5

Markov chains: Introduction

Let $X_n, n = 0, 1, 2, \dots$ be a "random process", taking values in some set I called the *state space*. That is, $X_0, X_1, X_2, \dots$ are random variables with $X_n \in I$ for all n.

Often X_n represents some quantity evolving in time. So far we have been working with random variables taking values which are real numbers of some kind, but there is no problem in considering a more general state space. For example, we might consider processes of the following kind:

- $I = \mathbb{Z}^2$, X_n=position at nth step of a "random walk" on the two-dimensional lattice.
- $I = \{A, B, C, \dots, a, b, c, \dots, ., ?, !, \dots\}$, X_n=nth character in a text or in an email.
- $I = \{C, G, A, T\}$ (representing cytosine, guanine, adenine, thymine, the four bases of DNA), X_n=base appearing in nth position in a DNA sequence.

We will assume that the state space I is finite or countably infinite (i.e. discrete). A **(probability) distribution** on I is a collection $\lambda = (\lambda_i, i \in I)$ with $\lambda_i \geq 0$ for all i, and $\sum \lambda_i = 1$. This is really just the same idea as the probability mass function of a discrete random variable. We will often think of λ as a **row vector**. We will say that a random variable Y taking values in I has distribution λ if $\mathbb{P}(Y = i) = \lambda_i$ for all i.

5.1 Markov chains

Let $X = (X_0, X_1, X_2, \dots)$ be a sequence of random variables taking values in I. The process X is called a **Markov chain** if for any $n \geq 0$ and any $i_0, i_1, \dots, i_{n+1} \in I$,

$$\mathbb{P}\left(X_{n+1} = i_{n+1} \middle| X_n = i_n, \dots, X_0 = i_0\right) = \mathbb{P}\left(X_{n+1} = i_{n+1} \middle| X_n = i_n\right). \tag{5.1}$$

(To be precise, we should restrict (5.1) to cases where these conditional probabilities are well-defined, i.e. where the event $\{X_n = i_n, \dots, X_0 = i_0\}$ has positive probability.)

The Markov chain is called **(time) homogeneous** if in addition $\mathbb{P}\left(X_{n+1} = j \middle| X_n = i\right)$ depends only on i and j, not on n. In that case we write

$$p_{ij} = \mathbb{P}\left(X_{n+1} = j \middle| X_n = i\right)$$

(or we will often write $p_{i,j}$ rather than p_{ij}, according to convenience). The quantities p_{ij} are known as the **transition probabilities** of the chain.

We will work almost always with homogeneous chains. To describe such a chain, it is enough to specify two things:

- the initial distribution λ of X_0. For each $i \in I$, $\lambda_i = \mathbb{P}(X_0 = i)$.
- the **transition matrix** $P = (p_{ij})_{i,j\in I}$.

P is a square (maybe infinite) matrix, whose rows and columns are indexed by I. P is a "stochastic matrix" which means that all its entries are non-negative and every row sums to 1. Equivalently, every row of P is a probability distribution. The ith row of P is the distribution of X_{n+1} given $X_n = i$.

Theorem 5.1. *For $i_0, i_1, \ldots, i_n \in I$,*

$$\boxed{\mathbb{P}\left(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n\right) = \lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n}.}$$

Proof. By the definition of conditional probabilities and cancellations,

$$\begin{aligned}
&\mathbb{P}\left(X_0 = i_0, X_1 = i_1, \ldots, X_n = i_n\right)\\
&\quad = \mathbb{P}\left(X_0 = i_0\right)\mathbb{P}\left(X_1 = i_1 \big| X_0 = i_0\right)\mathbb{P}\left(X_2 = i_2 \big| X_1 = i_1, X_0 = i_0\right) \times \ldots\\
&\qquad\qquad\qquad\qquad\qquad \cdots \times \mathbb{P}\left(X_n = i_n \big| X_{n-1} = i_{n-1}, \ldots, X_0 = i_0\right)\\
&\quad = \mathbb{P}\left(X_0 = i_0\right)\mathbb{P}\left(X_1 = i_1 \big| X_0 = i_0\right)\mathbb{P}\left(X_2 = i_2 \big| X_1 = i_1\right)\ldots\mathbb{P}\left(X_n = i_n \big| X_{n-1} = i_{n-1}\right)\\
&\quad = \lambda_{i_0} p_{i_0 i_1} p_{i_1 i_2} \ldots p_{i_{n-1} i_n},
\end{aligned}$$

where we used the definition of a Markov chain to get the penultimate line. □

If X is a Markov chain with initial distribution λ and transition matrix P, we will sometimes write "$X \sim \text{Markov}(\lambda, P)$".

Markov chains are "memoryless". If we know the current state, any information about previous states is irrelevant to the future evolution of the chain. We can say that "the future is independent of the past, given the present". This is known as the **Markov property**:

$$\begin{aligned}
&\mathbb{P}\left(X_{n+1} \in A_{n+1}, \ldots, X_{n+m} \in A_{n+m} \middle| X_0 \in A_0, \ldots, X_{n-1} \in A_{n-1}, X_n = i\right)\\
&= \mathbb{P}\left(X_{n+1} \in A_{n+1}, \ldots, X_{n+m} \in A_{n+m} \middle| X_n = i\right)\\
&= \mathbb{P}\left(X_1 \in A_{n+1}, \ldots, X_m \in A_{n+m} \middle| X_0 = i\right).
\end{aligned}$$

for all $A_0, \ldots, A_{m+n} \subseteq I$ with $\mathbb{P}(X_0 \in A_0, \ldots, X_{n-1} \in A_{n-1}, X_n = i) > 0$, or equivalently (by the definition of conditional probabilities and cancellations)

$$\begin{aligned}
&\mathbb{P}\left(X_0 \in A_0, \ldots, X_{n-1} \in A_{n-1}, X_{n+1} \in A_{n+1}, \ldots, X_{n+m} \in A_{n+m} \middle| X_n = i\right)\\
&= \mathbb{P}\left(X_0 \in A_0, \ldots, X_{n-1} \in A_{n-1} \middle| X_n = i\right)\mathbb{P}\left(X_{n+1} \in A_{n+1}, \ldots, X_{n+m} \in A_{n+m} \middle| X_n = i\right)\\
&= \mathbb{P}\left(X_0 \in A_0, \ldots, X_{n-1} \in A_{n-1} \middle| X_n = i\right)\mathbb{P}\left(X_1 \in A_{n+1}, \ldots, X_m \in A_{n+m} \middle| X_0 = i\right).
\end{aligned}$$

Notation: it will be convenient to write $\mathbb{P}_i$ for the distribution conditioned on $X_0 = i$. For example $\mathbb{P}_i(X_1 = j) = p_{ij}$. Similarly $\mathbb{E}_i$ for expectation conditioned on $X_0 = i$.

5.2 n-step transition probabilities

Write $p_{ij}^{(n)} = \mathbb{P}\left(X_{k+n} \middle| X_k = i\right)$. This is an **$n$-step transition probability** of the Markov chain.

Theorem 5.2. *(Chapman-Kolmogorov equations)*

(i) $p_{ik}^{(n+m)} = \sum_{j \in I} p_{ij}^{(n)} p_{jk}^{(m)}$.

(ii) $p_{ij}^{(n)} = (P^n)_{i,j}$.

Here P^n is the nth power of the transition matrix. As ever, matrix multiplication is given by $(AB)_{i,j} = \sum_k (A)_{i,k}(B)_{k,j}$, whether the matrices are finite or infinite.

Proof. (i) We condition on X_n, i.e. we consider the partition $\{X_n = j\}$, $j \in I$, and use the Law of Total Probability:

$$\begin{aligned}\mathbb{P}\left(X_{n+m} = k \middle| X_0 = i\right) &= \sum_j \mathbb{P}\left(X_n = j \middle| X_0 = i\right) \mathbb{P}\left(X_{n+m} = k \middle| X_n = j, X_0 = i\right) \\ &= \sum_j \mathbb{P}\left(X_n = j \middle| X_0 = i\right) \mathbb{P}\left(X_{n+m} = k \middle| X_n = j\right)\end{aligned}$$

(using the Markov property)

$$= \sum_j p_{ij}^{(n)} p_{jk}^{(m)}.$$

(ii) For $n = 1$, this holds by definition of P. Inductively, if this holds for any $n \geq 1$,

$$p_{ik}^{(n+1)} = \sum_j p_{ij}^{(n)} p_{jk}^{(1)} = \sum_j (P^n)_{i,j} \, (P)_{j,k} = (P^n P)_{i,k} = \left(P^{n+1}\right)_{i,k}.$$

□

Theorem 5.3. *Let λ be the initial distribution (i.e. the distribution of X_0). Then the distribution of X_1 is λP, and more generally the distribution of X_n is λP^n.*

Here we are thinking of λ as a row vector, so that λP^n is also a row vector; $(\lambda A)_i = \sum_k \lambda_k A_{ki}$ as usual, whether the dimensions are finite or infinite.

Proof. Just condition on the initial state, i.e. apply the Law of Total Probability for the partition $\{X_0 = i\}$, $i \in I$:

$$\begin{aligned}\mathbb{P}\left(X_1 = j\right) &= \sum_i \mathbb{P}\left(X_0 = i\right) \mathbb{P}\left(X_1 = j \middle| X_0 = i\right) \\ &= \sum_i \lambda_i p_{ij} \\ &= (\lambda P)_j,\end{aligned}$$

and similarly for X_n with p_{ij} and P replaced by $p_{ij}^{(n)}$ and P^n. □

Using this result and the Markov property it is easy to get the following property: if $(X_0, X_1, X_2, \dots)$ is a Markov chain with initial distribution λ and transition matrix P, then $(X_0, X_k, X_{2k}, \dots)$ is a Markov chain with initial distribution λ and transition matrix P^k.

Example 5.4 (General two-state Markov chain)**.** Let $I = \{1, 2\}$ and

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

What is $p_{11}^{(n)}$? Two approaches:

(1) P has eigenvalues 1 and $1 - \alpha - \beta$ (check! Every Markov transition matrix has 1 as an eigenvalue – why?). So we can diagonalise:

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

We get $(P^n)_{11} = A + B(1-\alpha-\beta)^n$ for some constants A and B.

Since we know $p_{11}^{(0)} = 1$ and we have $p_{11}^{(1)} = 1 - \alpha$, we can solve for A and B to get

$$p_{11}^{(n)} = \frac{\beta}{\alpha+\beta} + \frac{\alpha}{\alpha+\beta}(1-\alpha-\beta)^n. \tag{5.2}$$

(2) Alternatively, we can condition on the state at step $n - 1$:

$$\begin{aligned} p_{11}^{(n)} &= p_{11}^{(n-1)}(1-\alpha) + p_{12}^{(n-1)}\beta \\ &= p_{11}^{(n-1)}(1-\alpha) + \left(1 - p_{11}^{(n-1)}\right)\beta \\ &= (1-\alpha-\beta)p_{11}^{(n-1)} + \beta. \end{aligned}$$

This gives a linear recurrence relation for $p_{11}^{(n)}$, which we can solve using standard methods to give (5.2) again.

5.3 A few examples

Random walk on a cycle

$I = \{0, 1, 2, \dots, M-1\}$. At each step the walk increases by 1 (mod M) with probability p and decreases by 1 (mod M) with probability $1 - p$. That is,

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

or

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

Simple symmetric random walk on $\mathbb{Z}^d$

$I = \mathbb{Z}^d$. At each step the walk moves from its current site to one of its $2d$ neighbours chosen uniformly at random.

$$p_{ij} = \begin{cases} \frac{1}{2d} & \text{if } |i-j| = 1, \\ 0, & \text{otherwise} \end{cases}$$

where $|i-j| = |i_1 - j_1| + \cdots + |i_d - j_d|$ for states $i = (i_1, \ldots, i_d)$ and $j = (j_1, \ldots, j_d)$.

Card-shuffling

Let I be the set of orderings of 52 cards. We can regard I as the permutation group S_{52}. There are many interesting Markov chains on permutation groups. We can think of shuffling a pack of cards. A simple and not very practical example of a shuffle: at each step, choose a and b independently and uniformly in $\{1, 2, \ldots, 52\}$ and exchange the cards in positions a and b. This gives

$$p_{\alpha\beta} = \begin{cases} \dfrac{2}{52^2} & \text{if } \alpha = \beta\tau \text{ for some transposition } \tau, \\ \dfrac{1}{52} & \text{if } \alpha = \beta, \\ 0 & \text{otherwise.} \end{cases}$$

5.4 Exploring the Markov property

Let us look at a few examples of simple processes where the Markov property holds or fails. We can do this in the context of a simple random walk on $\mathbb{Z}$.

Let X_i be i.i.d. with $\mathbb{P}(X_i = 1) = p$ and $\mathbb{P}(X_i = -1) = 1 - p$.

Let $S_0 = 0$ and $S_n = \sum_{i=1}^n X_i$.

Then:

(1) X_n is a Markov chain. In fact, X_n are i.i.d., which is a stronger property. Given any history, the next state is equal to 1 with probability p and -1 with probability $1-p$. The matrix of the chain X_n (with rows and columns indexed by $\{-1, 1\}$) is $P = \begin{pmatrix} 1-p & p \\ 1-p & p \end{pmatrix}$.

(2) The random walk S_n is also a Markov chain. Its transition probabilities are $p_{i,i+1} = p$ and $p_{i,i-1} = p$ for all $i \in \mathbb{Z}$.

(3) Consider the process $M_n = \max_{0\leq k\leq n} S_k$. Try drawing some possible paths of the process S_n, and the corresponding paths of the "maximum process" M_n. Is this maximum process a Markov chain?

We can consider two different ways of arriving at the same state. Suppose we observe $(M_0, \dots, M_4) = (0, 0, 0, 1, 2)$. This implies $S_4 = 2$ (the maximum process has just increased, so now the walk must be at its current maximum.) In that case, if the random walk moves up at the next step, then the maximum will also increase. So

$$\mathbb{P}(M_5 = 3|(M_0, \dots, M_4) = (0, 0, 0, 1, 2)) = p.$$

Suppose instead that $(M_0, \dots, M_4) = (0, 1, 2, 2, 2)$. In that case, both $S_4 = 2$ and $S_4 = 0$ are possible (check! – find the corresponding paths). As a consequence, sometimes the maximum will stay the same at the next step, even when the random walk moves up. So we have

$$\mathbb{P}(M_5 = 3|(M_0, \dots, M_4) = (0, 1, 2, 2, 2)) < p.$$

We see that the *path* to $M_4 = 2$ affects the conditional probability of the next step of the process. So M_n is *not* a Markov chain.

The next result gives a criterion for the Markov property to hold.

Proposition 5.5. *Suppose that $(Y_n, n \geq 0)$ is a random process, and for some function f we can write, for each n,*

$$Y_{n+1} = f(Y_n, X_{n+1}),$$

where X_{n+1} is independent of $Y_0, Y_1, \dots, Y_n$. Then (Y_n) is a Markov chain.

Proof. The idea is that to update the chain, we use only the current state and some "new" randomness. We have

$$\begin{aligned}
&\mathbb{P}\big(Y_{n+1} = i_{n+1}\big|Y_n = i_n, \dots, Y_0 = i_0\big) \\
&\qquad = \mathbb{P}\left(f(i_n, X_{n+1}) = i_{n+1}\big|Y_n = i_n, \dots, Y_0 = i_0\right) \\
&\qquad = \mathbb{P}\left(f(i_n, X_{n+1}) = i_{n+1}\right) \quad \text{(by independence of } X_{n+1} \text{ from } Y_0, \dots, Y_n) \\
&\qquad = \mathbb{P}\left(f(i_n, X_{n+1}) = i_{n+1}\big|Y_n = i_n\right) \quad \text{(by independence of } X_{n+1} \text{ from } Y_n) \\
&\qquad = \mathbb{P}\left(Y_{n+1} = i_{n+1}\big|Y_n = i_n\right).
\end{aligned}$$

□

For example, for the simple random walk above, we can put $S_{n+1} = f(S_n, X_{n+1})$, where $f(s, x) = s + x$. For the card-shuffling example in the previous section, if $Y_n \in S_{52}$ is the permutation after step n, we can put $Y_{n+1} = f(Y_n, X_{n+1})$ where for a permutation β and a transposition τ, $f(\beta, \tau) = \beta\tau$, and where (X_n) is an i.i.d. sequence in which each member is uniform in the set of transpositions.

5.5 Class structure

Let $i, j \in I$. We say that "i leads to j" and write "$i \to j$" if $\mathbb{P}_i(X_n = j) > 0$ for some $n \geq 0$, i.e. $p_{ij}^{(n)} > 0$ for some $n \geq 0$.

If $i \to j$ and $j \to i$ then we say "i communicates with j" and write $i \leftrightarrow j$.

Then $\leftrightarrow$ is an equivalence relation (check!). It partitions the state space I into **communicating classes**.

A class C is called **closed** if $p_{ij} = 0$ whenever $i \in C, j \notin C$, or equivalently $i \not\to j$ for any $i \in C, j \notin C$. Once the chain enters a closed class, it can never escape from it. If $\{i\}$ is a closed class then $p_{ii} = 1$, and i is called an **absorbing state**. If C is not closed it is called **open**.

A chain (or more precisely a transition matrix) for which I consists of a single communicating class is called **irreducible**. Equivalently, $i \to j$ for all $i, j \in I$.

Example 5.6. Let $I = \{1, 2, 3, 4, 5, 6, 7\}$. The communicating classes for the transition matrix

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

are $\{1, 2, 3\}$, $\{4\}$, $\{5, 6\}$ and $\{7\}$. The closed classes are $\{5, 6\}$ and $\{7\}$ (so 7 is an absorbing state). Draw a diagram to visualise the chain!

5.6 Periodicity

Consider the transition matrix

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

Again, draw a diagram to visualise the chain. Note that $p_{ii}^{(n)} = 0$ whenever n is odd.

For a general chain and a state $i \in I$, the **period** of the state i is defined to be the greatest common divisor of the set $\left\{n \geq 1 \colon p_{ii}^{(n)} > 0\right\}$. (If $p_{ii}^{(n)} = 0$ for all $n > 0$, then the period is not defined). All the states in the chain above have period 2.

In Example 5.6, states 1, 2 and 3 have period 3, the period of state 4 is undefined, 5 and 6 have period 2 and the absorbing state 7 has period 1.

i is called **aperiodic** if this g.c.d. is 1 (and otherwise **periodic**). Equivalently (check!), i is aperiodic if $p_{ii}^{(n)} > 0$ for all sufficiently large n.

Fact. *All states in a communicating class have the same period.*

Proof. Suppose $i \leftrightarrow j$ and $d|n$ whenever $p_{ii}^{(n)} > 0$.

Since i and j communicate, we can find a and b with $p_{ij}^{(a)} > 0$ and $p_{ji}^{(b)} > 0$. Then also $p_{ii}^{(a+b)} > 0$.

Suppose $p_{jj}^{(m)} > 0$. Then also $p_{ii}^{(a+m+b)} > 0$.

Then $d|a+b$ and $d|a+m+b$, so also $d|m$.

This demonstrates that the sets $\left\{n \geq 1 \colon p_{ii}^{(n)} > 0\right\}$ and $\left\{m \geq 1 \colon p_{jj}^{(m)} > 0\right\}$ have the same divisors, and hence the same greatest common divisor. □

In particular, if a chain is irreducible, then all states have the same period. If this period is 1, we say that the chain is aperiodic (otherwise we say the chain is periodic).

Remark. Notice that both irreducibility and periodicity are "structural properties" in the following sense: they depend only on which transition probabilities p_{ij} are positive and which are zero, not on the particular values taken by those which are positive.

Example. Look back at the three examples in Section 5.3 and consider which are irreducible and which are periodic.

The random walk on the cycle is irreducible (since every site is accessible from every other). It has period 2 if M is even, and is aperiodic if M is odd.

The random walk on $\mathbb{Z}^d$ is irreducible and has period 2 for any d.

The card-shuffling chain is irreducible (because the set of transpositions is a set of generators for the group S_{52}). It is aperiodic, since there is a positive transition probability from any state to itself.

Remark. Later we will show results about convergence to equilibrium for Markov chains. The idea will be that after a long time, a Markov chain should more or less "forget where it started". There are essentially two reasons why this might not happen: (a) periodicity; for example if a chain has period 2, then it alternates between, say, "odd" and "even" states; even an arbitrarily long time, the chain will still remember whether it started at an "odd" or "even" state. (b) lack of irreducibility. A chain with more than one closed class can never move from one to the other, and so again will retain some memory of where it started, for ever. When we prove results about convergence to equilibrium, it will be under the condition that the chain is irreducible and aperiodic.

5.7 Hitting probabilities

Let A be a subset of the state space I. Define $h_i^A = \mathbb{P}_i\left(X_n \in A \text{ for some } n \geq 0\right)$, the **hitting probability** of A starting from state i.

If A is a closed class, we might call h_i^A the **absorption probability**.

Example. Let $I = \{1, 2, 3, 4\}$ and

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

Starting from 2, what is the probability of absorption at 4?

Write $h_i = \mathbb{P}_i(X_n = 4 \text{ for some } n \geq 0)$. Then $h_4 = 1$, and $h_1 = 0$ since 1 is itself absorbing. Also by conditioning on the first jump and applying the Markov property, we have

$$\begin{aligned}
h_2 =& \mathbb{P}_2(X_1 = 1)\mathbb{P}_2(X_n = 4 \text{ for some } n \geq 1 | X_1 = 1) \\
&+ \mathbb{P}_2(X_1 = 3)\mathbb{P}_2(X_n = 4 \text{ for some } n \geq 1 | X_1 = 3) \\
=& \tfrac{1}{2}h_1 + \tfrac{1}{2}h_3, \\
h_3 =& \tfrac{1}{2}h_2 + \tfrac{1}{2}h_4.
\end{aligned}$$

Solving, we get $h_2 = 1/3$ and $h_3 = 2/3$.

Theorem 5.7. *The vector of hitting probabilities $\left(h_i^A, i \in I\right)$ is the minimal non-negative solution to the equations*

$$h_i^A = \begin{cases} 1 & \text{if } i \in A \\ \sum_j p_{ij} h_j^A & \text{if } i \notin A \end{cases}. \tag{5.3}$$

Here by "minimal" we mean that if $(x_i, i \in I)$ is another non-negative solution to the system (5.3), then $h_i \leq x_i$ for all i.

Proof. To see that h_i^A satisfies (5.3), we condition on the first jump of the process. If $i \notin A$, then

$$\begin{aligned}
h_i^A &= \mathbb{P}_i\left(X_n \in A \text{ for some } n \geq 0\right) \\
&= \mathbb{P}_i\left(X_n \in A \text{ for some } n \geq 1\right) \\
&= \sum_j \mathbb{P}_i(X_1 = j)\mathbb{P}\left(X_n \in A \text{ for some } n \geq 1 \middle| X_0 = i, X_1 = j\right) \\
&= \sum_j p_{ij}\mathbb{P}\left(X_n \in A \text{ for some } n \geq 1 \middle| X_1 = j\right) \\
&= \sum_j p_{ij} h_j^A.
\end{aligned}$$

To obtain the penultimate line we applied the Markov property. Meanwhile for $i \in A$, $h_i^A = 1$ by definition. So indeed (5.3) holds.

To prove minimality, suppose $(x_i, i \in I)$ is any non-negative solution to (5.3). We want to show that $h_i^A \leq x_i$ for all i.

We make the following claim: for any $M \in \mathbb{N}$, and for all i,

$$x_i \geq \mathbb{P}_i(X_n \in A \text{ for some } n \leq M). \tag{5.4}$$

We will prove (5.4) by induction on M.

The case $M = 0$ is easy; the LHS is 1 for $i \in A$, while the RHS is 0 for $i \notin A$.

For the induction step, suppose that for all i, $x_i \geq \mathbb{P}_i(X_n \in A \text{ for some } n \leq M-1)$. If $i \in A$, then again $x_i = 1$ and (5.4) is clear. If $i \notin A$, then

$$\mathbb{P}_i(X_n \in A \text{ for some } n \leq M) = \sum_j p_{ij}\mathbb{P}_i(X_n \in A \text{ for some } n \in \{1, 2, \ldots, M\} | X_1 = j)$$

$$
\begin{aligned}
&= \sum_j p_{ij}\mathbb{P}_j\left(X_n \in A \text{ for some } n \in \{0,1,\dots,M-1\}\right) \\
&\le \sum_j p_{ij}x_j \\
&= x_i,
\end{aligned}
$$

and the induction step is complete. Hence (5.4) holds for all i and M as desired. Then, using the fact that the sequence of events in (5.4) is increasing in M, we have

$$
\begin{aligned}
x_i &\ge \lim_{M\to\infty} \mathbb{P}_i(X_n \in A \text{ for some } n \le M) \\
&= \mathbb{P}_i\big(\bigcup_M \{X_n \in A \text{ for some } n \le M\}\big) \\
&= \mathbb{P}_i(X_n \in A \text{ for some } n) \\
&= h_i^A.
\end{aligned}
$$

□

Important example: "Gambler's ruin"

Let $I = \{0, 1, 2 \dots\}$. Let $p \in (0,1)$ and $q = 1-p$, and consider the transition probabilities given by

$$
\begin{aligned}
p_{00} &= 1 \\
p_{i,i-1} &= q \text{ for } i \ge 1 \\
p_{i,i+1} &= p \text{ for } i \ge 1.
\end{aligned}
\tag{5.5}
$$

The name "gambler's ruin" comes from the interpretation where the state is the current capital of a gambler, who repeatedly bets 1 (against an infinitely rich bank). Will the gambler inevitably go broke? But chains like this come up in a wide range of settings. Chains on $\mathbb{Z}_+$ in which all transitions are steps up and down by 1 are called "birth-and-death chains" (modelling the size of a population). This is one of the simplest examples.

Let $h_i = \mathbb{P}_i(\text{hit } 0)$. To find h_i, we need the minimal non-negative solution to

$$h_0 = 1 \tag{5.6}$$

$$h_i = ph_{i+1} + qh_{i-1} \text{ for } i \ge 1. \tag{5.7}$$

If $p \ne q$, (5.7) has general solution

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

We look at three cases:

$\boxed{p<q}$ Jumps downwards are more likely than jumps upward. From (5.6), $A + B = 1$. Then for minimality, we take $A = 1$ and $B = 0$, since $\left(\frac{q}{p}\right)^i \ge 1$ for all i.

We obtain $h_i = 1$ for all i. So with probability 1, the chain will hit 0.

$\boxed{p > q}$ Again $A + B = 1$. Also $\left(\frac{q}{p}\right)^i \to 0$ as $i \to \infty$, so we need $A \geq 0$ for a non-negative solution. Then for a minimal solution, we will want $A = 0$, $B = 1$, since $1 \geq \left(\frac{q}{p}\right)^i$ for all i.

Hence $h_i = \left(\frac{q}{p}\right)^i$. The chain has a positive probability of "escaping to infinity".

$\boxed{p = q}$ Now the general solution of (5.7) is $h_i = A + Bi$. From $i = 0$ we get $A = 1$. For non-negativity we need $B \geq 0$, and then for minimality $B = 0$. We get $h_i = 1$ again. Now there is no drift, but still with probability 1 the chain will hit 0 eventually.

Remark. Notice that we could have seen $h_i = \alpha^i$ for some α, by a direct argument. Since the chain can only descend by one step at a time,

$$\mathbb{P}_i(\text{hit } 0) = \mathbb{P}_i(\text{hit } i-1)\mathbb{P}_{i-1}(\text{hit } i-2)\ldots\mathbb{P}_1(\text{hit } 0), \tag{5.8}$$

and all terms in the product are the same, since the transition probabilities are the same at every level.

5.8 Recurrence and transience

If the chain starts in state i, what is the chance that it returns to i at some point in the future? We can distinguish two possibilities:

(1)
$$\mathbb{P}_i(X_n = i \text{ for some } n \geq 1) = p < 1.$$

Then the total number of visits to i has geometric distribution with parameter $1 - p$ (since each time we return to i, we have chance $1 - p$ of never returning again). We have
$$\mathbb{P}_i(\text{hit } i \text{ infinitely often}) = 0.$$

The state i is called **transient**.

(2)
$$\mathbb{P}_i(X_n = i \text{ for some } n \geq 1) = 1.$$

Then
$$\mathbb{P}_i(\text{hit } i \text{ infinitely often}) = 1.$$

The state i is called **recurrent**.

The definition is very simple, but the concept of recurrence and transience is extremely rich (mainly for infinite chains).

There is an important criterion for recurrence and transience in terms of the transition probabilities:

Theorem 5.8. *State i is recurrent if and only if $\sum_{n=0}^{\infty} p_{ii}^{(n)} = \infty$.*

Proof. The total number of visits to i is $\sum_{n=0}^{\infty} \mathbf{1}\{X_n = i\}$ which has expectation

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

If i is transient, the number of visits to i is geometric with parameter $1-p$, and hence with mean $\frac{1}{1-p} < \infty$.

On the other hand if i is recurrent, the number of visits to i is infinite with probability 1, and so has mean ∞.

This gives the statement of the theorem. □

Theorem 5.9. *(a) Let C be a communicating class. Either all states in C are recurrent, or all are transient (so we may refer to the whole class as transient or recurrent).*

(b) Every recurrent class is closed.

(c) Every finite closed class is recurrent.

Proof. Exercises – see example sheet 3. For part (a), use Theorem 5.8 to show that if i is recurrent and $i \leftrightarrow j$, then j is also recurrent. □

The theorem tells us that recurrence and transience are quite boring for finite chains: state i is recurrent if and only if its communicating class is closed. But infinite chains are more interesting! An infinite closed class may be either transient or recurrent.

5.9 Random walk in $\mathbb{Z}^d$

Consider a simple symmetric random walk on the d-dimensional integer lattice. This is a Markov chain with state space $\mathbb{Z}^d$ and transition probabilities $p_{xy} = 1/2d$ if $|x-y| = 1$, and $p_{xy} = 0$ otherwise. The chain is irreducible, with period 2.

In this section we will show that the random walk is recurrent when $d = 1$ or $d = 2$ but transient in higher dimensions.

5.9.1 d=1

The analysis after (5.7) (for $p = q = 1/2$) shows us that for the simple symmetric random walk on $\mathbb{Z}$, the hitting probability of 0 from any $i > 0$ is 1. By symmetry, the same is true from any negative state. This shows that starting from 0, the probability of returning to 0 is 1. Hence state 0 is recurrent (and so by irreducibility the whole chain is recurrent).

An alternative approach uses Theorem 5.8. This gives a good warm-up for the approach we will use in higher dimensions.

We need to show that $\sum_{n=0}^{\infty} p_{00}^{(n)} = \infty$. We will use Stirling's formula, which tells us that

$$n! \sim \sqrt{2\pi}\, n^{n+1/2} e^{-n} \text{ as } n \to \infty. \tag{5.9}$$

(The constant $\sqrt{2\pi}$ will not be important.)

Suppose $X_0 = 0$. If n is odd, then $\mathbb{P}_0(X_n = 0) = 0$, since the chain has period 2. For $X_{2m} = 0$ we need m "ups" and m "downs" in the first $2m$ steps. Applying Stirling's formula to the binomial probability we obtain

$$\begin{aligned} p_{00}^{(2m)} &= \binom{2m}{m} \left(\frac{1}{2}\right)^{2m} \\ &= \frac{(2m)!}{m!m!} \left(\frac{1}{2}\right)^{2m} \\ &\sim \frac{1}{\sqrt{\pi}} \frac{1}{m^{1/2}}. \end{aligned} \tag{5.10}$$

Since $\sum m^{-1/2} = \infty$, we have $\sum p_{00}^{(n)} = \infty$ and the chain is recurrent.

Exercise. *Use Stirling's formula to show that if $p \neq q$, then the chain is transient. (We could also deduce this from the hitting probability analysis after (5.7).)*

5.9.2 d=2

Why should the walk be recurrent in 1 and 2 dimensions but transient in 3 dimensions? An intuitive answer is as follows. A d-dimensional random walk behaves in some sense like d independent 1-dimensional walks. For the d-dimensional walk to be back at the origin, we require all d of the 1-dimensional walks to be at 0. From (5.10), the probability that a 1-dimensional walk is at 0 decays like $m^{-1/2}$. Hence the probability that a 2-dimensional walk is at the origin decays like m^{-1}, which sums to infinity, leading to recurrence, while the corresponding probability for a 3-dimensional walk decays like $m^{-3/2}$ which has finite sum, leading to transience.

In two dimensions we can make this precise in a very direct way. Let X_n be the walk in $\mathbb{Z}^2$ and consider its projections onto the diagonal lines $x = y$ and $x = -y$ in the plane.

Each step of the walk increases or decreases the projection onto $x = y$ by $1/\sqrt{2}$, and also increases or decreases the projection onto $x = -y$ by $1/\sqrt{2}$. All four possibilities are equally likely.

Hence if we write W_n^+ and W_n^- for the two projections of X_n, we have that the processes W_n^+ and W_n^- are independent of each other, and both of them are simple symmetric random walks on $2^{-1/2}\mathbb{Z}$.

Then we have

$$\begin{aligned} \mathbb{P}(X_{2m} = 0) &= \mathbb{P}(W_{2m}^+ = 0)\mathbb{P}(W_{2m}^- = 0) \\ &\sim \left(\frac{1}{\sqrt{\pi}} \frac{1}{m^{1/2}}\right)^2 \\ &= \frac{1}{\pi m}. \end{aligned}$$

Hence $\sum p_{00}^{(2m)} = \infty$ and the walk is recurrent.

5.9.3 d=3

The trick from the previous section does not work in $d = 3$, so we need to do a little more combinatorics. As the walk has period 2 we have a positive chance of return to the origin only when n is even. Each step is $\pm e_1, \pm e_2$ or $\pm e_3$ where $e_i, i = 1, 2, 3$ are the three unit coordinate vectors. To return to the origin after $2m$ steps, we should have made, say, i steps in each of the directions $\pm e_1$, j steps in each of the directions $\pm e_2$, and k steps in each of the directions $\pm e_3$ for some i, j, k with $i + j + k = m$. Considering all the possible orderings of these steps among the first $2m$ steps of the walk, we get

$$\begin{aligned}
p_{00}^{(2m)} &= \sum_{\substack{i,j,k\geq 0\\ i+j+k=m}} \frac{(2m)!}{i!^2 j!^2 k!^2} \left(\frac{1}{6}\right)^{2m} \\
&= \binom{2m}{m} \left(\frac{1}{2}\right)^{2m} \sum_{\substack{i,j,k\geq 0\\ i+j+k=m}} \binom{m}{i,j,k}^2 \left(\frac{1}{3}\right)^{2m} \\
&\leq \binom{2m}{m} \left(\frac{1}{2}\right)^{2m} \left(\sum_{\substack{i,j,k\geq 0\\ i+j+k=m}} \binom{m}{i,j,k} \left(\frac{1}{3}\right)^{m} \right) \max_{\substack{i,j,k\geq 0\\ i+j+k=m}} \binom{m}{i,j,k} \left(\frac{1}{3}\right)^{m}. \qquad (5.11)
\end{aligned}$$

Here, if $i + j + k = m$, we write $\binom{m}{i,j,k} = \frac{m!}{i!j!k!}$. Note that

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

since it is the sum of the mass function of a "trinomial$(1/3, 1/3, 1/3)$" distribution (consider the number of ways of putting 3 balls into m boxes).

If m is divisible by 3, say $m = 3r$, then it is easy to check that the max in (5.11) is attained when $i = j = k = r$, giving

$$\begin{aligned}
p_{00}^{(2m)} &\leq \binom{2m}{m} \left(\frac{1}{2}\right)^{2m} \binom{m}{m/3, m/3, m/3} \left(\frac{1}{3}\right)^{m} \\
&\sim \frac{1}{\sqrt{2\pi}} \frac{1}{m^{1/2}} \times \frac{1}{2\pi} \frac{1}{m} \\
&\sim \frac{1}{(2\pi)^{3/2}} m^{-3/2},
\end{aligned}$$

where we used Stirling's formula again for the last line. Hence we have $\sum_{r=0}^{\infty} p_{00}^{(6r)} < \infty$.

Note also that $p_{00}^{(6r)} \geq \left(\frac{1}{6}\right)^2 p_{00}^{(6r-2)}$ and $p_{00}^{(6r)} \geq \left(\frac{1}{6}\right)^4 p_{00}^{(6r-4)}$, so overall, $\sum_{n=0}^{\infty} p_{00}^{(n)} < \infty$, and the walk is transient.

5.9.4 $\mathbf{d \geq 4}$

If we have a walk on $\mathbb{Z}^d$ for $d \geq 4$, we can obtain from it a walk on $\mathbb{Z}^3$ by looking only at the first 3 coordinates, and ignoring any transitions that do not change them. Since we know that a walk on $\mathbb{Z}^3$ only visits the origin finitely often, the same must be true for the walk in higher dimensions also. Hence we have transience for all $d \geq 3$.

5.9.5 Mean hitting time

Let $H^A = \inf\{n \geq 0 \colon X_n \in A\}$, the first hitting-time of the set A, with the convention that $H^A = \infty$ if $X_n \notin A$ for all $n \geq 1$. In fact if h_i^A is the hitting probability defined above, then $h_i^A = \mathbb{P}_i(H^A < \infty)$.

Let $k_i^A = \mathbb{E}_i(H^A)$, the mean hitting time of A from i. If $h_i^A < 1$, then $\mathbb{P}_i(H^A = \infty) > 0$ and certainly $k_i^A = \infty$. Also maybe $k_i^A = \infty$ even when $h_i^A = 1$.

Theorem 5.10. *The vector of mean hitting times $k^A = (k_i^A, i \in S)$ is the minimal non-negative solution to*

$$k_i^A = \begin{cases} 0 & \text{if } i \in A \\ 1 + \sum_j p_{ij} k_j^A & \text{if } i \notin A \end{cases}.$$

Proof. "Condition on the first jump" again. For $i \notin A$,

$$\begin{aligned} k_i^A = \mathbb{E}_i(H^A) &= \sum_j \mathbb{E}_i(H^A | X_1 = j) \mathbb{P}_i(X_1 = j) \\ &= \sum_j p_{ij}(1 + k_j^A) \\ &= 1 + \sum_j p_{ij} k_j^A. \end{aligned}$$

For the minimality, one can use a similar idea to that at (5.4) in the proof of Theorem 5.7 above. Specifically, one can show by induction that if (y_i) is any non-negative solution to the recursions, then $y_i \geq \mathbb{E}_i \min\left(H^A, m\right)$ for all $m \geq 0$; we omit the details. □

5.9.6 Gambler's ruin, continued

What is the expected hitting time of 0 from state i in the gambler's ruin chain at (5.5)?

Let k_i be the mean time to hit 0 starting from i. We give brief details (of course, one can be more formal!).

If $k_i < \infty$, one can see that $k_i = \beta i$ for some β, since (compare the remark at (5.8) above),

$$\mathbb{E}_i(\text{time to } 0) = \mathbb{E}_i(\text{time to } i-1) + \mathbb{E}_{i-1}(\text{time to } i-2) + \cdots + \mathbb{E}_1(\text{time to } 0).$$

To satisfy the recursion in Theorem (5.10), we need

$$k_i = 1 + qk_{i-1} + pk_{i+1}$$

which leads to $(q-p)\beta = 1$. We obtain:

$\boxed{p < q}$ $k_i = \frac{1}{q-p} i$; the chain takes a finite time on average to hit 0.

$\boxed{p > q}$ We already know $h_i < 1$, so certainly $k_i = \infty$.

$\boxed{p = q}$ There is no suitable β, so $k_i = \infty$ here also, even though $h_i = 1$. The chain hits 0 with probability 1, but the mean time to arrive there is infinite.

5.10 Null recurrence and positive recurrence

Define m_i, the **mean return time** to a state i by

$$\begin{aligned} m_i &:= \mathbb{E}_i \left(\inf\{n \geq 1 : X_n = i\} \right) \\ &= 1 + \sum p_{ij} k_j^{\{i\}} \end{aligned} \tag{5.12}$$

(where $k_j^{\{i\}}$ is the mean hitting time of i starting from j).

This quantity will be particularly important when we consider equilibrium behaviour of a Markov chain – loosely speaking, the long-run proportion of time spent in state i ought to be the reciprocal of the mean return time.

If i is transient, then certainly $m_i = \infty$ (since the return time itself is infinite with positive probability).

If i is recurrent, then the return time is also finite, but nonetheless the mean could be infinite.

If i is recurrent but $m_i = \infty$, the state i is said to be **null recurrent**.

If $m_i < \infty$ then the state i is said to be **positive recurrent**.

For similar reasons to those in Theorem 5.9, null recurrence and positive recurrence are *class properties*; if one state in a communicating class is null (resp. positive) recurrent, then every state in the class is null (resp. positive) recurrent.

If the chain is irreducible, we can therefore call the whole chain either transient, or null recurrent, or positive recurrent.

6

Markov chains: stationary distributions and convergence to equilibrium

6.1 Stationary distributions

Let $\pi = (\pi_i, i \in I)$ be a distribution on the state space I.

We say that π is a **stationary distribution**, or **invariant distribution**, or **equilibrium distribution**, for the transition matrix P if

$$\boxed{\pi P = \pi}.$$

That is, for all j, $\pi_j = \sum_i \pi_i p_{ij}$. The row vector π is a left eigenvector for the matrix P, with eigenvalue 1.

If X_0 has distribution π, then we know that X_n has distribution πP^n. Hence if π is stationary, then X_n has distribution π for all n. It follows that the sequence

$$(X_n, X_{n+1}, X_{n+2}, \dots)$$

has the same distribution as

$$(X_0, X_1, X_2, \dots)$$

for any n.

6.2 Main theorems

Theorem 6.1 (Existence and uniqueness of stationary distributions)**.** *Let P be an irreducible transition matrix.*

(a) P has a stationary distribution if and only if P is positive recurrent.

(b) In that case, the stationary distribution π is unique, and is given by $\pi_i = 1/m_i$ for all i (where m_i is the mean return time to state i defined at (5.12)).

Theorem 6.2 (Convergence to equilibrium)**.** *Suppose P is irreducible and aperiodic, with stationary distribution π. If X_n is a Markov chain with transition matrix P and any initial distribution, then for all $j \in I$,*

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

In particular,

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

Theorem 6.3 (Ergodic theorem)**.** *Let P be irreducible. Let $V_i(n)$ be the number of visits to state i before time n, that is*

$$V_i(n) = \sum_{r=0}^{n-1} I\left(X_r = i\right).$$

Then for any initial distribution, and for all $i \in I$,

$$\frac{V_i(n)}{n} \to \frac{1}{m_i} \text{ almost surely, as } n \to \infty.$$

That is,

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

The ergodic theorem concerns the "long-run proportion of time" spent in a state.

In the positive recurrent case, $1/m_i = \pi_i$ where π is the stationary distribution, so the ergodic theorem says that (with probability 1) the long-run proportion of time spent in a state is the stationary probability of that state.

In the null-recurrent or transient case, $1/m_i = 0$, so the ergodic theorem says that with probability 1 the long-run proportion of time spent in a state is 0.

We can see the ergodic theorem as a generalisation of the strong law of large numbers. If X_n is an i.i.d. sequence, then the strong law tells us that, with probability 1, the long run proportion of entries in the sequence which are equal to i is equal to the probability that any given entry is equal to i. The ergodic theorem can be seen as extending this to the case where X_n is not i.i.d. but is a Markov chain.

6.3 Examples of stationary distributions

Example 6.4. Let $P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}$. (Draw the diagram of the chain.)

For π to be stationary, we need

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

One of these equations is redundant, and we need the added relation $\pi_1 + \pi_2 + \pi_3 = 1$ to normalise the solution (so that π is a distribution).

Solving, we obtain $(\pi_1, \pi_2, \pi_3) = (1/5, 2/5, 2/5)$.

Correspondingly, the vector of mean return times is given by $(m_1, m_2, m_3) = (5, 5/2, 5/2)$.

Note that this chain is aperiodic and irreducible. Hence, from any initial state, the distribution at time n converges to π as $n \to \infty$. For example, $p_{11}^{(n)} \to 1/5$ as $n \to \infty$.

By the way, be careful to solve the equation $\pi P = \pi$ and not to solve $P\pi = \pi$ by mistake! For any transition matrix P, the equation $P\pi = \pi$ is solved by any vector π all of whose entries are the same (why is this?) which could trap you into thinking that the uniform distribution is stationary, which, of course, is not the case in general. We want the left eigenvector, rather than the right eigenvector.

Example 6.5. Recall the example of a simple symmetric random walk on a cycle of size M in Section 5.3. The distribution π with $\pi_i = 1/M$ for all i is stationary, since it solves

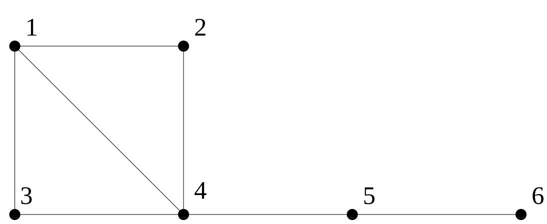
$$\pi_i = \tfrac{1}{2}\pi_{i+1} + \tfrac{1}{2}\pi_{i-1}$$

for each i. Because of the symmetry of the chain, it is not surprising that the stationary distribution is uniform.

Is it the true that $p_{00}^{(n)} \to 1/M$ as $n \to \infty$? If M is odd, then the chain is aperiodic (check this!), so the answer is yes.

However, if M is even then the chain has period 2. Then $p_{00}^{(n)} = 0$ whenever n is odd. In fact $p_{00}^{(2m)} \to 2\pi_0 = 2/M$ as $m \to \infty$ (exercise; consider the 2-step chain X_0, X_2, $X_4, \dots$ on the subset of the state space which consists just of even sites. Is it irreducible? What is its stationary distribution?)

Example 6.6 (Random walk on a graph)**.** A "graph" in the combinatorial sense is a collection of *vertices* joined by *edges*. For example, the following graph has 6 vertices and 7 edges.



Let I be the set of vertices. Two vertices are *neighbours* in the graph if they are joined by an edge. The *degree* of a vertex is its number of neighbours. Let d_i be the degree of vertex i. In the graph above, the vector of vertex degrees is $(d_i, i \in I) = (3, 2, 2, 4, 2, 1)$.

Assume $d_i > 0$ for all i. A *random walk* on the graph is a Markov chain with state space I, evolving as follows; if i is the current vertex, then at the next step move to each of the neighbours of i with probability $1/d_i$.

Assume irreducibility of the chain (equivalently, that there is a path between any two vertices in the graph). Then the stationary distribution of the chain π is unique.

In fact, the stationary probability of a vertex is proportional to its degree. To show this, we will check that $dP = d$ where d is the vector of vertex degrees and P is the transition matrix of the chain:

$$d_j = \sum_i \mathbf{1}(i \text{ is a neighbour of } j)$$

$$= \sum_i d_i \frac{1}{d_i} \mathbf{1}(i \text{ is a neighbour of } j)$$
$$= \sum_i d_i p_{ij},$$

as required.

To obtain the stationary distribution we simply need to normalise d. So we obtain $\pi_i = d_i / \sum_j d_j$.

For the graph above, $\sum_j d_j = 14$, and we obtain

$$\pi = \left(\frac{3}{14}, \frac{1}{7}, \frac{1}{7}, \frac{2}{7}, \frac{1}{7}, \frac{1}{14} \right).$$

From this we can deduce the mean return times. For example, $m_1 = 1/\pi_1 = 14/3$.

Notice that the chain is aperiodic. As a result, we also have convergence to the stationary distribution. For example, starting from any initial distribution, the probability that the walk is at vertex 1 at step n converges to $3/14$ as $n \to \infty$.

Example 6.7 (One-dimensional random walk)**.** Consider again the familiar example of a one-dimensional random walk. Let $I = \{0, 1, 2, \dots\}$ and let

$$\begin{aligned} p_{i,i+1} &= p \text{ for } i \geq 0, \\ p_{i,i-1} &= q = 1 - p \text{ for } i \geq 1, \\ p_{0,0} &= q. \end{aligned}$$

If $p > q$, we found previously that the walk is transient, so no stationary distribution will exist.

If $p = q$, the walk is recurrent, but the mean return time is infinite, so again there is no stationary distribution.

If $p < q$, the walk is positive recurrent. For stationarity, we need $\pi_i = \pi_{i-1} p + \pi_{i+1} q$ for $i \geq 1$. This is (not coincidentally) reminiscent of the hitting probability equation we previously found for the model (except the values of p and q are reversed). It has general solution $\pi_i = A + B(p/q)^i$.

We need $\sum \pi_i = 1$, which forces $A = 0$ and $B = (1 - p/q)$, giving

$$\pi_i = \left(1 - \frac{p}{q}\right) \left(\frac{p}{q}\right)^i.$$

That is, the stationary distribution of the walk is geometric with parameter $1 - \frac{p}{q}$.

Example 6.8 (A two-state chain and a non-irreducible chain)**.**

Consider the two-state chain on $\{1, 2\}$ with transition matrix $P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}$.

Solving $\pi P = \pi$ and normalising we obtain that $\pi = \left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta} \right)$.

Notice that this agrees with what we found in Example 5.4; the expression for $p_{11}^{(n)}$ given in (5.2) satisfies $p_{11}^{(n)} \to \pi_1 = \frac{\beta}{\alpha+\beta}$ as $n \to \infty$, as it should do because of the convergence to equilibrium in Theorem 6.2.

Now consider the chain on $\{1, 2, 3, 4\}$ whose transition matrix is

$$P = \begin{pmatrix} 1-\alpha & \alpha & 0 & 0 \\ \beta & 1-\beta & 0 & 0 \\ 0 & 0 & 1-\gamma & \gamma \\ 0 & 0 & \delta & 1-\delta \end{pmatrix}.$$

This chain is not irreducible. We can view it as two separate two-state chains, on $\{1, 2\}$ and $\{3, 4\}$, with no communication between them. Both $\left(\frac{\beta}{\alpha+\beta}, \frac{\alpha}{\alpha+\beta}, 0, 0\right)$ and $\left(0, 0, \frac{\delta}{\gamma+\delta}, \frac{\gamma}{\gamma+\delta}\right)$ are stationary distributions. But also any mixture of these is stationary (since if $\pi^{(1)}$ and $\pi^{(2)}$ are eigenvectors of P with eigenvalue 1, then so is any linear combination of $\pi^{(1)}$ and $\pi^{(2)}$).

So any distribution

$$\left(x\frac{\beta}{\alpha+\beta}, x\frac{\alpha}{\alpha+\beta}, (1-x)\frac{\delta}{\gamma+\delta}, (1-x)\frac{\gamma}{\gamma+\delta}\right),$$

where $x \in [0, 1]$, is stationary. (In fact, these are all the stationary distributions – exercise). The uniqueness result in Theorem 6.1 does not apply because the transition matrix is not irreducible.

6.4 Proof of Theorems 6.1, 6.2 and 6.3. *(non-examinable)*

The proofs below are given partly rather informally. They are not examinable; however, they are very helpful in developing your intuition about the results. The "coupling" idea used in the proof of Theorem 6.2 is particularly pretty and I certainly recommend thinking about it, but it will not be examined. (The results themselves are very much examinable!)

Proof of Theorem 6.3. This proof is essentially an application of the strong law of large numbers.

If the chain is transient, then with probability 1 there are only finitely many visits to any state, so $V_i(n)$ is bounded with probability 1. So

$$\mathbb{P}\left(\frac{V_i(n)}{n} \to 0 \text{ as } n \to \infty\right) = 1,$$

which is the result we want since $m_i = \infty$.

Suppose instead that the chain is recurrent. In this case we will visit state i infinitely often. Let R_k be the time between the kth and the $(k+1)$st visits to i. Then $R_1, R_2, R_3, \ldots$ are i.i.d. with mean m_i (which is finite in the positive recurrent case and infinite in the null recurrent case).

So by the strong law of large numbers,

$$\mathbb{P}\left(\frac{R_1 + R_2 + \cdots + R_K}{K} \to m_i \text{ as } K \to \infty\right) = 1. \tag{6.1}$$

Let T_K be the time of the Kth visit to i. Then $T_K = T_1 + X_1 + X_2 + \cdots + X_{K-1}$. It is easy to obtain that, for any c, $T_K/K \to c$ if and only if $(R_1 + \cdots + R_K)/K \to c$. Hence from (6.1) we have

$$\mathbb{P}\left(\frac{T_K}{K} \to m_i \text{ as } k \to \infty\right) = 1. \tag{6.2}$$

Notice that T_K/K is the time per visit (averaged over the first K visits) whereas $V_i(n)/n$ is the number of visits per unit time (averaged over the first n times). It is straightforward to obtain (check!) that, for any c, $T_K/K \to c$ as $K \to \infty$ if and only if $V_i(n)/n \to 1/c$ as $n \to \infty$.

Hence from (6.2) we have

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

as required. □

Lemma 6.9. *If P is positive recurrent, then it has stationary distribution π with $\pi_i = 1/m_i$.*

Proof. We give an informal version of the proof, which could quite easily be made rigorous.

From the ergodic theorem, we know that (with probability 1) the long-run proportion of visits to state i is $1/m_i$.

Each time the chain visits state i, it has probability p_{ij} of jumping from there to state j. We can obtain that the long-run proportion of jumps from i to j is $\frac{1}{m_i}p_{ij}$.

First consider the case where the state space I is finite. By summing over $i \in I$, we get that the long-run proportion of jumps into state j is $\sum_i \frac{1}{m_i}p_{ij}$.

But the long-run proportion of jumps into j is the same as the long-run proportion of visits to j, which (by the ergodic theorem) is $1/m_j$.

We obtain

$$\frac{1}{m_j} = \sum_i \frac{1}{m_i} p_{ij},$$

i.e. $\pi_j = \sum_i \pi_i p_{ij}$, so that π satisfies $\pi P = \pi$ and is stationary as desired.

If i is infinite, it is not immediate that the long-run proportion of jumps into j is the sum over i of the long-run proportions of jumps from i to j. However (by considering as large a finite set of i as desired) the second quantity does give an upper bound for the first, so we get

$$\frac{1}{m_j} \leq \sum_i \frac{1}{m_i} p_{ij}$$

for all $j \in I$. But summing both sides over j gives the same (finite) amount, since $\sum_j p_{ij} = 1$ for all i. So in fact we must have equality for all j as required. □

Lemma 6.10. *If π is any stationary distribution then $\pi_i = 1/m_i$.*

Proof. Suppose π is stationary for P, and let X be a Markov chain with initial distribution π and transition matrix P. Then by stationarity, $\mathbb{P}(X_n = i) = \pi_i$ for all n, and

$$\begin{aligned}
\frac{\mathbb{E}\, V_n(i)}{n} &= \frac{1}{n}\sum_{r=0}^{n-1} \mathbb{E}\left(\mathbf{1}\{X_n = i\}\right) \\
&= \frac{1}{n}\sum_{r=0}^{n-1} \mathbb{P}(X_n = i) \\
&= \pi_i. \qquad (6.3)
\end{aligned}$$

From the ergodic theorem, for any ϵ

$$\mathbb{P}\left(\left|\frac{V_n(i)}{n} - \frac{1}{m_i}\right| > \epsilon\right) < \epsilon \tag{6.4}$$

for large enough n (since almost sure convergence implies convergence in probability).

But since $V_n(i)/n$ is bounded between 0 and 1, it follows from (6.4) (check!) that

$$\frac{\mathbb{E}\, V_n(i)}{n} \to \frac{1}{m_i} \text{ as } n \to \infty.$$

Comparing to (6.3), we obtain $\pi_i = 1/m_i$. □

This gives uniqueness of the stationary distribution for positive recurrent chains, and shows that no stationary distribution can exist for null recurrent and transient chains. So we have proved Theorem 6.1.

Finally, we prove the result on convergence to equilibrium.

Proof of Theorem 6.2. Let P be irreducible and aperiodic, with stationary distribution π.

Let λ be any initial distribution, and let $(X_n, n \geq 0)$ be Markov(λ, P). We wish to show that $\mathbb{P}(X_n = j) \to \pi_j$ as $n \to \infty$, for any j.

Consider another chain $(Y_n, n \geq 0)$ which is Markov(π, P), and which is independent of Z. Since π is stationary, Y_n has distribution π for all n.

Let $T = \inf\{n \geq 0 : X_n = Y_n\}$. We will claim that $\mathbb{P}(T < \infty) = 1$; that is, the chains X and Y will meet at some point.

Suppose this claim is true. Then define another chain Z by

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

The idea is that Z starts in distribution λ, and evolves independently of the chain Y, until they first meet. As soon as that happens, Z copies the moves of Y exactly.

Then Z is also Markov(λ, P), since Z_n starts in distribution λ and each jump is done according to P, first by copying X up to time T, and then by copying Y after time T.

The idea is that the chain Y is "in equilibrium" (since it starts in the equilibrium distribution π) so that if there is high probability that $Y_n = Z_n$, then the distribution of Z_n must be close to π. More precisely:

$$\begin{aligned} |\mathbb{P}(Z_n = j) - \pi_j| &= |\mathbb{P}(Z_n = j) - \mathbb{P}(Y_n = j)| \\ &\leq \mathbb{P}(Z_n \neq Y_n) \\ &= \mathbb{P}(T > n) \\ &\to 0 \text{ as } n \to \infty. \end{aligned}$$

Then we have $\mathbb{P}(Z_n = j) \to \pi_j$.

But the chains X and Z have the same distribution (they are both Markov(λ, P)). So we have also shown that $\mathbb{P}(X_n = j) \to \pi_j$, as required.

It remains to prove the claim that T is finite with probability 1. Fix any state $b \in I$ and define $T_b = \inf\{n \geq 0 : X_n = Y_n = b\}$. Then $T \leq T_b$. We will show that T_b is finite with probability 1.

Consider the process $W_n = (X_n, Y_n)$. Since X_n and Y_n evolve independently, W_n is a Markov chain on the state space $I \times I$ with transition probabilities

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

and initial distribution

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

P is aperiodic and irreducible, so for all i, j, k, l, we have that

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

for all large enough n. So $\tilde{P}$ is irreducible.

$\tilde{P}$ has an invariant distribution given by

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

Hence $\tilde{P}$ is recurrent (by Theorem 6.1). But $T_b = \inf\{n \geq 0 : W_n = (b, b)\}$. Then indeed $\mathbb{P}(T_b < \infty) = 1$ (since an irreducible recurrent chain visits every state with probability 1). $\square$

Notice where the argument above fails when P is periodic. The chain $W_n = (X_n, Y_n)$ still has the stationary distribution of the form above, but it is not irreducible, so it may never reach the state (b, b). (For example, if P has period 2, and the chains X and Y start out with "opposite parity", then they will never meet).

7

Poisson processes

A Poisson process is a natural model for a stream of events occuring one by one in continuous time, in an uncoordinated way. For example: the process of times of detections by a Geiger counter near a radioactive source (a very accurate model); the process of times of arrivals of calls at a call centre (often a good model); the process of times of arrivals of buses at a bus stop (probably an inaccurate model; different buses are not really uncoordinated, for various reasons).

Consider a random process $N_t, t \in [0, \infty)$. (Note that "time" for our process is now a continuous rather than a discrete set!)

Such a process is called a *counting process* if N_t takes values in $\{0, 1, 2, \dots\}$, and $N_s \leq N_t$ whenever $s \leq t$. We will also assume that $t \mapsto N_t$ is right-continuous.

If N_t describes an arrival process, then $N_t = k$ means that there have been k arrivals in the time interval $[0, t]$. In fact we can describe the process by the sequence of arrival times, which we might call "points" of the process. Let $T_k = \inf\{t \geq 0 : N_t \geq k\}$ for $k \geq 0$. Then $T_0 = 0$ and T_k is the "kth arrival time", for $k \geq 1$. We also define $Y_k = T_k - T_{k-1}$ for $k \geq 1$. Y_k is the "interarrival time" between arrivals $k - 1$ and k.

For $s < t$, we write $N(s, t]$ for $N_t - N_s$, which we can think of as the number of points of the process which occur in the time-interval $(s, t]$. This is also called the "increment" of the process N on the interval $(s, t]$.

7.1 Poisson process: a choice of definitions

Let $\lambda > 0$. We will give two different definitions for what it means to be a **Poisson process of rate** λ. Afterwards we will show that these definitions are equivalent.

Definition 7.1 (Definition of Poisson process via exponential interarrival times)**.** $(N_t, t \geq 0)$ *is a Poisson process of rate* λ *if its interarrival times* $Y_1, Y_2, Y_3, \dots$ *are i.i.d. with* $\mathrm{Exp}(\lambda)$ *distribution.*

Definition 7.2 (Definition of Poisson process via Poisson distribution of increments)**.** $N_t, t \geq 0$ *is a Poisson process of rate* λ *if:*

(i) $N_0 = 0$.

(ii) If (s_1, t_1), $(s_2, t_2), \dots, (s_k, t_k)$ are disjoint intervals in $\mathbb{R}_+$, then the increments $N(s_1, t_1]$, $N(s_2, t_2], \dots, N(s_k, t_k]$ are independent, where $N(s_i, t_i] = N_{t_i} - N_{s_i}$.

(iii) For any $s < t$, the increment $N(s, t]$ has Poisson distribution with mean $\lambda(t - s)$.

Property (ii) in Definition 7.2 is called the **independent increments** property. The number of points falling in disjoint intervals is independent.

This can be seen as a version of the Markov property. For any t_0, the distribution of the process $(N(t_0, t_0 + t], t \geq 0)$, is independent of the process $(N_t, t \leq t_0)$. Put another way, the distribution of $(N_t, t > t_0)$ conditional on the process $(N_t, t \leq t_0)$ depends only on the value N_{t_0}.

7.2 Equivalence of the definitions

We wish to show that the properties listed in Definitions 7.1 and 7.2 are equivalent. The key idea is that the memoryless property for the exponential distribution and the independent increments property are telling us the same thing. The argument below is somewhat informal (but can be made completely rigorous).

Interrarival definition implies independent Poisson increments definition

Suppose we have Definition 7.1 in terms of i.i.d. exponential interarrival times. We wish to show that it implies the statements in Definition 7.2.

Property (i) is immediate: since $Y_1 = T_1 = \inf\{t \geq 0 : N_t \geq 1\}$ is strictly positive with probability 1, also $N_0 = 0$ with probability 1.

Now let us consider the distribution of the number of points in an interval. First let us take $s = 0$ in (iii), and consider $N(0, t]$. We want $N(0, t] \sim \text{Poisson}(\lambda t)$, i.e. that for all k,

$$\mathbb{P}\big(N(0, t] = k\big) = \frac{e^{-\lambda t}(\lambda t)^k}{k!}. \tag{7.1}$$

But we can rewrite the event on the LHS in terms of T_k and T_{k+1}. Since T_k is the sum of k independent exponentials of rate λ, we have $T_k \sim \text{Gamma}(k, \lambda)$, and similarly $T_{k+1} \sim \text{Gamma}(k + 1, \lambda)$. So

$$\begin{aligned}
\mathbb{P}\big(N(0, t] = k\big) &= \mathbb{P}\left(T_k \leq t, T_{k+1} > t\right) \\
&= \mathbb{P}\left(T_k \leq t\right) - \mathbb{P}\left(T_{k+1} \leq t\right) \\
&= \int_0^t \frac{(\lambda x)^{k-1} e^{-\lambda x}}{(k-1)!} dx - \int_0^t \frac{(\lambda x)^k e^{-\lambda x}}{k!} dx.
\end{aligned} \tag{7.2}$$

Now we can check that the RHS of (7.1) and (7.2) are the same (for example, either by integrating by parts in (7.2), or by differentiating in (7.1)). In this way we obtain that indeed $N(0, t] \sim \text{Poisson}(\lambda t)$.

Now we use the memoryless property of the exponential distribution to extend this to all intervals and to give the independent increments property.

Fix s, and suppose we condition on any outcome of the process on $[0, s]$. To be specific, condition on the event that

$$N_s = k, T_1 = t_1, T_2 = t_2, \ldots, T_k = t_k.$$

Equivalently,

$$Y_1 = t_1, Y_2 = t_2 - t_1, \ldots, Y_k = t_k - t_{k-1}, Y_{k+1} > s - t_k. \tag{7.3}$$

The memoryless property for Y_{k+1} tells us that conditional on $Y_{k+1} > s - t_k$, the distribution of $Y_{k+1} - (s - t_k)$ is again exponential with rate λ.

Combining this with the independence of the sequence Y_i, we have that conditional on (7.3), the sequence $Y_{k+1} - (s - t_k), Y_{k+2}, Y_{k+3}, \ldots$ is i.i.d. with $\text{Exp}(\lambda)$ distribution.

But this means that, conditional on (7.3), the distribution of the process $N(s, s+u], u \geq 0$ is the same as the original distribution of the process $N_u, u \geq 0$.

So indeed, the property (iii) extends to all s. Further, the increment on $(s, t]$ is independent of the whole process on $(0, s]$, and applying this repeatedly we get independence of any set of increments on disjoint intervals. So Definition 7.2 holds as desired.

Poisson definition characterises the distribution of the process

With some work we could show the reverse implication using a direct calculation. Instead we appeal to a general (although rather subtle) property. The Poisson definition specifies the joint distribution of $N_{t_1}, N_{t_2}, \ldots, N_{t_k}$ for any sequence $t_1, t_2, \ldots, t_k$. It turns out that such "finite dimensional distributions", along with the assumption that the process is right-continuous, are enough to characterise completely the distribution of the entire process. We will not delve any further here into this fact from stochastic process theory. But it means that at most one process could satisfy Definition 7.2, and since we have shown that a process defined by Definition 7.1 does so, we have that Definition 7.2 implies 7.1 as desired.

7.2.1 The Poisson process as a limit of discrete-time processes

The calculation showing that (7.1) and (7.2) are the same is perhaps not very illuminating. The case $k = 0$ is easy and is illustrated for example in Example 7.4(a) below. To get more intuition for the relation between Poisson increments and exponential interarrivals, one can also think about a related discrete-time process.

Let us recall some facts from earlier in the course:

(1) If $X_n \sim \text{Binomial}(n, \lambda/n)$, then $X_n \xrightarrow{d} \text{Poisson}(\lambda)$ as $n \to \infty$. (See Example 2.9.)

(2) If $Y_n \sim \text{Geometric}(\lambda/n)$, then $Y_n/n \xrightarrow{d} \text{Exp}(\lambda)$ as $n \to \infty$. (See Example 2.3.)

Now consider a sequence of independent Bernoulli trials. In each trial (or time-slot), suppose we see a success with probability p and no event with probability $1 - p$. Then in any run of M trials, the total number of successes has $\text{Binomial}(M, p)$ distribution. Meanwhile the distances between consecutive successes are i.i.d. with $\text{Geometric}(p)$ distribution.

Now consider n large. Let $p = \lambda/n$, and rescale time by a factor of $1/n$, so that a time-interval of length t corresponds to a run of tn trials. Then the number of events in a time-interval of length t has $\text{Binomial}(tn, \lambda/n)$ distribution, which is approximately $\text{Poisson}(\lambda t)$, while the times between consecutive successes have Geometric(λ/n) distribution rescaled by $1/n$, which is approximately Exp(λ).

So indeed, as $n \to \infty$, we obtain a continuous-time process in which the interarrival times are independent exponentials, and the increments on disjoint intervals are independent Poisson random variables. So we can see this exponential/Poisson relationship in the Poisson process as a limit of the geometric/binomial relationship which is already familiar from sequences of independent trials.

7.2.2 A third definition *(non-examinable)*

Reflecting some of the ideas in the previous section, there is in fact a third natural definition of the Poisson process, which we include for completeness. This involves the independent increments property as in the case of Definition 7.2, but instead of specifying that increments have Poisson distribution, it specifies the behaviour of the increments on small time-intervals. Namely, the probability of seeing an event in a small interval should behave like λ multiplied by the length of the interval, and it should be very unlikely that two or more events occur within the interval:

Definition 7.3 (Defintion of Poisson process via infinitesimal increments)**.** *$N_t, t \geq 0$ is a Poisson process of rate λ if:*

(i) $N_0 = 0$.

(ii) If (s_1, t_1), $(s_2, t_2), \ldots, (s_k, t_k)$ are disjoint intervals in $\mathbb{R}_+$, then the increments $N(s_1, t_1]$, $N(s_2, t_2], \ldots, N(s_k, t_k]$ are independent.

(iii) The distribution of $N(s, s+h]$ is the same for all s, and as $h \to 0$,

$$\begin{aligned} \mathbb{P}(N(s, s+h] = 0) &= 1 - \lambda h + o(h) \\ \mathbb{P}(N(s, s+h] = 1) &= \lambda h + o(h) \\ \mathbb{P}(N(s, s+h] \geq 2) &= o(h). \end{aligned} \tag{7.4}$$

Note that any two of the conditions of (7.4) imply the third.

This kind of formulation is very natural when moving to the context of more general continuous-time Markov jump processes (in which the rate at which jumps occur may depend on the present state). The definition can again be shown to be equivalent to Definitions 7.1 and 7.2.

7.3 Thinning and superposition of Poisson processes

Theorem 7.1 (Superposition of Poisson processes)**.** *Let L_t and M_t be independent Poisson processes of rate λ and μ respectively. Let $N_t = L_t + M_t$. Then N_t is a Poisson process of rate $\lambda + \mu$.*

Proof. We work from the definition of a Poisson process in terms of independent Poisson increments for disjoint intervals. Clearly, $N_0 = L_0 + M_0 = 0$ for property (i), and also

N_t satisfies property (ii) (independent increments) since L_t and M_t both have independent increments and are independent of each other.

So we need to show property (iii). Since $L(s,t] \sim \text{Poisson}(\lambda t)$ and $M(s,t] \sim \text{Poisson}(\mu t)$ independently of each other, we have $N(s,t] \sim \text{Poisson}((\lambda+\mu)t)$ as required, by familiar properties of the Poisson distribution. $\square$

Theorem 7.2 (Thinning of a Poisson process). *Let N_t be a Poisson process of rate λ. "Mark" each point of the process with probability p, independently for different points. Let M_t be the counting process of the marked points. Then M_t is a Poisson process of rate $p\lambda$.*

Proof. Again we will work with the definition in terms of independent Poisson increments. Properties (i) and (ii) for M follow from the same properties for N.

Now consider any interval $(s,t]$. We have $N(s,t] \sim \text{Poisson}(\lambda(t-s))$, and conditional on $N(s,t] = n$, we have $M(s,t] \sim \text{Binomial}(n,p)$.

But if $N \sim \text{Poisson}(\mu)$, and, conditional on $N = n$, $M \sim \text{Binomial}(n,p)$, then in fact $M \sim \text{Poisson}(p\mu)$. This fact was proved in two different ways in the Prelims course. For example, it can be done using generating functions: let $M = X_1 + X_2 + \cdots + X_N$ where X_i are i.i.d. Bernoulli random variables; then $G_M(s) = G_N(G_X(s))$. Alternatively, by direct calculation:

$$\begin{aligned}
\mathbb{P}(M = k) &= \sum_n \mathbb{P}(M = k | N = n)\mathbb{P}(N = n) \\
&= \sum_{n \geq k} \frac{e^{-\mu}\mu^n}{n!} \binom{n}{k} p^k (1-p)^{n-k} \\
&\vdots \\
&= \frac{e^{-p\mu}(p\mu)^k}{k!}.
\end{aligned}$$

Hence indeed we have here that $M(s,t] \sim \text{Poisson}(p\lambda(t-s))$. So indeed property (iii) holds as desired, and M is a Poisson process of rate $p\lambda$. $\square$

Remark 7.3. In fact, it is not too hard to prove something stronger. If L is the process of unmarked points, then L is a Poisson process of rate $(1-p)\lambda$, and the processes L and M are independent.

7.4 Poisson process examples

Example 7.4. A Geiger counter near a radioactive source detects particles at an average rate of 1 per 2 seconds. (a) What is the probability that there is no particle detected for 3 seconds after the detector is switched on? (b) What is the probability of detecting at least 3 particles in the first 4 seconds?

Solution: We model the process of detections as a Poisson process with rate $\lambda = 0.5$ (where the unit of time is 1 second).

For part (a), $\mathbb{P}(N_3 = 0) = e^{-3\lambda} = e^{-1.5}$, since N_3, the number of points up to time 3, has Poisson(3λ) distribution. Alternatively, we could calculate the same probability as $\mathbb{P}(T_1 > 3) = e^{-3\lambda}$ since T_1, the time of the first point of the process, has distribution Exp(λ).

For part (b), N_4 has Poisson distribution with mean $4\lambda = 2$. Then

$$\begin{aligned}\mathbb{P}(N_4 \geq 3) &= 1 - \mathbb{P}(N_4 = 0) - \mathbb{P}(N_4 = 1) - \mathbb{P}(N_4 = 2) \\ &= 1 - e^{-2} - 2e^{-2} - \frac{2^2 e^{-2}}{2!} \\ &= 1 - 5e^{-2}.\end{aligned}$$

Example 7.5. A call centre receives calls from existing customers at rate 1 per 20 seconds, and calls from potential new customers at rate 1 per 30 seconds. Assume that these form independent Poisson processes. (a) What is the distribution of the total number of calls in a given minute? (b) What is the probability that the next call to arrive is from a potential new customer? (c) Suppose each call from a potential new customer results in a contract with probability 1/4 independently. What is the distribution of the number of new contracts arising from calls in a given hour?

Solution: Let the unit of time be 1 minute, so that the Poisson processes in the question have rates 3 and 2.

(a) From Theorem 7.1, the combined process of all calls is a Poisson process of rate 5. The number of calls in a given minute has Poisson(5) distribution.

(b) From any given moment, the time until the next "existing" call, say U_1, is exponential with rate 3, and the time until the next "new" call, say V_1, is exponential with rate 2.

$$\begin{aligned}\mathbb{P}(U_1 < V_1) &= \int_{u=0}^{\infty} \int_{v=u}^{\infty} 3e^{-3u} \times 2e^{-2v} dv du \\ &= \int_{u=0}^{\infty} 3e^{-3u} e^{-2u} du \\ &= 3/(2+3) \\ &= 3/5.\end{aligned}$$

(In fact, it is not a coincidence that here the answer is the ratio of the rate of the "existing customer" process to the rate of the two processes combined. This fact follows from Remark 7.3; we can consider a single process of rate 5 and "mark" each point with probability 3/5, to arrive at two independent processes with rates 3 and 2. In particular, the probability that the first point is marked is then 3/5.)

(c) The process of calls resulting in contracts is a thinning of the process of calls from potential new customers. This gives us a new Poisson process of rate $1/4 \times 2 = 1/2$. So the total number of calls resulting in new contracts in a given time interval of length 60 has Poisson(30) distribution.

Example 7.6 (Genetic recombination model)**.** An illustration of genetic recombination is shown in the figure below. In most of our cells, we have two versions of each chromosome, one inherited from our mother and one from our father. Sex cells – sperm and ova – contain only one copy of each chromosome.

During *meiosis* – the process in which sperm and ova are created – the chromosomes are broken at certain random "crossover" or "recombination" points, to form new chromosomes

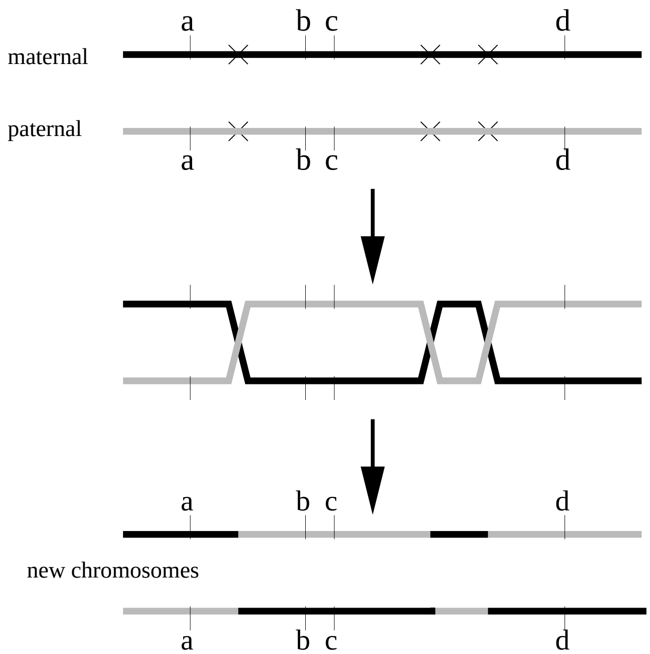


Figure 7.1: Recombination

out of pieces of the maternal and paternal chromosomes. The crossover points are shown as crosses in the top line of Figure 7.1.

Genes occur at particular positions along the chromosome. In early genetic research, biologists investigated the position of genes on chromosomes by looking at how likely the genes were to stay together, generation after generation. Genes on different chromosomes should be passed on independently. Genes that are close together on the same chromosome should almost always be passed on together, while genes that are on the same chromosome but further apart should be more likely than chance to be inherited together, but not certain.

In Figure 7.1, genes b, c and d stay together but a is separated from them.

As a simple model, we can imagine the chromosome as a continuous line, and model the recombination points as a Poisson process along it, of rate λ, say.

Consider two points a and b on the interval, representing the location of two genes. Let x be the distance between a and b. The probability of seeing no crossover at all between a and b is given by

$$\mathbb{P}\left(\text{no crossover in } (a,b)\right) = e^{-\lambda x}.$$

But what we really want to compute is the probability of seeing an *even number* of crossovers between a and b:

$$\begin{aligned} p = \mathbb{P}\left(\text{even number of crossovers in } (a,b)\right) &= \sum_{k=0}^{\infty} e^{-\lambda x} \frac{(\lambda x)^{2k}}{(2k)!} \\ &= e^{-\lambda x}\left(1 + \frac{(\lambda x)^2}{2!} + \frac{(\lambda x)^4}{4!} + \dots\right) \end{aligned}$$

$$
\begin{aligned}
&= e^{-\lambda x}\left(\frac{e^{\lambda x} + e^{-\lambda x}}{2}\right) \\
&= \frac{1 + e^{-2\lambda x}}{2}.
\end{aligned}
$$

If we observe that a and b are inherited together with probability $p > 1/2$, we can invert the expression above to estimate the distance between them by

$$x = -\frac{1}{2\lambda}\log(2p-1).$$