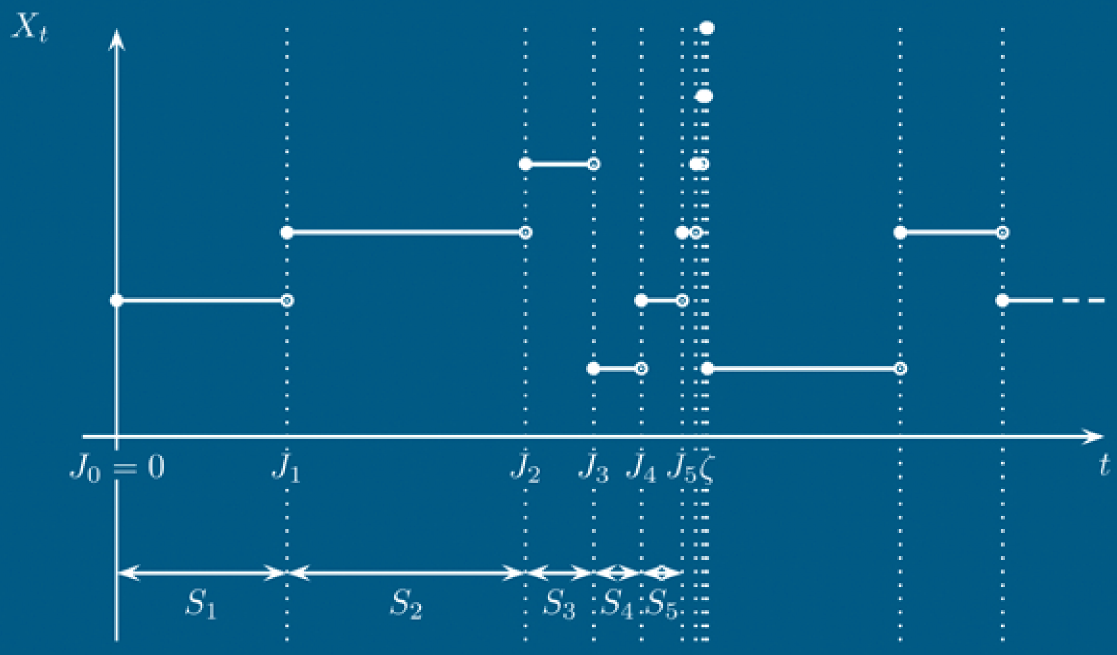
Cambridge Series in Statistical and Probabilistic Mathematics



Markov Chains

J.R. Norris

Markov Chains

Cambridge Series on Statistical and Probabilistic Mathematics

This series of high quality upper-division textbooks and expository monographs covers all areas of stochastic applicable mathematics. The topics range from pure and applied statistics to probability theory, operations research, mathematical programming, and optimzation. The books contain clear presentations of new developments in the field and also of the state of the art in classical methods. While emphasizing rigorous treatment of theoretical methods, the books contain important applications and discussions of new techniques made possible be advances in computational methods.

Markov Chains

J. R. Norris
University of Cambridge

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi

Cambridge University Press
32 Avenue of the Americas, New York, NY 10013-2473, USA

www.cambridge.org
Information on this title: www.cambridge.org/9780521481816

First published 1997
First paperback edition 1998
15th printing 2009

Printed in the United States of America

A catalog record for this publication is available from the British Library.

Library of Congress Cataloging in Publication Data

Norris, J. R. (James R.)
Markov chains / J. R. Norris.
p. cm. – (Cambridge series on statistical and probabilistic mathematics ; no. 2)
Includes bibliographical references and index.
ISBN 0-521-48181-3 (pbk.)
1. Markov processes. I. Title. II. Series.
QA274.7.N67 1997
519.2'33–dc20 96-31570
CIP

ISBN 978-0-521-48181-6 hardback
ISBN 978-0-521-63396-3 paperback

For my parents

Contents

Preface

Markov chains are the simplest mathematical models for random phenomena evolving in time. Their simple structure makes it possible to say a great deal about their behaviour. At the same time, the class of Markov chains is rich enough to serve in many applications. This makes Markov chains the first and most important examples of random processes. Indeed, the whole of the mathematical study of random processes can be regarded as a generalization in one way or another of the theory of Markov chains.

This book is an account of the elementary theory of Markov chains, with applications. It was conceived as a text for advanced undergraduates or master's level students, and is developed from a course taught to undergraduates for several years. There are no strict prerequisites but it is envisaged that the reader will have taken a course in elementary probability. In particular, measure theory is not a prerequisite.

The first half of the book is based on lecture notes for the undergraduate course. Illustrative examples introduce many of the key ideas. Careful proofs are given throughout. There is a selection of exercises, which forms the basis of classwork done by the students, and which has been tested over several years. Chapter 1 deals with the theory of discrete-time Markov chains, and is the basis of all that follows. You must begin here. The material is quite straightforward and the ideas introduced permeate the whole book. The basic pattern of Chapter 1 is repeated in Chapter 3 for continuous-time chains, making it easy to follow the development by analogy. In between, Chapter 2 explains how to set up the theory of continuous-time chains, beginning with simple examples such as the Poisson process and chains with finite state space.

The second half of the book comprises three independent chapters intended to complement the first half. In some sections the style is a little more demanding. Chapter 4 introduces, in the context of elementary Markov chains, some of the ideas crucial to the advanced study of Markov processes, such as martingales, potentials, electrical networks and Brownian motion. Chapter 5 is devoted to applications, for example to population growth, mathematical genetics, queues and networks of queues, Markov decision processes and Monte Carlo simulation. Chapter 6 is an appendix to the main text, where we explain some of the basic notions of probability and measure used in the rest of the book and give careful proofs of the few points where measure theory is really needed.

The following paragraph is directed primarily at an instructor and assumes some familiarity with the subject. Overall, the book is more focused on the Markovian context than most other books dealing with the elementary theory of stochastic processes. I believe that this restriction in scope is desirable for the greater coherence and depth it allows. The treatment of discrete-time chains in Chapter 1 includes the calculation of transition probabilities, hitting probabilities, expected hitting times and invariant distributions. Also treated are recurrence and transience, convergence to equilibrium, reversibility, and the ergodic theorem for long-run averages. All the results are proved, exploiting to the full the probabilistic viewpoint. For example, we use excursions and the strong Markov property to obtain conditions for recurrence and transience, and convergence to equilibrium is proved by the coupling method. In Chapters 2 and 3 we proceed via the jump chain/holding time construction to treat all right-continuous, minimal continuous-time chains, and establish analogues of all the main results obtained for discrete time. No conditions of uniformly bounded rates are needed. The student has the option to take Chapter 3 first, to study the *properties* of continuous-time chains before the technically more demanding *construction*. We have left measure theory in the background, but the proofs are intended to be rigorous, or very easily made rigorous, when considered in measure-theoretic terms. Some further details are given in Chapter 6.

It is a pleasure to acknowledge the work of colleagues from which I have benefitted in preparing this book. The course on which it is based has evolved over many years and under many hands – I inherited parts of it from Martin Barlow and Chris Rogers. In recent years it has been given by Doug Kennedy and Colin Sparrow. Richard Gibbens, Geoffrey Grimmett, Frank Kelly and Gareth Roberts gave expert advice at various stages. Meena Lakshmanan, Violet Lo and David Rose pointed out many typos and ambiguities. Brian Ripley and David Williams made constructive suggestions for improvement of an early version.

I am especially grateful to David Tranah at Cambridge University Press for his suggestion to write the book and for his continuing support, and to Sarah Shea-Simonds who typeset the whole book with efficiency, precision and good humour.

Cambridge, 1996 James Norris

Introduction

This book is about a certain sort of random process. The characteristic property of this sort of process is that it retains *no memory* of where it has been in the past. This means that only the current state of the process can influence where it goes next. Such a process is called a *Markov process*. We shall be concerned exclusively with the case where the process can assume only a finite or countable set of states, when it is usual to refer it as a *Markov chain*.

Examples of Markov chains abound, as you will see throughout the book. What makes them important is that not only do Markov chains model many phenomena of interest, but also the lack of memory property makes it possible to predict how a Markov chain may behave, and to compute probabilities and expected values which quantify that behaviour. In this book we shall present general techniques for the analysis of Markov chains, together with many examples and applications. In this introduction we shall discuss a few very simple examples and preview some of the questions which the general theory will answer.

We shall consider chains both in *discrete time*

$$n \in \mathbb{Z}^+ = \{0, 1, 2, \dots\}$$

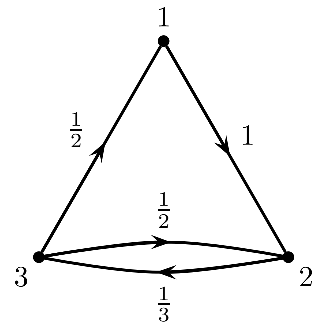
and *continuous time*

$$t \in \mathbb{R}^+ = [0, \infty).$$

The letters n, m, k will always denote integers, whereas t and s will refer to real numbers. Thus we write $(X_n)_{n\geq 0}$ for a discrete-time process and $(X_t)_{t\geq 0}$ for a continuous-time process.

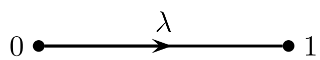
Markov chains are often best described by diagrams, of which we now give some simple examples:

(i) (*Discrete time*)



You move from state 1 to state 2 with probability 1. From state 3 you move either to 1 or to 2 with equal probability 1/2, and from 2 you jump to 3 with probability 1/3, otherwise stay at 2. We might have drawn a loop from 2 to itself with label 2/3. But since the total probability on jumping from 2 must equal 1, this does not convey any more information and we prefer to leave the loops out.

(ii) (*Continuous time*)

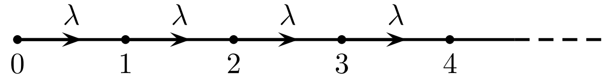


When in state 0 you wait for a random time with exponential distribution of parameter $\lambda \in (0, \infty)$, then jump to 1. Thus the density function of the waiting time T is given by

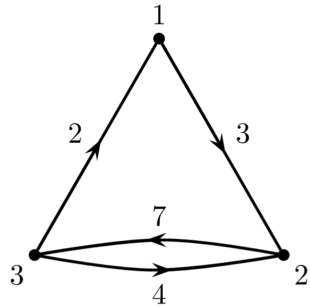
$$f_T(t) = \lambda e^{-\lambda t} \qquad \text{for } t \geq 0.$$

We write $T \sim E(\lambda)$ for short.

(iii) (*Continuous time*)



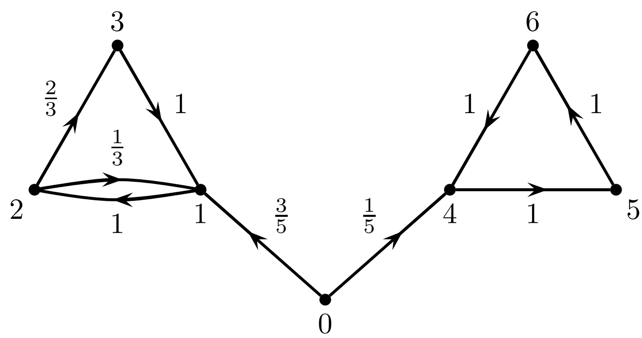
Here, when you get to 1 you do not stop but after another independent exponential time of parameter λ jump to 2, and so on. The resulting process is called the *Poisson process of rate* λ.



(iv) (*Continuous time*)

In state 3 you take two independent exponential times $T_1 \sim E(2)$ and $T_2 \sim E(4)$; if T_1 is the smaller you go to 1 after time T_1, and if T_2 is the smaller you go to 2 after time T_2. The rules for states 1 and 2 are as given in examples (ii) and (iii). It is a simple matter to show that the time spent in 3 is exponential of parameter $2 + 4 = 6$, and that the probability of jumping from 3 to 1 is $2/(2 + 4) = 1/3$. The details are given later.

(v) (*Discrete time*)



We use this example to anticipate some of the ideas discussed in detail in Chapter 1. The states may be partitioned into *communicating classes*, namely $\{0\}$, $\{1, 2, 3\}$ and $\{4, 5, 6\}$. Two of these classes are *closed*, meaning that you cannot escape. The closed classes here are *recurrent*, meaning that you return again and again to every state. The class $\{0\}$ is *transient*. The class $\{4, 5, 6\}$ is *periodic*, but $\{1, 2, 3\}$ is not. We shall show how to establish the following facts by solving some simple linear equations. You might like to try from first principles.

(a) Starting from 0, the probability of hitting 6 is $1/4$.

(b) Starting from 1, the probability of hitting 3 is 1.

(c) Starting from 1, it takes on average three steps to hit 3.

(d) Starting from 1, the long-run proportion of time spent in 2 is $3/8$.

Let us write $p_{ij}^{(n)}$ for the probability starting from i of being in state j after n steps. Then we have:

(e) $\lim_{n\to\infty} p_{01}^{(n)} = 9/32$;

(f) $p_{04}^{(n)}$ does not converge as $n \to \infty$;

(g) $\lim_{(n)\to\infty} p_{04}^{(3n)} = 1/124$.

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

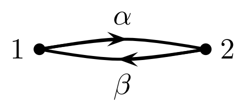
Let I be a countable set. Each $i \in I$ is called a *state* and I is called the *state-space*. We say that $\lambda = (\lambda_i : i \in I)$ is a *measure* on I if $0 \leq \lambda_i < \infty$ for all $i \in I$. If in addition the *total mass* $\sum_{i \in I} \lambda_i$ equals 1, then we call λ a *distribution*. 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

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

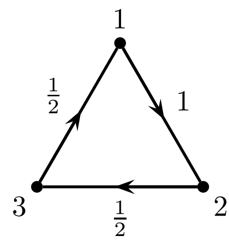
Then λ defines a distribution, the *distribution of* X. We think of X as modelling a random state which takes the value i with probability λ_i. There is a brief review of some basic facts about countable sets and probability spaces in Chapter 6.

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 $X_0, \dots, X_{n-1}$.

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

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

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 $n = 0, \dots, N-1$, 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* $i_1, \dots, i_N \in I$

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

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

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

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

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

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

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

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

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 $(X_n)_{n\ge 0}$ be Markov(λ, P). Then, conditional on $X_m = i$, $(X_{m+n})_{n\ge 0}$ is Markov(δ_i, P) and is independent of the random variables $X_0, \ldots, X_m$.*

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

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

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

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

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. □

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 $\mathbb{P}_i$, $(X_n)_{n \ge 0}$ is Markov(δ_i, P). So the behaviour of $(X_n)_{n \ge 0}$ under $\mathbb{P}_i$ does not depend on λ.

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

(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). □

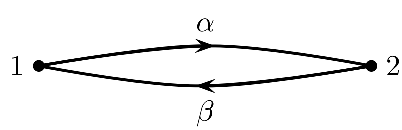
In light of this theorem we call $p_{ij}^{(n)}$ the *n-step transition probability from i to j*. The following examples give some methods for calculating $p_{ij}^{(n)}$.

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)

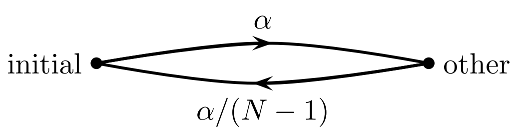
Suppose a virus can exist in N different strains and in each generation either stays the same, or with probability α mutates to another strain, which is chosen at random. What is the probability that the strain in the nth generation is the same as that in the 0th?

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



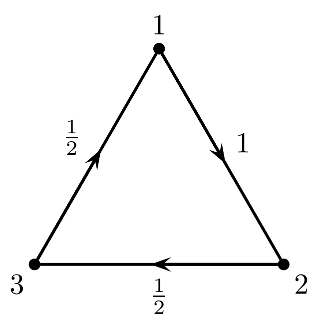
and by putting $\beta = \alpha/(N-1)$ in Example 1.1.4 we find that the desired probability is

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

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

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

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

State 1	$\frac{1}{2}$	$\frac{1}{2}$	0
State 2	$\frac{1}{2}$	$\frac{1}{12}$	$\frac{5}{12}$
State 3	0	0	1.

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*.

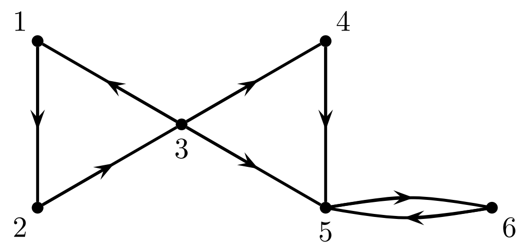
As the following example makes clear, when one can draw the diagram, the class structure of a chain is very easy to find.

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 $H^A : \Omega \to \{0, 1, 2, \dots\} \cup \{\infty\}$ 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

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

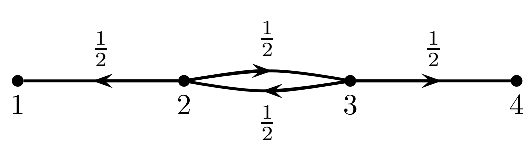
We shall often write less formally

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

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

$$\begin{aligned} x_i = \mathbb{P}_i(X_1 \in A) &+ \ldots + \mathbb{P}_i(X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A) \\ &+ \sum_{j_1\notin A} \cdots \sum_{j_n\notin A} p_{ij_1}p_{j_1j_2}\ldots p_{j_{n-1}j_n}x_{j_n}. \end{aligned}$$

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

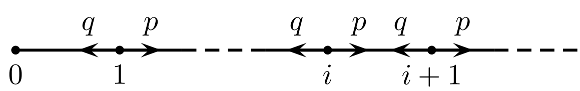
$$h_2 = \tfrac{1}{3} + \tfrac{2}{3}h_1, \; h_3 = \tfrac{2}{3} + \tfrac{1}{3}h_1.$$

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

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

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

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

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

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

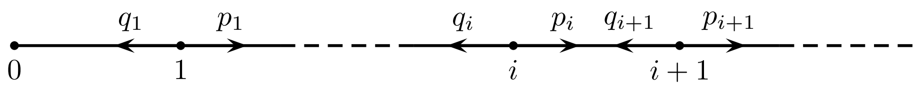
for a non-negative solution we must have $A \geq 0$, so the minimal non-negative solution is $h_i = (q/p)^i$. Finally, if $p = q$ the recurrence relation has a general solution

$$h_i = A + Bi$$

and again the restriction $0 \le h_i \le 1$ 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 $i = 1, 2, \dots$, 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

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

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

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

where the final equality defines γ_i. Then

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

so

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

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 $0 \le h_i \le 1$ 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

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

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 $i = 1, 2, \ldots$, 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 $k^A = (k_i^A : i \in I)$ 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

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

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

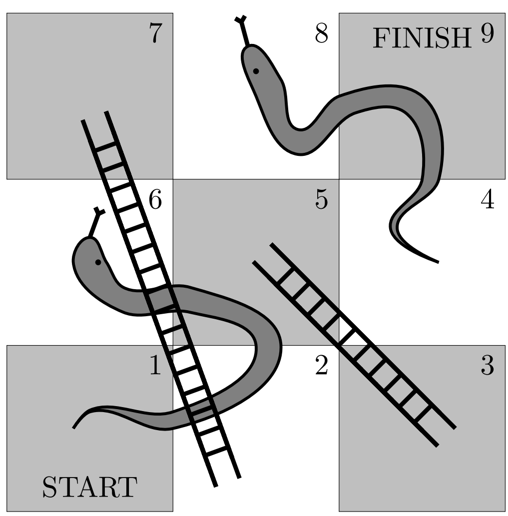
1.3.1 Prove the claims (a), (b) and (c) made in example (v) of the Introduction.

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 $X_0, X_1, \ldots, X_T$, then $B \cap \{T = m\}$ is determined by $X_0, X_1, \ldots, X_m$, for all $m = 0, 1, 2, \ldots$.

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 $X_0, X_1, \ldots, X_T$.*

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

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

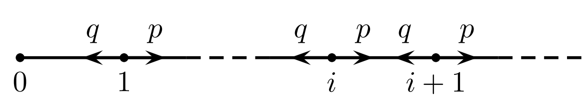
where we have used the condition $T = m$ to replace m by T. Now sum over $m = 0, 1, 2, \ldots$ and divide by $\mathbb{P}(T < \infty, X_T = i)$ to obtain

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

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

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

(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 $p \le q$, 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 $(X_n)_{n \ge 0}$ 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 $(Y_m)_{m \ge 0}$ may be obtained formally by setting $Y_m = X_{T_m}$, where

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

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

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

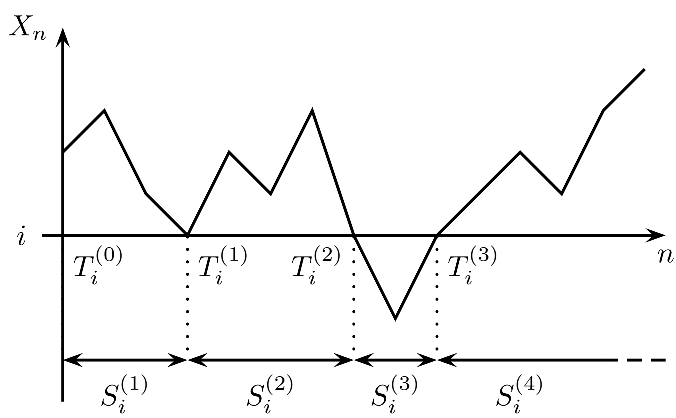
and, for $r = 0, 1, 2, \dots,$

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

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:

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

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. □

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

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

by Theorem 1.5.3. Hence j is also transient by Theorem 1.5.3. □

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 $m \geq 1$ 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. □

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

Proof. Suppose C is closed and finite and that $(X_n)_{n\geq 0}$ 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. □

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 $F_0, \ldots, F_n$ 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 $(3-\sqrt{5})/2$.

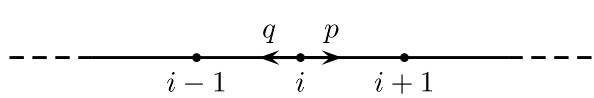
Deduce that $(X_n)_{n\geq 0}$ is transient. Show that, moreover, with probability 1, $F_n \to \infty$ as $n \to \infty$.

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

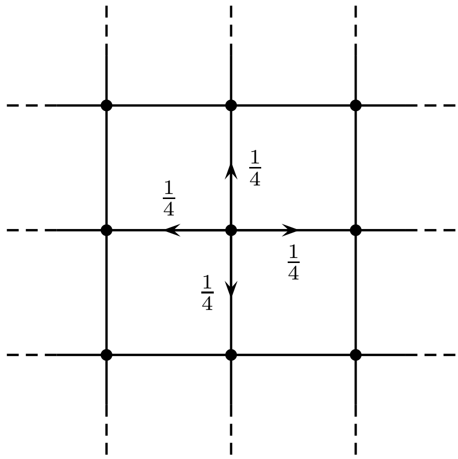
$$\sum_{n=N}^{\infty} p_{00}^{(2n)} \geq \frac{1}{2A} \sum_{n=N}^{\infty} \frac{1}{\sqrt{n}} = \infty$$

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.

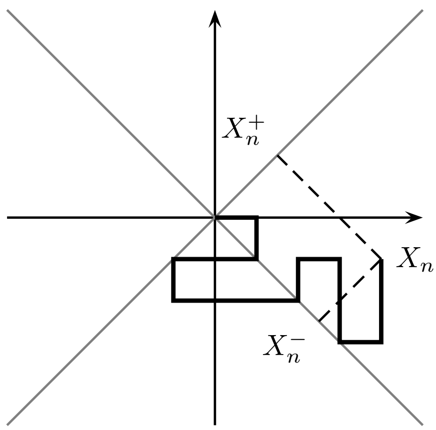
Example 1.6.2 (Simple symmetric random walk on $\mathbb{Z}^2$)
The simple symmetric random walk on $\mathbb{Z}^2$ has diagram



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*

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

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

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

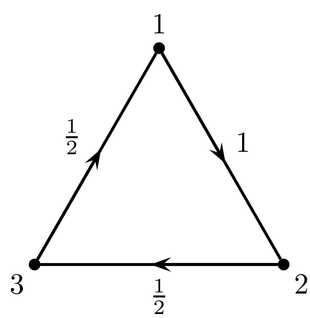
Ignore the trivial cases $\alpha = \beta = 0$ and $\alpha = \beta = 1$. Then, by Example 1.1.4

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

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 $(X_n)_{n \ge 0}$ 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) $\Rightarrow$ (i) armed with the knowledge that P is recurrent, so $\lambda = \gamma^k$ and the inequality (1.7) is in fact an equality. $\square$

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

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

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. □

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

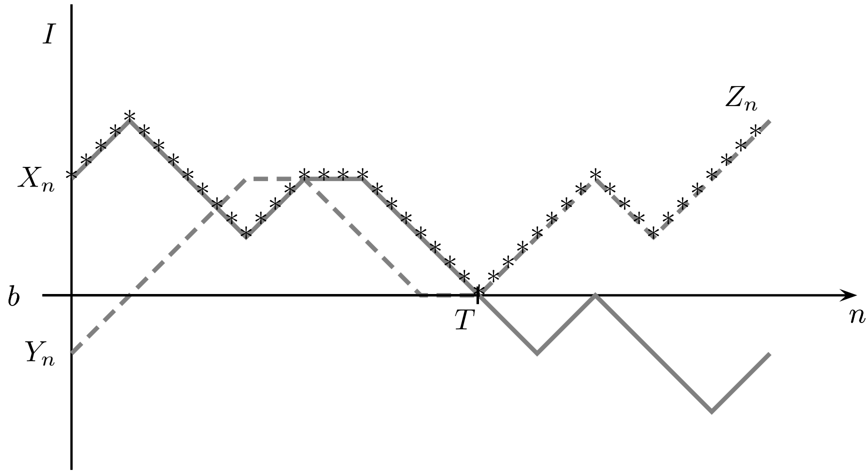
$$\widetilde{\pi}_{(i,k)} = \pi_i \pi_k$$

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), $(X_1, Y_1), \ldots, (X_T, Y_T)$. 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 $(X_0, Y_0), (X_1, Y_1), \ldots, (X_T, Y_T)$. 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:

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

Determine

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

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

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

(*Think of* $Y_1, Y_2, \ldots$ *as light-bulb lifetimes. A bulb is replaced when it fails. Thus the limiting probability that a bulb is replaced at time* n *is* $1/\mu$. *Although this appears to be a very special case of convergence to equilibrium, one can actually recover the full result by applying the renewal theorem to the excursion lengths* $S_i^{(1)}, S_i^{(2)}, \ldots$ *from state* i.)

1.9 Time reversal

For Markov chains, the past and future are independent given the present. This property is symmetrical in time and suggests looking at Markov chains with time running backwards. On the other hand, convergence to equilibrium shows behaviour which is asymmetrical in time: a highly organised state such as a point mass decays to a disorganised one, the invariant distribution. This is an example of entropy increasing. It suggests that if we want complete time-symmetry we must begin in equilibrium. The next result shows that a Markov chain in equilibrium, run backwards, is again a Markov chain. The transition matrix may however be different.

Theorem 1.9.1. *Let* P *be irreducible and have an invariant distribution* π. *Suppose that* $(X_n)_{0\le n\le N}$ *is Markov*(π, P) *and set* $Y_n = X_{N-n}$. *Then* $(Y_n)_{0\le n\le N}$ *is Markov*$(\pi, \widehat{P})$, *where* $\widehat{P} = (\widehat{p}_{ij})$ *is given by*

$$\pi_j \widehat{p}_{ji} = \pi_i p_{ij} \quad \text{for all } i, j$$

and $\widehat{P}$ *is also irreducible with invariant distribution* π.

Proof. First we check that $\widehat{P}$ is a stochastic matrix:

$$\sum_{i\in I} \widehat{p}_{ji} = \frac{1}{\pi_j} \sum_{i\in I} \pi_i p_{ij} = 1$$

since π is invariant for P. Next we check that π is invariant for $\widehat{P}$:

$$\sum_{j\in I} \pi_j \widehat{p}_{ji} = \sum_{j\in I} \pi_i p_{ij} = \pi_i$$

since P is a stochastic matrix.

We have

$$\begin{aligned}
&\mathbb{P}(Y_0 = i_1, Y_1 = i_2, \ldots, Y_N = i_N) \\
&\quad = \mathbb{P}(X_0 = i_N, X_1 = i_{N-1}, \ldots, X_N = i_1) \\
&\quad = \pi_{i_N} p_{i_N i_{N-1}} \ldots p_{i_2 i_1} = \pi_{i_1} \widehat{p}_{i_1 i_2} \ldots \widehat{p}_{i_{N-1} i_N}
\end{aligned}$$

so, by Theorem 1.1.1, $(Y_n)_{0\le n\le N}$ is Markov$(\pi, \widehat{P})$. Finally, since P is irreducible, for each pair of states i,j there is a chain of states $i_1 = i, i_2, \dots, i_{n-1}, i_n = j$ with $p_{i_1 i_2} \dots p_{i_{n-1} i_n} > 0$. Then

$$\widehat{p}_{i_n i_{n-1}} \dots \widehat{p}_{i_2 i_1} = \pi_{i_1} p_{i_1 i_2} \dots p_{i_{n-1} i_n} / \pi_{i_n} > 0$$

so $\widehat{P}$ is also irreducible. □

The chain $(Y_n)_{0\le n\le N}$ is called the *time-reversal* of $(X_n)_{0\le n\le N}$.

A stochastic matrix P and a measure λ are said to be in *detailed balance* if

$$\lambda_i p_{ij} = \lambda_j p_{ji} \quad \text{for all } i, j.$$

Though obvious, the following result is worth remembering because, when a solution λ to the detailed balance equations exists, it is often easier to find by the detailed balance equations than by the equation $\lambda = \lambda P$.

Lemma 1.9.2. *If P and λ are in detailed balance, then λ is invariant for P.*

Proof. We have $(\lambda P)_i = \sum_{j\in I} \lambda_j p_{ji} = \sum_{j\in I} \lambda_i p_{ij} = \lambda_i$. □

Let $(X_n)_{n\ge 0}$ be Markov(λ, P), with P irreducible. We say that $(X_n)_{n\ge 0}$ is *reversible* if, for all $N \ge 1$, $(X_{N-n})_{0\le n\le N}$ is also Markov(λ, P).

Theorem 1.9.3. *Let P be an irreducible stochastic matrix and let λ be a distribution. Suppose that $(X_n)_{n\ge 0}$ is Markov(λ, P). Then the following are equivalent:*

(a) *$(X_n)_{n\ge 0}$ is reversible;*

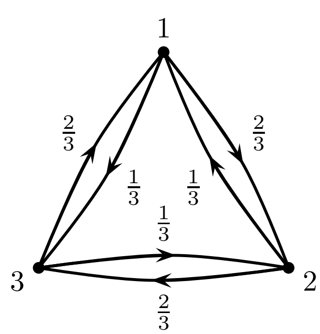
(b) *P and λ are in detailed balance.*

Proof. Both (a) and (b) imply that λ is invariant for P. Then both (a) and (b) are equivalent to the statement that $\widehat{P} = P$ in Theorem 1.9.1. □

We begin a collection of examples with a chain which is not reversible.

Example 1.9.4

Consider the Markov chain with diagram:



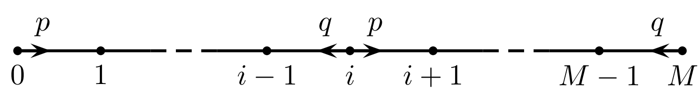
The transition matrix is

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

and $\pi = (1/3, 1/3, 1/3)$ is invariant. Hence $\widehat{P} = P^T$, the transpose of P. But P is not symmetric, so $P \neq \widehat{P}$ and this chain is not reversible. A patient observer would see the chain move clockwise in the long run: under time-reversal the clock would run backwards!

Example 1.9.5

Consider the Markov chain with diagram:



where $0 < p = 1 - q < 1$. The non-zero detailed balance equations read

$$\lambda_i p_{i,i+1} = \lambda_{i+1} p_{i+1,i} \quad \text{for } i = 0, 1, \dots, M-1.$$

So a solution is given by

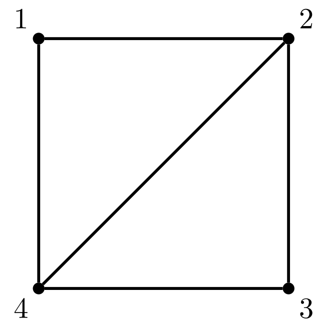
$$\lambda = \big((p/q)^i : i = 0, 1, \dots, M\big)$$

and this may be normalised to give a distribution in detailed balance with P. Hence this chain is reversible.

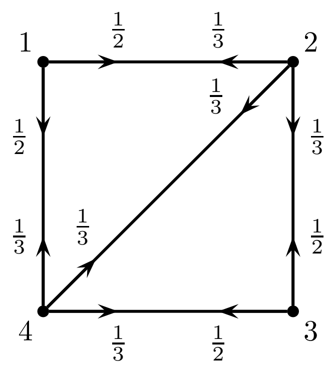
If p were much larger than q, one might argue that the chain would tend to move to the right and its time-reversal to the left. However, this ignores the fact that we reverse the chain *in equilibrium*, which in this case would be heavily concentrated near M. An observer would see the chain spending most of its time near M and making occasional brief forays to the left, which behaviour is symmetrical in time.

Example 1.9.6 (Random walk on a graph)

A *graph* G is a countable collection of states, usually called *vertices*, some of which are joined by *edges*, for example:



Thus a graph is a partially drawn Markov chain diagram. There is a natural way to complete the diagram which gives rise to the random walk on G. The *valency* v_i of vertex i is the number of edges at i. We have to assume that every vertex has finite valency. The random walk on G picks edges with equal probability:



Thus the transition probabilities are given by

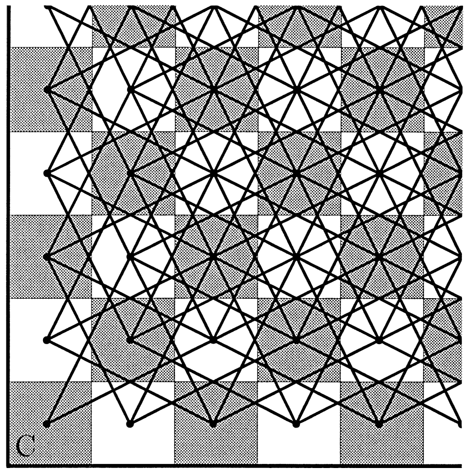
$$p_{ij} = \begin{cases} 1/v_i & \text{if } (i,j) \text{ is an edge} \\ 0 & \text{otherwise.} \end{cases}$$

We assume G is connected, so that P is irreducible. It is easy to see that P is in detailed balance with $v = (v_i : i \in G)$. So, if the total valency $\sigma = \sum_{i \in G} v_i$ is finite, then $\pi = v/\sigma$ is invariant and P is reversible.

Example 1.9.7 (Random chessboard knight)

A random knight makes each permissible move with equal probability. If it starts in a corner, how long on average will it take to return?

This is an example of a random walk on a graph: the vertices are the squares of the chessboard and the edges are the moves that the knight can take:



The diagram shows a part of the graph. We know by Theorem 1.7.7 and the preceding example that

$$\mathbb{E}_c(T_c) = 1/\pi_c = \sum_i (v_i/v_c)$$

so all we have to do is identify valencies. The four corner squares have valency 2, and the eight squares adjacent to the corners have valency 3. There are 20 squares of valency 4, 16 of valency 6, and the 16 central squares have valency 8. Hence

$$\mathbb{E}_c(T_c) = \frac{8 + 24 + 80 + 96 + 128}{2} = 168.$$

Alternatively, if you enjoy solving sets of 64 simultaneous linear equations, you might try finding π from $\pi P = \pi$, or calculating $\mathbb{E}_c(T_c)$ using Theorem 1.3.5!

Exercises

1.9.1 In each of the following cases determine whether the stochastic matrix P, which you may assume is irreducible, is reversible:

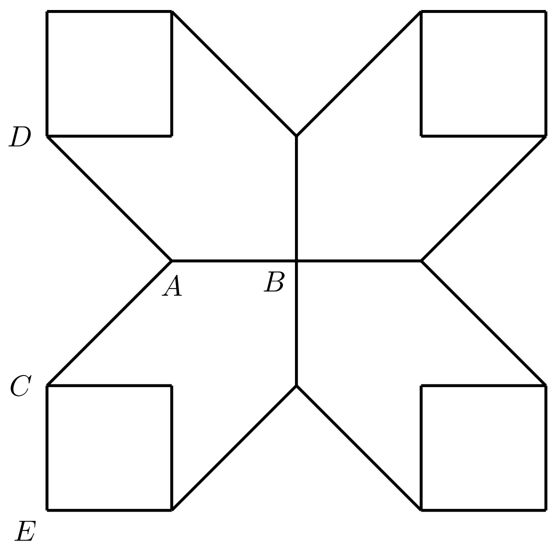
(a) $\begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$; (b) $\begin{pmatrix} 0 & p & 1-p \\ 1-p & 0 & p \\ p & 1-p & 0 \end{pmatrix}$;

(c) $I = \{0, 1, \dots, N\}$ and $p_{ij} = 0$ if $|j - i| \geq 2$;

(d) $I = \{0, 1, 2, \dots\}$ and $p_{01} = 1$, $p_{i,i+1} = p$, $p_{i,i-1} = 1 - p$ for $i \geq 1$;

(e) $p_{ij} = p_{ji}$ for all $i, j \in S$.

1.9.2 Two particles X and Y perform independent random walks on the graph shown in the diagram. So, for example, a particle at A jumps to B, C or D with equal probability 1/3.



Find the probability that X and Y ever meet at a vertex in the following cases:

(a) X starts at A and Y starts at B;

(b) X starts at A and Y starts at E.

For $I = B, D$ let M_I denote the expected time, when both X and Y start at I, until they are once again both at I. Show that $9M_D = 16M_B$.

1.10 Ergodic theorem

Ergodic theorems concern the limiting behaviour of averages over time. We shall prove a theorem which identifies for Markov chains the long-run proportion of time spent in each state. An essential tool is the following ergodic theorem for independent random variables which is a version of the strong law of large numbers.

Theorem 1.10.1 (Strong law of large numbers). *Let $Y_1, Y_2, \dots$ be a sequence of independent, identically distributed, non-negative random* *variables with* $\mathbb{E}(Y_1) = \mu$*. Then*

$$\mathbb{P}\left(\frac{Y_1 + \ldots + Y_n}{n} \to \mu \text{ as } n \to \infty\right) = 1.$$

Proof. A proof for the case $\mu < \infty$ may be found, for example, in *Probability with Martingales* by David Williams (Cambridge University Press, 1991). The case where $\mu = \infty$ is a simple deduction. Fix $N < \infty$ and set $Y_n^{(N)} = Y_n \wedge N$. Then

$$\frac{Y_1 + \ldots + Y_n}{n} \geq \frac{Y_1^{(N)} + \ldots + Y_n^{(N)}}{n} \to \mathbb{E}(Y_1 \wedge N) \qquad \text{as } n \to \infty$$

with probability one. As $N \uparrow \infty$ we have $\mathbb{E}(Y_1 \wedge N) \uparrow \mu$ by monotone convergence (see Section 6.4). So we must have, with probability 1

$$\frac{Y_1 + \ldots + Y_n}{n} \to \infty \qquad \text{as } n \to \infty. \qquad \square$$

We denote by $V_i(n)$ the *number of visits to* i *before* n:

$$V_i(n) = \sum_{k=0}^{n-1} 1_{\{X_k = i\}}.$$

Then $V_i(n)/n$ is the proportion of time before n spent in state i. The following result gives the long-run proportion of time spent by a Markov chain in each state.

Theorem 1.10.2 (Ergodic theorem). *Let* P *be irreducible and let* λ *be any distribution. If* $(X_n)_{n \geq 0}$ *is Markov*(λ, P) *then*

$$\mathbb{P}\left(\frac{V_i(n)}{n} \to \frac{1}{m_i} \text{ as } n \to \infty\right) = 1$$

where $m_i = \mathbb{E}_i(T_i)$ *is the expected return time to state* i*. Moreover, in the positive recurrent case, for any bounded function* $f : I \to \mathbb{R}$ *we have*

$$\mathbb{P}\left(\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) \to \overline{f} \text{ as } n \to \infty\right) = 1$$

where

$$\overline{f} = \sum_{i \in I} \pi_i f_i$$

and where $(\pi_i : i \in I)$ *is the unique invariant distribution.*

Proof. If P is transient, then, with probability 1, the total number V_i of visits to i is finite, so

$$\frac{V_i(n)}{n} \le \frac{V_i}{n} \to 0 = \frac{1}{m_i}.$$

Suppose then that P is recurrent and fix a state i. For $T = T_i$ we have $\mathbb{P}(T < \infty) = 1$ by Theorem 1.5.7 and $(X_{T+n})_{n\ge 0}$ is Markov(δ_i, P) and independent of $X_0, X_1, \dots, X_T$ by the strong Markov property. The long-run proportion of time spent in i is the same for $(X_{T+n})_{n\ge 0}$ and $(X_n)_{n\ge 0}$, so it suffices to consider the case $\lambda = \delta_i$.

Write $S_i^{(r)}$ for the length of the rth excursion to i, as in Section 1.5. By Lemma 1.5.1, the non-negative random variables $S_i^{(1)}, S_i^{(2)}, \dots$ are independent and identically distributed with $\mathbb{E}_i(S_i^{(r)}) = m_i$. Now

$$S_i^{(1)} + \dots + S_i^{(V_i(n)-1)} \le n - 1,$$

the left-hand side being the time of the last visit to i before n. Also

$$S_i^{(1)} + \dots + S_i^{(V_i(n))} \ge n,$$

the left-hand side being the time of the first visit to i after $n - 1$. Hence

$$\frac{S_i^{(1)} + \dots + S_i^{(V_i(n)-1)}}{V_i(n)} \le \frac{n}{V_i(n)} \le \frac{S_i^{(1)} + \dots + S_i^{(V_i(n))}}{V_i(n)}. \tag{1.8}$$

By the strong law of large numbers

$$\mathbb{P}\left(\frac{S_i^{(1)} + \dots + S_i^{(n)}}{n} \to m_i \text{ as } n \to \infty\right) = 1$$

and, since P is recurrent

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

So, letting $n \to \infty$ in (1.8), we get

$$\mathbb{P}\left(\frac{n}{V_i(n)} \to m_i \text{ as } n \to \infty\right) = 1,$$

which implies

$$\mathbb{P}\left(\frac{V_i(n)}{n} \to \frac{1}{m_i} \text{ as } n \to \infty\right) = 1.$$

Assume now that $(X_n)_{n\geq 0}$ has an invariant distribution $(\pi_i : i \in I)$. Let $f : I \to \mathbb{R}$ be a bounded function and assume without loss of generality that $|f| \leq 1$. For any $J \subseteq I$ we have

$$\begin{aligned}\left|\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) - \overline{f}\right| &= \left|\sum_{i\in I}\left(\frac{V_i(n)}{n} - \pi_i\right) f_i\right| \\ &\leq \sum_{i\in J}\left|\frac{V_i(n)}{n} - \pi_i\right| + \sum_{i\notin J}\left|\frac{V_i(n)}{n} - \pi_i\right| \\ &\leq \sum_{i\in J}\left|\frac{V_i(n)}{n} - \pi_i\right| + \sum_{i\notin J}\left(\frac{V_i(n)}{n} + \pi_i\right) \\ &\leq 2\sum_{i\in J}\left|\frac{V_i(n)}{n} - \pi_i\right| + 2\sum_{i\notin J}\pi_i.\end{aligned}$$

We proved above that

$$\mathbb{P}\left(\frac{V_i(n)}{n} \to \pi_i \text{ as } n \to \infty \text{ for all } i\right) = 1.$$

Given $\varepsilon > 0$, choose J finite so that

$$\sum_{i\notin J}\pi_i < \varepsilon/4$$

and then $N = N(\omega)$ so that, for $n \geq N(\omega)$

$$\sum_{i\in J}\left|\frac{V_i(n)}{n} - \pi_i\right| < \varepsilon/4.$$

Then, for $n \geq N(\omega)$, we have

$$\left|\frac{1}{n}\sum_{k=0}^{n-1} f(X_k) - \overline{f}\right| < \varepsilon,$$

which establishes the desired convergence. □

We consider now the statistical problem of estimating an unknown transition matrix P on the basis of observations of the corresponding Markov chain. Consider, to begin, the case where we have $N+1$ observations $(X_n)_{0\leq n\leq N}$. The log-likelihood function is given by

$$l(P) = \log(\lambda_{X_0} p_{X_0 X_1} \dots p_{X_{N-1} X_N}) = \sum_{i,j\in I} N_{ij}\log p_{ij}$$

up to a constant independent of P, where N_{ij} is the number of transitions from i to j. A standard statistical procedure is to find the *maximum likelihood estimate* $\widehat{P}$, which is the choice of P maximizing $l(P)$. Since P must satisfy the linear constraint $\sum_j p_{ij} = 1$ for each i, we first try to maximize

$$l(P) + \sum_{i,j \in I} \mu_i p_{ij}$$

and then choose $(\mu_i : i \in I)$ to fit the constraints. This is the method of Lagrange multipliers. Thus we find

$$\widehat{p}_{ij} = \sum_{n=0}^{N-1} 1_{\{X_n = i, X_{n+1} = j\}} \Big/ \sum_{n=0}^{N-1} 1_{\{X_n = i\}}$$

which is the proportion of jumps from i which go to j.

We now turn to consider the *consistency* of this sort of estimate, that is to say whether $\widehat{p}_{ij} \to p_{ij}$ with probability 1 as $N \to \infty$. Since this is clearly false when i is transient, we shall slightly modify our approach. Note that to find $\widehat{p}_{ij}$ we simply have to maximize

$$\sum_{j \in I} N_{ij} \log p_{ij}$$

subject to $\sum_j p_{ij} = 1$: the other terms and constraints are irrelevant. Suppose then that instead of $N+1$ observations we make enough observations to ensure the chain leaves state i a total of N times. In the transient case this may involve restarting the chain several times. Denote again by N_{ij} the number of transitions from i to j.

To maximize the likelihood for $(p_{ij} : j \in I)$ we still maximize

$$\sum_{j \in I} N_{ij} \log p_{ij}$$

subject to $\sum_j p_{ij} = 1$, which leads to the maximum likelihood estimate

$$\widehat{p}_{ij} = N_{ij}/N.$$

But $N_{ij} = Y_1 + \ldots + Y_N$, where $Y_n = 1$ if the nth transition from i is to j, and $Y_n = 0$ otherwise. By the strong Markov property $Y_1, \ldots, Y_N$ are independent and identically distributed random variables with mean p_{ij}. So, by the strong law of large numbers

$$\mathbb{P}(\widehat{p}_{ij} \to p_{ij} \text{ as } N \to \infty) = 1,$$

which shows that $\widehat{p}_{ij}$ is consistent.

Exercises

1.10.1 Prove the claim (d) made in example (v) of the Introduction.

1.10.2 A professor has N umbrellas. He walks to the office in the morning and walks home in the evening. If it is raining he likes to carry an umbrella and if it is fine he does not. Suppose that it rains on each journey with probability p, independently of past weather. What is the long-run proportion of journeys on which the professor gets wet?

1.10.3 Let $(X_n)_{n\geq 0}$ be an irreducible Markov chain on I having an invariant distribution π. For $J \subseteq I$ let $(Y_m)_{m\geq 0}$ be the Markov chain on J obtained by observing $(X_n)_{n\geq 0}$ whilst in J. (See Example 1.4.4.) Show that $(Y_m)_{m\geq 0}$ is positive recurrent and find its invariant distribution.

1.10.4 An opera singer is due to perform a long series of concerts. Having a fine artistic temperament, she is liable to pull out each night with probability 1/2. Once this has happened she will not sing again until the promoter convinces her of his high regard. This he does by sending flowers every day until she returns. Flowers costing x thousand pounds, $0 \leq x \leq 1$, bring about a reconciliation with probability $\sqrt{x}$. The promoter stands to make £750 from each successful concert. How much should he spend on flowers?

1.11 Appendix: recurrence relations

Recurrence relations often arise in the linear equations associated to Markov chains. Here is an account of the simplest cases. A more specialized case was dealt with in Example 1.3.4. In Example 1.1.4 we found a recurrence relation of the form

$$x_{n+1} = ax_n + b.$$

We look first for a constant solution $x_n = x$; then $x = ax + b$, so provided $a \neq 1$ we must have $x = b/(1-a)$. Now $y_n = x_n - b/(1-a)$ satisfies $y_{n+1} = ay_n$, so $y_n = a^n y_0$. Thus the general solution when $a \neq 1$ is given by

$$x_n = Aa^n + b/(1-a)$$

where A is a constant. When $a = 1$ the general solution is obviously

$$x_n = x_0 + nb.$$

In Example 1.3.3 we found a recurrence relation of the form

$$ax_{n+1} + bx_n + cx_{n-1} = 0$$

where a and c were both non-zero. Let us try a solution of the form $x_n = \lambda^n$; then $a\lambda^2 + b\lambda + c = 0$. Denote by α and β the roots of this quadratic. Then

$$y_n = A\alpha^n + B\beta^n$$

is a solution. If $\alpha \neq \beta$ then we can solve the equations

$$x_0 = A + B, \quad x_1 = A\alpha + B\beta$$

so that $y_0 = x_0$ and $y_1 = x_1$; but

$$a(y_{n+1} - x_{n+1}) + b(y_n - x_n) + c(y_{n-1} - x_{n-1}) = 0$$

for all n, so by induction $y_n = x_n$ for all n. If $\alpha = \beta \neq 0$, then

$$y_n = (A + nB)\alpha^n$$

is a solution and we can solve

$$x_0 = A\alpha^n, \quad x_1 = (A + B)\alpha^n$$

so that $y_0 = x_0$ and $y_1 = x_1$; then, by the same argument, $y_n = x_n$ for all n. The case $\alpha = \beta = 0$ does not arise. Hence the general solution is given by

$$x_n = \begin{cases} A\alpha^n + B\beta^n & \text{if } \alpha \neq \beta \\ (A + nB)\alpha^n & \text{if } \alpha = \beta. \end{cases}$$

1.12 Appendix: asymptotics for $n!$

Our analysis of recurrence and transience for random walks in Section 1.6 rested heavily on the use of the asymptotic relation

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

for some $A \in [1, \infty)$. Here is a derivation.

We make use of the power series expansions for $|t| < 1$

$$\log(1 + t) = t - \tfrac{1}{2}t^2 + \tfrac{1}{3}t^3 - \dots$$
$$\log(1 - t) = -t - \tfrac{1}{2}t^2 - \tfrac{1}{3}t^3 - \dots .$$

By subtraction we obtain

$$\tfrac{1}{2}\log\left(\frac{1+t}{1-t}\right) = t + \tfrac{1}{3}t^3 + \tfrac{1}{5}t^5 + \dots .$$

Set $A_n = n!/(n^{n+1/2}e^{-n})$ and $a_n = \log A_n$. Then, by a straightforward calculation

$$a_n - a_{n+1} = (2n+1)\frac{1}{2}\log\left(\frac{1+(2n+1)^{-1}}{1-(2n+1)^{-1}}\right) - 1.$$

By the series expansion written above we have

$$\begin{aligned}
a_n - a_{n+1} &= (2n+1)\left\{\frac{1}{(2n+1)} + \frac{1}{3}\frac{1}{(2n+1)^3} + \frac{1}{5}\frac{1}{(2n+1)^5} + \dots\right\} - 1 \\
&= \frac{1}{3}\frac{1}{(2n+1)^2} + \frac{1}{5}\frac{1}{(2n+1)^4} + \dots \\
&\le \frac{1}{3}\left\{\frac{1}{(2n+1)^2} + \frac{1}{(2n+1)^4} + \dots\right\} \\
&= \frac{1}{3}\frac{1}{(2n+1)^2 - 1} = \frac{1}{12n} - \frac{1}{12(n+1)}.
\end{aligned}$$

It follows that a_n decreases and $a_n - 1/(12n)$ increases as $n \to \infty$. Hence $a_n \to a$ for some $a \in [0, \infty)$ and hence $A_n \to A$, as $n \to \infty$, where $A = e^a$.

2

Continuous-time Markov chains I

The material on continuous-time Markov chains is divided between this chapter and the next. The theory takes some time to set up, but once up and running it follows a very similar pattern to the discrete-time case. To emphasise this we have put the setting-up in this chapter and the rest in the next. If you wish, you can begin with Chapter 3, provided you take certain basic properties on trust, which are reviewed in Section 3.1. The first three sections of Chapter 2 fill in some necessary background information and are independent of each other. Section 2.4 on the Poisson process and Section 2.5 on birth processes provide a gentle warm-up for general continuous-time Markov chains. These processes are simple and particularly important examples of continuous-time chains. Sections 2.6–2.8, especially 2.8, deal with the heart of the continuous-time theory. There is an irreducible level of difficulty at this point, so we advise that Sections 2.7 and 2.8 are read selectively at first. Some examples of more general processes are given in Section 2.9. As in Chapter 1 the exercises form an important part of the text.

2.1 Q-matrices and their exponentials

In this section we shall discuss some of the basic properties of Q-matrices and explain their connection with continuous-time Markov chains.

Let I be a countable set. A *Q-matrix* on I is a matrix $Q = (q_{ij} : i, j \in I)$ satisfying the following conditions:

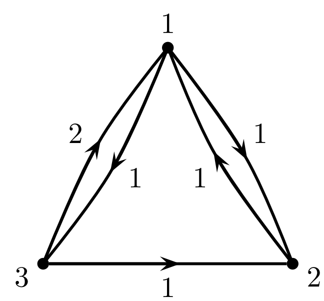
(i) $0 \le -q_{ii} < \infty$ for all i;
(ii) $q_{ij} \ge 0$ for all $i \ne j$;
(iii) $\sum_{j \in I} q_{ij} = 0$ for all i.

Thus in each row of Q we can choose the off-diagonal entries to be any non-negative real numbers, subject only to the constraint that the off-diagonal row sum is finite:

$$q_i = \sum_{j \ne i} q_{ij} < \infty.$$

The diagonal entry q_{ii} is then $-q_i$, making the total row sum zero.

A convenient way to present the data for a continuous-time Markov chain is by means of a diagram, for example:



Each diagram then corresponds to a unique Q-matrix, in this case

$$Q = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -1 & 0 \\ 2 & 1 & -3 \end{pmatrix}.$$

Thus each off-diagonal entry q_{ij} gives the value we attach to the (i, j) arrow on the diagram, which we shall interpret later as the *rate of going from i to j*. The numbers q_i are not shown on the diagram, but you can work them out from the other information given. We shall later interpret q_i as the *rate of leaving i*.

We may think of the discrete parameter space $\{0, 1, 2, \dots\}$ as embedded in the continuous parameter space $[0, \infty)$. For $p \in (0, \infty)$ a natural way to interpolate the discrete sequence $(p^n : n = 0, 1, 2, \dots)$ is by the function $(e^{tq} : t \ge 0)$, where $q = \log p$. Consider now a *finite* set I and a matrix

$P = (p_{ij} : i, j \in I)$. Is there a natural way to fill in the gaps in the discrete sequence $(P^n : n = 0, 1, 2, \dots)$?

For any matrix $Q = (q_{ij} : i, j \in I)$, the series

$$\sum_{k=0}^{\infty} \frac{Q^k}{k!}$$

converges componentwise and we denote its limit by e^Q. Moreover, if two matrices Q_1 and Q_2 commute, then

$$e^{Q_1+Q_2} = e^{Q_1} e^{Q_2}.$$

The proofs of these assertions follow the scalar case closely and are given in Section 2.10. Suppose then that we can find a matrix Q with $e^Q = P$. Then

$$e^{nQ} = (e^Q)^n = P^n$$

so $(e^{tQ} : t \geq 0)$ fills in the gaps in the discrete sequence.

Theorem 2.1.1. *Let Q be a matrix on a finite set I. Set $P(t) = e^{tQ}$. Then $(P(t) : t \geq 0)$ has the following properties:*

(i) *$P(s+t) = P(s)P(t)$ for all s, t (semigroup property);*

(ii) *$(P(t) : t \geq 0)$ is the unique solution to the forward equation*

$$\frac{d}{dt}P(t) = P(t)Q, \qquad P(0) = I\,;$$

(iii) *$(P(t) : t \geq 0)$ is the unique solution to the backward equation*

$$\frac{d}{dt}P(t) = QP(t), \qquad P(0) = I\,;$$

(iv) *for $k = 0, 1, 2, \dots,$ we have*

$$\left(\frac{d}{dt}\right)^k \bigg|_{t=0} P(t) = Q^k.$$

Proof. For any $s, t \in \mathbb{R}$, sQ and tQ commute, so

$$e^{sQ} e^{tQ} = e^{(s+t)Q}$$

proving the semigroup property. The matrix-valued power series

$$P(t) = \sum_{k=0}^{\infty} \frac{(tQ)^k}{k!}$$

has infinite radius of convergence (see Section 2.10). So each component is differentiable with derivative given by term-by-term differentiation:

$$P'(t) = \sum_{k=1}^{\infty} \frac{t^{k-1}Q^k}{(k-1)!} = P(t)Q = QP(t).$$

Hence $P(t)$ satisfies the forward and backward equations. Moreover by repeated term-by-term differentiation we obtain (iv). It remains to show that $P(t)$ is the only solution of the forward and backward equations. But if $M(t)$ satisfies the forward equation, then

$$\begin{aligned}\frac{d}{dt}(M(t)e^{-tQ}) &= \left(\frac{d}{dt}M(t)\right)e^{-tQ} + M(t)\left(\frac{d}{dt}e^{-tQ}\right)\\ &= M(t)Qe^{-tQ} + M(t)(-Q)e^{-tQ} = 0\end{aligned}$$

so $M(t)e^{-tQ}$ is constant, and so $M(t) = P(t)$. A similar argument proves uniqueness for the backward equation. □

The last result was about matrix exponentials in general. Now let us see what happens to Q-matrices. Recall that a matrix $P = (p_{ij} : i, j \in I)$ is stochastic if it satisfies

(i) $0 \le p_{ij} < \infty$ for all i, j;

(ii) $\sum_{j \in I} p_{ij} = 1$ for all i.

We recall the convention that in the limit $t \to 0$ the statement $f(t) = O(t)$ means that $f(t)/t \le C$ for all sufficiently small t, for some $C < \infty$. Later we shall also use the convention that $f(t) = o(t)$ means $f(t)/t \to 0$ as $t \to 0$.

Theorem 2.1.2. *A matrix Q on a finite set I is a Q-matrix if and only if $P(t) = e^{tQ}$ is a stochastic matrix for all $t \ge 0$.*

Proof. As $t \downarrow 0$ we have

$$P(t) = I + tQ + O(t^2)$$

so $q_{ij} \ge 0$ for $i \ne j$ if and only if $p_{ij}(t) \ge 0$ for all i, j and $t \ge 0$ sufficiently small. Since $P(t) = P(t/n)^n$ for all n, it follows that $q_{ij} \ge 0$ for $i \ne j$ if and only if $p_{ij}(t) \ge 0$ for all i, j and *all* $t \ge 0$.

If Q has zero row sums then so does Q^n for every n:

$$\sum_{k \in I} q_{ik}^{(n)} = \sum_{k \in I} \sum_{j \in I} q_{ij}^{(n-1)} q_{jk} = \sum_{j \in I} q_{ij}^{(n-1)} \sum_{k \in I} q_{jk} = 0.$$

So

$$\sum_{j\in I} p_{ij}(t) = 1 + \sum_{n=1}^{\infty} \frac{t^n}{n!} \sum_{j\in I} q_{ij}^{(n)} = 1.$$

On the other hand, if $\sum_{j\in I} p_{ij}(t) = 1$ for all $t \geq 0$, then

$$\sum_{j\in I} q_{ij} = \frac{d}{dt}\Big|_{t=0} \sum_{j\in I} p_{ij}(t) = 0. \qquad \square$$

Now, if P is a stochastic matrix of the form e^Q for some Q-matrix, we can do some sort of filling-in of gaps at the level of processes. Fix some large integer m and let $(X_n^m)_{n\geq 0}$ be discrete-time Markov$(\lambda, e^{Q/m})$. We define a process indexed by $\{n/m : n = 0, 1, 2, \dots\}$ by

$$X_{n/m} = X_n^m.$$

Then $(X_n : n = 0, 1, 2, \dots)$ is discrete-time Markov$(\lambda, (e^{Q/m})^m)$ (see Exercise 1.1.2) and

$$(e^{Q/m})^m = e^Q = P.$$

Thus we can find discrete-time Markov chains with arbitrarily fine grids $\{n/m : n = 0, 1, 2, \dots\}$ as time-parameter sets which give rise to Markov(λ, P) when sampled at integer times. It should not then be too surprising that there is, as we shall see in Section 2.8, a continuous-time process $(X_t)_{t\geq 0}$ which also has this property.

To anticipate a little, we shall see in Section 2.8 that a continuous-time Markov chain $(X_t)_{t\geq 0}$ with Q-matrix Q satisfies

$$\mathbb{P}(X_{t_{n+1}} = i_{n+1} \mid X_{t_0} = i_0, \dots, X_{t_n} = i_n) = p_{i_n i_{n+1}}(t_{n+1} - t_n)$$

for all $n = 0, 1, 2, \dots$, all times $0 \leq t_0 \leq \dots \leq t_{n+1}$ and all states $i_0, \dots, i_{n+1}$, where $p_{ij}(t)$ is the (i, j) entry in e^{tQ}. In particular, the *transition probability* from i to j in time t is given by

$$\mathbb{P}_i(X_t = j) := \mathbb{P}(X_t = j \mid X_0 = i) = p_{ij}(t).$$

(Recall that := means 'defined to equal'.) You should compare this with the defining property of a discrete-time Markov chain given in Section 1.1. We shall now give some examples where the transition probabilities $p_{ij}(t)$ may be calculated explicitly.

Example 2.1.3

We calculate $p_{11}(t)$ for the continuous-time Markov chain with Q-matrix

$$Q = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -1 & 0 \\ 2 & 1 & -3 \end{pmatrix}.$$

The method is similar to that of Example 1.1.6. We begin by writing down the characteristic equation for Q:

$$0 = \det(x - Q) = x(x+2)(x+4).$$

This shows that Q has distinct eigenvalues $0, -2, -4$. Then $p_{11}(t)$ has the form

$$p_{11}(t) = a + be^{-2t} + ce^{-4t}$$

for some constants a, b and c. (This is because we could diagonalize Q by an invertible matrix U:

$$Q = U \begin{pmatrix} 0 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & -4 \end{pmatrix} U^{-1}.$$

Then

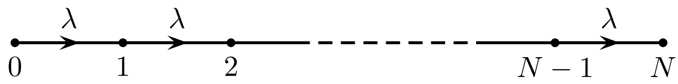
$$\begin{aligned} e^{tQ} &= \sum_{k=0}^{\infty} \frac{(tQ)^k}{k!} \\ &= U \sum_{k=0}^{\infty} \frac{1}{k!} \begin{pmatrix} 0^k & 0 & 0 \\ 0 & (-2t)^k & 0 \\ 0 & 0 & (-4t)^k \end{pmatrix} U^{-1} \\ &= U \begin{pmatrix} 1 & 0 & 0 \\ 0 & e^{-2t} & 0 \\ 0 & 0 & e^{-4t} \end{pmatrix} U^{-1}, \end{aligned}$$

so $p_{11}(t)$ must have the form claimed.) To determine the constants we use

$$\begin{aligned} 1 &= p_{11}(0) = a + b + c, \\ -2 &= q_{11} = p'_{11}(0) = -2b - 4c, \\ 7 &= q_{11}^{(2)} = p''_{11}(0) = 4b + 16c, \end{aligned}$$

so

$$p_{11}(t) = \tfrac{3}{8} + \tfrac{1}{4}e^{-2t} + \tfrac{3}{8}e^{-4t}.$$



Example 2.1.4

We calculate $p_{ij}(t)$ for the continuous-time Markov chain with diagram given above. The Q-matrix is

$$Q = \begin{pmatrix} -\lambda & \lambda & & & & \\ & -\lambda & \lambda & & & \\ & & \ddots & \ddots & & \\ & & & & \lambda & \\ & & & & -\lambda & \lambda \\ & & & & & 0 \end{pmatrix}$$

where entries off the diagonal and super-diagonal are all zero. The exponential of an upper-triangular matrix is upper-triangular, so $p_{ij}(t) = 0$ for $i > j$. In components the forward equation $P'(t) = P(t)Q$ reads

$$\begin{aligned} p'_{ii}(t) &= -\lambda p_{ii}(t), & p_{ii}(0) &= 1, & &\text{for } i < N, \\ p'_{ij}(t) &= -\lambda p_{ij}(t) + \lambda p_{i,j-1}(t), & p_{ij}(0) &= 0, & &\text{for } i < j < N, \\ p'_{iN}(t) &= \lambda p_{iN-1}(t), & p_{iN}(0) &= 0, & &\text{for } i < N. \end{aligned}$$

We can solve these equations. First, $p_{ii}(t) = e^{-\lambda t}$ for $i < N$. Then, for $i < j < N$

$$(e^{\lambda t} p_{ij}(t))' = e^{\lambda t} p_{i,j-1}(t)$$

so, by induction

$$p_{ij}(t) = e^{-\lambda t} \frac{(\lambda t)^{j-i}}{(j-i)!}.$$

If $i = 0$, these are the Poisson probabilities of parameter λt. So, starting from 0, the distribution of the Markov chain at time t is the same as the distribution of $\min\{Y_t, N\}$, where Y_t is a Poisson random variable of parameter λt.

Exercises

2.1.1 Compute $p_{11}(t)$ for $P(t) = e^{tQ}$, where

$$Q = \begin{pmatrix} -2 & 1 & 1 \\ 4 & -4 & 0 \\ 2 & 1 & -3 \end{pmatrix}.$$

2.1.2 Which of the following matrices is the exponential of a Q-matrix?

$$\text{(a)} \quad \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \text{(b)} \quad \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \qquad \text{(c)} \quad \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

What consequences do your answers have for the discrete-time Markov chains with these transition matrices?

2.2 Continuous-time random processes

Let I be a countable set. A *continuous-time random process*

$$(X_t)_{t\geq 0} = (X_t : 0 \leq t < \infty)$$

with values in I is a family of random variables $X_t : \Omega \to I$. We are going to consider ways in which we might specify the probabilistic behaviour (or *law*) of $(X_t)_{t\geq 0}$. These should enable us to find, at least in principle, any probability connected with the process, such as $\mathbb{P}(X_t = i)$ or $\mathbb{P}(X_{t_0} = i_0, \dots , X_{t_n} = i_n)$, or $\mathbb{P}(X_t = i \text{ for some } t)$. There are subtleties in this problem not present in the discrete-time case. They arise because, for a countable disjoint union

$$\mathbb{P}\left(\bigcup_n A_n\right) = \sum_n \mathbb{P}(A_n),$$

whereas for an uncountable union $\bigcup_{t\geq 0} A_t$ there is no such rule. To avoid these subtleties as far as possible we shall restrict our attention to processes $(X_t)_{t\geq 0}$ which are *right-continuous*. This means in this context that for all $\omega \in \Omega$ and $t \geq 0$ there exists $\varepsilon > 0$ such that

$$X_s(\omega) = X_t(\omega) \qquad \text{for } t \leq s \leq t + \varepsilon.$$

By a standard result of measure theory, which is proved in Section 6.6, the probability of any event depending on a right-continuous process can be determined from its *finite-dimensional distributions*, that is, from the probabilities

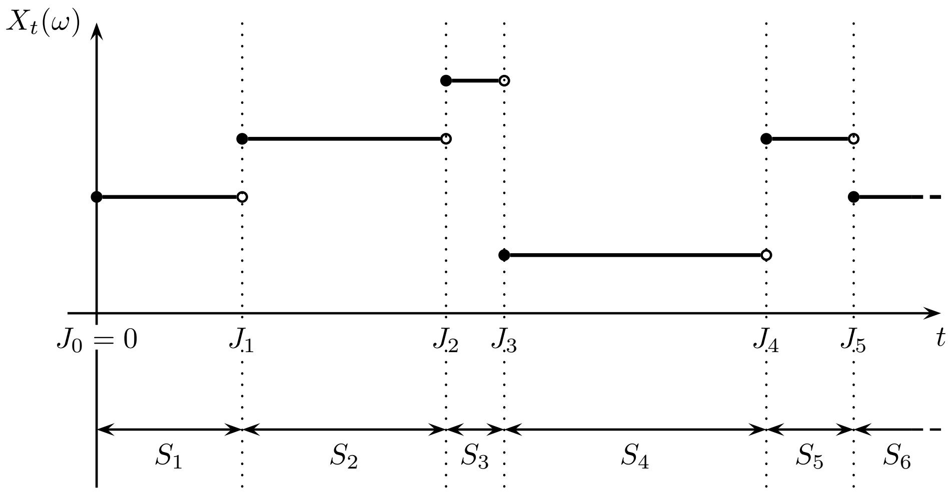
$$\mathbb{P}(X_{t_0} = i_0, X_{t_1} = i_1, \dots , X_{t_n} = i_n)$$

for $n \geq 0$, $0 \leq t_0 \leq t_1 \leq \dots \leq t_n$ and $i_0, \dots , i_n \in I$. For example

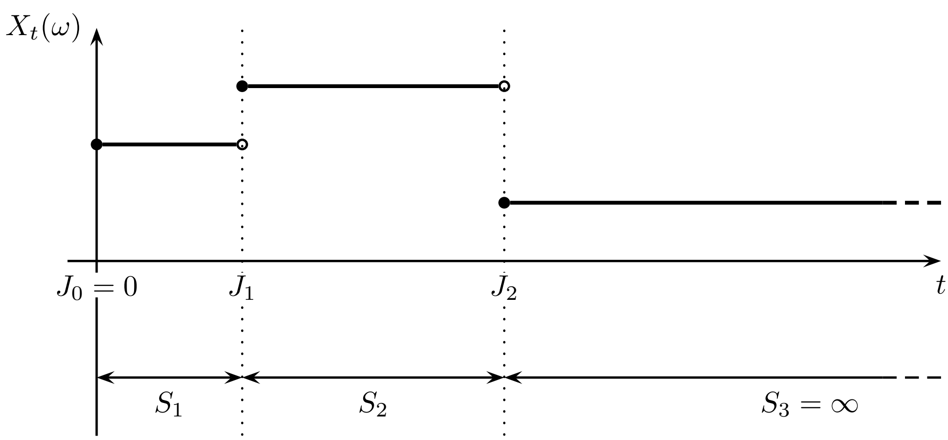
$$\mathbb{P}(X_t = i \text{ for some } t \in [0, \infty)) = 1 - \lim_{n\to\infty} \sum_{j_1,\dots,j_n \neq i} \mathbb{P}(X_{q_1} = j_i, \dots , X_{q_n} = j_n)$$

where $q_1, q_2, \dots$ is an enumeration of the rationals.

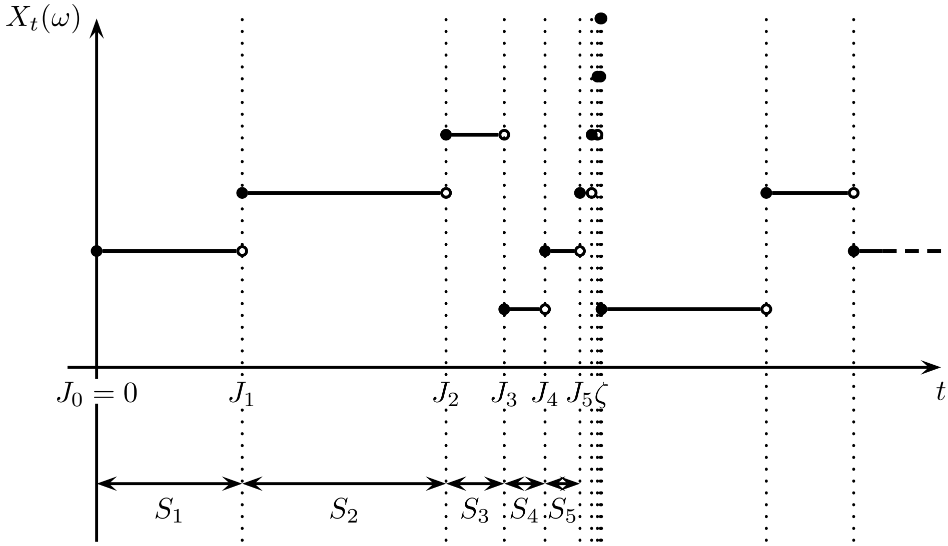
Every path $t \mapsto X_t(\omega)$ of a right-continuous process must remain constant for a while in each new state, so there are three possibilities for the sorts of path we get. In the first case the path makes infinitely many jumps, but only finitely many in any interval $[0, t]$:



The second case is where the path makes finitely many jumps and then becomes stuck in some state forever:



In the third case the process makes infinitely many jumps in a finite interval; this is illustrated below. In this case, after the explosion time ζ the process starts up again; it may explode again, maybe infinitely often, or it may not.



We call $J_0, J_1, \dots$ the *jump times* of $(X_t)_{t \geq 0}$ and $S_1, S_2, \dots$ the *holding times*. They are obtained from $(X_t)_{t \geq 0}$ by

$$J_0 = 0, \quad J_{n+1} = \inf\{t \geq J_n : X_t \neq X_{J_n}\}$$

for $n = 0, 1, \dots$, where $\inf \emptyset = \infty$, and, for $n = 1, 2, \dots$,

$$S_n = \begin{cases} J_n - J_{n-1} & \text{if } J_{n-1} < \infty \\ \infty & \text{otherwise.} \end{cases}$$

Note that right-continuity forces $S_n > 0$ for all n. If $J_{n+1} = \infty$ for some n, we define $X_\infty = X_{J_n}$, the final value, otherwise X_∞ is undefined. The (*first*) *explosion time* ζ is defined by

$$\zeta = \sup_n J_n = \sum_{n=1}^{\infty} S_n.$$

The discrete-time process $(Y_n)_{n \geq 0}$ given by $Y_n = X_{J_n}$ is called the *jump process* of $(X_t)_{t \geq 0}$, or the *jump chain* if it is a discrete-time Markov chain. This is simply the sequence of values taken by $(X_t)_{t \geq 0}$ up to explosion.

We shall not consider what happens to a process after explosion. So it is convenient to adjoin to I a new state, ∞ say, and require that $X_t = \infty$ if $t \geq \zeta$. Any process satisfying this requirement is called *minimal*. The terminology 'minimal' does not refer to the state of the process but to the interval of time over which the process is active. Note that a minimal process may be reconstructed from its holding times and jump process. Thus by specifying the joint distribution of $S_1, S_2, \dots$ and $(Y_n)_{n\geq 0}$ we have another 'countable' specification of the probabilistic behaviour of $(X_t)_{t\geq 0}$. For example, the probability that $X_t = i$ is given by

$$\mathbb{P}(X_t = i) = \sum_{n=0}^{\infty} \mathbb{P}(Y_n = i \text{ and } J_n \leq t < J_{n+1})$$

and

$$\mathbb{P}(X_t = i \text{ for some } t \in [0,\infty)) = \mathbb{P}(Y_n = i \text{ for some } n \geq 0).$$

2.3 Some properties of the exponential distribution

A random variable $T : \Omega \to [0,\infty]$ has *exponential distribution of parameter* λ $(0 \leq \lambda < \infty)$ if

$$\mathbb{P}(T > t) = e^{-\lambda t} \quad \text{for all } t \geq 0.$$

We write $T \sim E(\lambda)$ for short. If $\lambda > 0$, then T has density function

$$f_T(t) = \lambda e^{-\lambda t} 1_{t\geq 0}.$$

The mean of T is given by

$$\mathbb{E}(T) = \int_0^\infty \mathbb{P}(T > t)dt = \lambda^{-1}.$$

The exponential distribution plays a fundamental role in continuous-time Markov chains because of the following results.

Theorem 2.3.1 (Memoryless property). *A random variable $T : \Omega \to (0,\infty]$ has an exponential distribution if and only if it has the following memoryless property:*

$$\mathbb{P}(T > s + t \mid T > s) = \mathbb{P}(T > t) \quad \textit{for all } s, t \geq 0.$$

Proof. Suppose $T \sim E(\lambda)$, then

$$\mathbb{P}(T > s + t \mid T > s) = \frac{\mathbb{P}(T > s + t)}{\mathbb{P}(T > s)} = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = \mathbb{P}(T > t).$$

On the other hand, suppose T has the memoryless property whenever $\mathbb{P}(T > s) > 0$. Then $g(t) = \mathbb{P}(T > t)$ satisfies

$$g(s+t) = g(s)g(t) \quad \text{for all } s, t \geq 0.$$

We assumed $T > 0$ so that $g(1/n) > 0$ for some n. Then, by induction

$$g(1) = g\left(\frac{1}{n} + \ldots + \frac{1}{n}\right) = g\left(\frac{1}{n}\right)^n > 0$$

so $g(1) = e^{-\lambda}$ for some $0 \leq \lambda < \infty$. By the same argument, for integers $p, q \geq 1$

$$g(p/q) = g(1/q)^p = g(1)^{p/q}$$

so $g(r) = e^{-\lambda r}$ for all rationals $r > 0$. For real $t > 0$, choose rationals $r, s > 0$ with $r \leq t \leq s$. Since g is decreasing,

$$e^{-\lambda r} = g(r) \geq g(t) \geq g(s) = e^{-\lambda s}$$

and, since we can choose r and s arbitrarily close to t, this forces $g(t) = e^{-\lambda t}$, so $T \sim E(\lambda)$. $\square$

The next result shows that a sum of independent exponential random variables is either certain to be finite or certain to be infinite, and gives a criterion for deciding which is true. This will be used to determine whether or not certain continuous-time Markov chains can take infinitely many jumps in a finite time.

Theorem 2.3.2. *Let $S_1, S_2, \ldots$ be a sequence of independent random variables with $S_n \sim E(\lambda_n)$ and $0 < \lambda_n < \infty$ for all n.*

(i) *If* $\displaystyle\sum_{n=1}^{\infty} \frac{1}{\lambda_n} < \infty$, *then* $\displaystyle\mathbb{P}\left(\sum_{n=1}^{\infty} S_n < \infty\right) = 1.$

(ii) *If* $\displaystyle\sum_{n=1}^{\infty} \frac{1}{\lambda_n} = \infty$, *then* $\displaystyle\mathbb{P}\left(\sum_{n=1}^{\infty} S_n = \infty\right) = 1.$

Proof. (i) Suppose $\sum_{n=1}^{\infty} 1/\lambda_n < \infty$. Then, by monotone convergence

$$\mathbb{E}\left(\sum_{n=1}^{\infty} S_n\right) = \sum_{n=1}^{\infty} \frac{1}{\lambda_n} < \infty$$

so

$$\mathbb{P}\left(\sum_{n=1}^{\infty} S_n < \infty\right) = 1.$$

(ii) Suppose instead that $\sum_{n=1}^{\infty} 1/\lambda_n = \infty$. Then $\prod_{n=1}^{\infty}(1+1/\lambda_n) = \infty$. By monotone convergence and independence

$$\mathbb{E}\left(\exp\left\{-\sum_{n=1}^{\infty} S_n\right\}\right) = \prod_{n=1}^{\infty} \mathbb{E}\Big(\exp\{-S_n\}\Big) = \prod_{n=1}^{\infty}\left(1+\frac{1}{\lambda_n}\right)^{-1} = 0$$

so

$$\mathbb{P}\left(\sum_{n=1}^{\infty} S_n = \infty\right) = 1. \qquad \square$$

The following result is fundamental to continuous-time Markov chains.

Theorem 2.3.3. *Let I be a countable set and let $T_k, k \in I$, be independent random variables with $T_k \sim E(q_k)$ and $0 < q := \sum_{k\in I} q_k < \infty$. Set $T = \inf_k T_k$. Then this infimum is attained at a unique random value K of k, with probability 1. Moreover, T and K are independent, with $T \sim E(q)$ and $\mathbb{P}(K = k) = q_k/q$.*

Proof. Set $K = k$ if $T_k < T_j$ for all $j \neq k$, otherwise let K be undefined. Then

$$\begin{aligned}
&\mathbb{P}(K = k \text{ and } T \geq t)\\
&\quad = \mathbb{P}(T_k \geq t \text{ and } T_j > T_k \text{ for all } j \neq k)\\
&\quad = \int_t^{\infty} q_k e^{-q_k s}\mathbb{P}(T_j > s \text{ for all } j \neq k) ds\\
&\quad = \int_t^{\infty} q_k e^{-q_k s} \prod_{j\neq k} e^{-q_j s} ds\\
&\quad = \int_t^{\infty} q_k e^{-qs} ds = \frac{q_k}{q} e^{-qt}.
\end{aligned}$$

Hence $\mathbb{P}(K = k \text{ for some } k) = 1$ and T and K have the claimed joint distribution. $\square$

The following identity is the simplest case of an identity used in Section 2.8 in proving the forward equations for a continuous-time Markov chain.

Theorem 2.3.4. *For independent random variables $S \sim E(\lambda)$ and $R \sim E(\mu)$ and for $t \geq 0$, we have*

$$\mu\mathbb{P}(S \leq t < S + R) = \lambda\mathbb{P}(R \leq t < R + S).$$

Proof. We have

$$\mu\mathbb{P}(S \le t < S + R) = \mu \int_0^t \int_{t-s}^\infty \lambda\mu e^{-\lambda s} e^{-\mu r} dr ds = \lambda\mu \int_0^t e^{-\lambda s} e^{-\mu(t-s)} ds$$

from which the identity follows by symmetry. □

Exercises

2.3.1 Suppose S and T are independent exponential random variables of parameters α and β respectively. What is the distribution of $\min\{S,T\}$? What is the probability that $S \le T$? Show that the two events $\{S < T\}$ and $\{\min\{S,T\} \ge t\}$ are independent.

2.3.2 Let $T_1, T_2, \dots$ be independent exponential random variables of parameter λ and let N be an independent geometric random variable with

$$\mathbb{P}(N = n) = \beta(1-\beta)^{n-1}, \qquad n = 1, 2, \dots \, .$$

Show that $T = \sum_{i=1}^N T_i$ has exponential distribution of parameter $\lambda\beta$.

2.3.3 Let $S_1, S_2, \dots$ be independent exponential random variables with parameters $\lambda_1, \lambda_2, \dots$ respectively. Show that $\lambda_1 S_1$ is exponential of parameter 1.

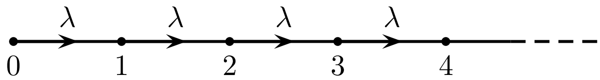
Use the strong law of large numbers to show, first in the special case $\lambda_n = 1$ for all n, and then subject only to the condition $\sup_n \lambda_n < \infty$, that

$$\mathbb{P}\left(\sum_{n=1}^\infty S_n = \infty\right) = 1 \, .$$

Is the condition $\sup_n \lambda_n < \infty$ absolutely necessary?

2.4 Poisson processes

Poisson processes are some of the simplest examples of continuous-time Markov chains. We shall also see that they may serve as building blocks for the most general continuous-time Markov chain. Moreover, a Poisson process is the natural probabilistic model for any uncoordinated stream of discrete events in continuous time. So we shall study Poisson processes first, both as a gentle warm-up for the general theory and because they are useful in themselves. The key result is Theorem 2.4.3, which provides three different descriptions of a Poisson process. The reader might well begin with the statement of this result and then see how it is used in the theorems and examples that follow. We shall begin with a definition in terms of jump chain and holding times (see Section 2.2). A right-continuous process $(X_t)_{t\geq 0}$ with values in $\{0, 1, 2, \dots\}$ is a *Poisson process of rate* λ $(0 < \lambda < \infty)$ if its holding times $S_1, S_2, \dots$ are independent exponential random variables of parameter λ and its jump chain is given by $Y_n = n$. Here is the diagram:

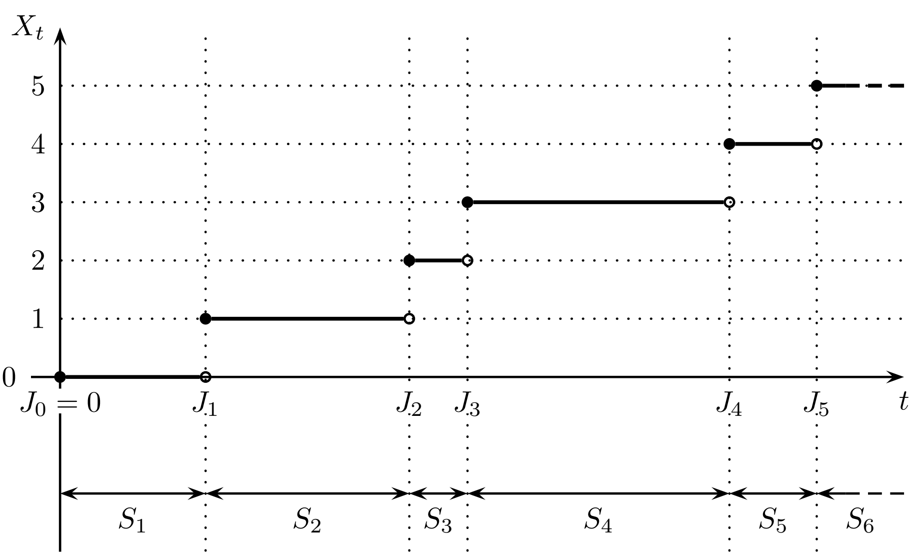


The associated Q-matrix is given by

$$Q = \begin{pmatrix} -\lambda & \lambda & & & \\ & -\lambda & \lambda & & \\ & & \ddots & \ddots & \\ & & & & \\ \end{pmatrix}.$$

By Theorem 2.3.2 (or the strong law of large numbers) we have $\mathbb{P}(J_n \to \infty) = 1$ so there is no explosion and the law of $(X_t)_{t\geq 0}$ is uniquely determined. A simple way to construct a Poisson process of rate λ is to take a sequence $S_1, S_2, \dots$ of independent exponential random variables of parameter λ, to set $J_0 = 0$, $J_n = S_1 + \ldots + S_n$ and then set

$$X_t = n \quad \text{if} \quad J_n \leq t < J_{n+1}.$$



The diagram illustrates a typical path. We now show how the memoryless property of the exponential holding times, Theorem 2.3.1, leads to a memoryless property of the Poisson process.

Theorem 2.4.1 (Markov property). *Let $(X_t)_{t\geq 0}$ be a Poisson process of rate λ. Then, for any $s \geq 0$, $(X_{s+t} - X_s)_{t\geq 0}$ is also a Poisson process of rate λ, independent of $(X_r : r \leq s)$.*

Proof. It suffices to prove the claim conditional on the event $X_s = i$, for each $i \geq 0$. Set $\widetilde{X}_t = X_{s+t} - X_s$. We have

$$\{X_s = i\} = \{J_i \leq s < J_{i+1}\} = \{J_i \leq s\} \cap \{S_{i+1} > s - J_i\}.$$

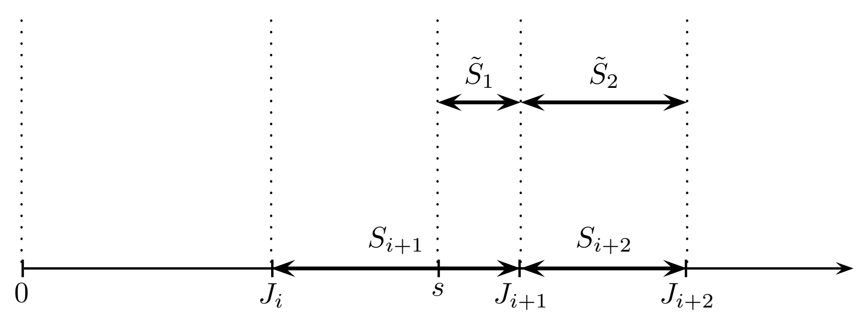
On this event

$$X_r = \sum_{j=1}^{i} 1_{\{S_j \leq r\}} \quad \text{for } r \leq s$$

and the holding times $\widetilde{S}_1, \widetilde{S}_2, \dots$ of $(\widetilde{X}_t)_{t\geq 0}$ are given by

$$\widetilde{S}_1 = S_{i+1} - (s - J_i), \quad \widetilde{S}_n = S_{i+n} \quad \text{for } n \geq 2$$

as shown in the diagram.



Recall that the holding times $S_1, S_2, \dots$ are independent $E(\lambda)$. Condition on $S_1, \dots, S_i$ and $\{X_s = i\}$, then by the memoryless property of S_{i+1} and independence, $\widetilde{S}_1, \widetilde{S}_2, \dots$ are themselves independent $E(\lambda)$. Hence, conditional on $\{X_s = i\}$, $\widetilde{S}_1, \widetilde{S}_2, \dots$ are independent $E(\lambda)$, and independent of $S_1, \dots, S_i$. Hence, conditional on $\{X_s = i\}$, $(\widetilde{X}_t)_{t\geq 0}$ is a Poisson process of rate λ and independent of $(X_r : r \leq s)$. $\square$

In fact, we shall see in Section 6.5, by an argument in essentially the same spirit that the result also holds with s replaced by any stopping time T of $(X_t)_{t\geq 0}$.

Theorem 2.4.2 (Strong Markov property). *Let $(X_t)_{t\geq 0}$ be a Poisson process of rate λ and let T be a stopping time of $(X_t)_{t\geq 0}$. Then, conditional on $T < \infty$, $(X_{T+t} - X_T)_{t\geq 0}$ is also a Poisson process of rate λ, independent of $(X_s : s \leq T)$.*

Here is some standard terminology. If $(X_t)_{t\geq 0}$ is a real-valued process, we can consider its *increment* $X_t - X_s$ over any interval $(s, t]$. We say that $(X_t)_{t\geq 0}$ has *stationary* increments if the distribution of $X_{s+t} - X_s$ depends only on $t \geq 0$. We say that $(X_t)_{t\geq 0}$ has *independent* increments if its increments over any finite collection of disjoint intervals are independent.

We come to the key result for the Poisson process, which gives two conditions equivalent to the jump chain/holding time characterization which we took as our original definition. Thus we have three alternative definitions of the same process.

Theorem 2.4.3. *Let $(X_t)_{t\geq 0}$ be an increasing, right-continuous integer-valued process starting from* 0. *Let* $0 < \lambda < \infty$. *Then the following three conditions are equivalent:*

(a) (jump chain/holding time definition) *the holding times $S_1, S_2, \dots$ of $(X_t)_{t\geq 0}$ are independent exponential random variables of parameter λ and the jump chain is given by $Y_n = n$ for all n;*

(b) (infinitesimal definition) *$(X_t)_{t\geq 0}$ has independent increments and, as $h \downarrow 0$, uniformly in t,*

$$\mathbb{P}(X_{t+h} - X_t = 0) = 1 - \lambda h + o(h), \quad \mathbb{P}(X_{t+h} - X_t = 1) = \lambda h + o(h);$$

(c) (transition probability definition) *$(X_t)_{t\geq 0}$ has stationary independent increments and, for each t, X_t has Poisson distribution of parameter λt.*

If $(X_t)_{t\geq 0}$ satisfies any of these conditions then it is called a *Poisson process of rate* λ.

Proof. (a) $\Rightarrow$ (b) If (a) holds, then, by the Markov property, for any $t, h \geq 0$, the increment $X_{t+h} - X_t$ has the same distribution as X_h and is independent of $(X_s : s \leq t)$. So $(X_t)_{t\geq 0}$ has independent increments and as $h \downarrow 0$

$$\begin{aligned} \mathbb{P}(X_{t+h} - X_t \geq 1) &= \mathbb{P}(X_h \geq 1) = \mathbb{P}(J_1 \leq h) = 1 - e^{-\lambda h} = \lambda h + o(h), \\ \mathbb{P}(X_{t+h} - X_t \geq 2) &= \mathbb{P}(X_h \geq 2) = \mathbb{P}(J_2 \leq h) \\ &\leq \mathbb{P}(S_1 \leq h \text{ and } S_2 \leq h) = (1 - e^{-\lambda h})^2 = o(h), \end{aligned}$$

which implies (b).

(b) $\Rightarrow$ (c) If (b) holds, then, for $i = 2, 3, \ldots$, we have $\mathbb{P}(X_{t+h} - X_t = i) = o(h)$ as $h \downarrow 0$, uniformly in t. Set $p_j(t) = \mathbb{P}(X_t = j)$. Then, for $j = 1, 2, \ldots$,

$$\begin{aligned} p_j(t+h) = \mathbb{P}(X_{t+h} = j) &= \sum_{i=0}^{j} \mathbb{P}(X_{t+h} - X_t = i)\,\mathbb{P}(X_t = j-i) \\ &= (1 - \lambda h + o(h))p_j(t) + (\lambda h + o(h))p_{j-1}(t) + o(h) \end{aligned}$$

so

$$\frac{p_j(t+h) - p_j(t)}{h} = -\lambda p_j(t) + \lambda p_{j-1}(t) + O(h).$$

Since this estimate is uniform in t we can put $t = s - h$ to obtain for all $s \geq h$

$$\frac{p_j(s) - p_j(s-h)}{h} = -\lambda p_j(s-h) + \lambda p_{j-1}(s-h) + O(h).$$

Now let $h \downarrow 0$ to see that $p_j(t)$ is first continuous and then differentiable and satisfies the differential equation

$$p_j'(t) = -\lambda p_j(t) + \lambda p_{j-1}(t).$$

By a simpler argument we also find

$$p_0'(t) = -\lambda p_0(t).$$

Since $X_0 = 0$ we have initial conditions

$$p_0(0) = 1, \quad p_j(0) = 0 \quad \text{for } j = 1, 2, \ldots .$$

As we saw in Example 2.1.4, this system of equations has a unique solution given by

$$p_j(t) = e^{-\lambda t} \frac{(\lambda t)^j}{j!}, \qquad j = 0, 1, 2, \ldots .$$

Hence $X_t \sim P(\lambda t)$. If $(X_t)_{t \geq 0}$ satisfies (b), then certainly $(X_t)_{t \geq 0}$ has independent increments, but also $(X_{s+t} - X_s)_{t \geq 0}$ satisfies (b), so the above argument shows $X_{s+t} - X_s \sim P(\lambda t)$, for any s, which implies (c).

(c) $\Rightarrow$ (a) There is a process satisfying (a) and we have shown that it must then satisfy (c). But condition (c) determines the finite-dimensional distributions of $(X_t)_{t \geq 0}$ and hence the distribution of jump chain and holding times. So if one process satisfying (c) also satisfies (a), so must every process satisfying (c). $\square$

The differential equations which appeared in the proof are really the forward equations for the Poisson process. To make this clear, consider the possibility of starting the process from i at time 0, writing $\mathbb{P}_i$ as a reminder, and set

$$p_{ij}(t) = \mathbb{P}_i(X_t = j).$$

Then, by spatial homogeneity $p_{ij}(t) = p_{j-i}(t)$, and we could rewrite the differential equations as

$$\begin{aligned} p'_{i0}(t) &= -\lambda p_{i0}(t), & p_{i0}(0) &= \delta_{i0}, \\ p'_{ij}(t) &= \lambda p_{i,j-1}(t) - \lambda p_{ij}(t), & p_{ij}(0) &= \delta_{ij} \end{aligned}$$

or, in matrix form, for Q as above,

$$P'(t) = P(t)Q, \qquad P(0) = I.$$

Theorem 2.4.3 contains a great deal of information about the Poisson process of rate λ. It can be useful when trying to decide whether a given process is a Poisson process as it gives you three alternative conditions to check, and it is likely that one will be easier to check than another. On the other hand it can also be useful when answering a question about a given Poisson process as this question may be more closely connected to one definition than another. For example, you might like to consider the difficulties in approaching the next result using the jump chain/holding time definition.

Theorem 2.4.4. *If $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ are independent Poisson processes of rates λ and μ, respectively, then $(X_t + Y_t)_{t\geq 0}$ is a Poisson process of rate $\lambda + \mu$.*

Proof. We shall use the infinitesimal definition, according to which $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ have independent increments and, as $h \downarrow 0$, uniformly in t,

$$\begin{aligned} \mathbb{P}(X_{t+h} - X_t = 0) = 1 - \lambda h + o(h), \quad \mathbb{P}(X_{t+h} - X_t = 1) = \lambda h + o(h), \\ \mathbb{P}(Y_{t+h} - Y_t = 0) = 1 - \mu h + o(h), \quad \mathbb{P}(Y_{t+h} - Y_t = 1) = \mu h + o(h). \end{aligned}$$

Set $Z_t = X_t + Y_t$. Then, since $(X_t)_{t\geq 0}$ and $(Y_t)_{t\geq 0}$ are independent, $(Z_t)_{t\geq 0}$ has independent increments and, as $h \downarrow 0$, uniformly in t,

$$\begin{aligned} \mathbb{P}(Z_{t+h} - Z_t = 0) &= \mathbb{P}(X_{t+h} - X_t = 0)\mathbb{P}(Y_{t+h} - Y_t = 0) \\ &= (1 - \lambda h + o(h))(1 - \mu h + o(h)) = 1 - (\lambda + \mu)h + o(h), \\ \mathbb{P}(Z_{t+h} - Z_t = 1) &= \mathbb{P}(X_{t+h} - X_t = 1)\mathbb{P}(Y_{t+h} - Y_t = 0) \\ &\quad + \mathbb{P}(X_{t+h} - X_t = 0)\mathbb{P}(Y_{t+h} - Y_t = 1) \\ &= (\lambda h + o(h))(1 - \mu h + o(h)) + (1 - \lambda h + o(h))(\mu h + o(h)) \\ &= (\lambda + \mu)h + o(h). \end{aligned}$$

Hence $(Z_t)_{t\geq 0}$ is a Poisson process of rate $\lambda + \mu$. $\square$

Next we establish some relations between Poisson processes and the uniform distribution. Notice that the conclusions are independent of the rate of the process considered. The results say in effect that the jumps of a Poisson process are as randomly distributed as possible.

Theorem 2.4.5. *Let $(X_t)_{t\geq 0}$ be a Poisson process. Then, conditional on $(X_t)_{t\geq 0}$ having exactly one jump in the interval $[s, s+t]$, the time at which that jump occurs is uniformly distributed on $[s, s+t]$.*

Proof. We shall use the finite-dimensional distribution definition. By stationarity of increments, it suffices to consider the case $s = 0$. Then, for $0 \leq u \leq t$,

$$\begin{aligned}\mathbb{P}(J_1 \leq u \mid X_t = 1) &= \mathbb{P}(J_1 \leq u \text{ and } X_t = 1)/\mathbb{P}(X_t = 1)\\ &= \mathbb{P}(X_u = 1 \text{ and } X_t - X_u = 0)/\mathbb{P}(X_t = 1)\\ &= \lambda u e^{-\lambda u} e^{-\lambda(t-u)}/(\lambda t e^{-\lambda t}) = u/t.\end{aligned}$$

□

Theorem 2.4.6. *Let $(X_t)_{t\geq 0}$ be a Poisson process. Then, conditional on the event $\{X_t = n\}$, the jump times $J_1, \dots, J_n$ have joint density function*

$$f(t_1, \dots, t_n) = n! t^{-n} 1_{\{0 \leq t_1 \leq \dots \leq t_n \leq t\}}.$$

Thus, conditional on $\{X_t = n\}$, the jump times $J_1, \dots, J_n$ have the same distribution as an ordered sample of size n from the uniform distribution on $[0, t]$.

Proof. The holding times $S_1, \dots, S_{n+1}$ have joint density function

$$\lambda^{n+1} e^{-\lambda(s_1 + \dots + s_{n+1})} 1_{\{s_1, \dots, s_{n+1} \geq 0\}}$$

so the jump times $J_1, \dots, J_{n+1}$ have joint density function

$$\lambda^{n+1} e^{-\lambda t_{n+1}} 1_{\{0 \leq t_1 \leq \dots \leq t_{n+1}\}}.$$

So for $A \subseteq \mathbb{R}^n$ we have

$$\begin{aligned}\mathbb{P}\big((J_1, \dots, J_n) \in A \text{ and } X_t = n\big) &= \mathbb{P}\big((J_1, \dots, J_n) \in A \text{ and } J_n \leq t < J_{n+1}\big)\\ &= e^{-\lambda t} \lambda^n \int_{(t_1, \dots, t_n) \in A} 1_{\{0 \leq t_1 \leq \dots \leq t_n \leq t\}} dt_1 \dots dt_n\end{aligned}$$

and since $\mathbb{P}(X_t = n) = e^{-\lambda t}(\lambda t)^n / n!$ we obtain

$$\mathbb{P}\big((J_1, \dots, J_n) \in A \mid X_t = n\big) = \int_A f(t_1, \dots, t_n) dt_1 \dots dt_n$$

as required. □

We finish with a simple example typical of many problems making use of a range of properties of the Poisson process.

Example 2.4.7

Robins and blackbirds make brief visits to my birdtable. The probability that in any small interval of duration h a robin will arrive is found to be $\rho h + o(h)$, whereas the corresponding probability for blackbirds is $\beta h + o(h)$. What is the probability that the first two birds I see are both robins? What is the distribution of the total number of birds seen in time t? Given that this number is n, what is the distribution of the number of blackbirds seen in time t?

By the infinitesimal characterization, the number of robins seen by time t is a Poisson process $(R_t)_{t\geq 0}$ of rate ρ, and the number of blackbirds is a Poisson process $(B_t)_{t\geq 0}$ of rate β. The times spent waiting for the first robin or blackbird are independent exponential random variables S_1 and T_1 of parameters ρ and β respectively. So a robin arrives first with probability $\rho/(\rho+\beta)$ and, by the memoryless property of T_1, the probability that the first two birds are robins is $\rho^2/(\rho+\beta)^2$. By Theorem 2.4.4 the total number of birds seen in an interval of duration t has Poisson distribution of parameter $(\rho+\beta)t$. Finally

$$\begin{aligned}\mathbb{P}(B_t = k \mid R_t + B_t = n) &= \mathbb{P}(B_t = k \text{ and } R_t = n-k)/\mathbb{P}(R_t + B_t = n)\\ &= \left(\frac{e^{-\beta}\beta^k}{k!}\right)\left(\frac{e^{-\rho}\rho^{n-k}}{(n-k)!}\right)\Big/\left(\frac{e^{-(\rho+\beta)}(\rho+\beta)^n}{n!}\right)\\ &= \binom{n}{k}\left(\frac{\beta}{\rho+\beta}\right)^k\left(\frac{\rho}{\rho+\beta}\right)^{n-k}\end{aligned}$$

so if n birds are seen in time t, then the distribution of the number of blackbirds is binomial of parameters n and $\beta/(\rho+\beta)$.

Exercises

2.4.1 State the transition probability definition of a Poisson process. Show directly from this definition that the first jump time J_1 of a Poisson process of rate λ is exponential of parameter λ.

Show also (from the same definition and without assuming the strong Markov property) that

$$\mathbb{P}(t_1 < J_1 \leq t_2 < J_2) = e^{-\lambda t_1}\lambda(t_2 - t_1)e^{-\lambda(t_2-t_1)}$$

and hence that $J_2 - J_1$ is also exponential of parameter λ and independent of J_1.

2.4.2 Show directly from the infinitesimal definition that the first jump time J_1 of a Poisson process of rate λ has exponential distribution of parameter λ.

2.4.3 Arrivals of the Number 1 bus form a Poisson process of rate one bus per hour, and arrivals of the Number 7 bus form an independent Poisson process of rate seven buses per hour.

(a) What is the probability that exactly three buses pass by in one hour?

(b) What is the probability that exactly three Number 7 buses pass by while I am waiting for a Number 1?

(c) When the maintenance depot goes on strike half the buses break down before they reach my stop. What, then, is the probability that I wait for 30 minutes without seeing a single bus?

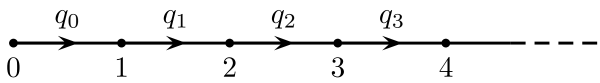
2.4.4 A radioactive source emits particles in a Poisson process of rate λ. The particles are each emitted in an independent random direction. A Geiger counter placed near the source records a fraction p of the particles emitted. What is the distribution of the number of particles recorded in time t?

2.4.5 A pedestrian wishes to cross a single lane of fast-moving traffic. Suppose the number of vehicles that have passed by time t is a Poisson process of rate λ, and suppose it takes time a to walk across the lane. Assuming that the pedestrian can foresee correctly the times at which vehicles will pass by, how long on average does it take to cross over safely? [*Consider the time at which the first car passes.*]

How long on average does it take to cross two similar lanes (a) when one must walk straight across (assuming that the pedestrian will not cross if, at any time whilst crossing, a car would pass in either direction), (b) when an island in the middle of the road makes it safe to stop half-way?

2.5 Birth processes

A birth process is a generalization of a Poisson process in which the parameter λ is allowed to depend on the current state of the process. The data for a birth process consist of *birth rates* $0 \leq q_j < \infty$, where $j = 0, 1, 2, \ldots$. We begin with a definition in terms of jump chain and holding times. A minimal right-continuous process $(X_t)_{t \geq 0}$ with values in $\{0, 1, 2, \ldots\} \cup \{\infty\}$ is a *birth process of rates* $(q_j : j \geq 0)$ if, conditional on $X_0 = i$, its holding times $S_1, S_2, \ldots$ are independent exponential random variables of parameters $q_i, q_{i+1}, \ldots$, respectively, and its jump chain is given by $Y_n = i + n$.



The flow diagram is shown above and the Q-matrix is given by:

$$Q = \begin{pmatrix} -q_0 & q_0 & & & \\ & -q_1 & q_1 & & \\ & & -q_2 & q_2 & \\ & & & \ddots & \ddots \\ & & & & \end{pmatrix}.$$

Example 2.5.1 (Simple birth process)

Consider a population in which each individual gives birth after an exponential time of parameter λ, all independently. If i individuals are present then the first birth will occur after an exponential time of parameter $i\lambda$. Then we have $i + 1$ individuals and, by the memoryless property, the process begins afresh. Thus the size of the population performs a birth process with rates $q_i = i\lambda$. Let X_t denote the number of individuals at time t and suppose $X_0 = 1$. Write T for the time of the first birth. Then

$$\begin{aligned} \mathbb{E}(X_t) &= \mathbb{E}(X_t 1_{T \le t}) + \mathbb{E}(X_t 1_{T > t}) \\ &= \int_0^t \lambda e^{-\lambda s} \mathbb{E}(X_t \mid T = s) ds + e^{-\lambda t}. \end{aligned}$$

Put $\mu(t) = \mathbb{E}(X_t)$, then $\mathbb{E}(X_t \mid T = s) = 2\mu(t - s)$, so

$$\mu(t) = \int_0^t 2\lambda e^{-\lambda s} \mu(t - s) ds + e^{-\lambda t}$$

and setting $r = t - s$

$$e^{\lambda t} \mu(t) = 2\lambda \int_0^t e^{\lambda r} \mu(r) dr + 1.$$

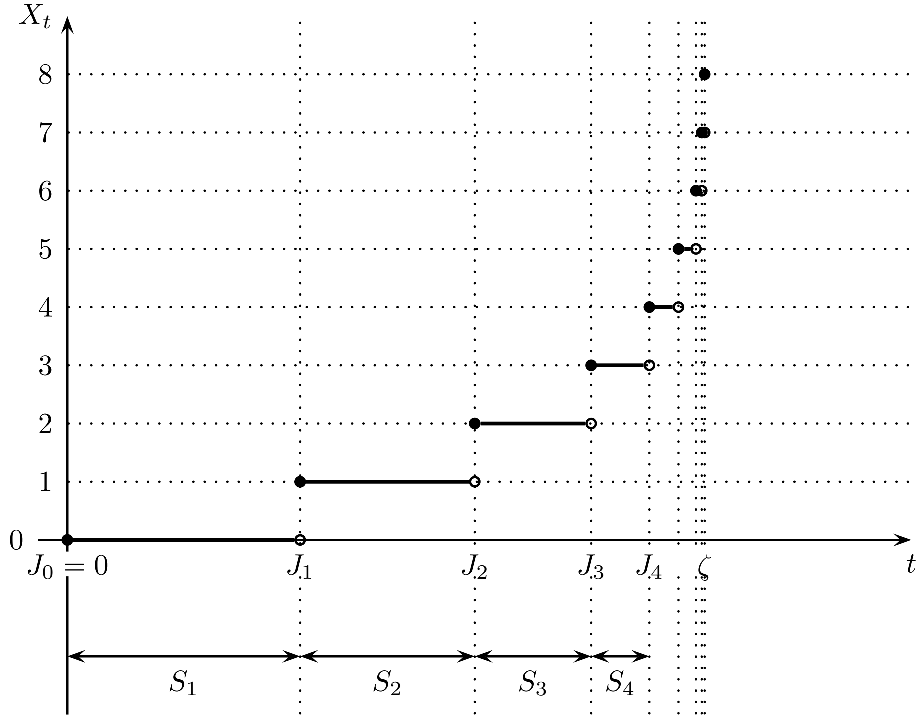
By differentiating we obtain

$$\mu'(t) = \lambda \mu(t)$$

so the mean population size grows exponentially:

$$\mathbb{E}(X_t) = e^{\lambda t}.$$

Much of the theory associated with the Poisson process goes through for birth processes with little change, except that some calculations can no longer be made so explicitly. The most interesting new phenomenon present in birth processes is the possibility of explosion. For certain choices of birth rates, a typical path will make infinitely many jumps in a finite time, as shown in the diagram. The convention of setting the process to equal ∞ after explosion is particularly appropriate for birth processes!



In fact, Theorem 2.3.2 tells us exactly when explosion will occur.

Theorem 2.5.2. *Let* $(X_t)_{t\geq 0}$ *be a birth process of rates* $(q_j : j \geq 0)$, *starting from* 0.

(i) *If* $\displaystyle\sum_{j=0}^{\infty} \frac{1}{q_j} < \infty$, *then* $\mathbb{P}(\zeta < \infty) = 1$.

(ii) *If* $\displaystyle\sum_{j=0}^{\infty} \frac{1}{q_j} = \infty$, *then* $\mathbb{P}(\zeta = \infty) = 1$.

Proof. Apply Theorem 2.3.2 to the sequence of holding times $S_1, S_2, \ldots$. □

The proof of the Markov property for the Poisson process is easily adapted to give the following generalization.

Theorem 2.5.3 (Markov property). *Let $(X_t)_{t\geq 0}$ be a birth process of rates $(q_j : j \geq 0)$. Then, conditional on $X_s = i$, $(X_{s+t})_{t\geq 0}$ is a birth process of rates $(q_j : j \geq 0)$ starting from i and independent of $(X_r : r \leq s)$.*

We shall shortly prove a theorem on birth processes which generalizes the key theorem on Poisson processes. First we must see what will replace the Poisson probabilities. In Theorem 2.4.3 these arose as the unique solution of a system of differential equations, which we showed were essentially the forward equations. Now we can still write down the forward equation

$$P'(t) = P(t)Q, \quad P(0) = I.$$

or, in components

$$p'_{i0}(t) = -p_{i0}(t)q_0, \quad p_{i0}(0) = \delta_{i0}$$

and, for $j = 1, 2, \dots$

$$p'_{ij}(t) = p_{i,j-1}(t)q_{j-1} - p_{ij}(t)q_j, \quad p_{ij}(0) = \delta_{ij}.$$

Moreover, these equations still have a unique solution; it is just not as explicit as before. For we must have

$$p_{i0}(t) = \delta_{i0}e^{-q_0 t}$$

which can be substituted in the equation

$$p'_{i1}(t) = p_{i0}(t)q_0 - p_{i1}(t)q_1, \quad p_{i1}(0) = \delta_{i1}$$

and this equation solved to give

$$p_{i1}(t) = \delta_{i1}e^{-q_1 t} + \delta_{i0}\int_0^t q_0 e^{-q_0 s}e^{-q_1(t-s)}ds.$$

Now we can substitute for $p_{i1}(t)$ in the next equation up the hierarchy and find an explicit expression for $p_{i2}(t)$, and so on.

Theorem 2.5.4. *Let $(X_t)_{t\geq 0}$ be an increasing, right-continuous process with values in $\{0, 1, 2, \dots\} \cup \{\infty\}$. Let $0 \leq q_j < \infty$ for all $j \geq 0$. Then the following three conditions are equivalent:*

(a) (jump chain/holding time definition) *conditional on $X_0 = i$, the holding times $S_1, S_2, \dots$ are independent exponential random variables of parameters $q_i, q_{i+1}, \dots$ respectively and the jump chain is given by $Y_n = i + n$ for all n;*

(b) (infinitesimal definition) *for all* $t, h \geq 0$, *conditional on* $X_t = i$, X_{t+h} *is independent of* $(X_s : s \leq t)$ *and, as* $h \downarrow 0$, *uniformly in* t,

$$\mathbb{P}(X_{t+h} = i \mid X_t = i) = 1 - q_i h + o(h),$$
$$\mathbb{P}(X_{t+h} = i + 1 \mid X_t = i) = q_i h + o(h);$$

(c) (transition probability definition) *for all* $n = 0, 1, 2, \dots$, *all times* $0 \leq t_0 \leq \dots \leq t_{n+1}$ *and all states* $i_0, \dots, i_{n+1}$

$$\mathbb{P}(X_{t_{n+1}} = i_{n+1} \mid X_{t_0} = i_0, \dots, X_{t_n} = i_n) = p_{i_n i_{n+1}}(t_{n+1} - t_n)$$

where $(p_{ij}(t) : i, j = 0, 1, 2, \dots)$ *is the unique solution of the forward equations.*

If $(X_t)_{t \geq 0}$ satisfies any of these conditions then it is called a *birth process of rates* $(q_j : j \geq 0)$.

Proof. (a) $\Rightarrow$ (b) If (a) holds, then, by the Markov property for any $t, h \geq 0$, conditional on $X_t = i$, X_{t+h} is independent of $(X_s : s \leq t)$ and, as $h \downarrow 0$, uniformly in t,

$$\begin{aligned}\mathbb{P}(X_{t+h} \geq i + 1 \mid X_t = i) &= \mathbb{P}(X_h \geq i + 1 \mid X_0 = i) \\ &= \mathbb{P}(J_1 \leq h \mid X_0 = i) = 1 - e^{-q_i h} = q_i h + o(h),\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}(X_{t+h} \geq i + 2 \mid X_t = i) &= \mathbb{P}(X_h \geq i + 2 \mid X_0 = i) \\ &= \mathbb{P}(J_2 \leq h \mid X_0 = i) \leq \mathbb{P}(S_1 \leq h \text{ and } S_2 \leq h \mid X_0 = i) \\ &= (1 - e^{-q_i h})(1 - e^{-q_{i+1} h}) = o(h),\end{aligned}$$

which implies (b).

(b) $\Rightarrow$ (c) If (b) holds, then certainly for $k = i + 2, i + 3, \dots$

$$\mathbb{P}(X_{t+h} = k \mid X_t = i) = o(h) \quad \text{as } h \downarrow 0, \text{ uniformly in } t.$$

Set $p_{ij}(t) = \mathbb{P}(X_t = j \mid X_0 = i)$. Then, for $j = 1, 2, \dots$

$$\begin{aligned}p_{ij}(t + h) &= \mathbb{P}(X_{t+h} = j \mid X_0 = i) \\ &= \sum_{k=i}^{j} \mathbb{P}(X_t = k \mid X_0 = i)\mathbb{P}(X_{t+h} = j \mid X_t = k) \\ &= p_{ij}(t)(1 - q_j h + o(h)) + p_{i,j-1}(t)(q_{j-1} h + o(h)) + o(h)\end{aligned}$$

so

$$\frac{p_{ij}(t+h)-p_{ij}(t)}{h}=p_{i,j-1}(t)q_{j-1}-p_{ij}(t)q_j+O(h).$$

As in the proof of Theorem 2.4.3, we can deduce that $p_{ij}(t)$ is differentiable and satisfies the differential equation

$$p'_{ij}(t)=p_{i,j-1}(t)q_{j-1}-p_{ij}(t)q_j.$$

By a simpler argument we also find

$$p'_{i0}(t)=-p_{i0}(t)q_0.$$

Thus $(p_{ij}(t):i,j=0,1,2,\dots)$ must be the unique solution to the forward equations. If $(X_t)_{t\geq 0}$ satisfies (b), then certainly

$$\mathbb{P}(X_{t_{n+1}}=i_{n+1}\mid X_0=i_0,\dots,X_{t_n}=i_n)=\mathbb{P}(X_{t_{n+1}}=i_{n+1}\mid X_{t_n}=i_n)$$

but also $(X_{t_n+t})_{t\geq 0}$ satisfies (b), so

$$\mathbb{P}(X_{t_{n+1}}=i_{n+1}\mid X_{t_n}=i_n)=p_{i_n i_{n+1}}(t_{n+1}-t_n)$$

by uniqueness for the forward equations. Hence $(X_t)_{t\geq 0}$ satisfies (c).

(c) $\Rightarrow$ (a) See the proof of Theorem 2.4.3. □

Exercise

2.5.1 Each bacterium in a colony splits into two identical bacteria after an exponential time of parameter λ, which then split in the same way but independently. Let X_t denote the size of the colony at time t, and suppose $X_0=1$. Show that the probability generating function $\phi(t)=\mathbb{E}(z^{X_t})$ satisfies

$$\phi(t)=ze^{-\lambda t}+\int_0^t \lambda e^{-\lambda s}\phi(t-s)^2 ds\,.$$

Make a change of variables $u=t-s$ in the integral and deduce that $d\phi/dt=\lambda\phi(\phi-1)$. Hence deduce that, for $q=1-e^{-\lambda t}$ and $n=1,2,\dots$

$$\mathbb{P}(X_t=n)=q^{n-1}(1-q)\,.$$

2.6 Jump chain and holding times

This section begins the theory of continuous-time Markov chains proper, which will occupy the remainder of this chapter and the whole of the next. The approach we have chosen is to introduce continuous-time chains in terms of the joint distribution of their jump chain and holding times. This provides the most direct mathematical description. It also makes possible a number of constructive realizations of a given Markov chain, which we shall describe, and which underlie many applications.

Let I be a countable set. The basic data for a continuous-time Markov chain on I are given in the form of a Q-matrix. Recall that a Q-matrix on I is any matrix $Q = (q_{ij} : i, j \in I)$ which satisfies the following conditions:

(i) $0 \leq -q_{ii} < \infty \quad \text{for all} \quad i;$

(ii) $q_{ij} \geq 0 \quad \text{for all} \quad i \neq j;$

(iii) $\displaystyle\sum_{j \in I} q_{ij} = 0 \quad \text{for all} \quad i.$

We will sometimes find it convenient to write q_i or $q(i)$ as an alternative notation for $-q_{ii}$.

We are going to describe a simple procedure for obtaining from a Q-matrix Q a stochastic matrix Π. The *jump matrix* $\Pi = (\pi_{ij} : i, j \in I)$ of Q is defined by

$$\pi_{ij} = \begin{cases} q_{ij}/q_i & \text{if } j \neq i \text{ and } q_i \neq 0 \\ 0 & \text{if } j \neq i \text{ and } q_i = 0, \end{cases}$$

$$\pi_{ii} = \begin{cases} 0 & \text{if } q_i \neq 0 \\ 1 & \text{if } q_i = 0. \end{cases}$$

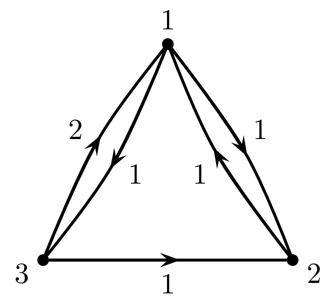
This procedure is best thought of row by row. For each $i \in I$ we take, where possible, the off-diagonal entries in the ith row of Q and scale them so they add up to 1, putting a 0 on the diagonal. This is only impossible when the off-diagonal entries are all 0, then we leave them alone and put a 1 on the diagonal. As you will see in the following example, the associated diagram transforms into a discrete-time Markov chain diagram simply by rescaling all the numbers on any arrows leaving a state so they add up to 1.

Example 2.6.1

The Q-matrix

$$Q = \begin{pmatrix} -2 & 1 & 1 \\ 1 & -1 & 0 \\ 2 & 1 & -3 \end{pmatrix}$$

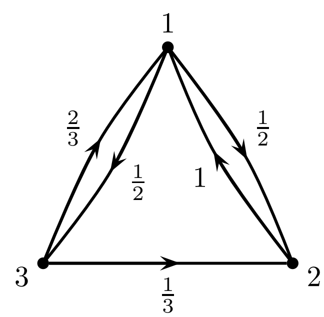
has diagram:



The jump matrix Π of Q is given by

$$\Pi = \begin{pmatrix} 0 & 1/2 & 1/2 \\ 1 & 0 & 0 \\ 2/3 & 1/3 & 0 \end{pmatrix}$$

and has diagram:



Here is the definition of a continuous-time Markov chain in terms of its jump chain and holding times. Recall that a minimal process is one which is set equal to ∞ after any explosion – see Section 2.2. A minimal right-continuous process $(X_t)_{t\geq 0}$ on I is a *Markov chain with initial distribution* λ *and generator matrix* Q if its jump chain $(Y_n)_{n\geq 0}$ is discrete-time Markov(λ, Π) and if for each $n \geq 1$, conditional on $Y_0, \dots, Y_{n-1}$, its holding times $S_1, \dots, S_n$ are independent exponential random variables of parameters $q(Y_0), \dots, q(Y_{n-1})$ respectively. We say $(X_t)_{t\geq 0}$ is *Markov*(λ, Q) for short. We can construct such a process as follows: let $(Y_n)_{n\geq 0}$ be discrete-time Markov(λ, Π) and let $T_1, T_2, \dots$ be independent exponential random variables of parameter 1, independent of $(Y_n)_{n\geq 0}$. Set $S_n = T_n/q(Y_{n-1})$, $J_n = S_1 + \ldots + S_n$ and

$$X_t = \begin{cases} Y_n & \text{if } J_n \leq t < J_{n+1} \text{ for some } n \\ \infty & \text{otherwise.} \end{cases}$$

Then $(X_t)_{t\geq 0}$ has the required properties.

We shall now describe two further constructions. You will need to understand these constructions in order to identify processes in applications which can be modelled as Markov chains. Both constructions make direct use of the entries in the Q-matrix, rather than proceeding first via the jump matrix. Here is the second construction.

We begin with an initial state $X_0 = Y_0$ with distribution λ, and with an array $(T_n^j : n \geq 1, j \in I)$ of independent exponential random variables of parameter 1. Then, inductively for $n = 0, 1, 2, \ldots$, if $Y_n = i$ we set

$$S_{n+1}^j = T_{n+1}^j / q_{ij}, \quad \text{for } j \neq i,$$
$$S_{n+1} = \inf_{j\neq i} S_{n+1}^j,$$
$$Y_{n+1} = \begin{cases} j & \text{if} \quad S_{n+1}^j = S_{n+1} < \infty \\ i & \text{if} \quad S_{n+1} = \infty. \end{cases}$$

Then, conditional on $Y_n = i$, the random variables S_{n+1}^j are independent exponentials of parameter q_{ij} for all $j \neq i$. So, conditional on $Y_n = i$, by Theorem 2.3.3, S_{n+1} is exponential of parameter $q_i = \sum_{j\neq i} q_{ij}$, Y_{n+1} has distribution $(\pi_{ij} : j \in I)$, and S_{n+1} and Y_{n+1} are independent, and independent of $Y_0, \ldots, Y_n$ and $S_1, \ldots, S_n$, as required. This construction shows why we call q_i the *rate of leaving* i and q_{ij} the *rate of going from* i *to* j.

Our third and final construction of a Markov chain with generator matrix Q and initial distribution λ is based on the Poisson process. Imagine the state-space I as a labyrinth of chambers and passages, each passage shut off by a single door which opens briefly from time to time to allow you through in one direction only. Suppose the door giving access to chamber j from chamber i opens at the jump times of a Poisson process of rate q_{ij} and you take every chance to move that you can, then you will perform a Markov chain with Q-matrix Q. In more mathematical terms, we begin with an initial state $X_0 = Y_0$ with distribution λ, and with a family of independent Poisson processes $\{(N_t^{ij})_{t\geq 0} : i, j \in I, i \neq j\}$, $(N_t^{ij})_{t\geq 0}$ having rate q_{ij}. Then set $J_0 = 0$ and define inductively for $n = 0, 1, 2, \ldots$

$$J_{n+1} = \inf\{t > J_n : N_t^{Y_n j} \neq N_{J_n}^{Y_n j} \quad \text{for some } j \neq Y_n\}$$
$$Y_{n+1} = \begin{cases} j & \text{if } J_{n+1} < \infty \text{ and } N_{J_{n+1}}^{Y_n j} \neq N_{J_n}^{Y_n j} \\ i & \text{if } J_{n+1} = \infty. \end{cases}$$

The first jump time of $(N_t^{ij})_{t\geq 0}$ is exponential of parameter q_{ij}. So, by Theorem 2.3.3, conditional on $Y_0 = i$, J_1 is exponential of parameter $q_i = \sum_{j\neq i} q_{ij}$, Y_1 has distribution $(\pi_{ij} : j \in I)$, and J_1 and Y_1 are independent.

Now suppose T is a stopping time of $(X_t)_{t\geq 0}$. If we condition on X_0 and on the processes $(N_t^{kl})_{t\geq 0}$ for $(k,l) \neq (i,j)$, which are independent of N_t^{ij}, then $\{T \leq t\}$ depends only on $(N_s^{ij} : s \leq t)$. So, by the strong Markov property of the Poisson process $\widetilde{N}_t^{ij} := N_{T+t}^{ij} - N_T^{ij}$ is a Poisson process of rate q_{ij} independent of $(N_s^{ij} : s \leq T)$, and independent of X_0 and $(N_t^{kl})_{t\geq 0}$ for $(k,l) \neq (i,j)$. Hence, conditional on $T < \infty$ and $X_T = i$, $(X_{T+t})_{t\geq 0}$ has the same distribution as $(X_t)_{t\geq 0}$ and is independent of $(X_s : s \leq T)$. In particular, we can take $T = J_n$ to see that, conditional on $J_n < \infty$ and $Y_n = i$, S_{n+1} is exponential of parameter q_i, Y_{n+1} has distribution $(\pi_{ij} : j \in I)$, and S_{n+1} and Y_{n+1} are independent, and independent of $Y_0, \dots, Y_n$ and $S_1, \dots, S_n$. Hence $(X_t)_{t\geq 0}$ is Markov(λ, Q) and, moreover, $(X_t)_{t\geq 0}$ has the strong Markov property. The conditioning on which this argument relies requires some further justification, especially when the state-space is infinite, so we shall not rely on this third construction in the development of the theory.

2.7 Explosion

We saw in the special case of birth processes that, although each holding time is strictly positive, one can run through a sequence of states with shorter and shorter holding times and end up taking infinitely many jumps in a finite time. This phenomenon is called explosion. Recall the notation of Section 2.2: for a process with jump times $J_0, J_1, J_2, \dots$ and holding times $S_1, S_2, \dots$, the explosion time ζ is given by

$$\zeta = \sup_n J_n = \sum_{n=1}^{\infty} S_n.$$

Theorem 2.7.1. *Let $(X_t)_{t\geq 0}$ be Markov(λ, Q). Then $(X_t)_{t\geq 0}$ does not explode if any one of the following conditions holds:*

(i) *I is finite;*
(ii) *$\sup_{i\in I} q_i < \infty$;*
(iii) *$X_0 = i$, and i is recurrent for the jump chain.*

Proof. Set $T_n = q(Y_{n-1})S_n$, then $T_1, T_2, \dots$ are independent $E(1)$ and independent of $(Y_n)_{n\geq 0}$. In cases (i) and (ii), $q = \sup_i q_i < \infty$ and

$$q\zeta \geq \sum_{n=1}^{\infty} T_n = \infty$$

with probability 1. In case (iii), we know that $(Y_n)_{n\geq 0}$ visits i infinitely often, at times $N_1, N_2, \dots$, say. Then

$$q_i\zeta \geq \sum_{m=1}^{\infty} T_{N_m+1} = \infty$$

with probability 1. □

We say that a Q-matrix Q is *explosive* if, for the associated Markov chain

$$\mathbb{P}_i(\zeta < \infty) > 0 \quad \text{for some } i \in I.$$

Otherwise Q is *non-explosive*. Here as in Chapter 1 we denote by $\mathbb{P}_i$ the conditional probability $\mathbb{P}_i(A) = \mathbb{P}(A|X_0 = i)$. It is a simple consequence of the Markov property for $(Y_n)_{n\geq 0}$ that under $\mathbb{P}_i$ the process $(X_t)_{t\geq 0}$ is Markov(δ_i, Q). The result just proved gives simple conditions for non-explosion and covers many cases of interest. As a corollary to the next result we shall obtain necessary and sufficient conditions for Q to be explosive, but these are not as easy to apply as Theorem 2.7.1.

Theorem 2.7.2. *Let $(X_t)_{t\geq 0}$ be a continuous-time Markov chain with generator matrix Q and write ζ for the explosion time of $(X_t)_{t\geq 0}$. Fix $\theta > 0$ and set $z_i = \mathbb{E}_i(e^{-\theta\zeta})$. Then $z = (z_i : i \in I)$ satisfies:*

(i) $|z_i| \leq 1$ *for all* i;
(ii) $Qz = \theta z$.

Moreover, if $\tilde{z}$ also satisfies (i) *and* (ii), *then $\tilde{z}_i \leq z_i$ for all i.*

Proof. Condition on $X_0 = i$. The time and place of the first jump are independent, J_1 is $E(q_i)$ and

$$\mathbb{P}_i(X_{J_1} = k) = \pi_{ik}.$$

Moreover, by the Markov property of the jump chain at time $n = 1$, conditional on $X_{J_1} = k$, $(X_{J_1+t})_{t\geq 0}$ is Markov(δ_k, Q) and independent of J_1. So

$$\begin{aligned}\mathbb{E}_i(e^{-\theta\zeta} \mid X_{J_1} = k) &= \mathbb{E}_i(e^{-\theta J_1} e^{-\theta\sum_{n=2}^{\infty} S_n} \mid X_{J_1} = k) \\ &= \int_0^{\infty} e^{-\theta t} q_i e^{-q_i t} dt\, \mathbb{E}_k(e^{-\theta\zeta}) = \frac{q_i z_k}{q_i + \theta}\end{aligned}$$

and

$$z_i = \sum_{k\neq i} \mathbb{P}_i(X_{J_1} = k)\mathbb{E}_i(e^{-\theta\zeta} \mid X_{J_1} = k) = \sum_{k\neq i} \frac{q_i \pi_{ik} z_k}{q_i + \theta}\,.$$

Recall that $q_i = -q_{ii}$ and $q_i \pi_{ik} = q_{ik}$. Then

$$(\theta - q_{ii}) z_i = \sum_{k \neq i} q_{ik} z_k$$

so

$$\theta z_i = \sum_{k \in I} q_{ik} z_k$$

and so z satisfies (i) and (ii). Note that the same argument also shows that

$$\mathbb{E}_i(e^{-\theta J_{n+1}}) = \sum_{k \neq i} \frac{q_i \pi_{ik}}{q_i + \theta} \mathbb{E}_k(e^{-\theta J_n}).$$

Suppose that $\widetilde{z}$ also satisfies (i) and (ii), then, in particular

$$\widetilde{z}_i \leq 1 = \mathbb{E}_i(e^{-\theta J_0})$$

for all i. Suppose inductively that

$$\widetilde{z}_i \leq \mathbb{E}_i(e^{-\theta J_n})$$

then, since $\widetilde{z}$ satisfies (ii)

$$\widetilde{z}_i = \sum_{k \neq i} \frac{q_i \pi_{ik}}{q_i + \theta} \widetilde{z}_k \leq \sum_{k \neq i} \frac{q_i \pi_{ik}}{q_i + \theta} \mathbb{E}_i(e^{-\theta J_n}) = \mathbb{E}_i(e^{-\theta J_{n+1}}).$$

Hence $\widetilde{z}_i \leq \mathbb{E}_i(e^{-\theta J_n})$ for all n. By monotone convergence

$$\mathbb{E}_i(e^{-\theta J_n}) \to \mathbb{E}_i(e^{-\theta \zeta})$$

as $n \to \infty$, so $\widetilde{z}_i \leq z_i$ for all i. □

Corollary 2.7.3. *For each $\theta > 0$ the following are equivalent:*

(a) *Q is non-explosive;*
(b) *$Qz = \theta z$ and $|z_i| \leq 1$ for all i imply $z = 0$.*

Proof. If (a) holds then $\mathbb{P}_i(\zeta = \infty) = 1$ so $\mathbb{E}_i(e^{-\theta\zeta}) = 0$. By the theorem, $Qz = \theta z$ and $|z| \leq 1$ imply $z_i \leq \mathbb{E}_i(e^{-\theta\zeta})$, hence $z \leq 0$, by symmetry $z \geq 0$, and hence (b) holds. On the other hand, if (b) holds, then by the theorem $\mathbb{E}_i(e^{-\theta\zeta}) = 0$ for all i, so $\mathbb{P}_i(\zeta = \infty) = 1$ and (a) holds. □

Exercise

2.7.1 Let $(X_t)_{t\geq 0}$ be a Markov chain on the integers with transition rates

$$q_{i,i+1} = \lambda q_i, \quad q_{i,i-1} = \mu q_i$$

and $q_{ij} = 0$ if $|j - i| \geq 2$, where $\lambda + \mu = 1$ and $q_i > 0$ for all i. Find for all integers i:

(a) the probability, starting from 0, that X_t hits i;

(b) the expected total time spent in state i, starting from 0.

In the case where $\mu = 0$, write down a necessary and sufficient condition for $(X_t)_{t\geq 0}$ to be explosive. Why is this condition necessary for $(X_t)_{t\geq 0}$ to be explosive for all $\mu \in [0, 1/2)$?

Show that, in general, $(X_t)_{t\geq 0}$ is non-explosive if and only if one of the following conditions holds:

(i) $\lambda = \mu$;

(ii) $\lambda > \mu$ and $\sum_{i=1}^{\infty} 1/q_i = \infty$;

(iii) $\lambda < \mu$ and $\sum_{i=1}^{\infty} 1/q_{-i} = \infty$.

2.8 Forward and backward equations

Although the definition of a continuous-time Markov chain in terms of its jump chain and holding times provides a clear picture of the process, it does not answer some basic questions. For example, we might wish to calculate $\mathbb{P}_i(X_t = j)$. In this section we shall obtain two more ways of characterizing a continuous-time Markov chain, which will in particular give us a means to find $\mathbb{P}_i(X_t = j)$. As for Poisson processes and birth processes, the first step is to deduce the *Markov property* from the jump chain/holding time definition. In fact, we shall give the *strong* Markov property as this is a fundamental result and the proof is not much harder. However, the proof of both results really requires the precision of measure theory, so we have deferred it to Section 6.5. If you want to understand what happens, Theorem 2.4.1 on the Poisson process gives the main idea in a simpler context.

Recall that a random variable T with values in $[0, \infty]$ is a stopping time of $(X_t)_{t\geq 0}$ if for each $t \in [0, \infty)$ the event $\{T \leq t\}$ depends only on $(X_s : s \leq t)$.

Theorem 2.8.1 (Strong Markov property). *Let $(X_t)_{t\geq 0}$ be Markov(λ, Q) and let T be a stopping time of $(X_t)_{t\geq 0}$. Then, conditional on $T < \infty$ and $X_T = i$, $(X_{T+t})_{t\geq 0}$ is Markov(δ_i, Q) and independent of $(X_s : s \leq T)$.*

We come to the key result for continuous-time Markov chains. We shall present first a version for the case of finite state-space, where there is a simpler proof. In this case there are three alternative definitions, just as for the Poisson process.

Theorem 2.8.2. *Let $(X_t)_{t\geq 0}$ be a right-continuous process with values in a finite set I. Let Q be a Q-matrix on I with jump matrix Π. Then the following three conditions are equivalent:*

(a) (jump chain/holding time definition) *conditional on $X_0 = i$, the jump chain $(Y_n)_{n\geq 0}$ of $(X_t)_{t\geq 0}$ is discrete-time Markov(δ_i, Π) and for each $n \geq 1$, conditional on $Y_0, \dots, Y_{n-1}$, the holding times $S_1, \dots, S_n$ are independent exponential random variables of parameters $q(Y_0), \dots, q(Y_{n-1})$ respectively;*

(b) (infinitesimal definition) *for all $t, h \geq 0$, conditional on $X_t = i$, X_{t+h} is independent of $(X_s : s \leq t)$ and, as $h \downarrow 0$, uniformly in t, for all j*

$$\mathbb{P}(X_{t+h} = j \mid X_t = i) = \delta_{ij} + q_{ij}h + o(h);$$

(c) (transition probability definition) *for all $n = 0, 1, 2, \dots$, all times $0 \leq t_0 \leq t_1 \leq \dots \leq t_{n+1}$ and all states $i_0, \dots, i_{n+1}$*

$$\mathbb{P}(X_{t_{n+1}} = i_{n+1} \mid X_{t_0} = i_0, \dots, X_{t_n} = i_n) = p_{i_n i_{n+1}}(t_{n+1} - t_n)$$

where $(p_{ij}(t) : i, j \in I, t \geq 0)$ is the solution of the forward equation

$$P'(t) = P(t)Q, \quad P(0) = I.$$

If $(X_t)_{t\geq 0}$ satisfies any of these conditions then it is called a *Markov chain with generator matrix Q*. We say that $(X_t)_{t\geq 0}$ is *Markov*(λ, Q) for short, where λ is the distribution of X_0.

Proof. (a) $\Rightarrow$ (b) Suppose (a) holds, then, as $h \downarrow 0$,

$$\mathbb{P}_i(X_h = i) \geq \mathbb{P}_i(J_1 > h) = e^{-q_i h} = 1 + q_{ii}h + o(h)$$

and for $j \neq i$ we have

$$\begin{aligned}\mathbb{P}_i(X_h = j) &\geq \mathbb{P}_(J_1 \leq h, Y_1 = j, S_2 > h) \\ &= (1 - e^{-q_i h})\pi_{ij} e^{-q_j h} = q_{ij}h + o(h).\end{aligned}$$

Thus for every state j there is an inequality

$$\mathbb{P}_i(X_h = j) \geq \delta_{ij} + q_{ij}h + o(h)$$

and by taking the finite sum over j we see that these must in fact be equalities. Then by the Markov property, for any $t, h \geq 0$, conditional on $X_t = i$, X_{t+h} is independent of $(X_s : s \leq t)$ and, as $h \downarrow 0$, uniformly in t

$$\mathbb{P}(X_{t+h} = j \mid X_t = i) = \mathbb{P}_i(X_h = j) = \delta_{ij} + q_{ij}h + o(h).$$

(b) ⇒ (c) Set $p_{ij}(t) = \mathbb{P}_i(X_t = j) = \mathbb{P}(X_t = j \mid X_0 = i)$. If (b) holds, then for all $t, h \geq 0$, as $h \downarrow 0$, uniformly in t

$$\begin{aligned} p_{ij}(t+h) &= \sum_{k \in I} \mathbb{P}_i(X_t = k)\mathbb{P}(X_{t+h} = j \mid X_t = k) \\ &= \sum_{k \in I} p_{ik}(t)(\delta_{kj} + q_{kj}h + o(h)). \end{aligned}$$

Since I is finite we have

$$\frac{p_{ij}(t+h) - p_{ij}(t)}{h} = \sum_{k \in I} p_{ik}(t)q_{kj} + O(h)$$

so, letting $h \downarrow 0$, we see that $p_{ij}(t)$ is differentiable on the right. Then by uniformity we can replace t by $t - h$ in the above and let $h \downarrow 0$ to see first that $p_{ij}(t)$ is continuous on the left, then differentiable on the left, hence differentiable, and satisfies the forward equations

$$p'_{ij}(t) = \sum_{k \in I} p_{ik}(t)q_{kj}, \quad p_{ij}(0) = \delta_{ij}.$$

Since I is finite, $p_{ij}(t)$ is then the unique solution by Theorem 2.1.1. Also, if (b) holds, then

$$\mathbb{P}(X_{t_{n+1}} = i_{n+1} \mid X_{t_0} = i_0, \dots, X_{t_n} = i_n) = \mathbb{P}(X_{t_{n+1}} = i_{n+1} \mid X_{t_n} = i_n)$$

and, moreover, (b) holds for $(X_{t_n+t})_{t \geq 0}$ so, by the above argument,

$$\mathbb{P}(X_{t_{n+1}} = i_{n+1} \mid X_{t_n} = i_n) = p_{i_n i_{n+1}}(t_{n+1} - t_n),$$

proving (c).

(c) ⇒ (a) See the proof of Theorem 2.4.3. □

We know from Theorem 2.1.1 that for I finite the forward and backward equations have the same solution. So in condition (c) of the result just proved we could replace the forward equation with the backward equation. Indeed, there is a slight variation of the argument from (b) to (c) which leads directly to the backward equation.

The deduction of (c) from (b) above can be seen as the matrix version of the following result: for $q \in \mathbb{R}$ we have

$$\left(1+\frac{q}{n}+o\left(\frac{1}{n}\right)\right)^n \to e^q \qquad \text{as } n \to \infty.$$

Suppose (b) holds and set

$$p_{ij}(t,t+h) = \mathbb{P}(X_{t+h}=j \mid X_t = i);$$

then $P(t,t+h) = (p_{ij}(t,t+h) : i,j \in I)$ satisfies

$$P(t,t+h) = I + Qh + o(h)$$

and

$$P(0,t) = P\left(0,\frac{t}{n}\right)P\left(\frac{t}{n},\frac{2t}{n}\right)\dots P\left(\frac{(n-1)t}{n},t\right) = \left(I+\frac{tQ}{n}+o\left(\frac{1}{n}\right)\right)^n.$$

Some care is needed in making this precise, since the $o(h)$ terms, though uniform in t, are not *a priori* identical. On the other hand, in (c) we see that

$$P(0,t) = e^{tQ}.$$

We turn now to the case of infinite state-space. The backward equation may still be written in the form

$$P'(t) = QP(t), \quad P(0) = I$$

only now we have an infinite system of differential equations

$$p'_{ij}(t) = \sum_{k \in I} q_{ik}p_{kj}(t), \quad p_{ij}(0) = \delta_{ij}$$

and the results on matrix exponentials given in Section 2.1 no longer apply. A solution to the backward equation is any matrix $(p_{ij}(t) : i,j \in I)$ of differentiable functions satisfying this system of differential equations.

Theorem 2.8.3. *Let Q be a Q-matrix. Then the backward equation*

$$P'(t) = QP(t), \quad P(0) = I$$

has a minimal non-negative solution $(P(t) : t \geq 0)$. This solution forms a matrix semigroup

$$P(s)P(t) = P(s+t) \qquad \textit{for all } s,t \geq 0.$$

We shall prove this result by a probabilistic method in combination with Theorem 2.8.4. Note that if I is finite we must have $P(t) = e^{tQ}$ by Theorem 2.1.1. We call $(P(t) : t \geq 0)$ the *minimal non-negative semigroup* associated to Q, or simply the *semigroup* of Q, the qualifications *minimal* and *non-negative* being understood.

Here is the key result for Markov chains with infinite state-space. There are just two alternative definitions now as the infinitesimal characterization becomes problematic for infinite state-space.

Theorem 2.8.4. *Let $(X_t)_{t\geq 0}$ be a minimal right-continuous process with values in I. Let Q be a Q-matrix on I with jump matrix Π and semigroup $(P(t) : t \geq 0)$. Then the following conditions are equivalent:*

(a) (jump chain/holding time definition) *conditional on $X_0 = i$, the jump chain $(Y_n)_{n\geq 0}$ of $(X_t)_{t\geq 0}$ is discrete-time Markov(δ_i, Π) and for each $n \geq 1$, conditional on $Y_0, \ldots, Y_{n-1}$, the holding times $S_1, \ldots, S_n$ are independent exponential random variables of parameters $q(Y_0), \ldots, q(Y_{n-1})$ respectively;*

(b) (transition probability definition) *for all $n = 0, 1, 2, \ldots$, all times $0 \leq t_0 \leq t_1 \leq \ldots \leq t_{n+1}$ and all states $i_0, i_1, \ldots, i_{n+1}$*

$$\mathbb{P}(X_{t_{n+1}} = i_{n+1} \mid X_{t_0} = i_0, \ldots, X_{t_n} = i_n) = p_{i_n i_{n+1}}(t_{n+1} - t_n).$$

If $(X_t)_{t\geq 0}$ satisfies any of these conditions then it is called a *Markov chain with generator matrix Q*. We say that $(X_t)_{t\geq 0}$ is *Markov(λ, Q)* for short, where λ is the distribution of X_0.

Proof of Theorems 2.8.3 and 2.8.4. We know that there exists a process $(X_t)_{t\geq 0}$ satisfying (a). So let us *define* $P(t)$ by

$$p_{ij}(t) = \mathbb{P}_i(X_t = j).$$

Step 1. *We show that $P(t)$ satisfies the backward equation.*

Conditional on $X_0 = i$ we have $J_1 \sim E(q_i)$ and $X_{J_1} \sim (\pi_{ik} : k \in I)$. Then conditional on $J_1 = s$ and $X_{J_1} = k$ we have $(X_{s+t})_{t\geq 0} \sim$ Markov(δ_k, Q). So

$$\mathbb{P}_i(X_t = j, t < J_1) = e^{-q_i t}\delta_{ij}$$

and

$$\mathbb{P}_i(J_1 \leq t, X_{J_1} = k, X_t = j) = \int_0^t q_i e^{-q_i s}\pi_{ik} p_{kj}(t-s)ds.$$

Therefore

$$\begin{aligned} p_{ij}(t) &= \mathbb{P}_i(X_t = j, t < J_1) + \sum_{k\neq i} \mathbb{P}_i(J_1 \leq t, X_{J_1} = k, X_t = j) \\ &= e^{-q_i t}\delta_{ij} + \sum_{k\neq i} \int_0^t q_i e^{-q_i s}\pi_{ik} p_{kj}(t-s)ds. \end{aligned} \tag{2.1}$$

Make a change of variable $u = t - s$ in each of the integrals, interchange sum and integral by monotone convergence and multiply by $e^{q_i t}$ to obtain

$$e^{q_i t} p_{ij}(t) = \delta_{ij} + \int_0^t \sum_{k\neq i} q_i e^{q_i u}\pi_{ik} p_{kj}(u)du. \tag{2.2}$$

This equation shows, firstly, that $p_{ij}(t)$ is continuous in t for all i, j. Secondly, the integrand is then a uniformly converging sum of continuous functions, hence continuous, and hence $p_{ij}(t)$ is differentiable in t and satisfies

$$e^{q_i t}(q_i p_{ij}(t) + p'_{ij}(t)) = \sum_{k \neq i} q_i e^{q_i t} \pi_{ik} p_{kj}(t).$$

Recall that $q_i = -q_{ii}$ and $q_{ik} = q_i \pi_{ik}$ for $k \neq i$. Then, on rearranging, we obtain

$$p'_{ij}(t) = \sum_{k \in I} q_{ik} p_{kj}(t) \tag{2.3}$$

so $P(t)$ satisfies the backward equation.

The integral equation (2.1) is called the *integral form of the backward equation.*

Step 2. *We show that if $\widetilde{P}(t)$ is another non-negative solution of the backward equation, then $P(t) \leq \widetilde{P}(t)$, hence $P(t)$ is the minimal non-negative solution.*

The argument used to prove (2.1) also shows that

$$\begin{aligned} &\mathbb{P}_i(X_t = j, t < J_{n+1}) \\ &\quad = e^{-q_i t}\delta_{ij} + \sum_{k \neq i} \int_0^t q_i e^{-q_i s} \pi_{ik} \mathbb{P}_k(X_{t-s} = j, t - s < J_n) ds. \end{aligned} \tag{2.4}$$

On the other hand, if $\widetilde{P}(t)$ satisfies the backward equation, then, by reversing the steps from (2.1) to (2.3), it also satisfies the integral form:

$$\widetilde{p}_{ij}(t) = e^{-q_i t}\delta_{ij} + \sum_{k \neq i} \int_0^t q_i e^{-q_i s} \pi_{ik} \widetilde{p}_{kj}(t - s) ds. \tag{2.5}$$

If $\widetilde{P}(t) \geq 0$, then

$$\mathbb{P}_i(X_t = j, t < J_0) = 0 \leq \widetilde{p}_{ij}(t) \qquad \text{for all } i, j \text{ and } t.$$

Let us suppose inductively that

$$\mathbb{P}_i(X_t = j, t < J_n) \leq \widetilde{p}_{ij}(t) \qquad \text{for all } i, j \text{ and } t,$$

then by comparing (2.4) and (2.5) we have

$$\mathbb{P}_i(X_t = j, t < J_{n+1}) \leq \widetilde{p}_{ij}(t) \qquad \text{for all } i, j \text{ and } t,$$

and the induction proceeds. Hence

$$p_{ij}(t) = \lim_{n\to\infty} \mathbb{P}_i(X_t = j, t < J_n) \le \widetilde{p}_{ij}(t) \qquad \text{for all } i, j \text{ and } t.$$

Step 3. Since $(X_t)_{t\ge 0}$ does not return from ∞ we have

$$\begin{aligned} p_{ij}(s+t) = \mathbb{P}_i(X_{s+t} = j) &= \sum_{k\in I} \mathbb{P}_i(X_{s+t} = j \mid X_s = k)\mathbb{P}_i(X_s = k) \\ &= \sum_{k\in I} \mathbb{P}_i(X_s = k)\mathbb{P}_k(X_t = j) = \sum_{k\in I} p_{ik}(s)p_{kj}(t) \end{aligned}$$

by the Markov property. Hence $(P(t) : t \ge 0)$ is a matrix semigroup. This completes the proof of Theorem 2.8.3.

Step 4. Suppose, as we have throughout, that $(X_t)_{t\ge 0}$ satisfies (a). Then, by the Markov property

$$\begin{aligned} &\mathbb{P}(X_{t_{n+1}} = i_{n+1} \,|X_{t_0} = i_0, \dots, X_{t_n} = i_n) \\ &\qquad\qquad = \mathbb{P}_{i_n}(X_{t_{n+1}-t_n} = i_{n+1}) = p_{i_n i_{n+1}}(t_{n+1} - t_n) \end{aligned}$$

so $(X_t)_{t\ge 0}$ satisfies (b). We complete the proof of Theorem 2.8.4 by the usual argument that (b) must now imply (a) (see the proof of Theorem 2.4.3, (c) $\Rightarrow$ (a)). □

So far we have said nothing about the forward equation in the case of infinite state-space. Remember that the finite state-space results of Section 2.1 are no longer valid. The forward equation may still be written

$$P'(t) = P(t)Q, \quad P(0) = I,$$

now understood as an infinite system of differential equations

$$p'_{ij}(t) = \sum_{k\in I} p_{ik}(t)q_{kj}, \quad p_{ij}(0) = \delta_{ij}.$$

A solution is then any matrix $\big(p_{ij}(t) : i, j \in I\big)$ of differentiable functions satisfying this system of equations. We shall show that the semigroup $\big(P(t) : t \ge 0\big)$ of Q does satisfy the forward equations, by a probabilistic argument resembling Step 1 of the proof of Theorems 2.8.3 and 2.8.4. This time, instead of conditioning on the first event, we condition on the last event before time t. The argument is a little longer because there is no reverse-time Markov property to give the conditional distribution. We need the following time-reversal identity, a simple version of which was given in Theorem 2.3.4.

Lemma 2.8.5. *We have*

$$q_{i_n}\mathbb{P}(J_n \le t < J_{n+1} \mid Y_0 = i_0, Y_1 = i_1, \dots, Y_n = i_n)$$
$$= q_{i_0}\mathbb{P}(J_n \le t < J_{n+1} \mid Y_0 = i_n, \dots, Y_{n-1} = i_1, Y_n = i_0).$$

Proof. Conditional on $Y_0 = i_0, \dots, Y_n = i_n$, the holding times $S_1, \dots, S_{n+1}$ are independent with $S_k \sim E(q_{i_{k-1}})$. So the left-hand side is given by

$$\int_{\Delta(t)} q_{i_n} \exp\{-q_{i_n}(t - s_1 - \dots - s_n)\} \prod_{k=1}^{n} q_{i_{k-1}} \exp\{-q_{i_{k-1}} s_k\} ds_k$$

where $\Delta(t) = \{(s_1, \dots, s_n) : s_1 + \dots + s_n \le t \text{ and } s_1, \dots, s_n \ge 0\}$. On making the substitutions $u_1 = t - s_1 - \dots - s_n$ and $u_k = s_{n-k+2}$, for $k = 2, \dots, n$, we obtain

$$q_{i_n}\mathbb{P}(J_n \le t < J_{n+1} \mid Y_0 = i_0, \dots, Y_n = i_n)$$
$$= \int_{\Delta(t)} q_{i_0} \exp\{-q_{i_0}(t - u_1 - \dots - u_n)\} \prod_{k=1}^{n} q_{i_{n-k+1}} \exp\{-q_{i_{n-k+1}} u_k\} du_k$$
$$= q_{i_0}\mathbb{P}(J_n \le t < J_{n+1} \mid Y_0 = i_n, \dots, Y_{n-1} = i_1, Y_n = i_0). \qquad \square$$

Theorem 2.8.6. *The minimal non-negative solution $\big(P(t) : t \ge 0\big)$ of the backward equation is also the minimal non-negative solution of the forward equation*

$$P'(t) = P(t)Q, \quad P(0) = I.$$

Proof. Let $(X_t)_{t\ge 0}$ denote the minimal Markov chain with generator matrix Q. By Theorem 2.8.4

$$p_{ij}(t) = \mathbb{P}_i(X_t = j)$$
$$= \sum_{n=0}^{\infty} \sum_{k \ne j} \mathbb{P}_i(J_n \le t < J_{n+1}, Y_{n-1} = k, Y_n = j).$$

Now by Lemma 2.8.5, for $n \ge 1$, we have

$$\mathbb{P}_i(J_n \le t < J_{n+1} \mid Y_{n-1} = k, Y_n = j)$$
$$= (q_i/q_j)\mathbb{P}_j(J_n \le t < J_{n+1} \mid Y_1 = k, Y_n = i)$$
$$= (q_i/q_j) \int_0^t q_j e^{-q_j s} \mathbb{P}_k(J_{n-1} \le t - s < J_n \mid Y_{n-1} = i) ds$$
$$= q_i \int_0^t e^{-q_j s} (q_k/q_i) \mathbb{P}_i(J_{n-1} \le t - s < J_n \mid Y_{n-1} = k) ds$$

where we have used the Markov property of $(Y_n)_{n\geq 0}$ for the second equality. Hence

$$\begin{aligned} p_{ij}(t) = \delta_{ij}e^{-q_i t} + \sum_{n=1}^{\infty}\sum_{k\neq j}\int_0^t \mathbb{P}_i(J_{n-1} \leq t-s < J_n \mid Y_{n-1} = k) \\ \times \mathbb{P}_i(Y_{n-1} = k, Y_n = j) q_k e^{-q_j s} ds \\ = \delta_{ij}e^{-q_i t} + \sum_{n=1}^{\infty}\sum_{k\neq j}\int_0^t \mathbb{P}_i(J_{n-1} \leq t-s < J_n, Y_{n-1} = k) q_k \pi_{kj} e^{-q_j s} ds \\ = \delta_{ij}e^{-q_i t} + \int_0^t \sum_{k\neq j} p_{ik}(t-s) q_{kj} e^{-q_j s} ds \end{aligned} \qquad (2.6)$$

where we have used monotone convergence to interchange the sum and integral at the last step. This is the *integral form of the forward equation.* Now make a change of variable $u = t - s$ in the integral and multiply by $e^{q_j t}$ to obtain

$$p_{ij}(t)e^{q_j t} = \delta_{ij} + \int_0^t \sum_{k\neq j} p_{ik}(u) q_{kj} e^{q_j u} du. \qquad (2.7)$$

We know by equation (2.2) that $e^{q_i t} p_{ik}(t)$ is *increasing* for all i, k. Hence either

$$\sum_{k\neq j} p_{ik}(u) q_{kj} \quad \text{converges uniformly for} \quad u \in [0, t]$$

or

$$\sum_{k\neq j} p_{ik}(u) q_{kj} = \infty \quad \text{for all } u \geq t.$$

The latter would contradict (2.7) since the left-hand side is finite for all t, so it is the former which holds. We know from the backward equation that $p_{ij}(t)$ is continuous for all i, j; hence by uniform convergence the integrand in (2.7) is continuous and we may differentiate to obtain

$$p'_{ij}(t) + p_{ij}(t) q_j = \sum_{k\neq j} p_{ik}(t) q_{kj}.$$

Hence $P(t)$ solves the forward equation.

To establish minimality let us suppose that $\widetilde{p}_{ij}(t)$ is another solution of the forward equation; then we also have

$$\widetilde{p}_{ij}(t) = \delta_{ij}e^{-q_i t} + \sum_{k\neq j}\int_0^t \widetilde{p}_{ik}(t-s) q_{kj} e^{-q_j s} ds.$$

A small variation of the argument leading to (2.6) shows that, for $n \geq 0$

$$\begin{aligned}\mathbb{P}_i(X_t = j, t < J_{n+1}) \\ = \delta_{ij} e^{-q_i t} + \sum_{k \neq j} \int_0^t \mathbb{P}_i(X_t = j, t < J_n) q_{kj} e^{-q_j s} ds. \end{aligned} \tag{2.8}$$

If $\widetilde{P}(t) \geq 0$, then

$$\mathbb{P}(X_t = j, t < J_0) = 0 \leq \widetilde{p}_{ij}(t) \quad \text{for all } i, j \text{ and } t.$$

Let us suppose inductively that

$$\mathbb{P}_i(X_t = j, t < J_n) \leq \widetilde{p}_{ij}(t) \quad \text{for all } i, j \text{ and } t;$$

then by comparing (2.7) and (2.8) we obtain

$$\mathbb{P}_i(X_t = j, t < J_{n+1}) \leq \widetilde{p}_{ij}(t) \quad \text{for all } i, j \text{ and } t$$

and the induction proceeds. Hence

$$p_{ij}(t) = \lim_{n \to \infty} \mathbb{P}_i(X_t = j, t < J_n) \leq \widetilde{p}_{ij}(t) \quad \text{for all } i, j \text{ and } t. \qquad \square$$

Exercises

2.8.1 Two fleas are bound together to take part in a nine-legged race on the vertices A, B, C of a triangle. Flea 1 hops at random times in the clockwise direction; each hop takes the pair from one vertex to the next and the times between successive hops of Flea 1 are independent random variables, each with with exponential distribution, mean $1/\lambda$. Flea 2 behaves similarly, but hops in the anticlockwise direction, the times between his hops having mean $1/\mu$. Show that the probability that they are at A at a given time $t > 0$ (starting from A at time $t = 0$) is

$$\frac{1}{3} + \frac{2}{3} \exp\left\{-\frac{3(\lambda + \mu)t}{2}\right\} \cos\left\{\frac{\sqrt{3}(\lambda - \mu)t}{2}\right\} .$$

2.8.2 Let $(X_t)_{t \geq 0}$ be a birth-and-death process with rates $\lambda_n = n\lambda$ and $\mu_n = n\mu$, and assume that $X_0 = 1$. Show that $h(t) = \mathbb{P}(X_t = 0)$ satisfies

$$h(t) = \int_0^t e^{-(\lambda + \mu)s} \{\mu + \lambda h(t - s)^2\} ds$$

and deduce that if $\lambda \neq \mu$ then

$$h(t) = (\mu e^{\mu t} - \mu e^{\lambda t})/(\mu e^{\mu t} - \lambda e^{\lambda t}) .$$

2.9 Non-minimal chains

This book concentrates entirely on processes which are right-continuous and minimal. These are the simplest sorts of process and, overwhelmingly, the ones of greatest practical application. We have seen in this chapter that we can associate to each distribution λ and Q-matrix Q a unique such process, the Markov chain with initial distribution λ and generator matrix Q. Indeed we have taken the liberty of defining Markov chains to be those processes which arise in this way. However, these processes do not by any means exhaust the class of memoryless continuous-time processes with values in a countable set I. There are many more exotic possibilities, the general theory of which goes very much deeper than the account given in this book. It is in the nature of things that these exotic cases have received the greater attention among mathematicians. Here are some examples to help you imagine the possibilities.

Example 2.9.1

Consider a birth process $(X_t)_{t\geq 0}$ starting from 0 with rates $q_i = 2^i$ for $i \geq 0$. We have chosen these rates so that

$$\sum_{i=0}^{\infty} q_i^{-1} = \sum_{i=0}^{\infty} 2^{-i} < \infty$$

which shows that the process explodes (see Theorems 2.3.2 and 2.5.2). We have until now insisted that $X_t = \infty$ for all $t \geq \zeta$, where ζ is the explosion time. But another obvious possibility is to start the process off again from 0 at time ζ, and do the same for all subsequent explosions. An argument based on the memoryless property of the exponential distribution shows that for $0 \leq t_0 \leq \ldots \leq t_{n+1}$ this process satisfies

$$\mathbb{P}\big(X_{t_{n+1}} = i_{n+1} \mid X_{t_0} = i_0, \ldots, X_{t_n} = i_n\big) = p_{i_n i_{n+1}}(t_{n+1} - t_n)$$

for a semigroup of stochastic matrices $(P(t) : t \geq 0)$ on I. This is the defining property for a more general class of Markov chains. Note that the chain is no longer determined by λ and Q alone; the rule for bringing $(X_t)_{t\geq 0}$ back into I after explosion also has to be given.

Example 2.9.2

We make a variation on the preceding example. Suppose now that the jump chain $(Y_n)_{n\geq 0}$ of $(X_t)_{t\geq 0}$ is the Markov chain on $\mathbb{Z}$ which moves one step away from 0 with probability 2/3 and one step towards 0 with probability 1/3, with $\pi_{01} = \pi_{0,-1} = 1/2$, and that $Y_0 = 0$. Let the transition rates for $(X_t)_{t\geq 0}$ be $q_i = 2^{|i|}$ for $i \in \mathbb{Z}$. Then $(X_t)_{t\geq 0}$ is again explosive. (A simple way to see this using some results of Chapter 3 is to check that $(Y_n)_{n\geq 0}$ is transient but $(X_t)_{t\geq 0}$ has an invariant distribution – by solution of the detailed balance equations. Then Theorem 3.5.3 makes explosion inevitable.) Now there are two ways in which $(X_t)_{t\geq 0}$ can explode, either $X_t \to -\infty$ or $X_t \to \infty$.

The process may again be restarted at 0 after explosion. Alternatively, we may choose the restart randomly, and according to the way that explosion occurred. For example

$$X_\zeta = \begin{cases} 0 & \text{if } X_t \to -\infty \text{ as } t \uparrow \zeta \\ Z & \text{if } X_t \to \infty \text{ as } t \uparrow \zeta \end{cases}$$

where Z takes values ± 1 with probability 1/2.

Example 2.9.3

The processes in the preceding two examples, though no longer minimal, were at least right-continuous. Here is an altogether more exotic example, due to P. Lévy, which is not even right-continuous. Consider

$$D_n = \{k2^{-n} : k \in \mathbb{Z}^+\} \qquad \text{for } n \geq 0$$

and set $I = \cup_n D_n$. With each $i \in D_n \backslash D_{n-1}$ we associate an independent exponential random variable S_i of parameter $(2^n)^2$. There are 2^{n-1} states in $(D^n \backslash D^{n-1}) \cap [0, 1)$, so, for all $i \in I$

$$\mathbb{E}\left(\sum_{j\leq i} S_j\right) \leq (i+1) \sum_{n=0}^{\infty} 2^{n-1}(2^{-2n}) < \infty$$

and

$$\mathbb{P}\left(\sum_{j\leq i} S_j \to \infty \text{ as } i \to \infty\right) = 1.$$

Now define

$$X_t = \begin{cases} i & \text{if } \sum_{j<i} S_j \leq t < \sum_{j\leq i} S_j \text{ for some } i \in I \\ \infty & \text{otherwise.} \end{cases}$$

This process runs through all the dyadic rationals $i \in I$ in the usual order. It remains in $i \in D_n \backslash D_{n-1}$ for an exponential time of parameter 1. Between any two distinct states i and j it makes infinitely many visits to ∞. The Lebesgue measure of the set of times t when $X_t = \infty$ is zero. There is a semigroup of stochastic matrices $(P(t) : t \geq 0)$ on I such that, for $0 \leq t_0 \leq \ldots \leq t_{n+1}$

$$\mathbb{P}\big(X_{t_{n+1}} = i_{n+1} \mid X_{t_0} = i_0, \ldots, X_{t_n} = i_n\big) = p_{i_n i_{n+1}}(t_{n+1} - t_n).$$

In particular, $\mathbb{P}(X_t = \infty) = 0$ for all $t \geq 0$. The details may be found in *Markov Chains* by D. Freedman (Holden-Day, San Francisco, 1971).

We hope these three examples will serve to suggest some of the possibilities for more general continuous-time Markov chains. For further reading, see Freedman's book, or else *Markov Chains with Stationary Transition Probabilities* by K.-L. Chung (Springer, Berlin, 2nd edition, 1967), or *Diffusions, Markov Processes and Martingales, Vol 1: Foundations* by L. C. G. Rogers and D. Williams (Wiley, Chichester, 2nd edition, 1994).

2.10 Appendix: matrix exponentials

Define two norms on the space of real-valued $N \times N$-matrices

$$|Q| = \sup_{v \neq 0} |Qv|/|v|, \quad \|Q\|_\infty = \sup_{i,j} |q_{ij}|.$$

Obviously, $\|Q\|_\infty$ is finite for all Q and controls the size of the entries in Q. We shall show that the two norms are equivalent and that $|Q|$ is well adapted to sums and products of matrices, which $\|Q\|_\infty$ is not.

Lemma 2.10.1. *We have*

(a) $\|Q\|_\infty \leq |Q| \leq N\|Q\|_\infty$;

(b) $|Q_1 + Q_2| \leq |Q_1| + |Q_2|$ *and* $|Q_1 Q_2| \leq |Q_1||Q_2|$.

Proof. (a) For any vector v we have $|Qv| \leq |Q||v|$. In particular, for the vector $\varepsilon_j = (0, \ldots, 1, \ldots, 0)$, with 1 in the jth place, we have $|Q\varepsilon_j| \leq |Q|$. The supremum defining $\|Q\|_\infty$ is achieved, at j say, so

$$\|Q\|_\infty^2 \leq \sum_i (q_{ij})^2 = |Q\varepsilon_j|^2 \leq |Q|^2.$$

On the other hand

$$\begin{aligned}|Qv|^2 &= \sum_i \left(\sum_j q_{ij} v_j\right)^2 \\ &\le \sum_i \left(\sum_j \|Q\|_\infty |v_j|\right)^2 \\ &= N\|Q\|_\infty^2 \left(\sum_j |v_j|\right)^2\end{aligned}$$

and, by the Cauchy–Schwarz inequality

$$\left(\sum_j |v_j|\right)^2 \le N \sum_j v_j^2$$

so $|Qv|^2 \le N^2 \|Q\|_\infty^2 |v|^2$. This implies that $|Q| \le N\|Q\|_\infty$.

(b) For any vector v we have

$$\begin{aligned}&|(Q_1 + Q_2)v| \le |Q_1 v| + |Q_2 v| \le (|Q_1| + |Q_2|)|v|, \\ &|Q_1 Q_2 v| \le |Q_1||Q_2 v| \le |Q_1||Q_2||v|.\end{aligned}$$ □

Now for $n = 0, 1, 2, \dots,$ consider the finite sum

$$E(n) = \sum_{k=0}^{n} \frac{Q^k}{k!}.$$

For each i and j, and $m \le n$, we have

$$\begin{aligned}|(E(n) - E(m))_{ij}| &\le \|E(n) - E(m)\|_\infty \le |E(n) - E(m)| \\ &= \left| \sum_{k=m+1}^{n} \frac{Q^k}{k!} \right| \\ &\le \sum_{k=m+1}^{n} \frac{|Q|^k}{k!}.\end{aligned}$$

Since $|Q| \le N\|Q\|_\infty < \infty$, $\sum_{k=0}^{\infty} |Q|^k / k!$ converges by the ratio test, so

$$\sum_{k=m+1}^{n} \frac{|Q|^k}{k!} \longrightarrow 0 \qquad \text{as } m, n \to \infty.$$

Hence each component of $E(n)$ forms a Cauchy sequence, which therefore converges, proving that

$$e^Q = \sum_{k=0}^{\infty} \frac{Q^k}{k!}$$

is well defined and, indeed, that the power series

$$(e^{tQ})_{ij} = \sum_{k=0}^{\infty} \frac{(tQ)_{ij}^k}{k!}$$

has infinite radius of convergence for all i, j. Finally, for two commuting matrices Q_1 and Q_2 we have

$$\begin{aligned} e^{Q_1+Q_2} &= \sum_{n=0}^{\infty} \frac{(Q_1+Q_2)^n}{n!} \\ &= \sum_{n=0}^{\infty} \frac{1}{n!} \sum_{k=0}^{n} \frac{n!}{k!(n-k)!} Q_1^k Q_2^{n-k} \\ &= \sum_{k=0}^{\infty} \frac{Q_1^k}{k!} \sum_{n=k}^{\infty} \frac{Q_2^{n-k}}{(n-k)!} \\ &= e^{Q_1} e^{Q_2}. \end{aligned}$$

3

Continuous-time Markov chains II

This chapter brings together the discrete-time and continuous-time theories, allowing us to deduce analogues, for continuous-time chains, of all the results given in Chapter 1. All the facts from Chapter 2 that are necessary to understand this synthesis are reviewed in Section 3.1. You will require a reasonable understanding of Chapter 1 here, but, given such an understanding, this chapter should look reassuringly familiar. Exercises remain an important part of the text.

3.1 Basic properties

Let I be a countable set. Recall that a *Q-matrix* on I is a matrix $Q = (q_{ij} : i, j \in I)$ satisfying the following conditions:

(i) $0 \le -q_{ii} < \infty \quad \text{for all} \quad i;$

(ii) $q_{ij} \ge 0 \quad \text{for all} \quad i \ne j;$

(iii) $\displaystyle\sum_{j \in I} q_{ij} = 0 \quad \text{for all} \quad i.$

We set $q_i = q(i) = -q_{ii}$. Associated to any Q-matrix is a *jump matrix* $\Pi = (\pi_{ij} : i, j \in I)$ given by

$$\pi_{ij} = \begin{cases} q_{ij}/q_i & \text{if } j \ne i \text{ and } q_i \ne 0 \\ 0 & \text{if } j \ne i \text{ and } q_i = 0, \end{cases}$$

$$\pi_{ii} = \begin{cases} 0 & \text{if } q_i \ne 0 \\ 1 & \text{if } q_i = 0. \end{cases}$$

Note that Π is a stochastic matrix.

A *sub-stochastic matrix* on I is a matrix $P = (p_{ij} : i, j \in I)$ with non-negative entries and such that

$$\sum_{j \in I} p_{ij} \leq 1 \quad \text{for all} \quad i.$$

Associated to any Q-matrix is a *semigroup* $(P(t) : t \geq 0)$ of sub-stochastic matrices $P(t) = (p_{ij}(t) : i, j \in I)$. As the name implies we have

$$P(s)P(t) = P(s+t) \quad \text{for all} \quad s, t \geq 0.$$

You will need to be familiar with the following terms introduced in Section 2.2: *minimal right-continuous random process, jump times, holding times, jump chain* and *explosion*. Briefly, a right-continuous process $(X_t)_{t \geq 0}$ runs through a sequence of states $Y_0, Y_1, Y_2, \ldots$, being held in these states for times $S_1, S_2, S_3, \ldots$ respectively and jumping to the next state at times $J_1, J_2, J_3, \ldots$. Thus $J_n = S_1 + \ldots + S_n$. The discrete-time process $(Y_n)_{n \geq 0}$ is the jump chain, $(S_n)_{n \geq 1}$ are the holding times and $(J_n)_{n \geq 1}$ are the jump times. The explosion time ζ is given by

$$\zeta = \sum_{n=1}^{\infty} S_n = \lim_{n \to \infty} J_n.$$

For a minimal process we take a new state ∞ and insist that $X_t = \infty$ for all $t \geq \zeta$. An important point is that a minimal right-continuous process is determined by its jump chain and holding times.

The data for a continuous-time Markov chain $(X_t)_{t \geq 0}$ are a distribution λ and a Q-matrix Q. The distribution λ gives the *initial distribution*, the distribution of X_0. The Q-matrix is known as the *generator matrix* of $(X_t)_{t \geq 0}$ and determines how the process evolves from its initial state. We established in Section 2.8 that there are two different, but equivalent, ways to describe how the process evolves.

The first, in terms of jump chain and holding times, states that

(a) $(Y_n)_{n \geq 0}$ is Markov(λ, Π);

(b) conditional on $Y_0 = i_0, \ldots, Y_{n-1} = i_{n-1}$, the holding times $S_1, \ldots, S_n$ are independent exponential random variables of parameters $q_{i_0}, \ldots, q_{i_{n-1}}$.

Put more simply, given that the chain starts at i, it waits there for an exponential time of parameter q_i and then jumps to a new state, choosing state j with probability π_{ij}. It then starts afresh, forgetting what has gone before.

The second description, in terms of the semigroup, states that the finite-dimensional distributions of the process are given by

(c) for all $n = 0, 1, 2, \dots$, all times $0 \le t_0 \le t_1 \le \dots \le t_{n+1}$ and all states $i_0, i_1, \dots, i_{n+1}$

$$\mathbb{P}(X_{t_{n+1}} = i_{n+1} \mid X_{t_0} = i_0, \dots, X_{t_n} = i_n) = p_{i_n i_{n+1}}(t_{n+1} - t_n).$$

Again, put more simply, given that the chain starts at i, by time t it is found in state j with probability $p_{ij}(t)$. It then starts afresh, forgetting what has gone before. In the case where

$$\widetilde{p}_{i\infty}(t) := 1 - \sum_{j\in I} p_{ij}(t) > 0$$

the chain is found at ∞ with probability $\widetilde{p}_{i\infty}(t)$. The semigroup $P(t)$ is referred to as the *transition matrix* of the chain and its entries $p_{ij}(t)$ are the *transition probabilities*. This description implies that for all $h > 0$ the discrete skeleton $(X_{nh})_{n\ge0}$ is Markov$(\lambda, P(h))$. Strictly, in the explosive case, that is, when $P(t)$ is strictly sub-stochastic, we should say Markov$(\widetilde{\lambda}, \widetilde{P}(h))$, where $\widetilde{\lambda}$ and $\widetilde{P}(h)$ are defined on $I \cup \{\infty\}$, extending λ and $P(h)$ by $\widetilde{\lambda}_\infty = 0$ and $\widetilde{p}_{\infty j}(h) = 0$. But there is no danger of confusion in using the simpler notation.

The information coming from these two descriptions is sufficient for most of the analysis of continuous-time chains done in this chapter. Note that we have not yet said how the semigroup $P(t)$ is associated to the Q-matrix Q, except via the process! This extra information will be required when we discuss reversibility in Section 3.7. So we recall from Section 2.8 that the semigroup is characterized as the minimal non-negative solution of the *backward equation*

$$P'(t) = QP(t), \quad P(0) = I$$

which reads in components

$$p'_{ij}(t) = \sum_{k\in I} q_{ik} p_{kj}(t), \quad p_{ij}(0) = \delta_{ij}.$$

The semigroup is also the minimal non-negative solution of the *forward equation*

$$P'(t) = P(t)Q, \quad P(0) = I.$$

In the case where I is finite, $P(t)$ is simply the matrix exponential e^{tQ}, and is the *unique* solution of the backward and forward equations.

3.2 Class structure

A first step in the analysis of a continuous-time Markov chain $(X_t)_{t\geq 0}$ is to identify its class structure. We emphasise that we deal only with minimal chains, those that die after explosion. Then the class structure is simply the discrete-time class structure of the jump chain $(Y_n)_{n\geq 0}$, as discussed in Section 1.2.

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

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

We say i *communicates with* j and write $i \leftrightarrow j$ if both $i \to j$ and $j \to i$. The notions of *communicating class, closed class, absorbing state* and *irreducibility* are inherited from the jump chain.

Theorem 3.2.1. *For distinct states i and j the following are equivalent:*

(i) $i \to j$;
(ii) $i \to j$ *for the jump chain;*
(iii) $q_{i_0 i_1} q_{i_1 i_2} \dots q_{i_{n-1} i_n} > 0$ *for some states* $i_0, i_1, \dots, i_n$ *with* $i_0 = i$, $i_n = j$;
(iv) $p_{ij}(t) > 0$ *for all* $t > 0$;
(v) $p_{ij}(t) > 0$ *for some* $t > 0$.

Proof. Implications (iv) $\Rightarrow$ (v) $\Rightarrow$ (i) $\Rightarrow$ (ii) are clear. If (ii) holds, then by Theorem 1.2.1, there are states $i_0, i_1, \dots, i_n$ with $i_0 = i$, $i_n = j$ and $\pi_{i_0 i_1} \pi_{i_1 i_2} \dots \pi_{i_{n-1} i_n} > 0$, which implies (iii). If $q_{ij} > 0$, then

$$p_{ij}(t) \geq \mathbb{P}_i(J_1 \leq t, Y_1 = j, S_2 > t) = (1 - e^{-q_i t})\pi_{ij} e^{-q_j t} > 0$$

for all $t > 0$, so if (iii) holds, then

$$p_{ij}(t) \geq p_{i_0 i_1}(t/n) \dots p_{i_{n-1} i_n}(t/n) > 0$$

for all $t > 0$, and (iv) holds. $\square$

Condition (iv) of Theorem 3.2.1 shows that the situation is simpler than in discrete-time, where it may be possible to reach a state, but only after a certain length of time, and then only periodically.

3.3 Hitting times and absorption probabilities

Let $(X_t)_{t\geq 0}$ be a Markov chain with generator matrix Q. The *hitting time* of a subset A of I is the random variable D^A defined by

$$D^A(\omega) = \inf\{t \geq 0 : X_t(\omega) \in A\}$$

with the usual convention that $\inf \emptyset = \infty$. We emphasise that $(X_t)_{t\geq 0}$ is minimal. So if H^A is the hitting time of A for the jump chain, then

$$\{H^A < \infty\} = \{D^A < \infty\}$$

and on this set we have

$$D^A = J_{H^A}.$$

The probability, starting from i, that $(X_t)_{t\geq 0}$ ever hits A is then

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

When A is a closed class, h_i^A is called the *absorption probability*. Since the hitting probabilities are those of the jump chain we can calculate them as in Section 1.3.

Theorem 3.3.1. *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*

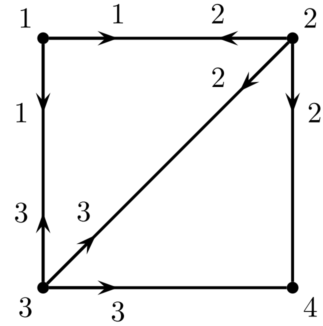
$$\begin{cases} h_i^A = 1 & \text{for } i \in A, \\ \sum_{j\in I} q_{ij} h_j^A = 0 & \text{for } i \notin A. \end{cases}$$

Proof. Apply Theorem 1.3.2 to the jump chain and rewrite (1.3) in terms of Q. □

The average time taken, starting from i, for $(X_t)_{t\geq 0}$ to reach A is given by

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

In calculating k_i^A we have to take account of the holding times so the relationship to the discrete-time case is not quite as simple.



Example 3.3.2

Consider the Markov chain $(X_t)_{t\geq 0}$ with the diagram given on the preceding page. How long on average does it take to get from 1 to 4?

Set $k_i = \mathbb{E}_i(\text{time to get to } 4)$. On starting in 1 we spend an average time $q_1^{-1} = 1/2$ in 1, then jump with equal probability to 2 or 3. Thus

$$k_1 = \tfrac{1}{2} + \tfrac{1}{2}k_2 + \tfrac{1}{2}k_3$$

and similarly

$$k_2 = \tfrac{1}{6} + \tfrac{1}{3}k_1 + \tfrac{1}{3}k_3, \quad k_3 = \tfrac{1}{9} + \tfrac{1}{3}k_1 + \tfrac{1}{3}k_2.$$

On solving these linear equations we find $k_1 = 17/12$.

Here is the general result. The proof follows the same lines as Theorem 1.3.5.

Theorem 3.3.3. *Assume that $q_i > 0$ for all $i \notin A$. The vector of expected hitting times $k^A = (k_i^A : i \in I)$ is the minimal non-negative solution to the system of linear equations*

$$\begin{cases} k_i^A = 0 & \text{for } i \in A \\ -\sum_{j\in I} q_{ij}k_j^A = 1 & \text{for } i \notin A. \end{cases} \tag{3.1}$$

Proof. First we show that k^A satisfies (3.1). If $X_0 = i \in A$, then $D^A = 0$, so $k_i^A = 0$. If $X_0 = i \notin A$, then $D^A \geq J_1$, so by the Markov property of the jump chain

$$\mathbb{E}_i(D^A - J_1 \mid Y_1 = j) = \mathbb{E}_j(D^A),$$

so

$$k_i^A = \mathbb{E}_i(D^A) = \mathbb{E}_i(J_1) + \sum_{j\neq i} \mathbb{E}(D^A - J_1 \mid Y_1 = j)\mathbb{P}_i(Y_1 = j) = q_i^{-1} + \sum_{j\neq i} \pi_{ij}k_j^A$$

and so

$$-\sum_{j\in I} q_{ij}k_j^A = 1.$$

Suppose now that $y = (y_i : i \in I)$ is another solution to (3.1). Then $k_i^A = y_i = 0$ for $i \in A$. Suppose $i \notin A$, then

$$\begin{aligned} y_i &= q_i^{-1} + \sum_{j\notin A} \pi_{ij}y_j = q_i^{-1} + \sum_{j\notin A} \pi_{ij}\left(q_j^{-1} + \sum_{k\notin A} \pi_{jk}y_k\right) \\ &= \mathbb{E}_i(S_1) + \mathbb{E}_i\big(S_2 1_{\{H^A\geq 2\}}\big) + \sum_{j\notin A}\sum_{k\notin A} \pi_{ij}\pi_{jk}y_k. \end{aligned}$$

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

$$y_i = \mathbb{E}_i(S_1) + \cdots + \mathbb{E}_i\big(S_n 1_{\{H^A \geq n\}}\big) + \sum_{j_1 \notin A} \cdots \sum_{j_n \notin A} \pi_{ij_1} \ldots \pi_{j_{n-1}j_n} y_{j_n}.$$

So, if y is non-negative

$$y_i \geq \sum_{m=1}^{n} \mathbb{E}_i\big(S_m 1_{H^A \geq m}\big) = \mathbb{E}_i\bigg(\sum_{m=1}^{H^A \wedge n} S_m\bigg)$$

where we use the notation $H^A \wedge n$ for the minimum of H^A and n. Now

$$\sum_{m=1}^{H^A} S_m = D_A$$

so, by monotone convergence, $y_i \geq \mathbb{E}_i(D_A) = k_i^A$, as required. $\square$

Exercise

3.3.1 Consider the Markov chain on $\{1, 2, 3, 4\}$ with generator matrix

$$Q = \begin{pmatrix} -1 & 1/2 & 1/2 & 0 \\ 1/4 & -1/2 & 0 & 1/4 \\ 1/6 & 0 & -1/3 & 1/6 \\ 0 & 0 & 0 & 0. \end{pmatrix}$$

Calculate (a) the probability of hitting 3 starting from 1, (b) the expected time to hit 4 starting from 1.

3.4 Recurrence and transience

Let $(X_t)_{t \geq 0}$ be Markov chain with generator matrix Q. Recall that we insist $(X_t)_{t \geq 0}$ be minimal. We say a state i is *recurrent* if

$$\mathbb{P}_i(\{t \geq 0 : X_t = i\} \text{ is unbounded}) = 1.$$

We say that i is *transient* if

$$\mathbb{P}_i(\{t \geq 0 : X_t = i\} \text{ is unbounded}) = 0.$$

Note that if $(X_t)_{t \geq 0}$ can explode starting from i then i is certainly not recurrent. The next result shows that, like class structure, recurrence and transience are determined by the jump chain.

Theorem 3.4.1. *We have:*

(i) *if i is recurrent for the jump chain $(Y_n)_{n\geq 0}$, then i is recurrent for $(X_t)_{t\geq 0}$;*

(ii) *if i is transient for the jump chain, then i is transient for $(X_t)_{t\geq 0}$;*

(iii) *every state is either recurrent or transient;*

(iv) *recurrence and transience are class properties.*

Proof. (i) Suppose i is recurrent for $(Y_n)_{n\geq 0}$. If $X_0 = i$ then $(X_t)_{t\geq 0}$ does not explode and $J_n \to \infty$ by Theorem 2.7.1. Also $X(J_n) = Y_n = i$ infinitely often, so $\{t \geq 0 : X_t = i\}$ is unbounded, with probability 1.
(ii) Suppose i is transient for $(Y_n)_{n\geq 0}$. If $X_0 = i$ then

$$N = \sup\{n \geq 0 : Y_n = i\} < \infty,$$

so $\{t \geq 0 : X_t = i\}$ is bounded by $J(N+1)$, which is finite, with probability 1, because $(Y_n : n \leq N)$ cannot include an absorbing state.
(iii) Apply Theorem 1.5.3 to the jump chain.
(iv) Apply Theorem 1.5.4 to the jump chain. □

The next result gives continuous-time analogues of the conditions for recurrence and transience found in Theorem 1.5.3. We denote by T_i the *first passage time* of $(X_t)_{t\geq 0}$ to state i, defined by

$$T_i(\omega) = \inf\{t \geq J_1(\omega) : X_t(\omega) = i\}.$$

Theorem 3.4.2. *The following dichotomy holds:*

(i) *if $q_i = 0$ or $\mathbb{P}_i(T_i < \infty) = 1$, then i is recurrent and $\int_0^\infty p_{ii}(t)dt = \infty$;*

(ii) *if $q_i > 0$ and $\mathbb{P}_i(T_i < \infty) < 1$, then i is transient and $\int_0^\infty p_{ii}(t)dt < \infty$.*

Proof. If $q_i = 0$, then $(X_t)_{t\geq 0}$ cannot leave i, so i is recurrent, $p_{ii}(t) = 1$ for all t, and $\int_0^\infty p_{ii}(t)dt = \infty$. Suppose then that $q_i > 0$. Let N_i denote the first passage time of the jump chain $(Y_n)_{n\geq 0}$ to state i. Then

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

so i is recurrent if and only if $\mathbb{P}_i(T_i < \infty) = 1$, by Theorem 3.4.1 and the corresponding result for the jump chain.

Write $\pi_{ij}^{(n)}$ for the (i,j) entry in Π^n. We shall show that

$$\int_0^\infty p_{ii}(t)dt = \frac{1}{q_i}\sum_{n=0}^\infty \pi_{ii}^{(n)} \tag{3.2}$$

so that i is recurrent if and only if $\int_0^\infty p_{ii}(t)dt = \infty$, by Theorem 3.4.1 and the corresponding result for the jump chain.

To establish (3.2) we use Fubini's theorem (see Section 6.4):

$$\begin{aligned}\int_0^\infty p_{ii}(t)dt &= \int_0^\infty \mathbb{E}_i(1_{\{X_t=i\}})dt = \mathbb{E}_i \int_0^\infty 1_{\{X_t=i\}}dt \\ &= \mathbb{E}_i \sum_{n=0}^\infty S_{n+1}1_{\{Y_n=i\}} \\ &= \sum_{n=0}^\infty \mathbb{E}_i(S_{n+1} \mid Y_n=i)\mathbb{P}_i(Y_n=i) = \frac{1}{q_i}\sum_{n=0}^\infty \pi_{ii}^{(n)}. \qquad \square\end{aligned}$$

Finally, we show that recurrence and transience are determined by any discrete-time sampling of $(X_t)_{t\geq 0}$.

Theorem 3.4.3. *Let $h>0$ be given and set $Z_n = X_{nh}$.*

(i) *If i is recurrent for $(X_t)_{t\geq 0}$ then i is recurrent for $(Z_n)_{n\geq 0}$.*

(ii) *If i is transient for $(X_t)_{t\geq 0}$ then i is transient for $(Z_n)_{n\geq 0}$.*

Proof. Claim (ii) is obvious. To prove (i) we use for $nh \leq t < (n+1)h$ the estimate

$$p_{ii}\big((n+1)h\big) \geq e^{-q_i h}p_{ii}(t)$$

which follows from the Markov property. Then, by monotone convergence

$$\int_0^\infty p_{ii}(t)dt \leq he^{q_i h}\sum_{n=1}^\infty p_{ii}(nh)$$

and the result follows by Theorems 1.5.3 and 3.4.2. $\square$

Exercise

3.4.1 Customers arrive at a certain queue in a Poisson process of rate λ. The single 'server' has two states A and B, state A signifying that he is 'in attendance' and state B that he is having a tea-break. Independently of how many customers are in the queue, he fluctuates between these states as a Markov chain Y on $\{A,B\}$ with Q-matrix

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

The total service time for any customer is exponentially distributed with parameter μ and is independent of the chain Y and of the service times of other customers.

Describe the system as a Markov chain X with state-space

$$\{A_0, A_1, A_2, \dots\} \cup \{B_0, B_1, B_2, \dots\},$$

A_n signifying that the server is in state A and there are n people in the queue (including anyone being served) and B_n signifying that the server is in state B and there are n people in the queue.

Explain why, for some θ in $(0,1]$, and $k = 0, 1, 2, \dots,$

$$P(X \text{ hits } A_0 | X_0 = A_k) = \theta^k .$$

Show that $(\theta - 1)f(\theta) = 0$, where

$$f(\theta) = \lambda^2\theta^2 - \lambda(\lambda + \mu + \alpha + \beta)\theta + (\lambda + \beta)\mu .$$

By considering $f(1)$ or otherwise, prove that X is transient if $\mu\beta < \lambda(\alpha+\beta)$, and explain why this is intuitively obvious.

3.5 Invariant distributions

Just as in the discrete-time theory, the notions of invariant distribution and measure play an important role in the study of continuous-time Markov chains. We say that λ is *invariant* if

$$\lambda Q = 0.$$

Theorem 3.5.1. *Let Q be a Q-matrix with jump matrix Π and let λ be a measure. The following are equivalent:*

(i) *λ is invariant;*
(ii) *$\mu\Pi = \mu$ where $\mu_i = \lambda_i q_i$.*

Proof. We have $q_i(\pi_{ij} - \delta_{ij}) = q_{ij}$ for all i, j, so

$$(\mu(\Pi - I))_j = \sum_{i \in I} \mu_i(\pi_{ij} - \delta_{ij}) = \sum_{i \in I} \lambda_i q_{ij} = (\lambda Q)_j. \qquad \square$$

This tie-up with measures invariant for the jump matrix means that we can use the existence and uniqueness results of Section 1.7 to obtain the following result.

Theorem 3.5.2. *Suppose that Q is irreducible and recurrent. Then Q has an invariant measure λ which is unique up to scalar multiples.*

Proof. Let us exclude the trivial case $I = \{i\}$; then irreducibility forces $q_i > 0$ for all i. By Theorems 3.2.1 and 3.4.1, Π is irreducible and recurrent. Then, by Theorems 1.7.5 and 1.7.6, Π has an invariant measure μ, which is unique up to scalar multiples. So, by Theorem 3.5.1, we can take $\lambda_i = \mu_i/q_i$ to obtain an invariant measure unique up to scalar multiples. □

Recall that a state i is recurrent if $q_i = 0$ or $\mathbb{P}_i(T_i < \infty) = 1$. If $q_i = 0$ or the *expected return time* $m_i = \mathbb{E}_i(T_i)$ is finite then we say i is *positive recurrent*. Otherwise a recurrent state i is called *null recurrent*. As in the discrete-time case positive recurrence is tied up with the existence of an invariant distribution.

Theorem 3.5.3. *Let Q be an irreducible Q-matrix. Then the following are equivalent:*

(i) *every state is positive recurrent;*
(ii) *some state i is positive recurrent;*
(iii) *Q is non-explosive and has an invariant distribution λ.*

Moreover, when (iii) holds we have $m_i = 1/(\lambda_i q_i)$ for all i.

Proof. Let us exclude the trivial case $I = \{i\}$; then irreducibility forces $q_i > 0$ for all i. It is obvious that (i) implies (ii). Define $\mu^i = (\mu_j^i : j \in I)$ by

$$\mu_j^i = \mathbb{E}_i \int_0^{T_i \wedge \zeta} 1_{\{X_s = j\}} ds,$$

where $T_i \wedge \zeta$ denotes the minimum of T_i and ζ. By monotone convergence,

$$\sum_{j \in I} \mu_j^i = \mathbb{E}_i(T_i \wedge \zeta).$$

Denote by N_i the first passage time of the jump chain to state i. By Fubini's theorem

$$\begin{aligned}
\mu_j^i &= \mathbb{E}_i \sum_{n=0}^{\infty} S_{n+1} 1_{\{Y_n = j, n < N_i\}} \\
&= \sum_{n=0}^{\infty} \mathbb{E}_i\big(S_{n+1} \mid Y_n = j\big) \mathbb{E}_i\big(1_{\{Y_n = j, n < N_i\}}\big) \\
&= q_j^{-1}\, \mathbb{E}_i \sum_{n=0}^{\infty} 1_{\{Y_n = j, n < N_i\}} \\
&= q_j^{-1}\, \mathbb{E}_i \sum_{n=0}^{N_i - 1} 1_{\{Y_n = j\}} = \gamma_j^i / q_j
\end{aligned}$$

where, in the notation of Section 1.7, γ_j^i is the expected time in j between visits to i for the jump chain.

Suppose (ii) holds, then i is certainly recurrent, so the jump chain is recurrent, and Q is non-explosive, by Theorem 2.7.1. We know that $\gamma^i \Pi = \gamma^i$ by Theorem 1.7.5, so $\mu^i Q = 0$ by Theorem 3.5.1. But μ^i has finite total mass

$$\sum_{j \in I} \mu_j^i = \mathbb{E}_i(T_i) = m_i$$

so we obtain an invariant distribution λ by setting $\lambda_j = \mu_j^i / m_i$.

On the other hand, suppose (iii) holds. Fix $i \in I$ and set $\nu_j = \lambda_j q_j / (\lambda_i q_i)$; then $\nu_i = 1$ and $\nu \Pi = \nu$ by Theorem 3.5.1, so $\nu_j \geq \gamma_j^i$ for all j by Theorem 1.7.6. So

$$\begin{aligned} m_i &= \sum_{j \in I} \mu_j^i = \sum_{j \in I} \gamma_j^i / q_j \leq \sum_{j \in I} \nu_j / q_j \\ &= \sum_{j \in I} \lambda_j / (\lambda_i q_i) = 1/(\lambda_i q_i) < \infty \end{aligned}$$

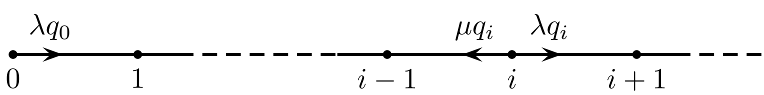
showing that i is positive recurrent.

To complete the proof we return to the preceding calculation armed with the knowledge that Q is recurrent, hence Π is recurrent, $\nu_j = \gamma_j^i$ and $m_i = 1/(\lambda_i q_i)$ for all i. $\square$

The following example is a caution that the existence of an invariant distribution for a continuous-time Markov chain is not enough to guarantee positive recurrence, or even recurrence.

Example 3.5.4

Consider the Markov chain $(X_t)_{t \geq 0}$ on $\mathbb{Z}^+$ with the following diagram, where $q_i > 0$ for all i and where $0 < \lambda = 1 - \mu < 1$:



The jump chain behaves as a simple random walk away from 0, so $(X_t)_{t \geq 0}$ is recurrent if $\lambda \leq \mu$ and transient if $\lambda > \mu$. To compute an invariant measure ν it is convenient to use the *detailed balance equations*

$$\nu_i q_{ij} = \nu_j q_{ji} \qquad \text{for all } i, j.$$

Look ahead to Lemma 3.7.2 to see that any solution is invariant. In this case the non-zero equations read

$$\nu_i \lambda q_i = \nu_{i+1} \mu q_{i+1} \qquad \text{for all } i.$$

So a solution is given by $\nu_i = q_i^{-1}(\lambda/\mu)^i$. If the jump rates q_i are constant then ν can be normalized to produce an invariant distribution precisely when $\lambda < \mu$.

Consider, on the other hand, the case where $q_i = 2^i$ for all i and $1 < \lambda/\mu < 2$. Then ν has finite total mass so $(X_t)_{t\geq 0}$ has an invariant distribution, but $(X_t)_{t\geq 0}$ is also transient. Given Theorem 3.5.3, the only possibility is that $(X_t)_{t\geq 0}$ is explosive.

The next result justifies calling measures λ with $\lambda Q = 0$ invariant.

Theorem 3.5.5. *Let Q be irreducible and recurrent, and let λ be a measure. Let $s > 0$ be given. The following are equivalent:*

(i) $\lambda Q = 0$;

(ii) $\lambda P(s) = \lambda$.

Proof. There is a very simple proof in the case of finite state-space: by the backward equation

$$\frac{d}{ds}\lambda P(s) = \lambda P'(s) = \lambda Q P(s)$$

so $\lambda Q = 0$ implies $\lambda P(s) = \lambda P(0) = \lambda$ for all s; $P(s)$ is also recurrent, so $\mu P(s) = \mu$ implies that μ is proportional to λ, so $\mu Q = 0$.

For infinite state-space, the interchange of differentiation with the summation involved in multiplication by λ is not justified and an entirely different proof is needed.

Since Q is recurrent, it is non-explosive by Theorem 2.7.1, and $P(s)$ is recurrent by Theorem 3.4.3. Hence any λ satisfying (i) or (ii) is unique up to scalar multiples; and from the proof of Theorem 3.5.3, if we fix i and set

$$\mu_j = \mathbb{E}_i \int_0^{T_i} 1_{\{X_t = j\}} dt,$$

then $\mu Q = 0$. Thus it suffices to show $\mu P(s) = \mu$. By the strong Markov property at T_i (which is a simple consequence of the strong Markov property of the jump chain)

$$\mathbb{E}_i \int_0^s 1_{\{X_t = j\}} dt = \mathbb{E}_i \int_{T_i}^{T_i + s} 1_{\{X_t = j\}} dt.$$

Hence, using Fubini's theorem,

$$\begin{aligned}\mu_j &= \mathbb{E}_i \int_s^{s+T_i} 1_{\{X_t=j\}} dt \\ &= \int_0^\infty \mathbb{P}_i(X_{s+t} = j, t < T_i) dt \\ &= \int_0^\infty \sum_{k\in I} \mathbb{P}_i(X_t = k, t < T_i) p_{kj}(s) dt \\ &= \sum_{k\in I} \left(\mathbb{E}_i \int_0^{T_i} 1_{\{X_t=k\}} dt \right) p_{kj}(s) \\ &= \sum_{k\in I} \mu_k p_{kj}(s)\end{aligned}$$

as required. □

Theorem 3.5.6. *Let Q be an irreducible non-explosive Q-matrix having an invariant distribution λ. If $(X_t)_{t\ge 0}$ is Markov(λ, Q) then so is $(X_{s+t})_{t\ge 0}$ for any $s \ge 0$.*

Proof. By Theorem 3.5.5, for all i,

$$\mathbb{P}(X_s = i) = (\lambda P(s))_i = \lambda_i$$

so, by the Markov property, conditional on $X_s = i$, $(X_{s+t})_{t\ge 0}$ is Markov(δ_i, Q). □

3.6 Convergence to equilibrium

We now investigate the limiting behaviour of $p_{ij}(t)$ as $t \to \infty$ and its relation to invariant distributions. You will see that the situation is analogous to the case of discrete-time, only there is no longer any possibility of periodicity.

We shall need the following estimate of uniform continuity for the transition probabilities.

Lemma 3.6.1. *Let Q be a Q-matrix with semigroup $P(t)$. Then, for all $t, h \ge 0$*

$$|p_{ij}(t+h) - p_{ij}(t)| \le 1 - e^{-q_i h}.$$

Proof. We have

$$\begin{aligned}|p_{ij}(t+h) - p_{ij}(t)| &= \left| \sum_{k\in I} p_{ik}(h) p_{kj}(t) - p_{ij}(t) \right| \\ &= \left| \sum_{k\ne i} p_{ik}(h) p_{kj}(t) - (1 - p_{ii}(h)) p_{ij}(t) \right| \\ &\le 1 - p_{ii}(h) \le \mathbb{P}_i(J_1 \le h) = 1 - e^{-q_i h}.\end{aligned}$$

□

Theorem 3.6.2 (Convergence to equilibrium). *Let Q be an irreducible non-explosive Q-matrix with semigroup $P(t)$, and having an invariant distribution λ. Then for all states i, j we have*

$$p_{ij}(t) \to \lambda_j \quad \text{as} \quad t \to \infty.$$

Proof. Let $(X_t)_{t\geq 0}$ be Markov(δ_i, Q). Fix $h > 0$ and consider the *h-skeleton* $Z_n = X_{nh}$. By Theorem 2.8.4

$$\mathbb{P}\big(Z_{n+1} = i_{n+1} \mid Z_0 = i_0, \dots, Z_n = i_n\big) = p_{i_n i_{n+1}}(h)$$

so $(Z_n)_{n\geq 0}$ is discrete-time Markov$(\delta_i, P(h))$. By Theorem 3.2.1 irreducibility implies $p_{ij}(h) > 0$ for all i, j so $P(h)$ is irreducible and aperiodic. By Theorem 3.5.5, λ is invariant for $P(h)$. So, by discrete-time convergence to equilibrium, for all i, j

$$p_{ij}(nh) \to \lambda_j \quad \text{as} \quad n \to \infty.$$

Thus we have a lattice of points along which the desired limit holds; we fill in the gaps using uniform continuity. Fix a state i. Given $\varepsilon > 0$ we can find $h > 0$ so that

$$1 - e^{-q_i s} \leq \varepsilon/2 \qquad \text{for } 0 \leq s \leq h$$

and then find N, so that

$$|p_{ij}(nh) - \lambda_j| \leq \varepsilon/2 \qquad \text{for } n \geq N.$$

For $t \geq Nh$ we have $nh \leq t < (n+1)h$ for some $n \geq N$ and

$$|p_{ij}(t) - \lambda_j| \leq |p_{ij}(t) - p_{ij}(nh)| + |p_{ij}(nh) - \lambda_j| \leq \varepsilon$$

by Lemma 3.6.1. Hence

$$p_{ij}(t) \to \lambda_j \quad \text{as} \quad n \to \infty. \qquad \square$$

The complete description of limiting behaviour for irreducible chains in continuous time is provided by the following result. It follows from Theorem 1.8.5 by the same argument we used in the preceding result. We do not give the details.

Theorem 3.6.3. *Let Q be an irreducible Q-matrix and let ν be any distribution. Suppose that $(X_t)_{t\geq 0}$ is Markov(ν, Q). Then*

$$\mathbb{P}(X_t = j) \to 1/(q_j m_j) \qquad \text{as } t \to \infty \quad \text{for all } j \in I$$

where m_j is the expected return time to state j.

Exercises

3.6.1 Find an invariant distribution λ for the Q-matrix

$$Q = \begin{pmatrix} -2 & 1 & 1 \\ 4 & -4 & 0 \\ 2 & 1 & -3 \end{pmatrix}$$

and verify that $\lim_{t\to\infty} p_{11}(t) = \lambda_1$ using your answer to Exercise 2.1.1.

3.6.2 In each of the following cases, compute $\lim_{t\to\infty} \mathbb{P}(X_t = 2 | X_0 = 1)$ for the Markov chain $(X_t)_{t\geq 0}$ with the given Q-matrix on $\{1, 2, 3, 4\}$:

$$\text{(a)} \begin{pmatrix} -2 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 1 & 0 & 0 & -1 \end{pmatrix} \qquad \text{(b)} \begin{pmatrix} -2 & 1 & 1 & 0 \\ 0 & -1 & 1 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

$$\text{(c)} \begin{pmatrix} -1 & 1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & -2 & 2 \\ 0 & 0 & 2 & -2 \end{pmatrix} \qquad \text{(d)} \begin{pmatrix} -2 & 1 & 0 & 1 \\ 0 & -2 & 2 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

3.6.3 Customers arrive at a single-server queue in a Poisson stream of rate λ. Each customer has a service requirement distributed as the sum of two independent exponential random variables of parameter μ. Service requirements are independent of one another and of the arrival process. Write down the generator matrix Q of a continuous-time Markov chain which models this, explaining what the states of the chain represent. Calculate the essentially unique invariant measure for Q, and deduce that the chain is positive recurrent if and only if $\lambda/\mu < 1/2$.

3.7 Time reversal

Time reversal of continuous-time chains has the same features found in the discrete-time case. Reversibility provides a powerful tool in the analysis of Markov chains, as we shall see in Section 5.2. Note in the following result how time reversal interchanges the roles of backward and forward equations. This echoes our proof of the forward equation, which rested on the time reversal identity of Lemma 2.8.5.

A small technical point arises in time reversal: right-continuous processes become left-continuous processes. For the processes we consider, this is unimportant. We could if we wished redefine the time-reversed process to equal its right limit at the jump times, thus obtaining again a right-continuous process. We shall suppose implicitly that this is done, and forget about the problem.

Theorem 3.7.1. *Let Q be irreducible and non-explosive and suppose that Q has an invariant distribution λ. Let $T \in (0,\infty)$ be given and let $(X_t)_{0\le t\le T}$ be Markov(λ, Q). Set $\widehat{X}_t = X_{T-t}$. Then the process $(\widehat{X}_t)_{0\le t\le T}$ is Markov$(\lambda, \widehat{Q})$, where $\widehat{Q} = (\widehat{q}_{ij} : i,j \in I)$ is given by $\lambda_j \widehat{q}_{ji} = \lambda_i q_{ij}$. Moreover, $\widehat{Q}$ is also irreducible and non-explosive with invariant distribution λ.*

Proof. By Theorem 2.8.6, the semigroup $\big(P(t) : t \ge 0\big)$ of Q is the minimal non-negative solution of the forward equation

$$P'(t) = P(t)Q, \quad P(0) = I.$$

Also, for all $t > 0$, $P(t)$ is an irreducible stochastic matrix with invariant distribution λ. Define $\widehat{P}(t)$ by

$$\lambda_j \widehat{p}_{ji}(t) = \lambda_i p_{ij}(t),$$

then $\widehat{P}(t)$ is an irreducible stochastic matrix with invariant distribution λ, and we can rewrite the forward equation transposed as

$$\widehat{P}'(t) = \widehat{Q}\widehat{P}(t).$$

But this is the backward equation for $\widehat{Q}$, which is itself a Q-matrix, and $\widehat{P}(t)$ is then its minimal non-negative solution. Hence $\widehat{Q}$ is irreducible and non-explosive and has invariant distribution λ.

Finally, for $0 = t_0 < \ldots < t_n = T$ and $s_k = t_k - t_{k-1}$, by Theorem 2.8.4 we have

$$\begin{aligned}
\mathbb{P}(\widehat{X}_{t_0} = i_0, \ldots, \widehat{X}_{t_n} = i_n) &= \mathbb{P}(X_{T-t_0} = i_0, \ldots, X_{T-t_n} = i_n) \\
&= \lambda_{i_n} p_{i_n i_{n-1}}(s_n) \ldots p_{i_1 i_0}(s_1) \\
&= \lambda_{i_0} \widehat{p}_{i_0 i_1}(s_1) \ldots \widehat{p}_{i_{n-1} i_n}(s_n)
\end{aligned}$$

so, by Theorem 2.8.4 again, $(\widehat{X}_t)_{0\le t\le T}$ is Markov$(\lambda, \widehat{Q})$. □

The chain $(\widehat{X}_t)_{0\le t\le T}$ is called the *time-reversal* of $(X_t)_{0\le t\le T}$.

A Q-matrix Q and a measure λ are said to be in *detailed balance* if

$$\lambda_i q_{ij} = \lambda_j q_{ji} \qquad \text{for all } i, j.$$

Lemma 3.7.2. *If Q and λ are in detailed balance then λ is invariant for Q.*

Proof. We have $(\lambda Q)_i = \sum_{j\in I} \lambda_j q_{ji} = \sum_{j\in I} \lambda_i q_{ij} = 0$. □

Let $(X_t)_{t\geq 0}$ be Markov(λ, Q), with Q irreducible and non-explosive. We say that $(X_t)_{t\geq 0}$ is *reversible* if, for all $T > 0$, $(X_{T-t})_{0\leq t\leq T}$ is also Markov(λ, Q).

Theorem 3.7.3. *Let Q be an irreducible and non-explosive Q-matrix and let λ be a distribution. Suppose that $(X_t)_{t\geq 0}$ is Markov(λ, Q). Then the following are equivalent:*

(a) $(X_t)_{t\geq 0}$ *is reversible;*

(b) Q *and* λ *are in detailed balance.*

Proof. Both (a) and (b) imply that λ is invariant for Q. Then both (a) and (b) are equivalent to the statement that $\widehat{Q} = Q$ in Theorem 3.7.1. □

Exercise

3.7.1 Consider a fleet of N buses. Each bus breaks down independently at rate μ, when it is sent to the depot for repair. The repair shop can only repair one bus at a time and each bus takes an exponential time of parameter λ to repair. Find the equilibrium distribution of the number of buses in service.

3.7.2 Calls arrive at a telephone exchange as a Poisson process of rate λ, and the lengths of calls are independent exponential random variables of parameter μ. Assuming that infinitely many telephone lines are available, set up a Markov chain model for this process.

Show that for large t the distribution of the number of lines in use at time t is approximately Poisson with mean λ/μ.

Find the mean length of the busy periods during which at least one line is in use.

Show that the expected number of lines in use at time t, given that n are in use at time 0, is $ne^{-\mu t} + \lambda(1 - e^{-\mu t})/\mu$.

Show that, in equilibrium, the number N_t of calls finishing in the time interval $[0, t]$ has Poisson distribution of mean λt.

Is $(N_t)_{t\geq 0}$ a Poisson process?

3.8 Ergodic theorem

Long-run averages for continuous-time chains display the same sort of behaviour as in the discrete-time case, and for similar reasons. Here is the result.

Theorem 3.8.1 (Ergodic theorem). *Let Q be irreducible and let ν be any distribution. If $(X_t)_{t\ge 0}$ is Markov(ν, Q), then*

$$\mathbb{P}\left(\frac{1}{t}\int_0^t 1_{\{X_s=i\}}ds \to \frac{1}{m_i q_i} \;\text{ as } t\to\infty\right) = 1$$

where $m_i = \mathbb{E}_i(T_i)$ is the expected return time to state i. Moreover, in the positive recurrent case, for any bounded function $f : I \to \mathbb{R}$ we have

$$\mathbb{P}\left(\frac{1}{t}\int_0^t f(X_s)ds \to \overline{f} \text{ as } t\to\infty\right) = 1$$

where

$$\overline{f} = \sum_{i\in I} \lambda_i f_i$$

and where $(\lambda_i : i \in I)$ is the unique invariant distribution.

Proof. If Q is transient then the total time spent in any state i is finite, so

$$\frac{1}{t}\int_0^t 1_{\{X_s=i\}}ds \le \frac{1}{t}\int_0^\infty 1_{\{X_s=i\}}ds \to 0 = \frac{1}{m_i}.$$

Suppose then that Q is recurrent and fix a state i. Then $(X_t)_{t\ge0}$ hits i with probability 1 and the long-run proportion of time in i equals the long-run proportion of time in i after first hitting i. So, by the strong Markov property (of the jump chain), it suffices to consider the case $\nu = \delta_i$.

Denote by M_i^n the length of the nth visit to i, by T_i^n the time of the nth return to i and by L_i^n the length of the nth excursion to i. Thus for $n = 0, 1, 2, \dots$, setting $T_i^0 = 0$, we have

$$\begin{aligned} M_i^{n+1} &= \inf\{t > T_i^n : X_t \ne i\} - T_i^n \\ T_i^{n+1} &= \inf\{t > T_i^n + M_i^{n+1} : X_t = i\} \\ L_i^{n+1} &= T_i^{n+1} - T_i^n. \end{aligned}$$

By the strong Markov property (of the jump chain) at the stopping times T_i^n for $n \ge 0$ we find that $L_i^1, L_i^2, \dots$ are independent and identically distributed with mean m_i, and that $M_i^1, M_i^2, \dots$ are independent and identically distributed with mean $1/q_i$. Hence, by the strong law of large numbers (see Theorem 1.10.1)

$$\begin{aligned} \frac{L_i^1 + \dots + L_i^n}{n} &\to m_i \quad \text{as } n\to\infty \\ \frac{M_i^1 + \dots + M_i^n}{n} &\to \frac{1}{q_i} \quad \text{as } n\to\infty \end{aligned}$$

and hence

$$\frac{M_i^1 + \cdots + M_i^n}{L_i^1 + \cdots + L_i^n} \to \frac{1}{m_i q_i} \quad \text{as } n \to \infty$$

with probability 1. In particular, we note that $T_i^n / T_i^{n+1} \to 1$ as $n \to \infty$ with probability 1. Now, for $T_i^n \le t < T_i^{n+1}$ we have

$$\frac{T_i^n}{T_i^{n+1}} \frac{M_i^1 + \cdots + M_i^n}{L_i^1 + \cdots + L_i^n} \le \frac{1}{t} \int_0^t 1_{\{X_s = i\}} ds \le \frac{T_i^{n+1}}{T_i^n} \frac{M_i^1 + \cdots + M_i^{n+1}}{L_i^1 + \cdots + L_i^{n+1}}$$

so on letting $t \to \infty$ we have, with probability 1

$$\frac{1}{t} \int_0^t 1_{\{X_s = i\}} ds \to \frac{1}{m_i q_i}.$$

In the positive recurrent case we can write

$$\frac{1}{t} \int_0^t f(X_s) ds - \overline{f} = \sum_{i \in I} f_i \left(\frac{1}{t} \int_0^t 1_{\{X_s = i\}} ds - \lambda_i \right)$$

where $\lambda_i = 1/(m_i q_i)$. We conclude that

$$\frac{1}{t} \int_0^t f(X_s) ds \to \overline{f} \quad \text{as } t \to \infty$$

with probability 1, by the same argument as was used in the proof of Theorem 1.10.2. □

Index