, so we recover $h_2 = 1/3$ as before. Of course, the extra boundary condition $h_1 = 0$ was obvious from the beginning so we built it into our system of equations and did not have to worry about minimal non-negative solutions.

In cases where the state-space is infinite it may not be possible to write down a corresponding extra boundary condition. Then, as we shall see in the next examples, the minimality condition is essential.

Example 1.3.3 (Gamblers' ruin)

Consider the Markov chain with diagram



where $0 < p = 1 - q < 1$. The transition probabilities are

$$p_{00} = 1,$$
$$p_{i,i-1} = q,\ p_{i,i+1} = p \quad \text{for } i = 1, 2, \ldots.$$

Imagine that you enter a casino with a fortune of £i and gamble, £1 at a time, with probability p of doubling your stake and probability q of losing it. The resources of the casino are regarded as infinite, so there is no upper limit to your fortune. But what is the probability that you leave broke?

Set $h_i = \mathbb{P}_i(\text{hit } 0)$, then h is the minimal non-negative solution to

$$h_0 = 1,$$
$$h_i = ph_{i+1} + qh_{i-1}, \quad \text{for } i = 1, 2, \ldots.$$

If $p \neq q$ this recurrence relation has a general solution

$$h_i = A + B\left(\frac{q}{p}\right)^i.$$

(See Section 1.11.) If $p < q$, which is the case in most successful casinos, then the restriction $0 \leq h_i \leq 1$ forces $B = 0$, so $h_i = 1$ for all i. If $p > q$, then since $h_0 = 1$ we get a family of solutions

$$h_i = \left(\frac{q}{p}\right)^i + A\left(1 - \left(\frac{q}{p}\right)^i\right);$$

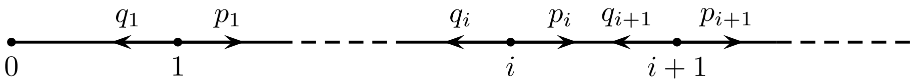
for a non-negative solution we must have $A \geq 0$, so the minimal non-negative solution is $h_i = (q/p)^i$. Finally, if $p = q$ the recurrence relation has a general solution

$$h_i = A + Bi$$

and again the restriction $0 \leq h_i \leq 1$ forces $B = 0$, so $h_i = 1$ for all i. Thus, even if you find a fair casino, you are certain to end up broke. This apparent paradox is called gamblers' ruin.

Example 1.3.4 (Birth-and-death chain)

Consider the Markov chain with diagram



where, for $i = 1, 2, \ldots$, we have $0 < p_i = 1 - q_i < 1$. As in the preceding example, 0 is an absorbing state and we wish to calculate the absorption probability starting from i. But here we allow p_i and q_i to depend on i.

Such a chain may serve as a model for the size of a population, recorded each time it changes, p_i being the probability that we get a birth before a death in a population of size i. Then $h_i = \mathbb{P}_i(\text{hit } 0)$ is the extinction probability starting from i.

We write down the usual system of equations

$$\begin{aligned} h_0 &= 1, \\ h_i &= p_i h_{i+1} + q_i h_{i-1}, \quad \text{for } i = 1, 2, \ldots . \end{aligned}$$

This recurrence relation has variable coefficients so the usual technique fails. But consider $u_i = h_{i-1} - h_i$, then $p_i u_{i+1} = q_i u_i$, so

$$u_{i+1} = \left(\frac{q_i}{p_i}\right) u_i = \left(\frac{q_i q_{i-1} \ldots q_1}{p_i p_{i-1} \ldots p_1}\right) u_1 = \gamma_i u_1$$

where the final equality defines γ_i. Then

$$u_1 + \ldots + u_i = h_0 - h_i$$

so

$$h_i = 1 - A(\gamma_0 + \ldots + \gamma_{i-1})$$

where $A = u_1$ and $\gamma_0 = 1$. At this point A remains to be determined. In the case $\sum_{i=0}^{\infty} \gamma_i = \infty$, the restriction $0 \leq h_i \leq 1$ forces $A = 0$ and $h_i = 1$ for all i. But if $\sum_{i=0}^{\infty} \gamma_i < \infty$ then we can take $A > 0$ so long as

$$1 - A(\gamma_0 + \ldots + \gamma_{i-1}) \geq 0 \quad \text{for all } i.$$

Thus the minimal non-negative solution occurs when $A = \left(\sum_{i=0}^{\infty} \gamma_i\right)^{-1}$ and then

$$h_i = \sum_{j=i}^{\infty} \gamma_j \Big/ \sum_{j=0}^{\infty} \gamma_j .$$

In this case, for $i = 1, 2, \dots$, we have $h_i < 1$, so the population survives with positive probability.

Here is the general result on mean hitting times. Recall that $k_i^A = \mathbb{E}_i(H^A)$, where H^A is the first time $(X_n)_{n\geq 0}$ hits A. We use the notation 1_B for the indicator function of B, so, for example, $1_{X_1=j}$ is the random variable equal to 1 if $X_1 = j$ and equal to 0 otherwise.

Theorem 1.3.5. *The vector of mean hitting times $k^A = (k^A : i \in I)$ is the minimal non-negative solution to the system of linear equations*

$$\begin{cases} k_i^A = 0 & \text{for } i \in A \\ k_i^A = 1 + \sum_{j \notin A} p_{ij} k_j^A & \text{for } i \notin A. \end{cases} \tag{1.4}$$

Proof. First we show that k^A satisfies (1.4). If $X_0 = i \in A$, then $H^A = 0$, so $k_i^A = 0$. If $X_0 = i \notin A$, then $H^A \geq 1$, so, by the Markov property,

$$\mathbb{E}_i(H^A \mid X_1 = j) = 1 + \mathbb{E}_j(H^A)$$

and

$$\begin{aligned} k_i^A = \mathbb{E}_i(H^A) &= \sum_{j \in I} \mathbb{E}_i(H^A 1_{X_1=j}) \\ &= \sum_{j \in I} \mathbb{E}_i(H^A \mid X_1 = j)\mathbb{P}_i(X_1 = j) = 1 + \sum_{j \notin A} p_{ij} k_j^A . \end{aligned}$$

Suppose now that $y = (y_i : i \in I)$ is any solution to (1.4). Then $k_i^A = y_i = 0$ for $i \in A$. If $i \notin A$, then

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

By repeated substitution for y in the final term we obtain after n steps

$$y_i = \mathbb{P}_i(H^A \geq 1) + \dots + \mathbb{P}_i(H^A \geq n) + \sum_{j_1 \notin A} \dots \sum_{j_n \notin A} p_{ij_1} p_{j_1 j_2} \dots p_{j_{n-1} j_n} y_{j_n} .$$

So, if y is non-negative,

$$y_i \geq \mathbb{P}_i(H^A \geq 1) + \ldots + \mathbb{P}_i(H^A \geq n)$$

and, letting $n \to \infty$,

$$y_i \geq \sum_{n=1}^{\infty} \mathbb{P}_i(H^A \geq n) = \mathbb{E}_i(H^A) = k_i^A. \qquad \square$$

Exercises

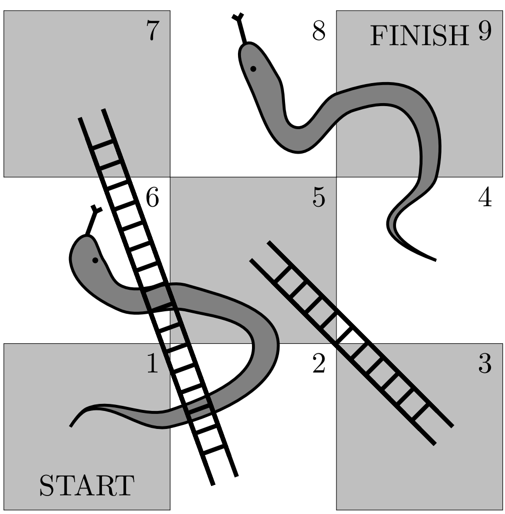
1.3.1 Prove the claims (a), (b) and (c) made in example (v) of the Introduction.

1.3.2 A gambler has £2 and needs to increase it to £10 in a hurry. He can play a game with the following rules: a fair coin is tossed; if a player bets on the right side, he wins a sum equal to his stake, and his stake is returned; otherwise he loses his stake. The gambler decides to use a bold strategy in which he stakes all his money if he has £5 or less, and otherwise stakes just enough to increase his capital, if he wins, to £10.

Let $X_0 = 2$ and let X_n be his capital after n throws. Prove that the gambler will achieve his aim with probability 1/5.

What is the expected number of tosses until the gambler either achieves his aim or loses his capital?

1.3.3 A simple game of 'snakes and ladders' is played on a board of nine squares.



At each turn a player tosses a fair coin and advances one or two places according to whether the coin lands heads or tails. If you land at the foot of a ladder you climb to the top, but if you land at the head of a snake you slide down to the tail. How many turns on average does it take to complete the game?

What is the probability that a player who has reached the middle square will complete the game without slipping back to square 1?

1.3.4 Let $(X_n)_{n\geq 0}$ be a Markov chain on $\{0, 1, \dots\}$ with transition probabilities given by

$$p_{01} = 1, \quad p_{i,i+1} + p_{i,i-1} = 1, \quad p_{i,i+1} = \left(\frac{i+1}{i}\right)^2 p_{i,i-1}, \quad i \geq 1\,.$$

Show that if $X_0 = 0$ then the probability that $X_n \geq 1$ for all $n \geq 1$ is $6/\pi^2$.

1.4 Strong Markov property

In Section 1.1 we proved the Markov property. This says that for each time m, conditional on $X_m = i$, the process after time m begins afresh from i. Suppose, instead of conditioning on $X_m = i$, we simply waited for the process to hit state i, at some random time H. What can one say about the process after time H? What if we replaced H by a more general random time, for example $H - 1$? In this section we shall identify a class of random times at which a version of the Markov property does hold. This class will include H but not $H - 1$; after all, the process after time $H - 1$ jumps straight to i, so it does not simply begin afresh.

A random variable $T : \Omega \to \{0, 1, 2, \dots\} \cup \{\infty\}$ is called a *stopping time* if the event $\{T = n\}$ depends only on $X_0, X_1, \dots, X_n$ for $n = 0, 1, 2, \dots$. Intuitively, by watching the process, you know at the time when T occurs. If asked to stop at T, you know when to stop.

Examples 1.4.1

(a) The *first passage time*

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

is a stopping time because

$$\{T_j = n\} = \{X_1 \neq j, \dots, X_{n-1} \neq j, X_n = j\}.$$

(b) The first hitting time H^A of Section 1.3 is a stopping time because

$$\{H^A = n\} = \{X_0 \notin A, \dots, X_{n-1} \notin A, X_n \in A\}.$$

(c) The *last exit time*

$$L^A = \sup\{n \geq 0 : X_n \in A\}$$

is not in general a stopping time because the event $\{L^A = n\}$ depends on whether $(X_{n+m})_{m\geq 1}$ visits A or not.

We shall show that the Markov property holds at stopping times. The crucial point is that, if T is a stopping time and $B \subseteq \Omega$ is determined by $X_0, X_1, \dots, X_T$, then $B \cap \{T = m\}$ is determined by $X_0, X_1, \dots, X_m$, for all $m = 0, 1, 2, \dots$.

Theorem 1.4.2 (Strong Markov property). *Let $(X_n)_{n\geq 0}$ be Markov(λ, P) and let T be a stopping time of $(X_n)_{n\geq 0}$. Then, conditional on $T < \infty$ and $X_T = i$, $(X_{T+n})_{n\geq 0}$ is Markov(δ_i, P) and independent of $X_0, X_1, \dots, X_T$.*

Proof. If B is an event determined by $X_0, X_1, \dots, X_T$, then $B \cap \{T = m\}$ is determined by $X_0, X_1, \dots, X_m$, so, by the Markov property at time m

$$\begin{aligned} &\mathbb{P}(\{X_T = j_0, X_{T+1} = j_1, \dots, X_{T+n} = j_n\} \cap B \cap \{T = m\} \cap \{X_T = i\}) \\ &\quad = \mathbb{P}_i(X_0 = j_0, X_1 = j_1, \dots, X_n = j_n)\mathbb{P}(B \cap \{T = m\} \cap \{X_T = i\}) \end{aligned}$$

where we have used the condition $T = m$ to replace m by T. Now sum over $m = 0, 1, 2, \dots$ and divide by $\mathbb{P}(T < \infty, X_T = i)$ to obtain

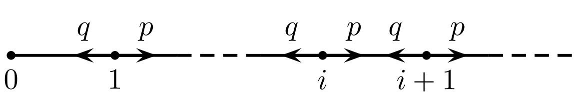
$$\begin{aligned} &\mathbb{P}(\{X_T = j_0, X_{T+1} = j_1, \dots, X_{T+n} = j_n\} \cap B \mid T < \infty, X_T = i) \\ &\quad = \mathbb{P}_i(X_0 = j_0, X_1 = j_1, \dots, X_n = j_n)\mathbb{P}(B \mid T < \infty, X_T = i). \qquad \square \end{aligned}$$

skip this

The following example uses the strong Markov property to get more information on the hitting times of the chain considered in Example 1.3.3.

Example 1.4.3

Consider the Markov chain $(X_n)_{n\geq 0}$ with diagram



where $0 < p = 1 - q < 1$. We know from Example 1.3.3 the probability of hitting 0 starting from 1. Here we obtain the complete distribution of the time to hit 0 starting from 1 in terms of its probability generating function. Set

$$H_j = \inf\{n \geq 0 : X_n = j\}$$

and, for $0 \leq s < 1$

$$\phi(s) = \mathbb{E}_1(s^{H_0}) = \sum_{n<\infty} s^n \mathbb{P}_1(H_0 = n).$$

Suppose we start at 2. Apply the strong Markov property at H_1 to see that under $\mathbb{P}_2$, conditional on $H_1 < \infty$, we have $H_0 = H_1 + \widetilde{H}_0$, where $\widetilde{H}_0$, the time taken after H_1 to get to 0, is independent of H_1 and has the (unconditioned) distribution of H_1. So

$$\begin{aligned}
\mathbb{E}_2(s^{H_0}) &= \mathbb{E}_2(s^{H_1} \mid H_1 < \infty)\mathbb{E}_2(s^{\widetilde{H}_0} \mid H_1 < \infty)\mathbb{P}_2(H_1 < \infty) \\
&= \mathbb{E}_2(s^{H_1} 1_{H_1<\infty})\mathbb{E}_2(s^{\widetilde{H}_0} \mid H_1 < \infty) \\
&= \mathbb{E}_2(s^{H_1})^2 = \phi(s)^2.
\end{aligned}$$

Then, by the Markov property at time 1, conditional on $X_1 = 2$, we have $H_0 = 1 + \overline{H}_0$, where $\overline{H}_0$, the time taken after time 1 to get to 0, has the same distribution as H_0 does under $\mathbb{P}_2$. So

$$\begin{aligned}
\phi(s) = \mathbb{E}_1(s^{H_0}) &= p\mathbb{E}_1(s^{H_0} \mid X_1 = 2) + q\mathbb{E}_1(s^{H_0} \mid X_1 = 0) \\
&= p\mathbb{E}_1(s^{1+\overline{H}_0} \mid X_1 = 2) + q\mathbb{E}_1(s \mid X_1 = 0) \\
&= ps\mathbb{E}_2(s^{H_0}) + qs \\
&= ps\phi(s)^2 + qs.
\end{aligned}$$

Thus $\phi = \phi(s)$ satisfies

$$ps\phi^2 - \phi + qs = 0 \tag{1.5}$$

and

$$\phi = (1 \pm \sqrt{1 - 4pqs^2})/2ps.$$

Since $\phi(0) \leq 1$ and ϕ is continuous we are forced to take the negative root at $s = 0$ and stick with it for all $0 \leq s < 1$.

To recover the distribution of H_0 we expand the square-root as a power series:

$$\begin{aligned}
\phi(s) &= \frac{1}{2ps}\left\{1 - \left(1 + \tfrac{1}{2}(-4pqs^2) + \tfrac{1}{2}(-\tfrac{1}{2})(-4pqs^2)^2/2! + \dots\right)\right\} \\
&= qs + pq^2s^3 + \dots \\
&= s\mathbb{P}_1(H_0 = 1) + s^2\mathbb{P}_1(H_0 = 2) + s^3\mathbb{P}_1(H_0 = 3) + \dots .
\end{aligned}$$

The first few probabilities $\mathbb{P}_1(H_0 = 1), \mathbb{P}_1(H_0 = 2), \ldots$ are readily checked from first principles.

On letting $s \uparrow 1$ we have $\phi(s) \to \mathbb{P}_1(H_0 < \infty)$, so

$$\mathbb{P}_1(H_0 < \infty) = \frac{1 - \sqrt{1 - 4pq}}{2p} = \begin{cases} 1 & \text{if } p \leq q \\ q/p & \text{if } p > q. \end{cases}$$

(Remember that $q = 1 - p$, so

$$\sqrt{1 - 4pq} = \sqrt{1 - 4p + 4p^2} = |1 - 2p| = |2q - 1|.)$$

We can also find the mean hitting time using

$$\mathbb{E}_1(H_0) = \lim_{s \uparrow 1} \phi'(s).$$

It is only worth considering the case $p \leq q$, where the mean hitting time has a chance of being finite. Differentiate (1.5) to obtain

$$2ps\phi\phi' + p\phi^2 - \phi' + q = 0$$

so

$$\phi'(s) = (p\phi(s)^2 + q)/(1 - 2ps\phi(s)) \to 1/(1 - 2p) = 1/(q - p) \quad \text{as } s \uparrow 1.$$

See Example 5.1.1 for a connection with branching processes.

Example 1.4.4

We now consider an application of the strong Markov property to a Markov chain $(X_n)_{n \geq 0}$ observed only at certain times. In the first instance suppose that J is some subset of the state-space I and that we observe the chain only when it takes values in J. The resulting process $(Y_m)_{m \geq 0}$ may be obtained formally by setting $Y_m = X_{T_m}$, where

$$T_0 = \inf\{n \geq 0 : X_n \in J\}$$

and, for $m = 0, 1, 2, \ldots$

$$T_{m+1} = \inf\{n > T_m : X_n \in J\}.$$

Let us assume that $\mathbb{P}(T_m < \infty) = 1$ for all m. For each m we can check easily that T_m, the time of the mth visit to J, is a stopping time. So the strong Markov property applies to show, for $i_1, \ldots, i_{m+1} \in J$, that

$$\begin{aligned} &\mathbb{P}(Y_{m+1} = i_{m+1} \mid Y_0 = i_1, \ldots, Y_m = i_m) \\ &\quad = \mathbb{P}(X_{T_{m+1}} = i_{m+1} \mid X_{T_0} = i_1, \ldots, X_{T_m} = i_m) \\ &\quad = \mathbb{P}_{i_m}(X_{T_1} = i_{m+1}) = \overline{p}_{i_m i_{m+1}} \end{aligned}$$

where, for $i, j \in J$

$$\overline{p}_{ij} = h_i^j$$

and where, for $j \in J$, the vector $(h_i^j : i \in I)$ is the minimal non-negative solution to

$$h_i^j = p_{ij} + \sum_{k \notin J} p_{ik} h_k^j. \tag{1.6}$$

Thus $(Y_m)_{m\geq 0}$ is a Markov chain on J with transition matrix $\overline{P}$.

A second example of a similar type arises if we observe the original chain $(X_n)_{n\geq 0}$ only when it moves. The resulting process $(Z_m)_{m\geq 0}$ is given by $Z_m = X_{S_m}$ where $S_0 = 0$ and for $m = 0, 1, 2, \dots$

$$S_{m+1} = \inf\{n \geq S_m : X_n \neq X_{S_m}\}.$$

Let us assume there are no absorbing states. Again the random times S_m for $m \geq 0$ are stopping times and, by the strong Markov property

$$\begin{aligned}
&\mathbb{P}(Z_{m+1} = i_{m+1} \mid Z_0 = i_1, \dots, Z_m = i_m) \\
&\quad = \mathbb{P}(X_{S_{m+1}} = i_{m+1} \mid X_{S_0} = i_1, \dots, X_{S_m} = i_m) \\
&\quad = \mathbb{P}_{i_m}(X_{S_1} = i_{m+1}) = \widetilde{p}_{i_m i_{m+1}}
\end{aligned}$$

where $\widetilde{p}_{ii} = 0$ and, for $i \neq j$

$$\widetilde{p}_{ij} = p_{ij} / \sum_{k \neq i} p_{ik}.$$

Thus $(Z_m)_{m\geq 0}$ is a Markov chain on I with transition matrix $\widetilde{P}$.

Exercises

1.4.1 Let $Y_1, Y_2, \dots$ be independent identically distributed random variables with
$\mathbb{P}(Y_1 = 1) = \mathbb{P}(Y_1 = -1) = 1/2$ and set $X_0 = 1$, $X_n = X_0 + Y_1 + \dots + Y_n$ for $n \geq 1$. Define

$$H_0 = \inf\{n \geq 0 : X_n = 0\}.$$

Find the probability generating function $\phi(s) = \mathbb{E}(s^{H_0})$.

Suppose the distribution of $Y_1, Y_2, \dots$ is changed to $\mathbb{P}(Y_1 = 2) = 1/2$, $\mathbb{P}(Y_1 = -1) = 1/2$. Show that ϕ now satisfies

$$s\phi^3 - 2\phi + s = 0.$$

1.4.2 Deduce carefully from Theorem 1.3.2 the claim made at (1.6).

1.5 Recurrence and transience

Let $(X_n)_{n\geq 0}$ be a Markov chain with transition matrix P. We say that a state i is *recurrent* if

$$\mathbb{P}_i(X_n = i \text{ for infinitely many } n) = 1.$$

We say that i is *transient* if

$$\mathbb{P}_i(X_n = i \text{ for infinitely many } n) = 0.$$

Thus a recurrent state is one to which you keep coming back and a transient state is one which you eventually leave for ever. We shall show that every state is either recurrent or transient.

Recall that the *first passage time* to state i is the random variable T_i defined by

$$T_i(\omega) = \inf\{n \geq 1 : X_n(\omega) = i\}$$

where $\inf \emptyset = \infty$. We now define inductively the *rth passage time* $T_i^{(r)}$ to state i by

$$T_i^{(0)}(\omega) = 0, \quad T_i^{(1)}(\omega) = T_i(\omega)$$

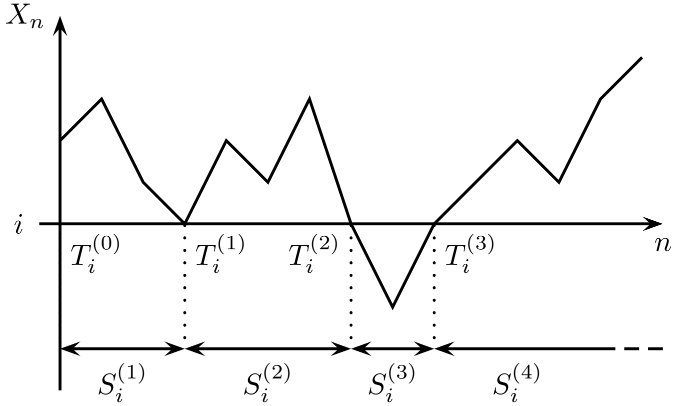
and, for $r = 0, 1, 2, \dots,$

$$T_i^{(r+1)}(\omega) = \inf\{n \geq T_i^{(r)}(\omega) + 1 : X_n(\omega) = i\}.$$

The length of the rth excursion to i is then

$$S_i^{(r)} = \begin{cases} T_i^{(r)} - T_i^{(r-1)} & \text{if } T_i^{(r-1)} < \infty \\ 0 & \text{otherwise.} \end{cases}$$

The following diagram illustrates these definitions:



Our analysis of recurrence and transience will rest on finding the joint distribution of these excursion lengths.

Lemma 1.5.1. *For $r = 2, 3, \dots$, conditional on $T_i^{(r-1)} < \infty$, $S_i^{(r)}$ is independent of $\{X_m : m \le T_i^{(r-1)}\}$ and*

$$\mathbb{P}(S_i^{(r)} = n \mid T_i^{(r-1)} < \infty) = \mathbb{P}_i(T_i = n).$$

Proof. Apply the strong Markov property at the stopping time $T = T_i^{(r-1)}$. It is automatic that $X_T = i$ on $T < \infty$. So, conditional on $T < \infty$, $(X_{T+n})_{n\ge 0}$ is Markov(δ_i, P) and independent of $X_0, X_1, \dots, X_T$. But

$$S_i^{(r)} = \inf\{n \ge 1 : X_{T+n} = i\},$$

so $S_i^{(r)}$ is the first passage time of $(X_{T+n})_{n\ge 0}$ to state i. $\square$

Recall that the indicator function $1_{\{X_1=j\}}$ is the random variable equal to 1 if $X_1 = j$ and 0 otherwise. Let us introduce the *number of visits* V_i to i, which may be written in terms of indicator functions as

$$V_i = \sum_{n=0}^{\infty} 1_{\{X_n=i\}}$$

and note that

$$\mathbb{E}_i(V_i) = \mathbb{E}_i \sum_{n=0}^{\infty} 1_{\{X_n=i\}} = \sum_{n=0}^{\infty} \mathbb{E}_i(1_{\{X_n=i\}}) = \sum_{n=0}^{\infty} \mathbb{P}_i(X_n = i) = \sum_{n=0}^{\infty} p_{ii}^{(n)}.$$

Also, we can compute the distribution of V_i under $\mathbb{P}_i$ in terms of the *return probability*

$$f_i = \mathbb{P}_i(T_i < \infty).$$

Lemma 1.5.2. *For $r = 0, 1, 2, \dots$, we have $\mathbb{P}_i(V_i > r) = f_i^r$.*

Proof. Observe that if $X_0 = i$ then $\{V_i > r\} = \{T_i^{(r)} < \infty\}$. When $r = 0$ the result is true. Suppose inductively that it is true for r, then

$$\begin{aligned}
\mathbb{P}_i(V_i > r+1) &= \mathbb{P}_i(T_i^{(r+1)} < \infty) \\
&= \mathbb{P}_i(T_i^{(r)} < \infty \text{ and } S_i^{(r+1)} < \infty) \\
&= \mathbb{P}_i(S_i^{(r+1)} < \infty \mid T_i^{(r)} < \infty)\mathbb{P}_i(T_i^{(r)} < \infty) \\
&= f_i f_i^r = f_i^{r+1}
\end{aligned}$$

by Lemma 1.5.1, so by induction the result is true for all r. $\square$

Recall that one can compute the expectation of a non-negative integer-valued random variable as follows:

$$\sum_{r=0}^{\infty} \mathbb{P}(V > r) = \sum_{r=0}^{\infty} \sum_{v=r+1}^{\infty} \mathbb{P}(V = v)$$
$$= \sum_{v=1}^{\infty} \sum_{r=0}^{v-1} \mathbb{P}(V = v) = \sum_{v=1}^{\infty} v\mathbb{P}(V = v) = \mathbb{E}(V).$$

The next theorem is the means by which we establish recurrence or transience for a given state. Note that it provides two criteria for this, one in terms of the return probability, the other in terms of the n-step transition probabilities. Both are useful.

Theorem 1.5.3. *The following dichotomy holds:*
(i) *if $\mathbb{P}_i(T_i < \infty) = 1$, then i is recurrent and $\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$;*
(ii) *if $\mathbb{P}_i(T_i < \infty) < 1$, then i is transient and $\sum_{n=0}^{\infty} p_{ii}^{(n)} < \infty$.*
In particular, every state is either transient or recurrent.

Proof. If $\mathbb{P}_i(T_i < \infty) = 1$, then, by Lemma 1.5.2,

$$\mathbb{P}_i(V_i = \infty) = \lim_{r\to\infty} \mathbb{P}_i(V_i > r) = 1$$

so i is recurrent and

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \mathbb{E}_i(V_i) = \infty.$$

On the other hand, if $f_i = \mathbb{P}_i(T_i < \infty) < 1$, then by Lemma 1.5.2

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \mathbb{E}_i(V_i) = \sum_{r=0}^{\infty} \mathbb{P}_i(V_i > r) = \sum_{r=0}^{\infty} f_i^r = \frac{1}{1-f_i} < \infty$$

so $\mathbb{P}_i(V_i = \infty) = 0$ and i is transient. $\square$

From this theorem we can go on to solve completely the problem of recurrence or transience for Markov chains with finite state-space. Some cases of infinite state-space are dealt with in the following chapter. First we show that recurrence and transience are *class properties*.

Theorem 1.5.4. *Let C be a communicating class. Then either all states in C are transient or all are recurrent.*

Proof. Take any pair of states $i, j \in C$ and suppose that i is transient. There exist $n, m \geq 0$ with $p_{ij}^{(n)} > 0$ and $p_{ji}^{(m)} > 0$, and, for all $r \geq 0$

$$p_{ii}^{(n+r+m)} \geq p_{ij}^{(n)} p_{jj}^{(r)} p_{ji}^{(m)}$$

so

$$\sum_{r=0}^{\infty} p_{jj}^{(r)} \le \frac{1}{p_{ij}^{(n)} p_{ji}^{(m)}} \sum_{r=0}^{\infty} p_{ii}^{(n+r+m)} < \infty$$

by Theorem 1.5.3. Hence j is also transient by Theorem 1.5.3. $\square$

In the light of this theorem it is natural to speak of a recurrent or transient class.

Theorem 1.5.5. *Every recurrent class is closed.*

Proof. Let C be a class which is not closed. Then there exist $i \in C$, $j \notin C$ and $m \ge 1$ with

$$\mathbb{P}_i(X_m = j) > 0.$$

Since we have

$$\mathbb{P}_i(\{X_m = j\} \cap \{X_n = i \text{ for infinitely many } n\}) = 0$$

this implies that

$$\mathbb{P}_i(X_n = i \text{ for infinitely many } n) < 1$$

so i is not recurrent, and so neither is C. $\square$

Theorem 1.5.6. *Every finite closed class is recurrent.*

Proof. Suppose C is closed and finite and that $(X_n)_{n \ge 0}$ starts in C. Then for some $i \in C$ we have

$$\begin{aligned} 0 &< \mathbb{P}(X_n = i \text{ for infinitely many } n) \\ &= \mathbb{P}(X_n = i \text{ for some } n)\mathbb{P}_i(X_n = i \text{ for infinitely many } n) \end{aligned}$$

by the strong Markov property. This shows that i is not transient, so C is recurrent by Theorems 1.5.3 and 1.5.4. $\square$

It is easy to spot closed classes, so the transience or recurrence of finite classes is easy to determine. For example, the only recurrent class in Example 1.2.2 is $\{5, 6\}$, the others being transient. On the other hand, infinite closed classes may be transient: see Examples 1.3.3 and 1.6.3.

We shall need the following result in Section 1.8. Remember that irreducibility means that the chain can get from any state to any other, with positive probability.

Theorem 1.5.7. *Suppose P is irreducible and recurrent. Then for all $j \in I$ we have $\mathbb{P}(T_j < \infty) = 1$.*

Proof. By the Markov property we have

$$\mathbb{P}(T_j < \infty) = \sum_{i \in I} \mathbb{P}(X_0 = i)\mathbb{P}_i(T_j < \infty)$$

so it suffices to show $\mathbb{P}_i(T_j < \infty) = 1$ for all $i \in I$. Choose m with $p_{ji}^{(m)} > 0$. By Theorem 1.5.3, we have

$$\begin{aligned} 1 &= \mathbb{P}_j(X_n = j \text{ for infinitely many } n) \\ &= \mathbb{P}_j(X_n = j \text{ for some } n \geq m+1) \\ &= \sum_{k \in I} \mathbb{P}_j(X_n = j \text{ for some } n \geq m+1 \mid X_m = k)\mathbb{P}_j(X_m = k) \\ &= \sum_{k \in I} \mathbb{P}_k(T_j < \infty)p_{jk}^{(m)} \end{aligned}$$

where the final equality uses the Markov property. But $\sum_{k \in I} p_{jk}^{(m)} = 1$ so we must have $\mathbb{P}_i(T_j < \infty) = 1$. □

Exercises

1.5.1 In Exercise 1.2.1, which states are recurrent and which are transient?

1.5.2 Show that, for the Markov chain $(X_n)_{n \geq 0}$ in Exercise 1.3.4 we have

$$\mathbb{P}(X_n \to \infty \text{ as } n \to \infty) = 1\,.$$

Suppose, instead, the transition probabilities satisfy

$$p_{i,i+1} = \left(\frac{i+1}{i}\right)^{\alpha} p_{i,i-1}.$$

For each $\alpha \in (0, \infty)$ find the value of $\mathbb{P}(X_n \to \infty \text{ as } n \to \infty)$.

1.5.3 (First passage decomposition). Denote by T_j the first passage time to state j and set

$$f_{ij}^{(n)} = \mathbb{P}_i(T_j = n).$$

Justify the identity

$$p_{ij}^{(n)} = \sum_{k=1}^{n} f_{ij}^{(k)} p_{jj}^{(n-k)} \qquad \text{for } n \geq 1$$

and deduce that

$$P_{ij}(s) = \delta_{ij} + F_{ij}(s)P_{jj}(s)$$

where

$$P_{ij}(s) = \sum_{n=0}^{\infty} p_{ij}^{(n)} s^n, \quad F_{ij}(s) = \sum_{n=0}^{\infty} f_{ij}^{(n)} s^n.$$

Hence show that $\mathbb{P}_i(T_i < \infty) = 1$ if and only if

$$\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$$

without using Theorem 1.5.3.

1.5.4 A random sequence of non-negative integers $(F_n)_{n\geq 0}$ is obtained by setting $F_0 = 0$ and $F_1 = 1$ and, once $F_0, \dots, F_n$ are known, taking F_{n+1} to be either the sum or the difference of F_{n-1} and F_n, each with probability $1/2$. Is $(F_n)_{n\geq 0}$ a Markov chain?

By considering the Markov chain $X_n = (F_{n-1}, F_n)$, find the probability that $(F_n)_{n\geq 0}$ reaches 3 before first returning to 0.

Draw enough of the flow diagram for $(X_n)_{n\geq 0}$ to establish a general pattern. Hence, using the strong Markov property, show that the hitting probability for $(1, 1)$, starting from $(1, 2)$, is $(3 - \sqrt{5})/2$.

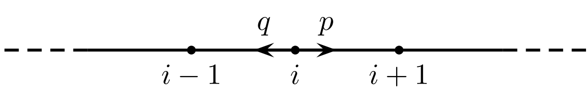
Deduce that $(X_n)_{n\geq 0}$ is transient. Show that, moreover, with probability 1, $F_n \to \infty$ as $n \to \infty$.

1.6 Recurrence and transience of random walks

In the last section we showed that recurrence was a class property, that all recurrent classes were closed and that all finite closed classes were recurrent. So the only chains for which the question of recurrence remains interesting are irreducible with infinite state-space. Here we shall study some simple and fundamental examples of this type, making use of the following criterion for recurrence from Theorem 1.5.3: a state i is recurrent if and only if $\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$.

Example 1.6.1 (Simple random walk on $\mathbb{Z}$)

The simple random walk on $\mathbb{Z}$ has diagram



where $0 < p = 1 - q < 1$. Suppose we start at 0. It is clear that we cannot return to 0 after an odd number of steps, so $p_{00}^{(2n+1)} = 0$ for all n. Any given sequence of steps of length $2n$ from 0 to 0 occurs with probability $p^n q^n$, there being n steps up and n steps down, and the number of such sequences is the number of ways of choosing the n steps up from $2n$. Thus

$$p_{00}^{(2n)} = \binom{2n}{n} p^n q^n.$$

Stirling's formula provides a good approximation to $n!$ for large n: it is known that

$$n! \sim \sqrt{2\pi n}(n/e)^n \qquad \text{as } n \to \infty$$

where $a_n \sim b_n$ means $a_n / b_n \to 1$. For a proof see W. Feller, *An Introduction to Probability Theory and its Applications, Vol I* (Wiley, New York, 3rd edition, 1968). At the end of this chapter we reproduce the argument used by Feller to show that

$$n! \sim A\sqrt{n}(n/e)^n \qquad \text{as } n \to \infty$$

for some $A \in [1, \infty)$. The additional work needed to show $A = \sqrt{2\pi}$ is omitted, as this fact is unnecessary to our applications.

For the n-step transition probabilities we obtain

$$p_{00}^{(2n)} = \frac{(2n)!}{(n!)^2}(pq)^n \sim \frac{(4pq)^n}{A\sqrt{n/2}} \qquad \text{as } n \to \infty.$$

In the symmetric case $p = q = 1/2$, so $4pq = 1$; then for some N and all $n \geq N$ we have

$$p_{00}^{(2n)} \geq \frac{1}{2A\sqrt{n}}$$

so

$$\sum_{n=N}^{\infty} p_{00}^{(2n)} \geq \frac{1}{2A} \sum_{n=N}^{\infty} \frac{1}{\sqrt{n}} = \infty$$

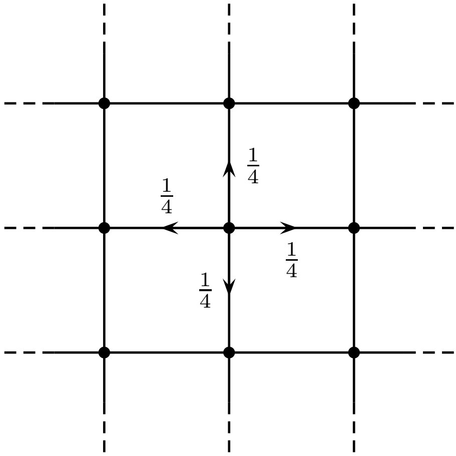
which shows that the random walk is recurrent. On the other hand, if $p \neq q$ then $4pq = r < 1$, so by a similar argument, for some N

$$\sum_{n=N}^{\infty} p_{00}^{(n)} \leq \frac{1}{A} \sum_{n=N}^{\infty} r^n < \infty$$

showing that the random walk is transient.

Example 1.6.2 (Simple symmetric random walk on $\mathbb{Z}^2$)

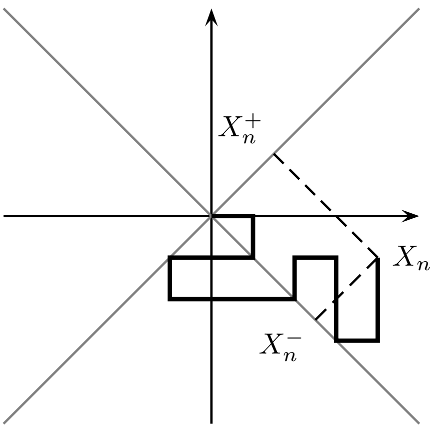
The simple symmetric random walk on $\mathbb{Z}^2$ has diagram



and transition probabilities

$$p_{ij} = \begin{cases} 1/4 & \text{if } |i - j| = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Suppose we start at 0. Let us call the walk X_n and write X_n^+ and X_n^- for the orthogonal projections of X_n on the diagonal lines $y = \pm x$:



Then X_n^+ and X_n^- are independent simple symmetric random walks on $2^{-1/2}\mathbb{Z}$ and $X_n = 0$ if and only if $X_n^+ = 0 = X_n^-$. This makes it clear that for X_n we have

$$p_{00}^{(2n)} = \left(\binom{2n}{n} \left(\frac{1}{2}\right)^{2n} \right)^2 \sim \frac{2}{A^2 n} \qquad \text{as } n \to \infty$$

by Stirling's formula. Then $\sum_{n=1}^{\infty} p_{00}^{(n)} = \infty$ by comparison with $\sum_{n=1}^{\infty} 1/n$ and the walk is recurrent.

Example 1.6.3 (Simple symmetric random walk on $\mathbb{Z}^3$)

The transition probabilities of the simple symmetric random walk on $\mathbb{Z}^3$ are given by

$$p_{ij} = \begin{cases} 1/6 & \text{if } |i-j| = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Thus the chain jumps to each of its nearest neighbours with equal probability. Suppose we start at 0. We can only return to 0 after an even number $2n$ of steps. Of these $2n$ steps there must be i up, i down, j north, j south, k east and k west for some $i, j, k \geq 0$, with $i + j + k = n$. By counting the ways in which this can be done, we obtain

$$p_{00}^{(2n)} = \sum_{\substack{i,j,k \geq 0 \\ i+j+k=n}} \frac{(2n)!}{(i!j!k!)^2} \left(\frac{1}{6}\right)^{2n} = \binom{2n}{n} \left(\frac{1}{2}\right)^{2n} \sum_{\substack{i,j,k \geq 0 \\ i+j+k=n}} \binom{n}{i\ j\ k}^2 \left(\frac{1}{3}\right)^{2n}.$$

Now

$$\sum_{\substack{i,j,k \geq 0 \\ i+j+k=n}} \binom{n}{i\ j\ k} \left(\frac{1}{3}\right)^{n} = 1$$

the left-hand side being the total probability of all the ways of placing n balls randomly into three boxes. For the case where $n = 3m$, we have

$$\binom{n}{i\ j\ k} = \frac{n!}{i!j!k!} \leq \binom{n}{m\ m\ m}$$

for all i, j, k, so

$$p_{00}^{(2n)} \leq \binom{2n}{n} \left(\frac{1}{2}\right)^{2n} \binom{n}{m\ m\ m} \left(\frac{1}{3}\right)^{n} \sim \frac{1}{2A^3} \left(\frac{6}{n}\right)^{3/2} \quad \text{as } n \to \infty$$

by Stirling's formula. Hence, $\sum_{m=0}^{\infty} p_{00}^{(6m)} < \infty$ by comparison with $\sum_{n=0}^{\infty} n^{-3/2}$. But $p_{00}^{(6m)} \geq (1/6)^2 p_{00}^{(6m-2)}$ and $p_{00}^{(6m)} \geq (1/6)^4 p_{00}^{(6m-4)}$ for all m so we must have

$$\sum_{n=0}^{\infty} p_{00}^{(n)} < \infty$$

and the walk is transient.

Exercises

1.6.1 The rooted binary tree is an infinite graph T with one distinguished vertex R from which comes a single edge; at every other vertex there are three edges and there are no closed loops. The random walk on T jumps from a vertex along each available edge with equal probability. Show that the random walk is transient.

1.6.2 Show that the simple symmetric random walk in $\mathbb{Z}^4$ is transient.

1.7 Invariant distributions

Many of the long-time properties of Markov chains are connected with the notion of an invariant distribution or measure. Remember that a measure λ is any row vector $(\lambda_i : i \in I)$ with non-negative entries. We say λ is *invariant* if

$$\lambda P = \lambda.$$

The terms *equilibrium* and *stationary* are also used to mean the same. The first result explains the term stationary.

Theorem 1.7.1. *Let $(X_n)_{n\geq 0}$ be Markov(λ, P) and suppose that λ is invariant for P. Then $(X_{m+n})_{n\geq 0}$ is also Markov(λ, P).*

Proof. By Theorem 1.1.3, $\mathbb{P}(X_m = i) = (\lambda P^m)_i = \lambda_i$ for all i and, clearly, conditional on $X_{m+n} = i$, X_{m+n+1} is independent of $X_m, X_{m+1}, \dots, X_{m+n}$ and has distribution $(p_{ij} : j \in I)$. $\square$

The next result explains the term equilibrium.

Theorem 1.7.2. *Let I be finite. Suppose for some $i \in I$ that*

$$p_{ij}^{(n)} \to \pi_j \quad \text{as} \quad n \to \infty \quad \text{for all } j \in I.$$

Then $\pi = (\pi_j : j \in I)$ is an invariant distribution.

Proof. We have

$$\sum_{j\in I} \pi_j = \sum_{j\in I} \lim_{n\to\infty} p_{ij}^{(n)} = \lim_{n\to\infty} \sum_{j\in I} p_{ij}^{(n)} = 1$$

and

$$\pi_j = \lim_{n\to\infty} p_{ij}^{(n)} = \lim_{n\to\infty} \sum_{k\in I} p_{ik}^{(n)} p_{kj} = \sum_{k\in I} \lim_{n\to\infty} p_{ik}^{(n)} p_{kj} = \sum_{k\in I} \pi_k p_{kj}$$

where we have used finiteness of I to justify interchange of summation and limit operations. Hence π is an invariant distribution. $\square$

Notice that for any of the random walks discussed in Section 1.6 we have $p_{ij}^{(n)} \to 0$ as $n \to \infty$ for all $i, j \in I$. The limit is certainly invariant, but it is not a distribution!

Theorem 1.7.2 is not a very useful result but it serves to indicate a relationship between invariant distributions and n-step transition probabilities. In Theorem 1.8.3 we shall prove a sort of converse, which is much more useful.

Example 1.7.3

Consider the two-state Markov chain with transition matrix

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

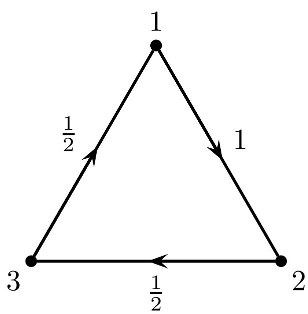
Ignore the trivial cases $\alpha = \beta = 0$ and $\alpha = \beta = 1$. Then, by Example 1.1.4

$$P^n \to \begin{pmatrix} \beta/(\alpha+\beta) & \alpha/(\alpha+\beta) \\ \beta/(\alpha+\beta) & \alpha/(\alpha+\beta) \end{pmatrix} \quad \text{as } n \to \infty,$$

so, by Theorem 1.7.2, the distribution $(\beta/(\alpha+\beta), \alpha/(\alpha+\beta))$ must be invariant. There are of course easier ways to discover this.

Example 1.7.4

Consider the Markov chain $(X_n)_{n\geq 0}$ with diagram



To find an invariant distribution we write down the components of the vector equation $\pi P = \pi$

$$\begin{aligned} \pi_1 &= \tfrac{1}{2}\pi_3 \\ \pi_2 &= \pi_1 + \tfrac{1}{2}\pi_2 \\ \pi_3 &= \tfrac{1}{2}\pi_2 + \tfrac{1}{2}\pi_3. \end{aligned}$$

In terms of the chain, the right-hand sides give the probabilities for X_1, when X_0 has distribution π, and the equations require X_1 also to have distribution π. The equations are homogeneous so one of them is redundant, and another equation is required to fix π uniquely. That equation is

$$\pi_1 + \pi_2 + \pi_3 = 1$$

and we find that $\pi = \big(1/5, 2/5, 2/5\big)$.

According to Example 1.1.6

$$p_{11}^{(n)} \to 1/5 \quad \text{as } n \to \infty$$

so this confirms Theorem 1.7.2. Alternatively, knowing that $p_{11}^{(n)}$ had the form

$$p_{11}^{(n)} = a + \left(\frac{1}{2}\right)^n \left(b \cos\frac{n\pi}{2} + c \sin\frac{n\pi}{2}\right)$$

we could have used Theorem 1.7.2 and knowledge of π_1 to identify $a = 1/5$, instead of working out $p_{11}^{(2)}$ in Example 1.1.6.

In the next two results we shall show that every irreducible and recurrent stochastic matrix P has an essentially unique positive invariant measure. The proofs rely heavily on the probabilistic interpretation so it is worth noting at the outset that, for a finite state-space I, the existence of an invariant row vector is a simple piece of linear algebra: the row sums of P are all 1, so the column vector of ones is an eigenvector with eigenvalue 1, so P must have a row eigenvector with eigenvalue 1.

For a fixed state k, consider for each i the *expected time spent in i between visits to k*:

$$\gamma_i^k = \mathbb{E}_k \sum_{n=0}^{T_k - 1} 1_{\{X_n = i\}}.$$

Here the sum of indicator functions serves to count the number of times n at which $X_n = i$ before the first passage time T_k.

Theorem 1.7.5. *Let P be irreducible and recurrent. Then*

(i) $\gamma_k^k = 1$;

(ii) $\gamma^k = (\gamma_i^k : i \in I)$ *satisfies* $\gamma^k P = \gamma^k$;

(iii) $0 < \gamma_i^k < \infty$ *for all* $i \in I$.

Proof. (i) This is obvious. (ii) For $n = 1, 2, \dots$ the event $\{n \le T_k\}$ depends only on $X_0, X_1, \dots, X_{n-1}$, so, by the Markov property at $n - 1$

$$\mathbb{P}_k(X_{n-1} = i, X_n = j \text{ and } n \le T_k) = \mathbb{P}_k(X_{n-1} = i \text{ and } n \le T_k) p_{ij}.$$

Since P is recurrent, under $\mathbb{P}_k$ we have $T_k < \infty$ and $X_0 = X_{T_k} = k$ with probability one. Therefore

$$\begin{aligned}
\gamma_j^k &= \mathbb{E}_k \sum_{n=1}^{T_k} 1_{\{X_n=j\}} = \mathbb{E}_k \sum_{n=1}^{\infty} 1_{\{X_n=j \text{ and } n\le T_k\}} \\
&= \sum_{n=1}^{\infty} \mathbb{P}_k(X_n = j \text{ and } n \le T_k) \\
&= \sum_{i\in I} \sum_{n=1}^{\infty} \mathbb{P}_k(X_{n-1} = i, X_n = j \text{ and } n \le T_k) \\
&= \sum_{i\in I} p_{ij} \sum_{n=1}^{\infty} \mathbb{P}_k(X_{n-1} = i \text{ and } n \le T_k) \\
&= \sum_{i\in I} p_{ij} \mathbb{E}_k \sum_{m=0}^{\infty} 1_{\{X_m=i \text{ and } m\le T_k-1\}} \\
&= \sum_{i\in I} p_{ij} \mathbb{E}_k \sum_{m=0}^{T_k-1} 1_{\{X_m=i\}} = \sum_{i\in I} \gamma_i^k p_{ij}.
\end{aligned}$$

(iii) Since P is irreducible, for each state i there exist $n, m \ge 0$ with $p_{ik}^{(n)}, p_{ki}^{(m)} > 0$. Then $\gamma_i^k \ge \gamma_k^k p_{ki}^{(m)} > 0$ and $\gamma_i^k p_{ik}^{(n)} \le \gamma_k^k = 1$ by (i) and (ii). □

Theorem 1.7.6. *Let P be irreducible and let λ be an invariant measure for P with $\lambda_k = 1$. Then $\lambda \ge \gamma^k$. If in addition P is recurrent, then $\lambda = \gamma^k$.*

Proof. For each $j \in I$ we have

$$\begin{aligned}
\lambda_j &= \sum_{i_1\in I} \lambda_{i_1} p_{i_1 j} = \sum_{i_1\ne k} \lambda_{i_1} p_{i_1 j} + p_{kj} \\
&= \sum_{i_1,i_2\ne k} \lambda_{i_2} p_{i_2 i_1} p_{i_1 j} + \Big(p_{kj} + \sum_{i_1\ne k} p_{k i_1} p_{i_1 j} \Big) \\
&\;\;\vdots \\
&= \sum_{i_1,\dots,i_n\ne k} \lambda_{i_n} p_{i_n i_{n-1}} \dots p_{i_1 j} \\
&\quad + \Big(p_{kj} + \sum_{i_1\ne k} p_{k i_1} p_{i_1 j} + \dots + \sum_{i_1,\dots,i_{n-1}\ne k} p_{k i_{n-1}} \dots p_{i_2 i_1} p_{i_1 j} \Big)
\end{aligned}$$

So for $j \ne k$ we obtain

$$\begin{aligned}
\lambda_j &\ge \mathbb{P}_k(X_1 = j \text{ and } T_k \ge 1) + \mathbb{P}_k(X_2 = j \text{ and } T_k \ge 2) \\
&\quad + \dots + \mathbb{P}_k(X_n = j \text{ and } T_k \ge n) \\
&\to \gamma_j^k \quad \text{as } n \to \infty.
\end{aligned}$$

So $\lambda \geq \gamma^k$. If P is recurrent, then γ^k is invariant by Theorem 1.7.5, so $\mu = \lambda - \gamma^k$ is also invariant and $\mu \geq 0$. Since P is irreducible, given $i \in I$, we have $p_{ik}^{(n)} > 0$ for some n, and $0 = \mu_k = \sum_{j \in I} \mu_j p_{jk}^{(n)} \geq \mu_i p_{ik}^{(n)}$, so $\mu_i = 0$. $\square$

Recall that a state i is recurrent if

$$\mathbb{P}_i(X_n = i \text{ for infinitely many } n) = 1$$

and we showed in Theorem 1.5.3 that this is equivalent to

$$\mathbb{P}_i(T_i < \infty) = 1.$$

If in addition the *expected return time*

$$m_i = \mathbb{E}_i(T_i)$$

is finite, then we say i is *positive recurrent*. A recurrent state which fails to have this stronger property is called *null recurrent*.

Theorem 1.7.7. *Let P be irreducible. Then the following are equivalent:*

(i) *every state is positive recurrent;*

(ii) *some state i is positive recurrent;*

(iii) *P has an invariant distribution, π say.*
Moreover, when (iii) *holds we have $m_i = 1/\pi_i$ for all i.*

Proof. (i) $\Rightarrow$ (ii) This is obvious.
(ii) $\Rightarrow$ (iii) If i is positive recurrent, it is certainly recurrent, so P is recurrent. By Theorem 1.7.5, γ^i is then invariant. But

$$\sum_{j \in I} \gamma_j^i = m_i < \infty$$

so $\pi_j = \gamma_j^i / m_i$ defines an invariant distribution.
(iii) $\Rightarrow$ (i) Take any state k. Since P is irreducible and $\sum_{i \in I} \pi_i = 1$ we have $\pi_k = \sum_{i \in I} \pi_i p_{ik}^{(n)} > 0$ for some n. Set $\lambda_i = \pi_i / \pi_k$. Then λ is an invariant measure with $\lambda_k = 1$. So by Theorem 1.7.6, $\lambda \geq \gamma^k$. Hence

$$m_k = \sum_{i \in I} \gamma_i^k \leq \sum_{i \in I} \frac{\pi_i}{\pi_k} = \frac{1}{\pi_k} < \infty \tag{1.7}$$

and k is positive recurrent.

To complete the proof we return to the argument for (iii) ⇒ (i) armed with the knowledge that P is recurrent, so $\lambda = \gamma^k$ and the inequality (1.7) is in fact an equality. □

Example 1.7.8 (Simple symmetric random walk on $\mathbb{Z}$)

The simple symmetric random walk on $\mathbb{Z}$ is clearly irreducible and, by Example 1.6.1, it is also recurrent. Consider the measure

$$\pi_i = 1 \quad \text{for all } i.$$

Then

$$\pi_i = \tfrac{1}{2}\pi_{i-1} + \tfrac{1}{2}\pi_{i+1}$$

so π is invariant. Now Theorem 1.7.6 forces any invariant measure to be a scalar multiple of π. Since $\sum_{i\in\mathbb{Z}} \pi_i = \infty$, there can be no invariant distribution and the walk is therefore null recurrent, by Theorem 1.7.7.

Example 1.7.9

The existence of an invariant measure does not guarantee recurrence: consider, for example, the simple symmetric random walk on $\mathbb{Z}^3$, which is transient by Example 1.6.3, but has invariant measure π given by $\pi_i = 1$ for all i.

Example 1.7.10

Consider the asymmetric random walk on $\mathbb{Z}$ with transition probabilities $p_{i,i-1} = q < p = p_{i,i+1}$. In components the invariant measure equation $\pi P = \pi$ reads

$$\pi_i = \pi_{i-1}p + \pi_{i+1}q.$$

This is a recurrence relation for π with general solution

$$\pi_i = A + B(p/q)^i.$$

So, in this case, there is a two-parameter family of invariant measures – uniqueness up to scalar multiples does not hold.

Example 1.7.11

Consider a *success-run chain* on $\mathbb{Z}^+$, whose transition probabilities are given by

$$p_{i,i+1} = p_i, \quad p_{i0} = q_i = 1 - p_i.$$

Then the components of the invariant measure equation $\pi P = \pi$ read

$$\begin{aligned} \pi_0 &= \sum_{i=0}^{\infty} q_i \pi_i, \\ \pi_i &= p_{i-1}\pi_{i-1}, \quad \text{for } i \geq 1. \end{aligned}$$

Suppose we choose p_i converging sufficiently rapidly to 1 so that

$$p = \prod_{i=0}^{\infty} p_i > 0.$$

Then for any invariant measure π we have

$$\pi_0 = \sum_{i=0}^{\infty} (1 - p_i) p_{i-1} \dots p_0 \pi_0 = (1 - p)\pi_0.$$

This equation forces either $\pi_0 = 0$ or $\pi_0 = \infty$, so there is no non-zero invariant measure.

Exercises

1.7.1 Find all invariant distributions of the transition matrix in Exercise 1.2.1.

1.7.2 Gas molecules move about randomly in a box which is divided into two halves symmetrically by a partition. A hole is made in the partition. Suppose there are N molecules in the box. Show that the number of molecules on one side of the partition just after a molecule has passed through the hole evolves as a Markov chain. What are the transition probabilities? What is the invariant distribution of this chain?

1.7.3 A particle moves on the eight vertices of a cube in the following way: at each step the particle is equally likely to move to each of the three adjacent vertices, independently of its past motion. Let i be the initial vertex occupied by the particle, o the vertex opposite i. Calculate each of the following quantities:

(i) the expected number of steps until the particle returns to i;

(ii) the expected number of visits to o until the first return to i;

(iii) the expected number of steps until the first visit to o.

1.7.4 Let $(X_n)_{n\geq 0}$ be a simple random walk on $\mathbb{Z}$ with $p_{i,i-1} = q < p = p_{i,i+1}$. Find

$$\gamma_i^0 = \mathbb{E}_0\left(\sum_{n=0}^{T_0-1} 1_{\{X_n=i\}}\right)$$

and verify that

$$\gamma_i^0 = \inf_\lambda \lambda_i \quad \text{for all } i$$

where the infimum is taken over all invariant measures λ with $\lambda_0 = 1$. (Compare with Theorem 1.7.6 and Example 1.7.10.)

1.7.5 Let P be a stochastic matrix on a finite set I. Show that a distribution π is invariant for P if and only if $\pi(I-P+A) = a$, where $A = (a_{ij} : i,j \in I)$ with $a_{ij} = 1$ for all i and j, and $a = (a_i : i \in I)$ with $a_i = 1$ for all i. Deduce that if P is irreducible then $I-P+A$ is invertible. *Note that this enables one to compute the invariant distribution by any standard method of inverting a matrix.*

1.8 Convergence to equilibrium

We shall investigate the limiting behaviour of the n-step transition probabilities $p_{ij}^{(n)}$ as $n \to \infty$. As we saw in Theorem 1.7.2, if the state-space is finite and if for some i the limit exists for all j, then it must be an invariant distribution. But, as the following example shows, the limit does not always exist.

Example 1.8.1

Consider the two-state chain with transition matrix

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

Then $P^2 = I$, so $P^{2n} = I$ and $P^{2n+1} = P$ for all n. Thus $p_{ij}^{(n)}$ fails to converge for all i,j.

Let us call a state i *aperiodic* if $p_{ii}^{(n)} > 0$ for all sufficiently large n. We leave it as an exercise to show that i is aperiodic if and only if the set $\{n \geq 0 : p_{ii}^{(n)} > 0\}$ has no common divisor other than 1. This is also a consequence of Theorem 1.8.4. The behaviour of the chain in Example 1.8.1 is connected with its periodicity.

Lemma 1.8.2. *Suppose P is irreducible and has an aperiodic state i. Then, for all states j and k, $p_{jk}^{(n)} > 0$ for all sufficiently large n. In particular, all states are aperiodic.*

Proof. There exist $r, s \geq 0$ with $p_{ji}^{(r)}, p_{ik}^{(s)} > 0$. Then

$$p_{jk}^{(r+n+s)} \geq p_{ji}^{(r)} p_{ii}^{(n)} p_{ik}^{(s)} > 0$$

for all sufficiently large n. $\square$

Here is the main result of this section. The method of proof, by coupling two Markov chains, is ingenious.

Theorem 1.8.3 (Convergence to equilibrium). *Let P be irreducible and aperiodic, and suppose that P has an invariant distribution π. Let λ be any distribution. Suppose that $(X_n)_{n\geq 0}$ is Markov(λ, P). Then*

$$\mathbb{P}(X_n = j) \to \pi_j \quad \text{as } n \to \infty \text{ for all } j.$$

In particular,

$$p_{ij}^{(n)} \to \pi_j \quad \text{as } n \to \infty \text{ for all } i, j.$$

Proof. We use a coupling argument. Let $(Y_n)_{n\geq 0}$ be Markov(π, P) and independent of $(X_n)_{n\geq 0}$. Fix a reference state b and set

$$T = \inf\{n \geq 1 : X_n = Y_n = b\}.$$

Step 1. We show $\mathbb{P}(T < \infty) = 1$. The process $W_n = (X_n, Y_n)$ is a Markov chain on $I \times I$ with transition probabilities

$$\widetilde{p}_{(i,k)(j,l)} = p_{ij} p_{kl}$$

and initial distribution

$$\mu_{(i,k)} = \lambda_i \pi_k.$$

Since P is aperiodic, for all states i, j, k, l we have

$$\widetilde{p}_{(i,k)(j,l)}^{(n)} = p_{ij}^{(n)} p_{kl}^{(n)} > 0$$

for all sufficiently large n; so $\widetilde{P}$ is irreducible. Also, $\widetilde{P}$ has an invariant distribution given by

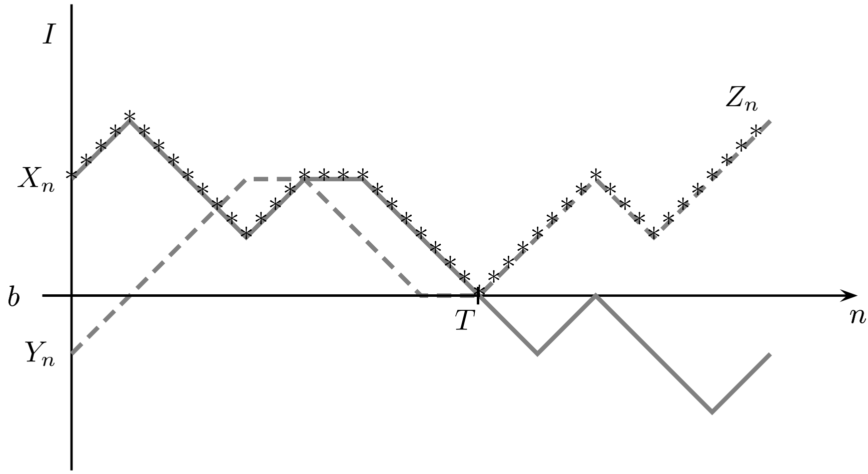
$$\widetilde{\pi}_{(i,k)} = \pi_i \pi_k$$

so, by Theorem 1.7.7, $\widetilde{P}$ is positive recurrent. But T is the first passage time of W_n to (b, b) so $\mathbb{P}(T < \infty) = 1$, by Theorem 1.5.7.

Step 2. Set

$$Z_n = \begin{cases} X_n & \text{if } n < T \\ Y_n & \text{if } n \geq T. \end{cases}$$

The diagram below illustrates the idea. We show that $(Z_n)_{n\geq 0}$ is Markov(λ, P).



The strong Markov property applies to $(W_n)_{n\geq 0}$ at time T, so $(X_{T+n}, Y_{T+n})_{n\geq 0}$ is Markov$(\delta_{(b,b)}, \widetilde{P})$ and independent of (X_0, Y_0), $(X_1, Y_1), \dots, (X_T, Y_T)$. By symmetry, we can replace the process $(X_{T+n}, Y_{T+n})_{n\geq 0}$ by $(Y_{T+n}, X_{T+n})_{n\geq 0}$ which is also Markov$(\delta_{(b,b)}, \widetilde{P})$ and remains independent of $(X_0, Y_0), (X_1, Y_1), \dots, (X_T, Y_T)$. Hence $W_n' = (Z_n, Z_n')$ is Markov$(\mu, \widetilde{P})$ where

$$Z_n' = \begin{cases} Y_n & \text{if } n < T \\ X_n & \text{if } n \geq T. \end{cases}$$

In particular, $(Z_n)_{n\geq 0}$ is Markov(λ, P).

Step 3. We have

$$\mathbb{P}(Z_n = j) = \mathbb{P}(X_n = j \text{ and } n < T) + \mathbb{P}(Y_n = j \text{ and } n \geq T)$$

so

$$\begin{aligned} |\mathbb{P}(X_n = j) - \pi_j| &= |\mathbb{P}(Z_n = j) - \mathbb{P}(Y_n = j)| \\ &= |\mathbb{P}(X_n = j \text{ and } n < T) - \mathbb{P}(Y_n = j \text{ and } n < T)| \\ &\leq \mathbb{P}(n < T) \end{aligned}$$

and $\mathbb{P}(n < T) \to 0$ as $n \to \infty$. $\square$

To understand this proof one should see what goes wrong when P is not aperiodic. Consider the two-state chain of Example 1.8.1 which has $(1/2, 1/2)$ as its unique invariant distribution. We start $(X_n)_{n\geq 0}$ from 0 and $(Y_n)_{n\geq 0}$ with equal probability from 0 or 1. However, if $Y_0 = 1$, then, because of periodicity, $(X_n)_{n\geq 0}$ and $(Y_n)_{n\geq 0}$ will never meet, and the proof fails. We move on now to the cases that were excluded in the last theorem, where $(X_n)_{n\geq 0}$ is periodic or transient or null recurrent. The remainder of this section might be omitted on a first reading.

Theorem 1.8.4. *Let P be irreducible. There is an integer $d \geq 1$ and a partition*

$$I = C_0 \cup C_1 \cup \ldots \cup C_{d-1}$$

such that (setting $C_{nd+r} = C_r$)

(i) $p_{ij}^{(n)} > 0$ *only if $i \in C_r$ and $j \in C_{r+n}$ for some r;*

(ii) $p_{ij}^{(nd)} > 0$ *for all sufficiently large n, for all $i, j \in C_r$, for all r.*

Proof. Fix a state k and consider $S = \{n \geq 0 : p_{kk}^{(n)} > 0\}$. Choose $n_1, n_2 \in S$ with $n_1 < n_2$ and such that $d := n_2 - n_1$ is as small as possible. (Here and throughout we use the symbol $:=$ to mean 'defined to equal'.) Define for $r = 0, \ldots, d-1$

$$C_r = \{i \in I : p_{ki}^{(nd+r)} > 0 \text{ for some } n \geq 0\}.$$

Then $C_0 \cup \ldots \cup C_{d-1} = I$, by irreducibility. Moreover, if $p_{ki}^{(nd+r)} > 0$ and $p_{ki}^{(nd+s)} > 0$ for some $r, s \in \{0, 1, \ldots, d-1\}$, then, choosing $m \geq 0$ so that $p_{ik}^{(m)} > 0$, we have $p_{kk}^{(nd+r+m)} > 0$ and $p_{kk}^{(nd+s+m)} > 0$ so $r = s$ by minimality of d. Hence we have a partition.

To prove (i) suppose $p_{ij}^{(n)} > 0$ and $i \in C_r$. Choose m so that $p_{ki}^{(md+r)} > 0$, then $p_{kj}^{(md+r+n)} > 0$ so $j \in C_{r+n}$ as required. By taking $i = j = k$ we now see that d must divide every element of S, in particular n_1.

Now for $nd \geq n_1^2$, we can write $nd = qn_1 + r$ for integers $q \geq n_1$ and $0 \leq r \leq n_1 - 1$. Since d divides n_1 we then have $r = md$ for some integer m and then $nd = (q-m)n_1 + mn_2$. Hence

$$p_{kk}^{(nd)} \geq (p_{kk}^{(n_1)})^{q-m}(p_{kk}^{(n_2)})^m > 0$$

and hence $nd \in S$. To prove (ii) for $i, j \in C_r$ choose m_1 and m_2 so that $p_{ik}^{(m_1)} > 0$ and $p_{kj}^{(m_2)} > 0$, then

$$p_{ij}^{(m_1+nd+m_2)} \geq p_{ik}^{(m_1)} p_{kk}^{(nd)} p_{kj}^{(m_2)} > 0$$

whenever $nd \geq n_1^2$. Since $m_1 + m_2$ is then necessarily a multiple of d, we are done. $\square$

We call d the *period* of P. The theorem just proved shows in particular for all $i \in I$ that d is the greatest common divisor of the set $\{n \geq 0 : p_{ii}^{(n)} > 0\}$. This is sometimes useful in identifying d.

Finally, here is a complete description of limiting behaviour for irreducible chains. This generalizes Theorem 1.8.3 in two respects since we require neither aperiodicity nor the existence of an invariant distribution. The argument we use for the null recurrent case was discovered recently by B. Fristedt and L. Gray.

Theorem 1.8.5. *Let P be irreducible of period d and let $C_0, C_1, \dots, C_{d-1}$ be the partition obtained in Theorem 1.8.4. Let λ be a distribution with $\sum_{i \in C_0} \lambda_i = 1$. Suppose that $(X_n)_{n\geq 0}$ is Markov(λ, P). Then for $r = 0, 1, \dots, d-1$ and $j \in C_r$ we have*

$$\mathbb{P}(X_{nd+r} = j) \to d/m_j \quad \text{as } n \to \infty$$

where m_j is the expected return time to j. In particular, for $i \in C_0$ and $j \in C_r$ we have

$$p_{ij}^{(nd+r)} \to d/m_j \quad \text{as } n \to \infty.$$

Proof

Step 1. We reduce to the aperiodic case. Set $\nu = \lambda P^r$, then by Theorem 1.8.4 we have

$$\sum_{i \in C_r} \nu_i = 1.$$

Set $Y_n = X_{nd+r}$, then $(Y_n)_{n\geq 0}$ is Markov(ν, P^d) and, by Theorem 1.8.4, P^d is irreducible and aperiodic on C_r. For $j \in C_r$ the expected return time of $(Y_n)_{n\geq 0}$ to j is m_j/d. So if the theorem holds in the aperiodic case, then

$$\mathbb{P}(X_{nd+r} = j) = \mathbb{P}(Y_n = j) \to d/m_j \quad \text{as } n \to \infty$$

so the theorem holds in general.

Step 2. Assume that P is aperiodic. If P is positive recurrent then $1/m_j = \pi_j$, where π is the unique invariant distribution, so the result follows from Theorem 1.8.3. Otherwise $m_j = \infty$ and we have to show that

$$\mathbb{P}(X_n = j) \to 0 \quad \text{as } n \to \infty.$$

If P is transient this is easy and we are left with the null recurrent case.

Step 3. Assume that P is aperiodic and null recurrent. Then

$$\sum_{k=0}^{\infty} \mathbb{P}_j(T_j > k) = \mathbb{E}_j(T_j) = \infty.$$

Given $\varepsilon > 0$ choose K so that

$$\sum_{k=0}^{K-1} \mathbb{P}_j(T_j > k) \geq \frac{2}{\varepsilon}\,.$$

Then, for $n \geq K - 1$

$$\begin{aligned}
1 &\geq \sum_{k=n-K+1}^{n} \mathbb{P}(X_k = j \text{ and } X_m \neq j \text{ for } m = k+1, \dots, n) \\
&= \sum_{k=n-K+1}^{n} \mathbb{P}(X_k = j)\mathbb{P}_j(T_j > n - k) \\
&= \sum_{k=0}^{K-1} \mathbb{P}(X_{n-k} = j)\mathbb{P}_j(T_j > k)
\end{aligned}$$

so we must have $\mathbb{P}(X_{n-k} = j) \leq \varepsilon/2$ for some $k \in \{0, 1, \dots, K-1\}$.

Return now to the coupling argument used in Theorem 1.8.3, only now let $(Y_n)_{n\geq 0}$ be Markov(μ, P), where μ is to be chosen later. Set $W_n = (X_n, Y_n)$. As before, aperiodicity of $(X_n)_{n\geq 0}$ ensures irreducibility of $(W_n)_{n\geq 0}$. If $(W_n)_{n\geq 0}$ is transient then, on taking $\mu = \lambda$, we obtain

$$\mathbb{P}(X_n = j)^2 = \mathbb{P}\big(W_n = (j, j)\big) \to 0$$

as required. Assume then that $(W_n)_{n\geq 0}$ is recurrent. Then, in the notation of Theorem 1.8.3, we have $\mathbb{P}(T < \infty) = 1$ and the coupling argument shows that

$$|\mathbb{P}(X_n = j) - \mathbb{P}(Y_n = j)| \to 0 \qquad \text{as } n \to \infty.$$

We exploit this convergence by taking $\mu = \lambda P^k$ for $k = 1, \dots, K-1$, so that $\mathbb{P}(Y_n = j) = \mathbb{P}(X_{n+k} = j)$. We can find N such that for $n \geq N$ and $k = 1, \dots, K-1$,

$$|\mathbb{P}(X_n = j) - \mathbb{P}(X_{n+k} = j)| \leq \frac{\varepsilon}{2}\,.$$

But for any n we can find $k \in \{0, 1, \dots, K-1\}$ such that $\mathbb{P}(X_{n+k} = j) \leq \varepsilon/2$. Hence, for $n \geq N$

$$\mathbb{P}(X_n = j) \leq \varepsilon.$$

Since $\varepsilon > 0$ was arbitrary, this shows that $\mathbb{P}(X_n = j) \to 0$ as $n \to \infty$, as required. $\square$

Exercises

1.8.1 Prove the claims (e), (f) and (g) made in example (v) of the Introduction.

1.8.2 Find the invariant distributions of the transition matrices in Exercise 1.1.7, parts (a), (b) and (c), and compare them with your answers there.

1.8.3 A fair die is thrown repeatedly. Let X_n denote the sum of the first n throws. Find

$$\lim_{n \to \infty} \mathbb{P}(X_n \text{ is a multiple of } 13)$$

quoting carefully any general theorems that you use.

1.8.4 Each morning a student takes one of the three books he owns from his shelf. The probability that he chooses book i is α_i, where $0 < \alpha_i < 1$ for $i = 1, 2, 3$, and choices on successive days are independent. In the evening he replaces the book at the left-hand end of the shelf. If p_n denotes the probability that on day n the student finds the books in the order 1,2,3, from left to right, show that, irrespective of the initial arrangement of the books, p_n converges as $n \to \infty$, and determine the limit.

1.8.5 (Renewal theorem). Let $Y_1, Y_2, \dots$ be independent, identically distributed random variables with values in $\{1, 2, \dots\}$. Suppose that the set of integers

$$\{n : \mathbb{P}(Y_1 = n) > 0\}$$

has greatest common divisor 1. Set $\mu = \mathbb{E}(Y_1)$. Show that the following process is a Markov chain:

$$X_n = \inf\{m \geq n : m = Y_1 + \ldots + Y_k \text{ for some } k \geq 0\} - n.$$

Determine

$$\lim_{n \to \infty} \mathbb{P}(X_n = 0)$$

and hence show that as $n \to \infty$

$$\mathbb{P}(n = Y_1 + \ldots + Y_k \text{ for some } k \geq 0) \to 1/\mu.$$