Random Processes (MTH6141), 2019

David Ellis

Introduction

Definition 0.1. A stochastic process is a collection of random variables \( \{X_t : t \in T\} \), indexed by a set \( T \), where each random variable \( X_t \) takes values in a (common) set \( S \). The index-set \( T \) is usually thought of as ‘time’. The set \( S \) is called the state space of the stochastic process. Note that, in this course, the state space \( S \) will always be either finite or countably infinite.

In the first half of the course, we will take \( T = \mathbb{N} \cup \{0\} \), so the stochastic process can be written as \( (X_0, X_1, X_2, \ldots) \). This is called a discrete-time stochastic process. It is a good model to use when we are only interested in the state of a system at ‘integer times’, e.g. at \( n \) seconds after the start of an experiment, where \( n \in \mathbb{N} \cup \{0\} \). For example, we might be observing the number of yeast cells in a culture, at the end of each hour after the start of an experiment, or we might be observing the number of rabbits in a warren, at the start of every month, from New Year’s Day 2016, onwards.

In the second half of the course, we will consider continuous-time stochastic processes, where \( T = \mathbb{R}_{\geq 0} = \{x \in \mathbb{R} : x \geq 0\} \). This is a good model to use if we are interested in the state of a system at any (real-valued) time, e.g. at 1.5 seconds, or at \( e = 2.718\ldots \) seconds, after the start of an experiment. For example, we might want to keep track of the number of yeast cells in a culture at all times after start of an experiment (meaning, we would want to know exactly when all the cell-divisions occurred).

∗These notes are adapted from lecture notes of Robert Johnson and of Mark Jerrum.
1 Discrete-time stochastic processes

The Markov property

Recall that if $A$ and $B$ are events with $P(A) \neq 0$, then we define the conditional probability of $B$ given $A$ as

$$P(B|A) = \frac{P(A \cap B)}{P(A)}.$$ 

This can be thought of as the probability that the event $B$ occurs, given that we know already that the event $A$ occurs.

**Definition 1.1.** Let $S$ be a finite or countable set. A discrete-time stochastic process $(X_0, X_1, X_2, \ldots)$ with state space $S$ is said to be a Markov chain if it satisfies, for all $t \in \mathbb{N} \cup \{0\}$,

$$P(X_{t+1} = i_{t+1} \mid \{X_t = i_t, X_{t-1} = i_{t-1}, \ldots, X_1 = i_1, X_0 = i_0\}) = P(X_{t+1} = i_{t+1} \mid X_t = i_t),$$

for all $i_{t+1}, i_t, i_{t-1}, \ldots, i_1, i_0 \in S$ for which $P(X_t = i_t, X_{t-1} = i_{t-1}, \ldots, X_1 = i_1, X_0 = i_0) \neq 0$. (Note that if $P(X_t = i_t, X_{t-1} = i_{t-1}, \ldots, X_1 = i_1, X_0 = i_0) = 0$, then the conditional probability on the left-hand side is not defined!)

This is called the Markov property. It says that if we know the state of the chain at a certain time, then we know the probability distribution of the state of the chain at any future time. Equivalently, it says that ‘the future, conditioned on the present, does not depend on the past.’

**Example 1** (A ‘rooms and doors’ Markov chain). Here is a simple example of a discrete-time Markov chain. There are 5 rooms, labelled 1, 2, 3, 4, 5. There are doors connecting certain pairs of rooms, indicated by the following diagram (we draw a line between two rooms if and only if there is a door connecting them):
We start in room 1. At the end of each minute, if we are in a certain room, we choose at random one of the doors leading from that room (we choose each door with equal probability), and we go through that door to the room beyond. We let \( X_t \) denote the label of the room we are in after \( t \) minutes (so \( X_0 = 1 \), and \( X_1 = 2 \) or 5, with \( \mathbb{P}(X_1 = 2) = \frac{1}{2} \) and \( \mathbb{P}(X_1 = 5) = \frac{1}{2} \)).

The stochastic process \( (X_0, X_1, X_2, \ldots) \) is clearly a discrete-time Markov chain, with state space \( S = \{1, 2, 3, 4, 5\} \). It is a Markov chain because for all \( t \in \mathbb{N} \cup \{0\} \), and for all \( i_t+1, i_t, \ldots, i_1, i_0 \in \{1, 2, 3, 4, 5\} \) for which \( \mathbb{P}(X_t = i_t, X_{t-1} = i_{t-1}, \ldots, X_1 = i_1, X_0 = i_0) \neq 0 \), we have

\[
\mathbb{P}(X_{t+1} = i_{t+1} \mid \{X_t = i_t, X_{t-1} = i_{t-1}, \ldots, X_1 = i_1, X_0 = i_0\}) = \begin{cases} \frac{1}{d(i_t)} & \text{if there is a door between room } i_t \text{ and room } i_{t+1}; \\ 0 & \text{otherwise,} \end{cases}
\]

where \( d(j) \) denotes the number of doors leading from room \( j \) (for each \( j \in \{1, 2, 3, 4, 5\} \)). Since the right-hand side depends only upon \( i_t \) (and not on \( i_{t-1}, i_{t-2}, \ldots, i_1 \) or \( i_0 \)), we have

\[
\mathbb{P}(X_{t+1} = i_{t+1} \mid \{X_t = i_t, X_{t-1} = i_{t-1}, \ldots, X_1 = i_1, X_0 = i_0\}) = \mathbb{P}(X_{t+1} = i_{t+1} \mid X_t = i_t),
\]

i.e. the Markov property is satisfied.

Some natural questions to ask are:

- What is the probability that we reach room 4 before reaching room 5?
- What is the expectation of the length of time we spend in room 3, before reaching room 4?
- What is the expectation of the time at which we first return to room 1?

By the end of the first half of the course, you should be able to answer all of these questions!

**Homogeneity**

**Definition 1.2.** If \( \mathbb{P}(X_{t+1} = j \mid X_t = i) \) is independent of \( t \), then we say that the Markov chain is *homogeneous*. (In fact, in this course, we will only consider homogeneous Markov chains.) In this case, we write \( \mathbb{P}(X_{t+1} = j \mid X_t = i) = p_{ij} \), and we call the \( p_{ij} \)'s the *transition probabilities*.

Note that the ‘rooms and doors’ Markov chain in Example 1 is homogeneous, as are many other useful Markov chains.

**Remark.** For the rest of the course, ‘Markov chain’ will always mean ‘homogeneous Markov chain’.
**Transition graphs**

A convenient way of visualising a Markov chain is by drawing its *transition graph*. The set of ‘vertices’ (or ‘nodes’) of the transition graph of a Markov chain is the set of states $S$. For each pair of states $i, j$ for which the transition probability $p_{ij}$ is non-zero, there is a directed ‘edge’ (or ‘arc’) from vertex $i$ to vertex $j$, labelled with the transition probability $p_{ij}$.

**Example 2** (The ‘grasshopper’ Markov chain). A grasshopper jumps between three flowers, labelled 1, 2 and 3. He starts at flower 1, and he jumps at the end of each minute. Whenever he is about to jump, he chooses at random one of the two flowers he is not currently sitting on, choosing each with probability $\frac{1}{2}$, and he jumps to that flower. Let $X_t$ be the label of the flower he is sitting on after $t$ minutes (so $X_0 = 1$, and $X_1 = 2$ or 3). Then $(X_0, X_1, X_2, \ldots)$ is a Markov chain with state space $S = \{1, 2, 3\}$ and with transition graph as follows.

![Transition Graph](attachment:image.png)

**Example 3.** Here is the transition graph of the ‘rooms and doors’ Markov chain in Example 1.
Calculating the probability distribution of the state of a Markov chain at time $t$.

Suppose $(X_0, X_1, X_2, \ldots)$ is a (homogeneous, discrete time) Markov chain, and we know the probability distribution of $X_0$, and all the transition probabilities. Then the following easy lemma tells us the joint probability distribution of the random variables $(X_0, X_1, \ldots, X_t)$, for any $t \in \mathbb{N}$.

**Lemma 1.1.** If $(X_0, X_1, X_2, \ldots)$ is a (homogeneous) Markov chain with state space $S$, with $\mathbb{P}(X_0 = i) = \mu^{(0)}_i$ for all $i \in S$, and with transition probabilities $p_{i,j}$ ($i, j \in S$), then

$$
\mathbb{P}(X_0 = i_0, X_1 = i_1, X_2 = i_2, \ldots, X_t = i_t) = \mu^{(0)}_{i_0} p_{i_0, i_1} p_{i_1, i_2} \ldots p_{i_{t-1}, i_t} \mu^{(0)}_{i_t},
$$

for all $i_0, i_1, \ldots, i_t \in S$. 

Proof.

\[ P(X_0 = i_0, X_1 = i_1, \ldots, X_t = i_t) \]
\[ = P(X_0 = i_0) \times P(X_1 = i_1 \mid X_0 = i_0) \times P(X_2 = i_2 \mid \{X_1 = i_1, X_0 = i_0\}) \]
\[ \times \cdots \times P(X_t = i_t \mid \{X_{t-1} = i_{t-1}, \ldots, X_0 = i_0\}) \]

(by the definition of conditional probability)
\[ = P(X_0 = i_0) P(X_1 = i_1 \mid X_0 = i_0) P(X_2 = i_2 \mid X_1 = i_1) \cdots P(X_t = i_t \mid X_{t-1} = i_{t-1}) \]

(by the Markov property)
\[ = \mu^{(0)}_{i_0} p_{i_0,i_1} p_{i_1,i_2} \cdots p_{i_{t-1},i_t}. \]

(by definition of \( p_{ij} \) and \( \mu^{(0)}_i \))

\[ \square \]

Definition 1.3. The initial distribution of a Markov chain \((X_0, X_1, X_2, \ldots)\) is the probability distribution of \(X_0\), i.e. the row-vector \(\mu^{(0)}\) with entries indexed by \(S\), where \(\mu^{(0)}_i = P(X_0 = i)\) for all \(i \in S\).

Note that Lemma 1.1 gives us a crude way of calculating \(P(X_t = i)\), for any \(t \in \mathbb{N} \cup \{0\}\) and any \(i \in S\): we can simply sum over all the possible ways of getting to the state \(i\) at time \(t\), giving

\[ P(X_t = i) = \sum_{i_0,i_1,\ldots,i_{t-1} \in S} P(X_0 = i_0, X_1 = i_1, \ldots, X_{t-1} = i_{t-1}, X_t = i) \]
\[ = \sum_{i_0,i_1,\ldots,i_{t-1} \in S} \mu^{(0)}_{i_0} p_{i_0,i_1} p_{i_1,i_2} \cdots p_{i_{t-1},i_t}. \]

Soon we will see a much cleaner way of calculating \(P(X_t = i)\), using transition matrices.

Transition matrices

Definition 1.4. Let \((X_0, X_1, X_2, \ldots)\) be a (homogeneous) Markov chain with state space \(S\) and transition probabilities \((p_{ij} : i, j \in S)\). Its transition matrix \(P\) is the matrix with rows and columns indexed by \(S\), with \((i,j)\)-th entry \(P_{ij} = p_{ij}\). In other words, it is just the matrix of transition probabilities.

Example 4. The grasshopper’s Markov chain in Example 2 has transition matrix

\[ P = \begin{pmatrix}
0 & \frac{1}{2} & \frac{1}{2} \\
\frac{1}{2} & 0 & \frac{1}{2} \\
\frac{1}{2} & \frac{1}{2} & 0
\end{pmatrix},
\]

(here, of course, the rows and columns are labelled by 1,2,3 in increasing order).

Example 5 (A simple weather model). The following Markov chain \((X_0, X_1, X_2, \ldots)\) is a simple model for the weather. There are two states, 1=’sunny’ and 2=’rainy’; \(X_t\)
represents the state of the weather on the \((t + 1)\)th day of the year, so \(X_t = 1\) or \(2\) for all \(t \in \mathbb{N} \cup \{0\}\). The transition probabilities are

\[ p_{1,1} = 0.7, \quad p_{1,2} = 0.3, \quad p_{2,1} = 0.6, \quad p_{2,2} = 0.4. \]

(\text{So the probability that it is rainy tomorrow given that it is sunny today is 0.3, etc.})

This Markov chain has transition graph

\[
\begin{array}{c}
0.7 \\
1 \\
0.6 \\
2 \\
0.4 \\
0.3 \\
0.7
\end{array}
\]

and transition matrix

\[
P = \begin{pmatrix}
0.7 & 0.3 \\
0.6 & 0.4
\end{pmatrix}.
\]

(\text{Here, again, the rows and columns are labelled by 1,2 in increasing order; this will always be the case when the state space is \(\{1,2,\ldots,n\}\) for some \(n \in \mathbb{N}\).})

Notice that a Markov chain is not a particularly realistic model for the weather in the UK, where the weather on a given day typically depends on a weather system (such as a low-pressure system or a warm or cold front) that evolves gradually over several days or even weeks, so it will not generally be dependent only on the weather the day before. For example, the weather the day before tells us nothing about whether the temperature is getting gradually ‘higher’, or gradually ‘lower’. But the model above may have some uses as a rough approximation.

**Example 6.** The ‘rooms and doors’ Markov chain in Example 1 has transition matrix

\[
P = \begin{pmatrix}
0 & 1/2 & 0 & 0 & 1/2 \\
1/3 & 0 & 1/3 & 0 & 1/3 \\
0 & 1/3 & 0 & 1/3 & 1/3 \\
0 & 0 & 1 & 0 & 0 \\
1/3 & 1/3 & 1/3 & 0 & 0
\end{pmatrix}.
\]

**Remark.** Note that for any transition matrix \(P\), all the row-sums are 1:

\[
\sum_{j \in S} p_{ij} = 1 \quad \forall i \in S.
\]

This is clear, as if \(X_0 = i\), then we must have \(X_1 = j\) for exactly one state \(j \in S\), so

\[
\sum_{j \in S} p_{ij} = \sum_{j \in S} P(X_1 = j \mid X_0 = i) = 1.
\]

A square matrix of non-negative numbers with all its row-sums equal to 1 is often called a **stochastic** matrix. Notice that any stochastic matrix is the transition matrix of some Markov chain. (Can you see why?)
**r-step transition probabilities**

To help our analysis of Markov chains using transition matrices, we need to introduce the following definition.

**Definition 1.5.** Let \((X_0, X_1, X_2, \ldots)\) be a (homogeneous) Markov chain with state space \(S\). Then for any \(r \in \mathbb{N}\), the *r-step transition probabilities* are defined by

\[
p^{(r)}_{ij} = \mathbb{P}(X_r = j \mid X_0 = i) \quad \forall i, j \in S.
\]

In other words, \(p^{(r)}_{ij}\) is the probability that the chain transitions from state \(i\) to state \(j\) in exactly \(r\) steps. Note that, since the chain is homogeneous, we have

\[
\mathbb{P}(X_{t+r} = j \mid X_t = i) = \mathbb{P}(X_r = j \mid X_0 = i) = p^{(r)}_{ij} \quad \forall t \in \mathbb{N}, \forall i, j \in S.
\]

The following relation between multi-step transition probabilities is very useful.

**Lemma 1.2** (Chapman-Kolmogorov relations).

\[
p^{(r+s)}_{i,j} = \sum_{k \in S} p^{(r)}_{i,k} p^{(s)}_{k,j}, \quad \forall i, j \in S, \forall r, s \in \mathbb{N} \cup \{0\}.
\]

**Proof.**

\[
p^{(r+s)}_{i,j} = \mathbb{P}(X_{r+s} = j \mid X_0 = i)
= \sum_{k \in S} \mathbb{P}(X_{r+s} = j, X_r = k \mid X_0 = i)
\quad \text{(since the events \(\{X_r = k\}\) partition the probability space)}
= \sum_{k \in S} \mathbb{P}(X_r = k \mid X_0 = i) \mathbb{P}(X_{r+s} = j \mid \{X_r = k, X_0 = i\})
\quad \text{(by the definition of conditional probability)}
= \sum_{k \in S} \mathbb{P}(X_r = k \mid X_0 = i) \mathbb{P}(X_{r+s} = j \mid X_r = k) \quad \text{(by the Markov property)}
= \sum_{k \in S} p^{(r)}_{i,k} p^{(s)}_{k,j}.
\]

This lemma enables us to relate the r-step transition probabilities to the \(r\)th power of the transition matrix, in the following corollary.

**Corollary 1.3.** Let \((X_0, X_1, X_2, \ldots)\) be a (homogeneous) Markov chain with state space \(S\) and transition probabilities \((p_{ij} : i, j \in S)\). If \(P\) is the transition matrix of a Markov...
chain, then $p^{(r)}_{ij} = (P^r)_{ij}$ for all $i, j \in S$. In other words, the matrix $P^r$ is the matrix of $r$-step transition probabilities. So if $S = \{1, 2, \ldots, n\}$, then

$$P^r = \begin{pmatrix}
p^{(r)}_{11} & p^{(r)}_{12} & \cdots & p^{(r)}_{1n} \\
p^{(r)}_{21} & p^{(r)}_{22} & \cdots & p^{(r)}_{2n} \\
\vdots & \vdots & \ddots & \vdots \\
p^{(r)}_{n1} & p^{(r)}_{n2} & \cdots & p^{(r)}_{nn}
\end{pmatrix}.$$ 

Proof. The proof is by induction on $r$, with the base case $r = 1$ being immediate. For the induction step, let $r \geq 2$ and assume the statement of the corollary holds with $r - 1$ in place of $r$. Then for any $i, j \in S$, we have

$$(P^r)_{i,j} = (P^{r-1}P)_{i,j}$$

$$= \sum_{k \in S} (P^{r-1})_{i,k} P_{k,j}$$

$$= \sum_{k \in S} p^{(r-1)}_{i,k} p^{(r)}_{k,j} \quad \text{(by inductive hypothesis)}$$

$$= p^{(r)}_{i,j} \quad \text{(by Theorem 1.2)}.$$ 

This completes the inductive step, proving the corollary. □

This corollary gives us a (relatively) clean way of calculating the probability distribution of $X_t$, for all $t$.

Definition 1.6. Let $(X_0, X_1, \ldots)$ be a Markov chain with state space $S$. Its distribution at time $t$ is the probability distribution of $X_t$, i.e. it is the row-vector $\mu^{(t)}$ with entries indexed by $S$, and with $\mu^{(t)}_i = \Pr(X_t = i)$ for all $i \in S$. (In particular, as before, $\mu^{(0)}$ is the initial distribution of the Markov chain.)

Corollary 1.4. Let $(X_0, X_1, \ldots)$ be a (homogeneous) Markov chain with state space $S$, initial distribution $\mu^{(0)}$ and transition matrix $P$. Then

$$\mu^{(t)} = \mu^{(0)} P^t.$$
Proof. For any $i \in S$, we have
\[
\mu_i^{(t)} = \mathbb{P}(X_t = i)
\]
\[
= \sum_{k \in S} \mathbb{P}(X_t = i, X_0 = k)
\]
\[
(\text{since the events } \{X_0 = k\} \text{ partition the probability space})
\]
\[
= \sum_{k \in S} \mathbb{P}(X_t = i | X_0 = k) \mathbb{P}(X_0 = k)
\]
\[
(\text{by the definition of conditional probabilities})
\]
\[
= \sum_{k \in S} p_{ki}^{(t)} \mu_k^{(0)}
\]
\[
(\text{by the definition of } t\text{-step transition probabilities})
\]
\[
= \sum_{k \in S} (P^t)_{k,i} \mu_k^{(0)}
\]
\[
(\text{by Corollary 1.3})
\]
\[
= (\mu^{(0)} P^t)_i.
\]

Hence, if we can calculate $P^t$, and we know the initial distribution (the probability distribution of $X_0$), then we can calculate the probability distribution of $X_t$, so we know everything there is to know about the Markov chain at time $t$. Sometimes it is possible to calculate $P^t$ fairly easily, but sometimes it is not, so we may need to fall back on other methods to analyse the Markov chain, as we will see later.

Absorption and first-step analysis

Some Markov chains have absorbing states: states which, once the chain has entered, it can never leave.

**Definition 1.7.** Let $(X_0, X_1, X_2, \ldots)$ be a Markov chain with state space $S$ and transition probabilities $(p_{ij} : i, j \in S)$. A state $i \in S$ is said to be absorbing if $p_{ii} = 1$.

**Remark.** When a Markov chain reaches an absorbing state, it cannot leave: it remains in this state forever.

**Definition 1.8.** Absorption is said to occur when a Markov chain first reaches an absorbing state.

When a Markov chain has one or more absorbing states, there are several questions we may wish to answer. For example, what is the probability of reaching a certain absorbing state? What is the expected time until absorption occurs? These questions can be answered by ‘first-step analysis’, also known as ‘conditioning on the first step’. The computation reduces to solving simultaneous linear equations.

**Theorem 1.5.** Suppose $(X_0, X_1, X_2, \ldots)$ is a Markov chain with state space $S$, transition probabilities $(p_{ij} : i, j \in S)$, and absorbing states $A \subseteq S$, and suppose $w : S \setminus A \to \mathbb{R}$ is a weight function on the non-absorbing states. Let $T = \min\{t : X_t \in A\}$ denote the time of first absorption, and let $W = \sum_{t=0}^{T-1} w(X_t)$. Then the following facts hold.
1. Fix $k \in A$. For each $i \in S$, let $a_i = \mathbb{P}(X_T = k | X_0 = i)$. Then the $(a_i : i \in S)$ satisfy the following system of simultaneous linear equations:

$$a_i = \begin{cases} 
1, & \text{if } i = k; \\
0, & \text{if } i \in A \setminus \{k\}; \\
p_i + \sum_{j \in S \setminus A} p_{ij} a_j & \text{if } i \in S \setminus A.
\end{cases}$$

2. For each $i \in S$, let $b_i = \mathbb{E}(W | X_0 = i)$. Then the $(b_i : i \in S)$ satisfy the following system of simultaneous linear equations:

$$b_i = \begin{cases} 
0, & \text{if } i \in A; \\
w(i) + \sum_{j \in S \setminus A} p_{ij} b_j & \text{if } i \notin A.
\end{cases}$$

3. If $S$ is finite and from every state it is possible to reach an absorbing state in some number of steps, then absorption occurs with probability 1, and the system of equations in (1) has a unique solution, and so does the system of equations in (2).

Proof. 1. If $i \in A$ (i.e., we start in an absorbing state) then $T = 0$; thus $a_k = \mathbb{P}(X_T = k | X_0 = k) = 1$.

If $i \in A \setminus \{k\}$, then $a_i = \mathbb{P}(X_T = k | X_0 = i) = 0$.

Otherwise (i.e., if $i \in S \setminus A$), then conditioning on the first step gives:

$$a_i = \sum_{j \in S} \mathbb{P}(X_1 = j | X_0 = i) \mathbb{P}(X_T = k | X_1 = j)$$

(using the Law of Total Probability)

$$= \sum_{j \in S} p_{ij} a_j$$

$$= p_i + \sum_{j \in S \setminus A} p_{ij} a_j$$

2. If $i \in A$ then $T = 0$ and $b_i = 0$. If $i \in S \setminus A$, then conditioning on the first step gives:

$$b_i = w(i) + \sum_{j \in S} \mathbb{P}(X_1 = j | X_0 = i) \mathbb{E} \left( \sum_{t=1}^{T-1} w(X_t) | X_1 = j \right)$$

(using the Law of Total Expectation)

$$= w(i) + \sum_{j \in S} p_{ij} b_j$$

$$= w(i) + \sum_{j \in S \setminus A} p_{ij} b_j.$$
3. Outside the scope of the module.

If we take \( w(i) = 1 \) for all \( i \in S \setminus A \) in the above theorem, then \( W = T \), the time of first absorption, so we get the following corollary.

**Corollary 1.6.** Suppose \((X_0, X_1, X_2, \ldots)\) is a Markov chain with state space \( S \), transition probabilities \((p_{ij} : i, j \in S)\), and absorbing states \( A \subseteq S \). Let \( T = \min\{ t : X_t \in A \} \) denote the time of first absorption. Then the following facts hold.

1. Let \( v_i = \mathbb{E}(T \mid X_0 = i) \). Then the \( v_i \) satisfy the following system of simultaneous linear equations:
   \[
   v_i = \begin{cases} 
   0, & \text{if } i \in A; \\
   1 + \sum_{j \in S \setminus A} p_{ij} v_j & \text{if } i \notin A.
   \end{cases}
   \]

2. If \( S \) is finite and from every state it is possible to reach some absorbing state in some number of steps, then absorption occurs with probability 1, and the above system of equations has a unique solution.

**Example 7.** Bill has £400 to gamble with at roulette. The roulette wheel has all the integers from 1 to 36, together with 0. Half of the non-zero numbers are red, half of them are black, and 0 is green. Any number between 0 and 36 is equally likely to occur each time the wheel is spun. Each time Bill spins the wheel, if red comes up he wins £200, and otherwise he loses £200. He decides to play until he either doubles his money (to £800), or loses all of it. Using first-step analysis on an appropriate Markov chain, find (to 4 significant figures) the probability he doubles his money, and the expectation of the number of times he spins the wheel.

Solution: let \( £200 \times X_t \) be the amount of money Bill has after \( t \) spins of the wheel. Then \((X_0, X_1, X_2, \ldots)\) is a Markov chain with state-space \( S = \{0, 1, 2, 3, 4\} \), and with transition graph as follows:

![Transition Graph](image)

The absorbing states are 0 and 4. Let \( T \) be the time of first absorption. For each \( i \in \{0, 1, 2, 3, 4\} \), let \( a_i = \mathbb{P}(X_T = 4 \mid X_0 = i) \). (We want to find \( a_2 \).) For brevity, write \( p = 18/37 \). Then by the above theorem, \( a_0, a_1, a_2, a_3, a_4 \) satisfy the following system of equations:
simultaneous linear equations:
\[
\begin{align*}
a_0 &= 0, \\
a_1 &= pa_2, \\
a_2 &= (1 - p)a_1 + pa_3, \\
a_3 &= (1 - p)a_2 + p, \\
a_4 &= 1.
\end{align*}
\]
We have three simultaneous linear equations in \(a_1, a_2, a_3\):
\[
\begin{align*}
a_1 &= pa_2, \\
a_2 &= (1 - p)a_1 + pa_3, \\
a_3 &= (1 - p)a_2 + p.
\end{align*}
\]
Substituting the first and third of these into the second gives:
\[
a_2 = (1 - p)(pa_2) + p((1 - p)a_2 + p) \Rightarrow a_2 = \frac{p^2}{1 - 2p + 2p^2} = \frac{324}{685} = 0.4730 \text{ (to 3 s.f.)}
\]
Now, for each \(i \in \{0, 1, 2, 3, 4\}\), let \(v_i = E(T|X_0 = i)\). (We want to find \(v_2\).) Then by the above corollary, \(v_0, v_1, v_2, v_3, v_4\) satisfy the following system of simultaneous linear equations:
\[
\begin{align*}
v_0 &= 0, \\
v_1 &= 1 + pv_2, \\
v_2 &= 1 + (1 - p)v_1 + pv_3, \\
v_3 &= 1 + (1 - p)v_2, \\
v_4 &= 0.
\end{align*}
\]
We have three simultaneous linear equations in \(v_1, v_2, v_3\):
\[
\begin{align*}
v_1 &= 1 + pv_2, \\
v_2 &= 1 + (1 - p)v_1 + pv_3, \\
v_3 &= 1 + (1 - p)v_2.
\end{align*}
\]
Substituting the first and third of these equations into the second, we get
\[
v_2 = 1 + (1 - p)(1 + pv_2) + p(1 + (1 - p)v_2) \Rightarrow v_2 = \frac{2}{1 - 2p + 2p^2} = \frac{2738}{685} = 3.997 \text{ (to 4 s.f.)}
\]
Here is another example.

**Example 8.** A fair coin is repeatedly tossed until the sequence \(HTH\) is seen. By considering an appropriate Markov chain, calculate the expectation of the total number of tosses.
Solution: this situation can be modelled as a Markov chain with $2^3 = 8$ states, one state for each possible sequence of three tosses ($HHH, HHT, \ldots, TTT$). However, there is a much more economical way of modelling it, as a Markov chain $(X_0, X_1, X_2, \ldots)$ with only 4 states:

**State 3:** The last three tosses were $HTH$;

**State 2:** The last two tosses were $HT$;

**State 1:** The last toss was an $H$, and the last three tosses were not $HTH$;

**State 0:** None of the above (i.e., the last two tosses were $TT$, or there has only been one toss so far and it was a $T$, or there have been no tosses so far).

Notice that state $i$ corresponds to the situation where ‘we have just seen the first $i$ tosses we need’. This Markov chain has transition graph as follows.

![Transition Graph](image)

There is just one absorbing state, namely the state 3. Let $T$ denote the time of first absorption. For each $i \in \{0, 1, 2, 3\}$, let $v_i = \mathbb{E}(T | X_0 = i)$. (We want to calculate $v_0$.) Then by the above corollary, $v_0, v_1, v_2, v_3$ satisfy the following system of simultaneous linear equations:

$v_0 = 1 + \frac{1}{2}v_0 + \frac{1}{2}v_1,$
$v_1 = 1 + \frac{1}{2}v_1 + \frac{1}{2}v_2,$
$v_2 = 1 + \frac{1}{2}v_0,$
$v_3 = 0.$

Simplifying the first two of these equations gives:

$\frac{1}{2}v_0 = 1 + \frac{1}{2}v_1,$
$\frac{1}{2}v_1 = 1 + \frac{1}{2}v_2,$
$v_2 = 1 + \frac{1}{2}v_0.$

Substituting the third of these equations into the second gives $v_1 = 3 + \frac{1}{2}v_0$, and substituting this into the first gives $v_0 = 5 + \frac{1}{2}v_0$, so $v_0 = 10.$
Computing $P^t$

By Corollary 1.4, if we know the initial distribution of a Markov chain and we can calculate $P^t$, then we can calculate the distribution at time $t$: we have $\mu^{(t)} = \mu^{(0)} P^t$.

This brings us on to the topic of calculating $P^t$.

Calculating $P^t$ is fairly easy if $P$ is diagonalisable. Recall that a square matrix $P$ is said to be diagonalizable (over $\mathbb{C}$) if there exists an invertible matrix $M$ (with complex entries) and a diagonal matrix $D$ (with complex entries) such that $M^{-1}DM = P$.

Calculating the $t$th power of a diagonal matrix is easy: if

$$D = \begin{pmatrix} d_{11} & 0 & \cdots & 0 \\ 0 & d_{22} & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn} \end{pmatrix},$$

then

$$D^t = \begin{pmatrix} d_{11}^t & 0 & \cdots & 0 \\ 0 & d_{22}^t & \cdots & 0 \\ 0 & 0 & \ddots & \vdots \\ 0 & 0 & \cdots & d_{nn}^t \end{pmatrix}.$$  

**Fact 1.** It follows that if $P$ is diagonalisable with $M^{-1}DM = P$, then

$$P^t = (M^{-1}DM)^t = (M^{-1}DM)(M^{-1}DM)\cdots(M^{-1}DM) = M^{-1}D^tM,$$

so we can calculate $P^t$ (fairly) easily.

Hence, if a Markov chain $(X_0, X_1, \ldots)$ has a diagonalizable transition matrix, we can calculate $P^t$ (and therefore $\mu^{(t)}$, if we also know the initial distribution $\mu^{(0)}$), fairly easily.

**Remark.** Unfortunately, the transition matrix of a Markov chain can fail to be diagonalizable over $\mathbb{C}$: for example,

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

is a stochastic matrix (so it is the transition matrix of some Markov chain), but it is not diagonalizable over $\mathbb{C}$. (Note that there was a typo in the definition of this matrix in the first edition of these notes.)

**Exercise 1** (for enthusiasts). Check that the above matrix is not diagonalisable over $\mathbb{C}$.

**Diagonalising the transition matrix $P$**

To diagonalise transition matrices, we need to recall (from Linear Algebra) some facts about eigenvalues.

One useful sufficient condition for diagonalizability is the following.
Fact 2. An $n \times n$ complex matrix is diagonalizable over $\mathbb{C}$ if it has $n$ distinct eigenvalues.

Definition 1.9. Recall that if $P$ is an $n \times n$ complex matrix, a complex number $\lambda \in \mathbb{C}$ is said to be an eigenvalue of $P$ if there exists a column-vector $v \in \mathbb{C}^n$ with $v \neq 0$, such that $Pv = \lambda v$; such a column-vector $v$ is said to be an eigenvector of $P$ with eigenvalue $\lambda$.

An equivalent definition is as follows: $\lambda$ is an eigenvalue of $P$ iff it is a root of the characteristic polynomial of $P$, defined by

$$\chi_P(t) = \det(P - tI).$$

Definition 1.10. If $P$ is an $n \times n$ complex matrix and $\lambda$ is an eigenvalue of $P$, then the $\lambda$-eigenspace of $P$ is the subspace $\{v \in \mathbb{C}^n : Pv = \lambda v\}$. (It is easy to check that this is a subspace of $\mathbb{C}^n$.)

Definition 1.11. If $P$ is an $n \times n$ complex matrix and $\lambda$ is an eigenvalue of $P$, then the algebraic multiplicity of $\lambda$ is the maximal integer $m$ such that $(t - \lambda)^m$ is a factor of $\chi_P(t)$. Informally, it is the ‘multiplicity of $\lambda$’ as a root of the characteristic polynomial $\chi_P(t)$.

Definition 1.12. If $P$ is an $n \times n$ complex matrix and $\lambda$ is an eigenvalue of $P$, then the geometric multiplicity of $\lambda$ is the dimension of the $\lambda$-eigenspace of $P$.

From now on, if $P$ is an $n \times n$ complex matrix and $\lambda$ is an eigenvalue of $P$, we will sometimes write $a_P(\lambda)$ for the algebraic multiplicity of $\lambda$, and $g_P(\lambda)$ for its geometric multiplicity.

Fact 3. If $P$ is an $n \times n$ complex matrix and $\lambda$ is an eigenvalue of $P$, then $g_P(\lambda) \leq a_P(\lambda)$.

You may recall the following necessary and sufficient condition for diagonalizability.

Fact 4. If $P$ is an $n \times n$ complex matrix, then $P$ is diagonalisable over $\mathbb{C}$ if and only if $g_P(\lambda) = a_P(\lambda)$ for every eigenvalue $\lambda$ of $P$.

The following three lemmas will be very useful for diagonalizing transition matrices.

Lemma 1.7. If $P$ is an $n \times n$ complex matrix which is diagonalizable with $n$ complex eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$ (repeated with their algebraic multiplicities), then we have

$$P = MDM^{-1},$$

where

$$D = \begin{pmatrix}
\lambda_1 & 0 & \cdots & 0 \\
0 & \lambda_2 & \cdots & 0 \\
0 & 0 & \ddots & \vdots \\
0 & 0 & \cdots & \lambda_n
\end{pmatrix},$$

and the $i$th column of $M$ is any eigenvector of $P$ with eigenvector $\lambda_i$. 

17
Proof. Omitted. □

Lemma 1.8. If \( P \) is the transition matrix of a Markov chain with finite state-space \( S \), then the all-1’s column-vector \((1, 1, \ldots, 1)^\top\) is an eigenvector of \( P \) with eigenvalue 1.

Proof. Let \( u = (1, 1, \ldots, 1)^\top \). Then for all \( i \in S \), we have \((Pu)_i = \sum_{j \in S} P_{i,j}u_j = \sum_{j \in S} P_{i,j} = 1 = u_i\), so \( Pu = u \), as required. □

Definition 1.13. If \( P \) is an \( n \times n \) complex matrix, the trace of \( P \) is the sum of all its diagonal entries:

\[
\text{Trace}(P) = \sum_{i=1}^{n} P_{i,i}.
\]

Lemma 1.9. If \( P \) is an \( n \times n \) complex matrix with complex eigenvalues \( \lambda_1, \lambda_2, \ldots, \lambda_n \) (repeated with their algebraic multiplicities), then

\[
\sum_{i=1}^{n} \lambda_i = \text{Trace}(P).
\]

Sketch-proof: Look at the coefficient of \( t^{n-1} \) in \( \chi_P(t) \). □

We will now use the above three lemmas to diagonalize \( P \) and hence calculate \( P^t \) for the ‘simple weather model’ we saw earlier.

Example 9. Our simple weather model has transition graph

![Transition Graph]

and transition matrix \[
P = \begin{pmatrix}
0.7 & 0.3 \\
0.6 & 0.4
\end{pmatrix};
\]

recall that the state 1 stands for ‘sunny’ and the state 2 stands for ‘rainy’.

Let \( \lambda_1, \lambda_2 \) be the eigenvalues of \( P \) (repeated with their algebraic multiplicities). By Lemma 1.8, 1 is an eigenvalue of \( P \) with eigenvector \((1, 1)\). So we may set \( \lambda_1 = 1 \). Using Lemma 1.9, we have

\[
1 + \lambda_2 = \lambda_1 + \lambda_2 = \text{Trace}(P) = 0.7 + 0.4 = 1.1,
\]

so \( \lambda_2 = 0.1 \). Since \( P \) has two distinct eigenvalues, we may conclude from Fact 2 that \( P \) is diagonalizable.
We now wish to find an eigenvector \( \begin{pmatrix} x \\ y \end{pmatrix} \) for the eigenvalue 0.1. Solving

\[
\begin{pmatrix} 0.7 & 0.3 \\ 0.6 & 0.4 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = 0.1 \begin{pmatrix} x \\ y \end{pmatrix}
\]
gives \( x = -y/2 \). Hence, we may choose \( \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} 1 \\ -2 \end{pmatrix} \).

Therefore, by Lemma 1.7, we have

\[
P = MDM^{-1},
\]

where

\[
D = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 0.1 \end{pmatrix}, \quad M = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix},
\]

and so

\[
M^{-1} = \frac{1}{\det(M)} \begin{pmatrix} m_{2,2} & -m_{1,2} \\ -m_{2,1} & m_{1,1} \end{pmatrix} = \frac{1}{-3} \begin{pmatrix} -2 & -1 \\ -1 & 1 \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix}.
\]

Therefore, by Fact 1, we have

\[
P^t = MD^tM^{-1} = \begin{pmatrix} 1 & 1 \\ 1 & -2 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 0.1 \end{pmatrix} \begin{pmatrix} 2 & 1 \\ 1 & -1 \end{pmatrix} = \begin{pmatrix} \frac{2}{3} + \frac{1}{3}10^{-t} & \frac{1}{3} - \frac{1}{3}10^{-t} \\ \frac{2}{3} - \frac{1}{3}10^{-t} & \frac{1}{3} + \frac{1}{3}10^{-t} \end{pmatrix}.
\]

We can now calculate \( P(X_t = j|X_0 = i) \), for any \( i, j \in \{1, 2\} \). Indeed, if \( X_0 = i \), then the initial distribution \( \mu^{(0)} \) is the row-vector \( e(i) \), where \( e(i) \) has a ‘1’ in the \( i \)th coordinate and a ‘0’ in the other coordinate (so \( e(1) = (1, 0) \) and \( e(2) = (0, 1) \)), so we have

\[
P(X_t = j|X_0 = i) = (\mu^{(t)})_j = (\mu^{(0)}P^t)_j = (e(i)P^t)_j = (P^t)_{i,j}.
\]

So, for example, if we know for sure that day 0 is sunny, then the probability that day \( t \) is rainy is

\[
P(X_t = 2|X_0 = 1) = (P^t)_{1,2} = \frac{1}{3} - \frac{2}{3}10^{-t},
\]

which converges to 1/3 as \( t \to \infty \). Similarly, if we know for sure that day 0 is rainy, then the probability that day \( t \) is rainy is

\[
P(X_t = 2|X_0 = 2) = (P^t)_{2,2} = \frac{1}{3} + \frac{2}{3}10^{-t},
\]

which also converges to 1/3 as \( t \to \infty \).

So according to this model, if we observe the weather over a long time, we should expect that about a third of the days will be rainy (so about two-thirds will be sunny), regardless of the weather on day 0. Do you think this is a realistic result for the UK? ;)

19
The long-term behaviour of a Markov chain

Limiting distributions

Definition 1.14. A probability vector is a row-vector whose entries are non-negative real numbers summing to 1.

Remark. Note that a probability vector with $n$ entries is equivalent to a probability distribution on a set of size $n$.

Definition 1.15. Let $(X_0, X_1, X_2, \ldots)$ be a Markov chain with finite state-space $S$. We say a probability vector $w$ (with entries indexed by $S$) is a limiting distribution for the chain if for any $i, j \in S$, we have

\[ P(X_t = j \mid X_0 = i) \to w_j \quad \text{as } t \to \infty. \]

Example 10. For the simple weather model above, we saw that for each $i \in \{1, 2\}$, we have

\[ P(X_t = 2 \mid X_0 = i) \to 1/3 \quad \text{as } t \to \infty, \]

so

\[ P(X_t = 1 \mid X_0 = i) = 1 - P(X_t = 2 \mid X_0 = i) \to 2/3 \quad \text{as } t \to \infty. \]

Hence, \((2/3, 1/3)\) is a limiting distribution for this Markov chain.

Some Markov chains do not have a limiting distribution:

Example 11. The Markov chain $(X_0, X_1, X_2, \ldots)$ with state-space \(\{1, 2\}\) and transition matrix

\[
\begin{pmatrix}
0 & 1 \\
1 & 0
\end{pmatrix}
\]

has no limiting distribution. Here is a proof of this fact. If $X_0 = 1$, then $X_t = 1$ for all even $t$ and $X_t = 2$ for all odd $t$, so

\[ P(X_t = 1 \mid X_0 = 1) = \begin{cases} 1 & \text{if } t \text{ is even;} \\ 0 & \text{if } t \text{ is odd.} \end{cases} \]

Hence, $P(X_t = 1 \mid X_0 = 1)$ does not converge as $t \to \infty$, so there is no limiting distribution.

You may object that the Markov chain in Example 11 involves no ‘genuine’ probability, because if we know the state at time 0, then we know (for sure) the state at time $t$ for any $t \in \mathbb{N}$: all the transition probabilities are either 1 or 0. Here is another slightly more complicated example of a Markov chain with no limiting distribution.

Example 12. The Markov chain $(X_0, X_1, X_2, \ldots)$ with state-space \(\{1, 2, 3, 4\}\) and transition matrix

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

20
has no limiting distribution. Here is a proof of this fact. Observe that at every step, with probability 1 the parity of the state changes, because the only non-zero transition probabilities are from odd-numbered states to even-numbered states, and from even-numbered states to odd-numbered states. (Recall that the parity of a number is said to be ‘odd’ if the number is odd, and ‘even’ if the number is even.)

Hence,
\[ P(X_t \text{ is odd} \mid X_0 = 1) = \begin{cases} 1 & \text{if } t \text{ is even;} \\ 0 & \text{if } t \text{ is odd}. \end{cases} \]

This means there can be no limiting distribution. Indeed, suppose for a contradiction that \( \mathbf{w} = (w_1, w_2, w_3, w_4) \) is a limiting distribution. Then we must have
\[ P(X_t \text{ is odd} \mid X_0 = 1) = P(X_t = 1 \mid X_0 = 1) + P(X_t = 3 \mid X_0 = 1) \]
\[ \rightarrow w_1 + w_3 \text{ as } t \rightarrow \infty, \]
but we have just seen that the left-hand side alternates between 1 and 0, so it cannot converge to any real number, so we have a contradiction.

The following simple lemma allows us to recognise that a Markov chain has a limiting distribution if we can calculate \( P^t \) (or, more precisely, we can find a nice formula for the entries of \( P^t \), as we did with the simple weather model). Recall that if \( M_0, M_1, M_2, \ldots \) is a sequence of \( n \times n \) complex matrices, and \( L \) is an \( n \times n \) complex matrix, then we say that \( M_t \) converges to \( L \) as \( t \rightarrow \infty \) (and we write \( M_t \rightarrow L \) as \( t \rightarrow \infty \)) if for each \( i, j \) we have \( (M_t)_{i,j} \rightarrow L_{i,j} \) as \( t \rightarrow \infty \).

**Lemma 1.10.** Let \( (X_0, X_1, X_2, \ldots) \) be a Markov chain with state-space \( S = \{1, 2, \ldots, n\} \) and transition matrix \( P \). Then a probability vector
\[ \mathbf{w} = (w_1, w_2, \ldots, w_n) \]
is a limiting distribution for the Markov chain if and only if
\[ P^t \rightarrow \begin{pmatrix} w_1 & w_2 & \cdots & w_n \\ w_1 & w_2 & \cdots & w_n \\ \vdots & \vdots & \ddots & \vdots \\ w_1 & w_2 & \cdots & w_n \end{pmatrix} \text{ as } t \rightarrow \infty. \]

**Proof.** Let \( \mathbf{w} = (w_1, w_2, \ldots, w_n) \) be a probability vector. Then \( \mathbf{w} \) is a limiting distribution for \( (X_0, X_1, X_2, \ldots) \) if and only if for all \( i, j \in S \) we have
\[ P(X_t = j \mid X_0 = i) \rightarrow w_j \text{ as } t \rightarrow \infty; \]
since \( P(X_t = j \mid X_0 = i) = (P^t)_{i,j} \) for all \( i, j \), this happens if and only if \( P^t \rightarrow L \) as \( t \rightarrow \infty \), where
\[ L = \begin{pmatrix} w_1 & w_2 & \cdots & w_n \\ w_1 & w_2 & \cdots & w_n \\ \vdots & \vdots & \ddots & \vdots \\ w_1 & w_2 & \cdots & w_n \end{pmatrix}. \]
\[ \square \]
Example 13. In the simple weather model in Example 9, we had
\[ P^t = \left( \frac{2}{3} + \frac{1}{3} 10^{-t}, \frac{1}{3} - \frac{1}{3} 10^{-t} \right) \rightarrow \left( \frac{2}{3}, \frac{1}{3} \right) \text{ as } t \rightarrow \infty, \]
so we can see from this also that $\left( \frac{2}{3}, \frac{1}{3} \right)$ is a limiting distribution for the Markov chain.

Here is another equivalent condition for a probability vector being a limiting distribution. Recall that if $v_0, v_1, v_2, \ldots$ is a sequence of complex row-vectors with $n$ entries, and $w$ is a complex row-vector with $n$ entries, then we say $v_t \rightarrow w$ as $t \rightarrow \infty$ if for all $i$, we have $(v_t)_i \rightarrow w_i$ as $t \rightarrow \infty$.

Lemma 1.11. Let $(X_0, X_1, X_2, \ldots)$ be a Markov chain with state-space $S = \{1, 2, \ldots, n\}$ and transition matrix $P$. Let $w = (w_1, w_2, \ldots, w_n)$ be a probability vector. Then $w$ is a limiting distribution for the Markov chain if and only if for any initial distribution $\mu^{(0)}$, the distributions $\mu^{(t)}$ satisfy
\[ \mu^{(t)} \rightarrow w \text{ as } t \rightarrow \infty. \]

Proof. Suppose $w$ is a limiting distribution. Let
\[ L = \begin{pmatrix} w_1 & \ldots & w_n \\ w_1 & \ldots & w_n \\ \vdots & \ddots & \vdots \\ w_1 & \ldots & w_n \end{pmatrix}. \]
Let $\mu^{(0)}$ be any initial distribution. Then
\[ \mu^{(t)} = \mu^{(0)} P^t \rightarrow \mu^{(0)} L \text{ as } t \rightarrow \infty, \]
using the fact that $P^t \rightarrow L$ as $t \rightarrow \infty$, and continuity. Now for all $j \in S$, we have
\[ (\mu^{(0)} L)_j = \sum_{i=1}^n (\mu^{(0)})_i L_{ij} = \sum_{i=1}^n (\mu^{(0)})_i w_j = w_j \sum_{i=1}^n (\mu^{(0)})_i = w_j \cdot 1 = w_j, \]
so $\mu^{(0)} L = w$. Hence, $\mu^{(t)} \rightarrow w$ as $t \rightarrow \infty$.

Conversely, if $\mu^{(t)} \rightarrow w$ as $t \rightarrow \infty$ for any initial distribution $\mu^{(0)}$, then taking $\mu^{(0)} = (0, \ldots, 0, 1, 0, \ldots, 0) = e(i)$ to be the probability vector with a 1 in the $i$th coordinate and a 0 in all other coordinates, we get
\[ P(X_t = j|X_0 = i) = (P^t)_{i,j} = (e(i) P^t)_j = (\mu^{(0)} P^t)_j = (\mu^{(t)})_j \rightarrow w_j \text{ as } t \rightarrow \infty, \]
so $w$ is a limiting distribution, as required.

Equilibrium distributions

Definition 1.16. Let \((X_0, X_1, X_2, \ldots)\) be a Markov chain with finite state-space \(S\) and transition matrix \(P\). We say a probability vector \(w\) (with entries indexed by \(S\)) is an equilibrium distribution for the chain if \(wP = w\).

Remark. An equilibrium distribution is also called an invariant distribution or a stationary distribution.

The name 'equilibrium distribution' is appropriate because if a Markov chain starts an at equilibrium distribution, then it has that distribution forever:

Lemma 1.12. Let \((X_0, X_1, X_2, \ldots)\) be a Markov chain with finite state-space \(S\) and transition matrix \(P\). Suppose \(w\) is an equilibrium distribution for the Markov chain. If \(\mu^{(0)} = w\), then \(\mu^{(t)} = w\) for all \(t \in \mathbb{N}\).

Proof. By induction on \(t\). For \(t = 0\), the statement holds by assumption. For the inductive step, let \(t \in \mathbb{N}\) and suppose that \(\mu^{(t-1)} = w\). Then
\[
\mu^{(t)} = \mu^{(t-1)}P = wP = w.
\]

Our next result says that if a Markov chain with a finite state space has a limiting distribution \(w\), then \(w\) is also its unique equilibrium distribution.

Lemma 1.13. Let \((X_0, X_1, X_2, \ldots)\) be a Markov chain with finite state-space \(S\) and transition matrix \(P\). Suppose \(w\) is a limiting distribution for the Markov chain. Then \(w\) is the unique equilibrium distribution.

Proof. Suppose \(w\) is a limiting distribution for the Markov chain. Then for any initial distribution \(\mu^{(0)}\), we have
\[
\mu^{(t)} = \mu^{(0)}P^t \to w \quad \text{as } t \to \infty,
\]
so by continuity,
\[
\mu^{(t+1)} = \mu^{(0)}P^{t+1} = (\mu^{(0)}P^t)P \to wP \quad \text{as } t \to \infty,
\]
but also
\[
\mu^{(t+1)} \to w \quad \text{as } t \to \infty,
\]
so \(wP = w\), i.e. \(w\) is an equilibrium distribution, as required.

Now suppose \(w'\) is another equilibrium distribution. Then if \(\mu^{(0)} = w'\), we have
\[
w' = w'P^t = \mu^{(0)}P^t \to w \quad \text{as } t \to \infty,
\]
so \(w' = w\). So \(w\) is indeed the unique equilibrium distribution.
A Markov chain may have an equilibrium distribution but no limiting distribution:

**Example 14.** Let \((X_0, X_1, X_2, \ldots)\) be the Markov chain with state-space \(\{1, 2, 3, 4\}\) and transition matrix
\[
P = \begin{pmatrix}
0 & 1/2 & 0 & 1/2 \\
1/2 & 0 & 1/2 & 0 \\
0 & 1/2 & 0 & 1/2 \\
1/2 & 0 & 1/2 & 0
\end{pmatrix}.
\]

A vector \(w = (w_1, w_2, w_3, w_4)\) is an equilibrium distribution if it is a probability vector (i.e. \(\sum_{i=1}^{4} w_i = 1\) and \(w_i \geq 0\) for all \(i\)), and also \(wP = w\). So it must satisfy the following equations:
\[
\begin{pmatrix}
0 & 1/2 & 0 & 1/2 \\
1/2 & 0 & 1/2 & 0 \\
0 & 1/2 & 0 & 1/2 \\
1/2 & 0 & 1/2 & 0
\end{pmatrix}
\begin{pmatrix}
w_1 \\
w_2 \\
w_3 \\
w_4
\end{pmatrix}
= \begin{pmatrix}w_1 \\
w_2 \\
w_3 \\
w_4\end{pmatrix},
\text{ where } \sum_{i=1}^{4} w_i = 1.
\]

This gives
\[
\frac{1}{2}w_2 + \frac{1}{2}w_4 = w_1,
\frac{1}{2}w_1 + \frac{1}{2}w_3 = w_2,
\frac{1}{2}w_2 + \frac{1}{2}w_4 = w_3,
\frac{1}{2}w_1 + \frac{1}{2}w_3 = w_4,
\]
\[
w_1 + w_2 + w_3 + w_4 = 1.
\]

Combining the first and third equations gives \(w_1 = w_3\), combining the second and fourth gives \(w_2 = w_4\), and then substituting \(w_2 = w_4\) into the first equation gives \(w_2 = w_1\), so \(w_1 = w_2 = w_3 = w_4\). Substituting these equalities into the fifth equation gives \(4w_1 = 1\), so the equations force \((w_1, w_2, w_3, w_4) = (1/4, 1/4, 1/4, 1/4)\). We easily check that this is indeed a solution of the equations above; it clearly also satisfies \(w_i \geq 0\) for all \(i\). Hence, \(w = (1/4, 1/4, 1/4, 1/4)\) is the unique equilibrium distribution for the Markov chain.

However, this Markov chain has no limiting distribution, by the same argument as in Example 12. As in Example 12, at every step, with probability 1 the parity of the state changes, because the only non-zero transition probabilities are from odd-numbered states to even-numbered states, and from even-numbered states to odd-numbered states.

Hence,
\[
\mathbb{P}(X_t \text{ is odd } | \ X_0 = 1) = \begin{cases} 1 & \text{if } t \text{ is even;} \\ 0 & \text{if } t \text{ is odd.} \end{cases}
\]

So by the same argument as in Example 12, there is no limiting distribution.
A sufficient condition for the existence of a unique equilibrium distribution: irreducibility

**Definition 1.17.** Let \((X_0, X_1, \ldots)\) be a Markov chain with state space \(S\) (finite or countable) and transition probabilities \((p_{i,j} : i, j \in S)\). We say the chain is *irreducible* if for any \(i, j \in S\), there exists \(r \in \mathbb{N} \cup \{0\}\) such that \(p_{i,j}^{(r)} > 0\).

Informally, a Markov chain is irreducible iff for any two states \(i\) and \(j\), we can get from \(i\) to \(j\) in some number \(r\) of steps (typically \(r\) will depend on the pair \((i, j)\)).

**Example 15.** Let \((X_0, X_1, X_2, \ldots)\) be the Markov chain with state-space \(S = \{0, 1, 2, 3, 4\}\) and transition graph as follows.

```
0 ----- 1/2 ----- 1 ----- 1/2 ----- 2 ----- 1/2 ----- 3 ----- 1/2 ----- 4 ----- 1/2
|           |           |           |           |           |           |           |           |
|           |           |           |           |           |           |           |           |
|           |           |           |           |           |           |           |           |
|           |           |           |           |           |           |           |           |
|           |           |           |           |           |           |           |           |

This Markov chain is irreducible. Indeed, for any \(i, j \in S\), we have

\[ p_{i,j}^{(|j-i|)} = \left(\frac{1}{2}\right)^{|j-i|} > 0. \]

**Example 16.** Let \((X_0, X_1, X_2, \ldots)\) be the Markov chain with state space \(\mathbb{Z}\) and transition probabilities defined by:

\[
\begin{align*}
p_{i,i+1} &= \frac{1}{2} & \forall i \in \mathbb{Z}; \\
p_{i,i-1} &= \frac{1}{2} & \forall i \in \mathbb{Z}; \\
p_{i,j} &= 0 & \forall i, j \in \mathbb{Z} : |i - j| \geq 2.
\end{align*}
\]

This is known as ‘the (symmetric) random walk on \(\mathbb{Z}\)’. This Markov chain is also irreducible. Indeed, for any \(i, j \in \mathbb{Z}\), we have

\[ p_{i,j}^{(|j-i|)} = \left(\frac{1}{2}\right)^{|j-i|} > 0. \]

Part of the importance of irreducibility stems from the following key theorem (which you should memorize).

**Theorem 1.14.** Let \((X_0, X_1, \ldots)\) be a Markov chain with finite state space \(S\) and transition probabilities \((p_{i,j} : i, j \in S)\). If it is irreducible, then it has a unique equilibrium distribution.

**Proof.** Non-examinable. (Next lecture.)
Example 17. Let \((X_0, X_1, \ldots)\) be the Markov chain with state-space \(S = \{1, 2\}\) and transition graph

\[
\begin{array}{c}
1/2 \quad 1/2 \\
\circ \quad \circ \\
1 \quad 2
\end{array}
\]

The transition matrix is

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

This is clearly not irreducible, as it has an absorbing state, the state 2. (If a Markov chain with more than one state has an absorbing state, \(i\) say, then it cannot be irreducible, since there is no way of getting from the state \(i\) to any other state, in any number of steps.)

However, it does have a unique equilibrium distribution, namely \((0, 1)\). This is clearly an equilibrium distribution: since the state 2 is absorbing, if we start in state 2 (with probability 1), then we remain in state 2 (with probability 1) at all future times. To see that it is unique, we solve the equations for an equilibrium distribution.

\[
\begin{pmatrix}
w_1 \\
w_2
\end{pmatrix}
\begin{pmatrix}
\frac{1}{2} & \frac{1}{2} \\
0 & 1
\end{pmatrix}
=
\begin{pmatrix}
w_1 \\
w_2
\end{pmatrix}, \\
w_1 + w_2 = 1.
\]

These give:

\[
\frac{1}{2}w_1 = w_1, \\
\frac{1}{2}w_1 + w_2 = w_2, \\
w_1 + w_2 = 1.
\]

The first equation gives \(w_1 = 0\); substituting this into the third gives \(w_2 = 1\), so we must have \((w_1, w_2) = (0, 1)\). Hence, \((0, 1)\) is the unique equilibrium distribution, as claimed.

Example 18. We know from Theorem 1.14 that the Markov chain in Example 15 has a unique equilibrium distribution. Let us find it. It has transition matrix

\[
P = \begin{pmatrix}
\frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 \\
\frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\
0 & 1/2 & 0 & \frac{1}{2} & 0 \\
0 & 0 & 1/2 & 0 & \frac{1}{2} \\
0 & 0 & 0 & \frac{1}{2} & 1/2
\end{pmatrix}.
\]

Since this matrix has all its column-sums equal to 1, the constant probability vector \(w = (1/5, 1/5, 1/5, 1/5, 1/5)\) satisfies

\[wP = w\]

so is an equilibrium distribution. We know that it is the unique such.
A sufficient condition for the existence of a limiting distribution: regularity

**Definition 1.18.** Let \((X_0, X_1, \ldots)\) be a Markov chain with state space \(S\) (finite or countable) and transition probabilities \((p_{i,j} : i, j \in S)\). We say the chain is regular if there exists \(r \in \mathbb{N}\) such that for any \(i, j \in S\), we have \(p_{i,j}^{(r)} > 0\).

**Remark.** Note that being regular is a stronger property than being irreducible, since for regularity, the same integer \(r\) must work for every pair of states \(i, j\).

**Remark.** Let \((X_0, X_1, \ldots)\) be a Markov chain with finite state space \(S\) and transition matrix \(P\). Then the Markov chain is regular if and only if there exists \(r \in \mathbb{N}\) such that \(P^r\) has all its entries greater than zero.

**Example 19.** Let \((X_0, X_1, \ldots)\) be the Markov chain with transition graph as follows.

The transition matrix is

\[
P = \begin{pmatrix}
0 & 1 & 0 & 0 \\
0 & 0 & 1/2 & 1/2 \\
1/2 & 0 & 0 & 1/2 \\
1/2 & 0 & 1/2 & 0
\end{pmatrix}.
\]

We will check that the chain is regular by showing that \(P^4\) has all its entries positive.
We have

\[ P^2 = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 1/2 & 0 & 0 & 1/2 \\ 1/2 & 0 & 1/2 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/4 & 1/4 \\ 1/4 & 1/2 & 1/4 & 0 \\ 1/4 & 1/2 & 0 & 1/4 \end{pmatrix} = \begin{pmatrix} 0 & 0 + + \\ + & 0 + + \\ + + + 0 \\ + + 0 + \end{pmatrix}, \]

\[ \text{sp} \]

\[ P^4 = (P^2)^2 = \begin{pmatrix} 0 & 0 + + \\ + & 0 + + \\ + + + 0 \\ + + 0 + \end{pmatrix} \begin{pmatrix} 0 & 0 + + \\ + & 0 + + \\ + + + 0 \\ + + 0 + \end{pmatrix} = \begin{pmatrix} + & + + + \\ + + + + \\ + + + + \\ + + + + \end{pmatrix} \]

has all its entries positive, as required.

Note that when checking for regularity we need only keep track of which entries are positive and which are zero (we need not calculate them exactly).

The following important theorem is what makes regularity useful in the study of Markov chains.

**Theorem 1.15.** Let \((X_0, X_1, \ldots)\) be a regular Markov chain with finite state space \(S\) and transition probabilities \((p_{i,j} : i, j \in S)\). Then the Markov chain has a limiting distribution.

**Proof.** Non-examinable. (See lectures for a sketch-proof using the Jordan Normal Form of a matrix.) \(\square\)

Since a Markov chain with limiting distribution \(w\) has \(w\) as its unique equilibrium distribution (see Lemma 1.13), Theorem 1.15 immediately gives the following useful corollary.

**Corollary 1.16.** Let \((X_0, X_1, \ldots)\) be a regular Markov chain with finite state space \(S\) and transition probabilities \((p_{i,j} : i, j \in S)\). Then the Markov chain has a limiting distribution which is also its unique equilibrium distribution.

Corollary 1.16 gives us an easy way of finding the limiting distribution of a regular Markov chain with a finite state space: we just need to calculate its unique equilibrium distribution, which we can do by solving a system of simultaneous linear equations! Here is an example.
Example 20. Let \((X_0, X_1, \ldots)\) be the Markov chain with state space \(\{1, 2, 3, 4\}\) and transition graph as follows.

First, we claim that this Markov chain is irreducible; indeed, we can get from any state to any other state in at most three steps:

\[
\begin{align*}
1 \to 3 \to 2; \\
1 \to 3; \\
1 \to 4; \\
2 \to 1; \\
2 \to 1 \to 3; \\
2 \to 1 \to 4; \\
3 \to 2 \to 1; \\
3 \to 2; \\
3 \to 2 \to 1 \to 4; \\
4 \to 3 \to 2 \to 1; \\
4 \to 3 \to 2; \\
4 \to 3.
\end{align*}
\]

Notice that we can get from any state \(i\) to state 1 in at most three steps, and from 1 to any state \(j\) in at most two steps.

We now claim that the chain is regular, indeed, that for any two states \(i, j\), we can get from \(i\) to \(j\) in exactly 5 steps. To do this, first go from state \(i\) to state 1 in \(s \leq 3\) steps, and then loop around the state 1 a total of \(5 - s - t\) times, and then go from state 1 to state \(j\) in \(t\) steps.

Since the chain is regular, by Corollary 1.16, it has a limiting distribution \(w = (w_1, w_2, w_3, w_4)\) which is equal to its unique equilibrium distribution. The transition
matrix is

\[ P = \begin{pmatrix} 1/8 & 0 & 1/2 & 3/8 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix}. \]

Hence, \( w \) satisfies

\[ wP = w, \quad \sum_{i=1}^{4} w_i = 1, \]

i.e.

\[
\begin{align*}
\frac{1}{8}w_1 + w_2 &= w_1, \\
& \quad w_3 = w_2, \\
\frac{1}{2}w_1 + w_4 &= w_3, \\
\frac{3}{8}w_1 &= w_4, \\
w_1 + w_2 + w_3 + w_4 &= 1.
\end{align*}
\]

Solving these simultaneous linear equations gives

\[ w = (8/25, 7/25, 7/25, 3/25). \]

**Remark.** Note that a Markov chain can have a limiting distribution without being regular, as can be seen from the following example.

**Example 21.** Let \((X_0, X_1, X_2, \ldots)\) be the Markov chain with state space \(S = \{1, 2\}\) and transition graph

```
1/2
1

2
```

Clearly, it is not regular; it is not even irreducible, as one cannot get from the state 2 to the state 1 in some number of steps (because the state 2 is an absorbing state). However, it does have a limiting distribution. To see this, let \(T\) be the time of first absorption in the state 2. Recall from Theorem 1.5 that for any initial distribution \(\mu^{(0)}\), we have \(\mathbb{P}(T < \infty) = 1\). We will now give a direct proof of this fact.

First, notice that \(\mathbb{P}(T = 0 | X_0 = 2) = 1\), so certainly \(\mathbb{P}(T < \infty | X_0 = 2) = 1\). Now we consider \(\mathbb{P}(T < \infty | X_0 = 1)\). Suppose that \(X_0 = 1\). Then for any \(t \in \mathbb{N}\), we have \(T \geq t\) if and only if the Markov chain loops around the state 1 at step 1, at step 2, \ldots, and at step \(t - 1\). Hence,

\[ \mathbb{P}(T \geq t) = \left(\frac{1}{2}\right)^{t-1} \quad \forall t \in \mathbb{N}. \]

Hence, for any \(t \in \mathbb{N}\), we have

\[ \mathbb{P}(T = t) = \mathbb{P}(T \geq t) - \mathbb{P}(T \geq t + 1) = \left(\frac{1}{2}\right)^{t-1} - \left(\frac{1}{2}\right)^t = \left(\frac{1}{2}\right)^t. \]
Hence,

\[ P(T < \infty) = \sum_{t=1}^{\infty} P(T = t) = \sum_{t=1}^{\infty} \left( \frac{1}{2} \right)^t = 1. \]

We therefore have \( P(T < \infty | X_0 = i) = 1 \) for each \( i \in \{ 1, 2 \} \). Hence, for any initial distribution \( \mu^{(0)} \), we have

\[
P(T < \infty) = P(T < \infty | X_0 = 1) \cdot P(X_0 = 1) + P(T < \infty | X_0 = 2) \cdot P(X_0 = 2)
= 1 \cdot \mu_1^{(0)} + 1 \cdot \mu_2^{(0)}
= 1,
\]

as claimed.

Now notice that

\[
P(X_t = 1 | X_0 = 1) = P(T \geq t + 1) = \left( \frac{1}{2} \right)^t \to 0 \quad \text{as} \quad t \to \infty,
\]

and obviously

\[
P(X_t = 1 | X_0 = 2) = 0 \quad \forall t \in \mathbb{N}.
\]

It follows that \((0, 1)\) is a limiting distribution for this Markov chain.

We now give a proof of Theorem 1.14 using Theorem 1.15.

**Proof of Theorem 1.14.** Let \((X_0, X_1, \ldots)\) be an irreducible Markov chain with finite state space \( S \) and transition matrix \( P \). Let \((Y_0, Y_1, \ldots)\) be the Markov chain with state space \( S \) and transition matrix

\[
Q = \frac{1}{2}(P + I),
\]

where \( I \) is the identity matrix. Then the Markov chain \((Y_0, Y_1, \ldots)\) is irreducible (since \((X_0, X_1, \ldots)\) is, and \( p_{i,j} > 0 \Rightarrow q_{i,j} > 0 \)), and moreover it has a loop at every state. By Question 2 on Exercise Sheet 3, an irreducible Markov chain with finite state space and a loop at some state, is regular. It follows that \((Y_0, Y_1, \ldots)\) is regular. Hence, by Corollary 1.16, it has a limiting distribution \( w \) which is also its unique equilibrium distribution.

Now observe that a probability vector \( u \) is an equilibrium distribution for \((X_0, X_1, \ldots)\) if and only if it is an equilibrium distribution for \((Y_0, Y_1, \ldots)\). Indeed, suppose \( u \) is an equilibrium distribution for \((X_0, X_1, \ldots)\); then \( uP = u \), so

\[
uQ = u\left( \frac{1}{2}(P + I) \right) = \frac{1}{2}uP + \frac{1}{2}uI = \frac{1}{2}u + \frac{1}{2}u = u,
\]

so \( u \) is an equilibrium distribution for \((Y_0, Y_1, \ldots)\). Conversely, suppose \( u \) is an equilibrium distribution for \((Y_0, Y_1, \ldots)\); then \( uQ = u \), so \( u\left( \frac{1}{2}(P + I) \right) = u \), so \( \frac{1}{2}uP + \frac{1}{2}uI = u \), so \( uP = u \), so \( u \) is an equilibrium distribution for \((X_0, X_1, \ldots)\).

It follows that \( w \) is the unique equilibrium distribution for \((X_0, X_1, \ldots)\). \( \square \)
The proportion of time spent in each state in the long run

**Definition 1.19.** Let \((X_0, X_1, \ldots)\) be a Markov chain with state space \(S\) and transition probabilities \((p_{i,j} : i, j \in S)\), and let \(i \in S\). We define

\[
V_i(T) = \# \{ t \in \{0, 1, 2, \ldots, T - 1\} : X_t = i \}
\]

to be the number of visits to state \(i\) before time \(T\).

Notice that \(V_i(T)\) is a random variable which depends only upon the random variables \(X_0, X_1, \ldots, X_{T-1}\). The random variable \(V_i(T)/T\) is the proportion of time which the Markov chain spends in the state \(i\), up until time \(T - 1\).

The following important (and useful) theorem tells us about the behaviour \(V_i(T)/T\), for large \(T\).

**Theorem 1.17 (Ergodic theorem).** Let \((X_0, X_1, \ldots)\) be an irreducible Markov chain with finite state space \(S\) and transition matrix \(P\). Let \(w\) be its unique equilibrium distribution. Then for any \(i \in S\),

\[
P \left( \frac{V_i(T)}{T} \to w_i \text{ as } T \to \infty \right) = 1.
\]

Moreover,

\[
E \left[ \frac{V_i(T)}{T} \right] \to w_i \text{ as } T \to \infty.
\]

**Proof.** Non-examinable. (Will be discussed later, time permitting.) \(\square\)

Informally, this theorem says that if a Markov chain with finite state space is irreducible, then in the long run, for each state \(i\) the proportion of time it spends in the state \(i\) is approximately \(w_i\), with probability 1.

Here is an application of this theorem.

**Example 22 (The Ehrenfest Urn model.).** Suppose there are two urns, urn A and urn B, containing a total of four balls. Every minute, I choose uniformly at random one of the four balls (so I choose each with probability \(1/4\)), and I move it from the urn it is currently in, to the other urn. Estimate the proportion of time for which urn A is empty, over a very long period of time.

We model this situation as a Markov chain \((X_0, X_1, \ldots)\) with 5 states; state \(i\) is where urn A contains exactly \(i\) balls. So \(S = \{0, 1, 2, 3, 4\}\), and the transition graph is
The transition matrix is therefore
\[
P = \begin{pmatrix}
0 & 1 & 0 & 0 & 0 \\
1/4 & 0 & 3/4 & 0 & 0 \\
0 & 1/2 & 0 & 1/2 & 0 \\
0 & 0 & 3/4 & 0 & 1/4 \\
0 & 0 & 0 & 1 & 0
\end{pmatrix}.
\]

This Markov chain is clearly irreducible, so it has a unique equilibrium distribution, \(w = (w_0, w_1, w_2, w_3, w_4)\) say, and \(w\) satisfies the equations \(wP = w\), \(\sum_{i=0}^{4} w_i = 1\), i.e.
\[
\begin{align*}
\frac{1}{4}w_1 &= w_0, \\
w_0 + \frac{1}{2}w_2 &= w_1, \\
\frac{3}{4}w_1 + \frac{3}{4}w_3 &= w_2; \\
\frac{1}{2}w_2 + w_4 &= w_3; \\
\frac{1}{4}w_3 &= w_4, \\
w_0 + w_1 + w_2 + w_3 + w_4 &= 1.
\end{align*}
\]
The first and fifth equations give \(w_1 = 4w_0\), \(w_3 = 4w_4\) respectively. Substituting the first equation into the second gives \(w_0 + \frac{1}{2}w_2 = 4w_0\) so \(w_2 = 6w_0\), and similarly substituting the fifth equation into the fourth gives \(w_2 = 6w_4\). Hence, \(w_4 = w_0\) and so \(w_1 = w_3 = 4w_0\), and \(w_2 = 6w_0\). Substituting these into the last equation gives \(16w_0 = 1\), so \(w_0 = 1/16\) and therefore
\[
w = (1/16, 4/16, 6/16, 4/16, 1/16) = (1/16, 1/4, 3/8, 1/4, 1/16).
\]

In particular, \(w_0 = 1/16\), so by the previous theorem, we have
\[
\mathbb{P} \left( \frac{\#\{t \in \{0,1,2,\ldots,T-1\} : \text{urn A is empty at time } t\}}{T} \to \frac{1}{16} \text{ as } T \to \infty \right) = 1,
\]
and moreover
\[
\mathbb{E} \left[ \frac{\#\{t \in \{0,1,2,\ldots,T-1\} : \text{urn A is empty at time } t\}}{T} \right] \to w_i \text{ as } T \to \infty.
\]
Hence, over a very long period of time, urn A will be empty roughly \((1/16)\)th of the time, with probability 1.

**Equilibrium distributions when the state-space is countably infinite**

**Definition 1.20.** Let \(S\) be a countably infinite state space. A **probability vector for** \(S\) is a row-vector \(w\) with entries indexed by \(S\), such that \(w_i \geq 0\) for all \(i \in S\) and \(\sum_{i \in S} w_i = 1\).

**Remark.** Note that, as in the case where \(S\) is finite, a probability vector for \(S\) represents a probability distribution on \(S\).
Definition 1.21. Let \((X_0, X_1, \ldots)\) be a Markov chain with state space \(S\) and transition probabilities \((p_{i,j} : \ i,j \in S)\), and let \(w\) be a probability vector for \(S\). We say \(w\) is an equilibrium distribution for the Markov chain if for all \(j \in S\), we have
\[
\sum_{i \in S} w_i p_{i,j} = w_j.
\]

Remark. Notice that these are exactly the same equations as in the case of finite \(S\).

The following lemma says that, as with the case of finite \(S\), a probability vector \(w\) is an equilibrium distribution if and only if, whenever the Markov chain has initial distribution \(w\), it has distribution \(w\) for all future times. (This justifies the name ‘equilibrium distribution’.)

Lemma 1.18. Let \((X_0, X_1, \ldots)\) be a Markov chain with state space \(S\) and transition probabilities \((p_{i,j} : \ i,j \in S)\), and let \(w\) be a probability vector for \(S\). Then \(w\) is an equilibrium distribution for the Markov chain if and only if \(\mu^{(0)} = w\) implies \(\mu^{(t)} = w\) for all \(t \in \mathbb{N}\).

Proof. Suppose \(w\) is an equilibrium distribution for the chain. Suppose \(\mu^{(0)} = w\). We claim that for all \(t \in \mathbb{N} \cup \{0\}\), we have \(\mu^{(t)} = w\). The proof is by induction on \(t\). The base case \(t = 0\) of this claim holds by assumption. Assume the statement holds for \(t - 1\). Then
\[
\mathbb{P}(X_t = j) = \sum_{i \in S} \mathbb{P}(X_t = j, X_{t-1} = i)
= \sum_{i \in S} \mathbb{P}(X_t = j \mid X_{t-1} = i) \mathbb{P}(X_{t-1} = i)
= \sum_{i \in S} p_{i,j} \mu^{(t-1)}_i
= \sum_{i \in S} p_{i,j} w_i
= w_j,
\]
so \(\mu^{(t)} = w\), so the statement holds for \(t\) as well. This proves the claim.

Conversely, suppose that \(\mu^{(0)} = w\) implies \(\mu^{(t)} = w\) for all \(t \in \mathbb{N}\). Suppose \(\mu^{(0)} = w\).
Then for all \( j \in S \), we have

\[
\begin{align*}
w_j &= \mu_j^{(1)} \\
&= \mathbb{P}(X_1 = j) \\
&= \sum_{i \in S} \mathbb{P}(X_1 = j, X_0 = i) \\
&= \sum_{i \in S} \mathbb{P}(X_1 = j | X_0 = i) \mathbb{P}(X_0 = i) \\
&= \sum_{i \in S} p_{i,j} \mu_j^{(0)} \\
&= \sum_{i \in S} p_{i,j} w_j,
\end{align*}
\]

as required.

**Example 23** (The success-run chain). Suppose a biased coin has probability \( p \) of coming up heads and probability \( 1 - p \) of coming up tails. We toss it repeatedly. For each \( t \in \mathbb{N} \), we define \( X_t \) to be the number of consecutive heads we have just seen, up to and including the \( t \)th toss (if the \( t \)th toss is a tail or \( t = 0 \), then we define \( X_t = 0 \)). For example, if we throw \( THHTHHHTT \), then \( X_0 = 0, X_1 = 0, X_2 = 1, X_3 = 2, X_4 = 0, X_5 = 1, X_6 = 2, X_7 = 3, X_8 = 0, X_9 = 0 \).

It is easy to see that \((X_0, X_1, \ldots)\) is an irreducible Markov chain; it is called the *success-run chain*. Its state space is \( \mathbb{N} \cup \{0\} \), and its transition probabilities are:

\[
p_{i,0} = 1 - p \quad \forall i \in \mathbb{N} \cup \{0\}; \quad p_{i,i+1} = p \quad \forall i \in \mathbb{N} \cup \{0\}; \quad p_{i,j} = 0 \text{ for all other } i, j.
\]

An equilibrium distribution \( w = (w_0, w_1, w_2, \ldots) \) must satisfy

\[
\sum_{i=0}^{\infty} w_i = 1, \quad w_j = \sum_{i=0}^{\infty} w_i p_{i,j} \quad \forall j \in \mathbb{N} \cup \{0\}.
\]

When \( j = 0 \), the second equation reads

\[
w_0 = \sum_{i=0}^{\infty} w_i (1 - p) = (1 - p) \sum_{i=0}^{\infty} w_i;
\]

together with \( \sum_{i=0}^{\infty} w_i = 1 \) this implies \( w_0 = 1 - p \). For each \( j \geq 1 \), the second equation reads

\[
w_j = pw_{j-1}.
\]

This yields

\[
w_j = p^j (1 - p) \quad \forall j \in \mathbb{N} \cup \{0\}.
\]

This solution \( w \) satisfies all our equations: indeed, we have

\[
\sum_{i=0}^{\infty} w_i = \sum_{i=0}^{\infty} p^i (1 - p) = \frac{1 - p}{1 - p} = 1,
\]

as required.
so the first equation is satisfied. The second equation is therefore clearly satisfied, for
each \( j \in \mathbb{N} \cup \{0\} \). Hence, the distribution \( w = (w_0, w_1, w_2 \ldots) \) defined by
\[
w_j = p^j(1 - p) \quad \forall j \in \mathbb{N} \cup \{0\}
\]
is the unique equilibrium distribution.

Note that this \( w \) is the probability distribution of \( Z - 1 \), where \( Z \) is a geometric
random variable with success-probability \( 1 - p \).

Here is an example of an irreducible Markov chain with infinite state space and no
equilibrium distribution.

**Example 24** (The symmetric random walk on \( \mathbb{Z} \)). Consider the symmetric random
walk on \( \mathbb{Z} \), i.e. the Markov chain with state space \( \mathbb{Z} \) and transition probabilities
\[
p_{i,i+1} = 1/2 \quad \forall i \in \mathbb{Z}, \quad p_{i,i-1} = 1/2 \quad \forall i \in \mathbb{Z}, \quad p_{i,j} = 0 \quad \text{for all other} \; i, j.
\]
This is clearly irreducible, but it has no equilibrium distribution! To prove this, suppose
for a contradiction that there exists an equilibrium distribution, \( w \) say. Then
\[
\sum_{i \in \mathbb{Z}} w_i = 1, \quad w_j = \sum_{i \in \mathbb{Z}} w_i p_{i,j} \quad \forall j \in \mathbb{Z}, \quad w_i \geq 0 \quad \forall i \in \mathbb{Z}.
\]
The second equation translates to
\[
w_j = \frac{1}{2}w_{j-1} + \frac{1}{2}w_{j+1} \quad \forall j \in \mathbb{Z},
\]
which implies
\[
w_{j+1} - w_j = w_j - w_{j-1} \quad \forall j \in \mathbb{Z}.
\]
In other words, the difference between any two consecutive \( w_i \)'s is the same: there exists
\( c \in \mathbb{R} \) such that
\[
w_j - w_{j-1} = c \quad \forall j \in \mathbb{Z}.
\]
Hence, for any \( i, j \in \mathbb{Z} \), we have
\[
w_j = w_i + c(j - i).
\]
We split into three cases.

*Case 1: \( c > 0 \).* In this case, choose any \( i \in \mathbb{Z} \). Then for all \( j \in \mathbb{Z} \), we have
\[
w_j = w_i + c(j - i) = w_i - c(i - j),
\]
and this is negative if \( i - j \) is sufficiently large, a contradiction, since we assumed \( w_j \geq 0 \)
for all \( j \in \mathbb{Z} \).

*Case 2: \( c < 0 \).* In this case, choose any \( i \in \mathbb{Z} \). Then for all \( j \in \mathbb{Z} \), we have
\[
w_j = w_i + c(j - i),
\]
and this is negative if \( j - i \) is sufficiently large, a contradiction.

**Case 3:** \( c = 0 \). Then we have \( w_i = w_j \) for all \( i, j \in \mathbb{Z} \), so there exists \( a \in \mathbb{R} \) such that

\[
w_j = a \quad \forall j \in \mathbb{Z}.
\]

Note that \( a \geq 0 \). If \( a = 0 \), then \( \sum_{i \in \mathbb{Z}} w_i = 0 \), contradicting our assumption that \( \sum_{i \in \mathbb{Z}} w_i = 1 \), and if \( a > 0 \), then

\[
\sum_{i \in \mathbb{Z}} w_i = \sum_{i \in \mathbb{Z}} a = \infty,
\]

again contradicting our assumption that \( \sum_{i \in \mathbb{Z}} w_i = 1 \).

Hence, in all cases we have a contradiction, so there exists no equilibrium distribution.

The above example shows that the assumption of \( S \) being finite was necessary, in Theorem 1.14.

**The communicating classes of a Markov chain**

**Definition 1.22.** Let \((X_0, X_1, \ldots)\) be a Markov chain with state space \( S \) and transition probabilities \((p_{i,j} : i, j \in S)\), and let \( i, j \in S \). We say that \( i \) **communicates with** \( j \), and we write \( i \to j \), if there exists \( r \in \mathbb{N} \cup \{0\} \) such that \( p_{i,j}^{(r)} > 0 \).

**Definition 1.23.** Let \((X_0, X_1, \ldots)\) be a Markov chain with state space \( S \) and transition probabilities \((p_{i,j} : i, j \in S)\), and let \( i, j \in S \). We say that \( i \) and \( j \) **intercommunicate**, and we write \( i \leftrightarrow j \), if we have both \( i \to j \) and \( j \to i \).

**Remark.** It is easy to see that the relation \( \leftrightarrow \) is an equivalence relation on the state space \( S \) (meaning, it is reflexive, symmetric and transitive).

Recall that if \( R \) is an equivalence relation on a set \( \Omega \), then the distinct equivalence classes of \( R \) form a partition of \( \Omega \) into disjoint sets. (Recall that if \( y \in \Omega \), the equivalence class of \( y \) is the set \( \{ x \in \Omega : xRy \} \).) Hence, the equivalence classes of \( \leftrightarrow \) form a partition of \( S \) into disjoint sets.

**Definition 1.24.** The equivalence classes of \( \leftrightarrow \) are called the **communicating classes** of the Markov chain. In other words, for each \( i \in S \), the communicating class of \( i \) is the set \( \{ j \in S : i \leftrightarrow j \} \).

**Remark.** The distinct communicating classes form a partition of \( S \) into disjoint sets.

**Example 25.** Find the communicating classes of the Markov chain with transition graph below.

37
There are four communicating classes: \{0, 1, 2\}, \{3, 4\}, \{5\} and \{6\}. (Note that any absorbing state always forms a singleton communicating class.)

**Remark.** A Markov chain with state space \(S\) is irreducible if and only if for any two states \(i, j \in S\), we have \(i \leftrightarrow j\), i.e. if and only if there is just one communicating class: the whole of \(S\).

**Recurrence and transience**

Somewhat surprisingly, it turns out that for any state \(i\) of a Markov chain, if the Markov chain starts at the state \(i\) then either it returns to the state \(i\) infinitely many times with probability 1, or else it only returns to the state \(i\) finitely many times with probability 1. The states with the first property are the **recurrent states**, and the states with the second property are the **transient states**. Below we give the formal definitions of recurrence and transience.

**Definition 1.25.** Let \((X_0, X_1, X_2, \ldots)\) be a Markov chain with state space \(S\), and let \(i \in S\). For each \(t \in \mathbb{N}\), we define

\[
f_{i,i}^{(t)} = \Pr(X_1 \neq i, X_2 \neq i, \ldots, X_{t-1} \neq i, X_t = i \mid X_0 = i).
\]

This is the probability that the Markov chain returns to the state \(i\) for the very first time at time \(t\), given that it starts in the state \(i\).

**Definition 1.26.** Let \((X_0, X_1, X_2, \ldots)\) be a Markov chain with state space \(S\), and let \(i \in S\). We define

\[
f_{i,i} = \sum_{t=1}^{\infty} f_{i,i}^{(t)} = \Pr(\{\text{there exists } t \geq 1 : X_t = i\} \mid X_0 = i).
\]

This is the probability that the Markov chain ever returns to the state \(i\), given that it starts in the state \(i\).
Definition 1.27. Let \((X_0, X_1, X_2, \ldots)\) be a Markov chain with state space \(S\), and let \(i \in S\). We say that the state \(i\) is recurrent if \(f_{i,i} = 1\); otherwise (i.e. if \(f_{i,i} < 1\)), we say that the state \(i\) is transient.

Example 26. Let \((X_0, X_1, X_2, \ldots)\) be the Markov chain with state space \(\{1, 2, 3, 4\}\) and transition graph below; for each state, determine whether it is recurrent or transient.

First, we look at the state 3. We have \(f^{(1)}_{3,3} = \frac{1}{2}\), \(f^{(2)}_{3,3} = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}\), and

\[
f^{(t)}_{3,3} = \frac{1}{2} \cdot \left(\frac{1}{2}\right)^{t-2} \cdot \frac{1}{2} = 2^{-t} \quad \forall t \geq 2,
\]

since the only way of starting at 3 and returning to 3 for the first time at the time \(t\), is to go

\[3 \rightarrow 4 \rightarrow 4 \rightarrow \ldots \rightarrow 4 \rightarrow 3,
\]

looping \(t - 2\) times around the state 4.

Hence, we have

\[f_{3,3} = \sum_{t=1}^{\infty} f^{(t)}_{3,3} = \sum_{t=1}^{\infty} 2^{-t} = 1,
\]

so the state 3 is recurrent.

A very similar calculation shows that the state 4 is also recurrent.

We now look at the state 1. We have \(f^{(1)}_{1,1} = 0\) since there is no loop at the state 1. We have \(f^{(2)}_{1,1} = \frac{1}{2} \cdot \frac{1}{3} = \frac{1}{6}\), since if we start at the state 1, then the only way of returning to the state 1 for the first time in exactly two steps is to go \(1 \rightarrow 2 \rightarrow 1\). In fact, for each \(t \geq 2\), we have

\[f^{(t)}_{1,1} = \frac{1}{2} \cdot \left(\frac{1}{2}\right)^{t-2} \cdot \frac{1}{3} = \frac{1}{2} \left(\frac{1}{3}\right)^{t-1},
\]

39
since if we start at the state 1, then the only way of returning to the state 1 for the first
time in exactly $t$ steps is to go

$$1 \rightarrow 2 \rightarrow 2 \rightarrow \ldots \rightarrow 2 \rightarrow 1,$$

looping $t - 2$ times around the state 2. (Note that if we ever jump to the state 3 or to
the state 4, then we can never ever return to the state 1.)

Hence,

$$f_{1,1} = \sum_{t=1}^{\infty} f^{(t)}_{1,1} = \sum_{t=2}^{\infty} \left(\frac{1}{2}\right)^{t-1} = \frac{1/6}{1 - 1/3} = \frac{1}{4} < 1,$$

so the state 1 is transient.

Here is way of seeing straight away (without any calculation) that the state 1 is
transient. Given that we start at the state 1, the probability of jumping to the state 3
at the first step is $1/2$, and once we have done this we can never ever return to the state
1. So given that we start at the state 1, the probability that we never ever return to the
state 1 is at least $1/2$, so we must have

$$f_{1,1} \leq \frac{1}{2} < 1,$$

so we see from this also that the state 1 is transient.

A similar argument shows straight away that the state 2 is transient. Can you give
it?

The following simple lemma gives us a way of calculating $f^{(1)}_{i,i}, \ldots, f^{(r)}_{i,i}$ if we know
$p_{i,i}^{(1)}, \ldots, p_{i,i}^{(r)}$.

**Lemma 1.19.** Let $(X_0, X_1, X_2, \ldots)$ be a Markov chain with state space $S$ and transition
probabilities $(p_{i,j} : i, j \in S)$. Then for any $t \in \mathbb{N}$, we have

$$p_{i,i}^{(t)} = \sum_{k=1}^{t} f^{(k)}_{i,i} p_{i,i}^{(t-k)}.$$
Proof. We have

\[ p_{i,i}^{(t)} = \mathbb{P}(X_t = i \mid X_0 = i) \]

\[ = \sum_{k=1}^{t} \mathbb{P}(X_1 \neq i, X_2 \neq i, \ldots, X_{k-1} \neq i, X_k = i, X_t = i \mid X_0 = i) \]

(law of total probability)

\[ = \sum_{k=1}^{t} \mathbb{P}(X_1 \neq i, X_2 \neq i, \ldots, X_{k-1} \neq i, X_k = i \mid X_0 = i) \cdot \mathbb{P}(X_t = i \mid X_k = i) \]

(by the Markov property)

\[ = \sum_{k=1}^{t} f_{i,i}^{(k)} p_{i,i}^{(t-k)} \cdot \]

(definition of \( f_{i,i}^{(k)} \))

So, for example, we have

\[ p_{i,i}^{(1)} = f_{i,i}^{(1)} p_{i,i}^{(0)} = f_{i,i}^{(1)} \]

\[ p_{i,i}^{(2)} = f_{i,i}^{(1)} p_{i,i}^{(1)} + f_{i,i}^{(2)} p_{i,i}^{(0)} = f_{i,i}^{(1)} p_{i,i}^{(1)} + f_{i,i}^{(2)} \]

so rearranging the second equation and substituting the first equation into it gives:

\[ f_{i,i}^{(2)} = p_{i,i}^{(2)} - f_{i,i}^{(1)} p_{i,i}^{(1)} = p_{i,i}^{(2)} - \left( p_{i,i}^{(1)} \right)^2. \]

Stopping times

To study recurrence and transience further, we need to study stopping times for Markov chains.

Definition 1.28. Let \((X_0, X_1, X_2, \ldots)\) be a Markov chain with state space \(S\). Let \(T\) be a random variable taking values in \(\mathbb{N} \cup \{0\} \cup \{\infty\}\). We say that \(T\) is a stopping time for the Markov chain if for any \(k \in \mathbb{N} \cup \{0\}\), the event \(\{T = k\}\) is determined by the values of \(X_0, X_1, \ldots, X_k\).

Usually, a stopping time tells us the time of occurrence of some important event associated with the Markov chain. Intuitively, a random variable is a stopping time if for any integer \(k\), we can tell whether or not \(T = k\) just by observing the Markov chain up until the time \(k\).
**Example 27.** Let \((X_0, X_1, X_2, \ldots)\) be a Markov chain with state space \(\mathbb{N}\), and suppose \(X_0 = 2\). Let \(T = \min\{t \geq 1 : X_t = 2\}\) be the time of first return to the state 2. (We define \(T = \infty\) if \(X_t \neq 2\) for all \(t \geq 1\), i.e. if the Markov chain never returns to the state 2.) Then \(T\) is a stopping time for the Markov chain.

**Example 28 (Non-example).** Let \((X_0, X_1, X_2, \ldots)\) be an irreducible Markov chain with state space \(\mathbb{N}\), and suppose \(X_0 = 3\). Let \(R = \max\{t \in \mathbb{N} \cup \{0\} : X_t = 3\}\) be the time of the last visit to the state 3. (We define \(R = \infty\) if the Markov chain returns to the state 3 infinitely many times.) Then \(R\) is not a stopping time for the Markov chain, since for example the occurrence (or not) of the event \(\{R = 2\}\) is not determined only by the values of \(X_0, X_1\) and \(X_2\).

The following important theorem is a rather strong ‘memoryless property’ for Markov chains. It says that if \((X_0, X_1, X_2, \ldots)\) is a Markov chain and \(T\) is a stopping time for the chain, then conditioned on \(T\) being finite and \(X_T = i\), the behaviour of the Markov chain after the time \(T\) is independent of its behaviour before the time \(T\). It is a considerable strengthening of the Markov property, so we call it the ‘strong Markov property’.

**Theorem 1.20 (Strong Markov property).** Let \((X_0, X_1, X_2, \ldots)\) be a Markov chain with state space \(S\) and transition probabilities \((p_{i,j} : i, j \in S)\). Let \(i \in S\). Let \(T\) be a stopping time for the chain. Then, conditional on \(T < \infty\) and \(X_T = i\), the random process \((X_T, X_{T+1}, X_{T+2}, \ldots)\) is a Markov chain with initial distribution \(e_i\) and is independent of \((X_0, X_1, \ldots, X_{T-1})\). (Recall that \(e_i = (0,0,\ldots,0,1,0,\ldots,0)\), where the 1 is in the \(i\)th coordinate.)

**Proof.** Not given. \(\square\)

We will now use the Strong Markov property to study recurrence and transience.

**Definition 1.29.** Let \((X_0, X_1, X_2, \ldots)\) be a Markov chain with state space \(S\) and transition probabilities \((p_{i,j} : i, j \in S)\). Let \(i \in S\). Suppose that \(X_0 = i\). For each \(k \in \mathbb{N}\), we define the random variable \(T_i^{(k)}\) to be the time of the \(k\)th return to the state \(i\). (If the Markov chain returns to the state \(i\) at most \(k-1\) times, then we define \(T_i^{(k)} = \infty\).)

Moreover, for each \(k \in \mathbb{N}\) such that \(T_i^{(k)} < \infty\), we define the random variable

\[
S_i^{(k)} = \begin{cases} 
T_i^{(1)} & \text{if } k = 1; \\
T_i^{(k)} - T_i^{(k-1)} & \text{if } k \geq 2 
\end{cases}
\]

to be the length of the \(k\)th ‘excursion’ from the state \(i\), back to the state \(i\).

**Remark.** Notice that for each \(k \in \mathbb{N}\), \(T_i^{(k)}\) is a stopping time, and we have \(f_{i,i} = \mathbb{P}(T_i^{(1)} < \infty)\).

**Definition 1.30.** Let \((X_0, X_1, X_2, \ldots)\) be a Markov chain with state space \(S\) and transition probabilities \((p_{i,j} : i, j \in S)\). Let \(i \in S\). Suppose that \(X_0 = i\). We define the random variable \(V_i\) to be the number of returns to the state \(i\), i.e.

\[V_i = \#\{t \in \mathbb{N}: X_t = i\} \]
We now use the Strong Markov property to prove the following.

**Lemma 1.21.** For any \(k \in \mathbb{N}\), we have

\[
P(V_i \geq k) = (f_{i,i})^k.\]

**Proof.** By induction on \(k\). For \(k = 1\), we have \(P(V_i \geq 1) = f_{i,i}\), by definition. For the inductive step, assume that \(k \geq 2\) and that \(P(V_i \geq k - 1) = (f_{i,i})^{k-1}\). Then we have

\[
P(V_i \geq k) = P(V_i \geq k \mid V_i \geq k - 1) \cdot P(V_i \geq k - 1)
\]

\[
= P(T_i^{(k)} < \infty \mid T_i^{(k-1)} < \infty) \cdot P(V_i \geq k - 1)
\]

\[
= P(T_i^{(k)} < \infty \mid T_i^{(k-1)} < \infty) \cdot (f_{i,i})^{k-1} \quad \text{(by the induction hypothesis)}
\]

\[
= P(S_i^{(k)} < \infty \mid \{T_i^{(k-1)} < \infty, X_{T_i^{(k-1)}} = i\}) \cdot (f_{i,i})^{k-1}
\]

\[
= P(T_i^{(1)} < \infty \mid X_0 = i) \cdot (f_{i,i})^{k-1}
\]

(by applying the Strong Markov Property to the stopping time \(T_i^{(k-1)}\))

\[
= f_{i,i} \cdot (f_{i,i})^{k-1}
\]

\[
= (f_{i,i})^k,
\]

completing the inductive step and proving the lemma. \(\square\)

It follows from this lemma that

\[
P(V_i = k) = P(V_i \geq k) - P(V_i \geq k + 1) = (f_{i,i})^k - (f_{i,i})^{k+1} = (f_{i,i})^k(1 - f_{i,i}) \quad \forall k \in \mathbb{N}.
\]

Hence, if \(i\) is transient (i.e. \(f_{i,i} < 1\)), then \(V_i + 1\) has the geometric distribution with parameter (success probability) equal to \(1 - f_{i,i}\), and therefore

\[
E[V_i + 1] = \frac{1}{1 - f_{i,i}},
\]

so

\[
E[V_i] = \frac{1}{1 - f_{i,i}} - 1 < \infty.
\]

It follows that

\[
P(V_i = \infty) = 0, \quad P(V_i < \infty) = 1,
\]

so with probability 1, if the Markov chain starts at the state \(i\) then it returns to the state \(i\) only finitely many times. This motivates the name 'transient'.

On the other hand, if \(i\) is recurrent (i.e. \(f_{i,i} = 1\)), then \(P(V_i \geq k) = 1\) for all \(k \in \mathbb{N}\), so \(P(V_i = k - 1) = 0\) for all \(k \in \mathbb{N}\)! It follows that \(P(V_i < \infty) = 0\), so \(P(V_i = \infty) = 1\). Hence, \(E[V_i] = \infty\), and with probability 1, if the Markov chain starts at the state \(i\) then it returns to the state \(i\) infinitely many times.

43
To summarize, we have the following criteria for deciding whether a state is recurrent or transient:

The state $i$ is transient
1. If and only if $f_{i,i} < 1$;
2. If and only if $E[V_i] < \infty$;
3. If and only if $P(V_i < \infty) = 1$.

The state $i$ is recurrent
1. If and only if $f_{i,i} = 1$;
2. If and only if $E[V_i] = \infty$;
3. If and only if $P(V_i = \infty) = 1$.

Criteria 1 and 2 are more useful for calculations, but criterion 3 is interesting theoretically.

Criterion 2 leads to the following very useful and important theorem (which you should memorize!).

**Theorem 1.22.** Let $(X_0, X_1, X_2, \ldots)$ be a Markov chain with state space $S$ and transition probabilities $(p_{i,j} : i, j \in S)$. Let $i \in S$. Then $i$ is recurrent if and only if $\sum_{t=1}^{\infty} p_{i,i}^{(t)} = \infty$, and $i$ is transient if and only if $\sum_{t=1}^{\infty} p_{i,i}^{(t)} < \infty$.

**Proof.** Suppose $X_0 = i$. It suffices to prove that

$$E[V_i] = \sum_{t=1}^{\infty} p_{i,i}^{(t)}.$$  

To do this, for each $t \in \mathbb{N}$ we define the random variable $A_t$ by

$$A_t = \begin{cases} 1 & \text{if } X_t = i; \\ 0 & \text{if } X_t \neq i. \end{cases}$$

(The random variable $A_t$ is called the *indicator random variable* or the *indicator function* of the event $\{X_t = i\}$.) Notice that

$$V_i = \sum_{t=1}^{\infty} A_t,$$

so by the linearity of expectation, we have

$$E[V_i] = \sum_{t=1}^{\infty} E[A_t].$$
Now notice that
\[ E[A_t] = 1 \times P(X_t = i) + 0 \times P(X_t \neq i) = P(X_t = i) = p_{i,i}^{(t)}, \]
since \( X_0 = i \). Hence,
\[ E[V_t] = \sum_{t=1}^{\infty} p_{i,i}^{(t)}. \]

We are now done by Criterion 2, above. \(\square\)

Our first application of Theorem 1.22 is the following.

**Theorem 1.23.** Let \((X_0, X_1, X_2, \ldots)\) be a Markov chain with state space \(S\). Let \(i, j \in S\) such that \(i \leftrightarrow j\). Then \(i\) is recurrent if and only if \(j\) is recurrent.

**Proof.** Let \(i, j \in S\) such that \(i \leftrightarrow j\). Suppose that \(i\) is transient. Then by Theorem 1.22, we have
\[ \sum_{t=1}^{\infty} p_{i,i}^{(t)} < \infty. \]

Since \(i \rightarrow j\), there exists \(r \in \mathbb{N} \cup \{0\}\) such that \(p_{i,j}^{(r)} > 0\). Since \(j \rightarrow i\), there exists \(s \in \mathbb{N} \cup \{0\}\) such that \(p_{j,i}^{(s)} > 0\). Notice that \[ p_{i,i}^{(t)} \geq p_{i,j}^{(r)} p_{j,j}^{(t-r-s)} p_{j,i}^{(s)} \quad \forall t \geq r + s, \]
as one way of going from \(i\) to \(i\) in \(t\) steps is to go from \(i\) to \(j\) in \(r\) steps, and then go from \(j\) to \(j\) in \(t - r - s\) steps, and then go from \(j\) to \(i\) in \(s\) steps. So we have
\[ \sum_{t=r+s}^{\infty} p_{i,i}^{(t)} \geq \sum_{t=r+s}^{\infty} p_{i,j}^{(r)} p_{j,j}^{(t-r-s)} p_{j,i}^{(s)} \]
\[ = p_{i,j}^{(r)} p_{j,i}^{(s)} \sum_{t=r+s}^{\infty} p_{j,j}^{(t-r-s)} \]
\[ = p_{i,j}^{(r)} p_{j,i}^{(s)} \sum_{u=0}^{\infty} p_{j,j}^{(u)} \]

Hence,
\[ \sum_{u=0}^{\infty} p_{j,j}^{(u)} \leq \frac{1}{p_{i,j}^{(r)} p_{j,i}^{(s)}} \sum_{t=r+s}^{\infty} p_{i,i}^{(t)} < \infty. \]

Hence,
\[ \sum_{t=1}^{\infty} p_{j,j}^{(t)} = \sum_{u=0}^{\infty} p_{j,j}^{(u)} - 1 < \infty. \]
So by Theorem 1.22 again, \(j\) is transient.
We have shown that if \(i\) is transient, then \(j\) is also transient. The contrapositive of this statement is: if \(j\) is recurrent, then \(i\) is also recurrent. Interchanging \(i\) and \(j\), we may conclude that if \(i\) is recurrent, then \(j\) is also recurrent. Hence, \(i\) is recurrent if and only if \(j\) is recurrent. This proves the theorem.

The following are immediate corollaries of this theorem.

**Corollary 1.24.** Recurrence and transience are class properties, meaning that if \(C\) is a communicating class of a Markov chain, then either all the states in \(C\) are recurrent, or all the states in \(C\) are transient.

**Corollary 1.25.** If \((X_0, X_1, X_2, \ldots)\) is an irreducible Markov chain with state space \(S\), then either all the states in \(S\) are recurrent, or all the states in \(S\) are transient.

We now use Theorem 1.22 to analyse recurrence/transience for the symmetric random walk on \(\mathbb{Z}\).

**Example 29.** Recall that the symmetric random walk on \(\mathbb{Z}\) is the Markov chain with state space \(\mathbb{Z}\) and transition probabilities given by
\[p_{i,i+1} = 1/2 \forall i \in \mathbb{Z}, \quad p_{i,i-1} = 1/2 \forall i \in \mathbb{Z}, \quad p_{i,j} = 0 \text{ for all other } i, j.\]

Determine whether the state 0 is recurrent or transient.

Solution: We'll show that the state 0 is recurrent by showing that
\[\sum_{t=1}^{\infty} p_{0,0}^{(t)} = \infty.\]

I claim that
\[p_{0,0}^{(t)} = \begin{cases} \left(\frac{t}{t/2}\right) 2^{-t} & \text{if } t \text{ is even;} \\ 0 & \text{if } t \text{ is odd}. \end{cases}\]

To see this, observe that if we start at the state 0, then we must be at odd-numbered states at odd times, and at even-numbered states at even times. So we cannot be back at state 0 at an odd time, and therefore \(p_{0,0}^{(t)} = 0\) if \(t\) is odd.

If \(t\) is even, then the only way we can be back at the state 0 at time \(t\) is if we took exactly \(t/2\) jumps to the right and exactly \(t/2\) jumps to the left (in some order). For any sequence of \(t\) jumps containing \(t/2\) leftward jumps and \(t/2\) rightward jumps, the probability of this sequence of jumps occurring is \((1/2)^{t/2}(1/2)^{t/2} = 2^{-t}\). And there are \(\binom{t}{t/2}\) different sequences of \(t\) jumps containing \(t/2\) leftward jumps and \(t/2\) rightward jumps. Each ‘sequence’ represents a different ‘route’ from 0 back to 0 in exactly \(t\) steps. So altogether, the transition probability is
\[p_{0,0}^{(t)} = \text{number of possible routes} \times \text{probability of each route} = \binom{t}{t/2} \times 2^{-t},\]
as I claimed.
So we have
\[ \sum_{t=1}^{\infty} P_{0,0}^{(t)} = \sum_{t \text{ even}} \left( \frac{t}{t/2} \right) 2^{-t} = \sum_{s=1}^{\infty} \left( \frac{2s}{s} \right) 2^{-2s}. \]

We now use a well-known lower bound for the middle binomial coefficient \( \binom{2s}{s} \): for any \( s \in \mathbb{N} \), we have
\[ \binom{2s}{s} \geq \frac{2^{2s}}{2\sqrt{2s}}. \]
This implies that
\[ \sum_{t=1}^{\infty} P_{0,0}^{(t)} \geq \sum_{s=1}^{\infty} \frac{2^{2s}}{2\sqrt{2s}} 2^{-2s} = \frac{1}{2\sqrt{2}} \sum_{s=1}^{\infty} \frac{1}{\sqrt{s}}. \]

It is a well-known fact from calculus that
\[ \sum_{n=1}^{\infty} \frac{1}{n^u} = \infty \]
for all \( u \in \mathbb{R} \) such that \( u \leq 1 \). (You can prove this using the Integral Comparison Test, for example.) In particular, applying this when \( u = 1/2 \) gives
\[ \sum_{s=1}^{\infty} \frac{1}{\sqrt{s}} = \infty. \]
Hence,
\[ \sum_{t=1}^{\infty} P_{0,0}^{(t)} = \infty, \]
so by Theorem 1.22, the state 0 is recurrent.

Since this Markov chain is irreducible, by Corollary 1.25 it follows that all its states are recurrent.

**Positive recurrence**

We now study a property which is slightly stronger than being recurrent.

**Definition 1.31.** Let \((X_0, X_1, X_2, \ldots)\) be a Markov chain with state space \( S \). Suppose \( i \in S \) is recurrent. We say that \( i \) is **positive recurrent** if \( \mathbb{E}[T_i^{(1)}] < \infty \). We say that \( i \) is **null recurrent** if \( \mathbb{E}[T_i^{(1)}] = \infty \).

Recall that the random variable \( T_i^{(1)} \) is the time of the first return to the state \( i \), given that the Markov chain starts at the state \( i \), i.e.
\[ T_i^{(1)} = \min\{t \geq 1 : X_t = i\}. \]

For brevity, we will often write \( R_i = T_i^{(1)} \).
Remark. We have
\[ E[R_i] (= E[T_i^{(1)}]) = \sum_{t=1}^{\infty} t f_i^{(t)} . \]

Remark. It turns out that positive recurrence and null recurrence (like recurrence) are *class-properties*: if \( C \) is a communicating class, then *either* all the states in \( C \) are positive recurrent, *or* all the states in \( C \) are null recurrent, *or* all the states in \( C \) are transient.

For irreducible Markov chains, it turns out that positive recurrence is equivalent to the existence of a unique equilibrium distribution!

**Theorem 1.26.** Let \((X_0, X_1, X_2, \ldots)\) be an irreducible Markov chain with state space \( S \). Then the following conditions are equivalent.

- The Markov chain has an equilibrium distribution.
- All the states are positive recurrent.
- There exists a positive recurrent state.

Moreover, in this case, there is a unique equilibrium distribution \( w \), and it is given by

\[ w_i = \frac{1}{E[R_i]} \quad \forall i \in S . \]

**Example 30** (The symmetric random walk on \( \mathbb{Z} \)). We saw (in example 29) that every state of the symmetric random walk on \( \mathbb{Z} \) is recurrent, but we also saw (in example 24) that it has no equilibrium distribution. It follows from Theorem 1.26 that all its states are null recurrent.

The following theorem is a useful analogue of Theorem 1.22 for positive recurrence and null recurrence.

**Theorem 1.27.** Let \((X_0, X_1, X_2, \ldots)\) be a Markov chain with state space \( S \). Let \( i \in S \) be a recurrent state. Then

- \( i \) is null recurrent if and only if \( p_{i,i}^{(t)} \to 0 \) as \( t \to \infty \);
- \( i \) is positive recurrent if and only if \( p_{i,i}^{(t)} \not\to 0 \) as \( t \to \infty \).

Remark. This theorem can be used to show that positive recurrence and null recurrence are class properties.

Remark. If \((X_0, X_1, X_2, \ldots)\) is an irreducible a Markov chain with *finite* state space \( S \), then all its states are positive recurrent. (We know from Theorems 1.14 and 1.26 that either all the states positive recurrent, or all the states are transient. Exercise: show that we cannot have all the states being transient!)
2 Continuous-time stochastic processes

The second (slightly shorter) part of the course is on continuous-time stochastic processes. Recall the definition from the first lecture.

**Definition 2.1.** A *continuous time stochastic process* is a collection of random variables \( (X(t) : t \geq 0) \).

(There is one random variable \( X(t) \) for each non-negative real number \( t \). Recall that the random variables in a stochastic process are ‘indexed’ by an ‘index-set’ \( T \), thought of as ‘time’; with a continuous-time stochastic process, \( T = \mathbb{R}_{\geq 0} = \{ t \in \mathbb{R} : t \geq 0 \} \).)

If the events in a random process can take place at any real-valued time, not just at integer times, then we model the process as a continuous-time stochastic process. Many real-world processes are modelled much better as continuous-time stochastic processes, than as discrete-time stochastic processes — for example, a continuous-time stochastic process is an appropriate model in the following situations:

- \( X(t) = \) the number of alpha particles emitted by a radioactive source between time 0 and time \( t \), where times are measured in seconds from the moment a particle detector is switched on.
- \( X(t) = \) the number of buses passing a bus stop outside Queen Mary between time 0 and time \( t \), where times are measured in minutes after 08:00 this morning (assuming there is some randomness in the times at which buses pass, due to variations in the traffic).
- \( X(t) = \) the number of yeast cells in a culture at \( t \) seconds after the start of an experiment (assuming there is some randomness in the times at which the cells divide).

**The Poisson process**

One of the most important continuous-time stochastic processes is the *Poisson Process*. Before defining this, recall the following.

**Definition 2.2.** An integer-valued random variable \( R \) is said to have the *Poisson distribution with parameter \( \mu \)* if

\[
\mathbb{P}\{R = k\} = e^{-\mu} \frac{\mu^k}{k!} \quad \forall k \in \mathbb{N} \cup \{0\}.
\]

In this case, we write \( R \sim \text{Po}(\mu) \).

**Remark.** If \( R \sim \text{Po}(\mu) \), then \( \mathbb{E}[R] = \mu \) and \( \text{Var}[R] = \mu \).
**Definition 2.3.** If $R_1, R_2, \ldots, R_n$ is a collection of random variables all taking values in $\mathbb{N} \cup \{0\}$, they are said to be **mutually independent** of one another if for any $A_1, A_2, \ldots, A_n \subset \mathbb{N} \cup \{0\}$, we have

$$
\mathbb{P}(R_1 \in A_1, R_2 \in A_2, \ldots, R_n \in A_n) = \prod_{i=1}^{n} \mathbb{P}(R_i \in A_i).
$$

**Remark.** Informally, $R_1, R_2, \ldots, R_n$ are mutually independent of one another if they behave as if they are physically ‘completely unrelated’ to one another.

**Definition 2.4.** Let $\lambda > 0$. A continuous-time stochastic process $(X(t) : t \geq 0)$ is said to be a **Poisson process with parameter** $\lambda$ if the following properties hold.

1. $X(0) = 0$.
2. For any $s, t \geq 0$, we have $X(s + t) - X(s) \sim \text{Po}(\lambda t)$.
3. For any $0 < t_1 < t_2 < t_3 < \ldots < t_n$, the random variables $X(t_1) - X(0), X(t_2) - X(t_1), X(t_3) - X(t_2), \ldots, X(t_n) - X(t_{n-1})$ are mutually independent of one another. (Note that $X(t_1) - X(0) = X(t_1)$, by property 1.)

**Remark.** The above definition implies that in a Poisson process $(X(t) : t \geq 0)$, each random variable $X(t)$ takes values in $\mathbb{N} \cup \{0\}$, and moreover it implies that $X(t_2) \geq X(t_1)$ whenever $t_2 \geq t_1$. Hence, a Poisson process can be thought of as ‘counting’ the number of events of a certain kind, which occur between time 0 and time $t$. For example, it could be counting the number of buses that have passed a stop between time 0 and time $t$.

**Remark.** The ‘differences’ $X(t_1) - X(0), X(t_2) - X(t_1), X(t_3) - X(t_2), \ldots, X(t_n) - X(t_{n-1})$ are also called ‘increments’ of the process.

**Remark.** The property 1 is just a technical condition which makes calculations easier; if we liked we could allow $X(0) = s$ for any $s \in \mathbb{N} \cup \{0\}$. The property 2 looks somewhat mysterious; we will see the reason for it later. The property 3 says that in a Poisson process, the ‘increments’ of $X(t)$ are independent over any disjoint intervals of time $[0, t_1], (t_1, t_2], (t_2, t_3], \ldots$. This is a rather strong ‘memoryless’ property. It implies that for any $s > 0$, the increments of the process after time $s$ are independent of the behaviour of the process up to time $s$. In particular, $X(t + s) - X(s)$ is independent of $X(s)$, for any $s, t \geq 0$.  

50
Example 31. If we have a very large block of radioactive material which decays very slowly (emitting alpha, beta or gamma particles as it decays), and $X(t)$ = the number of particles it emits between time 0 and time $t$, then $X(t)$ can be modelled as a Poisson process. (The assumptions that the block is very large, and decays very slowly, are necessary to ensure that the increments of the process are independent over disjoint intervals of time: the decay between time 0 and time $t$ does not affect the decay after time $t$, etc.)

Other real-world situations that can often be modelled as a Poisson process include:

- $X(t)$ = the number of phone calls arriving at a call centre between time 0 and time $t$.
- $X(t)$ = the number of faults that have arisen in a pipeline between time 0 and time $t$ (provided faults are independent of one another).
- $X(t)$ = the number of cars passing a particular spot on a road between time 0 and time $t$ (provided e.g. the road is fairly quiet, i.e. no traffic jams).

Here is an example of how to do calculations for a Poisson process.

Example 32. A very large block of Carbon-14 emits beta particles according to a Poisson process of rate 2 per hour. Find the following.

(a) The probability that exactly three beta particles are emitted between time 0 and time 0.25 (note that time is measured in hours after the start of an experiment).

(b) The probability that exactly three beta particles are emitted between time 1 and time 1.25, given that 100 beta particles were emitted between time 0 and time 1.

(c) The probability that exactly three beta particles are emitted between time 0 and time 0.25 and exactly one beta particle is emitted between time 0.25 and time 1.

Solution: Let $X(t)$ = the number of beta particles emitted between time 0 and time $t$ (measuring time in hours after the start of the experiment). Then $(X(t) : t \geq 0)$ is a Poisson process of rate $\lambda = 2$.

(a) This probability is $\mathbb{P}\{X(0.25) - X(0.00) = 3\}$. By property (2) of the Poisson process, we have $X(0.25) - X(0.00) \sim \text{Po}(0.25\lambda) = \text{Po}(0.25 \times 2) = \text{Po}(0.5)$, so

$$\mathbb{P}\{X(0.25) - X(0.00) = 3\} = \mathbb{P}\{\text{Po}(0.5) = 3\} = e^{-0.5} \frac{0.5^3}{3!} = \frac{1}{48\sqrt{e}}.$$ 

(b) This probability is $\mathbb{P}\{X(1.25) - X(1.00) = 3 \mid X(1.00) - X(0.00) = 100\}$. By property (3) of the Poisson process, we know that the random variables $X(1.25) - X(1.00)$ and $X(1.00) - X(0.00)$ are independent of one another, and by property (2) of the
Poisson process we know that $X(1.25) - X(1.00) = \text{Po}(\lambda(1.25 - 1.00)) = \text{Po}(0.25\lambda) = \text{Po}(0.25 \times 2) = \text{Po}(0.5)$, just like in part (a), so we have

$$
\mathbb{P}\{X(1.25) - X(1.00) = 3 \mid X(1.00) - X(0.00) = 100\} = \mathbb{P}\{X(1.25) - X(1.00) = 3\} = \mathbb{P}\{\text{Po}(0.5) = 3\} = \frac{1}{48\sqrt{e}}
$$

the same answer as for part (a).

(c) This probability is $\mathbb{P}\{X(0.25) - X(0.00) = 3 \text{ and } X(1.00) - X(0.25) = 1\}$. By property (3) of the Poisson process, we know that the random variables $X(1.00) - X(0.25)$ and $X(0.25) - X(0.00)$ are independent of one another, so we have

$$
\mathbb{P}\{X(0.25) - X(0.00) = 3 \text{ and } X(1.00) - X(0.25) = 1\} = \mathbb{P}\{X(0.25) - X(0.00) = 3\} \times \mathbb{P}\{X(1.00) - X(0.25) = 1\}
$$

$$
= \frac{1}{48\sqrt{e}} \times \mathbb{P}\{\text{Po}(0.75\lambda) = 1\}
$$

$$
= \frac{1}{48\sqrt{e}} \times \mathbb{P}\{\text{Po}(1.5) = 1\}
$$

$$
= \frac{1}{48\sqrt{e}} \times \frac{e^{-1.5} \times 1.5}{1!}
$$

$$
= \frac{1}{48\sqrt{e}} \times \frac{1}{e\sqrt{e}} \times \frac{3}{2}
$$

$$
= \frac{1}{32e^2},
$$

using the result of part (a) for the second equality.

**The infinitesimal definition of the Poisson process (non-examinable)**

It turns out that we can give an equivalent definition of the Poisson process (i.e. a definition equivalent to Definition 2.4) in terms of how the process behaves over small time-intervals; this is called the *infinitesimal definition* of the Poisson process. To do this we need the following notation.

**Notation.** We use the notation $o(h)$ to denote *any* function $f = f(h) : \mathbb{R} \to \mathbb{R}$ such that $\frac{f(h)}{h} \to 0$ as $h \to 0$.

**Example 33.** $h^2 = o(h)$, $100h^2 = o(h)$, but $\frac{1}{1000} h$ is not $o(h)$, and $\sqrt{h}$ is not $o(h)$.

**Remark.** For us, $o(h)$ will usually represent a small ‘error’ term which is ‘negligible’ for our purposes.
Theorem 2.1. Let \((X(t) : t \geq 0)\) be a continuous-time stochastic process with \(X(0) = 0\), with \(X(t) \in \mathbb{N} \cup \{0\}\) for all \(t \geq 0\), and with \(X(s) \leq X(t)\) whenever \(s \leq t\). Let \(\lambda > 0\).

Suppose the process has the following three properties:

1. For any \(m \in \mathbb{N} \cup \{0\}\),
   \[
   \mathbb{P}(X(t+h) = n \mid X(t) = m) = \begin{cases} 
   \lambda h + o(h) & \text{if } n = m + 1; \\
   1 - \lambda h + o(h) & \text{if } n = m; \\
   o(h) & \text{if } n \geq m + 2; \\
   0 & \text{if } n < m.
   \end{cases}
   \]

2. The probability distribution of the random variable \(X(s+t) - X(s)\) does not depend on \(s\) (This property is called the time-homogeneity condition.)

3. For any \(0 < t_1 < t_2 < t_3 < \ldots < t_n\), the random variables
   \[X(t_1) - X(0), X(t_2) - X(t_1), X(t_3) - X(t_2), \ldots, X(t_n) - X(t_{n-1})\]
   are mutually independent of one another.

Then the process \((X(t) : t \geq 0)\) is a Poisson process of rate \(\lambda\). Conversely, a Poisson process of rate \(\lambda\) has properties 1-3 above.

Remark. Properties 1-3 above are called the infinitesimal definition of the Poisson process.

Remark. When proving theorems about the Poisson process, sometimes it is better to use the basic definition (Definition 2.4), and sometimes it is better to use the infinitesimal definition.

Remark. The condition 1. in the ‘infinitesimal definition’ above says that if, for example, \(X(t)\) denotes the number of particles that arrive at a particle detector between time 0 and time \(t\), then the probability of exactly one new particle arriving in the time-interval \([t, t+h]\) is \(\lambda h + o(h)\) (so ‘approximately’ \(\lambda h\) if the time-interval is short) and the probability of more than one new particle arriving in that time-interval is \(o(h)\) (so ‘very small’, in fact ‘negligible’, if the time-interval is short). In other words, the probability of exactly one new particle arriving in a short time-interval is approximately proportional to the length of that time-interval, and the probability of more than one arriving in that time-interval is negligible. This is true of many processes that occur in nature — for example, the slow radioactive decay of a large piece of radioactive material, or the arrival of photons from a distance source of light. As we will see in the next lemma, this property (plus the time-homogeneity condition) implies that \(X(s+t) - X(s)\) has the Poisson distribution with parameter \(\lambda t\), i.e property 2 in Definition 2.4. This solves the ‘mystery’ behind property 2. of Definition 2.4.

1These three conditions mean that \(X(t)\) ‘counts’ the number of events of a certain type, which occur in the time-interval \([0, t]\).
Lemma 2.2. Let \((X(t) : t \geq 0)\) satisfy the conditions of Theorem 2.1. Then it satisfies property 2 of Definition 2.4.

**Proof.** We must show that for any \(s, t \geq 0\), \(X(s + t) - X(s) \sim \text{Po}(\lambda t)\). By the time-homogeneity condition, it suffices to prove this for \(s = 0\), i.e., we must show that \(X(t) \sim \text{Po}(\lambda t)\) for all \(t \geq 0\). For each \(k \in \mathbb{N} \cup \{0\}\), define a function \(p_k : \mathbb{R}_{\geq 0} \to [0, 1]\) by

\[
p_k(t) = \Pr(X(t) = k) \quad \forall t \geq 0.
\]

We will derive a system of differential equations for the functions \(p_k\) and then solve them to show that

\[
\Pr(X(t) = k) = p_k(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \quad \forall k \in \mathbb{N} \cup \{0\}, \ t \geq 0,
\]

which says precisely that \(X(t) \sim \text{Po}(\lambda t)\) for all \(t \geq 0\).

We have

\[
p_0(t + h) = \Pr(X(t + h) = 0)
= \Pr(X(t + h) = 0 | X(t) = 0) \cdot \Pr(X(t) = 0)
= (1 - \lambda h + o(h)) \Pr(X(t) = 0)
= (1 - \lambda h + o(h)) p_0(t)
= p_0(t) - \lambda hp_0(t) + o(h).
\]

Hence, rearranging, we have

\[
\frac{p_0(t + h) - p_0(t)}{h} = -\lambda p_0(t) + \frac{o(h)}{h}.
\]

Taking the limit of both sides as \(h \to 0\) gives

\[
p'_0(t) = -\lambda p_0(t).
\]

(Here, of course, \(p'_0(t)\) denotes \(\frac{d}{dt}(p_0(t))\).)

Now let \(k \geq 1\). We have

\[
p_k(t + h) = \Pr(X(t + h) = k)
= \sum_{m=0}^{k} \Pr(X(t + h) = k | X(t) = m) \cdot \Pr(X(t) = m)
= \sum_{m=0}^{k-2} (o(h)) \cdot \Pr(X(t) = m)
+ \Pr(X(t + h) = k | X(t) = k - 1) \cdot \Pr(X(t) = k - 1)
+ \Pr(X(t + h) = k | X(t) = k) \cdot \Pr(X(t) = k)
= o(h) + (\lambda h + o(h))p_{k-1}(t) + (1 - \lambda h + o(h)) p_k(t)
= \lambda hp_{k-1}(t) - \lambda hp_k(t) + o(h) + o(h)p_{k-1}(t) + o(h)p_k(t)
= \lambda hp_{k-1}(t) - \lambda hp_k(t) + o(h)
\]
Rearranging, we have
\[
\frac{p_k(t + h) - p_k(t)}{h} = \lambda p_{k-1}(t) - \lambda p_k(t) + \frac{o(h)}{h}.
\]
Taking the limit of both sides as \(h \to 0\) gives
\[p'_k(t) = \lambda p_{k-1}(t) - \lambda p_k(t).
\]
Hence, we must solve the following system of differential equations:
\[
p'_0(t) = -\lambda p_0(t), \quad p'_k(t) = \lambda p_{k-1}(t) - \lambda p_k(t) \quad \forall k \in \mathbb{N}.
\]
We do this inductively. First we solve the first equation (for \(p_0\)):
\[p'_0(t) = -\lambda p_0(t).
\]
This has general solution \(p_0(t) = Ae^{-\lambda t}\), where \(A\) is a constant of integration. Since \(X(0) = 0\), we have \(p_0(0) = \mathbb{P}(X(0) = 0) = 1\), so \(A = 1\), and therefore \(p_0(t) = e^{-\lambda t}\), which is the formula we want.

We can now solve the second differential equation, for \(p_1\). We have
\[p'_1(t) = \lambda p_0(t) - \lambda p_1(t) = \lambda e^{-\lambda t} - \lambda p_1(t).
\]
Rearranging, we have
\[p'_1(t) + \lambda p_1(t) = \lambda e^{-\lambda t}.
\]
Now we use a trick: we can rewrite the left-hand side of the above equation as
\[e^{-\lambda t} \frac{d}{dt}(p_1(t)e^{\lambda t}).
\]
So the equation can be rewritten as
\[
\frac{d}{dt}(p_1(t)e^{\lambda t}) = \lambda.
\]
Integrating both sides gives
\[p_1(t)e^{\lambda t} = \lambda t + C,
\]
where \(C\) is a constant of integration. Since \(X(0) = 0\), we have \(p_1(0) = \mathbb{P}(X(0) = 1) = 0\), and therefore \(C = 0\), so
\[p_1(t) = \lambda te^{-\lambda t}.
\]
which is the formula we want.

Now we proceed inductively; let \(k \geq 1\) and assume that we have proved that
\[p_{k-1}(t) = e^{-\lambda t} \frac{(\lambda t)^{k-1}}{(k-1)!}.
\]
Then the differential equation for $p_k(t)$ gives

$$p_k'(t) = \lambda p_{k-1}(t) - \lambda p_k(t) = \lambda \left( e^{-\lambda t} \frac{(\lambda t)^{k-1}}{(k-1)!} \right) - \lambda p_k(t).$$

Rearranging, we get

$$p_k'(t) + \lambda p_k(t) = e^{-\lambda t} \frac{\lambda^k t^{k-1}}{(k-1)!}.$$

We can rewrite the left-hand side of the above equation as

$$e^{-\lambda t} \frac{d}{dt}(p_k(t)e^{\lambda t}),$$

so the equation can be rewritten as

$$\frac{d}{dt}(p_k(t)e^{\lambda t}) = \frac{\lambda^k t^{k-1}}{(k-1)!}.$$

Integrating both sides gives

$$p_k(t)e^{\lambda t} = \frac{\lambda^k t^k}{k!} + C',$$

where $C'$ is a constant of integration. Since $X(0) = 0$, we have $p_k(0) = P(X(0) = k) = 0$, and therefore $C' = 0$. Hence,

$$p_k(t) = e^{-\lambda t} \frac{\lambda^k t^k}{k!}.$$

This proves (by induction on $k$) that

$$p_k(t) = e^{-\lambda t} \frac{(\lambda t)^k}{k!} \quad \forall k \in \mathbb{N} \cup \{0\},$$

and therefore $X(t) \sim \text{Po}(\lambda t)$, as required.

[End of non-examinable section.]

The Superposition Lemma and the Thinning Lemma

The following ‘Superposition Lemma’ says that when we add two independent Poisson processes together, we get another Poisson process whose rate is the sum of the rates of the two Poisson processes we added.

**Lemma 2.3 (The Superposition Lemma).** Let $\lambda, \mu > 0$. Let $(X(t) : t \geq 0)$ be a Poisson process of rate $\lambda$, and let $(Y(t) : t \geq 0)$ be a Poisson process of rate $\mu$. Suppose that $(X(t) : t \geq 0)$ and $(Y(t) : t \geq 0)$ are independent of one another. Then the stochastic process $(X(t) + Y(t) : t \geq 0)$ is a Poisson process of rate $\lambda + \mu$.

**Proof.** See Exercise Sheet!
Example 34. Suppose alpha particles arrive at a Geiger counter according to a Poisson process of rate 4 per hour, and beta particles arrive at the Geiger counter according to a Poisson process of rate 6 per hour; suppose further that these two processes are independent. What is the probability that exactly 3 particles in total (alpha particles plus beta particles) arrive at the Geiger counter in a 15-minute experiment?

Solution: let $X(t)$ denote the number of alpha particles arriving in the first $t$ hours of the experiment, and let $Y(t)$ denote the number of beta particles arriving in the first $t$ hours of the experiment. Then $(X(t) : t \geq 0)$ is a Poisson process of rate 4, and $(Y(t) : t \geq 0)$ is a Poisson process of rate 6. Hence, by the Superposition Lemma, $(X(t) + Y(t) : t \geq 0)$ is a Poisson process of rate $4 + 6 = 10$. The probability we want is therefore

$$P(X(0.25) + Y(0.25) = 3) = P(Po(10 \times 0.25) = 3)$$

$$= P(Po(2.5) = 3)$$

$$= e^{-2.5} \frac{2.5^3}{3!}$$

$$= e^{-5/2} \frac{5^3}{2^3 \cdot 6}$$

$$= \frac{125}{48e^{2\sqrt{e}}}.$$

The following ‘Thinning Lemma’ is also very important and useful. We state it in the context of counting particles, but it is applicable to any Poisson process, when the events being counted each independently ‘deleted’ with a certain fixed probability $p$.

Lemma 2.4 (The Thinning Lemma). Let $\lambda > 0$ and let $0 < p < 1$. Let $(X(t) : t \geq 0)$ be a Poisson process of rate $\lambda$. Suppose (for concreteness) that $X(t)$ denotes the number of particles arriving at a detector between time 0 and time $t$. Suppose that for each particle, the detector fails to detect it with probability $p$, independently for each particle. Let $Y(t)$ denote the number of particles detected by the particle detector between time 0 and time $t$. Then $(Y(t) : t \geq 0)$ is a Poisson process of rate $(1 - p)\lambda$.

Proof. See Exercise Sheet!

Example 35. Suppose alpha particles arrive at a Geiger counter according to a Poisson process of rate 4 per hour, but the Geiger counter is faulty and for each particle, fails to detect that particle with probability $1/4$, independently for each particle. Find the probability that the Geiger counter detects exactly 2 alpha particles during a 30-minute experiment.

Solution: let $X(t)$ denote the number of alpha particles arriving at the Geiger counter in the first $t$ hours of the experiment, and let $Y(t)$ denote the number of alpha particles that are detected by the Geiger counter in the first $t$ hours of the experiment. Then $(X(t) : t \geq 0)$ is a Poisson process of rate 4, so by the Thinning Lemma, $(Y(t) : t \geq 0)$
is a Poisson process of rate $4 \times \frac{3}{4} = 3$. Hence, the probability we want is

$$P(Y(0.5) = 2) = P(\text{Po}(3 \times 0.5) = 2)$$

$$= P(\text{Po}(1.5) = 2)$$

$$= e^{-1.5} \frac{1.5^2}{2!}$$

$$= \frac{2^2 \cdot 2 \cdot e}{8e}$$

$$= \frac{9}{8e}.$$

**Conditioning on $X(s) = n$**

Suppose $(X(t) : t \geq 0)$ is a Poisson process of rate $\lambda$, and we are told that $X(s) = n$ for some time $s > 0$ and some $n \in \mathbb{N}$. If $t < s$, what is the conditional distribution of $X(t)$, conditioned on $X(s) = n$? The following useful theorem tells us that this distribution is binomial with parameters $n$ and $\frac{t}{s}$ — independent of $\lambda$!

**Theorem 2.5.** Let $\lambda > 0$. Let $(X(t) : t \geq 0)$ be a Poisson process of rate $\lambda$. Let $s > 0$, and let $n \in \mathbb{N}$. Then for any $m \in \{0, 1, 2, \ldots, n\}$ and any $t < s$, we have

$$P\{X(t) = m \mid X(s) = n\} = \binom{n}{m} \left(\frac{t}{s}\right)^m (1 - \frac{t}{s})^{n-m}.$$

Hence, if $t < s$, then the conditional distribution of $X(t)$, conditioned on $X(s) = n$, is $\text{Bin}(n, \frac{t}{s})$, i.e. binomial with parameters $n$ and $\frac{t}{s}$. 

58
\begin{proof}
We have

\[
\mathbb{P}\{X(t) = m \mid X(s) = n\} = \frac{\mathbb{P}\{X(t) = m \text{ and } X(s) = n\}}{\mathbb{P}\{X(s) = n\}}
\]
(by the definition of conditional probability)

\[
= \frac{\mathbb{P}\{X(t) = m \} \cdot \mathbb{P}\{X(s) - X(t) = n - m\}}{\mathbb{P}\{X(s) = n\}}
\]
\[
= \mathbb{P}\{X(t) = m\} \cdot \mathbb{P}\{X(s) - X(t) = n - m\}
\]
\[
\mathbb{P}\{X(s) = n\}
\]
\[
(X(t) - X(0) \text{ and } X(s) - X(t) \text{ are independent of one another})
\]
\[
= \mathbb{P}\{\text{Po}(\lambda t) = m\} \cdot \mathbb{P}\{\text{Po}(\lambda(s - t)) = n - m\}
\]
\[
\mathbb{P}\{\text{Po}(\lambda s) = n\}
\]
(by property 2. of the basic definition of the Poisson process)

\[
e^{-\lambda t} \frac{\lambda^m}{m!} \cdot e^{-\lambda(s-t)} \frac{(\lambda(s-t))^{n-m}}{(n-m)!}
\]
\[
e^{-\lambda s} \frac{\lambda^n}{n!} \frac{n!}{m!(n-m)!} \frac{t^m(s-t)^{n-m}}{s^n}
\]
\[
= \binom{n}{m} \left(\frac{1}{s}\right)^m \left(\frac{s-t}{s}\right)^{n-m}
\]
\[
= \binom{n}{m} \left(\frac{1}{s}\right)^m (1 - \frac{t}{s})^{n-m}
\]
\[
= \mathbb{P}\{\text{Bin}(n, \frac{t}{s}) = m\}
\]
as required.
\end{proof}

**Example 36.** Suppose alpha particles arrive at a Geiger counter according to a Poisson process of rate 3 per hour. Suppose I perform a 2-hour experiment in which 8 alpha particles arrive. Conditioned on this, what is the conditional probability that exactly 3 alpha particles arrived in the first 20 minutes of the experiment?

Solution: let \(X(t)\) denote the number of particles that arrive in the first \(t\) hours of the experiment; then \((X(t) : t \geq 0)\) is a Poisson process of rate 3. Conditioning on the event \(X(2) = 8\), the conditional distribution of \(X\left(\frac{1}{3}\right)\) is \(\text{Bin}(8, \frac{1}{6})\), by Theorem 2.5. Hence,

\[
\mathbb{P}\{X\left(\frac{1}{3}\right) = 3\} = \mathbb{P}\{\text{Bin}(8, \frac{1}{6}) = 3\} = \binom{8}{3} \left(\frac{1}{6}\right)^3 \left(\frac{5}{6}\right)^5 = \frac{21875}{209952}.
\]
Random variables associated with the Poisson process: arrival times and interarrival times

Definition 2.5. Let $\lambda > 0$, and let $(X(t) : t \geq 0)$ be a Poisson process of rate $\lambda$. For each $n \in \mathbb{N}$, the $n$th arrival time of the Poisson process is the random variable $T_n$ defined by

$$T_n = \min\{t > 0 : X(t) \geq n\}.$$ 

In other words, if $X(t)$ counts the number of particles which have arrived by time $t$, then $T_n$ is the time at which the $n$th particle arrives.

For each $n \in \mathbb{N}$, the $n$th interarrival time of the Poisson process is the random variable $S_n$ defined by

$$S_n = \begin{cases} T_1 & \text{if } n = 1; \\ T_n - T_{n-1} & \text{if } n \geq 2. \end{cases}$$

In other words, if $X(t)$ counts the number of particles which have arrived by time $t$, then $S_n$ is the length of time between the arrival of the $(n-1)$th particle and the arrival of the $n$th particle.

Our aim is to investigate the probability distributions of these random variables. First recall the following definition from Probability Models.

Definition 2.6. A continuous random variable $R$ is said to have the exponential distribution with parameter $\mu$ if $R$ has probability density function (pdf) $f_R$ given by

$$f_R(t) = \begin{cases} \mu e^{-\mu t} & \text{if } t \geq 0; \\ 0 & \text{if } t < 0 \end{cases}$$

or equivalently, if it has cumulative distribution function $F_R(t) = \mathbb{P}\{R \leq t\}$ given by

$$F_R(t) = \begin{cases} 1 - e^{-\mu t} & \text{if } t \geq 0; \\ 0 & \text{if } t < 0. \end{cases}$$

In this case, we write $R \sim \text{Exp}($\mu$)$ for short.

It turns out that the first arrival time $T_1$ of a Poisson process of rate $\lambda$, has the exponential distribution with parameter $\lambda$.

Lemma 2.6. Let $\lambda > 0$, and let $(X(t) : t \geq 0)$ be a Poisson process of rate $\lambda$. Then the first arrival time $T_1$ of the process has the exponential distribution with parameter $\lambda$.

Proof. We calculate the cdf of $T_1$, denoted $F_{T_1}$. If $t > 0$, then

$$F_{T_1}(t) = \mathbb{P}\{T_1 \leq t\} = 1 - \mathbb{P}\{T_1 > t\} = 1 - \mathbb{P}\{X(t) = 0\} = 1 - e^{-\lambda t}.$$
Clearly, if \( t < 0 \), then \( F_{T_1}(t) = \Pr\{T_1 \leq t\} = 0 \). Hence,

\[
F_{T_1}(t) = \begin{cases} 
1 - e^{-\lambda t} & \text{if } t \geq 0; \\
0 & \text{if } t < 0,
\end{cases}
\]

so \( T_1 \sim \text{Exp}(\lambda) \), as required. \( \square \)

Recall from Probability Models that if \( R \) is a continuous random variable with \( R \sim \text{Exp}(\mu) \), then \( \mathbb{E}[R] = 1/\mu \). It follows that if \( (X(t) : t \geq 0) \) is a Poisson process of rate \( \lambda \), then \( \mathbb{E}[T_1] = 1/\lambda \). This makes sense: the higher the rate of a Poisson process, the shorter the time we expect to wait before seeing the first ‘event’ (e.g. the arrival of the first particle, if the Poisson process counts the number of particles that arrive between time 0 and time \( t \)).

Recall that the exponential distribution has the following ‘memoryless’ property.

**Lemma 2.7.** Let \( \mu > 0 \), and suppose \( R \) is a continuous random variable with \( R \sim \text{Exp}(\mu) \). Then for any \( s, t > 0 \),

\[
\Pr(R > s + t \mid R > s) = \Pr(R > t).
\]

**Proof.** If \( R \sim \text{Exp}(\mu) \), then for all \( t > 0 \), we have

\[
\Pr(R > t) = 1 - \Pr(R \leq t) = 1 - F_R(t) = e^{-\mu t}.
\]

Hence,

\[
\Pr(R > s + t \mid R > s) = \frac{\Pr(R > s + t \text{ and } R > s)}{\Pr(R > s)}
= \frac{\Pr(R > s + t)}{\Pr(R > s)}
= \frac{e^{-\mu(s+t)}}{e^{-\mu s}}
= e^{-\mu t}
= \Pr(R > t),
\]

as required. \( \square \)

It follows that if \( (X(t) : t \geq 0) \) is a Poisson process of rate \( \lambda \), then the first arrival time \( T_1 \) has this ‘memoryless’ property: for any \( s, t > 0 \), we have

\[
\Pr(T_1 > s + t \mid T_1 > s) = \Pr(T_1 > t).
\]

In other words, if we have already waited a time \( s \) and no particle has arrived, the ‘further’ time we have to wait is just the same as if were starting from time 0. This makes sense, given property 3. of the basic definition of the Poisson process: what happens over the time interval \([s, t+s]\) is independent of what happened over the time interval \([0, s]\).

We now move on to studying the distribution of \( T_n \) for \( n \geq 2 \).
Theorem 2.8. Let \((X(t) : t \geq 0)\) be a Poisson process of rate \(\lambda\). Then for each \(n \in \mathbb{N}\), the \(n\)th arrival time \(T_n\) has the Gamma distribution with parameters \(n\) and \(\lambda\), i.e. it has pdf
\[
f_{T_n}(t) = \begin{cases} 
  e^{-\lambda t} \frac{\lambda^n t^{n-1}}{(n-1)!} & \text{if } t \geq 0; \\
  0 & \text{if } t < 0.
\end{cases}
\]

Proof. (Non-examinable.) Let \(n \in \mathbb{N}\). Consider the cdf \(F_{T_n}(t)\) of \(T_n\). Let \(t > 0\). We have
\[
F_{T_n}(t+h) - F_{T_n}(t) = \mathbb{P}(T_n \leq t+h) - \mathbb{P}(T_n \leq t) \\
= \mathbb{P}(T_n \in (t,t+h)) \\
= \mathbb{P}(X(t) \leq n-1, X(t+h) \geq n) \\
= \sum_{k=0}^{n-1} \mathbb{P}(X(t) = k) \mathbb{P}(X(t+h) \geq n \mid X(t) = k) \\
\text{(by the law of total probability, and conditioning on } X(t) \text{)} \\
= \mathbb{P}(X(t) = n-1) \mathbb{P}(X(t+h) \geq n \mid X(t) = n-1) + o(h) \\
\text{(since } \mathbb{P}(X(t+h) - X(t) \geq 2) = o(h), \text{ by the infinitesimal definition of the Poisson process)} \\
= e^{-\lambda t} \frac{(\lambda t)^{n-1}}{(n-1)!} (\lambda h + o(h) + o(h)) + o(h) \\
\text{(since } X(t) \sim \text{Po}(\lambda t), \text{ and using the infinitesimal definition again)} \\
= he^{-\lambda t} \frac{\lambda^n t^{n-1}}{(n-1)!} + o(h).
\]
Dividing both sides by \(h\) gives
\[
\frac{F_{T_n}(t+h) - F_{T_n}(t)}{h} = e^{-\lambda t} \frac{\lambda^n t^{n-1}}{(n-1)!} + \frac{o(h)}{h}.
\]
Taking the limit as \(h \to 0\) gives
\[
f_{T_n}(t) = \frac{d}{dt} F_{T_n}(t) \\
= \lim_{h \to 0} \left[ \frac{F_{T_n}(t+h) - F_{T_n}(t)}{h} \right] \\
= \lim_{h \to 0} \left[ e^{-\lambda t} \frac{\lambda^n t^{n-1}}{(n-1)!} + \frac{o(h)}{h} \right] \\
= e^{-\lambda t} \frac{\lambda^n t^{n-1}}{(n-1)!}.
\]
Clearly, we have \(f_{T_n}(t) = 0\) for all \(t < 0\). Hence, \(T_n\) has the pdf claimed. \(\Box\)
The interarrival times $S_1, S_2, \ldots$ have a very nice and useful property:

**Theorem 2.9.** Let $(X(t) : t \geq 0)$ be a Poisson process of rate $\lambda$. Then the interarrival times $S_1, S_2, S_3, \ldots$ all have the exponential distribution with parameter $\lambda$, and are all mutually independent of one another.

**Proof.** Non-examinable; for a sketch, see lectures.

In fact, the condition in the above theorem gives yet another equivalent definition of the Poisson process!

**Theorem 2.10.** Let $(X(t) : t \geq 0)$ be a continuous-time stochastic process with $X(0) = 0$, with $X(t) \in \mathbb{N} \cup \{0\}$ for all $t \geq 0$, and with $X(s) \leq X(t)$ whenever $s \leq t$. (These three conditions mean that $X(t)$ ‘counts’ the number of events of a certain type, which occur in the time-interval $[0, t]$.) Suppose for concreteness that $X(t)$ counts the number of particles that arrive at a detector in the time-interval $[0, t]$.

Let $\lambda > 0$. Let $S_1, S_2, \ldots$ be the interarrival times of the process, defined as in Definition 2.5. Suppose $S_1, S_2, \ldots$ all have the exponential distribution with parameter $\lambda$, and are all mutually independent of one another. Then $(X(t) : t \geq 0)$ is a Poisson process of rate $\lambda$.

**Proof.** Non-examinable.

**Example 37.** Suppose beta particles arrive at a detector according to a Poisson process of rate 10 per hour. Suppose we perform an experiment, and the first particle arrives at the detector 50 minutes after the start of the experiment. Conditional on this fact, find the expected time at which the second particle arrives at the detector.

Solution: let us measure time in minutes. Let $X(t)$ be the number of particles that arrive between time 0 and time $t$. Then $(X(t) : t \geq 0)$ is a Poisson process of rate $\lambda = 1/6$. We are told that $T_1 = 50$. But $S_2$ is independent of $T_1 (= S_1)$, and moreover $S_2$ has the exponential distribution with parameter $\lambda = 1/6$, so

\[\mathbb{E}[S_2 | T_1 = 50] = \mathbb{E}[S_2] = 1/\lambda = 1/(1/6) = 6.\]

The time at which the second particle arrives is precisely $T_2 = T_1 + S_2$; we are being asked to find $\mathbb{E}[T_2 | T_1 = 50]$. We have

\[\mathbb{E}[T_2 | T_1 = 50] = \mathbb{E}[T_1 + S_2 | T_1 = 50] = \mathbb{E}[T_1 | T_1 = 50] + \mathbb{E}[S_2 | T_1 = 50] = 50 + 6 = 56.\]

**More on conditioning**

Suppose $(X(t) : t \geq 0)$ is a Poisson process of rate $\lambda$, and we are told that $X(s) = 1$ for some time $s > 0$. Then, conditional on this fact, the first arrival time of the process is uniformly distributed on the time-interval $[0, s]$. 

63
Lemma 2.11. Let $\lambda > 0$, and let $(X(t) : t \geq 0)$ be a Poisson process of rate $\lambda$. Let $s > 0$. Then conditional on $X(s) = 1$, the first arrival time $T_1$ is uniformly distributed on $[0, s]$, i.e the conditional pdf $f_{T_1|X(s)=1}(t)$ satisfies

$$f_{T_1|X(s)=1}(t) = \begin{cases} \frac{1}{s} & \text{if } 0 \leq t \leq s; \\ 0 & \text{otherwise}. \end{cases}$$

Proof. We first calculate the conditional cdf $F_{T_1|X(s)=1}(t)$. Let $t \in [0, s]$. Then

$$F_{T_1|X(s)=1}(t) = \mathbb{P}(T_1 \leq t \mid X(s) = 1)$$

$$= \mathbb{P}(X(t) = 1 \mid X(s) = 1)$$

$$= \mathbb{P}(\text{Bin}(1, \frac{t}{s}) = 1) \quad \text{(by Theorem 2.5)}$$

$$= \frac{t}{s}.$$  

Clearly, if $t > s$, then

$$F_{T_1|X(s)=1}(t) = \mathbb{P}(T_1 \leq t \mid X(s) = 1) = 1,$$

and if $t < 0$, then

$$F_{T_1|X(s)=1}(t) = \mathbb{P}(T_1 \leq t \mid X(s) = 1) = 0.$$  

Therefore,

$$F_{T_1|X(s)=1}(t) = \begin{cases} 0 & \text{if } t < 0; \\ \frac{t}{s} & \text{if } 0 \leq t \leq s; \\ 1 & \text{if } t > s. \end{cases}$$

So, (by differentiating the function $\frac{t}{s}$, or recalling that the above is the cdf of the uniform distribution on $[0, s]$), the conditional pdf is

$$f_{T_1|X(s)=1}(t) = \begin{cases} \frac{1}{s} & \text{if } 0 \leq t \leq s; \\ 0 & \text{otherwise}, \end{cases}$$

proving the lemma.

Example 38. Suppose beta particles arrive at a detector according to a Poisson process of rate 20 per hour. Suppose that, in a 50-minute experiment, just one particle arrives. Conditional on this fact, find the probability that the particle actually arrived in the first 20 minutes of the experiment.

Solution: let us measure time in minutes, from the start of the experiment. Let $X(t)$ be the number of particles that arrive between time 0 and time $t$. Then $(X(t) : t \geq 0)$ is a Poisson process of rate 1/3. We are told that $X(50) = 1$, and we are asked to find $\mathbb{P}(T_1 \leq 20 \mid X(50) = 1)$. Conditional on $X(50) = 1$, the first arrival time $T_1$ is uniformly distributed on the time-interval $[0, 50]$, so

$$\mathbb{P}(T_1 \leq 20 \mid X(50) = 1) = 20/50 = 2/5.$$  

Notice that the rate (1/3 per minute) didn’t enter into this calculation at all!
The following theorem tells us, more generally, what happens to the joint distribution of the arrival times $T_1, T_2, \ldots, T_n$ if we condition on $X(s) = n$ for some time $s > 0$.

**Theorem 2.12.** Let $\lambda > 0$, and let $(X(t) : t \geq 0)$ be a Poisson process of rate $\lambda$. Let $s > 0$ and let $n \in \mathbb{N}$. Suppose we condition on $X(s) = n$. Then the (conditional) joint distribution of the arrival times $(T_1, T_2, \ldots, T_n)$ can be found as follows. Let $U_1, U_2, \ldots, U_n$ be mutually independent random variables which are each uniformly distributed on $[0, s]$. Choose a permutation $\pi \in S_n$ such that

$$U_{\pi(1)} \leq U_{\pi(2)} \leq \ldots \leq U_{\pi(n)},$$

i.e. the permutation $\pi$ puts the random variables $U_1, U_2, \ldots, U_n$ in non-decreasing order of the values they take. Then the (conditional) joint distribution of the arrival times $(T_1, T_2, \ldots, T_n)$ is identical to the joint distribution of $(U_{\pi(1)}, U_{\pi(2)}, \ldots, U_{\pi(n)})$.

This theorem says that, if, for example, particles arrive at a detector according to a Poisson process of rate $\lambda$ per minute, and we are told that after 60 minutes (say), exactly 100 particles have arrived, then we can ‘generate’ the first 100 arrival times $T_1, T_2, \ldots, T_{100}$ (with the correct conditional distribution) by taking 100 independent random variables $U_1, U_2, \ldots, U_{100}$ each uniformly distributed on the time-interval $[0, 60]$, and then reordering them so that they are in increasing order. This is useful for calculation, as we will see below.

**Example 39.** Suppose telephone calls arrive at a call centre according to a Poisson process of rate 100 per hour, and each caller stays on the phone for exactly 10 minutes. The call centre opens at 9 am one morning. By 10 am, just five people have called, and there were enough customer service advisors to answer each call as soon as it came in. However, there is a power-cut at 10 am and the call centre does not reopen for the rest of the morning. Conditional on this information, find the expected total duration of all the calls that morning.

**Solution:** let us measure time in minutes from 9am. Let $X(t)$ be the number of calls that arrive between time 0 and time $t$. Then $(X(t) : t \geq 0)$ is a Poisson process of rate $1/2$. We are told that $X(60) = 5$. By Theorem 2.12, conditional on $X(60) = 5$, we can generate the arrival-times of the five calls by generating five independent random variables each uniformly distributed on the time-interval $[0, 60]$. Let these random variables be $U_1, U_2, U_3, U_4, U_5$; then the arrival-times of the five calls are $U_1, U_2, U_3, U_4, U_5$. For each $i \in \{1, 2, 3, 4, 5\}$, let $D_i$ be the duration of the call that comes in at time $U_i$. We are being asked to find

$$E[D_1 + D_2 + D_3 + D_4 + D_5].$$

Since each $D_i$ has the same distribution, by the linearity of expectation we have

$$E[D_1 + D_2 + D_3 + D_4 + D_5] = 5E[D_1].$$

So we just have to find $E[D_1]$. If $U_1 \leq 50$, then $D_1 = 10$, but if $U_1 > 50$, then $D_1 = 60 - U_1$. Recall that the random variable $U_1$ is uniformly distributed on the
time-interval \([0, 60]\). Hence,
\[
E[D_1] = E[D_1 \mid U_1 \leq 50] \times P(U_1 \leq 50) + E[D_1 \mid U_1 > 50] \times P(U_1 > 50)
\]
\[
= 10 \times \frac{50}{60} + E[D_1 \mid U_1 > 50] \times \frac{10}{60}
\]
\[
= \frac{25}{3} + \frac{1}{6} E[D_1 \mid U_1 > 50]
\] (1)

Now, conditional on \(U_1 > 50\), the conditional distribution of \(U_1\) is uniform on the time-interval \((50, 60]\). (Check this!) So \(E[U_1 \mid U_1 > 50] = 55\). Since \(D_1 = 60 - U_1\) whenever \(U_1 > 50\), we have
\[
E[D_1 \mid U_1 > 50] = E[60 - U_1 \mid U_1 > 50] = 60 - E[U_1 \mid U_1 > 50] = 60 - 55 = 5.
\]

Substituting this into (1) gives
\[
E[D_1] = \frac{25}{3} + \frac{1}{6} \times 5 = \frac{55}{6}.
\]

Hence,
\[
E[D_1 + D_2 + D_3 + D_4 + D_5] = 5E[D_1] = 5 \times \frac{55}{6} = \frac{275}{6}.
\]

So the expected total duration of all the calls that morning is 45 minutes and 50 seconds.

**Birth processes (non-examinable).**

A *birth process* is an important generalisation of the Poisson process. If the ‘arrivals’ we are counting (e.g. the arrival of particles at a detector), are arriving according to a Poisson process, then the ‘rate’ of arrival does not depend on the number of particles that have already arrived. However, if the arrivals we are counting are arriving according to a birth process, then the ‘rate’ of arrival is allowed to depend on the number that have already arrived. This is often the case in nature — for example, the birth-rate in a population almost always depends on the number of individuals already present in the population.

To define a birth process formally, we generalise the infinitesimal definition of the Poisson process.

**Definition 2.7.** Let \((X(t) : t \geq 0)\) be a continuous-time stochastic process. Let \(\lambda_0, \lambda_1, \lambda_2, \ldots > 0\). Suppose the process has the following three properties:

1. \(X(0) \in \mathbb{N} \cup \{0\}\);

2. For any \(m \in \mathbb{N} \cup \{0\},
   \[
   P(X(t+h) = n \mid X(t) = m) = \begin{cases} 
   \lambda_m h + o(h) & \text{if } n = m + 1; \\
   1 - \lambda_m h + o(h) & \text{if } n = m; \\
   o(h) & \text{if } n \geq m + 2; \\
   0 & \text{if } n < m.
   \end{cases}
   \]

66
3. If $0 < s < t$, then conditional on $X(s)$, the random variable $X(t) - X(s)$ is independent of the process before time $s$, i.e. it is independent of $(X(u) : 0 \leq u < s)$.

Then the process $(X(t) : t \geq 0)$ is said to be a birth process with birth parameters $\lambda_0, \lambda_1, \lambda_2, \ldots > 0$.

**Remark.** The above properties imply that if $(X(t) : t \geq 0)$ is a birth process, then $X(t) \in \mathbb{N} \cup \{0\}$ for all $t \geq 0$, and $X(s) \leq X(t)$ whenever $0 \leq s \leq t$, so like the Poisson process, a birth process can be viewed as counting the number of events of a certain kind that occur between time 0 and time $t$. Often (but not always), these events are actual births in a population! Hence the name.

**Remark.** The birth parameter $\lambda_m$ can be understood as the ‘rate’ of arrival of the things we are counting, when the number of arrivals so far is $m$. (When we are counting births, for example, $\lambda_m$ is the ‘instantaneous birth rate’ when the population size is $m$.) It replaces $\lambda$ in the infinitesimal definition of the Poisson process.

**Example 40.** Let $\lambda > 0$. A Poisson process of rate $\lambda$ is precisely a birth process with $X(0) = 0$ and with birth parameters given by $\lambda_n = \lambda$ for all $n \in \mathbb{N} \cup \{0\}$.

**Example 41 (The ‘Linear’ Birth Process (also called the ‘Yule’ Birth Process)).** Let $\lambda > 0$. A Linear Birth Process with birth-rate $\lambda$ is a birth process with $X(0) = 1$ and with birth parameters given by $\lambda_n = n\lambda$ for all $n \in \mathbb{N} \cup \{0\}$.

The linear birth process is a very good model for the growth of a population of cells (e.g., yeast cells) where reproduction is asexual, and where population growth is not limited by a lack of food, space or other resources, and where cell-death does not occur. (This is the case, for example, when one yeast cell is placed in a large vat with plenty of space and food, in the early stages of population growth before cells start to die.)

Under these conditions, we can model each cell in the population as reproducing according to a Poisson process of some fixed rate $\lambda$, independently of all the other cells (and so in particular, independently of the number of other cells already present). The following lemma says that this does indeed result in a linear birth process.

**Lemma 2.13.** Let $\lambda > 0$ Suppose a population starts with one individual (member) at time 0. Suppose each individual in the population produces offspring (who are new members of the population) via asexual reproduction, according to a Poisson process of rate $\lambda$. Suppose these Poisson processes are mutually independent. Finally, suppose that no member of the population dies. Let $X(t)$ be the number of individuals in the population at time $t$. Then $(X(t) : t \geq 0)$ is a linear birth process with birth rate $\lambda$, and therefore birth parameters $\lambda_n = n\lambda$ for all $n \in \mathbb{N}$.

**Proof.** We will Property 2 of the definition of the birth process. (Property 3 is clear from Property 3 of the basic definition of the Poisson process.) Let $m \in \mathbb{N} \cup \{0\}$, and
let $t > 0$. If $X(t) = m$, let $(Y_i(t) : t \geq 0)$ be the Poisson process counting the number of offspring of the $i$th member of the population at time $t$. Then

$$\mathbb{P}(X(t+h) = m+1 \mid X(t) = m)$$

$$= \mathbb{P}\{\text{there exists } i \text{ s.t. } Y_i(t+h) - Y_i(t) = 1, Y_j(t+h) - Y_j(t) = 0 \text{ for all } j \neq i\}$$

$$= m \mathbb{P}\{Y_1(t+h) - Y_1(t) = 1, Y_j(t+h) - Y_i(t) = 0 \text{ for all } j \neq 1\}$$

$$= m(\lambda h + o(h))(1 - \lambda h + o(h))^{m-1}$$

(using the independence of the Poisson processes)

$$= m(\lambda h + o(h))(1 - (m - 1)\lambda h + o(h))$$

(using a binomial expansion)

$$= m\lambda h + o(h).$$

Moreover, we have

$$\mathbb{P}(X(t+h) = m \mid X(t) = m) = \mathbb{P}\{Y_j(t+h) = Y_i(t) = 0 \text{ for all } j\}$$

$$= (1 - \lambda h + o(h))^m$$

(again using the independence of the Poisson processes)

$$= 1 - m\lambda h + o(h).$$

Clearly, the size of the population cannot decrease, so $\mathbb{P}(X(t+h) < m \mid X(t) = m) = 0$. This verifies Property 2.

**Example 42.** Suppose we perform an experiment in which one yeast cell is placed in a large vat with unlimited food and space, and it and its descendants reproduce asexually by cell-division, so that yeast each cell produces offspring according to a Poisson process of rate $\lambda$ per hour, with these Poisson processes being mutually independent. Suppose that the yeast cells never die. Let $X(t)$ denote the number of yeast cells at time $t$ hours after the start of the experiment. Then $(X(t) : t \geq 0)$ is a linear birth process with birth parameters given by $\lambda_n = n\lambda$ for all $n \in \mathbb{N}$.

**Differential equations satisfied by a birth process**

We have the following useful theorem.

**Theorem 2.14.** Let $(X(t) : t \geq 0)$ be a birth process with $X(0) = s \in \mathbb{N} \cup \{0\}$ and with birth parameters $\lambda_0, \lambda_1, \lambda_2, \ldots$. For each $n \geq s$, define a function $p_n(t)$ by

$$p_n(t) = \mathbb{P}(X(t) = n).$$

Then the functions $p_n(t)$ satisfy the following system of differential equations.

$$p'_n(t) = -\lambda_n p_n(t),$$

$$p'_n(t) = -\lambda_n p_n(t) + \lambda_{n-1} p_{n-1}(t) \quad \forall n \geq s + 1.$$ 

Moreover, these equations have a unique solutions subject to the initial conditions $p_s(0) = 1$, $p_n(0) = 0 \forall n \geq s + 1$. 68
Proof. Let \( n \geq s + 1 \). Then

\[
p_n(t + h) = \mathbb{P}(X(t + h) = n)
\]

\[
= \sum_{k=0}^{n} \mathbb{P}(X(t + h) = n | X(t) = k) \mathbb{P}(X(t) = k)
\]

(by the law of total probability)

\[
= \mathbb{P}(X(t + h) = n | X(t) = n) \mathbb{P}(X(t) = n)
+ \mathbb{P}(X(t + h) = n | X(t) = n - 1) \mathbb{P}(X(t) = n - 1)
+ \sum_{k=0}^{n-2} \mathbb{P}(X(t + h) = n | X(t) = k) \mathbb{P}(X(t) = k)
\]

\[
= (1 - \lambda_n h + o(h))p_n(t) + (\lambda_{n-1} h + o(h))p_{n-1}(t) + o(h)
\]

(using property 2. of the definition of a birth process)

\[
= (1 - \lambda_n h)p_n(t) + \lambda_{n-1} hp_{n-1}(t) + o(h).
\]

Rearranging gives

\[
\frac{p_n(t + h) - p_n(t)}{h} = -\lambda_n p_n(t) + \lambda_{n-1} p_{n-1}(t) + \frac{o(h)}{h}.
\]

Taking the limit as \( h \to 0 \) gives

\[
p_n'(t) = -\lambda_n p_n(t) + \lambda_{n-1} p_{n-1}(t) \quad \forall n \geq s + 1,
\]

as required.

Now we deal with the case \( n = s \). If \( X(t + h) = s \), then we must have \( X(t) = s \) as well, so we have

\[
p_s(t + h) = \mathbb{P}(X(t + h) = s)
\]

\[
= \mathbb{P}(X(t + h) = s | X(t) = s) \mathbb{P}(X(t) = s)
\]

\[
= (1 - \lambda_s h + o(h))p_s(t)
\]

\[
= (1 - \lambda_s h)p_s(t) + o(h).
\]

Rearranging gives

\[
\frac{p_s(t + h) - p_s(t)}{h} = -\lambda_s p_s(t) + \frac{o(h)}{h}.
\]

Taking the limit as \( h \to 0 \) gives

\[
p_s'(t) = -\lambda_s p_s(t),
\]

as required.

Hence, the system of differential equations in the theorem is indeed satisfied. (We do not prove the uniqueness statement here.) \( \Box \)
We can use the differential equations in the above theorem to obtain an explicit formula for \( p_n(t) \), for any \( n \geq s \), in much the same way as we did for the Poisson process in Lemma 2.2. Here is an example of how to do this.

**Example 43.** Individuals in a population reproduce according to the following rule. At time 0, there are two individuals. At any time \( t \), if there are \( m \) individuals in the population, then each of the \( \binom{m}{2} \) possible pairs give birth to offspring according to a Poisson process of rate 3 per second. (All of the Poisson processes are independent of one another.) Let \( X(t) \) denote the number of individuals in the population at time \( t \) seconds. For each \( n \geq 2 \), define \( p_n(t) = \Pr(X(t) = n) \). Write down two differential equations satisfied by \( p_2(t) \) and \( p_3(t) \) (and involving only \( p_2(t) \), \( p_3(t) \) and their derivatives). Solve them to find explicit formulae for \( p_2(t) \) and \( p_3(t) \). Hence calculate the probability that there are at least four members of the population after five seconds.

Solution: \((X(t): t \geq 0)\) is a birth process with \( X(0) = 2 \), and with birth parameters given by \( \lambda_n = 3 \binom{n}{2} \) for all \( n \geq 2 \). (This is because if the population has \( n \) members, then each of the \( \binom{n}{2} \) possible pairs are separately giving birth to offspring according to independent Poisson processes of parameter \( \lambda = 3 \). The formal proof is very similar to the proof of Lemma 2.13.) In particular, \( \lambda_2 = 3 \) and \( \lambda_3 = 9 \). Hence, by the previous theorem, the functions \( p_2(t) \) and \( p_3(t) \) satisfy the following differential equations.

\[
\begin{align*}
p_2'(t) &= -3p_2(t), \\
p_3'(t) &= -9p_3(t) + 3p_2(t).
\end{align*}
\]

Integrating the first differential equation gives

\[
p_2(t) = Ce^{-3t}
\]

for some constant \( C \). Substituting in the initial condition \( p_2(0) = 1 \) gives \( 1 = C \), so

\[
p_2(t) = e^{-3t}.
\]

Substituting this into the second differential equation gives

\[
p_3'(t) = -9p_3(t) + 3e^{-3t}.
\]

Rearranging gives

\[
p_3'(t) + 9p_3(t) = 3e^{-3t}.
\]

Now observe, similarly to in the proof of Lemma 2.2, that we can rewrite the left-hand side as

\[
e^{-9t} \frac{d}{dt} (p_3(t)e^{9t}),
\]

because

\[
\frac{d}{dt} (e^{9t}p_3(t)) = 9e^{9t}p_3(t) + e^{9t}p_3'(t).
\]

So we get

\[
e^{-9t} \frac{d}{dt} (e^{9t}p_3(t)) = 3e^{-3t}.
\]
Multiplying both sides by $e^{9t}$ gives
\[
\frac{d}{dt} \left( e^{9t} p_3(t) \right) = 3e^{6t}.
\]
Integrating both sides gives
\[
e^{9t} p_3(t) = \frac{1}{2} e^{6t} + C',
\]
where $C'$ is a constant of integration. Substituting in the initial condition $p_3(0) = 0$ gives
\[
0 = \frac{1}{2} + C',
\]
so $C' = -\frac{1}{2}$, so
\[
p_3(t) = \frac{1}{2} e^{-3t} - \frac{1}{2} e^{-9t} = \frac{1}{2} e^{-3t}(1 - e^{-6t}).
\]

So to summarise, we have
\[
p_2(t) = e^{-3t}, \quad p_3(t) = \frac{1}{2} e^{-3t}(1 - e^{-6t}) \quad \forall t \geq 0.
\]
This implies that
\[
\mathbb{P}(X(t) \geq 4) = 1 - \mathbb{P}(X(t) \leq 3)
= 1 - \mathbb{P}(X(t) = 2) - \mathbb{P}(X(t) = 3)
= 1 - p_2(t) - p_3(t)
= 1 - e^{-3t} - \frac{1}{2} e^{-3t}(1 - e^{-6t})
= 1 - \frac{3}{2} e^{-3t} + \frac{1}{2} e^{-9t}.
\]

Hence, the probability that there are at least four members of the population after five seconds is
\[
\mathbb{P}(X(5) \geq 4) = 1 - \frac{3}{2} e^{-15} + \frac{1}{2} e^{-45}.
\]

The arrival times and interarrival times of a birth process

**Definition 2.8.** Let $(X(t): t \geq 0)$ be a birth process with $X(0) = s \in \mathbb{N} \cup \{0\}$ and with birth parameters $\lambda_0, \lambda_1, \lambda_2, \ldots$. For each $n \geq s + 1$, we define the *arrival time* $T_n$ by
\[
T_n = \min\{t \geq 0: X(t) \geq n\}.
\]
In other words, if $X(t)$ is the size of a certain population at time $t$, then the continuous random variable $T_n$ is the time of birth of the $n$th oldest member of the population, i.e. the time of the $(n - s)$th birth. (Note that at time 0, the population had $s$ members, so the birth of the $(s + 1)$th oldest member of the population is the time of the first birth, etc.)

For each $n \geq s + 1$, we define the *interarrival time* $S_n$ by
\[
S_n = \begin{cases} 
T_n & \text{if } n = s + 1; \\
T_n - T_{n-1} & \text{if } n \geq s + 2.
\end{cases}
\]

71
In other words, \( S_n \) is the length of time between the birth of the \((n-1)\)th oldest member of the population and the birth of the \(n\)th oldest member of the population, i.e. the length of time between the \((n-s-1)\)th birth and the \((n-s)\)th birth.

Similarly to in a Poisson process, it turns out that the interarrival times are mutually independent exponential random variables, but (unlike with the Poisson process), their parameters may differ: \( S_n \) is an exponential random variable with parameter \( \lambda_{n-1} \), the birth parameter corresponding to when there are \( n-1 \) individuals in the population. This is the content of the following useful theorem.

**Theorem 2.15.** Let \((X(t) : t \geq 0)\) be a birth process with \(X(0) = s \in \mathbb{N} \cup \{0\}\) and with birth parameters \( \lambda_0, \lambda_1, \lambda_2, \ldots \). Then the interarrival times \( S_{s+1}, S_{s+2}, S_{s+3}, \ldots \) are mutually independent random variables. Moreover, for each \( n \geq s+1 \), the random variable \( S_n \) is exponentially distributed with parameter \( \lambda_{n-1} \). So in particular, \( E[S_n] = \frac{1}{\lambda_{n-1}} \).

**Proof.** Non-examinable. \( \square \)

**Explosion**

In a (mathematical) birth process, it is theoretically possible for \( \sum_{n=s+1}^{\infty} S_n < \infty \), i.e. the sum of all the (infinitely many) interarrival times is finite! In this case, let

\[
r = \sum_{n=s+1}^{\infty} S_n.
\]

Then there are infinitely many births before time \( r \), so the arrival times \( T_{s+1}, T_{s+2}, \ldots \) are all at most \( r \), so they form an increasing sequence of real numbers all bounded above by \( r \), so they must tend to a limit, \( \zeta \) say. We then have \( T_n \to \zeta \) as \( n \to \infty \), so there are infinitely many births before time \( \zeta \), so \( X(t) \to \infty \) as \( t \to \zeta \). We then say that the birth process **explodes** at time \( \zeta \), and \( \zeta \) is called the **time of explosion**. (Informally, this means that the population reaches an infinite size at time \( \zeta \).

If, on the other hand, the birth process has \( \sum_{n=s+1}^{\infty} S_n = \infty \), then for any \( t > 0 \), there exists \( N \in \mathbb{N} \) such that \( X(t) \leq N \) for all \( u \leq t \), so the birth process does **not** explode. (This means that the population never reaches infinite size.)

Obviously, a birth process which realistically models a real-world population cannot explode, since there are only a finite number of particles on earth, so the number of individuals in the population cannot possibly exceed this number! However, explosion is certainly possible, mathematically. We would like to be able to decide whether or not a mathematical birth process is likely to explode.

Notice, from the previous theorem, that

\[
E \left[ \sum_{n=s+1}^{\infty} S_n \right] = \sum_{n=s+1}^{\infty} E[S_n] = \sum_{n=s+1}^{\infty} \frac{1}{\lambda_{n-1}} = \sum_{n=s}^{\infty} \frac{1}{\lambda_n}.
\]
If
\[ \sum_{n=s}^{\infty} \frac{1}{\lambda_n} < \infty, \]
then \( \mathbb{E} \left[ \sum_{n=s+1}^{\infty} S_n \right] < \infty \), so we must have \( \sum_{n=s+1}^{\infty} S_n < \infty \) with probability 1, so the birth process explodes with probability 1.

If, on the other hand, we have
\[ \sum_{n=s}^{\infty} \frac{1}{\lambda_n} = \infty, \]
i.e. \( \sum_{n=s+1}^{\infty} S_n = \infty \), then it turns out that with probability 1, the birth process will not explode. (Though we will not prove this in our course.) To summarize, we have the following useful theorem.

**Theorem 2.16.** Let \((X(t) : t \geq 0)\) be a birth process with \(X(0) = s \in \mathbb{N} \cup \{0\}\) and with birth parameters \(\lambda_0, \lambda_1, \lambda_2, \ldots\). Then

1. If \(\sum_{n=s}^{\infty} \frac{1}{\lambda_n} = \infty\), then the probability of explosion is 0.
2. If \(\sum_{n=s}^{\infty} \frac{1}{\lambda_n} < \infty\), then the probability of explosion is 1.

**Example 44.** A Poisson process of rate \(\lambda\) is a birth process with \(X(0) = 0\) and with birth parameters given by \(\lambda_n = \lambda\) for all \(n \in \mathbb{N} \cup \{0\}\). Hence,
\[ \sum_{n=0}^{\infty} \frac{1}{\lambda_n} = \sum_{n=0}^{\infty} \frac{1}{\lambda} = \infty, \]
so a Poisson process explodes with probability zero, by the previous theorem.

**Example 45.** A Linear (Yule) birth process with birth-rate \(\lambda\) has \(X(0) = 1\) and has birth parameters given by \(\lambda_n = n\lambda\) for all \(n \in \mathbb{N} \cup \{0\}\). Hence,
\[ \sum_{n=1}^{\infty} \frac{1}{\lambda_n} = \sum_{n=1}^{\infty} \frac{1}{n\lambda} = \frac{1}{\lambda} \sum_{n=1}^{\infty} \frac{1}{n} = \infty, \]
so this birth process explodes with probability zero, by the previous theorem.

**Example 46.** Let \((X(t) : t \geq 0)\) be the birth process in Example 43. Then \(X(0) = 2\) and \(\lambda_n = 3 \binom{n}{2}\) for all \(n \geq 2\), so
\[ \sum_{n=2}^{\infty} \frac{1}{\lambda_n} = \sum_{n=2}^{\infty} \frac{1}{3 \binom{n}{2}} = \frac{2}{3} \sum_{n=2}^{\infty} \frac{1}{n(n-1)} \leq \frac{4}{3} \sum_{n=1}^{\infty} \frac{1}{n^2} < \infty, \]
so this birth process explodes with probability 1, by the previous theorem.