# 2

# Continuous-time Markov chains I

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

## 2.1 $Q$-matrices and their exponentials

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

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

https://doi.org/10.1017/CBO9780511810633.004 Published online by Cambridge University Press

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

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

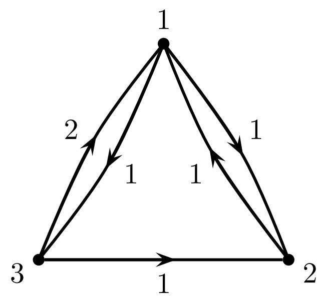
(iii) $\sum_{j \in I} q_{ij} = 0$ for all $i$.

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

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

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

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




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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

proving the semigroup property. The matrix-valued power series

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

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

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

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

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

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

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

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

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

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

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

*Proof.* As $t \downarrow 0$ we have

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

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

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

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

So

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

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

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

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

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

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

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

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

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

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

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

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

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

**Example 2.1.3**

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

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

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

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

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

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

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

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

Then

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

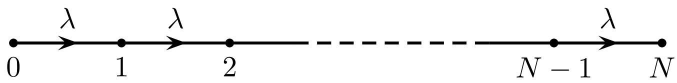
so $p_{11}(t)$ must have the form claimed.) To determine the constants we use

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

so

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

https://doi.org/10.1017/CBO9780511810633.004 Published online by Cambridge University Press




**Example 2.1.4**

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

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

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

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

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

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

so, by induction

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

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

**Exercises**

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

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

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

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

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

## 2.2 Continuous-time random processes

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

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

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

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

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

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

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

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

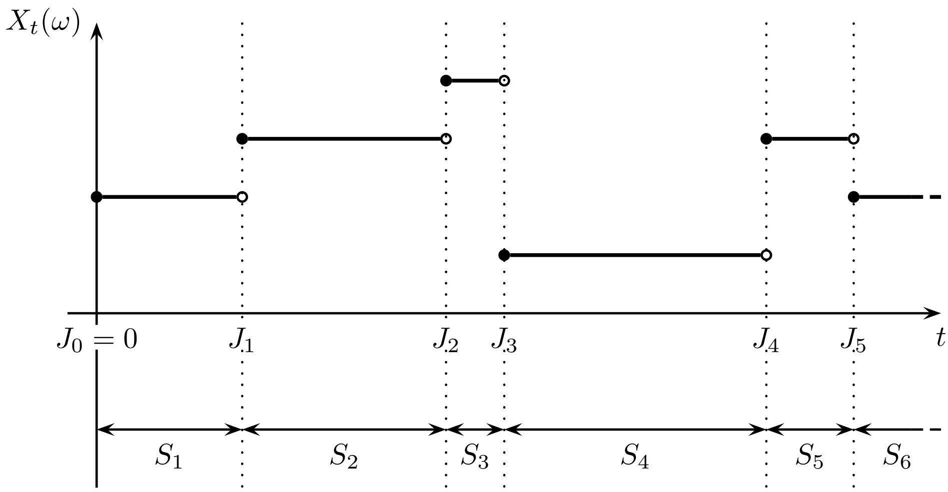
for $n \geq 0$, $0 \leq t_0 \leq t_1 \leq \ldots \leq t_n$ and $i_0, \dots, i_n \in I$. For example

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

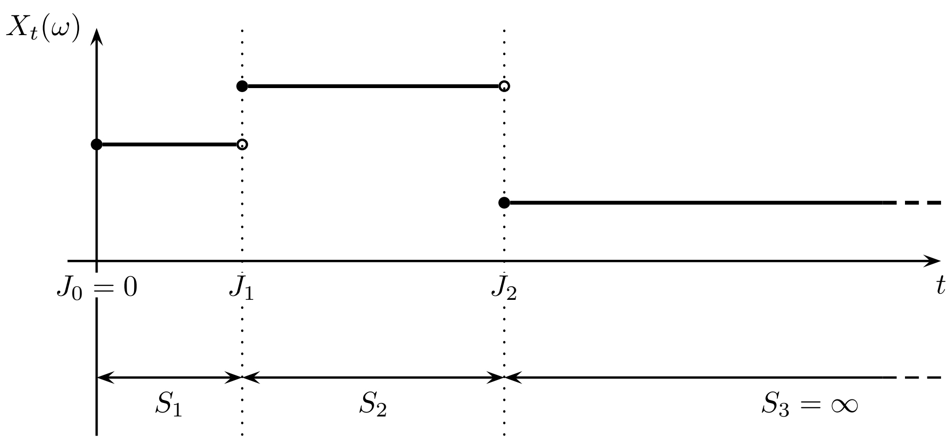
where $q_1, q_2, \ldots$ is an enumeration of the rationals.

https://doi.org/10.1017/CBO9780511810633.004 Published online by Cambridge University Press

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

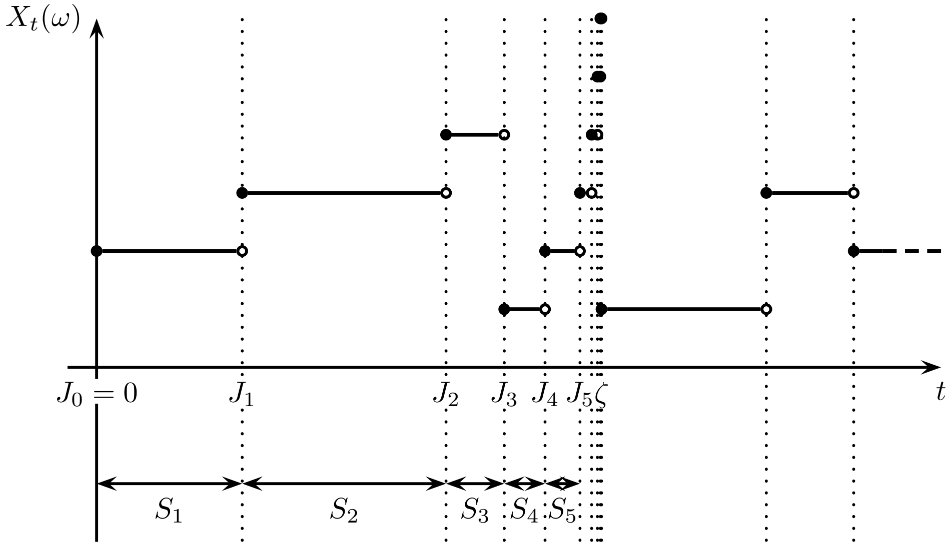



The second case is where the path makes finitely many jumps and then becomes stuck in some state forever:




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

https://doi.org/10.1017/CBO9780511810633.004 Published online by Cambridge University Press




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

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

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

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

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

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

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

We shall not consider what happens to a process after explosion. So it is convenient to adjoin to $I$ a new state, $\infty$ say, and require that $X_t = \infty$ if $t \geq \zeta$. Any process satisfying this requirement is called *minimal*. The terminology 'minimal' does not refer to the state of the process but to the

https://doi.org/10.1017/CBO9780511810633.004 Published online by Cambridge University Press

interval of time over which the process is active. Note that a minimal process may be reconstructed from its holding times and jump process. Thus by specifying the joint distribution of $S_1, S_2, \dots$ and $(Y_n)_{n\geq 0}$ we have another 'countable' specification of the probabilistic behaviour of $(X_t)_{t\geq 0}$. For example, the probability that $X_t = i$ is given by

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

and

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

## 2.3 Some properties of the exponential distribution

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

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

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

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

The mean of $T$ is given by

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

so

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

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

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

so

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

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

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

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

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

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

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

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

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

*Proof.* We have

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

from which the identity follows by symmetry. □

**Exercises**

**2.3.1** Suppose $S$ and $T$ are independent exponential random variables of parameters $\alpha$ and $\beta$ respectively. What is the distribution of $\min\{S,T\}$? What is the probability that $S \leq T$? Show that the two events $\{S < T\}$ and $\{\min\{S,T\} \geq t\}$ are independent.

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

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

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

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

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

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

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

## 2.4 Poisson processes

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

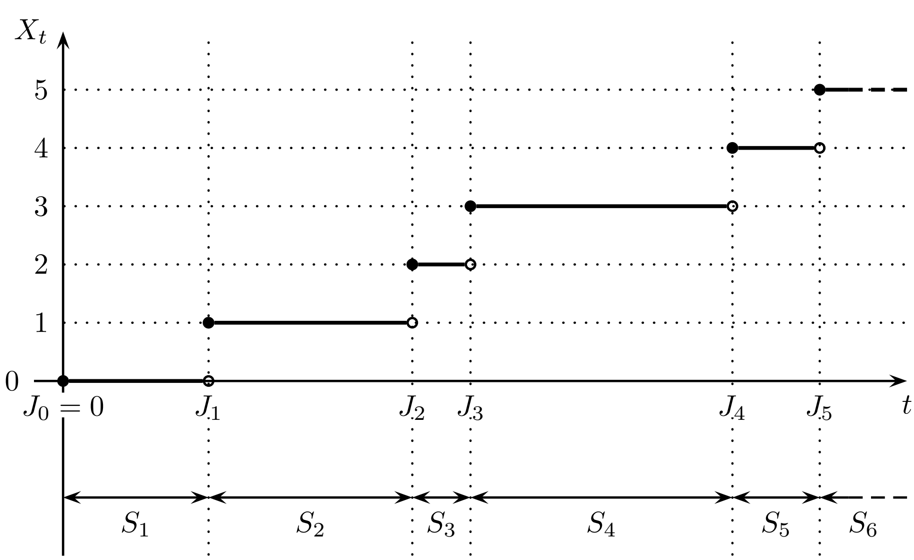
$$0 \xrightarrow{\lambda} 1 \xrightarrow{\lambda} 2 \xrightarrow{\lambda} 3 \xrightarrow{\lambda} 4 \text{ --- - - -}$$

The associated $Q$-matrix is given by

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

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

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




https://doi.org/10.1017/CBO9780511810633.004 Published online by Cambridge University Press

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

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

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

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

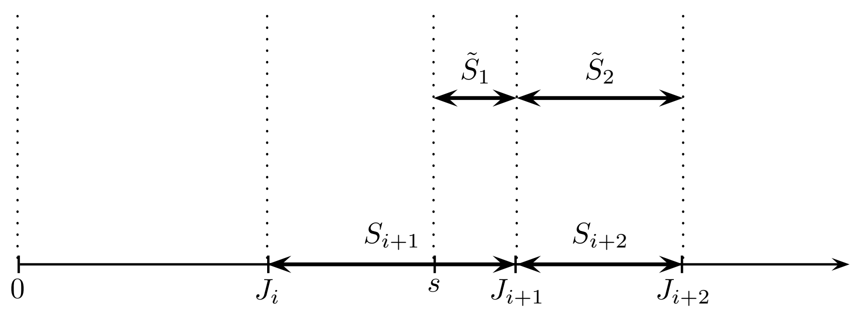
On this event

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

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

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

as shown in the diagram.




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

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

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

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

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

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

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

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

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

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

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

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

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

which implies (b).

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

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

so

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

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

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

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

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

By a simpler argument we also find

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

Since $X_0 = 0$ we have initial conditions

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

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

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

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

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

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

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

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

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

or, in matrix form, for $Q$ as above,

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

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

**Theorem 2.4.4.** *If $(X_t)_{t\ge0}$ and $(Y_t)_{t\ge0}$ are independent Poisson processes of rates $\lambda$ and $\mu$, respectively, then $(X_t+Y_t)_{t\ge0}$ is a Poisson process of rate $\lambda+\mu$.*

*Proof.* We shall use the infinitesimal definition, according to which $(X_t)_{t\ge0}$ and $(Y_t)_{t\ge0}$ have independent increments and, as $h \downarrow 0$, uniformly in $t$,

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

Set $Z_t = X_t + Y_t$. Then, since $(X_t)_{t\ge0}$ and $(Y_t)_{t\ge0}$ are independent, $(Z_t)_{t\ge0}$ has independent increments and, as $h \downarrow 0$, uniformly in $t$,

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

Hence $(Z_t)_{t\ge0}$ is a Poisson process of rate $\lambda+\mu$. $\square$

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

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

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

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

□

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

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

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

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

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

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

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

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

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

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

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

as required. □

https://doi.org/10.1017/CBO9780511810633.004 Published online by Cambridge University Press

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

**Example 2.4.7**

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

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

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

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

**Exercises**

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

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

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

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

https://doi.org/10.1017/CBO9780511810633.004 Published online by Cambridge University Press

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

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

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

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

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

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

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

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

## 2.5 Birth processes

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

$$0 \xrightarrow{q_0} 1 \xrightarrow{q_1} 2 \xrightarrow{q_2} 3 \xrightarrow{q_3} 4 \longrightarrow \cdots$$

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

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

**Example 2.5.1 (Simple birth process)**

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

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

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

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

and setting $r = t - s$

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

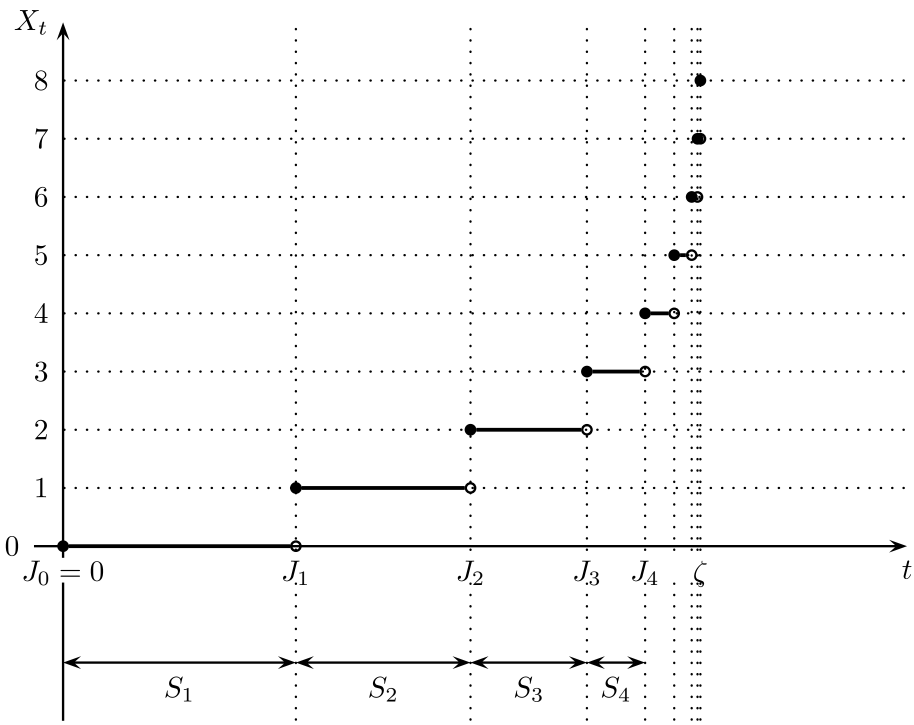
By differentiating we obtain

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

so the mean population size grows exponentially:

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

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




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

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

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

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

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

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

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

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

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

or, in components

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

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

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

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

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

which can be substituted in the equation

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

and this equation solved to give

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

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

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

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

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

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

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

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

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

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

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

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

and

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

which implies (b).

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

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

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

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

so

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

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

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

By a simpler argument we also find

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

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

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

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

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

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

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

### Exercise

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

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

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

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

https://doi.org/10.1017/CBO9780511810633.004 Published online by Cambridge University Press

## 2.6 Jump chain and holding times

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

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

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

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

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

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

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

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

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

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

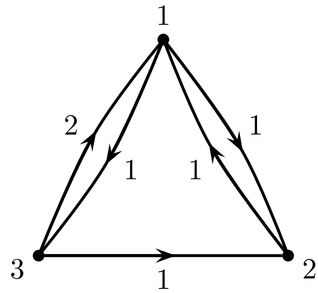
### Example 2.6.1

The $Q$-matrix

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

https://doi.org/10.1017/CBO9780511810633.004 Published online by Cambridge University Press

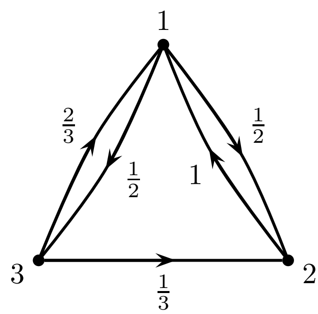
has diagram:




The jump matrix $\Pi$ of $Q$ is given by

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

and has diagram:




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

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

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

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

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

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

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

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

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

https://doi.org/10.1017/CBO9780511810633.004 Published online by Cambridge University Press

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

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

## 2.7 Explosion

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

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

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

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

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

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

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

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

with probability 1. □

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

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

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

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

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

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

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

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

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

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

and

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

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

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

so

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

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

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

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

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

for all $i$. Suppose inductively that

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

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

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

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

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

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

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

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

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

**Exercise**

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

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

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

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

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

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

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

(i) $\lambda = \mu$;

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

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

## 2.8 Forward and backward equations

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

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

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

We come to the key result for continuous-time Markov chains. We shall present first a version for the case of finite state-space, where there is a

https://doi.org/10.1017/CBO9780511810633.004 Published online by Cambridge University Press

simpler proof. In this case there are three alternative definitions, just as for the Poisson process.

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

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

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

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

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

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

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

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

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

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

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

and for $j \neq i$ we have

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

Thus for every state $j$ there is an inequality

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

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

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

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

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

Since $I$ is finite we have

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

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

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

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

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

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

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

proving (c).

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

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

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

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

Suppose (b) holds and set

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

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

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

and

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

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

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

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

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

only now we have an infinite system of differential equations

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

and

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

Therefore

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

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

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

https://doi.org/10.1017/CBO9780511810633.004 Published online by Cambridge University Press

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

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

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

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

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

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

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

The argument used to prove (2.1) also shows that

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

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

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

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

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

Let us suppose inductively that

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

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

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

and the induction proceeds. Hence

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

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

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

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

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

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

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

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

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

now understood as an infinite system of differential equations

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

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

**Lemma 2.8.5.** *We have*

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

or

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

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

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

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

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

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

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

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

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

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

Let us suppose inductively that

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

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

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

and the induction proceeds. Hence

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

## Exercises

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

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

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

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

https://doi.org/10.1017/CBO9780511810633.004 Published online by Cambridge University Press

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

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

## 2.9 Non-minimal chains

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

**Example 2.9.1**

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

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

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

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

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

**Example 2.9.2**

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

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

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

where $Z$ takes values $\pm 1$ with probability $1/2$.

**Example 2.9.3**

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

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

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

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

and

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

Now define

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

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

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

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

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

## 2.10 Appendix: matrix exponentials

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

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

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

**Lemma 2.10.1.** *We have*

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

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

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

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

On the other hand

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

and, by the Cauchy–Schwarz inequality

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

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

(b) For any vector $v$ we have

$$\begin{aligned}&|(Q_1+Q_2)v| \le |Q_1v| + |Q_2v| \le (|Q_1| + |Q_2|)|v|,\\ &|Q_1Q_2v| \le |Q_1||Q_2v| \le |Q_1||Q_2||v|.\end{aligned}$$ □

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

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

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

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

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

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

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

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

is well defined and, indeed, that the power series

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

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

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