6

Markov chains

Summary. A Markov chain is a random process with the property that, conditional on its present value, the future is independent of the past. The Chapman–Kolmogorov equations are derived, and used to explore the persistence and transience of states. Stationary distributions are studied at length, and the ergodic theorem for irreducible chains is proved using coupling. The reversibility of Markov chains is discussed. After a section devoted to branching processes, the theory of Poisson processes and birth–death processes is considered in depth, and the theory of continuous-time chains is sketched. The technique of imbedding a discrete-time chain inside a continuous-time chain is exploited in different settings. The basic properties of spatial Poisson processes are described, and the chapter ends with an account of the technique of Markov chain Monte Carlo.

6.1 Markov processes

The simple random walk (5.3) and the branching process (5.4) are two examples of sequences of random variables that evolve in some random but prescribed manner. Such collections are called† 'random processes'. A typical random process X is a family $\{X_t : t \in T\}$ of random variables indexed by some set T. In the above examples $T = \{0, 1, 2, \dots\}$ and we call the process a 'discrete-time' process; in other important examples $T = \mathbb{R}$ or $T = [0, \infty)$ and we call it a 'continuous-time' process. In either case we think of a random process as a family of variables that evolve as time passes. These variables may even be independent of each other, but then the evolution is not very surprising and this very special case is of little interest to us in this chapter. Rather, we are concerned with more general, and we hope realistic, models for random evolution. Simple random walks and branching processes shared the following property: conditional on their values at the nth step, their future values did not depend on their previous values. This property proved to be very useful in their analysis, and it is to the general theory of processes with this property that we turn our attention now.

Until further notice we shall be interested in discrete-time processes. Let $\{X_0, X_1, \dots\}$ be a sequence of random variables which take values in some countable set S, called the *state*

†Such collections are often called 'stochastic' processes. The Greek verb 'στοχάζομαι' means 'to shoot at, aim at, guess at', and the adjective 'στοχαστικός' was used, for example by Plato, to mean 'proceeding by guesswork'.

space†. Each X_n is a discrete random variable that takes one of N possible values, where $N = |S|$; it may be the case that $N = \infty$.

(1) Definition. The process X is a **Markov chain**‡ if it satisfies the **Markov condition**:

$$\mathbb{P}\big(X_n = s \,\big|\, X_0 = x_0,\ X_1 = x_1, \dots,\ X_{n-1} = x_{n-1}\big) = \mathbb{P}(X_n = s \mid X_{n-1} = x_{n-1})$$

for all $n \ge 1$ and all $s, x_1, \dots, x_{n-1} \in S$.

A proof that the random walk is a Markov chain was given in Lemma (3.9.5). The reader can check that the Markov property is equivalent to each of the stipulations (2) and (3) below: for each $s \in S$ and for every sequence $\{x_i : i \ge 0\}$ in S,

(2) $$\mathbb{P}\big(X_{n+1} = s \,\big|\, X_{n_1} = x_{n_1},\ X_{n_2} = x_{n_2},\ X_{n_k} = x_{n_k}\big) = \mathbb{P}\big(X_{n+1} = s \,\big|\, X_{n_k} = x_{n_k}\big)$$ for all $n_1 < n_2 < \cdots < n_k \le n$,

(3) $$\mathbb{P}\big(X_{m+n} = s \,\big|\, X_0 = x_0,\ X_1 = x_1, \dots,\ X_m = x_m\big) = \mathbb{P}\big(X_{m+n} = s \,\big|\, X_m = x_m\big)$$ for any $m, n \ge 0$.

We have assumed that X takes values in some *countable set* S. The reason for this is essentially the same as the reason for treating discrete and continuous variables separately. Since S is assumed countable, it can be put in one–one correspondence with some subset S' of the integers, and without loss of generality we can assume that S *is* this set S' of integers. If $X_n = i$, then we say that the chain is in the 'ith state at the nth step'; we can also talk of the chain as 'having the value i', 'visiting i', or 'being in state i', depending upon the context of the remark.

The evolution of a chain is described by its 'transition probabilities' $\mathbb{P}(X_{n+1} = j \mid X_n = i)$; it can be quite complicated in general since these probabilities depend upon the three quantities n, i, and j. We shall restrict our attention to the case when they do not depend on n but only upon i and j.

(4) Definition. The chain X is called **homogeneous** if

$$\mathbb{P}(X_{n+1} = j \mid X_n = i) = \mathbb{P}(X_1 = j \mid X_0 = i)$$

for all n, i, j. The **transition matrix** $\mathbf{P} = (p_{ij})$ is the $|S| \times |S|$ matrix of **transition probabilities**

$$p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i).$$

Some authors write p_{ji} in place of p_{ij} here, so beware; sometimes we write $p_{i,j}$ for p_{ij}. Henceforth, *all Markov chains are assumed homogeneous* unless otherwise specified; we assume that the process X is a Markov chain, and we denote the transition matrix of such a chain by $\mathbf{P}$.

---

†There is, of course, an underlying probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and each X_n is an $\mathcal{F}$-measurable function which maps Ω into S.

‡The expression 'stochastically determined process' was in use until around 1930, when Khinchin suggested this more functional label.

(5) Theorem. *The transition matrix* $\mathbf{P}$ *is a* stochastic matrix, *which is to say that*:

(a) $\mathbf{P}$ *has non-negative entries, or* $p_{ij} \geq 0$ *for all* i, j,

(b) $\mathbf{P}$ *has row sums equal to one, or* $\sum_j p_{ij} = 1$ *for all* i.

Proof. An easy *exercise*. ■

We can easily see that (5) characterizes transition matrices.

Broadly speaking, we are interested in the evolution of X over two different time scales, the 'short term' and the 'long term'. In the short term the random evolution of X is described by $\mathbf{P}$, whilst long-term changes are described in the following way.

(6) Definition. The **n-step transition matrix** $\mathbf{P}(m, m+n) = (p_{ij}(m, m+n))$ is the matrix of **n-step transition probabilities** $p_{ij}(m, m+n) = \mathbb{P}(X_{m+n} = j \mid X_m = i)$.

By the assumption of homogeneity, $\mathbf{P}(m, m+1) = \mathbf{P}$. That $\mathbf{P}(m, m+n)$ does not depend on m is a consequence of the following important fact.

(7) Theorem. Chapman–Kolmogorov equations.

$$p_{ij}(m, m+n+r) = \sum_k p_{ik}(m, m+n) p_{kj}(m+n, m+n+r).$$

Therefore, $\mathbf{P}(m, m+n+r) = \mathbf{P}(m, m+n)\mathbf{P}(m+n, m+n+r)$, *and* $\mathbf{P}(m, m+n) = \mathbf{P}^n$, *the nth power of* $\mathbf{P}$.

Proof. We have as required that

$$\begin{aligned}
p_{ij}(m, m+n+r) &= \mathbb{P}(X_{m+n+r} = j \mid X_m = i) \\
&= \sum_k \mathbb{P}(X_{m+n+r} = j,\ X_{m+n} = k \mid X_m = i) \\
&= \sum_k \mathbb{P}(X_{m+n+r} = j \mid X_{m+n} = k,\ X_m = i)\mathbb{P}(X_{m+n} = k \mid X_m = i) \\
&= \sum_k \mathbb{P}(X_{m+n+r} = j \mid X_{m+n} = k)\mathbb{P}(X_{m+n} = k \mid X_m = i),
\end{aligned}$$

where we have used the fact that $\mathbb{P}(A \cap B \mid C) = \mathbb{P}(A \mid B \cap C)\mathbb{P}(B \mid C)$, proved in Exercise (1.4.2), together with the Markov property (2). The established equation may be written in matrix form as $\mathbf{P}(m, m+n+r) = \mathbf{P}(m, m+n)\mathbf{P}(m+n, m+n+r)$, and it follows by iteration that $\mathbf{P}(m, m+n) = \mathbf{P}^n$. ■

It is a consequence of Theorem (7) that $\mathbf{P}(m, m+n) = \mathbf{P}(0, n)$, and we write henceforth $\mathbf{P}_n$ for $\mathbf{P}(m, m+n)$, and $p_{ij}(n)$ for $p_{ij}(m, m+n)$. This theorem relates long-term development to short-term development, and tells us how X_n depends on the initial variable X_0. Let $\mu_i^{(n)} = \mathbb{P}(X_n = i)$ be the mass function of X_n, and write $\boldsymbol{\mu}^{(n)}$ for the row vector with entries $(\mu_i^{(n)} : i \in S)$.

(8) Lemma. $\boldsymbol{\mu}^{(m+n)} = \boldsymbol{\mu}^{(m)}\mathbf{P}_n$, *and hence* $\boldsymbol{\mu}^{(n)} = \boldsymbol{\mu}^{(0)}\mathbf{P}^n$.

Proof. We have that

$$\mu_j^{(m+n)} = \mathbb{P}(X_{m+n} = j) = \sum_i \mathbb{P}(X_{m+n} = j \mid X_m = i)\mathbb{P}(X_m = i)$$
$$= \sum_i \mu_i^{(m)} p_{ij}(n) = (\boldsymbol{\mu}^{(m)}\mathbf{P}_n)_j$$

and the result follows from Theorem (7). ■

Thus we reach the important conclusion that the random evolution of the chain is determined by the transition matrix $\mathbf{P}$ and the initial mass function $\boldsymbol{\mu}^{(0)}$. Many questions about the chain can be expressed in terms of these quantities, and the study of the chain is thus largely reducible to the study of algebraic properties of matrices.

(9) Example. Simple random walk. The simple random walk on the integers has state space $S = \{0, \pm 1, \pm 2, \dots\}$ and transition probabilities

$$p_{ij} = \begin{cases} p & \text{if } j = i + 1, \\ q = 1 - p & \text{if } j = i - 1, \\ 0 & \text{otherwise.} \end{cases}$$

The argument leading to equation (3.10.2) shows that

$$p_{ij}(n) = \begin{cases} \dbinom{n}{\frac{1}{2}(n + j - i)} p^{\frac{1}{2}(n+j-i)} q^{\frac{1}{2}(n-j+i)} & \text{if } n + j - i \text{ is even,} \\ 0 & \text{otherwise.} \end{cases}$$ ●

(10) Example. Branching process. As in Section 5.4, $S = \{0, 1, 2, \dots\}$ and p_{ij} is the coefficient of s^j in $G(s)^i$. Also, $p_{ij}(n)$ is the coefficient of s^j in $G_n(s)^i$. ●

(11) Example. Gene frequencies. One of the most interesting and extensive applications of probability theory is to genetics, and particularly to the study of gene frequencies. The problem may be inadequately and superficially described as follows. For definiteness suppose the population is human. Genetic information is (mostly) contained in chromosomes, which are strands of chemicals grouped in cell nuclei. In humans ordinary cells carry 46 chromosomes, 44 of which are homologous pairs. For our purposes a chromosome can be regarded as an ordered set of n sites, the states of which can be thought of as a sequence of random variables $C_1, C_2, \dots, C_n$. The possible values of each C_i are certain combinations of chemicals, and these values influence (or determine) some characteristic of the owner such as hair colour or leg length.

Now, suppose that A is a possible value of C_1, say, and let X_n be the number of individuals in the nth generation for which C_1 has the value A. What is the behaviour of the sequence $X_1, X_2, \dots, X_n, \dots$? The first important (and obvious) point is that the sequence is random, because of the following factors.

(a) The value A for C_1 may affect the owner's chances of contributing to the next generation. If A gives you short legs, you stand a better chance of being caught by a sabre-toothed tiger. The breeding population is randomly selected from those born, but there may be bias for or against the gene A.

(b) The breeding population is randomly combined into pairs to produce offspring. Each parent contributes 23 chromosomes to its offspring, but here again, if A gives you short legs you may have a smaller (or larger) chance of catching a mate.

(c) Sex cells having half the normal complement of chromosomes are produced by a special and complicated process called 'meiosis'. We shall not go into details, but essentially the homologous pairs of the parent are shuffled to produce new and different chromosomes for offspring. The sex cells from each parent (with 23 chromosomes) are then combined to give a new cell (with 46 chromosomes).

(d) Since meiosis involves a large number of complex chemical operations it is hardly surprising that things go wrong occasionally, producing a new value for C_1, $\widehat{A}$ say. This is a 'mutation'.

The reader can now see that if generations are segregated (in a laboratory, say), then we can suppose that $X_1, X_2, \dots$ is a Markov chain with a finite state space. If generations are not segregated and $X(t)$ is the frequency of A in the population at time t, then $X(t)$ may be a continuous-time Markov chain.

For a simple example, suppose that the population size is N, a constant. If $X_n = i$, it may seem reasonable that any member of the $(n+1)$th generation carries A with probability i/N, independently of the others. Then

$$p_{ij} = \mathbb{P}(X_{n+1} = j \mid X_n = i) = \binom{N}{j}\left(\frac{i}{N}\right)^j\left(1-\frac{i}{N}\right)^{N-j}.$$

Even more simply, suppose that at each stage exactly one individual dies and is replaced by a new individual; each individual is picked for death with probability $1/N$. If $X_n = i$, we assume that the probability that the replacement carries A is i/N. Then

$$p_{ij} = \begin{cases} \dfrac{i(N-i)}{N^2} & \text{if } j = i \pm 1, \\ 1 - 2\dfrac{i(N-i)}{N^2} & \text{if } j = i, \\ 0 & \text{otherwise.} \end{cases}$$ ●

(12) Example. Recurrent events. Suppose that X is a Markov chain on S, with $X_0 = i$. Let $T(1)$ be the time of the first return of the chain to i: that is, $T(1) = \min\{n \geq 1 : X_n = i\}$, with the convention that $T(1) = \infty$ if $X_n \neq i$ for all $n \geq 1$. Suppose that you tell me that $T(1) = 3$, say, which is to say that $X_n \neq i$ for $n = 1, 2$, and $X_3 = i$. The future evolution of the chain $\{X_3, X_4, \dots\}$ depends, by the Markov property, only on the fact that the new starting point X_3 equals i, and does not depend further on the values of X_0, X_1, X_2. Thus the future process $\{X_3, X_4, \dots\}$ has the same distribution as had the original process $\{X_0, X_1, \dots\}$ starting from state i. The same argument is valid for any given value of $T(1)$, and we are therefore led to the following observation. Having returned to its starting point for the first time, the future of the chain has the same distribution as had the original chain. Let $T(2)$ be the time which elapses between the first and second return of the chain to its starting point. Then $T(1)$ and $T(2)$ must be independent and identically distributed random variables. Arguing similarly for future returns, we deduce that the time of the nth return of the chain to its starting point may be represented as $T(1) + T(2) + \cdots + T(n)$, where $T(1), T(2), \dots$ are independent identically distributed random variables. That is to say, the return times of the chain form a 'recurrent-event process'; see Example (5.2.15). Some care is needed in order to make this argument fully rigorous, and this is the challenge of Exercise (5).

A problem arises with the above argument if $T(1)$ takes the value ∞ with strictly positive probability, which is to say that the chain is not (almost) certain to return to its starting point. For the moment we overlook this difficulty, and suppose not only that $\mathbb{P}(T(1) < \infty) = 1$, but also that $\mu = \mathbb{E}(T(1))$ satisfies $\mu < \infty$. It is now an immediate consequence of the renewal theorem (5.2.24) that

$$p_{ii}(n) = \mathbb{P}(X_n = i \mid X_0 = i) \to \frac{1}{\mu} \quad \text{as } n \to \infty$$

so long as the distribution of $T(1)$ is non-arithmetic; the latter condition is certainly satisfied if, say, $p_{ii} > 0$. ●

(13) Example. Bernoulli process. Let $S = \{0, 1, 2, \dots\}$ and define the Markov chain Y by $Y_0 = 0$ and

$$\mathbb{P}(Y_{n+1} = s + 1 \mid Y_n = s) = p, \quad \mathbb{P}(Y_{n+1} = s \mid Y_n = s) = 1 - p,$$

for all $n \ge 0$, where $0 < p < 1$. You may think of Y_n as the number of heads thrown in n tosses of a coin. It is easy to see that

$$\mathbb{P}(Y_{m+n} = j \mid Y_m = i) = \binom{n}{j-i} p^{j-i}(1-p)^{n-j+i}, \quad 0 \le j - i \le n.$$

Viewed as a Markov chain, Y is not a very interesting process. Suppose, however, that the value of Y_n is counted using a conventional digital decimal meter, and let X_n be the final digit of the reading, $X_n = Y_n$ modulo 10. It may be checked that $X = \{X_n : n \ge 0\}$ is a Markov chain on the state space $S' = \{0, 1, 2, \dots, 9\}$ with transition matrix

$$\mathbf{P} = \begin{pmatrix} 1-p & p & 0 & \cdots & 0 \\ 0 & 1-p & p & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ p & 0 & 0 & \cdots & 1-p \end{pmatrix}.$$

There are various ways of studying the behaviour of X. If we are prepared to use the renewal theorem (5.2.24), then we might argue as follows. The process X passes through the values $0, 1, 2, \dots, 9, 0, 1, \dots$ sequentially. Consider the times at which X takes the value i, say. These times form a recurrent-event process for which a typical inter-occurrence time T satisfies

$$T = \begin{cases} 1 & \text{with probability } 1-p, \\ 1 + Z & \text{with probability } p, \end{cases}$$

where Z has the negative binomial distribution with parameters 9 and p. Therefore $\mathbb{E}(T) = 1 + p\mathbb{E}(Z) = 1 + p(9/p) = 10$. It is now an immediate consequence of the renewal theorem that $\mathbb{P}(X_n = i) \to \frac{1}{10}$ for $i = 0, 1, \dots, 9$, as $n \to \infty$. ●

(14) Example. Markov's other chain (1910). Let $Y_1, Y_3, Y_5, \dots$ be a sequence of independent identically distributed random variables such that

(15) $$\mathbb{P}(Y_{2k+1} = -1) = \mathbb{P}(Y_{2k+1} = 1) = \tfrac{1}{2}, \quad k = 0, 1, 2, \dots,$$

and define $Y_{2k} = Y_{2k-1}Y_{2k+1}$, for $k = 1, 2, \ldots$. You may check that $Y_2, Y_4, \ldots$ is a sequence of independent identically distributed variables with the same distribution (15). Now $\mathbb{E}(Y_{2k}Y_{2k+1}) = \mathbb{E}(Y_{2k-1}Y_{2k+1}^2) = \mathbb{E}(Y_{2k-1}) = 0$, and so (by the result of Problem (3.11.12)) the sequence $Y_1, Y_2, \ldots$ is pairwise independent. Hence $p_{ij}(n) = \mathbb{P}(Y_{m+n} = j \mid Y_m = i)$ satisfies $p_{ij}(n) = \frac{1}{2}$ for all n and $i, j = \pm 1$, and it follows easily that the Chapman–Kolmogorov equations are satisfied.

Is Y a Markov chain? *No*, because $\mathbb{P}(Y_{2k+1} = 1 \mid Y_{2k} = -1) = \frac{1}{2}$, whereas

$$\mathbb{P}(Y_{2k+1} = 1 \mid Y_{2k} = -1,\ Y_{2k-1} = 1) = 0.$$

Thus, whilst the Chapman–Kolmogorov equations are *necessary* for the Markov property, they are not *sufficient*; this is for much the same reason that pairwise independence is weaker than independence.

Although Y is not a Markov chain, we can find a Markov chain by enlarging the state space. Let $Z_n = (Y_n, Y_{n+1})$, taking values in $S = \{-1, +1\}^2$. It is an *exercise* to check that Z is a (non-homogeneous) Markov chain with, for example,

$$\mathbb{P}\big(Z_{n+1} = (1,1) \,\big|\, Z_n = (1,1)\big) = \begin{cases} \frac{1}{2} & \text{if } n \text{ even}, \\ 1 & \text{if } n \text{ odd}. \end{cases}$$

This technique of 'imbedding' Y in a Markov chain on a larger state space turns out to be useful in many contexts of interest. ●

Exercises for Section 6.1

1. Show that any sequence of independent random variables taking values in the countable set S is a Markov chain. Under what condition is this chain homogeneous?

2. A die is rolled repeatedly. Which of the following are Markov chains? For those that are, supply the transition matrix.

(a) The largest number X_n shown up to the nth roll.

(b) The number N_n of sixes in n rolls.

(c) At time r, the time C_r since the most recent six.

(d) At time r, the time B_r until the next six.

3. Let $\{S_n : n \geq 0\}$ be a simple random walk with $S_0 = 0$, and show that $X_n = |S_n|$ defines a Markov chain; find the transition probabilities of this chain. Let $M_n = \max\{S_k : 0 \leq k \leq n\}$, and show that $Y_n = M_n - S_n$ defines a Markov chain. What happens if $S_0 \neq 0$?

4. Let X be a Markov chain and let $\{n_r : r \geq 0\}$ be an unbounded increasing sequence of positive integers. Show that $Y_r = X_{n_r}$ constitutes a (possibly inhomogeneous) Markov chain. Find the transition matrix of Y when $n_r = 2r$ and X is: (a) simple random walk, and (b) a branching process.

5. Let X be a Markov chain on S, and let $I : S^n \to \{0, 1\}$. Show that the distribution of $X_n, X_{n+1}, \ldots$, conditional on $\{I(X_1, \ldots, X_n) = 1\} \cap \{X_n = i\}$, is identical to the distribution of $X_n, X_{n+1}, \ldots$ conditional on $\{X_n = i\}$.

6. Strong Markov property. Let X be a Markov chain on S, and let T be a random variable taking values in $\{0, 1, 2, \ldots\}$ with the property that the indicator function $I_{\{T=n\}}$, of the event that $T = n$, is a function of the variables $X_1, X_2, \ldots, X_n$. Such a random variable T is called a *stopping time*, and the above definition requires that it is decidable whether or not $T = n$ with a knowledge only of the past and present, $X_0, X_1, \ldots, X_n$, and with no further information about the future.

Show that

$$\mathbb{P}\big(X_{T+m} = j \,\big|\, X_k = x_k \text{ for } 0 \leq k < T,\ X_T = i\big) = \mathbb{P}(X_{T+m} = j \mid X_T = i)$$

for $m \geq 0$, $i, j \in S$, and all sequences (x_k) of states.

7. Let X be a Markov chain with state space S, and suppose that $h : S \to T$ is one–one. Show that $Y_n = h(X_n)$ defines a Markov chain on T. Must this be so if h is not one–one?

8. Let X and Y be Markov chains on the set $\mathbb{Z}$ of integers. Is the sequence $Z_n = X_n + Y_n$ necessarily a Markov chain?

9. Let X be a Markov chain. Which of the following are Markov chains?
(a) X_{m+r} for $r \geq 0$.
(b) X_{2m} for $m \geq 0$.
(c) The sequence of pairs (X_n, X_{n+1}) for $n \geq 0$.

10. Let X be a Markov chain. Show that, for $1 < r < n$,

$$\mathbb{P}(X_r = k \mid X_i = x_i \text{ for } i = 1, 2, \dots, r-1, r+1 \dots, n) \\ = \mathbb{P}(X_r = k \mid X_{r-1} = x_{r-1}, X_{r+1} = x_{r+1}).$$

11. Let $\{X_n : n \geq 1\}$ be independent identically distributed integer-valued random variables. Let $S_n = \sum_{r=1}^n X_r$, with $S_0 = 0$, $Y_n = X_n + X_{n-1}$ with $X_0 = 0$, and $Z_n = \sum_{r=0}^n S_r$. Which of the following constitute Markov chains: (a) S_n, (b) Y_n, (c) Z_n, (d) the sequence of pairs (S_n, Z_n)?

12. A stochastic matrix $\mathbf{P}$ is called *doubly stochastic* if $\sum_i p_{ij} = 1$ for all j. It is called *sub-stochastic* if $\sum_i p_{ij} \leq 1$ for all j. Show that, if $\mathbf{P}$ is stochastic (respectively, doubly stochastic, sub-stochastic), then $\mathbf{P}^n$ is stochastic (respectively, doubly stochastic, sub-stochastic) for all n.

6.2 Classification of states

We can think of the development of the chain as the motion of a notional particle which jumps between the states of the state space S at each epoch of time. As in Section 5.3, we may be interested in the (possibly infinite) time which elapses before the particle returns to its starting point. We saw there that it sufficed to find the distribution of the length of time until the particle returns for the first time, since other interarrival times are merely independent copies of this. However, need the particle ever return to its starting point? With this question in mind we make the following definition.

(1) Definition. State i is called **persistent** (or **recurrent**) if

$$\mathbb{P}(X_n = i \text{ for some } n \geq 1 \mid X_0 = i) = 1,$$

which is to say that the probability of eventual return to i, having started from i, is 1. If this probability is strictly less than 1, the state i is called **transient**.

As in Section 5.3, we are interested in the *first passage times* of the chain. Let

$$f_{ij}(n) = \mathbb{P}\big(X_1 \neq j,\ X_2 \neq j, \dots,\ X_{n-1} \neq j, X_n = j \,\big|\, X_0 = i\big)$$

be the probability that the first visit to state j, starting from i, takes place at the nth step. Define

(2)
$$f_{ij} = \sum_{n=1}^{\infty} f_{ij}(n)$$

to be the probability that the chain ever visits j, starting from i. Of course, j is persistent if and only if $f_{jj} = 1$. We seek a criterion for persistence in terms of the n-step transition probabilities. Following our random walk experience, we define the generating functions

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

with the conventions that $p_{ij}(0) = \delta_{ij}$, the Kronecker delta, and $f_{ij}(0) = 0$ for all i and j. Clearly $f_{ij} = F_{ij}(1)$. We usually assume that $|s| < 1$, since $P_{ij}(s)$ is then guaranteed to converge. On occasions when we require properties of $P_{ij}(s)$ as $s \uparrow 1$, we shall appeal to Abel's theorem (5.1.15).

(3) Theorem.
(a) $P_{ii}(s) = 1 + F_{ii}(s)P_{ii}(s)$.
(b) $P_{ij}(s) = F_{ij}(s)P_{jj}(s)$ *if* $i \neq j$.

Proof. The proof is exactly as that of Theorem (5.3.1). Fix $i, j \in S$ and let $A_m = \{X_m = j\}$ and B_m be the event that the first visit to j (after time 0) takes place at time m; that is, $B_m = \{X_r \neq j \text{ for } 1 \le r < m,\ X_m = j\}$. The B_m are disjoint, so that

$$\mathbb{P}(A_m \mid X_0 = i) = \sum_{r=1}^{m} \mathbb{P}(A_m \cap B_r \mid X_0 = i).$$

Now, using the Markov condition (as found in Exercises (6.1.5) or (6.1.6)),

$$\begin{aligned}\mathbb{P}(A_m \cap B_r \mid X_0 = i) &= \mathbb{P}(A_m \mid B_r,\ X_0 = i)\mathbb{P}(B_r \mid X_0 = i)\\ &= \mathbb{P}(A_m \mid X_r = j)\mathbb{P}(B_r \mid X_0 = i).\end{aligned}$$

Hence

$$p_{ij}(m) = \sum_{r=1}^{m} f_{ij}(r)p_{jj}(m-r), \qquad m = 1, 2, \dots .$$

Multiply throughout by s^m, where $|s| < 1$, and sum over m (≥ 1) to find that $P_{ij}(s) - \delta_{ij} = F_{ij}(s)P_{jj}(s)$ as required. ■

(4) Corollary.
(a) *State j is persistent if $\sum_n p_{jj}(n) = \infty$, and if this holds then $\sum_n p_{ij}(n) = \infty$ for all i such that $f_{ij} > 0$.*
(b) *State j is transient if $\sum_n p_{jj}(n) < \infty$, and if this holds then $\sum_n p_{ij}(n) < \infty$ for all i.*

Proof. First we show that j is persistent if and only if $\sum_n p_{jj}(n) = \infty$. From (3a),

$$P_{jj}(s) = \frac{1}{1 - F_{jj}(s)} \quad \text{if } |s| < 1.$$

Hence, as $s \uparrow 1$, $P_{jj}(s) \to \infty$ if and only if $f_{jj} = F_{jj}(1) = 1$. Now use Abel's theorem (5.1.15) to obtain $\lim_{s\uparrow 1} P_{jj}(s) = \sum_n p_{jj}(n)$ and our claim is shown. Use (3b) to complete the proof. ■

(5) Corollary. *If j is transient then $p_{ij}(n) \to 0$ as $n \to \infty$ for all i.*

Proof. This is immediate from (4). ■

An important application of Theorem (4) is to the persistence of symmetric random walk; see Problem (5.12.5).

Thus each state is either persistent or transient. It is intuitively clear that the number $N(i)$ of times which the chain visits its starting point i satisfies

$$\text{(6)} \qquad \mathbb{P}\big(N(i) = \infty\big) = \begin{cases} 1 & \text{if } i \text{ is persistent,} \\ 0 & \text{if } i \text{ is transient,} \end{cases}$$

since after each such visit, subsequent return is assured if and only if $f_{ii} = 1$ (see Problem (6.15.5) for a more detailed argument).

Here is another important classification of states. Let

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

be the time of the first visit to j, with the convention that $T_j = \infty$ if this visit never occurs; $\mathbb{P}(T_i = \infty \mid X_0 = i) > 0$ if and only if i is transient, and in this case $\mathbb{E}(T_i \mid X_0 = i) = \infty$.

(7) Definition. The **mean recurrence time** μ_i of a state i is defined as

$$\mu_i = \mathbb{E}(T_i \mid X_0 = i) = \begin{cases} \sum_n n f_{ii}(n) & \text{if } i \text{ is persistent,} \\ \infty & \text{if } i \text{ is transient.} \end{cases}$$

Note that μ_i may be infinite even if i is persistent.

(8) Definition. For a persistent state i,

$$i \text{ is called } \begin{cases} \textbf{null} & \text{if } \mu_i = \infty, \\ \textbf{non-null} \text{ (or } \textbf{positive}) & \text{if } \mu_i < \infty. \end{cases}$$

There is a simple criterion for nullity in terms of the transition probabilities.

(9) Theorem. *A persistent state is null if and only if $p_{ii}(n) \to 0$ as $n \to \infty$; if this holds then $p_{ji}(n) \to 0$ for all j.*

Proof. We defer this until note (a) after Theorem (6.4.17). ■

Finally, for technical reasons we shall sometimes be interested in the epochs of time at which return to the starting point is possible.

(10) Definition. The **period** $d(i)$ of a state i is defined by $d(i) = \gcd\{n : p_{ii}(n) > 0\}$, the greatest common divisor of the epochs at which return is possible. We call i **periodic** if $d(i) > 1$ and **aperiodic** if $d(i) = 1$.

This to say, $p_{ii}(n) = 0$ unless n is a multiple of $d(i)$, and $d(i)$ is maximal with this property.

(11) Definition. A state is called **ergodic** if it is persistent, non-null, and aperiodic.

(12) Example. Random walk. Corollary (5.3.4) and Problem (5.12.5) show that the states of the simple random walk are all periodic with period 2, and

(a) transient, if $p \neq \frac{1}{2}$,

(b) null persistent, if $p = \frac{1}{2}$. ●

(13) Example. Branching process. Consider the branching process of Section 5.4 and suppose that $\mathbb{P}(Z_1 = 0) > 0$. Then 0 is called an *absorbing* state, because the chain never leaves it once it has visited it; all other states are transient. ●

Exercises for Section 6.2

1. Last exits. Let $l_{ij}(n) = \mathbb{P}(X_n = j, X_k \neq i \text{ for } 1 \leq k < n \mid X_0 = i)$, the probability that the chain passes from i to j in n steps without revisiting i. Writing

$$L_{ij}(s) = \sum_{n=1}^{\infty} s^n l_{ij}(n),$$

show that $P_{ij}(s) = P_{ii}(s)L_{ij}(s)$ if $i \neq j$. Deduce that the first passage times and last exit times have the same distribution for any Markov chain for which $P_{ii}(s) = P_{jj}(s)$ for all i and j. Give an example of such a chain.

2. Let X be a Markov chain containing an absorbing state s with which all other states i communicate, in the sense that $p_{is}(n) > 0$ for some $n = n(i)$. Show that all states other than s are transient.

3. Show that a state i is persistent if and only if the mean number of visits of the chain to i, having started at i, is infinite.

4. Visits. Let $V_j = |\{n \geq 1 : X_n = j\}|$ be the number of visits of the Markov chain X to j, and define $\eta_{ij} = \mathbb{P}(V_j = \infty \mid X_0 = i)$. Show that:

(a) $\eta_{ii} = \begin{cases} 1 & \text{if } i \text{ is persistent,} \\ 0 & \text{if } i \text{ is transient,} \end{cases}$

(b) $\eta_{ij} = \begin{cases} \mathbb{P}(T_j < \infty \mid X_0 = i) & \text{if } j \text{ is persistent,} \\ 0 & \text{if } j \text{ is transient,} \end{cases}$ where $T_j = \min\{n \geq 1 : X_n = j\}$.

5. Symmetry. The distinct pair i, j of states of a Markov chain is called *symmetric* if

$$\mathbb{P}(T_j < T_i \mid X_0 = i) = \mathbb{P}(T_i < T_j \mid X_0 = j),$$

where $T_i = \min\{n \geq 1 : X_n = i\}$. Show that, if $X_0 = i$ and i, j is symmetric, the expected number of visits to j before the chain revisits i is 1.

6.3 Classification of chains

We consider next the ways in which the states of a Markov chain are related to one other. This investigation will help us to achieve a full classification of the states in the language of the previous section.

(1) Definition. We say i **communicates with** j, written $i \to j$, if the chain may ever visit state j with positive probability, having started from i. That is, $i \to j$ if $p_{ij}(m) > 0$ for some $m \geq 0$. We say i and j **intercommunicate** if $i \to j$ and $j \to i$, in which case we write $i \leftrightarrow j$.

If $i \neq j$, then $i \to j$ if and only if $f_{ij} > 0$. Clearly $i \to i$ since $p_{ii}(0) = 1$, and it follows that $\leftrightarrow$ is an equivalence relation (*exercise*: if $i \leftrightarrow j$ and $j \leftrightarrow k$, show that $i \leftrightarrow k$). The state space S can be partitioned into the equivalence classes of $\leftrightarrow$. Within each equivalence class all states are of the same type.

(2) Theorem. *If $i \leftrightarrow j$ then:*
(a) *i and j have the same period,*
(b) *i is transient if and only if j is transient,*
(c) *i is null persistent if and only if j is null persistent.*

Proof. (b) If $i \leftrightarrow j$ then there exist $m, n \geq 0$ such that $\alpha = p_{ij}(m)p_{ji}(n) > 0$. By the Chapman–Kolmogorov equations (6.1.7),

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

for any non-negative integer r. Now sum over r to obtain

$$\sum_r p_{jj}(r) < \infty \quad \text{if} \quad \sum_r p_{ii}(r) < \infty.$$

Thus, by Corollary (6.2.4), j is transient if i is transient. The converse holds similarly and (b) is shown.
(a) This proof is similar and proceeds by way of Definition (6.2.10).
(c) We defer this until the next section. A possible route is by way of Theorem (6.2.9), but we prefer to proceed differently in order to avoid the danger of using a circular argument. ■

(3) Definition. A set C of states is called:
(a) **closed** if $p_{ij} = 0$ for all $i \in C$, $j \notin C$,
(b) **irreducible** if $i \leftrightarrow j$ for all $i, j \in C$.

Once the chain takes a value in a closed set C of states then it never leaves C subsequently. A closed set containing exactly one state is called *absorbing*; for example, the state 0 is absorbing for the branching process. It is clear that the equivalence classes of $\leftrightarrow$ are irreducible. We call an irreducible set C *aperiodic* (or *persistent, null,* and so on) if all the states in C have this property; Theorem (2) ensures that this is meaningful. If the whole state space S is irreducible, then we speak of the chain itself as having the property in question.

(4) Decomposition theorem. *The state space S can be partitioned uniquely as*

$$S = T \cup C_1 \cup C_2 \cup \cdots$$

where T is the set of transient states, and the C_i are irreducible closed sets of persistent states.

Proof. Let $C_1, C_2, \ldots$ be the persistent equivalence classes of $\leftrightarrow$. We need only show that each C_r is closed. Suppose on the contrary that there exist $i \in C_r$, $j \notin C_r$, such that $p_{ij} > 0$. Now $j \not\to i$, and therefore

$$\mathbb{P}\big(X_n \neq i \text{ for all } n \geq 1 \,\big|\, X_0 = i\big) \geq \mathbb{P}(X_1 = j \mid X_0 = i) > 0,$$

in contradiction of the assumption that i is persistent. ■

The decomposition theorem clears the air a little. For, on the the one hand, if $X_0 \in C_r$, say, the chain never leaves C_r and we might as well take C_r to be the whole state space. On the other hand, if $X_0 \in T$ then the chain either stays in T for ever or moves eventually to one of the C_k where it subsequently remains. Thus, either the chain always takes values in the set of transient states or it lies eventually in some irreducible closed set of persistent states. For the special case when S is finite the first of these possibilities cannot occur.

(5) Lemma. *If S is finite, then at least one state is persistent and all persistent states are non-null.*

Proof. If all states are transient, then take the limit through the summation sign to obtain the contradiction

$$1 = \lim_{n\to\infty} \sum_j p_{ij}(n) = 0$$

by Corollary (6.2.5). The same contradiction arises by Theorem (6.2.9) for the closed set of all null persistent states, should this set be non-empty. ■

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

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

The sets $\{1, 2\}$ and $\{5, 6\}$ are irreducible and closed, and therefore contain persistent non-null states. States 3 and 4 are transient because $3 \to 4 \to 6$ but return from 6 is impossible. All states have period 1 because $p_{ii}(1) > 0$ for all i. Hence, 3 and 4 are transient, and 1, 2, 5, and 6 are ergodic. Easy calculations give

$$f_{11}(n) = \begin{cases} p_{11} = \frac{1}{2} & \text{if } n = 1, \\ p_{12}(p_{22})^{n-2}p_{21} = \frac{1}{2}(\frac{3}{4})^{n-2}\frac{1}{4} & \text{if } n \geq 2, \end{cases}$$

and hence $\mu_1 = \sum_n nf_{11}(n) = 3$. Other mean recurrence times can be found similarly. The next section gives another way of finding the μ_i which usually requires less computation. ●

Exercises for Section 6.3

1. Let X be a Markov chain on $\{0, 1, 2, \dots\}$ with transition matrix given by $p_{0j} = a_j$ for $j \geq 0$, $p_{ii} = r$ and $p_{i,i-1} = 1 - r$ for $i \geq 1$. Classify the states of the chain, and find their mean recurrence times.

2. Determine whether or not the random walk on the integers having transition probabilities $p_{i,i+2} = p$, $p_{i,i-1} = 1 - p$, for all i, is persistent.

3. Classify the states of the Markov chains with transition matrices

(a)
$$\begin{pmatrix} 1-2p & 2p & 0 \\ p & 1-2p & p \\ 0 & 2p & 1-2p \end{pmatrix},$$

(b)
$$\begin{pmatrix} 0 & p & 0 & 1-p \\ 1-p & 0 & p & 0 \\ 0 & 1-p & 0 & p \\ p & 0 & 1-p & 0 \end{pmatrix}.$$

In each case, calculate $p_{ij}(n)$ and the mean recurrence times of the states.

4. A particle performs a random walk on the vertices of a cube. At each step it remains where it is with probability $\frac{1}{4}$, or moves to one of its neighbouring vertices each having probability $\frac{1}{4}$. Let v and w be two diametrically opposite vertices. If the walk starts at v, find:
(a) the mean number of steps until its first return to v,
(b) the mean number of steps until its first visit to w,
(c) the mean number of visits to w before its first return to v.

5. **Visits.** With the notation of Exercise (6.2.4), show that
(a) if $i \to j$ and i is persistent, then $\eta_{ij} = \eta_{ji} = 1$,
(b) $\eta_{ij} = 1$ if and only if $\mathbb{P}(T_j < \infty \mid X_0 = i) = \mathbb{P}(T_j < \infty \mid X_0 = j) = 1$.

6. **First passages.** Let $T_A = \min\{n \geq 0 : X_n \in A\}$, where X is a Markov chain and A is a subset of the state space S, and let $\eta_j = \mathbb{P}(T_A < \infty \mid X_0 = j)$. Show that

$$\eta_j = \begin{cases} 1 & \text{if } j \in A, \\ \sum_{k \in S} p_{jk}\eta_k & \text{if } j \notin A. \end{cases}$$

Show further that if $\mathbf{x} = (x_j : j \in S)$ is any non-negative solution of these equations then $x_j \geq \eta_j$ for all j.

7. **Mean first passage.** In the notation of Exercise (6), let $\rho_j = \mathbb{E}(T_A \mid X_0 = j)$. Show that

$$\rho_j = \begin{cases} 0 & \text{if } j \in A, \\ 1 + \sum_{k \in S} p_{jk}\rho_k & \text{if } j \notin A, \end{cases}$$

and that if $\mathbf{x} = (x_j : j \in S)$ is any non-negative solution of these equations then $x_j \geq \rho_j$ for all j.

8. Let X be an irreducible Markov chain and let A be a subset of the state space. Let S_r and T_r be the successive times at which the chain enters A and visits A respectively. Are the sequences $\{X_{S_r} : r \geq 1\}$, $\{X_{T_r} : r \geq 1\}$ Markov chains? What can be said about the times at which the chain exits A?

9. (a) Show that for each pair i, j of states of an irreducible aperiodic chain, there exists $N = N(i, j)$ such that $p_{ij}(r) > 0$ for all $r \geq N$.
(b) Show that there exists a function f such that, if $\mathbf{P}$ is the transition matrix of an irreducible aperiodic Markov chain with n states, then $p_{ij}(r) > 0$ for all states i, j, and all $r \geq f(n)$.
(c) Show further that $f(4) \geq 6$ and $f(n) \geq (n-1)(n-2)$.
[Hint: The postage stamp lemma asserts that, for a, b coprime, the smallest n such that all integers strictly exceeding n have the form $\alpha a + \beta b$ for some integers $\alpha, \beta \geq 0$ is $(a-1)(b-1)$.]

10. An urn initially contains n green balls and $n+2$ red balls. A ball is picked at random: if it is green then a red ball is also removed and both are discarded; if it is red then it is replaced together with an extra red and an extra green ball. This is repeated until there are no green balls in the urn. Show that the probability the process terminates is $1/(n+1)$.

Now reverse the rules: if the ball is green, it is replaced together with an extra green and an extra red ball; if it is red it is discarded along with a green ball. Show that the expected number of iterations until no green balls remain is $\sum_{j=1}^{n}(2j+1) = n(n+2)$. [Thus, a minor perturbation of a simple symmetric random walk can be non-null persistent, whereas the original is null persistent.]

6.4 Stationary distributions and the limit theorem

How does a Markov chain X_n behave after a long time n has elapsed? The sequence $\{X_n\}$ cannot generally, of course, converge to some particular state s since it enjoys the inherent random fluctuation which is specified by the transition matrix. However, we might hold out some hope that the *distribution* of X_n settles down. Indeed, subject to certain conditions this turns out to be the case. The classical study of limiting distributions proceeds by algebraic manipulation of the generating functions of Theorem (6.2.3); we shall avoid this here, contenting ourselves for the moment with results which are not quite the best possible but which have attractive probabilistic proofs. This section is in two parts, dealing respectively with stationary distributions and limit theorems.

(A) Stationary distributions. We shall see that the existence of a limiting distribution for X_n, as $n \to \infty$, is closely bound up with the existence of so-called 'stationary distributions'.

(1) Definition. The vector $\boldsymbol{\pi}$ is called a **stationary distribution** of the chain if $\boldsymbol{\pi}$ has entries $(\pi_j : j \in S)$ such that:

(a) $\pi_j \geq 0$ for all j, and $\sum_j \pi_j = 1$,

(b) $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$, which is to say that $\pi_j = \sum_i \pi_i p_{ij}$ for all j.

Such a distribution is called stationary for the following reason. Iterate (1b) to obtain $\boldsymbol{\pi}\mathbf{P}^2 = (\boldsymbol{\pi}\mathbf{P})\mathbf{P} = \boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$, and so

(2) $$\boldsymbol{\pi}\mathbf{P}^n = \boldsymbol{\pi} \qquad \text{for all } n \geq 0.$$

Now use Lemma (6.1.8) to see that if X_0 has distribution $\boldsymbol{\pi}$ then X_n has distribution $\boldsymbol{\pi}$ for all n, showing that the distribution of X_n is 'stationary' as time passes; in such a case, of course, $\boldsymbol{\pi}$ is also the limiting distribution of X_n as $n \to \infty$.

Following the discussion after the decomposition theorem (6.3.4), we shall assume henceforth that the chain is irreducible and shall investigate the existence of stationary distributions. No assumption of aperiodicity is required at this stage.

(3) Theorem. *An irreducible chain has a stationary distribution $\boldsymbol{\pi}$ if and only if all the states are non-null persistent; in this case, $\boldsymbol{\pi}$ is the unique stationary distribution and is given by $\pi_i = \mu_i^{-1}$ for each $i \in S$, where μ_i is the mean recurrence time of i.*

Stationary distributions $\boldsymbol{\pi}$ satisfy $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$. We may display a root $\mathbf{x}$ of the matrix equation $\mathbf{x} = \mathbf{x}\mathbf{P}$ explicitly as follows, whenever the chain is irreducible and persistent. Fix a state k and let $\rho_i(k)$ be the mean number of visits of the chain to the state i between two successive visits to state k; that is, $\rho_i(k) = \mathbb{E}(N_i \mid X_0 = k)$ where

$$N_i = \sum_{n=1}^{\infty} I_{\{X_n=i\}\cap\{T_k\geq n\}}$$

and T_k is the time of the first return to state k, as before. Note that $N_k = 1$ so that $\rho_k(k) = 1$, and that

$$\rho_i(k) = \sum_{n=1}^{\infty} \mathbb{P}(X_n = i,\ T_k \geq n \mid X_0 = k).$$

We write $\boldsymbol{\rho}(k)$ for the vector $(\rho_i(k) : i \in S)$. Clearly $T_k = \sum_{i\in S} N_i$, since the time between visits to k must be spent somewhere; taking expectations, we find that

$$\textbf{(4)} \qquad \mu_k = \sum_{i\in S} \rho_i(k),$$

so that the vector $\boldsymbol{\rho}(k)$ contains terms whose sum equals the mean recurrence time μ_k.

(5) Lemma. *For any state k of an irreducible persistent chain, the vector $\boldsymbol{\rho}(k)$ satisfies $\rho_i(k) < \infty$ for all i, and furthermore $\boldsymbol{\rho}(k) = \boldsymbol{\rho}(k)\mathbf{P}$.*

Proof. We show first that $\rho_i(k) < \infty$ when $i \neq k$. Write

$$l_{ki}(n) = \mathbb{P}(X_n = i,\ T_k \geq n \mid X_0 = k),$$

the probability that the chain reaches i in n steps but with no intermediate return to its starting point k. Clearly $f_{kk}(m+n) \geq l_{ki}(m)f_{ik}(n)$; this holds since the first return time to k equals $m+n$ if: (a) $X_m = i$, (b) there is no return to k prior to time m, and (c) the next subsequent visit to k takes place after another n steps. By the irreducibility of the chain, there exists n such that $f_{ik}(n) > 0$. With this choice of n, we have that $l_{ki}(m) \leq f_{kk}(m+n)/f_{ik}(n)$, and so

$$\rho_i(k) = \sum_{m=1}^{\infty} l_{ki}(m) \leq \frac{1}{f_{ik}(n)} \sum_{m=1}^{\infty} f_{kk}(m+n) \leq \frac{1}{f_{ik}(n)} < \infty$$

as required.

For the second statement of the lemma, we argue as follows. We have that $\rho_i(k) = \sum_{n=1}^{\infty} l_{ki}(n)$. Now $l_{ki}(1) = p_{ki}$, and

$$l_{ki}(n) = \sum_{j:j\neq k} \mathbb{P}\big(X_n = i,\ X_{n-1} = j,\ T_k \geq n \,\big|\, X_0 = k\big) = \sum_{j:j\neq k} l_{kj}(n-1)p_{ji} \quad \text{for } n \geq 2,$$

by conditioning on the value of X_{n-1}. Summing over $n \geq 2$, we obtain

$$\rho_i(k) = p_{ki} + \sum_{j:j\neq k}\Big(\sum_{n\geq 2} l_{kj}(n-1)\Big)p_{ji} = \rho_k(k)p_{ki} + \sum_{j:j\neq k} \rho_j(k)p_{ji}$$

since $\rho_k(k) = 1$. The lemma is proved. ■

We have from equation (4) and Lemma (5) that, for any irreducible persistent chain, the vector $\boldsymbol{\rho}(k)$ satisfies $\boldsymbol{\rho}(k) = \boldsymbol{\rho}(k)\mathbf{P}$, and furthermore that the components of $\boldsymbol{\rho}(k)$ are non-negative with sum μ_k. Hence, if $\mu_k < \infty$, the vector π with entries $\pi_i = \rho_i(k)/\mu_k$ satisfies $\pi = \pi\mathbf{P}$ and furthermore has non-negative entries which sum to 1; that is to say, π is a stationary distribution. We have proved that every non-null persistent irreducible chain has a stationary distribution, an important step towards the proof of the main theorem (3).

Before continuing with the rest of the proof of (3), we note a consequence of the results so far. Lemma (5) implies the existence of a root of the equation $\mathbf{x} = \mathbf{xP}$ whenever the chain is irreducible and persistent. Furthermore, there exists a root whose components are strictly positive (certainly there exists a non-negative root, and it is not difficult—see the argument after (8)—to see that this root may be taken strictly positive). It may be shown that this root is unique up to a multiplicative constant (Problem (6.15.7)), and we arrive therefore at the following useful conclusion.

(6) Theorem. *If the chain is irreducible and persistent, there exists a positive root* $\mathbf{x}$ *of the equation* $\mathbf{x} = \mathbf{xP}$, *which is unique up to a multiplicative constant. The chain is non-null if* $\sum_i x_i < \infty$ *and null if* $\sum_i x_i = \infty$.

Proof of (3). Suppose that π is a stationary distribution of the chain. If all states are transient then $p_{ij}(n) \to 0$, as $n \to \infty$, for all i and j by Corollary (6.2.5). From (2),

(7) $$\pi_j = \sum_i \pi_i p_{ij}(n) \to 0 \quad \text{as} \quad n \to \infty, \qquad \text{for all } i \text{ and } j,$$

which contradicts (1a). Thus all states are persistent. To see the validity of the limit in (7)†, let F be a finite subset of S and write

$$\begin{aligned}\sum_i \pi_i p_{ij}(n) &\le \sum_{i\in F} \pi_i p_{ij}(n) + \sum_{i\notin F} \pi_i \\ &\to \sum_{i\notin F} \pi_i \quad \text{as } n \to \infty, \qquad \text{since } F \text{ is finite} \\ &\to 0 \quad \text{as} \quad F \uparrow S.\end{aligned}$$

We show next that the existence of π implies that all states are non-null and that $\pi_i = \mu_i^{-1}$ for each i. Suppose that X_0 has distribution $\boldsymbol{\pi}$, so that $\mathbb{P}(X_0 = i) = \pi_i$ for each i. Then, by Problem (3.11.13a),

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

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

$$\begin{aligned}\mathbb{P}(T_j \ge n, X_0 = j) &= \mathbb{P}(X_0 = j,\ X_m \ne j \text{ for } 1 \le m \le n-1) \\ &= \mathbb{P}(X_m \ne j \text{ for } 1 \le m \le n-1) - \mathbb{P}(X_m \ne j \text{ for } 0 \le m \le n-1) \\ &= \mathbb{P}(X_m \ne j \text{ for } 0 \le m \le n-2) - \mathbb{P}(X_m \ne j \text{ for } 0 \le m \le n-1) \\ &\qquad\qquad \text{by homogeneity} \\ &= a_{n-2} - a_{n-1}\end{aligned}$$

†Actually this argument is a form of the bounded convergence theorem (5.6.12) applied to sums instead of to integrals. We shall make repeated use of this technique.

where $a_n = \mathbb{P}(X_m \neq j \text{ for } 0 \leq m \leq n)$. Sum over n to obtain

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

However, $a_n \to \mathbb{P}(X_m \neq j \text{ for all } m) = 0$ as $n \to \infty$, by the persistence of j (and surreptitious use of Problem (6.15.6)). We have shown that

$$\textbf{(8)} \qquad \pi_j \mu_j = 1,$$

so that $\mu_j = \pi_j^{-1} < \infty$ if $\pi_j > 0$. To see that $\pi_j > 0$ for all j, suppose on the contrary that $\pi_j = 0$ for some j. Then

$$0 = \pi_j = \sum_i \pi_i p_{ij}(n) \geq \pi_i p_{ij}(n) \qquad \text{for all } i \text{ and } n,$$

yielding that $\pi_i = 0$ whenever $i \to j$. The chain is assumed irreducible, so that $\pi_i = 0$ for all i in contradiction of the fact that the π_i have sum 1. Hence $\mu_j < \infty$ and all states of the chain are non-null. Furthermore, (8) specifies π_j uniquely as μ_j^{-1}.

Thus, if $\boldsymbol{\pi}$ exists then it is unique and all the states of the chain are non-null persistent. Conversely, if the states of the chain are non-null persistent then the chain has a stationary distribution given by (5). ■

We may now complete the proof of Theorem (6.3.2c).

Proof of (6.3.2c). Let $C(i)$ be the irreducible closed equivalence class of states which contains the non-null persistent state i. Suppose that $X_0 \in C(i)$. Then $X_n \in C(i)$ for all n, and (5) and (3) combine to tell us that all states in $C(i)$ are non-null. ■

(9) Example (6.3.6) revisited. To find μ_1 and μ_2 consider the irreducible closed set $C = \{1, 2\}$. If $X_0 \in C$, then solve the equation $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}_C$ for $\boldsymbol{\pi} = (\pi_1, \pi_2)$ in terms of

$$\mathbf{P}_C = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

to find the unique stationary distribution $\boldsymbol{\pi} = (\frac{1}{3}, \frac{2}{3})$, giving that $\mu_1 = \pi_1^{-1} = 3$ and $\mu_2 = \pi_2^{-1} = \frac{3}{2}$. Now find the other mean recurrence times yourself (*exercise*). ●

Theorem (3) provides a useful criterion for deciding whether or not an irreducible chain is non-null persistent: just look for a stationary distribution†. There is a similar criterion for the transience of irreducible chains.

(10) Theorem. *Let $s \in S$ be any state of an irreducible chain. The chain is transient if and only if there exists a non-zero solution $\{y_j : j \neq s\}$, satisfying $|y_j| \leq 1$ for all j, to the equations*

$$\textbf{(11)} \qquad y_i = \sum_{j:j\neq s} p_{ij} y_j, \qquad i \neq s.$$

†We emphasize that a stationary distribution is a *left* eigenvector of the transition matrix, not a *right* eigenvector.

Proof. The chain is transient if and only if s is transient. First suppose s is transient and define

$$\text{(12)} \qquad \begin{aligned} \tau_i(n) &= \mathbb{P}(\text{no visit to } s \text{ in first } n \text{ steps} \mid X_0 = i) \\ &= \mathbb{P}(X_m \neq s \text{ for } 1 \leq m \leq n \mid X_0 = i). \end{aligned}$$

Then

$$\tau_i(1) = \sum_{j:j\neq s} p_{ij}, \qquad \tau_i(n+1) = \sum_{j:j\neq s} p_{ij}\tau_j(n).$$

Furthermore, $\tau_i(n) \geq \tau_i(n+1)$, and so

$$\tau_i = \lim_{n\to\infty} \tau_i(n) = \mathbb{P}(\text{no visit to } s \text{ ever} \mid X_0 = i) = 1 - f_{is}$$

satisfies (11). (Can *you* prove this? Use the method of proof of (7).) Also $\tau_i > 0$ for some i, since otherwise $f_{is} = 1$ for all $i \neq s$, and therefore

$$f_{ss} = p_{ss} + \sum_{i:i\neq s} p_{si} f_{is} = \sum_i p_{si} = 1$$

by conditioning on X_1; this contradicts the transience of s.

Conversely, let $\mathbf{y}$ satisfy (11) with $|y_i| \leq 1$. Then

$$|y_i| \leq \sum_{j:j\neq s} p_{ij}|y_j| \leq \sum_{j:j\neq s} p_{ij} = \tau_i(1),$$

$$|y_i| \leq \sum_{j:j\neq s} p_{ij}\tau_j(1) = \tau_i(2),$$

and so on, where the $\tau_i(n)$ are given by (12). Thus $|y_i| \leq \tau_i(n)$ for all n. Let $n \to \infty$ to show that $\tau_i = \lim_{n\to\infty} \tau_i(n) > 0$ for some i, which implies that s is transient by the result of Problem (6.15.6). ■

This theorem provides a necessary and sufficient condition for persistence: an irreducible chain is persistent if and only if the only bounded solution to (11) is the zero solution. This combines with (3) to give a condition for null persistence. Another condition is the following (see Exercise (6.4.10)); a corresponding result holds for any countably infinite state space S.

(13) Theorem. *Let $s \in S$ be any state of an irreducible chain on $S = \{0, 1, 2, \dots\}$. The chain is persistent if there exists a solution $\{y_j : j \neq s\}$ to the inequalities*

$$\text{(14)} \qquad y_i \geq \sum_{j:j\neq s} p_{ij} y_j, \qquad i \neq s,$$

such that $y_i \to \infty$ as $i \to \infty$.

(15) Example. Random walk with retaining barrier. A particle performs a random walk on the non-negative integers with a retaining barrier at 0. The transition probabilities are

$$p_{0,0} = q, \qquad p_{i,i+1} = p \quad \text{for } i \geq 0, \qquad p_{i,i-1} = q \quad \text{for } i \geq 1,$$

where $p + q = 1$. Let $\rho = p/q$.

(a) If $q < p$, take $s = 0$ to see that $y_j = 1 - \rho^{-j}$ satisfies (11), and so the chain is transient.

(b) Solve the equation $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$ to find that there exists a stationary distribution, with $\pi_j = \rho^j(1-\rho)$, if and only if $q > p$. Thus the chain is non-null persistent if and only if $q > p$.

(c) If $q = p = \frac{1}{2}$, take $s = 0$ in (13) and check that $y_j = j$, $j \geq 1$, solves (14). Thus the chain is null persistent. Alternatively, argue as follows. The chain is persistent since symmetric random walk is persistent (just reflect negative excursions of a symmetric random walk into the positive half-line). Solve the equation $\mathbf{x} = \mathbf{x}\mathbf{P}$ to find that $x_i = 1$ for all i provides a root, unique up to a multiplicative constant. However, $\sum_i x_i = \infty$ so that the chain is null, by Theorem (6).

These conclusions match our intuitions well. ●

(B) Limit theorems. Next we explore the link between the existence of a stationary distribution and the limiting behaviour of the probabilities $p_{ij}(n)$ as $n \to \infty$. The following example indicates a difficulty which arises from periodicity.

(16) Example. If $S = \{1, 2\}$ and $p_{12} = p_{21} = 1$, then

$$p_{11}(n) = p_{22}(n) = \begin{cases} 0 & \text{if } n \text{ is odd,} \\ 1 & \text{if } n \text{ is even.} \end{cases}$$

Clearly $p_{ii}(n)$ does not converge as $n \to \infty$; the reason is that both states are periodic with period 2. ●

Until further notice we shall deal only with irreducible *aperiodic* chains. The principal result is the following theorem.

(17) Theorem. *For an irreducible aperiodic chain, we have that*

$$p_{ij}(n) \to \frac{1}{\mu_j} \quad \textit{as} \quad n \to \infty, \qquad \textit{for all } i \textit{ and } j.$$

We make the following remarks.

(a) If the chain is *transient* or *null persistent* then $p_{ij}(n) \to 0$ for all i and j, since $\mu_j = \infty$. We are now in a position to prove Theorem (6.2.9). Let $C(i)$ be the irreducible closed set of states which contains the persistent state i. If $C(i)$ is aperiodic then the result is an immediate consequence of (17); the periodic case can be treated similarly, but with slightly more difficulty (see note (d) following).

(b) If the chain is *non-null persistent* then $p_{ij}(n) \to \pi_j = \mu_j^{-1}$, where $\boldsymbol{\pi}$ is the unique stationary distribution by (3).

(c) It follows from (17) that the limit probability, $\lim_{n\to\infty} p_{ij}(n)$, does not depend on the starting point $X_0 = i$; that is, the chain forgets its origin. It is now easy to check that

$$\mathbb{P}(X_n = j) = \sum_i \mathbb{P}(X_0 = i) p_{ij}(n) \to \frac{1}{\mu_j} \quad \text{as} \quad n \to \infty$$

by Lemma (6.1.8), irrespective of the distribution of X_0.

(d) If $X = \{X_n\}$ is an irreducible chain with period d, then $Y = \{Y_n = X_{nd} : n \geq 0\}$ is an aperiodic chain, and it follows that

$$p_{jj}(nd) = \mathbb{P}(Y_n = j \mid Y_0 = j) \to \frac{d}{\mu_j} \quad \text{as} \quad n \to \infty.$$

Proof of (17). If the chain is transient then the result holds from Corollary (6.2.5). The persistent case is treated by an important technique known as 'coupling' which we met first in Section 4.12. Construct a 'coupled chain' $Z = (X, Y)$, being an ordered pair $X = \{X_n : n \geq 0\}$, $Y = \{Y_n : n \geq 0\}$ of *independent* Markov chains, each having state space S and transition matrix $\mathbf{P}$. Then $Z = \{Z_n = (X_n, Y_n) : n \geq 0\}$ takes values in $S \times S$, and it is easy to check that Z is a Markov chain with transition probabilities

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

Since X is irreducible and aperiodic, for any states i, j, k, l there exists $N = N(i, j, k, l)$ such that $p_{ik}(n)p_{jl}(n) > 0$ for all $n \geq N$; thus Z also is irreducible (see Exercise (6.3.9) or Problem (6.15.4); *only here* do we require that X be aperiodic).

Suppose that X is non-null persistent. Then X has a unique stationary distribution $\boldsymbol{\pi}$, by (3), and it is easy to see that Z has a stationary distribution $\boldsymbol{\nu} = (\nu_{ij} : i, j \in S)$ given by $\nu_{ij} = \pi_i \pi_j$; thus Z is also non-null persistent, by (3). Now, suppose that $X_0 = i$ and $Y_0 = j$, so that $Z_0 = (i, j)$. Choose any state $s \in S$ and let

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

denote the time of the first passage of Z to (s, s); from Problem (6.15.6) and the persistence of Z, $\mathbb{P}(T < \infty) = 1$. The central idea of the proof is the following observation. If $m \leq n$ and $X_m = Y_m$, then X_n and Y_n are identically distributed since the distributions of X_n and Y_n depend only upon the shared transition matrix $\mathbf{P}$ and upon the shared value of the chains at the mth stage. Thus, conditional on $\{T \leq n\}$, X_n and Y_n have the same distribution. We shall use this fact, together with the finiteness of T, to show that the ultimate distributions of X and Y are independent of their starting points. More precisely, starting from $Z_0 = (X_0, Y_0) = (i, j)$,

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

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

$$|p_{ik}(n) - p_{jk}(n)| \leq \mathbb{P}(T > n) \to 0 \quad \text{as} \quad n \to \infty$$

because $\mathbb{P}(T < \infty) = 1$; therefore,

(18) $$p_{ik}(n) - p_{jk}(n) \to 0 \quad \text{as} \quad n \to \infty \qquad \text{for all } i,\, j, \text{ and } k.$$

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

(19) $$\pi_k - p_{jk}(n) = \sum_i \pi_i \big(p_{ik}(n) - p_{jk}(n)\big) \to 0 \quad \text{as} \quad n \to \infty,$$

giving the result. To see that the limit in (19) follows from (18), use the bounded convergence argument in the proof of (7); for any finite subset F of S,

$$\begin{aligned}\sum_i \pi_i |p_{ik}(n) - p_{jk}(n)| &\le \sum_{i\in F} |p_{ik}(n) - p_{jk}(n)| + 2\sum_{i\notin F} \pi_i \\ &\to 2\sum_{i\notin F}\pi_i \quad \text{as} \quad n \to \infty\end{aligned}$$

which in turn tends to zero as $F \uparrow S$.

Finally, suppose that X is null persistent; the argument is a little trickier in this case. If Z is transient, then from Corollary (6.2.5) applied to Z,

$$\mathbb{P}\big(Z_n = (j,j) \,\big|\, Z_0 = (i,i)\big) = p_{ij}(n)^2 \to 0 \quad \text{as} \quad n \to \infty$$

and the result holds. If Z is non-null persistent then, starting from $Z_0 = (i,i)$, the epoch T_{ii}^Z of the first return of Z to (i,i) is no smaller than the epoch T_i of the first return of X to i; however, $\mathbb{E}(T_i) = \infty$ and $\mathbb{E}(T_{ii}^Z) < \infty$ which is a contradiction. Lastly, suppose that Z is null persistent. The argument which leads to (18) still holds, and we wish to deduce that

$$p_{ij}(n) \to 0 \quad \text{as} \quad n \to \infty \qquad \text{for all } i \text{ and } j.$$

If this does not hold then there exists a subsequence $n_1, n_2, \dots$ along which

(20) $$p_{ij}(n_r) \to \alpha_j \quad \text{as} \quad r \to \infty \qquad \text{for all } i \text{ and } j,$$

for some $\boldsymbol{\alpha}$, where the α_j are not all zero and are independent of i by (18); this is an application of the principle of 'diagonal selection' (see Billingsley 1995, Feller 1968, p. 336, or Exercise (5)). Equation (20) implies that, for any finite set F of states,

$$\sum_{j\in F} \alpha_j = \lim_{r\to\infty} \sum_{j\in F} p_{ij}(n_r) \le 1$$

and so $\alpha = \sum_j \alpha_j$ satisfies $0 < \alpha \le 1$. Furthermore

$$\sum_{k\in F} p_{ik}(n_r) p_{kj} \le p_{ij}(n_r + 1) = \sum_k p_{ik} p_{kj}(n_r);$$

let $r \to \infty$ here to deduce from (20) and bounded convergence (as used in the proof of (19)) that

$$\sum_{k\in F} \alpha_k p_{kj} \le \sum_k p_{ik} \alpha_j = \alpha_j,$$

and so, letting $F \uparrow S$, we obtain $\sum_k \alpha_k p_{kj} \le \alpha_j$ for each $j \in S$. However, equality must hold here, since if strict inequality holds for some j then

$$\sum_k \alpha_k = \sum_{k,j} \alpha_k p_{kj} < \sum_j \alpha_j,$$

which is a contradiction. Therefore

$$\sum_k \alpha_k p_{kj} = \alpha_j \qquad \text{for each } j \in S,$$

giving that $\pi = \{\alpha_j/\alpha : j \in S\}$ is a stationary distribution for X; this contradicts the nullity of X by (3). ■

The original and more general version of the ergodic theorem (17) for Markov chains does *not* assume that the chain is irreducible. We state it here; for a proof see Theorem (5.2.24) or Example (10.4.20).

(21) Theorem. *For any aperiodic state j of a Markov chain, $p_{jj}(n) \to \mu_j^{-1}$ as $n \to \infty$. Furthermore, if i is any other state then $p_{ij}(n) \to f_{ij}/\mu_j$ as $n \to \infty$.*

(22) Corollary. *Let*

$$\tau_{ij}(n) = \frac{1}{n} \sum_{m=1}^{n} p_{ij}(m)$$

be the mean proportion of elapsed time up to the nth step during which the chain was in state j, starting from i. If j is aperiodic, $\tau_{ij}(n) \to f_{ij}/\mu_j$ as $n \to \infty$.

Proof. *Exercise*: prove and use the fact that, as $n \to \infty$, $n^{-1}\sum_1^n x_i \to x$ if $x_n \to x$. ■

(23) Example. The coupling game. You may be able to amaze your friends and break the ice at parties with the following card 'trick'. A pack of cards is shuffled, and you deal the cards (face up) one by one. You instruct the audience as follows. Each person is to select some card, secretly, chosen from the first six or seven cards, say. If the face value of this card is m (aces count 1 and court cards count 10), let the next $m - 1$ cards pass and note the face value of the mth. Continuing according to this rule, there will arrive a last card in this sequence, face value X say, with fewer than X cards remaining. Call X the 'score'. Each person's score is known to that person but not to you, and can generally be any number between 1 and 10. At the end of the game, using an apparently fiendishly clever method you announce to the audience a number between 1 and 10. If few errors have been made, the majority of the audience will find that your number agrees with their score. Your popularity will then be assured, for a short while at least.

This is the 'trick'. You follow the same rules as the audience, beginning for the sake of simplicity with the first card. You will obtain a 'score' of Y, say, and it happens that there is a large probability that any given person obtains the score Y also; therefore you announce the score Y.

Why does the game often work? Suppose that someone picks the m_1th card, m_2th card, and so on, and you pick the n_1 $(= 1)$th, n_2th, etc. If $n_i = m_j$ for some i and j, then the two of you are 'stuck together' forever after, since the rules of the game require you to follow the same pattern henceforth; when this happens first, we say that 'coupling' has occurred. Prior to coupling, each time you read the value of a card, there is a positive probability that you will arrive at the next stage on exactly the same card as the other person. If the pack of cards were infinitely large, then coupling would certainly take place sooner or later, and it turns out that there is a good chance that coupling takes place before the last card of a regular pack has been dealt.

You may recognize the argument above as being closely related to that used near the beginning of the proof of Theorem (17). ●

Exercises for Section 6.4

1. The proof copy of a book is read by an infinite sequence of editors checking for mistakes. Each mistake is detected with probability p at each reading; between readings the printer corrects the detected mistakes but introduces a random number of new errors (errors may be introduced even if no mistakes were detected). Assuming as much independence as usual, and that the numbers of new errors after different readings are identically distributed, find an expression for the probability generating function of the stationary distribution of the number X_n of errors after the nth editor–printer cycle, whenever this exists. Find it explicitly when the printer introduces a Poisson-distributed number of errors at each stage.

2. Do the appropriate parts of Exercises (6.3.1)–(6.3.4) again, making use of the new techniques at your disposal.

3. **Dams.** Let X_n be the amount of water in a reservoir at noon on day n. During the 24 hour period beginning at this time, a quantity Y_n of water flows into the reservoir, and just before noon on each day exactly one unit of water is removed (if this amount can be found). The maximum capacity of the reservoir is K, and excessive inflows are spilled and lost. Assume that the Y_n are independent and identically distributed random variables and that, by rounding off to some laughably small unit of volume, all numbers in this exercise are non-negative integers. Show that (X_n) is a Markov chain, and find its transition matrix and an expression for its stationary distribution in terms of the probability generating function G of the Y_n.

Find the stationary distribution when Y has probability generating function $G(s) = p(1-qs)^{-1}$.

4. Show by example that chains which are not irreducible may have many different stationary distributions.

5. **Diagonal selection.** Let $(x_i(n) : i, n \geq 1)$ be a bounded collection of real numbers. Show that there exists an increasing sequence $n_1, n_2, \dots$ of positive integers such that $\lim_{r\to\infty} x_i(n_r)$ exists for all i. Use this result to prove that, for an irreducible Markov chain, if it is not the case that $p_{ij}(n) \to 0$ as $n \to \infty$ for all i and j, then there exists a sequence $(n_r : r \geq 1)$ and a vector $\boldsymbol{\alpha}$ ($\neq \mathbf{0}$) such that $p_{ij}(n_r) \to \alpha_j$ as $r \to \infty$ for all i and j.

6. **Random walk on a graph.** A particle performs a random walk on the vertex set of a connected graph G, which for simplicity we assume to have neither loops nor multiple edges. At each stage it moves to a neighbour of its current position, each such neighbour being chosen with equal probability. If G has η ($< \infty$) edges, show that the stationary distribution is given by $\pi_v = d_v/(2\eta)$, where d_v is the degree of vertex v.

7. Show that a random walk on the infinite binary tree is transient.

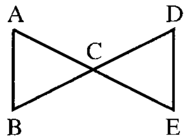
8. At each time $n = 0, 1, 2, \dots$ a number Y_n of particles enters a chamber, where $\{Y_n : n \geq 0\}$ are independent and Poisson distributed with parameter λ. Lifetimes of particles are independent and geometrically distributed with parameter p. Let X_n be the number of particles in the chamber at time n. Show that X is a Markov chain, and find its stationary distribution.

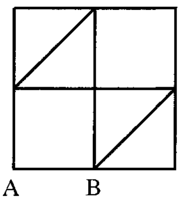
9. A random sequence of convex polygons is generated by picking two edges of the current polygon at random, joining their midpoints, and picking one of the two resulting smaller polygons at random to be the next in the sequence. Let $X_n + 3$ be the number of edges of the nth polygon thus constructed. Find $\mathbb{E}(X_n)$ in terms of X_0, and find the stationary distribution of the Markov chain X.

10. Let s be a state of an irreducible Markov chain on the non-negative integers. Show that the chain is persistent if there exists a solution $\mathbf{y}$ to the equations $y_i \geq \sum_{j:j\neq s} p_{ij} y_j$, $i \neq s$, satisfying $y_i \to \infty$.

11. Bow ties. A particle performs a random walk on a bow tie ABCDE drawn beneath on the left, where C is the knot. From any vertex its next step is equally likely to be to any neighbouring vertex. Initially it is at A. Find the expected value of:

(a) the time of first return to A,
(b) the number of visits to D before returning to A,
(c) the number of visits to C before returning to A,
(d) the time of first return to A, given no prior visit by the particle to E,
(e) the number of visits to D before returning to A, given no prior visit by the particle to E.





12. A particle starts at A and executes a symmetric random walk on the graph drawn above on the right. Find the expected number of visits to B before it returns to A.

6.5 Reversibility

Most laws of physics have the property that they would make the same assertions if the universal clock were reversed and time were made to run backwards. It may be implausible that nature works in such ways (have *you* ever seen the fragments of a shattered teacup re-assemble themselves on the table from which it fell?), and so one may be led to postulate a non-decreasing quantity called 'entropy'. However, never mind such objections; let us think about the reversal of the time scale of a Markov chain.

Suppose that $\{X_n : 0 \leq n \leq N\}$ is an irreducible non-null persistent Markov chain, with transition matrix $\mathbf{P}$ and stationary distribution $\boldsymbol{\pi}$. Suppose further that X_n has distribution $\boldsymbol{\pi}$ for every n. Define the 'reversed chain' Y by $Y_n = X_{N-n}$ for $0 \leq n \leq N$. We first check as follows that Y is a Markov chain.

(1) Theorem. *The sequence Y is a Markov chain with $\mathbb{P}(Y_{n+1} = j \mid Y_n = i) = (\pi_j/\pi_i)p_{ji}$.*

Proof. We have as required that

$$\begin{aligned}
\mathbb{P}\big(Y_{n+1} = i_{n+1} \,\big|\, Y_n = i_n,\ Y_{n-1} = i_{n-1}, \dots,\ Y_0 = i_0\big) \\
&= \frac{\mathbb{P}(Y_k = i_k,\ 0 \leq k \leq n+1)}{\mathbb{P}(Y_k = i_k,\ 0 \leq k \leq n)} \\
&= \frac{\mathbb{P}(X_{N-n-1} = i_{n+1},\ X_{N-n} = i_n, \dots,\ X_N = i_0)}{\mathbb{P}(X_{N-n} = i_n, \dots,\ X_N = i_0)} \\
&= \frac{\pi_{i_{n+1}} p_{i_{n+1},i_n} p_{i_n,i_{n-1}} \cdots p_{i_1,i_0}}{\pi_{i_n} p_{i_n,i_{n-1}} \cdots p_{i_1,i_0}} = \frac{\pi_{i_{n+1}} p_{i_{n+1},i_n}}{\pi_{i_n}}. \qquad \blacksquare
\end{aligned}$$

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

(2) Definition. Let $X = \{X_n : 0 \le n \le N\}$ be an irreducible Markov chain such that X_n has the stationary distribution $\boldsymbol{\pi}$ for all n. The chain is called **reversible** if the transition matrices of X and its time-reversal Y are the same, which is to say that

$$\pi_i p_{ij} = \pi_j p_{ji} \qquad \text{for all } i, j. \tag{3}$$

Equations (3) are called the *detailed balance* equations, and they are pivotal to the study of reversible chains. More generally we say that a transition matrix $\mathbf{P}$ and a distribution $\boldsymbol{\lambda}$ are *in detailed balance* if $\lambda_i p_{ij} = \lambda_j p_{ji}$ for all $i, j \in S$. An irreducible chain X having a stationary distribution $\boldsymbol{\pi}$ is called *reversible in equilibrium* if its transition matrix $\mathbf{P}$ is in detailed balance with $\boldsymbol{\pi}$. It may be noted that a chain having a tridiagonal transition matrix is reversible in equilibrium; see Exercise (1) and Problem (6.15.16c).

The following theorem provides a useful way of finding the stationary distribution of an irreducible chain whose transition matrix $\mathbf{P}$ is in detailed balance with some distribution $\boldsymbol{\lambda}$.

(4) Theorem. *Let* $\mathbf{P}$ *be the transition matrix of an irreducible chain* X*, and suppose that there exists a distribution* $\boldsymbol{\pi}$ *such that* $\pi_i p_{ij} = \pi_j p_{ji}$ *for all* $i, j \in S$*. Then* $\boldsymbol{\pi}$ *is a stationary distribution of the chain. Furthermore,* X *is reversible in equilibrium.*

Proof. Suppose that $\boldsymbol{\pi}$ satisfies the conditions of the theorem. Then

$$\sum_i \pi_i p_{ij} = \sum_i \pi_j p_{ji} = \pi_j \sum_i p_{ji} = \pi_j$$

and so $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$, whence $\boldsymbol{\pi}$ is stationary. The reversibility in equilibrium of X follows from the definition (2). ■

Although the above definition of reversibility applies to a Markov chain defined on only finitely many time points $0, 1, 2, \dots, N$, if is easily seen to apply to the infinite time set $0, 1, 2, \dots$. It may be extended also to the doubly-infinite time set $\dots, -2, -1, 0, 1, 2, \dots$. In the last case it is necessary to note the following fact. Let $X = \{X_n : -\infty < n < \infty\}$ be a Markov chain with stationary distribution $\boldsymbol{\pi}$. In order that X_n have distribution $\boldsymbol{\pi}$ for all n, it is not generally sufficient that X_0 has distribution $\boldsymbol{\pi}$.

(5) Example. Ehrenfest model of diffusion†. Two containers A and B are placed adjacent to each other and gas is allowed to pass through a small aperture joining them. A total of m gas molecules is distributed between the containers. We assume that at each epoch of time one molecule, picked uniformly at random from the m available, passes through this aperture. Let X_n be the number of molecules in container A after n units of time has passed. Clearly $\{X_n\}$ is a Markov chain with transition matrix

$$p_{i,i+1} = 1 - \frac{i}{m}, \quad p_{i,i-1} = \frac{i}{m} \quad \text{if} \quad 0 \le i \le m.$$

Rather than solve the equation $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$ to find the stationary distribution, we note that such a reasonable diffusion model should be reversible in equilibrium. Look for solutions of the detailed balance equations $\pi_i p_{ij} = \pi_j p_{ji}$ to obtain $\pi_i = \binom{m}{i}(\frac{1}{2})^m$. ●

†Originally introduced by Paul and Tatiana Ehrenfest as the 'dog–flea model'.

Here is a way of thinking about reversibility and the equations $\pi_i p_{ij} = \pi_j p_{ji}$. Suppose we are provided with a Markov chain with state space S and stationary distribution $\boldsymbol{\pi}$. To this chain there corresponds a 'network' as follows. The nodes of the network are the states in S, and arrows are added between certain pairs of nodes; an arrow is added pointing from state i to state j whenever $p_{ij} > 0$. We are provided with one unit of material (disease, water, or sewage, perhaps) which is distributed about the nodes of the network and allowed to flow along the arrows. The transportation rule is as follows: at each epoch of time a proportion p_{ij} of the amount of material at node i is transported to node j. Initially the material is distributed in such a way that exactly π_i of it is at node i, for each i. It is a simple calculation that the amount at node i after one epoch of time is $\sum_j \pi_j p_{ji}$, which equals π_i, since $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$. Therefore the system is in equilibrium: there is a 'global balance' in the sense that the total quantity leaving each node equals the total quantity arriving there. There may or may not be a 'local balance', in the sense that, for all i, j, the amount flowing from i to j equals the amount flowing from j to i. Local balance occurs if and only if $\pi_i p_{ij} = \pi_j p_{ji}$ for all i, j, which is to say that $\mathbf{P}$ and $\boldsymbol{\pi}$ are in detailed balance.

Exercises for Section 6.5

1. A random walk on the set $\{0, 1, 2, \dots, b\}$ has transition matrix given by $p_{00} = 1 - \lambda_0$, $p_{bb} = 1 - \mu_b$, $p_{i,i+1} = \lambda_i$ and $p_{i+1,i} = \mu_{i+1}$ for $0 \le i < b$, where $0 < \lambda_i, \mu_i < 1$ for all i, and $\lambda_i + \mu_i = 1$ for $1 \le i < b$. Show that this process is reversible in equilibrium.

2. Kolmogorov's criterion for reversibility. Let X be an irreducible non-null persistent aperiodic Markov chain. Show that X is reversible in equilibrium if and only if

$$p_{j_1 j_2} p_{j_2 j_3} \cdots p_{j_{n-1} j_n} p_{j_n j_1} = p_{j_1 j_n} p_{j_n j_{n-1}} \cdots p_{j_2 j_1}$$

for all n and all finite sequences $j_1, j_2, \dots, j_n$ of states.

3. Let X be a reversible Markov chain, and let C be a non-empty subset of the state space S. Define the Markov chain Y on S by the transition matrix $\mathbf{Q} = (q_{ij})$ where

$$q_{ij} = \begin{cases} \beta p_{ij} & \text{if } i \in C \text{ and } j \notin C, \\ p_{ij} & \text{otherwise,} \end{cases}$$

for $i \neq j$, and where β is a constant satisfying $0 < \beta < 1$. The diagonal terms q_{ii} are arranged so that $\mathbf{Q}$ is a stochastic matrix. Show that Y is reversible in equilibrium, and find its stationary distribution. Describe the situation in the limit as $\beta \downarrow 0$.

4. Can a reversible chain be periodic?

5. Ehrenfest dog–flea model. The dog–flea model of Example (6.5.5) is a Markov chain X on the state space $\{0, 1, \dots, m\}$ with transition probabilities

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

Show that, if $X_0 = i$,

$$\mathbb{E}\left(X_n - \frac{m}{2}\right) = \left(i - \frac{m}{2}\right)\left(1 - \frac{2}{m}\right)^n \to 0 \quad \text{as } n \to \infty.$$

6. Which of the following (when stationary) are reversible Markov chains?

(a) The chain $X = \{X_n\}$ having transition matrix $\mathbf{P} = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$ where $\alpha + \beta > 0$.

(b) The chain $Y = \{Y_n\}$ having transition matrix $\mathbf{P} = \begin{pmatrix} 0 & p & 1-p \\ 1-p & 0 & p \\ p & 1-p & 0 \end{pmatrix}$ where $0 < p < 1$.

(c) $Z_n = (X_n, Y_n)$, where X_n and Y_n are independent and satisfy (a) and (b).

7. Let X_n, Y_n be independent simple random walks. Let Z_n be (X_n, Y_n) truncated to lie in the region $X_n \geq 0$, $Y_n \geq 0$, $X_n + Y_n \leq a$ where a is integral. Find the stationary distribution of Z_n.

8. Show that an irreducible Markov chain with a finite state space and transition matrix $\mathbf{P}$ is reversible in equilibrium if and only if $\mathbf{P} = \mathbf{DS}$ for some symmetric matrix $\mathbf{S}$ and diagonal matrix $\mathbf{D}$ with strictly positive diagonal entries. Show further that for reversibility in equilibrium to hold, it is necessary but not sufficient that $\mathbf{P}$ has real eigenvalues.

9. **Random walk on a graph.** Let G be a finite connected graph with neither loops nor multiple edges, and let X be a random walk on G as in Exercise (6.4.6). Show that X is reversible in equilibrium.

6.6 Chains with finitely many states

The theory of Markov chains is much simplified by the condition that S be finite. By Lemma (6.3.5), if S is finite and irreducible then it is necessarily non-null persistent. It may even be possible to calculate the n-step transition probabilities explicitly. Of central importance here is the following algebraic theorem, in which $i = \sqrt{-1}$. Let N denote the cardinality of S.

(1) Theorem (Perron–Frobenius). *If* $\mathbf{P}$ *is the transition matrix of a finite irreducible chain with period* d *then*:

(a) $\lambda_1 = 1$ *is an eigenvalue of* $\mathbf{P}$,

(b) *the* d *complex roots of unity*

$$\lambda_1 = \omega^0,\ \lambda_2 = \omega^1, \dots,\ \lambda_d = \omega^{d-1} \quad \textit{where } \omega = e^{2\pi i/d},$$

are eigenvalues of $\mathbf{P}$,

(c) *the remaining eigenvalues* $\lambda_{d+1}, \dots, \lambda_N$ *satisfy* $|\lambda_j| < 1$.

If the eigenvalues $\lambda_1, \dots, \lambda_N$ are distinct then it is well known that there exists a matrix $\mathbf{B}$ such that $\mathbf{P} = \mathbf{B}^{-1}\mathbf{\Lambda}\mathbf{B}$ where $\mathbf{\Lambda}$ is the diagonal matrix with entries $\lambda_1, \dots, \lambda_N$. Thus

$$\mathbf{P}^n = \mathbf{B}^{-1}\mathbf{\Lambda}^n\mathbf{B} = \mathbf{B}^{-1}\begin{pmatrix} \lambda_1^n & 0 & \cdots & 0 \\ 0 & \lambda_2^n & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_N^n \end{pmatrix}\mathbf{B}.$$

The rows of $\mathbf{B}$ are left eigenvectors of $\mathbf{P}$. We can use the Perron–Frobenius theorem to explore the properties of $\mathbf{P}^n$ for large n. For example, if the chain is aperiodic then $d = 1$ and

$$\mathbf{P}^n \to \mathbf{B}^{-1}\begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix}\mathbf{B} \quad \text{as} \quad n \to \infty.$$

When the eigenvalues of the matrix $\mathbf{P}$ are not distinct, then $\mathbf{P}$ cannot always be reduced to the diagonal canonical form in this way. The best that we may be able to do is to rewrite $\mathbf{P}$ in its 'Jordan canonical form' $\mathbf{P} = \mathbf{B}^{-1}\mathbf{M}\mathbf{B}$ where

$$\mathbf{M} = \begin{pmatrix} \mathbf{J}_1 & \mathbf{0} & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{J}_2 & \mathbf{0} & \cdots \\ \mathbf{0} & \mathbf{0} & \mathbf{J}_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

and $\mathbf{J}_1, \mathbf{J}_2, \ldots$ are square matrices given as follows. Let $\lambda_1, \lambda_2, \ldots, \lambda_m$ be the distinct eigenvalues of $\mathbf{P}$ and let k_i be the multiplicity of λ_i. Then

$$\mathbf{J}_i = \begin{pmatrix} \lambda_i & 1 & 0 & 0 & \cdots \\ 0 & \lambda_i & 1 & 0 & \cdots \\ 0 & 0 & \lambda_i & 1 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

is a $k_i \times k_i$ matrix with each diagonal term λ_i, each superdiagonal term 1, and all other terms 0. Once again we have that $\mathbf{P}^n = \mathbf{B}^{-1}\mathbf{M}^n\mathbf{B}$, where $\mathbf{M}^n$ has quite a simple form (see Cox and Miller (1965, p. 118 *et seq.*) for more details).

(2) Example. Inbreeding. Consider the genetic model described in Example (6.1.11c) and suppose that C_1 can take the values A or a on each of two homologous chromosomes. Then the possible types of individuals can be denoted by

$$AA, \ Aa \ (\equiv aA), \ aa,$$

and mating between types is denoted by

$$AA \times AA, \ AA \times Aa, \quad \text{and so on.}$$

As described in Example (6.1.11c), meiosis causes the offspring's chromosomes to be selected randomly from each parent; in the simplest case (since there are two choices for each of two places) each outcome has probability $\frac{1}{4}$. Thus for the offspring of $AA \times Aa$ the four possible outcomes are

$$AA, \ Aa, \ AA, \ Aa$$

and $\mathbb{P}(AA) = \mathbb{P}(Aa) = \frac{1}{2}$. For the cross $Aa \times Aa$,

$$\mathbb{P}(AA) = \mathbb{P}(aa) = \tfrac{1}{2}\mathbb{P}(Aa) = \tfrac{1}{4}.$$

Clearly the offspring of $AA \times AA$ can only be AA, and those of $aa \times aa$ can only be aa.

We now construct a Markov chain by mating an individual with itself, then crossing a single resulting offspring with itself, and so on. (This scheme is possible with plants.) The genetic types of this sequence of individuals constitute a Markov chain with three states, AA, Aa, aa. In view of the above discussion, the transition matrix is

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

and the reader may verify that

$$\mathbf{P}^n = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} - (\frac{1}{2})^{n+1} & (\frac{1}{2})^n & \frac{1}{2} - (\frac{1}{2})^{n+1} \\ 0 & 0 & 1 \end{pmatrix} \to \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \\ 0 & 0 & 1 \end{pmatrix} \quad \text{as} \quad n \to \infty.$$

Thus, ultimately, inbreeding produces a pure (AA or aa) line for which all subsequent offspring have the same type. In like manner one can consider the progress of many different breeding schemes which include breeding with rejection of unfavourable genes, back-crossing to encourage desirable genes, and so on. ●

Exercises for Section 6.6

The first two exercises provide proofs that a Markov chain with finitely many states has a stationary distribution.

1. The Markov–Kakutani theorem asserts that, for any convex compact subset C of $\mathbb{R}^n$ and any linear continuous mapping T of C into C, T has a fixed point (in the sense that $T(x) = x$ for some $x \in C$). Use this to prove that a finite stochastic matrix has a non-negative non-zero left eigenvector corresponding to the eigenvalue 1.

2. Let $\mathbf{T}$ be a $m \times n$ matrix and let $\mathbf{v} \in \mathbb{R}^n$. Farkas's theorem asserts that exactly one of the following holds:

(i) there exists $\mathbf{x} \in \mathbb{R}^m$ such that $\mathbf{x} \geq \mathbf{0}$ and $\mathbf{xT} = \mathbf{v}$,

(ii) there exists $\mathbf{y} \in \mathbb{R}^n$ such that $\mathbf{yv}' < 0$ and $\mathbf{Ty}' \geq \mathbf{0}$.

Use this to prove that a finite stochastic matrix has a non-negative non-zero left eigenvector corresponding to the eigenvalue 1.

3. Arbitrage. Suppose you are betting on a race with m possible outcomes. There are n bookmakers, and a unit stake with the ith bookmaker yields t_{ij} if the jth outcome of the race occurs. A vector $\mathbf{x} = (x_1, x_2, \dots, x_n)$, where $x_r \in (-\infty, \infty)$ is your stake with the rth bookmaker, is called a *betting scheme*. Show that exactly one of (a) and (b) holds:

(a) there exists a probability mass function $\mathbf{p} = (p_1, p_2, \dots, p_m)$ such that $\sum_{j=1}^{m} t_{ij} p_j = 0$ for all values of i,

(b) there exists a betting scheme $\mathbf{x}$ for which you surely win, that is, $\sum_{i=1}^{n} x_i t_{ij} > 0$ for all j.

4. Let X be a Markov chain with state space $S = \{1, 2, 3\}$ and transition matrix

$$\mathbf{P} = \begin{pmatrix} 1-p & p & 0 \\ 0 & 1-p & p \\ p & 0 & 1-p \end{pmatrix}$$

where $0 < p < 1$. Prove that

$$\mathbf{P}^n = \begin{pmatrix} a_{1n} & a_{2n} & a_{3n} \\ a_{3n} & a_{1n} & a_{2n} \\ a_{2n} & a_{3n} & a_{1n} \end{pmatrix}$$

where $a_{1n} + \omega a_{2n} + \omega^2 a_{3n} = (1 - p + p\omega)^n$, ω being a complex cube root of 1.

5. Let $\mathbf{P}$ be the transition matrix of a Markov chain with finite state space. Let $\mathbf{I}$ be the identity matrix, $\mathbf{U}$ the $|S| \times |S|$ matrix with all entries unity, and $\mathbf{1}$ the row $|S|$-vector with all entries unity. Let $\boldsymbol{\pi}$ be a non-negative vector with $\sum_i \pi_i = 1$. Show that $\boldsymbol{\pi}\mathbf{P} = \boldsymbol{\pi}$ if and only if $\boldsymbol{\pi}(\mathbf{I} - \mathbf{P} + \mathbf{U}) = \mathbf{1}$. Deduce that if $\mathbf{P}$ is irreducible then $\boldsymbol{\pi} = \mathbf{1}(\mathbf{I} - \mathbf{P} + \mathbf{U})^{-1}$.

6. Chess. A chess piece performs a random walk on a chessboard; at each step it is equally likely to make any one of the available moves. What is the mean recurrence time of a corner square if the piece is a: (a) king? (b) queen? (c) bishop? (d) knight? (e) rook?

7. Chess continued. A rook and a bishop perform independent symmetric random walks with synchronous steps on a 4×4 chessboard (16 squares). If they start together at a corner, show that the expected number of steps until they meet again at the same corner is 448/3.

8. Find the n-step transition probabilities $p_{ij}(n)$ for the chain X having transition matrix

$$\mathbf{P} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{3} & \frac{1}{4} & \frac{5}{12} \\ \frac{2}{3} & \frac{1}{4} & \frac{1}{12} \end{pmatrix}.$$

6.7 Branching processes revisited

The foregoing general theory is an attractive and concise account of the evolution through time of a Markov chain. Unfortunately, it is an inadequate description of many specific Markov chains. Consider for example a branching process $\{Z_0, Z_1, \dots\}$ where $Z_0 = 1$. If there is strictly positive probability $\mathbb{P}(Z_1 = 0)$ that each family is empty then 0 is an absorbing state. Hence 0 is persistent non-null, and all other states are transient. The chain is not irreducible but there exists a unique stationary distribution $\boldsymbol{\pi}$ given by $\pi_0 = 1$, $\pi_i = 0$ if $i > 0$. These facts tell us next to nothing about the behaviour of the process, and we must look elsewhere for detailed information. The difficulty is that the process may behave in one of various qualitatively different ways depending, for instance, on whether or not it ultimately becomes extinct. One way of approaching the problem is to study the behaviour of the process *conditional* upon the occurrence of some event, such as extinction, or on the value of some random variable, such as the total number $\sum_i Z_i$ of progeny. This section contains an outline of such a method.

Let f and G be the mass function and generating function of a typical family size Z_1:

$$f(k) = \mathbb{P}(Z_1 = k), \qquad G(s) = \mathbb{E}(s^{Z_1}).$$

Let $T = \inf\{n : Z_n = 0\}$ be the time of extinction, with the convention that the infimum of the empty set is $+\infty$. Roughly speaking, if $T = \infty$ then the process will grow beyond all possible bounds, whilst if $T < \infty$ then the size of the process never becomes very large and subsequently reduces to zero. Think of $\{Z_n\}$ as a fluctuating sequence which either becomes so large that it escapes to ∞ or is absorbed at 0 during one of its fluctuations. From the results of Section 5.4, the probability $\mathbb{P}(T < \infty)$ of ultimate extinction is the smallest non-negative root of the equation $s = G(s)$. Now let

$$E_n = \{n < T < \infty\}$$

be the event that extinction occurs at some time after n. We shall study the distribution of Z_n conditional upon the occurrence of E_n. Let

$${}_0p_j^{(n)} = \mathbb{P}(Z_n = j \mid E_n)$$

be the conditional probability that $Z_n = j$ given the future extinction of Z. We are interested in the limiting value

$${}_0\pi_j = \lim_{n\to\infty} {}_0p_j(n),$$

if this limit exists. To avoid certain trivial cases we assume henceforth that

$$0 < f(0) + f(1) < 1, \qquad f(0) > 0;$$

these conditions imply for example that $0 < \mathbb{P}(E_n) < 1$ and that the probability η of ultimate extinction satisfies $0 < \eta \le 1$.

(1) Lemma. *If* $\mathbb{E}(Z_1) < \infty$ *then* $\lim_{n\to\infty} {}_0p_j(n) = {}_0\pi_j$ *exists. The generating function*

$$G^\pi(s) = \sum_j {}_0\pi_j s^j$$

satisfies the functional equation

$$G^\pi\big(\eta^{-1}G(s\eta)\big) = mG^\pi(s) + 1 - m \tag{2}$$

where η *is the probability of ultimate extinction and* $m = G'(\eta)$.

Note that if $\mu = \mathbb{E}Z_1 \le 1$ then $\eta = 1$ and $m = \mu$. Thus (2) reduces to

$$G^\pi(G(s)) = \mu G^\pi(s) + 1 - \mu.$$

Whatever the value of μ, we have that $G'(\eta) \le 1$, with equality if and only if $\mu = 1$.

Proof. For $s \in [0, 1)$, let

$$\begin{aligned} G_n^\pi(s) &= \mathbb{E}(s^{Z_n} \mid E_n) = \sum_j {}_0p_j(n)s^j \\ &= \sum_j s^j \frac{\mathbb{P}(Z_n = j,\ E_n)}{\mathbb{P}(E_n)} = \frac{G_n(s\eta) - G_n(0)}{\eta - G_n(0)} \end{aligned}$$

where $G_n(s) = \mathbb{E}(s^{Z_n})$ as before, since

$$\begin{aligned} \mathbb{P}(Z_n = j,\ E_n) &= \mathbb{P}(Z_n = j \text{ and all subsequent lines die out}) \\ &= \mathbb{P}(Z_n = j)\eta^j \quad \text{if} \quad j \ge 1, \end{aligned}$$

and $\mathbb{P}(E_n) = \mathbb{P}(T < \infty) - \mathbb{P}(T \le n) = \eta - G_n(0)$. Let

$$H_n(s) = \frac{\eta - G_n(s)}{\eta - G_n(0)}, \quad h(s) = \frac{\eta - G(s)}{\eta - s}, \quad 0 \le s < \eta,$$

so that

$$G_n^\pi(s) = 1 - H_n(s\eta). \tag{3}$$

Note that H_n has domain $[0, \eta)$ and G_n^π has domain $[0, 1)$. By Theorem (5.4.1),

$$\frac{H_n(s)}{H_{n-1}(s)} = \frac{h(G_{n-1}(s))}{h(G_{n-1}(0))}.$$

However, G_{n-1} is non-decreasing, and h is non-decreasing because G is convex on $[0, \eta)$, giving that $H_n(s) \ge H_{n-1}(s)$ for $s < \eta$. Hence, by (3), the limits

$$\lim_{n\to\infty} G_n^\pi(s) = G^\pi(s) \quad \text{and} \quad \lim_{n\to\infty} H_n(s\eta) = H(s\eta)$$

exist for $s \in [0, 1)$ and satisfy

(4) $$G^{\pi}(s) = 1 - H(s\eta) \quad \text{if} \quad 0 \le s < 1.$$

Thus the coefficient ${}_0\pi_j$ of s^j in $G^{\pi}(s)$ exists for all j as required. Furthermore, if $0 \le s < \eta$,

(5) $$\begin{aligned} H_n(G(s)) &= \frac{\eta - G_n(G(s))}{\eta - G_n(0)} = \frac{\eta - G(G_n(0))}{\eta - G_n(0)} \cdot \frac{\eta - G_{n+1}(s)}{\eta - G_{n+1}(0)} \\ &= h(G_n(0))H_{n+1}(s). \end{aligned}$$

As $n \to \infty$, $G_n(0) \uparrow \eta$ and so

$$h(G_n(0)) \to \lim_{s \uparrow \eta} \frac{\eta - G(s)}{\eta - s} = G'(\eta).$$

Let $n \to \infty$ in (5) to obtain

(6) $$H(G(s)) = G'(\eta)H(s) \quad \text{if} \quad 0 \le s < \eta$$

and (2) follows from (4). ■

(7) Corollary. *If* $\mu \neq 1$, *then* $\sum_j {}_0\pi_j = 1$.
If $\mu = 1$, *then* ${}_0\pi_j = 0$ *for all* j.

Proof. We have that $\mu = 1$ if and only if $G'(\eta) = 1$. If $\mu \neq 1$ then $G'(\eta) \neq 1$ and letting s increase to η in (6) gives $\lim_{s\uparrow\eta} H(s) = 0$; therefore, from (4), $\lim_{s\uparrow 1} G^{\pi}(s) = 1$, or

$$\sum_j {}_0\pi_j = 1.$$

If $\mu = 1$ then $G'(\eta) = 1$, and (2) becomes $G^{\pi}(G(s)) = G^{\pi}(s)$. However, $G(s) > s$ for all $s < 1$ and so $G^{\pi}(s) = G^{\pi}(0) = 0$ for all $s < 1$. Thus ${}_0\pi_j = 0$ for all j. ■

So long as $\mu \neq 1$, the distribution of Z_n, conditional on future extinction, converges as $n \to \infty$ to some limit $\{{}_0\pi_j\}$ which is a proper distribution. The so-called 'critical' branching process with $\mu = 1$ is more difficult to study in that, for $j \ge 1$,

$$\begin{aligned} \mathbb{P}(Z_n = j) &\to 0 \quad \text{because extinction is certain,} \\ \mathbb{P}(Z_n = j \mid E_n) &\to 0 \quad \text{because } Z_n \to \infty, \text{conditional on } E_n. \end{aligned}$$

However, it is possible to show, in the spirit of the discussion at the end of Section 5.4, that the distribution of

$$Y_n = \frac{Z_n}{n\sigma^2} \quad \text{where} \quad \sigma^2 = \operatorname{var} Z_1,$$

conditional on E_n, converges as $n \to \infty$.

(8) Theorem. *If* $\mu = 1$ *and* $G''(1) < \infty$ *then* $Y_n = Z_n/(n\sigma^2)$ *satisfies*

$$\mathbb{P}(Y_n \le y \mid E_n) \to 1 - e^{-2y}, \quad \text{as} \quad n \to \infty.$$

Proof. See Athreya and Ney (1972, p. 20). ■

So, if $\mu = 1$, the distribution of Y_n, given E_n, is asymptotically exponential with parameter 2. In this case, the branching process is called *critical*; the cases $\mu < 1$ and $\mu > 1$ are called *subcritical* and *supercritical* respectively. See Athreya and Ney (1972) for further details.

Exercises for Section 6.7

1. Let Z_n be the size of the nth generation of a branching process with $Z_0 = 1$ and $\mathbb{P}(Z_1 = k) = 2^{-k}$ for $k \geq 0$. Show directly that, as $n \to \infty$, $\mathbb{P}(Z_n \leq 2yn \mid Z_n > 0) \to 1 - e^{-2y}$, $y > 0$, in agreement with Theorem (6.7.8).

2. Let Z be a supercritical branching process with $Z_0 = 1$ and family-size generating function G. Assume that the probability η of extinction satisfies $0 < \eta < 1$. Find a way of describing the process Z, *conditioned on its ultimate extinction*.

3. Let Z_n be the size of the nth generation of a branching process with $Z_0 = 1$ and $\mathbb{P}(Z_1 = k) = qp^k$ for $k \geq 0$, where $p + q = 1$ and $p > \frac{1}{2}$. Use your answer to Exercise (2) to show that, if we condition on the ultimate extinction of Z, then the process grows in the manner of a branching process with generation sizes $\widetilde{Z}_n$ satisfying $\widetilde{Z}_0 = 1$ and $\mathbb{P}(\widetilde{Z}_1 = k) = pq^k$ for $k \geq 0$.

4. (a) Show that $\mathbb{E}(X \mid X > 0) \leq \mathbb{E}(X^2)/\mathbb{E}(X)$ for any random variable X taking non-negative values.

(b) Let Z_n be the size of the nth generation of a branching process with $Z_0 = 1$ and $\mathbb{P}(Z_1 = k) = qp^k$ for $k \geq 0$, where $p > \frac{1}{2}$. Use part (a) to show that $\mathbb{E}(Z_n/\mu^n \mid Z_n > 0) \leq 2p/(p - q)$, where $\mu = p/q$.

(c) Show that, in the notation of part (b), $\mathbb{E}(Z_n/\mu^n \mid Z_n > 0) \to p/(p - q)$ as $n \to \infty$.

6.8 Birth processes and the Poisson process

Many processes in nature may change their values at any instant of time rather than at certain specified epochs only. Such a process is a family $\{X(t) : t \geq 0\}$ of random variables indexed by the half-line $[0, \infty)$ and taking values in a state space S. Depending on the underlying random mechanism, X may or may not be a Markov process. Before attempting to study any general theory of continuous-time processes we explore one simple but non-trivial example in detail.

Given the right equipment, we should have no difficulty in observing that the process of emission of particles from a radioactive source seems to behave in a manner which is not totally predictable. If we switch on our Geiger counter at time zero, then the reading $N(t)$ which it shows at a later time t is the outcome of some random process. This process $\{N(t) : t \geq 0\}$ has certain obvious properties, such as:

(a) $N(0) = 0$, and $N(t) \in \{0, 1, 2, \dots\}$,

(b) if $s < t$ then $N(s) \leq N(t)$,

but it is not so easy to specify more detailed properties. We might use the following description. In the time interval $(t, t + h)$ there may or may not be some emissions. If h is small then the likelihood of an emission is roughly proportional to h; it is not very likely that two or more emissions will occur in a small interval. More formally, we make the following definition of a Poisson process†.

†Developed separately but contemporaneously by Erlang, Bateman, and Campbell in 1909, and named after Poisson by Feller before 1940.

(1) Definition. A **Poisson process with intensity** λ is a process $N = \{N(t) : t \geq 0\}$ taking values in $S = \{0, 1, 2, \dots\}$ such that:

(a) $N(0) = 0$; if $s < t$ then $N(s) \leq N(t)$,

(b) $$\mathbb{P}\big(N(t+h) = n+m \mid N(t) = n\big) = \begin{cases} \lambda h + o(h) & \text{if } m = 1, \\ o(h) & \text{if } m > 1, \\ 1 - \lambda h + o(h) & \text{if } m = 0, \end{cases}$$

(c) if $s < t$, the number $N(t) - N(s)$ of emissions in the interval $(s, t]$ is independent of the times of emissions during $[0, s]$.

We speak of $N(t)$ as the number of 'arrivals' or 'occurrences' or 'events', or in this example 'emissions', of the process by time t. The process N is called a 'counting process' and is one of the simplest examples of continuous-time Markov chains. We shall consider the general theory of such processes in the next section; here we study special properties of Poisson processes and their generalizations.

We are interested first in the distribution of $N(t)$.

(2) Theorem. *$N(t)$ has the Poisson distribution with parameter λt; that is to say,*

$$\mathbb{P}\big(N(t) = j\big) = \frac{(\lambda t)^j}{j!} e^{-\lambda t}, \qquad j = 0, 1, 2, \dots .$$

Proof. Condition $N(t+h)$ on $N(t)$ to obtain

$$\begin{aligned} \mathbb{P}\big(N(t+h) = j\big) &= \sum_i \mathbb{P}\big(N(t) = i\big)\mathbb{P}\big(N(t+h) = j \mid N(t) = i\big) \\ &= \sum_i \mathbb{P}\big(N(t) = i\big)\mathbb{P}\big((j-i) \text{ arrivals in } (t, t+h]\big) \\ &= \mathbb{P}\big(N(t) = j-1\big)\mathbb{P}(\text{one arrival}) + \mathbb{P}\big(N(t) = j\big)\mathbb{P}(\text{no arrivals}) + o(h). \end{aligned}$$

Thus $p_j(t) = \mathbb{P}(N(t) = j)$ satisfies

$$\begin{aligned} p_j(t+h) &= \lambda h p_{j-1}(t) + (1 - \lambda h)p_j(t) + o(h) \quad \text{if} \quad j \neq 0, \\ p_0(t+h) &= (1 - \lambda h)p_0(t) + o(h). \end{aligned}$$

Subtract $p_j(t)$ from each side of the first of these equations, divide by h, and let $h \downarrow 0$ to obtain

(3) $$p_j'(t) = \lambda p_{j-1}(t) - \lambda p_j(t) \quad \text{if} \quad j \neq 0;$$

likewise

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

The boundary condition is

(5) $$p_j(0) = \delta_{j0} = \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{if } j \neq 0. \end{cases}$$

Equations (3) and (4) form a collection of differential–difference equations for the $p_j(t)$. Here are two methods of solution, both of which have applications elsewhere.

Method A. Induction. Solve (4) subject to the condition $p_0(0) = 1$ to obtain $p_0(t) = e^{-\lambda t}$. Substitute this into (3) with $j = 1$ to obtain $p_1(t) = \lambda t e^{-\lambda t}$ and iterate, to obtain by induction that

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

Method B. Generating functions. Define the generating function

$$G(s, t) = \sum_{j=0}^{\infty} p_j(t) s^j = \mathbb{E}(s^{N(t)}).$$

Multiply (3) by s^j and sum over j to obtain

$$\frac{\partial G}{\partial t} = \lambda(s-1)G$$

with the boundary condition $G(s, 0) = 1$. The solution is, as required,

$$G(s, t) = e^{\lambda(s-1)t} = e^{-\lambda t} \sum_{j=0}^{\infty} \frac{(\lambda t)^j}{j!} s^j. \tag{6}$$ ■

This result seems very like the account in Example (3.5.4) that the binomial bin(n, p) distribution approaches the Poisson distribution if $n \to \infty$ and $np \to \lambda$. Why is this no coincidence?

There is an important alternative and equivalent formulation of a Poisson process which provides much insight into its behaviour. Let $T_0, T_1, \dots$ be given by

$$T_0 = 0, \quad T_n = \inf\{t : N(t) = n\}. \tag{7}$$

Then T_n is the time of the nth arrival. The *interarrival times* are the random variables $X_1, X_2, \dots$ given by

$$X_n = T_n - T_{n-1}. \tag{8}$$

From knowledge of N, we can find the values of $X_1, X_2, \dots$ by (7) and (8). Conversely, we can reconstruct N from a knowledge of the X_i by

$$T_n = \sum_{1}^{n} X_i, \quad N(t) = \max\{n : T_n \le t\}. \tag{9}$$

Figure 6.1 is an illustration of this.

(10) Theorem. *The random variables $X_1, X_2, \dots$ are independent, each having the exponential distribution with parameter λ.*

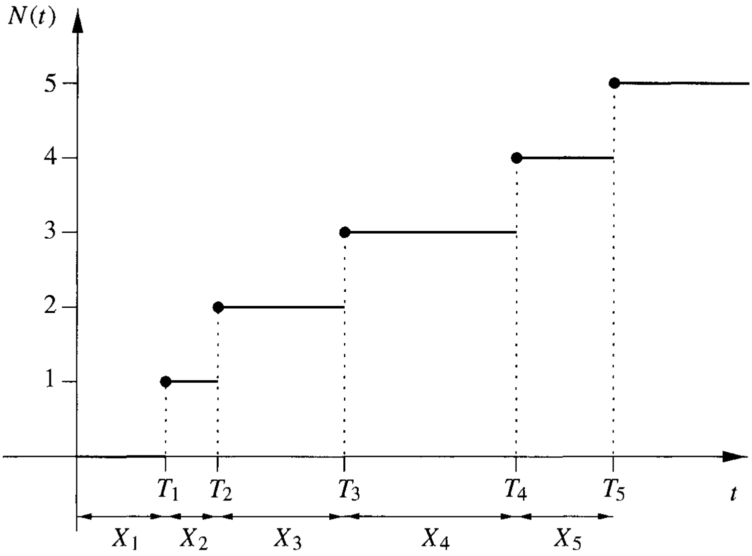


Figure 6.1. A typical realization of a Poisson process $N(t)$.

There is an important generalization of this result to arbitrary continuous-time Markov chains with countable state space. We shall investigate this in the next section.

Proof. First consider X_1:

$$\mathbb{P}(X_1 > t) = \mathbb{P}\big(N(t) = 0\big) = e^{-\lambda t}$$

and so X_1 is exponentially distributed. Now, conditional on X_1,

$$\mathbb{P}(X_2 > t \mid X_1 = t_1) = \mathbb{P}\big(\text{no arrival in } (t_1, t_1 + t] \,\big|\, X_1 = t_1\big).$$

The event $\{X_1 = t_1\}$ relates to arrivals during the time interval $[0, t_1]$, whereas the event $\{\text{no arrival in } (t_1, t_1 + t]\}$ relates to arrivals after time t_1. These events are independent, by (1c), and therefore

$$\mathbb{P}\big(X_2 > t \,\big|\, X_1 = t_1\big) = \mathbb{P}\big(\text{no arrival in } (t_1, t_1 + t]\big) = e^{-\lambda t}.$$

Thus X_2 is independent of X_1, and has the same distribution. Similarly,

$$\mathbb{P}\big(X_{n+1} > t \,\big|\, X_1 = t_1, \dots, X_n = t_n\big) = \mathbb{P}\big(\text{no arrival in } (T, T + t]\big)$$

where $T = t_1 + t_2 + \cdots + t_n$, and the claim of the theorem follows by induction on n. ■

It is not difficult to see that the process N, constructed by (9) from a sequence $X_1, X_2, \dots$, is a Poisson process if and only if the X_i are independent identically distributed exponential variables (*exercise*: use the lack-of-memory property of Problem (4.14.5)). If the X_i form such a sequence, it is a simple matter to deduce the distribution of $N(t)$ directly, as follows. In this case, $T_n = \sum_1^n X_i$ is $\Gamma(\lambda, n)$ and $N(t)$ is specified by the useful remark that

$$N(t) \geq j \quad \text{if and only if} \quad T_j \leq t.$$

Therefore

$$\mathbb{P}\big(N(t) = j\big) = \mathbb{P}(T_j \le t < T_{j+1}) = \mathbb{P}(T_j \le t) - \mathbb{P}(T_{j+1} \le t)$$
$$= \frac{(\lambda t)^j}{j!} e^{-\lambda t}$$

using the properties of gamma variables and integration by parts (see Problem (4.14.11c)).

The Poisson process is a very satisfactory model for radioactive emissions from a sample of uranium-235 since this isotope has a half-life of 7×10^8 years and decays fairly slowly. However, for a newly produced sample of strontium-92, which has a half-life of 2.7 hours, we need a more sophisticated process which takes into account the retardation in decay rate over short time intervals. We might suppose that the rate λ at which emissions are detected depends on the number detected already.

(11) Definition. A **birth process with intensities** $\lambda_0, \lambda_1, \dots$ is a process $\{N(t) : t \ge 0\}$ taking values in $S = \{0, 1, 2, \dots\}$ such that:

(a) $N(0) \ge 0$; if $s < t$ then $N(s) \le N(t)$,

(b) $\mathbb{P}\big(N(t+h) = n + m \mid N(t) = n\big) = \begin{cases} \lambda_n h + o(h) & \text{if } m = 1, \\ o(h) & \text{if } m > 1, \\ 1 - \lambda_n h + o(h) & \text{if } m = 0, \end{cases}$

(c) if $s < t$ then, conditional on the value of $N(s)$, the increment $N(t) - N(s)$ is independent of all arrivals prior to s.

Here are some interesting special cases.

(a) **Poisson process.** $\lambda_n = \lambda$ for all n. ●

(b) **Simple birth.** $\lambda_n = n\lambda$. This models the growth of a population in which living individuals give birth independently of one another, each giving birth to a new individual with probability $\lambda h + o(h)$ in the interval $(t, t+h)$. No individuals may die. The number M of births in the interval $(t, t+h)$ satisfies:

$$\mathbb{P}\big(M = m \mid N(t) = n\big) = \binom{n}{m} (\lambda h)^m (1 - \lambda h)^{n-m} + o(h)$$
$$= \begin{cases} 1 - n\lambda h + o(h) & \text{if } m = 0, \\ n\lambda h + o(h) & \text{if } m = 1, \\ o(h) & \text{if } m > 1. \end{cases}$$ ●

(c) **Simple birth with immigration.** $\lambda_n = n\lambda + \nu$. This models a simple birth process which experiences immigration at constant rate ν from elsewhere. ●

Suppose that N is a birth process with positive intensities $\lambda_0, \lambda_1, \dots$. Let us proceed as for the Poisson process. Define the transition probabilities

$$p_{ij}(t) = \mathbb{P}\big(N(s+t) = j \mid N(s) = i\big) = \mathbb{P}\big(N(t) = j \mid N(0) = i\big);$$

now condition $N(t+h)$ on $N(t)$ and let $h \downarrow 0$ as we did for (3) and (4), to obtain the so-called

(12) Forward system of equations: $p'_{ij}(t) = \lambda_{j-1}p_{i,j-1}(t) - \lambda_j p_{ij}(t)$ for $j \geq i$,

with the convention that $\lambda_{-1} = 0$, and the boundary condition $p_{ij}(0) = \delta_{ij}$. Alternatively we might condition $N(t+h)$ on $N(h)$ and let $h \downarrow 0$ to obtain the so-called

(13) Backward system of equations: $p'_{ij}(t) = \lambda_i p_{i+1,j}(t) - \lambda_i p_{ij}(t)$ for $j \geq i$,

with the boundary condition $p_{ij}(0) = \delta_{ij}$.

Can we solve these equations as we did for the Poisson process?

(14) Theorem. *The forward system has a unique solution, which satisfies the backward system.*

Proof. Note first that $p_{ij}(t) = 0$ if $j < i$. Solve the forward equation with $j = i$ to obtain $p_{ii}(t) = e^{-\lambda_i t}$. Substitute into the forward equation with $j = i+1$ to find $p_{i,i+1}(t)$. Continue this operation to deduce that the forward system has a unique solution. To obtain more information about this solution, define the Laplace transforms†

$$\widehat{p}_{ij}(\theta) = \int_0^\infty e^{-\theta t} p_{ij}(t)\, dt.$$

Transform the forward system to obtain

$$(\theta + \lambda_j)\widehat{p}_{ij}(\theta) = \delta_{ij} + \lambda_{j-1}\widehat{p}_{i,j-1}(\theta);$$

this is a difference equation which is readily solved to obtain

(15)
$$\widehat{p}_{ij}(\theta) = \frac{1}{\lambda_j}\frac{\lambda_i}{\theta+\lambda_i}\frac{\lambda_{i+1}}{\theta+\lambda_{i+1}}\cdots\frac{\lambda_j}{\theta+\lambda_j} \quad \text{for} \quad j \geq i.$$

This determines $p_{ij}(t)$ uniquely by the inversion theorem for Laplace transforms.

To see that this solution satisfies the backward system, transform this system similarly to obtain that any solution $\pi_{ij}(t)$ to the backward equation, with Laplace transform

$$\widehat{\pi}_{ij}(\theta) = \int_0^\infty e^{-\theta t}\pi_{ij}(t)\, dt,$$

satisfies

$$(\theta + \lambda_i)\widehat{\pi}_{ij}(\theta) = \delta_{ij} + \lambda_i\widehat{\pi}_{i+1,j}(\theta).$$

The $\widehat{p}_{ij}$, given by (15), satisfy this equation, and so the p_{ij} satisfy the backward system. ■

We have not been able to show that the backward system has a unique solution, for the very good reason that this may not be true. All we can show is that it has a minimal solution.

(16) Theorem. *If $\{p_{ij}(t)\}$ is the unique solution of the forward system, then any solution $\{\pi_{ij}(t)\}$ of the backward system satisfies $p_{ij}(t) \leq \pi_{ij}(t)$ for all i, j, t.*

Proof. See Feller (1968, pp. 475–477). ■

†See Section F of Appendix I for some properties of Laplace transforms.

There may seem something wrong here, because the condition

$$\sum_j p_{ij}(t) = 1 \tag{17}$$

in conjunction with the result of (16) would constrain $\{p_{ij}(t)\}$ to be the *unique* solution of the backward system which is a proper distribution. The point is that (17) may fail to hold. A problem arises when the birth rates λ_n increase sufficiently quickly with n that the process N may pass through all (finite) states in bounded time, and we say that *explosion* occurs if this happens with a strictly positive probability. Let $T_\infty = \lim_{n\to\infty} T_n$ be the limit of the arrival times of the process.

(18) Definition. We call the process N **honest** if $\mathbb{P}(T_\infty = \infty) = 1$ for all t, and **dishonest** otherwise.

Equation (17) is equivalent to $\mathbb{P}(T_\infty > t) = 1$, whence (17) holds for all t if and only if N is honest.

(19) Theorem. *The process N is honest if and only if $\sum_n \lambda_n^{-1} = \infty$.*

This beautiful theorem asserts that if the birth rates are small enough then $N(t)$ is almost surely finite, but if they are sufficiently large that $\sum \lambda_n^{-1}$ converges then births occur so frequently that there is positive probability of infinitely many births occurring in a finite interval of time; thus $N(t)$ may take the value $+\infty$ instead of a non-negative integer. Think of the deficit $1 - \sum_j p_{ij}(t)$ as the probability $\mathbb{P}(T_\infty \le t)$ of escaping to infinity by time t, starting from i.

Theorem (19) is a immediate consequence of the following lemma.

(20) Lemma. *Let $X_1, X_2, \dots$ be independent random variables, X_n having the exponential distribution with parameter λ_{n-1}, and let $T_\infty = \sum_n X_n$. We have that*

$$\mathbb{P}(T_\infty < \infty) = \begin{cases} 0 & \text{if } \sum_n \lambda_n^{-1} = \infty, \\ 1 & \text{if } \sum_n \lambda_n^{-1} < \infty. \end{cases}$$

Proof. We have by equation (5.6.13) that

$$\mathbb{E}(T_\infty) = \mathbb{E}\left(\sum_{n=1}^{\infty} X_n\right) = \sum_{n=1}^{\infty} \frac{1}{\lambda_{n-1}}.$$

If $\sum_n \lambda_n^{-1} < \infty$ then $\mathbb{E}(T_\infty) < \infty$, whence $\mathbb{P}(T_\infty = \infty) = 0$.

In order to study the atom of T_∞ at ∞ we work with the bounded random variable e^{-T_∞}, defined as the limit as $n \to \infty$ of e^{-T_n}. By monotone convergence (5.6.12),

$$\begin{aligned} \mathbb{E}(e^{-T_\infty}) &= \mathbb{E}\left(\prod_{n=1}^{\infty} e^{-X_n}\right) = \lim_{N\to\infty} \mathbb{E}\left(\prod_{n=1}^{N} e^{-X_n}\right) \\ &= \lim_{N\to\infty} \prod_{n=1}^{N} \mathbb{E}(e^{-X_n}) \qquad \text{by independence} \\ &= \lim_{N\to\infty} \prod_{n=1}^{N} \frac{1}{1+\lambda_{n-1}^{-1}} = \left\{\prod_{n=1}^{\infty}(1+\lambda_{n-1}^{-1})\right\}^{-1}. \end{aligned}$$

The last product† equals ∞ if $\sum_n \lambda_n^{-1} = \infty$, implying in turn that $\mathbb{E}(e^{-T_\infty}) = 0$. However, $e^{-T_\infty} \geq 0$, and therefore $\mathbb{P}(T_\infty = \infty) = \mathbb{P}(e^{-T_\infty} = 0) = 1$ as required. ■

In summary, we have considered several random processes, indexed by continuous time, which model phenomena occurring in nature. However, certain dangers arise unless we take care in the construction of such processes. They may even find a way to the so-called 'boundary' of the state space by exploding in finite time.

We terminate this section with a brief discussion of the Markov property for birth processes. Recall that a sequence $X = \{X_n : n \geq 0\}$ is said to satisfy the Markov property if, conditional on the event $\{X_n = i\}$, events relating to the collection $\{X_m : m > n\}$ are independent of events relating to $\{X_m : m < n\}$. Birth processes have a similar property. Let N be a birth process and let T be a fixed time. Conditional on the event $\{N(T) = i\}$, the evolution of the process subsequent to time T is independent of that prior to T; this is an immediate consequence of (11c), and is called the 'weak Markov property'. It is often desirable to make use of a stronger property, in which T is allowed to be a *random variable* rather than merely a constant. On the other hand, such a conclusion cannot be valid for all random T, since if T 'looks into the future' as well as the past, then information about the past may generally be relevant to the future (*exercise*: find a random variable T for which the desired conclusion is false). A useful class of random times are those whose values depend only on the past, and here is a formal definition. We call the random time T a *stopping time* for the process N if, for all $t \geq 0$, the indicator function of the event $\{T \leq t\}$ is a function of the values $\{N(s) : s \leq t\}$ of the process up to time t; that is to say, we require that it be decidable whether or not T has occurred by time t knowing only the values of the process up to time t. Examples of stopping times are the times $T_1, T_2, \dots$ of arrivals; examples of times which are not stopping times are $T_4 - 2$, $\frac{1}{2}(T_1 + T_2)$, and other random variables which 'look into the future'.

(21) Theorem. Strong Markov property. *Let N be a birth process and let T be a stopping time for N. Let A be an event which depends on $\{N(s) : s > T\}$ and B be an event which depends on $\{N(s) : s \leq T\}$. Then*

$$\mathbb{P}\big(A \mid N(T) = i,\ B\big) = \mathbb{P}\big(A \mid N(T) = i\big) \quad \textit{for all } i. \tag{22}$$

Proof. The following argument may be made rigorous. The event B contains information about the process N prior to T; the 'worst' such event is one which tells everything. Assume then that B is a complete description of $\{N(s) : s \leq T\}$ (problems of measurability may in general arise here, but these are not serious in this case since birth processes have only countably many arrivals). Since B is a complete description, knowledge of B carries with it knowledge of the value of the stopping time T, which we write as $T = T(B)$. Therefore

$$\mathbb{P}\big(A \mid N(T) = i,\ B\big) = \mathbb{P}\big(A \mid N(T) = i,\ B,\ T = T(B)\big).$$

The event $\{N(T) = i\} \cap B \cap \{T = T(B)\}$ specifies: (i) the value of T, (ii) the value of $N(T)$, and (iii) the history of the process up to time T; it is by virtue of the fact that T is a stopping time that this event is defined in terms of $\{N(s) : s \leq T(B)\}$. By the *weak* Markov property, since T is constant on this event, we may discount information in (iii), so that

$$\mathbb{P}\big(A \mid N(T) = i,\ B\big) = \mathbb{P}\big(A \mid N(T) = i,\ T = T(B)\big).$$

†See Subsection (8) of Appendix I for some notes about infinite products.

Now, the process is temporally homogeneous, and A is defined in terms of $\{N(s) : s > T\}$; it follows that the (conditional) probability of A depends only on the value of $N(T)$, which is to say that

$$\mathbb{P}\big(A \mid N(T) = i,\ T = T(B)\big) = \mathbb{P}\big(A \mid N(T) = i\big)$$

and (22) follows.

To obtain (22) for more general events B than that given above requires a small spot of measure theory. For those readers who want to see this, we note that, for general B,

$$\begin{aligned}\mathbb{P}\big(A \mid N(T) = i,\ B\big) &= \mathbb{E}\big(I_A \mid N(T) = i,\ B\big)\\ &= \mathbb{E}\Big\{\mathbb{E}\big(I_A \mid N(T) = i,\ B,\ H\big)\,\Big|\, N(T) = i,\ B\Big\}\end{aligned}$$

where $H = \{N(s) : s \le T\}$. The inner expectation equals $\mathbb{P}(A \mid N(T) = i)$, by the argument above, and the claim follows. ∎

We used two properties of birth processes in our proof of the strong Markov property: temporal homogeneity and the weak Markov property. The strong Markov property plays an important role in the study of continuous-time Markov chains and processes, and we shall encounter it in a more general form later. When applied to a birth process N, it implies that the new process N', defined by $N'(t) = N(t + T) - N(T), t \ge 0$, conditional on $\{N(T) = i\}$ is also a birth process, whenever T is a stopping time for N; it is easily seen that this new birth process has intensities $\lambda_i, \lambda_{i+1}, \dots$. In the case of the Poisson process, we have that $N'(t) = N(t + T) - N(T)$ is a Poisson process also.

(23) Example. A Poisson process N is said to have 'stationary independent increments', since: (a) the distribution of $N(t) - N(s)$ depends only on $t - s$, and (b) the increments $\{N(t_i) - N(s_i) : i = 1, 2, \dots, n\}$ are independent if $s_1 \le t_1 \le s_2 \le t_2 \le \cdots \le t_n$. this property is nearly a characterization of the Poisson process. Suppose that $M = \{M(t) : t \ge 0\}$ is a non-decreasing right-continuous integer-valued process with $M(0) = 0$, having stationary independent increments, and with the extra property that M has only jump discontinuities of size 1. Note first that, for $u, v \ge 0$,

$$\mathbb{E}M(u + v) = \mathbb{E}M(u) + \mathbb{E}\big[M(u + v) - M(u)\big] = \mathbb{E}M(u) + \mathbb{E}M(v)$$

by the assumption of stationary increments. Now $\mathbb{E}M(u)$ is non-decreasing in u, so that there exists λ such that

(23) $$\mathbb{E}M(u) = \lambda u, \qquad u \ge 0.$$

Let $T = \sup\{t : M(t) = 0\}$ be the time of the first jump of M. We have from the right-continuity of M that $M(T) = 1$ (almost surely), so that T is a stopping time for M. Now

(24) $$\mathbb{E}M(s) = \mathbb{E}\big\{\mathbb{E}(M(s) \mid T)\big\}.$$

Certainly $\mathbb{E}(M(s) \mid T) = 0$ if $s < T$, and for $s \ge t$

$$\begin{aligned}\mathbb{E}\big(M(s) \mid T = t\big) &= \mathbb{E}\big(M(t) \mid T = t\big) + \mathbb{E}\big(M(s) - M(t) \mid T = t\big)\\ &= 1 + \mathbb{E}\Big(M(s) - M(t)\,\Big|\, M(t) = 1,\ M(u) = 0 \text{ for } u < v\Big)\\ &= 1 + \mathbb{E}M(s - t)\end{aligned}$$

by the assumption of stationary independent increments. We substitute this into (24) to obtain

$$\mathbb{E}M(s) = \int_0^s \big[1 + \mathbb{E}M(s-t)\big]\, dF(t)$$

where F is the distribution function of T. Now $\mathbb{E}M(s) = \lambda s$ for all s, so that

$$\lambda s = F(s) + \lambda \int_0^s (s-t)\, dF(t), \tag{25}$$

an integral equation for the unknown function F. One of the standard ways of solving such an equation is to use Laplace transforms. We leave it as an *exercise* to deduce from (25) that $F(t) = 1 - e^{-\lambda t}$, $t \geq 0$, so that T has the exponential distribution. An argument similar to that used for Theorem (10) now shows that the 'inter-jump' times of M are independent and have the exponential distribution. Hence M is a Poisson process with intensity λ. ●

Exercises for Section 6.8

1. Superposition. Flies and wasps land on your dinner plate in the manner of independent Poisson processes with respective intensities λ and μ. Show that the arrivals of flying objects form a Poisson process with intensity $\lambda + \mu$.

2. Thinning. Insects land in the soup in the manner of a Poisson process with intensity λ, and each such insect is green with probability p, independently of the colours of all other insects. Show that the arrivals of green insects form a Poisson process with intensity λp.

3. Let T_n be the time of the nth arrival in a Poisson process N with intensity λ, and define the excess lifetime process $E(t) = T_{N(t)+1} - t$, being the time one must wait subsequent to t before the next arrival. Show by conditioning on T_1 that

$$\mathbb{P}\big(E(t) > x\big) = e^{-\lambda(t+x)} + \int_0^t \mathbb{P}\big(E(t-u) > x\big)\lambda e^{-\lambda u}\, du.$$

Solve this integral equation in order to find the distribution function of $E(t)$. Explain your conclusion.

4. Let B be a simple birth process (6.8.11b) with $B(0) = I$; the birth rates are $\lambda_n = n\lambda$. Write down the forward system of equations for the process and deduce that

$$\mathbb{P}\big(B(t) = k\big) = \binom{k-1}{I-1} e^{-I\lambda t}\big(1 - e^{-\lambda t}\big)^{k-I}, \qquad k \geq I.$$

Show also that $\mathbb{E}(B(t)) = Ie^{\lambda t}$ and $\mathrm{var}(B(t)) = Ie^{2\lambda t}(1 - e^{-\lambda t})$.

5. Let B be a process of simple birth with immigration (6.8.11c) with parameters λ and ν, and with $B(0) = 0$; the birth rates are $\lambda_n = n\lambda + \nu$. Write down the sequence of differential–difference equations for $p_n(t) = \mathbb{P}(B(t) = n)$. Without solving these equations, use them to show that $m(t) = \mathbb{E}(B(t))$ satisfies $m'(t) = \lambda m(t) + \nu$, and solve for $m(t)$.

6. Let N be a birth process with intensities $\lambda_0, \lambda_1, \ldots$, and let $N(0) = 0$. Show that $p_n(t) = \mathbb{P}(N(t) = n)$ is given by

$$p_n(t) = \frac{1}{\lambda_n} \sum_{i=0}^{n} \lambda_i e^{-\lambda_i t} \prod_{\substack{j=0 \\ j \neq i}}^{n} \frac{\lambda_j}{\lambda_j - \lambda_i}$$

provided that $\lambda_i \neq \lambda_j$ whenever $i \neq j$.

7. Suppose that the general birth process of the previous exercise is such that $\sum_n \lambda_n^{-1} < \infty$. Show that $\lambda_n p_n(t) \to f(t)$ as $n \to \infty$ where f is the density function of the random variable $T = \sup\{t : N(t) < \infty\}$. Deduce that $\mathbb{E}(N(t) \mid N(t) < \infty)$ is finite or infinite depending on the convergence or divergence of $\sum_n n\lambda_n^{-1}$.

Find the Laplace transform of f in closed form for the case when $\lambda_n = (n + \frac{1}{2})^2$, and deduce an expression for f.

6.9 Continuous-time Markov chains

Let $X = \{X(t) : t \geq 0\}$ be a family of random variables taking values in some countable state space S and indexed by the half-line $[0, \infty)$. As before, we shall assume that S is a subset of the integers. The process X is called a (continuous-time) *Markov chain* if it satisfies the following condition.

(1) Definition. The process X satisfies the **Markov property** if

$$\mathbb{P}\big(X(t_n) = j \,\big|\, X(t_1) = i_1, \ldots, X(t_{n-1}) = i_{n-1}\big) = \mathbb{P}\big(X(t_n) = j \,\big|\, X(t_{n-1}) = i_{n-1}\big)$$

for all $j, i_1, \ldots, i_{n-1} \in S$ and any sequence $t_1 < t_2 < \cdots < t_n$ of times.

The evolution of continuous-time Markov chains can be described in very much the same terms as those used for discrete-time processes. Various difficulties may arise in the analysis, especially when S is infinite. The way out of these difficulties is too difficult to describe in detail here, and the reader should look elsewhere (see Chung 1960, or Freedman 1971 for example). The general scheme is as follows. For discrete-time processes we wrote the n-step transition probabilities in matrix form and expressed them in terms of the one-step matrix $\mathbf{P}$. In continuous time there is no exact analogue of $\mathbf{P}$ since there is no implicit unit length of time. The infinitesimal calculus offers one way to plug this gap; we shall see that there exists a matrix $\mathbf{G}$, called the *generator* of the chain, which takes over the role of $\mathbf{P}$. An alternative way of approaching the question of continuous time is to consider the imbedded discrete-time process obtained by listing the changes of state of the original process.

First we address the basics.

(2) Definition. The **transition probability** $p_{ij}(s, t)$ is defined to be

$$p_{ij}(s, t) = \mathbb{P}\big(X(t) = j \,\big|\, X(s) = i\big) \quad \text{for} \quad s \leq t.$$

The chain is called **homogeneous** if $p_{ij}(s, t) = p_{ij}(0, t - s)$ for all i, j, s, t, and we write $p_{ij}(t - s)$ for $p_{ij}(s, t)$.

Henceforth we suppose that X is a homogeneous chain, and we write $\mathbf{P}_t$ for the $|S| \times |S|$ matrix with entries $p_{ij}(t)$. The family $\{\mathbf{P}_t : t \geq 0\}$ is called the *transition semigroup* of the chain.

(3) Theorem. *The family $\{\mathbf{P}_t : t \geq 0\}$ is a* stochastic semigroup; *that is, it satisfies the following*:

(a) $\mathbf{P}_0 = \mathbf{I}$, *the identity matrix*,

(b) $\mathbf{P}_t$ *is stochastic, that is* $\mathbf{P}_t$ *has non-negative entries and row sums* 1,

(c) *the* Chapman–Kolmogorov equations, $\mathbf{P}_{s+t} = \mathbf{P}_s\mathbf{P}_t$ *if* $s, t \geq 0$.

Proof. Part (a) is obvious.

(b) With $\mathbf{1}$ a row vector of ones, we have that

$$(\mathbf{P}_t\mathbf{1}')_i = \sum_j p_{ij}(t) = \mathbb{P}\Big(\bigcup_j \{X(t) = j\} \,\Big|\, X(0) = i\Big) = 1.$$

(c) Using the Markov property,

$$\begin{aligned} p_{ij}(s+t) &= \mathbb{P}\big(X(s+t) = j \mid X(0) = i\big) \\ &= \sum_k \mathbb{P}\big(X(s+t) = j \mid X(s) = k,\ X(0) = i\big)\mathbb{P}\big(X(s) = k \mid X(0) = i\big) \\ &= \sum_k p_{ik}(s)p_{kj}(t) \qquad \text{as for Theorem (6.1.7).} \end{aligned}$$

■

As before, the evolution of $X(t)$ is specified by the stochastic semigroup $\{\mathbf{P}_t\}$ and the distribution of $X(0)$. Most questions about X can be rephrased in terms of these matrices and their properties.

Many readers will not be very concerned with the general theory of these processes, but will be much more interested in specific examples and their stationary distributions. Therefore, we present only a broad outline of the theory in the remaining part of this section and hope that it is sufficient for most applications. Technical conditions are usually omitted, with the consequence that *some of the statements which follow are false in general*; such statements are marked with an asterisk. Indications of how to fill in the details are given in the next section. We shall always suppose that the transition probabilities are continuous.

(4) **Definition.** The semigroup $\{\mathbf{P}_t\}$ is called **standard** if $\mathbf{P}_t \to \mathbf{I}$ as $t \downarrow 0$, which is to say that $p_{ii}(t) \to 1$ and $p_{ij}(t) \to 0$ for $i \neq j$ as $t \downarrow 0$.

Note that the semigroup is standard if and only if its elements $p_{ij}(t)$ are continuous functions of t. In order to see this, observe that $p_{ij}(t)$ is continuous for all t whenever the semigroup is standard; we just use the Chapman–Kolmogorov equations (3c) (see Problem (6.15.14)). Henceforth we consider only Markov chains with standard semigroups of transition probabilities.

Suppose that the chain is in state $X(t) = i$ at time t. Various things may happen in the small time interval $(t, t+h)$:

(a) nothing may happen, with probability $p_{ii}(h) + o(h)$, the error term taking into account the possibility that the chain moves out of i and back to i in the interval,

(b) the chain may move to a new state j with probability $p_{ij}(h) + o(h)$.

We are assuming here that the probability of two or more transitions in the interval $(t, t+h)$ is $o(h)$; this can be proved. Following (a) and (b), we are interested in the behaviour of $p_{ij}(h)$ for small h; it turns out that $p_{ij}(h)$ is approximately linear in h when h is small. That is, there exist constants $\{g_{ij} : i, j \in S\}$ such that

(5) $$p_{ij}(h) \simeq g_{ij}h \quad \textit{if} \quad i \neq j, \qquad p_{ii}(h) \simeq 1 + g_{ii}h.$$

Clearly $g_{ij} \geq 0$ for $i \neq j$ and $g_{ii} \leq 0$ for all i; the matrix† $\mathbf{G} = (g_{ij})$ is called the

†Some writers use the notation q_{ij} in place of g_{ij}, and term the resulting matrix $\mathbf{Q}$ the 'Q-matrix' of the process.

generator of the chain and takes over the role of the transition matrix $\mathbf{P}$ for discrete-time chains. Combine (5) with (a) and (b) above to find that, starting from $X(t) = i$,

(a) nothing happens during $(t, t+h)$ with probability $1 + g_{ii}h + o(h)$,

(b) the chain jumps to state j ($\neq i$) with probability $g_{ij}h + o(h)$.

One may expect that $\sum_j p_{ij}(t) = 1$, and so

$$1 = \sum_j p_{ij}(h) \simeq 1 + h \sum_j g_{ij}$$

leading to the equation

$$\text{(6*)} \qquad \sum_j g_{ij} = 0 \quad \text{for all } i, \quad \text{or} \quad \mathbf{G1}' = \mathbf{0}',$$

where $\mathbf{1}$ and $\mathbf{0}$ are row vectors of ones and zeros. Treat (6) with care; there are some chains for which it fails to hold.

(7) Example. Birth process (6.8.11). From the definition of this process, it is clear that

$$g_{ii} = -\lambda_i, \quad g_{i,i+1} = \lambda_i, \quad g_{ij} = 0 \quad \text{if} \quad j < i \quad \text{or} \quad j > i+1.$$

Thus

$$\mathbf{G} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \cdots \\ 0 & -\lambda_1 & \lambda_1 & 0 & 0 & \cdots \\ 0 & 0 & -\lambda_2 & \lambda_2 & 0 & \cdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}. \qquad \bullet$$

Relation (5) is usually written as

$$\text{(8)} \qquad \lim_{h \downarrow 0} \frac{1}{h}(\mathbf{P}_h - \mathbf{I}) = \mathbf{G},$$

and amounts to saying that $\mathbf{P}_t$ is differentiable at $t = 0$. It is clear that $\mathbf{G}$ can be found from knowledge of the $\mathbf{P}_t$. The converse also is usually true. We argue roughly as follows. Suppose that $X(0) = i$, and condition $X(t+h)$ on $X(t)$ to find that

$$\begin{aligned} p_{ij}(t+h) &= \sum_k p_{ik}(t) p_{kj}(h) \\ &\simeq p_{ij}(t)(1 + g_{jj}h) + \sum_{k:k\neq j} p_{ik}(t) g_{kj} h \quad \text{by (5)} \\ &= p_{ij}(t) + h \sum_k p_{ik}(t) g_{kj}, \end{aligned}$$

giving that

$$\frac{1}{h}\left[p_{ij}(t+h) - p_{ij}(t)\right] \simeq \sum_k p_{ik}(t) g_{kj} = (\mathbf{P}_t\mathbf{G})_{ij}.$$

Let $h \downarrow 0$ to obtain the forward equations. We write $\mathbf{P}_t'$ for the matrix with entries $p_{ij}'(t)$.

(9*) Forward equations. *We have that* $\mathbf{P}_t' = \mathbf{P}_t\mathbf{G}$, *which is to say that*

$$p_{ij}'(t) = \sum_k p_{ik}(t)g_{kj} \quad \textit{for all } i, j \in S.$$

A similar argument, by conditioning $X(t+h)$ on $X(h)$, yields the backward equations.

(10*) Backward equations. *We have that* $\mathbf{P}_t' = \mathbf{G}\mathbf{P}_t$, *which is to say that*

$$p_{ij}'(t) = \sum_k g_{ik}p_{kj}(t) \quad \textit{for all } i, j \in S.$$

These equations are general forms of equations (6.8.12) and (6.8.13) and relate $\{\mathbf{P}_t\}$ to $\mathbf{G}$. Subject to the boundary condition $\mathbf{P}_0 = \mathbf{I}$, they often have a unique solution given by the infinite sum

(11*) $$\mathbf{P}_t = \sum_{n=0}^{\infty} \frac{t^n}{n!}\mathbf{G}^n$$

of powers of matrices (remember that $\mathbf{G}^0 = \mathbf{I}$). Equation (11) is deducible from (9) or (10) in very much the same way as we might show that the function of the single variable $p(t) = e^{gt}$ solves the differential equation $p'(t) = gp(t)$. The representation (11) for $\mathbf{P}_t$ is very useful and is usually written as

(12*) $$\mathbf{P}_t = e^{t\mathbf{G}} \quad \textit{or} \quad \mathbf{P}_t = \exp(t\mathbf{G}).$$

where $e^{\mathbf{A}}$ is the natural abbreviation for $\sum_{n=0}^{\infty}(1/n!)\mathbf{A}^n$ whenever $\mathbf{A}$ is a square matrix.

So, subject to certain technical conditions, a continuous-time chain has a generator $\mathbf{G}$ which specifies the transition probabilities. Several examples of such generators are given in Section 6.11. In the last section we saw that a Poisson process (this is Example (7) with $\lambda_i = \lambda$ for all $i \geq 0$) can be described in terms of its interarrival times; an equivalent remark holds here. Suppose that $X(s) = i$. The future development of $X(s+t)$, for $t \geq 0$, goes roughly as follows. Let $U = \inf\{t \geq 0 : X(s+t) \neq i\}$ be the further time until the chain changes its state; U is called a 'holding time'.

(13*) Claim. *The random variable* U *is exponentially distributed with parameter* $-g_{ii}$.

Thus the exponential distribution plays a central role in the theory of Markov processes.

Sketch proof. The distribution of U has the 'lack of memory' property (see Problem (4.14.5)) because

$$\begin{aligned}\mathbb{P}\big(U > x+y \,\big|\, U > x\big) &= \mathbb{P}\big(U > x+y \,\big|\, X(t+x) = i\big)\\ &= \mathbb{P}(U > y) \quad \text{if} \quad x, y \geq 0\end{aligned}$$

by the Markov property and the homogeneity of the chain. It follows that the distribution function F_U of U satisfies $1 - F_U(x+y) = [1 - F_U(x)][1 - F_U(y)]$, and so $1 - F_U(x) = e^{-\lambda x}$ where $\lambda = F_U'(0) = -g_{ii}$. ■

Therefore, if $X(s) = i$, the chain remains in state i for an exponentially distributed time U, after which it jumps to some other state j.

(14*) Claim. *The probability that the chain jumps to j ($\neq i$) is $-g_{ij}/g_{ii}$.*

Sketch proof. Roughly speaking, suppose that $x < U \leq x + h$ and suppose that the chain jumps only once in $(x, x+h]$. Then

$$\mathbb{P}(\text{jumps to } j \mid \text{it jumps}) \simeq \frac{p_{ij}(h)}{1 - p_{ii}(h)} \to -\frac{g_{ij}}{g_{ii}} \quad \text{as} \quad h \downarrow 0. \qquad \blacksquare$$

(15) Example. Consider a two-state chain X with $S = \{1, 2\}$; X jumps between 1 and 2 as time passes. There are two equivalent ways of describing the chain, depending on whether we specify $\mathbf{G}$ or we specify the holding times:

(a) X has generator $\mathbf{G} = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}$;

(b) if the chain is in state 1 (or 2), then it stays in this state for a length of time which is exponentially distributed with parameter α (or β) before jumping to 2 (or 1).

The forward equations (9), $\mathbf{P}_t' = \mathbf{P}_t\mathbf{G}$, take the form

$$p_{11}'(t) = -\alpha p_{11}(t) + \beta p_{12}(t)$$

and are easily solved to find the transition probabilities of the chain (*exercise*). ●

We move on to the classification of states; this is not such a chore as it was for discrete-time chains. It turns out that for any pair i, j of states

(16) $$\text{either } p_{ij}(t) = 0 \text{ for all } t > 0, \text{ or } p_{ij}(t) > 0 \text{ for all } t > 0,$$

and this leads to a definition of irreducibility.

(17) Definition. The chain is called **irreducible** if for any pair i, j of states we have that $p_{ij}(t) > 0$ for some t.

Any time $t > 0$ will suffice in (17), because of (16). The birth process is *not* irreducible, since it is non-decreasing. See Problem (6.15.15) for a condition for irreducibility in terms of the generator $\mathbf{G}$ of the chain.

As before, the asymptotic behaviour of $X(t)$ for large t is closely bound up with the existence of stationary distributions. Compare their definition with Definition (6.4.1).

(18) Definition. The vector $\boldsymbol{\pi}$ is a **stationary distribution** of the chain if $\pi_j \geq 0$, $\sum_j \pi_j = 1$, and $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}_t$ for all $t \geq 0$.

If $X(0)$ has distribution $\boldsymbol{\mu}^{(0)}$ then the distribution $\boldsymbol{\mu}^{(t)}$ of $X(t)$ is given by

(19) $$\boldsymbol{\mu}^{(t)} = \boldsymbol{\mu}^{(0)}\mathbf{P}_t.$$

If $\boldsymbol{\mu}^{(0)} = \boldsymbol{\mu}$, a stationary distribution, then $X(t)$ has distribution $\boldsymbol{\mu}$ for all t. For discrete-time chains we found stationary distributions by solving the equations $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}$; the corresponding equations $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}_t$ for continuous-time chains may seem complicated but they amount to a simple condition relating $\boldsymbol{\pi}$ and $\mathbf{G}$.

(20*) Claim. *We have that $\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P}_t$ for all t if and only if $\boldsymbol{\pi}\mathbf{G} = \mathbf{0}$.*

Sketch proof. From (11), and remembering that $\mathbf{G}^0 = \mathbf{I}$,

$$\begin{aligned}
\boldsymbol{\pi}\mathbf{G} = \mathbf{0} &\Leftrightarrow \boldsymbol{\pi}\mathbf{G}^n = \mathbf{0} && \text{for all } n \geq 1 \\
&\Leftrightarrow \sum_{n=1}^{\infty} \frac{t^n}{n!} \boldsymbol{\pi}\mathbf{G}^n = \mathbf{0} && \text{for all } t \\
&\Leftrightarrow \boldsymbol{\pi} \sum_{n=0}^{\infty} \frac{t^n}{n!} \mathbf{G}^n = \boldsymbol{\pi} && \text{for all } t \\
&\Leftrightarrow \boldsymbol{\pi}\mathbf{P}_t = \boldsymbol{\pi} && \text{for all } t.
\end{aligned}$$
■

This provides a useful collection of equations which specify stationary distributions, whenever they exist. The ergodic theorem for continuous-time chains is as follows; it holds exactly as stated, and requires no extra conditions.

(21) Theorem. *Let X be irreducible with a standard semigroup $\{\mathbf{P}_t\}$ of transition probabilities.*

(a) *If there exists a stationary distribution $\boldsymbol{\pi}$ then it is unique and*

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

(b) *If there is no stationary distribution then $p_{ij}(t) \to 0$ as $t \to \infty$, for all i and j.*

Sketch proof. Fix $h > 0$ and let $Y_n = X(nh)$. Then $Y = \{Y_n\}$ is an irreducible aperiodic discrete-time Markov chain; Y is called a *skeleton* of X. If Y is non-null persistent, then it has a unique stationary distribution $\boldsymbol{\pi}^h$ and

$$p_{ij}(nh) = \mathbf{P}(Y_n = j \mid Y_0 = i) \to \pi_j^h \qquad \text{as } n \to \infty;$$

otherwise $p_{ij}(nh) \to 0$ as $n \to \infty$. Use this argument for two rational values h_1 and h_2 and observe that the sequences $\{nh_1 : n \geq 0\}$, $\{nh_2 : n \geq 0\}$ have infinitely many points in common to deduce that $\boldsymbol{\pi}^{h_1} = \boldsymbol{\pi}^{h_2}$ in the non-null persistent case. Thus the limit of $p_{ij}(t)$ exists along all sequences $\{nh : n \geq 0\}$ of times, for rational h; now use the continuity of $p_{ij}(t)$ to fill in the gaps. The proof is essentially complete. ■

As noted earlier, an alternative approach to a continuous-time chain X is to concentrate on its changes of state at the times of jumps. Indeed one may extend the discussion leading to (13) and (14) to obtain the following, subject to conditions of regularity not stated here. Let T_n be the time of the nth change in value of the chain X, and set $T_0 = 0$. The values $Z_n = X(T_n+)$ of X immediately after its jumps constitute a discrete-time Markov chain Z with transition matrix $h_{ij} = g_{ij}/g_i$, when $g_i = -g_{ii}$ satisfies $g_i > 0$; if $g_i = 0$, the chain remains forever in state i once it has arrived there for the first time. Furthermore, if $Z_n = j$, the holding time $T_{n+1} - T_n$ has the exponential distribution with parameter g_j. The chain Z is called the *jump chain* of X. There is an important and useful converse to this statement, which illuminates the interplay between X and its jump chain. Given a discrete-time chain Z, one may construct a continuous-time chain X having Z as its jump chain; indeed many such chains X exist. We make this more formal as follows.

Let S be a countable state space, and let $\mathbf{H} = (h_{ij})$ be the transition matrix of a discrete-time Markov chain Z taking values in S. We shall assume that $h_{ii} = 0$ for all $i \in S$; this is not an essential assumption, but recognizes the fact that jumps from any state i to itself will be invisible in continuous time. Let g_i, $i \in S$, be non-negative constants. We define

(22)
$$g_{ij} = \begin{cases} g_i h_{ij} & \text{if } i \neq j, \\ -g_i & \text{if } i = j. \end{cases}$$

We now construct a continuous-time chain X as follows. First, let $X(0) = Z_0$. After a holding time U_0 having the exponential distribution with parameter g_{Z_0}, the process jumps to the state Z_1. After a further holding time U_1 having the exponential distribution with parameter g_{Z_1}, the chain jumps to Z_2, and so on.

We argue more fully as follows. Conditional on the values Z_n of the chain Z, let $U_0, U_1, \dots$ be independent random variables having the respective exponential distributions with parameters $g_{Z_0}, g_{Z_1}, \dots$, and set $T_n = U_0 + U_1 + \dots + U_n$. We now define

(23)
$$X(t) = \begin{cases} Z_n & \text{if } T_n \le t < T_{n+1} \text{ for some } n, \\ \infty & \text{otherwise.} \end{cases}$$

The special state denoted ∞ is introduced in case $T_\infty = \lim_{n\to\infty} T_n$ satisfies $T_\infty \le t < \infty$. The time T_∞ is called the *explosion time* of the chain X, and the chain is said to *explode* if $\mathbb{P}(T_\infty < \infty) > 0$. It may be seen that X is a continuous-time chain on the augmented state space $S \cup \{\infty\}$, and the generator of X, up to the explosion time T_∞, is the matrix $\mathbf{G} = (g_{ij})$. Evidently Z is the jump chain of X. No major difficulty arises in verifying these assertions when S is finite.

The definition of X in (23) is only one of many possibilities, the others imposing different behaviours at times of explosion. The process X in (23) is termed the *minimal* process, since it is 'active' for a minimal interval of time. It is important to have verifiable conditions under which a chain does not explode.

(24) Theorem. *The chain X constructed above does not explode if any of the following three conditions holds:*

(a) *S is finite;*

(b) $\sup_i g_i < \infty$;

(c) *$X(0) = i$ where i is a persistent state for the jump chain Z.*

Proof. First we prove that (b) suffices, noting in advance that (a) implies (b). Suppose that $g_i < \gamma < \infty$ for all i. The nth holding time U_n of the chain has the exponential distribution with parameter g_{Z_n}. If $g_{Z_n} > 0$, it is an easy exercise to show that $V_n = g_{Z_n} U_n$ has the exponential distribution with parameter 1. If $g_{Z_n} = 0$, then $U_n = \infty$ almost surely. Therefore,

$$\gamma T_\infty = \begin{cases} \infty & \text{if } g_{Z_n} = 0 \text{ for some } n, \\ \displaystyle\sum_{n=1}^{\infty} \gamma U_n \ge \sum_{n=1}^{\infty} V_n & \text{otherwise.} \end{cases}$$

It follows by Lemma (6.8.20) that the last sum is almost surely infinite; therefore, explosion does not occur.

Suppose now that (c) holds. If $g_i = 0$, then $X(t) = i$ for all t, and there is nothing to prove. Suppose that $g_i > 0$. Since $Z_0 = i$ and i is persistent for Z, there exists almost surely an infinity of times $N_0 < N_1 < \cdots$ at which Z takes the value i. Now,

$$g_i T_\infty \geq \sum_{i=0}^{\infty} g_i U_{N_i},$$

and we may once again appeal to Lemma (6.8.20). ■

(25) Example. Let Z be a discrete-time chain with transition matrix $\mathbf{H} = (h_{ij})$ satisfying $h_{ii} = 0$ for all $i \in S$, and let N be a Poisson process with intensity λ. We define X by $X(t) = Z_n$ if $T_n \leq t < T_{n+1}$ where T_n is the time of the nth arrival in the Poisson process (and $T_0 = 0$). The process X has transition semigroup $\mathbf{P}_t = (p_{ij}(t))$ given by

$$\begin{aligned} p_{ij}(t) &= \mathbb{P}\big(X(t) = j \bigm| X(0) = i\big) \\ &= \sum_{n=0}^{\infty} \mathbb{P}\big(X(t) = j,\ N(t) = n \bigm| X(0) = i\big) \\ &= \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n!} e^{-\lambda t} \mathbb{P}(Z_n = j \mid Z_0 = i) = \sum_{n=0}^{\infty} e^{-\lambda t} \frac{(\lambda t)^n (\mathbf{H}^n)_{ij}}{n!}. \end{aligned}$$

We note that

$$e^{-\lambda t \mathbf{I}} = \sum_{n=0}^{\infty} \frac{(-\lambda t)^n}{n!} \mathbf{I}^n = \mathbf{I} e^{-\lambda t},$$

whence $\mathbf{P}_t = e^{\lambda t(\mathbf{H}-\mathbf{I})}$. ●

The interplay between a continuous-time chain X and its jump chain Z provides a basic tool for the study of the former. We present just one example of this statement; others may be found in the exercises. We call the state i

(26)
$$\begin{aligned} &\textit{persistent} \text{ for } X \text{ if } \mathbb{P}\big(\text{the set } \{t : X(t) = i\} \text{ is unbounded} \bigm| X(0) = i\big) = 1, \\ &\textit{transient} \text{ for } X \text{ if } \mathbb{P}\big(\text{the set } \{t : X(t) = i\} \text{ is unbounded} \bigm| X(0) = i\big) = 0. \end{aligned}$$

(27) Theorem. *Consider the chain X constructed above.*

(a) *If $g_i = 0$, the state i is persistent for the continuous-time chain X.*

(b) *Assume that $g_i > 0$. State i is persistent for the continuous-time chain X if and only if it is persistent for the jump chain Z. Furthermore, i is persistent if the transition probabilities $p_{ii}(t) = \mathbb{P}(X(t) = i \mid X(0) = i)$ satisfy $\int_0^\infty p_{ii}(t)\,dt = \infty$, and is transient otherwise.*

Proof. It is trivial that i is persistent if $g_i = 0$, since then the chain X remains in the state i once it has first visited it.

Assume $g_i > 0$. If i is transient for the jump chain Z, there exists almost surely a last visit of Z to i; this implies the almost sure boundedness of the set $\{t : X(t) = i\}$, whence i is transient for X. Suppose i is persistent for Z, and $X(0) = i$. It follows from Theorem

(24c) that the chain X does not explode. By the persistence of i, there exists almost surely an infinity of values n with $Z_n = i$. Since there is no explosion, the times T_n of these visits are unbounded, whence i is persistent for X.

Now, the integrand being positive, we may interchange limits to obtain

$$\begin{aligned}\int_0^\infty p_{ii}(t)\,dt &= \int_0^\infty \mathbb{E}\big(I_{\{X(t)=i\}} \mid X(0)=i\big)\,dt \\ &= \mathbb{E}\left[\int_0^\infty I_{\{X(t)=i\}}\,dt \,\middle|\, X(0)=i\right] \\ &= \mathbb{E}\left[\sum_{n=0}^\infty U_n I_{\{Z_n=i\}} \,\middle|\, Z_0 = i\right]\end{aligned}$$

where $\{U_n : n \ge 1\}$ are the holding times of X. The right side equals

$$\sum_{n=0}^\infty \mathbb{E}(U_0 \mid X(0)=i)h_{ii}(n) = \frac{1}{g_i}\sum_{n=0}^\infty h_{ii}(n)$$

where $h_{ii}(n)$ is the appropriate n-step transition probability of Z. By Corollary (6.2.4), the last sum diverges if and only if i is persistent for Z. ■

Exercises for Section 6.9

1. Let $\lambda\mu > 0$ and let X be a Markov chain on $\{1, 2\}$ with generator

$$\mathbf{G} = \begin{pmatrix} -\mu & \mu \\ \lambda & -\lambda \end{pmatrix}.$$

(a) Write down the forward equations and solve them for the transition probabilities $p_{ij}(t)$, $i, j = 1, 2$.

(b) Calculate $\mathbf{G}^n$ and hence find $\sum_{n=0}^\infty (t^n/n!)\mathbf{G}^n$. Compare your answer with that to part (a).

(c) Solve the equation $\boldsymbol{\pi}\mathbf{G} = \mathbf{0}$ in order to find the stationary distribution. Verify that $p_{ij}(t) \to \pi_j$ as $t \to \infty$.

2. As a continuation of the previous exercise, find:

(a) $\mathbb{P}(X(t) = 2 \mid X(0) = 1, X(3t) = 1)$,

(b) $\mathbb{P}(X(t) = 2 \mid X(0) = 1, X(3t) = 1, X(4t) = 1)$.

3. Jobs arrive in a computer queue in the manner of a Poisson process with intensity λ. The central processor handles them one by one in the order of their arrival, and each has an exponentially distributed runtime with parameter μ, the runtimes of different jobs being independent of each other and of the arrival process. Let $X(t)$ be the number of jobs in the system (either running or waiting) at time t, where $X(0) = 0$. Explain why X is a Markov chain, and write down its generator. Show that a stationary distribution exists if and only if $\lambda < \mu$, and find it in this case.

4. Pasta property. Let $X = \{X(t) : t \ge 0\}$ be a Markov chain having stationary distribution $\boldsymbol{\pi}$. We may sample X at the times of a Poisson process: let N be a Poisson process with intensity λ, independent of X, and define $Y_n = X(T_n+)$, the value taken by X immediately after the epoch T_n of the nth arrival of N. Show that $Y = \{Y_n : n \ge 0\}$ is a discrete-time Markov chain with the same stationary distribution as X. (This exemplifies the 'Pasta' property: Poisson arrivals see time averages.)

[The full assumption of the independence of N and X is not necessary for the conclusion. It suffices that $\{N(s) : s \geq t\}$ be independent of $\{X(s) : s \leq t\}$, a property known as 'lack of anticipation'. It is not even necessary that X be Markov; the Pasta property holds for many suitable ergodic processes.]

5. Let X be a continuous-time Markov chain with generator $\mathbf{G}$ satisfying $g_i = -g_{ii} > 0$ for all i. Let $H_A = \inf\{t \geq 0 : X(t) \in A\}$ be the hitting time of the set A of states, and let $\eta_j = \mathbb{P}(H_A < \infty \mid X(0) = j)$ be the chance of ever reaching A from j. By using properties of the jump chain, which you may assume to be well behaved, show that $\sum_k g_{jk}\eta_k = 0$ for $j \notin A$.

6. In continuation of the preceding exercise, let $\mu_j = \mathbb{E}(H_A \mid X(0) = j)$. Show that the vector $\boldsymbol{\mu}$ is the minimal non-negative solution of the equations

$$\mu_j = 0 \quad \text{if } j \in A, \qquad 1 + \sum_{k \in S} g_{jk}\mu_k = 0 \quad \text{if } j \notin A.$$

7. Let X be a continuous-time Markov chain with transition probabilities $p_{ij}(t)$ and define $F_i = \inf\{t > T_1 : X(t) = i\}$ where T_1 is the time of the first jump of X. Show that, if $g_{ii} \neq 0$, then $\mathbb{P}(F_i < \infty \mid X(0) = i) = 1$ if and only if i is persistent.

8. Let X be the simple symmetric random walk on the integers in continuous time, so that

$$p_{i,i+1}(h) = p_{i,i-1}(h) = \tfrac{1}{2}\lambda h + \mathrm{o}(h).$$

Show that the walk is persistent. Let T be the time spent visiting m during an excursion from 0. Find the distribution of T.

9. Let i be a transient state of a continuous-time Markov chain X with $X(0) = i$. Show that the total time spent in state i has an exponential distribution.

10. Let X be an asymmetric simple random walk in continuous time on the non-negative integers with retention at 0, so that

$$p_{ij}(h) = \begin{cases} \lambda h + \mathrm{o}(h) & \text{if } j = i+1,\ i \geq 0, \\ \mu h + \mathrm{o}(h) & \text{if } j = i-1,\ i \geq 1. \end{cases}$$

Suppose that $X(0) = 0$ and $\lambda > \mu$. Show that the total time V_r spent in state r is exponentially distributed with parameter $\lambda - \mu$.

Assume now that $X(0)$ has some general distribution with probability generating function G. Find the expected amount of time spent at 0 in terms of G.

11. Let $X = \{X(t) : t \geq 0\}$ be a non-explosive irreducible Markov chain with generator $\mathbf{G}$ and unique stationary distribution $\boldsymbol{\pi}$. The mean recurrence time μ_k is defined as follows. Suppose $X(0) = k$, and let $U = \inf\{s : X(s) \neq k\}$. Then $\mu_k = \mathbb{E}(\inf\{t > U : X(t) = k\})$. Let $Z = \{Z_n : n \geq 0\}$ be the imbedded 'jump chain' given by $Z_0 = X(0)$ and Z_n is the value of X just after its nth jump.

(a) Show that Z has stationary distribution $\widehat{\boldsymbol{\pi}}$ satisfying

$$\widehat{\pi}_k = \frac{\pi_k g_k}{\sum_i \pi_i g_i},$$

where $g_i = -g_{ii}$, provided $\sum_i \pi_i g_i < \infty$. When is it the case that $\widehat{\boldsymbol{\pi}} = \boldsymbol{\pi}$?

(b) Show that $\pi_i = 1/(\mu_i g_i)$ if $\mu_i < \infty$, and that the mean recurrence time $\widehat{\mu}_k$ of the state k in the jump chain Z satisfies $\widehat{\mu}_k = \mu_k \sum_i \pi_i g_i$ if the last sum is finite.

12. Let Z be an irreducible discrete-time Markov chain on a countably infinite state space S, having transition matrix $\mathbf{H} = (h_{ij})$ satisfying $h_{ii} = 0$ for all states i, and with stationary distribution $\boldsymbol{\nu}$. Construct a continuous-time process X on S for which Z is the imbedded chain, such that X has no stationary distribution.

6.10 Uniform semigroups

This section is not for lay readers and may be omitted; it indicates where some of the difficulties lie in the heuristic discussion of the last section (see Chung (1960) or Freedman (1971) for the proofs of the following results).

Perhaps the most important claim is equation (6.9.5), that $p_{ij}(h)$ is approximately linear in h when h is small.

(1) Theorem. *If $\{P_t\}$ is a standard stochastic semigroup then there exists an $|S| \times |S|$ matrix $\mathbf{G} = (g_{ij})$ such that, as $t \downarrow 0$,*

(a) $p_{ij}(t) = g_{ij}t + \mathrm{o}(t)$ *for* $i \neq j$,

(b) $p_{ii}(t) = 1 + g_{ii}t + \mathrm{o}(t)$.

Also, $0 \leq g_{ij} < \infty$ if $i \neq j$, and $0 \geq g_{ii} \geq -\infty$. The matrix $\mathbf{G}$ is called the generator *of the semigroup $\{P_t\}$.*

Equation (1b) is fairly easy to demonstrate (see Problem (6.15.14)); the proof of (1a) is considerably more difficult. The matrix $\mathbf{G}$ has non-negative entries off the diagonal and non-positive entries (*which may be* $-\infty$) on the diagonal. We normally write

$$\mathbf{G} = \lim_{t \downarrow 0} \frac{1}{t}(\mathbf{P}_t - \mathbf{1}). \tag{2}$$

If S is finite then

$$\mathbf{G1}' = \lim_{t \downarrow 0} \frac{1}{t}(\mathbf{P}_t - \mathbf{I})\mathbf{1}' = \lim_{t \downarrow 0} \frac{1}{t}(\mathbf{P}_t\mathbf{1}' - \mathbf{1}') = \mathbf{0}'$$

from (6.9.3b), and so the row sums of $\mathbf{G}$ equal 0. If S is infinite, all we can assert is that

$$\sum_j g_{ij} \leq 0.$$

In the light of Claim (6.9.13), states i with $g_{ii} = -\infty$ are called *instantaneous*, since the chain leaves them at the same instant that it arrives in them. Otherwise, state i is called *stable* if $0 > g_{ii} > -\infty$ and *absorbing* if $g_{ii} = 0$.

We cannot proceed much further unless we impose a stronger condition on the semigroup $\{\mathbf{P}_t\}$ than that it be standard.

(3) Definition. We call the semigroup $\{\mathbf{P}_t\}$ **uniform** if $\mathbf{P}_t \to \mathbf{1}$ uniformly as $t \downarrow 0$, which is to say that

$$p_{ii}(t) \to 1 \quad \text{as} \quad t \downarrow 0, \quad \text{uniformly in } i \in S. \tag{4}$$

Clearly (4) implies that $p_{ij}(t) \to 0$ for $i \neq j$, since $p_{ij}(t) \leq 1 - p_{ii}(t)$. A uniform semigroup is standard; the converse is not generally true, but holds if S is finite. The uniformity of the semigroup depends upon the sizes of the diagonal elements of its generator $\mathbf{G}$.

(5) Theorem. *The semigroup $\{\mathbf{P}_t\}$ is uniform if and only if* $\sup_i\{-g_{ii}\} < \infty$.

We consider uniform semigroups only for the rest of this section. Here is the main result, which vindicates equations (6.9.9)–(6.9.11).

(6) Theorem. Kolmogorov's equations. *If* $\{\mathbf{P}_t\}$ *is a uniform semigroup with generator* $\mathbf{G}$, *then it is the unique solution to the*:

(7) forward equation: $\mathbf{P}_t' = \mathbf{P}_t\mathbf{G}$,

(8) backward equation: $\mathbf{P}_t' = \mathbf{G}\mathbf{P}_t$,

subject to the boundary condition $\mathbf{P}_0 = \mathbf{I}$. *Furthermore*

$$\mathbf{P}_t = e^{t\mathbf{G}} \quad \textit{and} \quad \mathbf{G}\mathbf{1}' = \mathbf{0}'. \tag{9}$$

The backward equation is more fundamental than the forward equation since it can be derived subject to the condition that $\mathbf{G}\mathbf{1}' = \mathbf{0}'$, which is a weaker condition than that the semigroup be uniform. This remark has some bearing on the discussion of dishonesty in Section 6.8. (Of course, a dishonest birth process is not even a Markov chain in our sense, unless we augment the state space $\{0, 1, 2, \dots\}$ by adding the point $\{\infty\}$.) You can prove (6) yourself. Just use the argument which established equations (6.9.9) and (6.9.10) with an eye to rigour; then show that (9) gives a solution to (7) and (8), and finally prove uniqueness.

Thus uniform semigroups are characterized by their generators; but which matrices are generators of uniform semigroups? Let $\mathcal{M}$ be the collection of $|S| \times |S|$ matrices $\mathbf{A} = (a_{ij})$ for which

$$\|\mathbf{A}\| = \sup_i \sum_{j\in S} |a_{ij}| \quad \text{satisfies} \quad \|A\| < \infty.$$

(10) Theorem. $\mathbf{A} \in \mathcal{M}$ *is the generator of a uniform semigroup* $\mathbf{P}_t = e^{t\mathbf{A}}$ *if and only if*

$$a_{ij} \ge 0 \quad \textit{for} \quad i \ne j, \qquad \textit{and} \qquad \sum_j a_{ij} = 0 \quad \textit{for all } i.$$

Next we discuss irreducibility. Observation (6.9.16) amounts to the following.

(11) Theorem. *If* $\{\mathbf{P}_t\}$ *is standard (but not necessarily uniform) then*:

(a) $p_{ii}(t) > 0$ *for all* $t \ge 0$.

(b) Lévy dichotomy: *If* $i \ne j$, *either* $p_{ij}(t) = 0$ *for all* $t > 0$,
or $p_{ij}(t) > 0$ *for all* $t > 0$.

Partial proof. (a) $\{\mathbf{P}_t\}$ is assumed standard, so $p_{ii}(t) \to 1$ as $t \downarrow 0$. Pick $h > 0$ such that $p_{ii}(s) > 0$ for all $s \le h$. For any real t pick n large enough so that $t \le hn$. By the Chapman–Kolmogorov equations,

$$p_{ii}(t) \ge p_{ii}(t/n)^n > 0 \quad \text{because} \quad t/n \le h.$$

(b) The proof of this is quite difficult, though the method of (a) can easily be adapted to show that if $\alpha = \inf\{t : p_{ij}(t) > 0\}$ then $p_{ij}(t) > 0$ for all $t > \alpha$. The full result asserts that either $\alpha = 0$ or $\alpha = \infty$. ■

(12) Example (6.9.15) revisited. If $\alpha, \beta > 0$ and $S = \{1, 2\}$, then

$$\mathbf{G} = \begin{pmatrix} -\alpha & \alpha \\ \beta & -\beta \end{pmatrix}$$

is the generator of a uniform stochastic semigroup $\{\mathbf{P}_t\}$ given by the following calculation. Diagonalize $\mathbf{G}$ to obtain $\mathbf{G} = \mathbf{B}\mathbf{\Lambda}\mathbf{B}^{-1}$ where

$$\mathbf{B} = \begin{pmatrix} \alpha & 1 \\ -\beta & 1 \end{pmatrix}, \qquad \mathbf{\Lambda} = \begin{pmatrix} -(\alpha+\beta) & 0 \\ 0 & 0 \end{pmatrix}.$$

Therefore

$$\begin{aligned}\mathbf{P}_t &= \sum_{n=0}^{\infty} \frac{t^n}{n!}\mathbf{G}^n = \mathbf{B}\left(\sum_{n=0}^{\infty} \frac{t^n}{n!}\mathbf{\Lambda}^n\right)\mathbf{B}^{-1} \\ &= \mathbf{B}\begin{pmatrix} h(t) & 0 \\ 0 & 1 \end{pmatrix}\mathbf{B}^{-1} \quad \text{since} \quad \mathbf{\Lambda}^0 = \mathbf{I} \\ &= \frac{1}{\alpha+\beta}\begin{pmatrix} \alpha h(t) + \beta & \alpha[1-h(t)] \\ \beta[1-h(t)] & \alpha + \beta h(t) \end{pmatrix}\end{aligned}$$

where $h(t) = e^{-t(\alpha+\beta)}$. Let $t \to \infty$ to obtain

$$\mathbf{P}_t \to \begin{pmatrix} 1-\rho & \rho \\ 1-\rho & \rho \end{pmatrix} \quad \text{where} \quad \rho = \frac{\alpha}{\alpha+\beta}$$

and so

$$\mathbb{P}\big(X(t) = i\big) \to \begin{cases} 1-\rho & \text{if } i = 1, \\ \rho & \text{if } i = 2, \end{cases}$$

irrespective of the initial distribution of $X(0)$. This shows that $\boldsymbol{\pi} = (1-\rho, \rho)$ is the limiting distribution. Check that $\boldsymbol{\pi}\mathbf{G} = \mathbf{0}$. The method of Example (6.9.15) provides an alternative and easier route to these results. ●

(13) Example. Birth process. Recall the birth process of Definition (6.8.11), and suppose that $\lambda_i > 0$ for all i. The process is uniform if and only if $\sup_i\{-g_{ii}\} = \sup_i\{\lambda_i\} < \infty$, and this is a sufficient condition for the forward and backward equations to have unique solutions. We saw in Section 6.8 that the weaker condition $\sum_i \lambda_i^{-1} = \infty$ is necessary and sufficient for this to hold. ●

6.11 Birth–death processes and imbedding

A birth process is a non-decreasing Markov chain for which the probability of moving from state n to state $n+1$ in the time interval $(t, t+h)$ is $\lambda_n h + o(h)$. More realistic continuous-time models for population growth incorporate death also. Suppose then that the number $X(t)$ of individuals alive in some population at time t evolves in the following way:

(a) X is a Markov chain taking values in $\{0, 1, 2, \dots\}$,

(b) the infinitesimal transition probabilities are given by

$$\text{(1)} \qquad \mathbb{P}\big(X(t+h) = n+m \,\big|\, X(t) = n\big) = \begin{cases} \lambda_n h + o(h) & \text{if } m = 1, \\ \mu_n h + o(h) & \text{if } m = -1, \\ o(h) & \text{if } |m| > 1, \end{cases}$$

(c) the 'birth rates' $\lambda_0, \lambda_1, \dots$ and the 'death rates' $\mu_0, \mu_1, \dots$ satisfy $\lambda_i \geq 0$, $\mu_i \geq 0$, $\mu_0 = 0$.

Then X is called a *birth–death process*. It has generator $\mathbf{G} = (g_{ij} : i, j \geq 0)$ given by

$$\mathbf{G} = \begin{pmatrix} -\lambda_0 & \lambda_0 & 0 & 0 & 0 & \dots \\ \mu_1 & -(\lambda_1 + \mu_1) & \lambda_1 & 0 & 0 & \dots \\ 0 & \mu_2 & -(\lambda_2 + \mu_2) & \lambda_2 & 0 & \dots \\ 0 & 0 & \mu_3 & -(\lambda_3 + \mu_3) & \lambda_3 & \dots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}.$$

The chain is uniform if and only if $\sup_i\{\lambda_i + \mu_i\} < \infty$. In many particular cases we have that $\lambda_0 = 0$, and then 0 is an absorbing state and the chain is not irreducible.

The transition probabilities $p_{ij}(t) = \mathbb{P}(X(t) = j \mid X(0) = i)$ may in principle be calculated from a knowledge of the birth and death rates, although in practice these functions rarely have nice forms. It is an easier matter to determine the asymptotic behaviour of the process as $t \to \infty$. Suppose that $\lambda_i > 0$ and $\mu_i > 0$ for all relevant i. A stationary distribution $\boldsymbol{\pi}$ would satisfy $\boldsymbol{\pi}\mathbf{G} = \mathbf{0}$ which is to say that

$$\begin{aligned} -\lambda_0\pi_0 + \mu_1\pi_1 &= 0, \\ \lambda_{n-1}\pi_{n-1} - (\lambda_n + \mu_n)\pi_n + \mu_{n+1}\pi_{n+1} &= 0 \quad \text{if} \quad n \geq 1. \end{aligned}$$

A simple induction† yields that

(2)
$$\pi_n = \frac{\lambda_0\lambda_1 \cdots \lambda_{n-1}}{\mu_1\mu_2 \cdots \mu_n}\pi_0, \qquad n \geq 1.$$

Such a vector $\boldsymbol{\pi}$ is a stationary distribution if and only if $\sum_n \pi_n = 1$; this may happen if and only if

(3)
$$\sum_{n=0}^{\infty} \frac{\lambda_0\lambda_1 \cdots \lambda_{n-1}}{\mu_1\mu_2 \cdots \mu_n} < \infty,$$

where the term $n = 0$ is interpreted as 1; if this holds, then

(4)
$$\pi_0 = \left(\sum_{n=0}^{\infty} \frac{\lambda_0\lambda_1 \cdots \lambda_{n-1}}{\mu_1\mu_2 \cdots \mu_n} \right)^{-1}.$$

We have from Theorem (6.9.21) that the process settles into equilibrium (with stationary distribution given by (2) and (4)) if and only if the summation in (3) is finite, a condition requiring that the birth rates are not too large relative to the death rates.

Here are some examples of birth–death processes.

(5) Example. Pure birth. The death rates satisfy $\mu_n = 0$ for all n. ●

---

†Alternatively, note that the matrix is tridiagonal, whence the chain is reversible in equilibrium (see Problem (6.15.16c)). Now seek a solution to the detailed balance equations.

(6) Example. Simple death with immigration. Let us model a population which evolves in the following way. At time zero the size $X(0)$ of the population equals I. Individuals do not reproduce, but new individuals immigrate into the population at the arrival times of a Poisson process with intensity $\lambda > 0$. Each individual may die in the time interval $(t, t+h)$ with probability $\mu h + \mathrm{o}(h)$, where $\mu > 0$. The transition probabilities of $X(t)$ satisfy

$$\begin{aligned} p_{ij}(h) &= \mathbb{P}\big(X(t+h) = j \,\big|\, X(t) = i\big) \\ &= \begin{cases} \mathbb{P}(j-i \text{ arrivals, no deaths}) + \mathrm{o}(h) & \text{if } j \geq i, \\ \mathbb{P}(i-j \text{ deaths, no arrivals}) + \mathrm{o}(h) & \text{if } j < i, \end{cases} \end{aligned}$$

since the probability of two or more changes occurring during the interval $(t, t+h)$ is $\mathrm{o}(h)$. Therefore

$$\begin{aligned} p_{i,i+1}(h) &= \lambda h(1-\mu h)^i + \mathrm{o}(h) = \lambda h + \mathrm{o}(h), \\ p_{i,i-1}(h) &= i(\mu h)(1-\mu h)^{i-1}(1-\lambda h) + \mathrm{o}(h) = (i\mu)h + \mathrm{o}(h), \\ p_{ij}(h) &= \mathrm{o}(h) \quad \text{if} \quad |j-i| > 1, \end{aligned}$$

and we recognize X as a birth–death process with parameters

$$\lambda_n = \lambda, \qquad \mu_n = n\mu. \tag{7}$$

It is an irreducible continuous-time Markov chain and, by Theorem (6.10.5), it is not uniform. We may ask for the distribution of $X(t)$ and for the limiting distribution of the chain as $t \to \infty$. The former question is answered by solving the forward equations; this is Problem (6.15.18). The latter question is answered by the following. ●

(8) Theorem. *In the limit as $t \to \infty$, $X(t)$ is asymptotically Poisson distributed with parameter $\rho = \lambda/\mu$. That is,*

$$\mathbb{P}\big(X(t) = n\big) \to \frac{\rho^n}{n!}e^{-\rho}, \qquad n = 0, 1, 2, \dots .$$

Proof. Either substitute (7) into (2) and (4), or solve the equation $\boldsymbol{\pi}\mathbf{G} = \mathbf{0}$ directly. ■●

(9) Example. Simple birth–death. Assume that each individual who is alive in the population at time t either dies in the interval $(t, t+h)$ with probability $\mu h + \mathrm{o}(h)$ or splits into two in the interval with probability $\lambda h + \mathrm{o}(h)$. Different individuals behave independently of one another. The transition probabilities satisfy equations such as

$$\begin{aligned} p_{i,i+1}(h) &= \mathbb{P}(\text{one birth, no deaths}) + \mathrm{o}(h) \\ &= i(\lambda h)(1-\lambda h)^{i-1}(1-\mu h)^i + \mathrm{o}(h) \\ &= (i\lambda)h + \mathrm{o}(h) \end{aligned}$$

and it is easy to check that the number $X(t)$ of living individuals at time t satisfies (1) with

$$\lambda_n = n\lambda, \qquad \mu_n = n\mu.$$

We shall explore this model in detail. The chain $X = \{X(t)\}$ is standard but not uniform. We shall assume that $X(0) = I > 0$; the state 0 is absorbing. We find the distribution of $X(t)$ through its generating function.

(10) Theorem. *The generating function of $X(t)$ is*

$$G(s,t) = \mathbb{E}(s^{X(t)}) = \begin{cases} \left(\dfrac{\lambda t(1-s)+s}{\lambda t(1-s)+1}\right)^I & \textit{if } \mu = \lambda, \\ \left(\dfrac{\mu(1-s)-(\mu-\lambda s)e^{-t(\lambda-\mu)}}{\lambda(1-s)-(\mu-\lambda s)e^{-t(\lambda-\mu)}}\right)^I & \textit{if } \mu \neq \lambda. \end{cases}$$

Proof. This is like Proof B of Theorem (6.8.2). Write $p_j(t) = \mathbb{P}(X(t) = j)$ and condition $X(t+h)$ on $X(t)$ to obtain the forward equations

$$\begin{aligned} p_j'(t) &= \lambda(j-1)p_{j-1}(t) - (\lambda+\mu)jp_j(t) + \mu(j+1)p_{j+1}(t) \quad \text{if } j \geq 1, \\ p_0'(t) &= \mu p_1(t). \end{aligned}$$

Multiply the jth equation by s^j and sum to obtain

$$\begin{aligned} \sum_{j=0}^{\infty} s^j p_j'(t) = \lambda s^2 \sum_{j=1}^{\infty}(j-1)s^{j-2}p_{j-1}(t) - (\lambda+\mu)s\sum_{j=0}^{\infty} js^{j-1}p_j(t) \\ + \mu \sum_{j=0}^{\infty}(j+1)s^j p_{j+1}(t). \end{aligned}$$

Put $G(s,t) = \sum_0^\infty s^j p_j(t) = \mathbb{E}(s^{X(t)})$ to obtain

$$\begin{aligned} \frac{\partial G}{\partial t} &= \lambda s^2 \frac{\partial G}{\partial s} - (\lambda+\mu)s\frac{\partial G}{\partial s} + \mu\frac{\partial G}{\partial s} \\ &= (\lambda s - \mu)(s-1)\frac{\partial G}{\partial s} \end{aligned} \tag{11}$$

with boundary condition $G(s,0) = s^I$. The solution to this partial differential equation is given by (10); to see this either solve (11) by standard methods, or substitute the conclusion of (10) into (11). ■

Note that X is honest for all λ and μ since $G(1,t) = 1$ for all t. To find the mean and variance of $X(t)$, differentiate G:

$$\mathbb{E}(X(t)) = Ie^{(\lambda-\mu)t}, \qquad \operatorname{var}(X(t)) = \begin{cases} 2I\lambda t & \text{if } \lambda = \mu, \\ I\dfrac{\lambda+\mu}{\lambda-\mu}e^{(\lambda-\mu)t}[e^{(\lambda-\mu)t}-1] & \text{if } \lambda \neq \mu. \end{cases}$$

Write $\rho = \lambda/\mu$ and notice that

$$\mathbb{E}(X(t)) \to \begin{cases} 0 & \text{if } \rho < 1, \\ \infty & \text{if } \rho > 1. \end{cases}$$

(12) Corollary. *The extinction probabilities $\eta(t) = \mathbb{P}(X(t) = 0)$ satisfy, as $t \to \infty$,*

$$\eta(t) \to \begin{cases} 1 & \text{if } \rho \leq 1, \\ \rho^{-I} & \text{if } \rho > 1. \end{cases}$$

Proof. We have that $\eta(t) = G(0, t)$. Substitute $s = 0$ in $G(s, t)$ to find $\eta(t)$ explicitly. ∎

The observant reader will have noticed that these results are almost identical to those obtained for the branching process, except in that they pertain to a process in continuous time. There are (at least) two discrete Markov chains imbedded in X.

(A) Imbedded random walk. We saw in Claims (6.9.13) and (6.9.14) that if $X(s) = n$, say, then the length of time $T = \inf\{t > 0 : X(s+t) \neq n\}$ until the next birth or death is exponentially distributed with parameter $-g_{nn} = n(\lambda + \mu)$. When this time is complete, X moves from state n to state $n + M$ where

$$\mathbb{P}(M = 1) = -\frac{g_{n,n+1}}{g_{nn}} = \frac{\lambda}{\lambda + \mu}, \qquad \mathbb{P}(M = -1) = \frac{\mu}{\lambda + \mu}.$$

Think of this transition as the movement of a particle from the integer n to the new integer $n + M$, where $M = \pm 1$. Such a particle performs a simple random walk with parameter $p = \lambda/(\lambda + \mu)$ and initial position I. We know already (see Example (3.9.6)) that the probability of ultimate absorption at 0 is given by (12). Other properties of random walks (see Sections 3.9 and 5.3) are applicable also.

(B) Imbedded branching process. We can think of the birth–death process in the following way. After birth an individual lives for a certain length of time which is exponentially distributed with parameter $\lambda + \mu$. When this period is over it dies, leaving behind it either no individuals, with probability $\mu/(\lambda + \mu)$, or two individuals, with probability $\lambda/(\lambda + \mu)$. This is just an age-dependent branching process with age density function

(13)
$$f_T(u) = (\lambda + \mu)e^{-(\lambda+\mu)u}, \qquad u \geq 0,$$

and family-size generating function

(14)
$$G(s) = \frac{\mu + \lambda s^2}{\mu + \lambda},$$

in the notation of Section 5.5 (do not confuse G in (14) with $G(s, t) = \mathbb{E}(s^{X(t)})$). Thus if $I = 1$, the generating function $G(s, t) = \mathbb{E}(s^{X(t)})$ satisfies the differential equation

(15)
$$\frac{\partial G}{\partial t} = \lambda G^2 - (\lambda + \mu)G + \mu.$$

After (11), this is the *second* differential equation for $G(s, t)$. Needless to say, (15) is really just the backward equation of the process; the reader should check this and verify that it has the same solution as the forward equation (11). Suppose we lump together the members of each generation of this age-dependent branching process. Then we obtain an ordinary branching process with family-size generating function $G(s)$ given by (14). From the general theory, the extinction probability of the process is the smallest non-negative root of the equation $s = G(s)$, and we can verify easily that this is given by (12) with $I = 1$. ●

(16) Example. A more general branching process. Finally, we consider a more general type of age-dependent branching process than that above, and we investigate its honesty. Suppose that each individual in a population lives for an exponentially distributed time with parameter λ say. After death it leaves behind it a (possibly empty) family of offspring: the size N of this family has mass function $f(k) = \mathbb{P}(N = k)$ and generating function G_N. Let $X(t)$ be the size of the population at time t; we assume that $X(0) = 1$. From Section 5.5 the backward equation for $G(s,t) = \mathbb{E}(s^{X(t)})$ is

$$\frac{\partial G}{\partial t} = \lambda\big(G_N(G) - G\big)$$

with boundary condition $G(s, 0) = s$; the solution is given by

(17)
$$\int_s^{G(s,t)} \frac{du}{G_N(u) - u} = \lambda t$$

provided that $G_N(u) - u$ has no zeros within the domain of the integral. There are many interesting questions about this process; for example, is it honest in the sense that

$$\sum_{j=0}^{\infty} \mathbb{P}\big(X(t) = j\big) = 1?$$

(18) Theorem. *The process X is honest if and only if*

(19)
$$\int_{1-\epsilon}^{1} \frac{du}{G_N(u) - u} \quad \textit{diverges for all } \epsilon > 0.$$

Proof. See Harris (1963, p. 107). ■

If condition (19) fails then the population size may explode to $+\infty$ in finite time.

(20) Corollary. *X is honest if* $\mathbb{E}(N) < \infty$.

Proof. Expand $G_N(u) - u$ about $u = 1$ to find that

$$G_N(u) - u = [\mathbb{E}(N) - 1](u - 1) + \mathrm{o}(u - 1) \quad \text{as} \quad u \uparrow 1.$$ ■●

Exercises for Section 6.11

1. Describe the jump chain for a birth–death process with rates λ_n and μ_n.

2. Consider an immigration–death process X, being a birth–death process with birth rates $\lambda_n = \lambda$ and death rates $\mu_n = n\mu$. Find the transition matrix of the jump chain Z, and show that it has as stationary distribution

$$\pi_n = \frac{1}{2(n!)}\left(1 + \frac{n}{\rho}\right)\rho^n e^{-\rho}$$

where $\rho = \lambda/\mu$. Explain why this differs from the stationary distribution of X.

3. Consider the birth–death process X with $\lambda_n = n\lambda$ and $\mu_n = n\mu$ for all $n \geq 0$. Suppose $X(0) = 1$ and let $\eta(t) = \mathbb{P}(X(t) = 0)$. Show that η satisfies the differential equation

$$\eta'(t) + (\lambda + \mu)\eta(t) = \mu + \lambda\eta(t)^2.$$

Hence find $\eta(t)$, and calculate $\mathbb{P}(X(t) = 0 \mid X(u) = 0)$ for $0 < t < u$.

4. For the birth–death process of the previous exercise with $\lambda < \mu$, show that the distribution of $X(t)$, conditional on the event $\{X(t) > 0\}$, converges as $t \to \infty$ to a geometric distribution.

5. Let X be a birth–death process with $\lambda_n = n\lambda$ and $\mu_n = n\mu$, and suppose $X(0) = 1$. Show that the time T at which $X(t)$ first takes the value 0 satisfies

$$\mathbb{E}(T \mid T < \infty) = \begin{cases} \dfrac{1}{\lambda}\log\left(\dfrac{\mu}{\mu - \lambda}\right) & \text{if } \lambda < \mu, \\[2ex] \dfrac{1}{\mu}\log\left(\dfrac{\lambda}{\lambda - \mu}\right) & \text{if } \lambda > \mu. \end{cases}$$

What happens when $\lambda = \mu$?

6. Let X be the birth–death process of Exercise (5) with $\lambda \neq \mu$, and let $V_r(t)$ be the total amount of time the process has spent in state $r \geq 0$, up to time t. Find the distribution of $V_1(\infty)$ and the generating function $\sum_r s^r \mathbb{E}(V_r(t))$. Hence show in two ways that $\mathbb{E}(V_1(\infty)) = [\max\{\lambda, \mu\}]^{-1}$. Show further that $\mathbb{E}(V_r(\infty)) = \lambda^{r-1} r^{-1} [\max\{\lambda, \mu\}]^{-r}$.

7. Repeat the calculations of Exercise (6) in the case $\lambda = \mu$.

6.12 Special processes

There are many more general formulations of the processes which we modelled in Sections 6.8 and 6.11. Here is a very small selection of some of them, with some details of the areas in which they have been found useful.

(1) Non-homogeneous chains. We may relax the assumption that the transition probabilities $p_{ij}(s, t) = \mathbb{P}(X(t) = j \mid X(s) = i)$ satisfy the homogeneity condition $p_{ij}(s, t) = p_{ij}(0, t - s)$. This leads to some very difficult problems. We may make some progress in the special case when X is the simple birth–death process of the previous section, for which $\lambda_n = n\lambda$ and $\mu_n = n\mu$. The parameters λ and μ are now assumed to be non-constant functions of t. (After all, most populations have birth and death rates which vary from season to season.) It is easy to check that the forward equation (6.11.11) remains unchanged:

$$\frac{\partial G}{\partial t} = [\lambda(t)s - \mu(t)](s - 1)\frac{\partial G}{\partial s}.$$

The solution is

$$G(s,t) = \left[1 + \left(\frac{e^{r(t)}}{s-1} - \int_0^t \lambda(u)e^{r(u)}\,du\right)^{-1}\right]^I$$

where $I = X(0)$ and

$$r(t) = \int_0^t \big[\mu(u) - \lambda(u)\big]\,du.$$

The extinction probability $\mathbb{P}(X(t) = 0)$ is the coefficient of s^0 in $G(s,t)$, and it is left as an *exercise* for the reader to prove the next result.

(2) Theorem. $\mathbb{P}(X(t) = 0) \to 1$ *if and only if*

$$\int_0^T \mu(u)e^{r(u)}\,du \to \infty \quad \textit{as} \quad T \to \infty. \qquad \bullet$$

(3) A bivariate branching process. We advertised the branching process as a feasible model for the growth of cell populations; we should also note one of its inadequacies in this role. Even the age-dependent process cannot meet the main objection, which is that the time of division of a cell may depend rather more on the *size* of the cell than on its *age*. So here is a model for the growth and degradation of long-chain polymers†.

A population comprises *particles*. Let $N(t)$ be the number of particles present at time t, and suppose that $N(0) = 1$. We suppose that the $N(t)$ particles are partitioned into $W(t)$ groups of size $N_1, N_2, \dots, N_W$ such that the particles in each group are aggregated into a *unit cell*. Think of the cells as a collection of $W(t)$ polymers, containing $N_1, N_2, \dots, N_W$ particles respectively. As time progresses each cell grows and divides. We suppose that each cell can accumulate one particle from outside the system with probability $\lambda h + o(h)$ in the time interval $(t, t+h)$. As cells become larger they are more likely to divide. We assume that the probability that a cell of size N divides into two cells of sizes M and $N - M$, for some $0 < M < N$, during the interval $(t, t+h)$, is $\mu(N-1)h + o(h)$. The assumption that the probability of division is a *linear* function of the cell size N is reasonable for polymer degradation since the particles are strung together in a line and any of the $N-1$ 'links' between pairs of particles may sever. At time t there are $N(t)$ particles and $W(t)$ cells, and the process is said to be in state $X(t) = (N(t), W(t))$. During the interval $(t, t+h)$ various transitions for $X(t)$ are possible. Either some cell grows or some cell divides, or more than one such event occurs. The probability that some cell grows is $\lambda Wh + o(h)$ since there are W chances of this happening; the probability of a division is

$$\mu(N_1 + \cdots + N_W - W)h + o(h) = \mu(N - W)h + o(h)$$

since there are $N - W$ links in all; the probability of more than one such occurrence is $o(h)$. Putting this information together results in a Markov chain $X(t) = (N(t), W(t))$ with state space $\{1, 2, \dots\}^2$ and transition probabilities

$$\mathbb{P}\big(X(t+h) = (n,w) + \epsilon \,\big|\, X(t) = (n,w)\big)$$
$$= \begin{cases} \lambda wh + o(h) & \text{if } \epsilon = (1,0), \\ \mu(n-w)h + o(h) & \text{if } \epsilon = (0,1), \\ 1 - [w(\lambda - \mu) + \mu n]h + o(h) & \text{if } \epsilon = (0,0), \\ o(h) & \text{otherwise.} \end{cases}$$

†In physical chemistry, a *polymer* is a chain of molecules, neighbouring pairs of which are joined by bonds.

Write down the forward equations as usual to obtain that the joint generating function

$$G(x, y; t) = \mathbb{E}(x^{N(t)} y^{W(t)})$$

satisfies the partial differential equation

$$\frac{\partial G}{\partial t} = \mu x(y-1)\frac{\partial G}{\partial x} + y\big[\lambda(x-1) - \mu(y-1)\big]\frac{\partial G}{\partial y}$$

with $G(x, y; 0) = xy$. The joint moments of N and W are easily derived from this equation. More sophisticated techniques show that $N(t) \to \infty$, $W(t) \to \infty$, and $N(t)/W(t)$ approaches some constant as $t \to \infty$.

Unfortunately, most cells in nature are irritatingly non-Markovian! ●

(4) A non-linear epidemic. Consider a population of constant size $N + 1$, and watch the spread of a disease about its members. Let $X(t)$ be the number of healthy individuals at time t and suppose that $X(0) = N$. We assume that if $X(t) = n$ then the probability of a new infection during $(t, t + h)$ is proportional to the number of possible encounters between ill folk and healthy folk. That is,

$$\mathbb{P}\big(X(t+h) = n-1 \,\big|\, X(t) = n\big) = \lambda n(N+1-n)h + \mathrm{o}(h).$$

Nobody ever gets better. In the usual way, the reader can show that

$$G(s, t) = \mathbb{E}(s^{X(t)}) = \sum_{n=0}^{N} s^n \mathbb{P}\big(X(t) = n\big)$$

satisfies

$$\frac{\partial G}{\partial t} = \lambda(1-s)\left(N\frac{\partial G}{\partial s} - s\frac{\partial^2 G}{\partial s^2}\right)$$

with $G(s, 0) = s^N$. There is no simple way of solving this equation, though a lot of information is available about approximate solutions. ●

(5) Birth–death with immigration. We saw in Example (6.11.6) that populations are not always closed and that there is sometimes a chance that a new process will be started by an arrival from outside. This may be due to mutation (if we are counting genes), or leakage (if we are counting neutrons), or irresponsibility (if we are counting cases of rabies).

Suppose that there is one individual in the population at time zero; this individual is the founding member of some birth–death process N with fixed but unspecified parameters. Suppose further that other individuals immigrate into the population in the manner of a Poisson process I with intensity ν. Each immigrant starts a new birth–death process which is an independent identically distributed copy of the original process N but displaced in time according to its time of arrival. Let T_0 $(= 0)$, $T_1, T_2, \dots$ be the times at which immigrants arrive, and let $X_1, X_2, \dots$ be the interarrival times $X_n = T_n - T_{n-1}$. The total population at time t is the aggregate of the processes generated by the $I(t) + 1$ immigrants up to time t. Call this total $Y(t)$ to obtain

(6) $$Y(t) = \sum_{i=0}^{I(t)} N_i(t - T_i)$$

where $N_1, N_2, \ldots$ are independent copies of $N = N_0$. The problem is to find how the distribution of Y depends on the typical process N and the immigration rate ν; this is an example of the problem of compounding discussed in Theorem (5.1.25).

First we prove an interesting result about order statistics. Remember that I is a Poisson process and $T_n = \inf\{t : I(t) = n\}$ is the time of the nth immigration.

(7) Theorem. *The conditional joint distribution of $T_1, T_2, \ldots, T_n$, conditional on the event $\{I(t) = n\}$, is the same as the joint distribution of the order statistics of a family of n independent variables which are uniformly distributed on* $[0, t]$.

This is something of a mouthful, and asserts that if we know that n immigrants have arrived by time t then their actual arrival times are indistinguishable from a collection of n points chosen uniformly at random in the interval $[0, t]$.

Proof. We want the conditional density function of $\mathbf{T} = (T_1, T_2, \ldots, T_n)$ given that $I(t) = n$. First note that $X_1, X_2, \ldots, X_n$ are independent exponential variables with parameter ν so that

$$f_{\mathbf{X}}(\mathbf{x}) = \nu^n \exp\left(-\nu \sum_{i=1}^{n} x_i\right).$$

Make the transformation $\mathbf{X} \mapsto \mathbf{T}$ and use the change of variable formula (4.7.4) to find that

$$f_{\mathbf{T}}(\mathbf{t}) = \nu^n e^{-\nu t_n} \quad \text{if} \quad t_1 < t_2 < \cdots < t_n.$$

Let $C \subset \mathbb{R}^n$. Then

$$\mathbb{P}\big(\mathbf{T} \in C \,\big|\, I(t) = n\big) = \frac{\mathbb{P}(I(t) = n \text{ and } \mathbf{T} \in C)}{\mathbb{P}(I = n)}, \tag{8}$$

but

$$\begin{aligned} \mathbb{P}\big(I(t) = n \text{ and } \mathbf{T} \in C\big) &= \int_C \mathbb{P}\big(I(t) = n \,\big|\, \mathbf{T} = \mathbf{t}\big) f_{\mathbf{T}}(\mathbf{t})\, d\mathbf{t} \\ &= \int_C \mathbb{P}\big(I(t) = n \,\big|\, T_n = t_n\big) f_{\mathbf{T}}(\mathbf{t})\, d\mathbf{t} \end{aligned} \tag{9}$$

and

$$\mathbb{P}\big(I(t) = n \,\big|\, T_n = t_n\big) = \mathbb{P}(X_{n+1} > t - t_n) = e^{-\nu(t - t_n)} \tag{10}$$

so long as $t_n \le t$. Substitute (10) into (9) and (9) into (8) to obtain

$$\mathbb{P}\big(\mathbf{T} \in C \,\big|\, I(t) = n\big) = \int_C L(\mathbf{t}) n!\, t^{-n}\, d\mathbf{t}$$

where

$$L(\mathbf{t}) = \begin{cases} 1 & \text{if } t_1 < t_2 < \cdots < t_n, \\ 0 & \text{otherwise.} \end{cases}$$

We recognize $g(\mathbf{t}) = L(\mathbf{t}) n!\, t^{-n}$ from the result of Problem (4.14.23) as the joint density function of the order statistics of n independent uniform variables on $[0, t]$. ■

We are now ready to describe $Y(t)$ in terms of the constituent processes N_i.

(11) Theorem. *If $N(t)$ has generating function $G_N(s,t) = \mathbb{E}(s^{N(t)})$ then the generating function $G(s,t) = \mathbb{E}(s^{Y(t)})$ satisfies*

$$G(s,t) = G_N(s,t)\exp\left(\nu\int_0^t [G_N(s,u)-1]\,du\right).$$

Proof. Let $U_1, U_2, \dots$ be a sequence of independent uniform variables on $[0,t]$. By (6),

$$\mathbb{E}(s^{Y(t)}) = \mathbb{E}\big(s^{N_0(t)+N_1(t-T_1)+\cdots+N_I(t-T_I)}\big)$$

where $I = I(t)$. By independence, conditional expectation, and (7),

(12)
$$\begin{aligned}\mathbb{E}(s^{Y(t)}) &= \mathbb{E}(s^{N_0(t)})\mathbb{E}\big\{\mathbb{E}(s^{N_1(t-T_1)+\cdots+N_I(t-T_I)} \mid I)\big\}\\ &= G_N(s,t)\mathbb{E}\big\{\mathbb{E}(s^{N_1(t-U_1)+\cdots+N_I(t-U_I)} \mid I)\big\}\\ &= G_N(s,t)\mathbb{E}\{\mathbb{E}(s^{N_1(t-U_1)})^I\}.\end{aligned}$$

However,

(13)
$$\begin{aligned}\mathbb{E}(s^{N_1(t-U_1)}) &= \mathbb{E}\big\{\mathbb{E}(s^{N_1(t-U_1)} \mid U_1)\big\}\\ &= \int_0^t \frac{1}{t}G_N(s,t-u)\,du = H(s,t), \quad \text{and}\end{aligned}$$

(14)
$$\mathbb{E}(H^I) = \sum_{k=0}^{\infty} H^k \frac{(\nu t)^k}{k!}e^{-\nu t} = e^{\nu t(H-1)}.$$

Substitute (13) and (14) into (12) to obtain the result. ■●

(15) Branching random walk. Another characteristic of many interesting populations is their distribution about the space that they inhabit. We introduce this spatial aspect gently, by assuming that each individual lives at some point on the real line. (This may seem a fair description of a sewer, river, or hedge.) Let us suppose that the evolution proceeds as follows. After its birth, a typical individual inhabits a randomly determined spot X in $\mathbb{R}$ for a random time T. After this time has elapsed it dies, leaving behind a family containing N offspring which it distributes at points $X+Y_1, X+Y_2, \dots, X+Y_N$ where $Y_1, Y_2, \dots$ are independent and identically distributed. These individuals then behave as their ancestor did, producing the next generation offspring after random times at points $X+Y_i+Y_{ij}$, where Y_{ij} is the displacement of the jth offspring of the ith individual, and the Y_{ij} are independent and identically distributed. We shall be interested in the way that living individuals are distributed about $\mathbb{R}$ at some time t.

Suppose that the process begins with a single newborn individual at the point 0. We require some notation. Write $G_N(s)$ for the generating function of a typical family size N and let F be the distribution function of a typical Y. Let $Z(x,t)$ be the number of living individuals at points in the interval $(-\infty, x]$ at time t. We shall study the generating function

$$G(s;x,t) = \mathbb{E}(s^{Z(x,t)}).$$

Let T be the lifetime of the initial individual, N its family size, and $Y_1, Y_2, \ldots, Y_N$ the positions of its offspring. We shall condition Z on all these variables to obtain a type of backward equation. We must be careful about the order in which we do this conditioning, because the length of the sequence $Y_1, Y_2, \ldots$ depends on N. Hold your breath, and note from Problem (4.14.29) that

$$G(s; x, t) = \mathbb{E}\Big\{\mathbb{E}\Big[\mathbb{E}\big(\mathbb{E}(s^Z \mid T, N, \mathbf{Y}) \,\big|\, T, N\big) \,\Big|\, T\Big]\Big\}.$$

Clearly

$$Z(x,t) = \begin{cases} Z(x,0) & \text{if } T > t, \\ \sum_{i=1}^{N} Z_i(x - Y_i, t - T) & \text{if } T \le t, \end{cases}$$

where the processes $Z_1, Z_2, \ldots$ are independent copies of Z. Hence

$$\mathbb{E}(s^Z \mid T, N, \mathbf{Y}) = \begin{cases} G(s; x, 0) & \text{if } T > t, \\ \prod_{i=1}^{N} G(s; x - Y_i, t - T) & \text{if } T \le t. \end{cases}$$

Thus, if $T \le t$ then

$$\begin{aligned} \mathbb{E}\Big[\mathbb{E}\big(\mathbb{E}(s^Z \mid T, N, \mathbf{Y}) \,\big|\, T, N\big) \,\Big|\, T\Big] &= \mathbb{E}\left[\left(\int_{-\infty}^{\infty} G(s; x - y, t - T)\, dF(y)\right)^N \,\middle|\, T\right] \\ &= G_N\left(\int_{-\infty}^{\infty} G(s; x - y, t - T)\, dF(y)\right). \end{aligned}$$

Now breathe again. We consider here only the Markovian case when T is exponentially distributed with some parameter μ. Then

$$G(s; x, t) = \int_0^t \mu e^{-\mu u} G_N\left(\int_{-\infty}^{\infty} G(s; x - y, t - u)\, dF(y)\right) du + e^{-\mu t} G(s; x, 0).$$

Substitute $v = t - u$ inside the integral and differentiate with respect to t to obtain

$$\frac{\partial G}{\partial t} + \mu G = \mu G_N\left(\int_{-\infty}^{\infty} G(s; x - y, t)\, dF(y)\right).$$

It is not immediately clear that this is useful. However, differentiate with respect to s at $s = 1$ to find that $m(x, t) = \mathbb{E}(Z(x, t))$ satisfies

$$\frac{\partial m}{\partial t} + \mu m = \mu \mathbb{E}(N) \int_{-\infty}^{\infty} m(x - y, t)\, dF(y)$$

which equation is approachable by Laplace transform techniques. Such results can easily be generalized to higher dimensions. ●

(16) Spatial growth. Here is a simple model for skin cancer. Suppose that each point (x, y) of the two-dimensional square lattice $\mathbb{Z}^2 = \{(x, y) : x, y = 0, \pm 1, \pm 2, \ldots\}$ is a skin cell. There are two types of cell, called b-cells (*benign* cells) and m-cells (*malignant* cells). Each cell lives for an exponentially distributed period of time, parameter β for b-cells and parameter

μ for m-cells, after which it splits into two similar cells, one of which remains at the point of division and the other displaces one of the four nearest neighbours, each chosen at random with probability $\frac{1}{4}$. The displaced cell moves out of the system. Thus there are two competing types of cell. We assume that m-cells divide at least as fast as b-cells; the ratio $\kappa = \mu/\beta \geq 1$ is the 'carcinogenic advantage'.

Suppose that there is only one m-cell initially and that all other cells are benign. What happens to the resulting tumour of malignant cells?

(17) Theorem. *If* $\kappa = 1$, *the m-cells die out with probability* 1, *but the mean time until extinction is infinite. If* $\kappa > 1$, *there is probability* κ^{-1} *that the m-cells die out, and probability* $1 - \kappa^{-1}$ *that their number grows beyond all bounds.*

Thus there is strictly positive probability of the malignant cells becoming significant if and only if the carcinogenic advantage exceeds one.

Proof. Let $X(t)$ be the number of m-cells at time t, and let T_0 $(= 0)$, $T_1, T_2, \dots$ be the sequence of times at which X changes its value. Consider the imbedded discrete-time process $X = \{X_n\}$, where $X_n = X(T_n+)$ is the number of m-cells just after the nth transition; X is a Markov chain taking values in $\{0, 1, 2, \dots\}$. Remember the imbedded random walk of the birth–death process, Example (6.11.9); in the case under consideration a little thought shows that X has transition probabilities

$$p_{i,i+1} = \frac{\mu}{\mu+\beta} = \frac{\kappa}{\kappa+1}, \quad p_{i,i-1} = \frac{1}{\kappa+1} \quad \text{if} \quad i \neq 0, \quad p_{0,0} = 1.$$

Therefore X_n is simply a random walk with parameter $p = \kappa/(\kappa+1)$ and with an absorbing barrier at 0. The probability of ultimate extinction from the starting point $X(0) = 1$ is κ^{-1}. The walk is symmetric and null persistent if $\kappa = 1$ and all non-zero states are transient if $\kappa > 1$. ■

If $\kappa = 1$, the same argument shows that the m-cells certainly die out whenever there is a finite number of them to start with. However, suppose that they are distributed initially at the points of some (possibly infinite) set. It is possible to decide what happens after a long length of time; roughly speaking this depends on the relative densities of benign and malignant cells over large distances. One striking result is the following.

(18) Theorem. *If* $\kappa = 1$, *the probability that a specified finite collection of points contains only one type of cell approaches one as* $t \to \infty$.

Sketch proof. If two cells have a common ancestor then they are of the same type. Since offspring displace any neighbour with equal probability, the line of ancestors of any cell performs a symmetric random walk in two dimensions stretching backwards in time. Therefore, given any two cells at time t, the probability that they have a common ancestor is the probability that two symmetric and independent random walks S_1 and S_2 which originate at these points have met by time t. The difference $S_1 - S_2$ is also a type of symmetric random walk, and, as in Theorem (5.10.17), $S_1 - S_2$ almost certainly visits the origin sooner or later, implying that $\mathbb{P}(S_1(t) = S_2(t) \text{ for some } t) = 1$. ■●

(19) Example. Simple queue. Here is a simple model for a queueing system. Customers enter a shop in the manner of a Poisson process, parameter λ. They are served in the order of their arrival by a single assistant; each service period is a random variable which we assume to be exponential with parameter μ and which is independent of all other considerations. Let $X(t)$ be the length of the waiting line at time t (including any person being served). It is easy to see that X is a birth–death process with parameters $\lambda_n = \lambda$ for $n \geq 0$, $\mu_n = \mu$ for $n \geq 1$. The server would be very unhappy indeed if the queue length $X(t)$ were to tend to infinity as $t \to \infty$, since then he or she would have very few tea breaks. It is not difficult to see that the distribution of $X(t)$ settles down to a limit distribution, as $t \to \infty$, if and only if $\lambda < \mu$, which is to say that arrivals occur more slowly than departures on average (see condition (6.11.3)). We shall consider this process in detail in Chapter 11, together with other more complicated queueing models. ●

Exercises for Section 6.12

1. Customers entering a shop are served in the order of their arrival by the single server. They arrive in the manner of a Poisson process with intensity λ, and their service times are independent exponentially distributed random variables with parameter μ. By considering the jump chain, show that the expected duration of a busy period B of the server is $(\mu - \lambda)^{-1}$ when $\lambda < \mu$. (The busy period runs from the moment a customer arrives to find the server free until the earliest subsequent time when the server is again free.)

2. **Disasters.** Immigrants arrive at the instants of a Poisson process of rate ν, and each independently founds a simple birth process of rate λ. At the instants of an independent Poisson process of rate δ, the population is annihilated. Find the probability generating function of the population $X(t)$, given that $X(0) = 0$.

3. **More disasters.** In the framework of Exercise (2), suppose that each immigrant gives rise to a simple birth–death process of rates λ and μ. Show that the mean population size stays bounded if and only if $\delta > \lambda - \mu$.

4. **The queue M/G/∞.** (See Section 11.1.) An ftp server receives clients at the times of a Poisson process with parameter λ, beginning at time 0. The ith client remains connected for a length S_i of time, where the S_i are independent identically distributed random variables, independent of the process of arrivals. Assuming that the server has an infinite capacity, show that the number of clients being serviced at time t has the Poisson distribution with parameter $\lambda \int_0^t [1 - G(x)]\,dx$, where G is the common distribution function of the S_i.

6.13 Spatial Poisson processes

The Poisson process of Section 6.8 is a cornerstone of the theory of continuous-time Markov chains. It is also a beautiful process in its own right, with rich theory and many applications. While the process of Section 6.8 was restricted to the time axis $\mathbb{R}_+ = [0, \infty)$, there is a useful generalization to the Euclidean space $\mathbb{R}^d$ where $d \geq 1$.

We begin with a technicality. Recall that the essence of the Poisson process of Section 6.8 was the set of arrival times, a random countable subset of $\mathbb{R}_+$. Similarly, a realization of a Poisson process on $\mathbb{R}^d$ will be a countable subset Π of $\mathbb{R}^d$. We shall study the distribution of Π through the number $|\Pi \cap A|$ of its points lying in a typical subset A of $\mathbb{R}^d$. Some regularity will be assumed about such sets A, namely that there is a well defined notion of the 'volume' of A. Specifically, we shall assume that $A \in \mathcal{B}^d$, where $\mathcal{B}^d$ denotes the Borel σ-field of $\mathbb{R}^d$, being the smallest σ-field containing all *boxes* of the form $\prod_{i=1}^d (a_i, b_i]$. Members of $\mathcal{B}^d$ are called *Borel sets*, and we write $|A|$ for the volume (or Lebesgue measure) of the Borel set A.

(1) Definition. The random countable subset Π of $\mathbb{R}^d$ is called a **Poisson process with (constant) intensity** λ if, for all $A \in \mathcal{B}^d$, the random variables $N(A) = |\Pi \cap A|$ satisfy:

(a) $N(A)$ has the Poisson distribution with parameter $\lambda|A|$, and

(b) if $A_1, A_2, \dots, A_n$ are disjoint sets in $\mathcal{B}^d$, then $N(A_1), N(A_2), \dots, N(A_n)$ are independent random variables.

We often refer to the counting process N as being itself a Poisson process if it satisfies (a) and (b) above. In the case when $\lambda > 0$ and $|A| = \infty$, the number $|\Pi \cap A|$ has the Poisson distribution with parameter ∞, a statement to be interpreted as $\mathbb{P}(|\Pi \cap A| = \infty) = 1$.

It is not difficult to see the equivalence of (1) and Definition (6.8.1) when $d = 1$. That is, if $d = 1$ and N satisfies (1), then

$$M(t) = N\big([0, t]\big), \quad t \geq 0, \tag{2}$$

satisfies (6.8.1). Conversely, if M satisfies (6.8.1), one may find a process N satisfying (1) such that (2) holds. Attractive features of the above definition include the facts that the origin plays no special role, and that the definition may be extended to sub-regions of $\mathbb{R}^d$ as well as to σ-fields of subsets of general measure spaces.

There are many stochastic models based on the Poisson process. One-dimensional processes were used by Bateman, Erlang, Geiger, and Rutherford around 1909 in their investigations of practical situations involving radioactive particles and telephone calls. Examples in two and higher dimensions include the positions of animals in their habitat, the distribution of stars in a galaxy or of galaxies in the universe, the locations of active sites in a chemical reaction or of the weeds in your lawn, and the incidence of thunderstorms and tornadoes. Even when a Poisson process is not a perfect description of such a system, it can provide a relatively simple yardstick against which to measure the improvements which may be offered by more sophisticated but often less tractable models.

Definition (1) utilizes as reference measure the Lebesgue measure on $\mathbb{R}^d$, in the sense that the volume of a set A is its Euclidean volume. It is useful to have a definition of a Poisson process with other measures than Lebesgue measure, and such processes are termed 'non-homogeneous'. Replacing the Euclidean element $\lambda\, d\mathbf{x}$ with the element $\lambda(\mathbf{x})\, d\mathbf{x}$, we obtain the following, in which $\Lambda(A)$ is given by

$$\Lambda(A) = \int_A \lambda(\mathbf{x})\, d\mathbf{x}, \quad A \in \mathcal{B}^d. \tag{3}$$

(4) Definition. Let $d \geq 1$ and let $\lambda : \mathbb{R}^d \to \mathbb{R}$ be a non-negative measurable function such that $\Lambda(A) < \infty$ for all bounded A. The random countable subset Π of $\mathbb{R}^d$ is called a **non-homogeneous Poisson process with intensity function** λ if, for all $A \in \mathcal{B}^d$, the random variables $N(A) = |\Pi \cap A|$ satisfy:

(a) $N(A)$ has the Poisson distribution with parameter $\Lambda(A)$, and

(b) if $A_1, A_2, \dots, A_n$ are disjoint sets in $\mathcal{B}^d$, then $N(A_1), N(A_2), \dots, N(A_n)$ are independent random variables.

We call the function $\Lambda(A)$, $A \in \mathcal{B}^d$, the *mean measure* of the process Π. We have constructed Λ as the integral (3) of the intensity function λ; one may in fact dispense altogether with the function λ, working instead with measures Λ which 'have no atoms' in the sense that $\Lambda(\{\mathbf{x}\}) = 0$ for all $\mathbf{x} \in \mathbb{R}^d$.

Our first theorem states that the union of two independent Poisson processes is also a Poisson process. A similar result is valid for the union of countably many independent Poisson processes.

(5) Superposition theorem. *Let Π' and Π'' be independent Poisson processes on $\mathbb{R}^d$ with respective intensity functions λ' and λ''. The set $\Pi = \Pi' \cup \Pi''$ is a Poisson process with intensity function $\lambda = \lambda' + \lambda''$.*

Proof. Let $N'(A) = |\Pi' \cap A|$ and $N''(A) = |\Pi'' \cap A|$. Then $N'(A)$ and $N''(A)$ are independent Poisson-distributed random variables with respective parameters $\Lambda'(A)$ and $\Lambda''(A)$, the integrals (3) of λ' and λ''. It follows that the sum $S(A) = N'(A) + N''(A)$ has the Poisson distribution with parameter $\Lambda'(A) + \Lambda''(A)$. Furthermore, if $A_1, A_2, \dots$ are disjoint, the random variables $S(A_1), S(A_2), \dots$ are independent. It remains to show that, almost surely, $S(A) = |\Pi \cap A|$ for all A, which is to say that no point of Π' coincides with a point of Π''. This is a rather technical step, and the proof may be omitted on a first read.

Since $\mathbb{R}^d$ is a countable union of bounded sets, it is enough to show that, for every bounded $A \subseteq \mathbb{R}^d$, A contains almost surely no point common to Π' and Π''. Let $n \geq 1$ and, for $\mathbf{k} = (k_1, k_2, \dots, k_d) \in \mathbb{Z}^d$, let $B_{\mathbf{k}}(n) = \prod_{i=1}^d (k_i 2^{-n}, (k_i + 1)2^{-n}]$; cubes of this form are termed *n-cubes* or *n-boxes*. Let A be a bounded subset of $\mathbb{R}^d$, and $\overline{A}$ be the (bounded) union of all $B_{\mathbf{k}}(0)$ which intersect A. The probability that A contains a point common to Π' and Π'' is bounded for all n by the probability that some $B_{\mathbf{k}}(n)$ lying in $\overline{A}$ contains a common point. This is no greater than the mean number of such boxes, whence

$$
\begin{aligned}
\mathbb{P}(\Pi' \cap \Pi'' \cap A \neq \varnothing) &\leq \sum_{\mathbf{k}: B_{\mathbf{k}}(n) \subseteq \overline{A}} \mathbb{P}\big(N'(B_{\mathbf{k}}(n)) \geq 1,\ N''(B_{\mathbf{k}}(n)) \geq 1\big) \\
&= \sum_{\mathbf{k}: B_{\mathbf{k}}(n) \subseteq \overline{A}} \big(1 - e^{-\Lambda'(B_{\mathbf{k}}(n))}\big)\big(1 - e^{-\Lambda''(B_{\mathbf{k}}(n))}\big) \\
&\leq \sum_{\mathbf{k}: B_{\mathbf{k}}(n) \subseteq \overline{A}} \Lambda'(B_{\mathbf{k}}(n))\Lambda''(B_{\mathbf{k}}(n)) \quad \text{since } 1 - e^{-x} \leq x \text{ for } x \geq 0 \\
&\leq \max_{\mathbf{k}: B_{\mathbf{k}}(n) \subseteq \overline{A}} \{\Lambda'(B_{\mathbf{k}}(n))\} \sum_{\mathbf{k}: B_{\mathbf{k}}(n) \subseteq \overline{A}} \Lambda''(B_{\mathbf{k}}(n)) \\
&= M_n(A)\Lambda''(\overline{A})
\end{aligned}
$$

where

$$
M_n(A) = \max_{\mathbf{k}: B_{\mathbf{k}}(n) \subseteq \overline{A}} \Lambda'(B_{\mathbf{k}}(n)).
$$

It is the case that $M_n(A) \to 0$ as $n \to \infty$. This is easy to prove when λ' is a constant function, since then $M_n(A) \propto |B_{\mathbf{k}}(n)| = 2^{-nd}$. It is not quite so easy to prove for general λ'. Since we shall need a slightly more general argument later, we state next the required result.

(6) Lemma. *Let μ be a measure on the pair $(\mathbb{R}^d, \mathcal{B}^d)$ which has no atoms, which is to say that $\mu(\{\mathbf{y}\}) = 0$ for all $\mathbf{y} \in \mathbb{R}^d$. Let $n \geq 1$, and $B_{\mathbf{k}}(n) = \prod_{i=1}^d (k_i 2^{-n}, (k_i + 1)2^{-n}]$, $\mathbf{k} \in \mathbb{Z}^d$. For any bounded set A, we have that*

$$
\max_{\mathbf{k}: B_{\mathbf{k}}(n) \subseteq A} \mu(B_{\mathbf{k}}(n)) \to 0 \qquad \text{as } n \to \infty.
$$

Returning to the proof of the theorem, it follows by Lemma (6) applied to the set $\overline{A}$ that $M_n(A) \to 0$ as $n \to \infty$, and the proof is complete. ■

Proof of Lemma (6). We may assume without loss of generality that A is a finite union of 0-cubes. Let

$$M_n(A) = \max_{\mathbf{k}: B_{\mathbf{k}}(n) \subseteq A} \mu(B_{\mathbf{k}}(n)),$$

and note that $M_n \geq M_{n+1}$. Suppose that $M_n(A) \not\to 0$. There exists $\delta > 0$ such that $M_n(A) > \delta$ for all n, and therefore, for every $n \geq 0$, there exists an n-cube $B_{\mathbf{k}}(n) \subseteq A$ with $\mu(B_{\mathbf{k}}(n)) > \delta$. We colour an m-cube C *black* if, for all $n \geq m$, there exists an n-cube $C' \subseteq C$ such that $\mu(C') > \delta$. Now A is the union of finitely many translates of $(0, 1]^d$, and for at least one of these, B_0 say, there exist infinitely many n such that B_0 contains some n-cube B' with $\mu(B') > \delta$. Since $\mu(\cdot)$ is monotonic, the 0-cube B_0 is black. By a similar argument, B_0 contains some black 1-cube B_1. Continuing similarly, we obtain an infinite decreasing sequence $B_0, B_1, \ldots$ such that each B_r is a black r-cube. In particular, $\mu(B_r) > \delta$ for all r, whereas†

$$\lim_{r\to\infty} \mu(B_r) = \mu\left(\bigcap_r B_r\right) = \mu(\{\mathbf{y}\}) = 0$$

by assumption, where $\mathbf{y}$ is the unique point in the intersection of the B_r. The conclusion of the theorem follows from this contradiction. ■

It is possible to avoid the complication of Lemma (6) at this stage, but we introduce the lemma here since it will be useful in the forthcoming proof of Rényi's theorem (17). In an alternative and more general approach to Poisson processes, instead of the 'random set' Π one studies the 'random measure' N. This leads to substantially easier proofs of results corresponding to (5) and the forthcoming (8), but at the expense of of extra abstraction.

The following 'mapping theorem' enables us to study the image of a Poisson process Π under a (measurable) mapping $f : \mathbb{R}^d \to \mathbb{R}^s$. Suppose that Π is a non-homogeneous Poisson process on $\mathbb{R}^d$ with intensity function λ, and consider the set $f(\Pi)$ of images of Π under f. We shall need that $f(\Pi)$ contains (with probability 1) no multiple points, and this imposes a constraint on the pair λ, f. The subset $B \subseteq \mathbb{R}^s$ contains the images of points of Π lying in $f^{-1}B$, whose cardinality is a random variable having the Poisson distribution with parameter $\Lambda(f^{-1}B)$. The key assumption on the pair λ, f will therefore be that

(7) $$\Lambda(f^{-1}\{\mathbf{y}\}) = 0 \quad \text{for all } \mathbf{y} \in \mathbb{R}^s,$$

where Λ is the integral (3) of λ.

(8) Mapping theorem. *Let Π be a non-homogeneous Poisson process on $\mathbb{R}^d$ with intensity function λ, and let $f : \mathbb{R}^d \to \mathbb{R}^s$ satisfy (7). Assume further that*

(9) $$\mu(B) = \Lambda(f^{-1}B) = \int_{f^{-1}B} \lambda(\mathbf{x})\, d\mathbf{x}, \qquad B \in \mathcal{B}^s,$$

satisfies $\mu(B) < \infty$ for all bounded sets B. Then $f(\Pi)$ is a non-homogeneous Poisson process on $\mathbb{R}^s$ with mean measure μ.

†We use here a property of continuity of general measures, proved in the manner of Lemma (1.3.5).

Proof. Assume for the moment that the points in $f(\Pi)$ are distinct. The number of points of $f(\Pi)$ lying in the set B ($\subseteq \mathbb{R}^s$) is $|\Pi \cap f^{-1}B|$, which has the Poisson distribution with parameter $\mu(B)$, as required. If $B_1, B_2, \dots$ are disjoint, their pre-images $f^{-1}B_1, f^{-1}B_2, \dots$ are disjoint also, whence the numbers of points in the B_i are independent. It follows that $f(\Pi)$ is a Poisson process, and it remains only to show the assumed distinctness of $f(\Pi)$. The proof of this is similar to that of (5), and may be omitted on first read.

We shall work with the set $\Pi \cap U$ of points of Π lying within the unit cube $U = (0, 1]^d$ of $\mathbb{R}^d$. This set is a Poisson process with intensity function

$$\lambda_U(\mathbf{x}) = \begin{cases} \lambda(\mathbf{x}) & \text{if } \mathbf{x} \in U, \\ 0 & \text{otherwise,} \end{cases}$$

and with finite total mass $\Lambda(U) = \int_U \lambda(\mathbf{x})\,d\mathbf{x}$. (This is easy to prove, and is in any case a very special consequence of the forthcoming colouring theorem (14).) We shall prove that $f(\Pi \cap U)$ is a Poisson process on $\mathbb{R}^s$ with mean measure

$$\mu_U(B) = \int_{f^{-1}B} \lambda_U(\mathbf{x})\,d\mathbf{x}.$$

A similar conclusion will hold for the set $f(\Pi \cap U_\mathbf{k})$ where $U_\mathbf{k} = \mathbf{k} + U$ for $\mathbf{k} \in \mathbb{Z}^d$, and the result will follow by the superposition theorem (5) (in a version for the sum of *countably* many Poisson processes) on noting that the sets $f(\Pi \cap U_\mathbf{k})$ are independent, and that, in the obvious notation,

$$\sum_\mathbf{k} \mu_{U_\mathbf{k}}(B) = \int_{f^{-1}B} \left\{ \sum_\mathbf{k} \lambda_{U_\mathbf{k}}(\mathbf{x}) \right\} d\mathbf{x} = \int_{f^{-1}B} \lambda(\mathbf{x})\,d\mathbf{x}.$$

Write $\Pi' = \Pi \cap U$, and assume for the moment that the points $f(\Pi')$ are almost surely distinct. The number of points lying in the subset B of $\mathbb{R}^s$ is $|\Pi' \cap f^{-1}B|$, which has the Poisson distribution with parameter $\mu_U(B)$ as required. If $B_1, B_2, \dots$ are disjoint, their pre-images $f^{-1}B_1, f^{-1}B_2, \dots$ are disjoint also, whence the numbers of points in the B_i are independent. It follows that $f(\Pi')$ is a Poisson process with mean measure μ_U.

It remains to show that the points in $f(\Pi')$ are almost surely distinct under hypothesis (7). The probability that the small box $B_\mathbf{k} = \prod_{i=1}^s (k_i 2^{-n}, (k_i + 1)2^{-n}]$ of $\mathbb{R}^s$ contains two or more points of $f(\Pi')$ is

$$1 - e^{-\mu_\mathbf{k}} - \mu_\mathbf{k} e^{-\mu_\mathbf{k}} \le 1 - (1 + \mu_\mathbf{k})(1 - \mu_\mathbf{k}) = \mu_\mathbf{k}^2,$$

where $\mu_\mathbf{k} = \mu_U(B_\mathbf{k})$, and we have used the fact that $e^{-x} \ge 1 - x$ for $x \ge 0$. The mean number of such boxes within the unit cube $U_s = (0, 1]^s$ is no greater than

$$\sum_\mathbf{k} \mu_\mathbf{k}^2 \le M_n \sum_\mathbf{k} \mu_\mathbf{k} = M_n \mu_U(U_s)$$

where

$$M_n = \max_\mathbf{k} \mu_\mathbf{k} \to 0 \qquad \text{as } n \to \infty$$

by hypothesis (7) and Lemma (6). Now $\mu_U(U_s) \le \Lambda(U) < \infty$, and we deduce as in the proof of Theorem (5) that U_s contains almost surely no repeated points. Since $\mathbb{R}^s$ is the union of countably many translates of U_s to each of which the above argument may be applied, we deduce that $\mathbb{R}^s$ contains almost surely no repeated points of $f(\Pi')$. The proof is complete. ■

(10) Example. Polar coordinates. Let Π be a Poisson process on $\mathbb{R}^2$ with constant rate λ, and let $f : \mathbb{R}^2 \to \mathbb{R}^2$ be the polar coordinate function $f(x, y) = (r, \theta)$ where

$$r = \sqrt{x^2 + y^2}, \quad \theta = \tan^{-1}(y/x).$$

It is straightforward to check that (7) holds, and we deduce that $f(\Pi)$ is a Poisson process on $\mathbb{R}^2$ with mean measure

$$\mu(B) = \int_{f^{-1}B} \lambda \, dx \, dy = \int_{B \cap S} \lambda r \, dr \, d\theta,$$

where $S = f(\mathbb{R}^2)$ is the strip $\{(r, \theta) : r \ge 0, \ 0 \le \theta < 2\pi\}$. We may think of $f(\Pi)$ as a Poisson process on the strip S having intensity function λr. ●

We turn now to one of the most important attributes of the Poisson process, which unlocks the door to many other useful results. This is the so-called 'conditional property', of which we saw a simple version in Theorem (6.12.7).

(11) Theorem. Conditional property. *Let Π be a non-homogeneous Poisson process on $\mathbb{R}^d$ with intensity function λ, and let A be a subset of $\mathbb{R}^d$ such that $0 < \Lambda(A) < \infty$. Conditional on the event that $|\Pi \cap A| = n$, the n points of the process lying in A have the same distribution as n points chosen independently at random in A according to the common probability measure*

$$\mathbb{Q}(B) = \frac{\Lambda(B)}{\Lambda(A)}, \qquad B \subseteq A.$$

Since

$$\mathbb{Q}(B) = \int_B \frac{\lambda(\mathbf{x})}{\Lambda(A)} \, d\mathbf{x}, \tag{12}$$

the relevant density function is $\lambda(\mathbf{x})/\Lambda(A)$ for $\mathbf{x} \in A$. When Π has constant intensity λ, the theorem implies that, given $|\Pi \cap A| = n$, the n points in question are distributed uniformly and independently at random in A.

Proof. Let $A_1, A_2, \dots, A_k$ be a partition of A. It is an elementary calculation that, if $n_1 + n_2 + \cdots + n_k = n$,

$$\begin{aligned}
(13) \quad \mathbb{P}\big(N(A_1) = n_1, \ N(A_2) = n_2, \dots, N(A_k) = n_k \,\big|\, N(A) = n\big) \\
&= \frac{\prod_i \mathbb{P}(N(A_i) = n_i)}{\mathbb{P}(N(A) = n)} \qquad \text{by independence} \\
&= \frac{\prod_i \Lambda(A_i)^{n_i} e^{-\Lambda(A_i)}/n_i!}{\Lambda(A)^n e^{-\Lambda(A)}/n!} \\
&= \frac{n!}{n_1! \, n_2! \cdots n_k!} \mathbb{Q}(A_1)^{n_1} \mathbb{Q}(A_2)^{n_2} \cdots \mathbb{Q}(A_k)^{n_k}.
\end{aligned}$$

The conditional distribution of the positions of the n points is specified by this function of $A_1, A_2, \dots, A_n$.

We recognize the multinomial distribution of (13) as the joint distribution of n points selected independently from A according to the probability measure $\mathbb{Q}$. It follows that the joint distribution of the points in $\Pi \cap A$, conditional on there being exactly n of them, is the same as that of the independent sample. ■

The conditional property enables a proof of the existence of Poisson processes and aids the simulation thereof. Let $\lambda > 0$ and let $A_1, A_2, \dots$ be a partition of $\mathbb{R}^d$ into Borel sets of finite Lebesgue measure. For each i, we simulate a random variable N_i having the Poisson distribution with parameter $\lambda|A_i|$. Then we sample n independently chosen points in A_i, each being uniformly distributed on A_i. The union over i of all such sets of points is a Poisson process with constant intensity λ. A similar construction is valid for a non-homogeneous process. The method may be facilitated by a careful choice of the A_i, perhaps as unit cubes of $\mathbb{R}^d$.

The following colouring theorem may be viewed as complementary to the superposition theorem (5). As in the latter case, there is a version of the theorem in which points are marked with one of countably many colours rather than just two.

(14) Colouring theorem. *Let Π be a non-homogeneous Poisson process on $\mathbb{R}^d$ with intensity function λ. We colour the points of Π in the following way. A point of Π at position $\mathbf{x}$ is coloured green with probability $\gamma(\mathbf{x})$; otherwise it is coloured scarlet (with probability $\sigma(\mathbf{x}) = 1 - \gamma(\mathbf{x})$). Points are coloured independently of one another. Let Γ and Σ be the sets of points coloured green and scarlet, respectively. Then Γ and Σ are independent Poisson processes with respective intensity functions $\gamma(\mathbf{x})\lambda(\mathbf{x})$ and $\sigma(\mathbf{x})\lambda(\mathbf{x})$.*

Proof. Let $A \subseteq \mathbb{R}^d$ with $\Lambda(A) < \infty$. By the conditional property (11), if $|\Pi \cap A| = n$, these points have the same distribution as n points chosen independently at random from A according to the probability measure $\mathbb{Q}(B) = \Lambda(B)/\Lambda(A)$. We may therefore consider n points chosen in this way. By the independence of the points, their colours are independent of one another. The chance that a given point is coloured green is $\overline{\gamma} = \int_A \gamma(\mathbf{x})\, d\mathbb{Q}$, the corresponding probability for the colour scarlet being $\overline{\sigma} = 1 - \overline{\gamma} = \int_A \sigma(\mathbf{x})\, d\mathbb{Q}$. It follows that, conditional on $|\Pi \cap A| = n$, the numbers N_g and N_s of green and scarlet points in A have, jointly, the binomial distribution

$$\mathbb{P}\big(N_g = g,\ N_s = s \,\big|\, N(A) = n\big) = \frac{n!}{g!\,s!}\overline{\gamma}^g\overline{\sigma}^s, \quad \text{where } g + s = n.$$

The unconditional probability is therefore

$$\begin{aligned}\mathbb{P}(N_g = g,\ N_s = s) &= \frac{(g+s)!}{g!\,s!}\overline{\gamma}^g\overline{\sigma}^s\frac{\Lambda(A)^{g+s}e^{-\Lambda(A)}}{(g+s)!}\\ &= \frac{(\overline{\gamma}\Lambda(A))^g e^{-\overline{\gamma}\Lambda(A)}}{g!}\cdot\frac{(\overline{\sigma}\Lambda(A))^s e^{-\overline{\sigma}\Lambda(A)}}{s!}\end{aligned}$$

which is to say that the numbers of green and scarlet points in A are independent. Furthermore they have, by (12), Poisson distributions with parameters

$$\overline{\gamma}\Lambda(A) = \int_A \gamma(\mathbf{x})\Lambda(A)\, d\mathbb{Q} = \int_A \gamma(\mathbf{x})\lambda(\mathbf{x})\, d\mathbf{x},$$

$$\overline{\sigma}\Lambda(A) = \int_A \sigma(\mathbf{x})\Lambda(A)\, d\mathbb{Q} = \int_A \sigma(\mathbf{x})\lambda(\mathbf{x})\, d\mathbf{x}.$$

Independence of the counts of points in disjoint regions follows trivially from the fact that Π has this property. ■

(15) Example. The Alternative Millennium Dome contains n zones, and visitors are required to view them all in sequence. Visitors arrive at the instants of a Poisson process on $\mathbb{R}_+$ with constant intensity λ, and the rth visitor spends time $X_{r,s}$ in the sth zone, where the random variables $X_{r,s}$, $r \geq 1$, $1 \leq s \leq n$, are independent.

Let $t \geq 0$, and let $V_s(t)$ be the number of visitors in zone s at time t. Show that, for fixed t, the random variables $V_s(t)$, $1 \leq s \leq n$, are independent, each with a Poisson distribution.

Solution. Let $T_1 < T_2 < \cdots$ be the times of arrivals of visitors, and let $c_1, c_2, \dots, c_n, \delta$ be distinct colours. A point of the Poisson process at time x is coloured c_s if and only if

(16) $$x + \sum_{v=1}^{s-1} X_v \leq t < x + \sum_{v=1}^{s} X_v$$

where $X_1, X_2, \dots, X_n$ are the times to be spent in the zones by a visitor arriving at time x. If (16) holds for no s, we colour the point at x with the colour δ; at time t, such a visitor has either not yet arrived or has already departed. Note that the colours of different points of the Poisson process are independent, and that a visitor arriving at time x is coloured c_s if and only if this individual is in zone s at time t.

The required independence follows by a version of the colouring theorem with $n + 1$ available colours instead of just two. ●

Before moving to other things, we note yet another characterization of the Poisson process. It turns out that one needs only check that the probability that a given region is empty is given by the Poisson formula. Recall from the proof of (5) that a *box* is a region of $\mathbb{R}^d$ of the form $B_{\mathbf{k}}(n) = \prod_{i=1}^{d}(k_i 2^{-n}, (k_i + 1)2^{-n}]$ for some $\mathbf{k} \in \mathbb{Z}^d$ and $n \geq 1$.

(17) Rényi's theorem. *Let Π be a random countable subset of $\mathbb{R}^d$, and let $\lambda : \mathbb{R}^d \to \mathbb{R}$ be a non-negative integrable function satisfying $\Lambda(A) = \int_A \lambda(\mathbf{x})\,d\mathbf{x} < \infty$ for all bounded A. If*

(18) $$\mathbb{P}(\Pi \cap A = \varnothing) = e^{-\Lambda(A)}$$

for any finite union A of boxes, then Π is a Poisson process with intensity function λ.

Proof. Let $n \geq 1$, and denote by $I_{\mathbf{k}}(n)$ the indicator function of the event that $B_{\mathbf{k}}(n)$ is non-empty. It follows by (18) that the events $I_{\mathbf{k}}(n)$, $\mathbf{k} \in \mathbb{Z}^d$, are independent.

Let A be a bounded open set in $\mathbb{R}^d$, and let $\mathcal{K}_n(A)$ be the set of all $\mathbf{k}$ such that $B_{\mathbf{k}}(n) \subseteq A$. Since A is open, we have that

(19) $$N(A) = |\Pi \cap A| = \lim_{n\to\infty} T_n(A) \quad \text{where} \quad T_n(A) = \sum_{\mathbf{k}\in\mathcal{K}_n(A)} I_{\mathbf{k}}(n);$$

note that, by the nesting of the boxes $B_{\mathbf{k}}(n)$, this is a monotone increasing limit. We have also that

(20) $$\Lambda(A) = \lim_{n\to\infty} \sum_{\mathbf{k}\in\mathcal{K}_n(A)} \Lambda(B_{\mathbf{k}}(n)).$$

The quantity $T_n(A)$ is the sum of independent variables, and has probability generating function

$$\mathbb{E}(s^{T_n(A)}) = \prod_{\mathbf{k} \in \mathcal{K}_n(A)} \left\{ s + (1-s)e^{-\Lambda(B_{\mathbf{k}}(n))} \right\}. \tag{21}$$

We have by Lemma (6) that $\Lambda(B_{\mathbf{k}}(n)) \to 0$ uniformly in $\mathbf{k} \in \mathcal{K}_n(A)$, as $n \to \infty$. Also, for fixed $s \in [0, 1)$, there exists $\phi(\delta)$ satisfying $\phi(\delta) \uparrow 1$ as $\delta \downarrow 0$ such that

$$e^{-(1-s)\alpha} \le s + (1-s)e^{-\alpha} \le e^{-(1-s)\phi(\delta)\alpha} \quad \text{if } 0 \le \alpha \le \delta. \tag{22}$$

[The left inequality holds by the convexity of e^{-x}, and the right inequality by Taylor's theorem.] It follows by (19), (21), and monotone convergence, that

$$\mathbb{E}(s^{N(A)}) = \lim_{n\to\infty} \prod_{\mathbf{k} \in \mathcal{K}_n(A)} \left\{ s + (1-s)e^{-\Lambda(B_{\mathbf{k}}(n))} \right\} \qquad \text{for } 0 \le s < 1,$$

and by (20) and (22) that, for fixed $s \in [0, 1)$,

$$e^{-(1-s)\Lambda(A)} \le \mathbb{E}(s^{N(A)}) \le e^{-(1-s)\phi(\delta)\Lambda(A)} \quad \text{for all } \delta > 0.$$

We take the limit as $\delta \downarrow 0$ to obtain the Poisson distribution of $N(A)$.

It remains to prove the independence of the variables $N(A_1), N(A_2), \dots$ for disjoint $A_1, A_2, \dots$. This is an immediate consequence of the facts that $T_n(A_1), T_n(A_2), \dots$ are independent, and $T_n(A_i) \to N(A_i)$ as $n \to \infty$. ■

There are many applications of the theory of Poisson processes in which the points of a process have an effect elsewhere in the space. A well-known practical example concerns the fortune of someone who plays a lottery. The player wins prizes at the times of a Poisson process Π on $\mathbb{R}_+$, and the amounts won are independent identically distributed random variables. Gains are discounted at rate α. The total gain $G(t)$ by time t may be expressed in the form

$$G(t) = \sum_{T \in \Pi,\, T \le t} \alpha^{t-T} W_T,$$

where W_T is the amount won at time T ($\in \Pi$). We may write

$$G(t) = \sum_{T \in \Pi} r(t-T) W_T$$

where

$$r(u) = \begin{cases} 0 & \text{if } u < 0, \\ \alpha^u & \text{if } u \ge 0. \end{cases}$$

Such sums may be studied by way of the next theorem. We state this in the special case of a homogeneous Poisson process on the half-line $\mathbb{R}_+$, but it is easily generalized. The one-dimensional problem is sometimes termed *shot noise*, since one may think of the sum as the cumulative effect of pulses which arrive in a system, and whose amplitudes decay exponentially.

(23) Theorem†. *Let Π be a Poisson process on $\mathbb{R}$ with constant intensity λ, let $r : \mathbb{R} \to \mathbb{R}$ be a smooth function, and let $\{W_x : x \in \Pi\}$ be independent identically distributed random variables, independent of Π. The sum*

$$G(t) = \sum_{x \in \Pi,\ x \geq 0} r(t - x) W_x$$

has characteristic function

$$\mathbb{E}(e^{i\theta G(t)}) = \exp\left\{\lambda \int_0^t \left(\mathbb{E}(e^{i\theta W r(s)}) - 1\right) ds\right\},$$

where W has the common distribution of the W_x. In particular,

$$\mathbb{E}(G(t)) = \lambda \mathbb{E}(W) \int_0^t r(s)\, ds.$$

Proof. This runs just like that of Theorem (6.12.11), which is in fact a special case. It is left as an *exercise* to check the details. The mean of $G(t)$ is calculated from its characteristic function by differentiating the latter at $\theta = 0$. ■

A similar idea works in higher dimensions, as the following demonstrates.

(24) Example. Olbers's paradox. Suppose that stars occur in $\mathbb{R}^3$ at the points $\{\mathbf{R}_i : i \geq 1\}$ of a Poisson process with constant intensity λ. The star at $\mathbf{R}_i$ has brightness B_i, where the B_i are independent and identically distributed with mean β. The intensity of the light striking an observer at the origin O from a star of brightness B, distance r away, is (in the absence of intervening clouds of dust) equal to cB/r^2, for some absolute constant c. Hence the total illumination at O from stars within a large ball S with radius a is

$$I_a = \sum_{i : |\mathbf{R}_i| \leq a} \frac{cB_i}{|\mathbf{R}_i|^2}.$$

Conditional on the event that the number N_a of such stars satisfies $N_a = n$, we have from the conditional property (11) that these n stars are uniformly and independently distributed over S. Hence

$$\mathbb{E}(I_a \mid N_a) = N_a c\beta \frac{1}{|S|} \int_S \frac{1}{|\mathbf{r}|^2}\, dV.$$

Now $\mathbb{E}(N_a) = \lambda|S|$, whence

$$\mathbb{E} I_a = \lambda c\beta \int_S \frac{1}{|\mathbf{r}|^2}\, dV = \lambda c\beta(4\pi a).$$

The fact that this is unbounded as $a \to \infty$ is called 'Olbers's paradox', and suggests that the celestial sphere should be uniformly bright at night. The fact that it is not is a problem whose resolution is still a matter for debate. One plausible explanation relies on a sufficiently fast rate of expansion of the Universe. ●

---

†This theorem is sometimes called the Campbell–Hardy theorem. See also Exercise (6.13.2).

Exercises for Section 6.13

1. In a certain town at time $t = 0$ there are no bears. Brown bears and grizzly bears arrive as independent Poisson processes B and G with respective intensities β and γ.

(a) Show that the first bear is brown with probability $\beta/(\beta+\gamma)$.

(b) Find the probability that between two consecutive brown bears, there arrive exactly r grizzly bears.

(c) Given that $B(1) = 1$, find the expected value of the time at which the first bear arrived.

2. Campbell–Hardy theorem. Let Π be the points of a non-homogeneous Poisson process on $\mathbb{R}^d$ with intensity function λ. Let $S = \sum_{\mathbf{x}\in\Pi} g(\mathbf{x})$ where g is a smooth function which we assume for convenience to be non-negative. Show that $\mathbb{E}(S) = \int_{\mathbb{R}^d} g(\mathbf{u})\lambda(\mathbf{u})\,d\mathbf{u}$ and $\operatorname{var}(S) = \int_{\mathbb{R}^d} g(\mathbf{u})^2\lambda(\mathbf{u})\,d\mathbf{u}$, provided these integrals converge.

3. Let Π be a Poisson process with constant intensity λ on the surface of the sphere of $\mathbb{R}^3$ with radius 1. Let P be the process given by the (X, Y) coordinates of the points projected on a plane passing through the centre of the sphere. Show that P is a Poisson process, and find its intensity function.

4. Repeat Exercise (3), when Π is a homogeneous Poisson process on the ball $\{(x_1, x_2, x_3) : x_1^2 + x_2^2 + x_3^2 \le 1\}$.

5. You stick pins in a Mercator projection of the Earth in the manner of a Poisson process with constant intensity λ. What is the intensity function of the corresponding process on the globe? What would be the intensity function on the map if you formed a Poisson process of constant intensity λ of meteorite strikes on the surface of the Earth?

6. Shocks. The rth point T_r of a Poisson process N of constant intensity λ on $\mathbb{R}_+$ gives rise to an effect $X_r e^{-\alpha(t-T_r)}$ at time $t \ge T_r$, where the X_r are independent and identically distributed with finite variance. Find the mean and variance of the total effect $S(t) = \sum_{r=1}^{N(t)} X_r e^{-\alpha(t-T_r)}$ in terms of the first two moments of the X_r, and calculate $\operatorname{cov}(S(s), S(t))$.

What is the behaviour of the correlation $\rho(S(s), S(t))$ as $s \to \infty$ with $t - s$ fixed?

7. Let N be a non-homogeneous Poisson process on $\mathbb{R}_+$ with intensity function λ. Find the joint density of the first two inter-event times, and deduce that they are not in general independent.

8. Competition lemma. Let $\{N_r(t) : r \ge 1\}$ be a collection of independent Poisson processes on $\mathbb{R}_+$ with respective constant intensities $\{\lambda_r : r \ge 1\}$, such that $\sum_r \lambda_r = \lambda < \infty$. Set $N(t) = \sum_r N_r(t)$, and let I denote the index of the process supplying the first point in N, occurring at time T. Show that

$$\mathbb{P}(I = i,\ T \ge t) = \mathbb{P}(I = i)\mathbb{P}(T \ge t) = \frac{\lambda_i}{\lambda} e^{-\lambda t}, \qquad i \ge 1.$$

6.14 Markov chain Monte Carlo

In applications of probability and statistics, we are frequently required to compute quantities of the form $\int_\Theta g(\theta)\pi(\theta)\,d\theta$ or $\sum_{\theta\in\Theta} g(\theta)\pi(\theta)$, where $g : \Theta \to \mathbb{R}$ and π is a density or mass function, as appropriate. When the domain Θ is large and π is complicated, it can be beyond the ability of modern computers to perform such a computation, and we may resort to 'Monte Carlo' methods (recall Section 2.6). Such situations arise surprisingly frequently in areas as disparate as statistical inference and physics. Monte Carlo techniques do not normally yield exact answers, but instead give a sequence of approximations to the required quantity.

(1) Example. Bayesian inference. A prior mass function $\pi(\theta)$ is postulated on the discrete set Θ of possible values of θ, and data x is collected. The posterior mass function $\pi(\theta \mid x)$ is given by

$$\pi(\theta \mid x) = \frac{f(x \mid \theta)\pi(\theta)}{\sum_{\psi \in \Theta} f(x \mid \psi)\pi(\psi)}.$$

It is required to compute some characteristic of the posterior, of the form

$$\mathbb{E}(g(\theta) \mid x) = \sum_{\theta} g(\theta)\pi(\theta \mid x).$$

Depending on the circumstances, such a quantity can be hard to compute. This problem arises commonly in statistical applications including the theory of image analysis, spatial statistics, and more generally in the analysis of large structured data sets. ●

(2) Example. Ising model†. We are given a finite graph $G = (V, E)$ with vertex set V and edge set E. Each vertex may be in either of two states, -1 or 1, and a *configuration* is a vector $\theta = \{\theta_v : v \in V\}$ lying in the state space $\Theta = \{-1, 1\}^V$. The configuration θ is assigned the probability

$$\pi(\theta) = \frac{1}{Z}\exp\left\{\sum_{v \neq w,\ v \sim w} \theta_v \theta_w\right\}$$

where the sum is over all pairs v, w of distinct neighbours in the graph G (the relation $\sim$ denoting adjacency), and Z is the appropriate normalizing constant, or 'partition function',

$$Z = \sum_{\theta \in \Theta} \exp\left\{\sum_{v \neq w,\ v \sim w} \theta_v \theta_w\right\}.$$

For $t, u \in V$, the chance that t and u have the same state is

$$\sum_{\theta : \theta_t = \theta_u} \pi(\theta) = \sum_{\theta} \tfrac{1}{2}(\theta_t \theta_u + 1)\pi(\theta).$$

The calculation of such probabilities can be strangely difficult. ●

It can be difficult to calculate the sums in such examples, even with the assistance of ordinary Monte Carlo methods. For example, the elementary Monte Carlo method of Section 2.6 relied upon having a supply of independent random variables with mass function π. In practice, Θ is often large and highly structured, and π may have complicated form, with the result that it may be hard to simulate directly from π. The 'Markov chain Monte Carlo' (McMC) approach is to construct a Markov chain having the following properties:

(a) the chain has π as unique stationary distribution,

(b) the transition probabilities of the chain have a simple form.

Property (b) ensures the easy simulation of the chain, and property (a) ensures that the distribution thereof approaches the required distribution as time passes. Let $X = \{X_n : n \geq 0\}$ be such a chain. Subject to weak conditions, the Cesàro averages of $g(X_r)$ satisfy

$$\frac{1}{n}\sum_{r=0}^{n-1} g(X_r) \to \sum_{\theta} g(\theta)\pi(\theta).$$

†This famous model of ferromagnetism was proposed by Lenz, and was studied by Ising around 1924.

The convergence is usually in mean square and almost surely (see Problem (6.15.44) and Chapter 7), and thus the Cesàro averages provide the required approximations.

Although the methods of this chapter may be adapted to *continuous* spaces Θ, we consider here only the case when Θ is finite. Suppose then that we are given a finite set Θ and a mass function $\boldsymbol{\pi} = (\pi_i : i \in \Theta)$, termed the 'target distribution'. Our task is to discuss how to construct an ergodic discrete-time Markov chain X on Θ with transition matrix $\mathbf{P} = (p_{ij})$, having given stationary distribution $\boldsymbol{\pi}$, and with the property that realizations of the X may be readily simulated.

There is a wide choice of such Markov chains. Computation and simulation is easier for reversible chains, and we shall therefore restrict out attention to chains whose transition probabilities p_{ij} satisfy the detailed balance equations

$$\pi_k p_{kj} = \pi_j p_{jk}, \quad j, k \in \Theta; \tag{3}$$

recall Definition (6.5.2). Producing a suitable chain X turns out to be remarkably straightforward. There are two steps in the following simple algorithm. Suppose that $X_n = i$, and it is required to construct X_{n+1}.

(i) Let $\mathbf{H} = (h_{ij} : i, j \in \Theta)$ be an arbitrary stochastic matrix, called the 'proposal matrix'. We pick $Y \in \Theta$ according to the probabilities $\mathbb{P}(Y = j \mid X_n = i) = h_{ij}$.

(ii) Let $\mathbf{A} = (a_{ij} : i, j \in \Theta)$ be a matrix with entries satisfying $0 \le a_{ij} \le 1$; the a_{ij} are called 'acceptance probabilities'. Given that $Y = j$, we set

$$X_{n+1} = \begin{cases} j & \text{with probability } a_{ij}, \\ X_n & \text{with probability } 1 - a_{ij}. \end{cases}$$

How do we determine the matrices $\mathbf{H}, \mathbf{A}$? The proposal matrix $\mathbf{H}$ is chosen in such a way that it is easy and cheap to simulate according to it. The acceptance matrix $\mathbf{A}$ is chosen in such a way that the detailed balance equations (3) hold. Since p_{ij} is given by

$$p_{ij} = \begin{cases} h_{ij} a_{ij} & \text{if } i \ne j, \\ 1 - \sum_{k:k\ne i} h_{ik} a_{ik} & \text{if } i = j, \end{cases} \tag{4}$$

the detailed balance equations (3) will be satisfied if we choose

$$a_{ij} = 1 \wedge \left(\frac{\pi_j h_{ji}}{\pi_i h_{ij}} \right) \tag{5}$$

where $x \wedge y = \min\{x, y\}$ as usual. This choice of $\mathbf{A}$ leads to an algorithm called the *Hastings algorithm*†. It may be considered desirable to accept as many proposals as possible, and this may be achieved as follows. Let (t_{ij}) be a symmetric matrix with non-negative entries satisfying $a_{ij} t_{ij} \le 1$ for all $i, j \in \Theta$, and let a_{ij} be given by (5). It is easy to see that one may choose any acceptance probabilities a'_{ij} given by $a'_{ij} = a_{ij} t_{ij}$. Such a generalization is termed *Hastings's general algorithm*.

While the above provides a general approach to McMC, further ramifications are relevant in practice. It is often the case in applications that the space Θ is a product space. For example,

†Or the *Metropolis–Hastings algorithm*; see Example (8).

it was the case in (2) that $\Theta = \{-1, 1\}^V$ where V is the vertex set of a certain graph; in the statistical analysis of images, one may take $\Theta = S^V$ where S is the set of possible states of a given pixel and V is the set of all pixels. It is natural to exploit this product structure in devising the required Markov chain, and this may be done as follows.

Suppose that S is a finite set of 'local states', that V is a finite index set, and set $\Theta = S^V$. For a given target distribution $\boldsymbol{\pi}$ on Θ, we seek to construct an approximating Markov chain X. One way to proceed is to restrict ourselves to transitions which flip the value of the current state at only one coordinate $v \in V$; this is called 'updating at v'. That is, given that $X_n = i = (i_w : w \in V)$, we decide that X_{n+1} takes a value in the set of all $j = (j_w : w \in V)$ such that $j_w = i_w$ whenever $w \neq v$. This may be achieved by following the above recipe in a way specific to the choice of the index v.

How do we decide on the choice of v? Several ways present themselves, of which the following two are obvious examples. One way is to select v uniformly at random from V at each step of the chain X. Another is to cycle through the elements of V is some deterministic manner.

(6) Example. Gibbs sampler, or heat bath algorithm. As in Example (2), take $\Theta = S^V$ where the 'local state space' S and the index set V are finite. For $i = (i_w : w \in V) \in \Theta$ and $v \in V$, let $\Theta_{i,v} = \{j \in \Theta : j_w = i_w \text{ for } w \neq v\}$. Suppose that $X_n = i$ and that we have decided to update at v. We take

(7)
$$h_{ij} = \frac{\pi_j}{\sum_{k \in \Theta_{i,v}} \pi_k}, \qquad j \in \Theta_{i,v},$$

which is to say that the proposal Y is chosen from $\Theta_{i,v}$ according to the conditional distribution given the other components i_w, $w \neq v$.

We have from (5) that $a_{ij} = 1$ for all $j \in \Theta_{i,v}$, on noting that $\Theta_{i,v} = \Theta_{j,v}$ if $j \in \Theta_{i,v}$. Therefore $\mathbb{P}_v(X_{n+1} = j \mid X_n = i) = h_{ij}$ for $j \in \Theta_{i,v}$, where $\mathbb{P}_v$ denotes the probability measure associated with updating at v.

We may choose the value of v either by flipping coins or by cycling through V in some pre-determined manner. ●

(8) Example. Metropolis algorithm. If the matrix $\mathbf{H}$ is symmetric, equation (5) gives $a_{ij} = 1 \wedge (\pi_j/\pi_i)$, whence $p_{ij} = h_{ij}\{1 \wedge (\pi_j/\pi_i)\}$ for $i \neq j$.

A simple choice for the proposal probabilities h_{ij} would be to sample the proposal 'uniformly at random' from the set of available changes. In the notation of Example (6), we might take

$$h_{ij} = \begin{cases} \dfrac{1}{|\Theta_{i,v}| - 1} & \text{if } j \neq i,\ j \in \Theta_{i,v}, \\ 0 & \text{if } j = i. \end{cases}$$ ●

The accuracy of McMC hinges on the rate at which the Markov chain X approaches its stationary distribution $\boldsymbol{\pi}$. In practical cases, it is notoriously difficult to decide whether or not X_n is close to its equilibrium, although certain theoretical results are available. The choice of distribution $\boldsymbol{\alpha}$ of X_0 is relevant, and it is worthwhile to choose $\boldsymbol{\alpha}$ in such a way that X_0 has strictly positive probability of lying in any part of the set Θ where $\boldsymbol{\pi}$ has positive weight. One might choose to estimate $\sum_\theta g(\theta)\pi(\theta)$ by $n^{-1}\sum_{r=M}^{M+n-1} g(X_r)$ for some large 'mixing time' M. We do not pursue here the determination of suitable M.

This section closes with a precise mathematical statement concerning the rate of convergence of the distribution $\boldsymbol{\alpha}\mathbf{P}^n$ to the stationary distribution $\boldsymbol{\pi}$. We assume for simplicity that X is aperiodic and irreducible. Recall from the Perron–Frobenius theorem (6.6.1) that $\mathbf{P}$ has $T = |\Theta|$ eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_T$ such that $\lambda_1 = 1$ and $|\lambda_j| < 1$ for $j \neq 1$. We write λ_2 for the eigenvalue with second largest modulus. It may be shown in some generality that

$$\mathbf{P}^n = \mathbf{I}\boldsymbol{\pi}' + \mathrm{O}(n^{m-1}|\lambda_2|^n),$$

where $\mathbf{I}$ is the identity matrix, $\boldsymbol{\pi}'$ is the column vector $(\pi_i : i \in \Theta)$, and m is the multiplicity of λ_2. Here is a concrete result in the reversible case.

(9) Theorem. *Let X be an aperiodic irreducible reversible Markov chain on the finite state space Θ, with transition matrix $\mathbf{P}$ and stationary distribution $\boldsymbol{\pi}$. Then*

$$\text{(10)} \qquad \sum_{k\in\Theta} |p_{ik}(n) - \pi_k| \le |\Theta| \cdot |\lambda_2|^n \sup\{|v_r(i)| : r \in \Theta\}, \qquad i \in \Theta,\ n \ge 1,$$

where $v_r(i)$ is the ith term of the rth right-eigenvector $\mathbf{v}_r$ of $\mathbf{P}$.

We note that the left side of (10) is the total variation distance (see equation (4.12.7)) between the mass functions $p_{i\cdot}(n)$ and $\boldsymbol{\pi}$.

Proof. Let $T = |\Theta|$ and number the states in Θ as $1, 2, \dots, T$. Using the notation and result of Exercise (6.14.1), we have that $\mathbf{P}$ is self-adjoint. Therefore the right eigenvectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_T$ corresponding to the eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_T$, are real. We may take $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_T$ to be an orthonormal basis of $\mathbb{R}^T$ with respect to the given scalar product. The unit vector $\mathbf{e}_k$, having 1 in its kth place and 0 elsewhere, may be written

$$\text{(11)} \qquad \mathbf{e}_k = \sum_{r=1}^{T} \langle \mathbf{e}_k, \mathbf{v}_r \rangle \mathbf{v}_r = \sum_{r=1}^{T} v_r(k)\pi_k \mathbf{v}_r.$$

Now $\mathbf{P}^n\mathbf{e}_k = (p_{1k}(n), p_{2k}(n), \dots, p_{Tk}(n))'$, and $\mathbf{P}^n\mathbf{v}_r = \lambda_r^n\mathbf{v}_r$. We pre-multiply (11) by $\mathbf{P}^n$ and deduce that

$$p_{ik}(n) = \sum_{r=1}^{T} v_r(k)\pi_k\lambda_r^n v_r(i).$$

Now $\mathbf{v}_1 = \mathbf{1}$ and $\lambda_1 = 1$, so that the term of the sum corresponding to $r = 1$ is simply π_k. It follows that

$$\sum_k |p_{ik}(n) - \pi_k| \le \sum_{r=2}^{T} |\lambda_r|^n |v_r(i)| \sum_k \pi_k |v_r(k)|.$$

By the Cauchy–Schwarz inequality,

$$\sum_k \pi_k|v_r(k)| \le \sqrt{\sum_k \pi_k |v_r(k)|^2} = 1,$$

and (10) follows. ■

Despite the theoretical appeal of such results, they are not always useful when $\mathbf{P}$ is large, because of the effort required to compute the right side of (10). It is thus important to establish readily computed bounds for $|\lambda_2|$, and bounds on $|p_{ik}(n) - \pi_k|$, which do not depend on the $\mathbf{v}_j$. We give a representative bound without proof.

(12) Theorem. Conductance bound. *We have under the assumptions of Theorem (9) that* $1 - 2\Psi \le \lambda_2 \le 1 - \frac{1}{2}\Psi^2$ *where*

$$\Psi = \inf\left\{ \sum_{i \in B} \pi_i p_{ij} \Big/ \sum_{i \in B} \pi_i : B \subseteq \Theta,\ 0 < \sum_{i \in B} \pi_i \le \tfrac{1}{2} \right\}.$$

Exercises for Section 6.14

1. Let $\mathbf{P}$ be a stochastic matrix on the finite set Θ with stationary distribution π. Define the inner product $\langle \mathbf{x}, \mathbf{y} \rangle = \sum_{k \in \Theta} x_k y_k \pi_k$, and let $l^2(\pi) = \{\mathbf{x} \in \mathbb{R}^\Theta : \langle \mathbf{x}, \mathbf{x} \rangle < \infty\}$. Show, in the obvious notation, that $\mathbf{P}$ is reversible with respect to π if and only if $\langle \mathbf{x}, \mathbf{P}\mathbf{y} \rangle = \langle \mathbf{P}\mathbf{x}, \mathbf{y} \rangle$ for all $\mathbf{x}, \mathbf{y} \in l^2(\pi)$.

2. Barker's algorithm. Show that a possible choice for the acceptance probabilities in Hastings's general algorithm is

$$b_{ij} = \frac{\pi_j g_{ji}}{\pi_i g_{ij} + \pi_j g_{ji}},$$

where $\mathbf{G} = (g_{ij})$ is the proposal matrix.

3. Let S be a countable set. For each $j \in S$, the sets A_{jk}, $k \in S$, form a partition of the interval $[0, 1]$. Let $g : S \times [0, 1] \to S$ be given by $g(j, u) = k$ if $u \in A_{jk}$. The sequence $\{X_n : n \ge 0\}$ of random variables is generated recursively by $X_{n+1} = g(X_n, U_{n+1})$, $n \ge 0$, where $\{U_n : n \ge 1\}$ are independent random variables with the uniform distribution on $[0, 1]$. Show that X is a Markov chain, and find its transition matrix.

4. Dobrushin's bound. Let $\mathbf{U} = (u_{st})$ be a finite $|S| \times |T|$ stochastic matrix. *Dobrushin's ergodic coefficient* is defined to be

$$d(\mathbf{U}) = \tfrac{1}{2} \sup_{i,j \in S} \sum_{t \in T} |u_{it} - u_{jt}|.$$

(a) Show that, if $\mathbf{V}$ is a finite $|T| \times |U|$ stochastic matrix, then $d(\mathbf{UV}) \le d(\mathbf{U}) d(\mathbf{V})$.

(b) Let X and Y be discrete-time Markov chains with the same transition matrix $\mathbf{P}$, and show that

$$\sum_k \big|\mathbb{P}(X_n = k) - \mathbb{P}(Y_n = k)\big| \le d(\mathbf{P})^n \sum_k \big|\mathbb{P}(X_0 = k) - \mathbb{P}(Y_0 = k)\big|.$$

6.15 Problems

1. Classify the states of the discrete-time Markov chains with state space $S = \{1, 2, 3, 4\}$ and transition matrices

$$\text{(a)} \quad \begin{pmatrix} \frac{1}{3} & \frac{2}{3} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{2} \\ 0 & 0 & 0 & 1 \end{pmatrix} \qquad \text{(b)} \quad \begin{pmatrix} 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}.$$

In case (a), calculate $f_{34}(n)$, and deduce that the probability of ultimate absorption in state 4, starting from 3, equals $\frac{2}{3}$. Find the mean recurrence times of the states in case (b).

2. A transition matrix is called *doubly stochastic* if all its column sums equal 1, that is, if $\sum_i p_{ij} = 1$ for all $j \in S$.

(a) Show that if a finite chain has a doubly stochastic transition matrix, then all its states are non-null persistent, and that if it is, in addition, irreducible and aperiodic then $p_{ij}(n) \to N^{-1}$ as $n \to \infty$, where N is the number of states.

(b) Show that, if an infinite irreducible chain has a doubly stochastic transition matrix, then its states are either all null persistent or all transient.

3. Prove that intercommunicating states of a Markov chain have the same period.

4. (a) Show that for each pair i, j of states of an irreducible aperiodic chain, there exists $N = N(i, j)$ such that $p_{ij}(n) > 0$ for all $n \geq N$.

(b) Let X and Y be independent irreducible aperiodic chains with the same state space S and transition matrix $\mathbf{P}$. Show that the bivariate chain $Z_n = (X_n, Y_n)$, $n \geq 0$, is irreducible and aperiodic.

(c) Show that the bivariate chain Z may be reducible if X and Y are periodic.

5. Suppose $\{X_n : n \geq 0\}$ is a discrete-time Markov chain with $X_0 = i$. Let N be the total number of visits made subsequently by the chain to the state j. Show that

$$\mathbb{P}(N = n) = \begin{cases} 1 - f_{ij} & \text{if } n = 0, \\ f_{ij}(f_{jj})^{n-1}(1 - f_{jj}) & \text{if } n \geq 1, \end{cases}$$

and deduce that $\mathbb{P}(N = \infty) = 1$ if and only if $f_{ij} = f_{jj} = 1$.

6. Let i and j be two states of a discrete-time Markov chain. Show that if i communicates with j, then there is positive probability of reaching j from i without revisiting i in the meantime. Deduce that, if the chain is irreducible and persistent, then the probability f_{ij} of ever reaching j from i equals 1 for all i and j.

7. Let $\{X_n : n \geq 0\}$ be a persistent irreducible discrete-time Markov chain on the state space S with transition matrix $\mathbf{P}$, and let $\mathbf{x}$ be a positive solution of the equation $\mathbf{x} = \mathbf{xP}$.

(a) Show that

$$q_{ij}(n) = \frac{x_j}{x_i} p_{ji}(n), \qquad i, j \in S,\ n \geq 1,$$

defines the n-step transition probabilities of a persistent irreducible Markov chain on S whose first-passage probabilities are given by

$$g_{ij}(n) = \frac{x_j}{x_i} l_{ji}(n), \qquad i \neq j,\ n \geq 1,$$

where $l_{ji}(n) = \mathbb{P}(X_n = i, T > n \mid X_0 = j)$ and $T = \min\{m > 0 : X_m = j\}$.

(b) Show that $\mathbf{x}$ is unique up to a multiplicative constant.

(c) Let $T_j = \min\{n \geq 1 : X_n = j\}$ and define $h_{ij} = \mathbb{P}(T_j \leq T_i \mid X_0 = i)$. Show that $x_i h_{ij} = x_j h_{ji}$ for all $i, j \in S$.

8. **Renewal sequences.** The sequence $u = \{u_n : n \geq 0\}$ is called a 'renewal sequence' if

$$u_0 = 1, \qquad u_n = \sum_{i=1}^{n} f_i u_{n-i} \qquad \text{for } n \geq 1,$$

for some collection $f = \{f_n : n \geq 1\}$ of non-negative numbers summing to 1.

(a) Show that u is a renewal sequence if and only if there exists a Markov chain X on a countable state space S such that $u_n = \mathbb{P}(X_n = s \mid X_0 = s)$, for some persistent $s \in S$ and all $n \geq 1$.

(b) Show that if u and v are renewal sequences then so is $\{u_n v_n : n \geq 0\}$.

9. Consider the symmetric random walk in three dimensions on the set of points $\{(x, y, z) : x, y, z = 0, \pm 1, \pm 2, \dots\}$; this process is a sequence $\{\mathbf{X}_n : n \geq 0\}$ of points such that $\mathbb{P}(\mathbf{X}_{n+1} = \mathbf{X}_n + \boldsymbol{\epsilon}) = \frac{1}{6}$ for $\boldsymbol{\epsilon} = (\pm 1, 0, 0), (0, \pm 1, 0), (0, 0, \pm 1)$. Suppose that $\mathbf{X}_0 = (0, 0, 0)$. Show that

$$\mathbb{P}\big(\mathbf{X}_{2n} = (0,0,0)\big) = \left(\frac{1}{6}\right)^{2n} \sum_{i+j+k=n} \frac{(2n)!}{(i!\,j!\,k!)^2} = \left(\frac{1}{2}\right)^{2n} \binom{2n}{n} \sum_{i+j+k=n} \left(\frac{n!}{3^n i!\,j!\,k!}\right)^2$$

and deduce by Stirling's formula that the origin is a transient state.

10. Consider the three-dimensional version of the cancer model (6.12.16). If $\kappa = 1$, are the empires of Theorem (6.12.18) inevitable in this case?

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

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

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

12. Another diffusion model. N black balls and N white balls are placed in two urns so that each contains N balls. After each unit of time one ball is selected at random from each urn, and the two balls thus selected are interchanged. Let the number of black balls in the first urn denote the state of the system. Write down the transition matrix of this Markov chain and find the unique stationary distribution. Is the chain reversible in equilibrium?

13. Consider a Markov chain on the set $S = \{0, 1, 2, \dots\}$ with transition probabilities $p_{i,i+1} = a_i$, $p_{i,0} = 1 - a_i$, $i \geq 0$, where $(a_i : i \geq 0)$ is a sequence of constants which satisfy $0 < a_i < 1$ for all i. Let $b_0 = 1$, $b_i = a_0 a_1 \cdots a_{i-1}$ for $i \geq 1$. Show that the chain is

(a) persistent if and only if $b_i \to 0$ as $i \to \infty$,

(b) non-null persistent if and only if $\sum_i b_i < \infty$,

and write down the stationary distribution if the latter condition holds.

Let A and β be positive constants and suppose that $a_i = 1 - Ai^{-\beta}$ for all large i. Show that the chain is

(c) transient if $\beta > 1$,

(d) non-null persistent if $\beta < 1$.

Finally, if $\beta = 1$ show that the chain is

(e) non-null persistent if $A > 1$,

(f) null persistent if $A \leq 1$.

14. Let X be a continuous-time Markov chain with countable state space S and standard semigroup $\{\mathbf{P}_t\}$. Show that $p_{ij}(t)$ is a continuous function of t. Let $g(t) = -\log p_{ii}(t)$; show that g is a continuous function, $g(0) = 0$, and $g(s+t) \leq g(s) + g(t)$. We say that g is 'subadditive', and a well known theorem gives the result that

$$\lim_{t \downarrow 0} \frac{g(t)}{t} = \lambda \quad \text{exists and} \quad \lambda = \sup_{t>0} \frac{g(t)}{t} \leq \infty.$$

Deduce that $g_{ii} = \lim_{t \downarrow 0} t^{-1}\{p_{ii}(t) - 1\}$ exists, but may be $-\infty$.

15. Let X be a continuous-time Markov chain with generator $\mathbf{G} = (g_{ij})$ and suppose that the transition semigroup $\mathbf{P}_t$ satisfies $\mathbf{P}_t = \exp(t\mathbf{G})$. Show that X is irreducible if and only if for any pair i, j of states there exists a sequence $k_1, k_2, \dots, k_n$ of states such that $g_{i,k_1} g_{k_1,k_2} \cdots g_{k_n,j} \neq 0$.

16. (a) Let $X = \{X(t) : -\infty < t < \infty\}$ be a Markov chain with stationary distribution $\boldsymbol{\pi}$, and suppose that $X(0)$ has distribution $\boldsymbol{\pi}$. We call X *reversible* if X and Y have the same joint distributions, where $Y(t) = X(-t)$.

(i) If $X(t)$ has distribution $\boldsymbol{\pi}$ for all t, show that Y is a Markov chain with transition probabilities $p'_{ij}(t) = (\pi_j/\pi_i)p_{ji}(t)$, where the $p_{ji}(t)$ are the transition probabilities of the chain X.

(ii) If the transition semigroup $\{\mathbf{P}_t\}$ of X is standard with generator $\mathbf{G}$, show that $\pi_i g_{ij} = \pi_j g_{ji}$ (for all i and j) is a necessary condition for X to be reversible.

(iii) If $\mathbf{P}_t = \exp(t\mathbf{G})$, show that $X(t)$ has distribution $\boldsymbol{\pi}$ for all t and that the condition in (ii) is sufficient for the chain to be reversible.

(b) Show that every irreducible chain X with exactly two states is reversible in equilibrium.

(c) Show that every birth–death process X having a stationary distribution is reversible in equilibrium.

17. Show that not every discrete-time Markov chain can be imbedded in a continuous-time chain. More precisely, let

$$\mathbf{P} = \begin{pmatrix} \alpha & 1-\alpha \\ 1-\alpha & \alpha \end{pmatrix} \quad \text{for some } 0 < \alpha < 1$$

be a transition matrix. Show that there exists a uniform semigroup $\{\mathbf{P}_t\}$ of transition probabilities in continuous time such that $\mathbf{P}_1 = \mathbf{P}$, if and only if $\frac{1}{2} < \alpha < 1$. In this case show that $\{\mathbf{P}_t\}$ is unique and calculate it in terms of α.

18. Consider an immigration–death process $X(t)$, being a birth–death process with rates $\lambda_n = \lambda$, $\mu_n = n\mu$. Show that its generating function $G(s,t) = \mathbb{E}(s^{X(t)})$ is given by

$$G(s,t) = \left\{1 + (s-1)e^{-\mu t}\right\}^I \exp\left\{\rho(s-1)(1-e^{-\mu t})\right\}$$

where $\rho = \lambda/\mu$ and $X(0) = I$. Deduce the limiting distribution of $X(t)$ as $t \to \infty$.

19. Let N be a non-homogeneous Poisson process on $\mathbb{R}_+ = [0, \infty)$ with intensity function λ. Write down the forward and backward equations for N, and solve them.

Let $N(0) = 0$, and find the density function of the time T until the first arrival in the process. If $\lambda(t) = c/(1+t)$, show that $\mathbb{E}(T) < \infty$ if and only if $c > 1$.

20. Successive offers for my house are independent identically distributed random variables X_1, $X_2, \dots$, having density function f and distribution function F. Let $Y_1 = X_1$, let Y_2 be the first offer exceeding Y_1, and generally let Y_{n+1} be the first offer exceeding Y_n. Show that $Y_1, Y_2, \dots$ are the times of arrivals in a non-homogeneous Poisson process with intensity function $\lambda(t) = f(t)/(1 - F(t))$. The Y_i are called 'record values'.

Now let Z_1 be the first offer received which is the second largest to date, and let Z_2 be the second such offer, and so on. Show that the Z_i are the arrival times of a non-homogeneous Poisson process with intensity function λ.

21. Let N be a Poisson process with constant intensity λ, and let $Y_1, Y_2, \dots$ be independent random variables with common characteristic function ϕ and density function f. The process $N^*(t) = Y_1 + Y_2 + \cdots + Y_{N(t)}$ is called a *compound* Poisson process. Y_n is the change in the value of N^* at the nth arrival of the Poisson process N. Think of it like this. A 'random alarm clock' rings at the arrival times of a Poisson process. At the nth ring the process N^* accumulates an extra quantity Y_n. Write down a forward equation for N^* and hence find the characteristic function of $N^*(t)$. Can you see directly why it has the form which you have found?

22. If the intensity function λ of a non-homogeneous Poisson process N is itself a random process, then N is called a *doubly stochastic* Poisson process (or *Cox process*). Consider the case when $\lambda(t) = \Lambda$ for all t, and Λ is a random variable taking either of two values λ_1 or λ_2, each being picked with equal probability $\frac{1}{2}$. Find the probability generating function of $N(t)$, and deduce its mean and variance.

23. Show that a simple birth process X with parameter λ is a doubly stochastic Poisson process with intensity function $\lambda(t) = \lambda X(t)$.

24. The Markov chain $X = \{X(t) : t \geq 0\}$ is a birth process whose intensities $\lambda_k(t)$ depend also on the time t and are given by

$$\mathbb{P}\big(X(t+h) = k+1 \,\big|\, X(t) = k\big) = \frac{1+\mu k}{1+\mu t} h + \mathrm{o}(h)$$

as $h \downarrow 0$. Show that the probability generating function $G(s,t) = \mathbb{E}(s^{X(t)})$ satisfies

$$\frac{\partial G}{\partial t} = \frac{s-1}{1+\mu t}\left\{G + \mu s \frac{\partial G}{\partial s}\right\}, \qquad 0 < s < 1.$$

Hence find the mean and variance of $X(t)$ when $X(0) = I$.

25. (a) Let X be a birth–death process with strictly positive birth rates $\lambda_0, \lambda_1, \ldots$ and death rates $\mu_1, \mu_2, \ldots$. Let η_i be the probability that $X(t)$ ever takes the value 0 starting from $X(0) = i$. Show that

$$\lambda_j \eta_{j+1} - (\lambda_j + \mu_j)\eta_j + \mu_j \eta_{j-1} = 0, \qquad j \geq 1,$$

and deduce that $\eta_i = 1$ for all i so long as $\sum_1^\infty e_j = \infty$ where $e_j = \mu_1\mu_2 \cdots \mu_j/(\lambda_1\lambda_2\cdots\lambda_j)$.
(b) For the discrete-time chain on the non-negative integers with

$$p_{j,j+1} = \frac{(j+1)^2}{j^2 + (j+1)^2} \quad \text{and} \quad p_{j,j-1} = \frac{j^2}{j^2 + (j+1)^2},$$

find the probability that the chain ever visits 0, starting from 1.

26. Find a good necessary condition and a good sufficient condition for the birth–death process X of Problem (6.15.25a) to be honest.

27. Let X be a simple symmetric birth–death process with $\lambda_n = \mu_n = n\lambda$, and let T be the time until extinction. Show that

$$\mathbb{P}(T \leq x \mid X(0) = I) = \left(\frac{\lambda x}{1+\lambda x}\right)^I,$$

and deduce that extinction is certain if $\mathbb{P}(X(0) < \infty) = 1$.

Show that $\mathbb{P}(\lambda T/I \leq x \mid X(0) = I) \to e^{-1/x}$ as $I \to \infty$.

28. Immigration–death with disasters. Let X be an immigration–death–disaster process, that is, a birth–death process with parameters $\lambda_i = \lambda$, $\mu_i = i\mu$, and with the additional possibility of 'disasters' which reduce the population to 0. Disasters occur at the times of a Poisson process with intensity δ, independently of all previous births and deaths.

(a) Show that X has a stationary distribution, and find an expression for the generating function of this distribution.

(b) Show that, in equilibrium, the mean of $X(t)$ is $\lambda/(\delta + \mu)$.

29. With any sufficiently nice (Lebesgue measurable, say) subset B of the real line $\mathbb{R}$ is associated a random variable $X(B)$ such that

(i) $X(B)$ takes values in $\{0, 1, 2, \ldots\}$,

(ii) if $B_1, B_2, \ldots, B_n$ are disjoint then $X(B_1), X(B_2), \ldots, X(B_n)$ are independent, and furthermore $X(B_1 \cup B_2) = X(B_1) + X(B_2)$,

(iii) the distribution of $X(B)$ depends only on B through its Lebesgue measure ('length') $|B|$, and

$$\frac{\mathbb{P}(X(B) \geq 1)}{\mathbb{P}(X(B) = 1)} \to 1 \quad \text{as } |B| \to 0.$$

Show that X is a Poisson process.

30. Poisson forest. Let N be a Poisson process in $\mathbb{R}^2$ with constant intensity λ, and let $R_{(1)} < R_{(2)} < \cdots$ be the ordered distances from the origin of the points of the process.
(a) Show that $R_{(1)}^2, R_{(2)}^2, \dots$ are the points of a Poisson process on $\mathbb{R}_+ = [0, \infty)$ with intensity $\lambda\pi$.
(b) Show that $R_{(k)}$ has density function

$$f(r) = \frac{2\pi\lambda r(\lambda\pi r^2)^{k-1}e^{-\lambda\pi r^2}}{(k-1)!}, \qquad r > 0.$$

31. Let X be a n-dimensional Poisson process with constant intensity λ. Show that the volume of the largest (n-dimensional) sphere centred at the origin which contains no point of X is exponentially distributed. Deduce the density function of the distance R from the origin to the nearest point of X. Show that $\mathbb{E}(R) = \Gamma(1/n)/\{n(\lambda c)^{1/n}\}$ where c is the volume of the unit ball of $\mathbb{R}^n$ and Γ is the gamma function.

32. A village of $N+1$ people suffers an epidemic. Let $X(t)$ be the number of ill people at time t, and suppose that $X(0) = 1$ and X is a birth process with rates $\lambda_i = \lambda i(N+1-i)$. Let T be the length of time required until every member of the population has succumbed to the illness. Show that

$$\mathbb{E}(T) = \frac{1}{\lambda}\sum_{k=1}^{N}\frac{1}{k(N+1-k)}$$

and deduce that

$$\mathbb{E}(T) = \frac{2(\log N + \gamma)}{\lambda(N+1)} + \mathrm{O}(N^{-2})$$

where γ is Euler's constant. It is striking that $\mathbb{E}(T)$ decreases with N, for large N.

33. A particle has velocity $V(t)$ at time t, where $V(t)$ is assumed to take values in $\{n + \frac{1}{2} : n \geq 0\}$. Transitions during $(t, t+h)$ are possible as follows:

$$\mathbb{P}\big(V(t+h) = w \mid V(t) = v\big) = \begin{cases} (v+\frac{1}{2})h + \mathrm{o}(h) & \text{if } w = v+1, \\ 1 - 2vh + \mathrm{o}(h) & \text{if } w = v, \\ (v-\frac{1}{2})h + \mathrm{o}(h) & \text{if } w = v-1. \end{cases}$$

Initially $V(0) = \frac{1}{2}$. Let

$$G(s,t) = \sum_{n=0}^{\infty} s^n \mathbb{P}\big(V(t) = n + \tfrac{1}{2}\big).$$

(a) Show that

$$\frac{\partial G}{\partial t} = (1-s)^2\frac{\partial G}{\partial s} - (1-s)G$$

and deduce that $G(s,t) = \{1 + (1-s)t\}^{-1}$.
(b) Show that the expected length $m_n(T)$ of time for which $V = n + \frac{1}{2}$ during the time interval $[0, T]$ is given by

$$m_n(T) = \int_0^T \mathbb{P}\big(V(t) = n + \tfrac{1}{2}\big)\,dt$$

and that, for fixed k, $m_k(T) - \log T \to -\sum_{i=1}^{k} i^{-1}$ as $T \to \infty$.
(c) What is the expected velocity of the particle at time t?

34. A random sequence of non-negative integers $\{X_n : n \geq 0\}$ begins $X_0 = 0$, $X_1 = 1$, and is produced by

$$X_{n+1} = \begin{cases} X_n + X_{n-1} & \text{with probability } \frac{1}{2}, \\ |X_n - X_{n-1}| & \text{with probability } \frac{1}{2}. \end{cases}$$

Show that $Y_n = (X_{n-1}, X_n)$ is a transient Markov chain, and find the probability of ever reaching $(1, 1)$ from $(1, 2)$.

35. Take a regular hexagon and join opposite corners by straight lines meeting at the point C. A particle performs a symmetric random walk on these 7 vertices, starting at A. Find:

(a) the probability of return to A without hitting C,

(b) the expected time to return to A,

(c) the expected nmber of visits to C before returning to A,

(d) the expected time to return to A, given that there is no prior visit to C.

36. Diffusion, osmosis. Markov chains are defined by the following procedures at any time n:

(a) **Bernoulli model.** Two adjacent containers A and B each contain m particles; m are of type I and m are of type II. A particle is selected at random in each container. If they are of opposite types they are exchanged with probability α if the type I is in A, or with probability β if the type I is in B. Let X_n be the number of type I particles in A at time n.

(b) **Ehrenfest dog–flea model.** Two adjacent containers contain m particles in all. A particle is selected at random. If it is in A it is moved to B with probability α, if it is in B it is moved to A with probability β. Let Y_n be the number of particles in A at time n.

In each case find the transition matrix and stationary distribution of the chain.

37. Let X be an irreducible continuous-time Markov chain on the state space S with transition probabilities $p_{jk}(t)$ and unique stationary distribution $\boldsymbol{\pi}$, and write $\mathbb{P}(X(t) = j) = a_j(t)$. If $c(x)$ is a concave function, show that $d(t) = \sum_{j \in S} \pi_j c(a_j(t)/\pi_j)$ increases to $c(1)$ as $t \to \infty$.

38. With the notation of the preceding problem, let $u_k(t) = \mathbb{P}(X(t) = k \mid X(0) = 0)$, and suppose the chain is reversible in equilibrium (see Problem (6.15.16)). Show that $u_0(2t) = \sum_j (\pi_0/\pi_j) u_j(t)^2$, and deduce that $u_0(t)$ decreases to π_0 as $t \to \infty$.

39. Perturbing a Poisson process. Let Π be the set of points in a Poisson process on $\mathbb{R}^d$ with constant intensity λ. Each point is displaced, where the displacements are independent and identically distributed. Show that the resulting point process is a Poisson process with intensity λ.

40. Perturbations continued. Suppose for convenience in Problem (6.15.39) that the displacements have a continuous distribution function and finite mean, and that $d = 1$. Suppose also that you are at the origin originally, and you move to a in the perturbed process. Let $\mathrm{L_R}$ be the number of points formerly on your left that are now on your right, and $\mathrm{R_L}$ the number of points formerly on your right that are now on your left. Show that $\mathbb{E}(\mathrm{L_R}) = \mathbb{E}(\mathrm{R_L})$ if and only if $a = \mu$ where μ is the mean displacement of a particle.

Deduce that if cars enter the start of a long road at the instants of a Poisson process, having independent identically distributed velocities, then, if you travel at the average speed, in the long run the rate at which you are overtaken by other cars equals the rate at which you overtake other cars.

41. Ants enter a kitchen at the instants of a Poisson process N of rate λ; they each visit the pantry and then the sink, and leave. The rth ant spends time X_r in the pantry and Y_r in the sink (and $X_r + Y_r$ in the kitchen altogether), where the vectors $V_r = (X_r, Y_r)$ and V_s are independent for $r \neq s$. At time $t = 0$ the kitchen is free of ants. Find the joint distribution of the numbers $A(t)$ of ants in the pantry and $B(t)$ of ants in the sink at time t. Now suppose the ants arrive in pairs at the times of the Poisson process, but then separate to behave independently as above. Find the joint distribution of the numbers of ants in the two locations.

42. Let $\{X_r : r \geq 1\}$ be independent exponential random variables with parameter λ, and set $S_n = \sum_{r=1}^{n} X_r$. Show that:

(a) $Y_k = S_k/S_n$, $1 \leq k \leq n-1$, have the same distribution as the order statistics of independent variables $\{U_k : 1 \leq k \leq n-1\}$ which are uniformly distributed on $(0, 1)$,

(b) $Z_k = X_k/S_n$, $1 \leq k \leq n$, have the same joint distribution as the coordinates of a point $(U_1, \dots, U_n)$ chosen uniformly at random on the simplex $\sum_{r=1}^{n} u_r = 1$, $u_r \geq 0$ for all r.

43. Let X be a discrete-time Markov chain with a finite number of states and transition matrix $\mathbf{P} = (p_{ij})$ where $p_{ij} > 0$ for all i, j. Show that there exists $\lambda \in (0, 1)$ such that $|p_{ij}(n) - \pi_j| < \lambda^n$, where $\boldsymbol{\pi}$ is the stationary distribution.

44. Under the conditions of Problem (6.15.43), let $V_i(n) = \sum_{r=0}^{n-1} I_{\{X_r=i\}}$ be the number of visits of the chain to i before time n. Show that

$$\mathbb{E}\left(\left|\frac{1}{n}V_i(n) - \pi_i\right|^2\right) \to 0 \quad \text{as } n \to \infty.$$

Show further that, if f is any bounded function on the state space, then

$$\mathbb{E}\left(\left|\frac{1}{n}\sum_{r=0}^{n-1} f(X_r) - \sum_{i \in S} f(i)\pi_i\right|^2\right) \to 0.$$

45. Conditional entropy. Let A and $\mathbf{B} = (B_0, B_1, \dots, B_n)$ be a discrete random variable and vector, respectively. The *conditional entropy* of A with respect to $\mathbf{B}$ is defined as $H(A \mid \mathbf{B}) = \mathbb{E}\big(\mathbb{E}\{-\log f(A \mid \mathbf{B}) \bigm| \mathbf{B}\}\big)$ where $f(a \mid \mathbf{b}) = \mathbb{P}(A = a \mid \mathbf{B} = \mathbf{b})$. Let X be an aperiodic Markov chain on a finite state space. Show that

$$H(X_{n+1} \mid X_0, X_1 \dots, X_n) = H(X_{n+1} \mid X_n),$$

and that

$$H(X_{n+1} \mid X_n) \to -\sum_i \pi_i \sum_j p_{ij} \log p_{ij} \qquad \text{as } n \to \infty,$$

if X is aperiodic with a unique stationary distribution $\boldsymbol{\pi}$.

46. Coupling. Let X and Y be independent persistent birth–death processes with the same parameters (and no explosions). It is not assumed that $X_0 = Y_0$. Show that:
(a) for any $A \subseteq \mathbb{R}$, $|\mathbb{P}(X_t \in A) - \mathbb{P}(Y_t \in A)| \to 0$ as $t \to \infty$,
(b) if $\mathbb{P}(X_0 \le Y_0) = 1$, then $\mathbb{E}[g(X_t)] \le \mathbb{E}[g(Y_t)]$ for any increasing function g.

47. Resources. The number of birds in a wood at time t is a continuous-time Markov process X. Food resources impose the constraint $0 \le X(t) \le n$. Competition entails that the transition probabilities obey

$$p_{k,k+1}(h) = \lambda(n-k)h + o(h), \qquad p_{k,k-1}(h) = \mu k h + o(h).$$

Find $\mathbb{E}(s^{X(t)})$, together with the mean and variance of $X(t)$, when $X(0) = r$. What happens as $t \to \infty$?

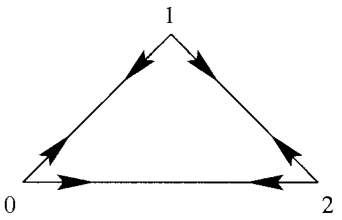
48. Parrando's paradox. A counter performs an irreducible random walk on the vertices 0, 1, 2 of the triangle in the figure beneath, with transition matrix

$$\mathbf{P} = \begin{pmatrix} 0 & p_0 & q_0 \\ q_1 & 0 & p_1 \\ p_2 & q_2 & 0 \end{pmatrix}$$

where $p_i + q_i = 1$ for all i. Show that the stationary distribution $\boldsymbol{\pi}$ has

$$\pi_0 = \frac{1 - q_2 p_1}{3 - q_1 p_0 - q_2 p_1 - q_0 p_2},$$

with corresponding formulae for π_1, π_2.



Suppose that you gain one peseta for each clockwise step of the walk, and you lose one peseta for each anticlockwise step. Show that, in equilibrium, the mean yield per step is

$$\gamma = \sum_i (2p_i - 1)\pi_i = \frac{3(2p_0p_1p_2 - p_0p_1 - p_1p_2 - p_2p_0 + p_0 + p_1 + p_2 - 1)}{3 - q_1p_0 - q_2p_1 - q_0p_2}.$$

Consider now three cases of this process:

A. We have $p_i = \frac{1}{2} - a$ for each i, where $a > 0$. Show that the mean yield per step satisfies $\gamma_A < 0$.

B. We have that $p_0 = \frac{1}{10} - a$, $p_1 = p_2 = \frac{3}{4} - a$, where $a > 0$. Show that $\gamma_B < 0$ for sufficiently small a.

C. At each step the counter is equally likely to move according to the transition probabilities of case A or case B, the choice being made independently at every step. Show that, in this case, $p_0 = \frac{3}{10} - a$, $p_1 = p_2 = \frac{5}{8} - a$. Show that $\gamma_C > 0$ for sufficiently small a.

The fact that two systematically unfavourable games may be combined to make a favourable game is called Parrando's paradox. Such bets are not available in casinos.

49. Cars arrive at the beginning of a long road in a Poisson stream of rate λ from time $t = 0$ onwards. A car has a fixed velocity $V > 0$ which is a random variable. The velocities of cars are independent and identically distributed, and independent of the arrival process. Cars can overtake each other freely. Show that the number of cars on the first x miles of the road at time t has the Poisson distribution with parameter $\lambda\mathbb{E}[V^{-1}\min\{x, Vt\}]$.

50. Events occur at the times of a Poisson process with intensity λ, and you are offered a bet based on the process. Let $t > 0$. You are required to say the word 'now' immediately after the event which you think will be the last to occur prior to time t. You win if you succeed, otherwise you lose. If no events occur before t you lose. If you have not selected an event before time t you lose.

Consider the strategy in which you choose the first event to occur after a specified time s, where $0 < s < t$.

(a) Calculate an expression for the probability that you win using this strategy.

(b) Which value of s maximizes this probability?

(c) If $\lambda t \geq 1$, show that the probability that you win using this value of s is e^{-1}.

51. A new Oxbridge professor wishes to buy a house, and can afford to spend up to one million pounds. Declining the services of conventional estate agents, she consults her favourite internet property page on which houses are announced at the times of a Poisson process with intensity λ per day. House prices may be assumed to be independent random variables which are uniformly distributed over the interval (800,000, 2,000,000). She decides to view every affordable property announced during the next 30 days. The time spent viewing any given property is uniformly distributed over the range (1, 2) hours. What is the moment generating function of the total time spent viewing houses?