# Discrete-Time Markov Chains

James Neill

*Supervisor: Professor Dmitry Korshunov*

April 2022

# Contents

**Abstract**

Markov Chains are a stochastic process where the probabilities of future values are independent of past values. Homogeneous Markov Chains are introduced on a countable state space and over discrete time, using transition probability matrices and distributions to classify states, calculate hitting times, and calculate long-run probabilities of being in a given state (asymptotic distributions). Applications of Markov Chains to simple random walks, branching processes, and Monte Carlo methods are also explored.

# 1 Introduction

A Markov Chain is a sequence of random variables where the probabilities of future values are independent of past values (given the present value). This report considers homogeneous Markov Chains over discrete time, and with countable (either finite or countably infinite) state space. We analyse Markov Chains primarily by using transition probability matrices, which consist of the probabilities of moving between specific states, and distributions, which consist of the probabilities of being in each state after a given number of steps. In Sections 2 and 3 we define other key terms, such as communicating classes, recurrent states, and aperiodic states.

We explore further aspects of Markov Chains in later sections. In Section 4 we explore hitting times – the number of transitions it takes a Markov Chain to reach a given subset of states – including a discussion of the Gambler's Ruin problem. In Section 5 we define invariant and asymptotic distributions, and prove the conditions of their existence. Finally, in Section 6 we explore applications of Markov Chains: we determine when symmetric simple random walks on $\mathbb{Z}^d$ are recurrent, use Markov Chains to model population growth (branching processes), and investigate Markov Chain Monte Carlo methods.

The majority of this report is based on Chapter 1 of [Nor97] and Chapter 12 of [GGDW14], including many of the proofs. This report expands on this material by adding examples (created directly for this report unless otherwised noted) and additional explanation.

# 2 Basic Definitions

## 2.1 The Markov Property

We begin with the definition of a discrete-time Markov Chain.

**Definition 2.1.1.** Consider a sequence of random variables $(X_n)_{n\geq 0}$. $(X_n)_{n\geq 0}$ is a *Markov Chain (MC)* with *state space I* if

$$\mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}, X_{n-2} = x_{n-2}, \dots, X_0 = x_0) = \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1})$$

for all $n \in \mathbb{N}$, $x_0, \dots, x_n \in I$.

This means the value of $X_n$ is dependent on the value of $X_{n-1}$, but independent of $X_{n-2}, \dots, X_1, X_0$ – the future is independent of the past given the current time state. This is called the *Markov property*.

**Definition 2.1.2.** A MC is *homogeneous* if the probabilities of transition between states are constant over time. This means for any $i, j \in I$, the value of $\mathbb{P}(X_n = j | X_{n-1} = i)$ is the same for all values of $n$.

*Remark* 2.1.3. We will assume that all Markov Chains are homogeneous, and that $I$ is countable (either finite or countably infinite), unless stated otherwise.

This means we can define the following matrix element-wise for each pair $i, j$:

**Definition 2.1.4.** The *transition probability matrix (TPM) P* of a MC $(X_n)$ is defined element-wise by

$$P_{i,j} = \mathbb{P}(X_1 = j | X_0 = i)$$

for all $i, j \in I$.

When $I$ is finite (with $|I| = n$ for some $n \in \mathbb{N}$) then we clearly have a well-defined $n$-by-$n$ matrix. When $I$ is countably infinite we cannot write out the matrix in full, but we can define $P$ element-wise. In this case we still have for any $i, j \in I$ that

$$(P^2)_{i,j} = \sum_{k \in I} P_{i,k} P_{k,j},$$

where $\sum_{k \in I} P_{i,k} P_{k,j}$ is a convergent series. By iteration we may calculate $P^m$ for any $m \in \mathbb{N}$, noting that $P^0$ is the identity matrix.

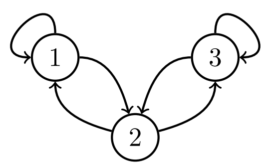
For any row $i$, $\sum_{j \in I} P_{i,j}$ is the sum of the probabilities of moving to every possible value of $X_1$, given that $X_0 = i$. This means $\sum_{j \in I} P_{i,j} = 1$.

We often draw graphs to visualise Markov Chains, with vertices representing states and directed lines from vertex $i$ to $j$ whenever $P_{i,j} > 0$. Sometimes it is useful to label the directed lines with the corresponding probabilities.

*Example* 2.1.5. A Markov Chain with $I = \{1, 2, 3\}$ and TPM

$$P = \begin{pmatrix} 0.7 & 0.3 & 0 \\ 0.5 & 0 & 0.5 \\ 0 & 0.3 & 0.7 \end{pmatrix}$$

can be represented by the following graph:




The element $P_{i,j}$ is the probability that, starting from state $i$, the MC is in state $j$ after one step. We will now generalise this to the probability of being in state $j$ after $m$ steps, for some $m \in \mathbb{N}$.

**Definition 2.1.6.** The *$m$-step transition probability matrix* $P^{(m)}$ is defined element-wise by

$$P_{i,j}^{(m)} = \mathbb{P}(X_m = j | X_0 = i)$$

for all $i, j \in I$.

The $m$-step transition probability matrix can be calculated using the following theorem, the proof of which is generalised from [Kor20] pg 44. This theorem means that we can use $P^{(m)}$ and $P^m$ interchangeably, and directly leads to the following corollary – the *Chapman–Kolmogorov equation.*

**Theorem 2.1.7.** *For all $m \in \mathbb{N}$, $P^{(m)} = P^m$.*

*Proof.* We will prove by induction. Clearly the base case $m = 1$ is true, since

$$P^{(1)} = \mathbb{P}(X_1 = j | X_0 = i) = P = P^1.$$

We will assume the $n$th case is true, so

$$P^{(n)} = P^n, \tag{2.1.1}$$

and attempt to show the $(n+1)$th case is true:

$$P_{i,j}^{(n+1)} = \mathbb{P}(X_{n+1} = j | X_0 = i)$$

$$
\begin{aligned}
&= \sum_{k\in I} \mathbb{P}(X_{n+1} = j, X_n = k | X_0 = i) \\
&= \sum_{k\in I} \mathbb{P}(X_{n+1} = j | X_n = k, X_0 = i)\mathbb{P}(X_n = k | X_0 = i),
\end{aligned}
$$

and so by the Markov Property (Definition 2.1.1), we have

$$
\begin{aligned}
P_{i,j}^{(n+1)} &= \sum_{k\in I} \mathbb{P}(X_{n+1} = j | X_n = k)\mathbb{P}(X_n = k | X_0 = i) \\
&= \sum_{k\in I} P_{i,k} P_{k,j}^{(n)} \\
&= (P\ P^{(n)})_{i,j}.
\end{aligned}
$$

Thus by our assumption (2.1.1),

$$
\begin{aligned}
P_{i,j}^{(n+1)} &= (P\ P^n)_{i,j} \\
&= P_{i,j}^{n+1}.
\end{aligned}
$$

This means $P^{(n+1)} = P^{n+1}$, and so we have proved $P^{(m)} = P^m$ for all $m$. ☐

**Corollary 2.1.8** (Chapman—Kolmogorov equation)**.** *For any $m, n \in \mathbb{N}_0$, we have*

$$P^{(m+n)} = P^{(m)} P^{(n)}.$$

*Proof.* From Theorem 2.1.7, we have $P^{(m+n)} = P^{m+n} = P^m P^n = P^{(m)} P^{(n)}$. ☐

## 2.2 Communicating Classes

Now we consider $\leftrightarrow$, an important relation between states of a Markov Chain.

**Definition 2.2.1.** We say that $i$ *leads to* $j$, $i \to j$, if

$$\mathbb{P}(X_n = j \text{ for some } n \geq 0 | X_0 = i) > 0.$$

We say that $i$ *communicates with* $j$, $i \leftrightarrow j$, if both $i \to j$ and $j \to i$.

**Theorem 2.2.2.** *The relation $\leftrightarrow$ is an equivalence relation on the state space $I$.*

*Proof.* For any $i \in I$, $\mathbb{P}(X_0 = i | X_0 = i) = 1 > 0$, which means $i \to i$, and so $i \leftrightarrow i$. Therefore $\leftrightarrow$ is reflexive.

If $i \leftrightarrow j$, then clearly $j \leftrightarrow i$. Therefore $\leftrightarrow$ is symmetric.

We take the last argument from [GGDW14] pg 212. If $i \to j$ and $j \to k$, then for some $m, n \geq 0$, we have $P_{i,j}^{(m)} > 0$ and $P_{j,k}^{(n)} > 0$. By Corollary 2.1.8, $P^{(m+n)} = P^{(m)} P^{(n)}$. Since all the entries of $P$ are probabilities (and so are non-negative), this means that $P_{i,k}^{(m+n)} \geq P_{i,j}^{(m)} P_{j,k}^{(n)} > 0$, and so $i \to k$. Therefore if $i \leftrightarrow j$ and $j \leftrightarrow k$, then $i \leftrightarrow k$, so $\leftrightarrow$ is transitive.

Therefore $\leftrightarrow$ is an equivalence relation. ☐

We call the equivalence classes of $\leftrightarrow$ in $I$ *communicating classes*, which partition $I$.

**Definition 2.2.3.** A MC where $I$ consists of one communicating class is called *irreducible.*

Now we consider *closedness*, a property communicating classes may have.

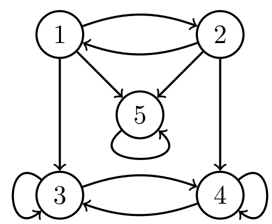
**Definition 2.2.4.** Let $C$ be a communicating class in $I$, $i \in C$, $j \in I$, and $i \to j$. Then $C$ is *closed* if $j \in C$. (Otherwise, $C$ is *open*.)

**Definition 2.2.5.** Let $i \in I$. Then if $\{i\}$ is closed we call $i$ *absorbing*.

*Example* 2.2.6. Consider a Markov Chain with $I = \{1, 2, 3, 4, 5\}$, TPM

$$P = \begin{pmatrix} 0 & 0.8 & 0.1 & 0 & 0.1 \\ 0.5 & 0 & 0 & 0.4 & 0.1 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0.2 & 0.8 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

and graph:




The communicating classes are $\{1, 2\}$ (which is open), $\{3, 4\}$ (which is closed), and $\{5\}$ (which is closed). This means the MC is not irreducible, and 5 is an absorbing state.

# 3 Classification of States

## 3.1 The Strong Markov Property

In this section we will classify states in Markov Chains as either recurrent or transient, and as either periodic or aperiodic. These definitions explore important behaviour of Markov Chains, and are used in Section 5 to determine the existence of invariant and asymptotic distributions. Before defining these terms we will expand our definition of the Markov Property by Theorem 3.1.2, which will be used in Section 3.2.

**Definition 3.1.1.** Let $(X_n)$ be a MC with sample space $\Omega$. A random variable $T : \Omega \to \mathbb{N}_0 \cup \{\infty\}$ is a *stopping time* if for all $n \in \mathbb{N}_0$ the event $\{T = n\}$ is independent of $X_{n+1}, X_{n+2}, \dots$ (dependent on $X_0, \dots, X_n$ only).

Examples of stopping times include hitting times and return times, which will be explored in Section 4. This definition leads to the following theorem (with proof from [GGDW14] pg 225):

**Theorem 3.1.2** (Strong Markov property)**.** *Let $(X_n)$ be a Markov Chain with TPM $P$, let $T$ be a stopping time, and let $(Y_n)$ be a sequence of random variables defined by $Y_n = X_{T+n}$ for all $n \in \mathbb{N}_0$. If $T < \infty$ and $X_T = i_0$, then $(Y_n)$ is a Markov Chain with TPM $P$ and $Y_0 = i_0$, and is independent of $X_0, X_1, \dots, X_T$.*

*Proof.* Let $i_1, i_2, \dots \in I$. Then for any $n \in \mathbb{N}$

$$\begin{aligned} &\mathbb{P}(Y_n = i_n | Y_{n-1} = i_{n-1}, Y_{n-2} = i_{n-2}, \dots, Y_0 = i_0) \\ &= \mathbb{P}(X_{T+n} = i_n | X_{T+n-1} = i_{n-1}, X_{T+n-2} = i_{n-2}, \dots, X_T = i_0) \\ &= \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}, X_{n-2} = i_{n-2}, \dots, X_0 = i_0), \end{aligned}$$

where again the final step is by the assumption that the MC is homogeneous. By the Markov Property (Definition 2.1.1), we have

$$\mathbb{P}(Y_n = i_n | Y_{n-1} = i_{n-1}, Y_{n-2} = i_{n-2}, \dots, Y_0 = i_0) = \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}),$$

$$= \mathbb{P}(X_{T+n} = i_n | X_{T+n-1} = i_{n-1}),$$

where the final step is by the assumption that the MC is homogeneous. This means

$$\mathbb{P}(Y_n = i_n | Y_{n-1} = i_{n-1}, Y_{n-2} = i_{n-2}, \dots, Y_0 = i_0) = \mathbb{P}(Y_n = i_n | Y_{n-1} = i_{n-1}).$$

Therefore $(Y_n)$ is a Markov Chain (clearly with TPM $P$ and $Y_0 = i_0$).

Let $H$ be an event dependent on $X_0, X_1, \dots, X_{T-1}$ only. Then we have

$$\begin{aligned}
&\mathbb{P}(Y_1 = i_1, Y_2 = i_2, \dots, H | T < \infty, X_T = i_0) \\
&= \sum_{m=0}^{\infty} \mathbb{P}(Y_1 = i_1, Y_2 = i_2, \dots, H, T = m | T < \infty, X_T = i_0) \\
&= \sum_{m=0}^{\infty} \mathbb{P}(X_{T+1} = i_1, X_{T+2} = i_2, \dots, H, T = m | T < \infty, X_T = i_0) \\
&= \sum_{m=0}^{\infty} \mathbb{P}(X_{m+1} = i_1, X_{m+2} = i_2, \dots, H \cap \{T = m\} | T < \infty, X_T = i_0).
\end{aligned}$$

Since $T$ is a stopping time, the event $H \cap \{T = m\}$ depends on $X_0, \dots, X_{m-1}$ only. Therefore

$$\begin{aligned}
&\mathbb{P}(Y_1 = i_1, Y_2 = i_2, \dots, H | T < \infty, X_T = i_0) \\
&= \sum_{m=0}^{\infty} \mathbb{P}(X_{m+1} = i_1, X_{m+2} = i_2, \dots) \mathbb{P}(H \cap \{T = m\} | T < \infty, X_T = i_0) \\
&= \sum_{m=0}^{\infty} \mathbb{P}(X_{T+1} = i_1, X_{T+2} = i_2, \dots, T = m) \mathbb{P}(H, T = m | T < \infty, X_T = i_0) \\
&= \sum_{m=0}^{\infty} \mathbb{P}(Y_1 = i_1, Y_2 = i_2, \dots, T = m) \mathbb{P}(H, T = m | T < \infty, X_T = i_0) \\
&= \mathbb{P}(Y_1 = i_1, Y_2 = i_2, \dots) \mathbb{P}(H | T < \infty, X_T = i_0),
\end{aligned}$$

which means $(Y_n)$ is independent of $X_0, X_1, \dots, X_T$, as required. □

This means the Markov Property holds at a random stopping time $T$, rather than simply at a known state $i$ – this is known as the *Strong Markov property.*

## 3.2 Recurrent and Transient States

Firstly, we will classify states in a Markov Chain by the likelihood that, after leaving the state, the chain will eventually return to that state.

**Definition 3.2.1.** For a MC $(X_n)$, a state $i$ is *recurrent* if $\mathbb{P}(X_n = i \text{ for some } n \geq 1 | X_0 = i) = 1$, and is *transient* if $\mathbb{P}(X_n = i \text{ for some } n \geq 1 | X_0 = i) < 1$.

This means a state is recurrent if the time taken to return is finite – we will return to this in Section 4.3. Since the probability of any event is at most 1, all states are either recurrent or transient.

*Remark* 3.2.2. Consider a recurrent state $i$. If $X_0 = i$, we know there exists some time $T_1 \in \mathbb{N}$ such that $X_{T_1} = i$. This is clearly a stopping time, since it is random, and dependent on $X_0, \dots, X_{T_1}$ only.

Therefore by the Strong Markov Property (Theorem 3.1.2), we may consider a new MC starting from $X_{T_1} = i$, and so we know there exists some random $T_2$ such that $X_{T_1+T_2} = i$.

Continuing this process, we see that $(X_n)$ is guaranteed to enter state $i$ infinitely many times, and so

$$\mathbb{P}(X_n = i \text{ for infinitely many } n | X_0 = i) = 1.$$

*Remark* 3.2.3. Conversely, if $i$ is transient, then

$$\mathbb{P}(X_n = i \text{ for infinitely many } n | X_0 = i) = \lim_{m\to\infty} \mathbb{P}(X_n = i \text{ for some } n \geq 1 | X_0 = i)^m = 0,$$

since $\mathbb{P}(X_n = i \text{ for some } n \geq 1 | X_0 = i) < 1$.

*Example* 3.2.4. We continue from Example 2.2.6. States 1 and 2 are transient, since if they first move to state 5 (with probability 0.1) then they will never return. Also, state 5 is clearly recurrent, since $\mathbb{P}(X_1 = 5 | X_0 = 5) = 1$.

Now consider state 3. Both states 3 and 4 can only go to states 3 and 4, so the probability of starting in state 3, then never returning, is 0.5 (from state 3 to state 4) times 0.5 (staying in state 4) times 0.5 (staying in state 4), and so on – this converges to 0. This means the probability of eventually returning to state 3 is 1, and so state 3 is recurrent. Similarly, state 4 is recurrent.

The following theorem (with proof from [Ros14] pg 196-197) is used to determine whether a state is recurrent or transient:

**Theorem 3.2.5.** *A state $i$ is recurrent if and only if $\sum_{n=0}^{\infty} P_{i,i}^{(n)} = \infty$, and a state $i$ is transient if and only if $\sum_{n=0}^{\infty} P_{i,i}^{(n)} < \infty$.*

*Proof.* Let $i$ be a recurrent state, and define $I_n$ by $I_n = 1$ if $X_n = i$ and by $I_n = 0$ if $X_n \neq i$. Then $\sum_{n=0}^{\infty} I_n$ is the number of times the MC is in state $i$. By Remark 3.2.2, $(X_n)$ will be at state $i$ infinitely many times, so $\mathbb{E}(\sum_{n=0}^{\infty} I_n | X_0 = i) = \infty$.

We also have, by the linearity of expectation,

$$\begin{aligned}
\mathbb{E}\left(\sum_{n=0}^{\infty} I_n \middle| X_0 = i\right) &= \sum_{n=0}^{\infty} \mathbb{E}(I_n | X_0 = i) \\
&= \sum_{n=0}^{\infty} (1 \cdot \mathbb{P}(I_n = 1 | X_0 = i) + 0 \cdot \mathbb{P}(I_n = 0 | X_0 = i)) \\
&= \sum_{n=0}^{\infty} \mathbb{P}(X_n = i | X_0 = i) \\
&= \sum_{n=0}^{\infty} P_{i,i}^{(n)}.
\end{aligned}$$

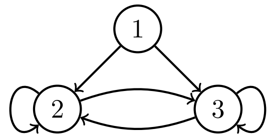
Therefore $i$ is recurrent if and only if $\sum_{n=0}^{\infty} P_{i,i}^{(n)} = \infty$, and so $i$ is transient if and only if $\sum_{n=0}^{\infty} P_{i,i}^{(n)} < \infty$. □

This is a very useful theorem, as by Theorem 2.1.7 we can easily calculate $P^{(n)}$ for any $n \in \mathbb{N}$. We will now use this theorem to show recurrence and transient of states for a specific example, and prove the link between between communicating classes and recurrence/transience of states.

*Example* 3.2.6. Consider a Markov Chain with $I = \{1, 2, 3\}$, TPM

$$P = \begin{pmatrix} 0 & 0.5 & 0.5 \\ 0 & 0.5 & 0.5 \\ 0 & 0.5 & 0.5 \end{pmatrix},$$

and graph:




Clearly $P^2 = P$, which means that $P^n = P$ for all $n$. This means

$$\sum_{n=0}^{\infty} P_{1,1}^n = \sum_{n=0}^{\infty} 0 = 0,$$

so by Theorem 3.2.5, state 1 is transient, and for $i \in \{2, 3\}$,

$$\sum_{n=0}^{\infty} P_{i,i}^n = \sum_{n=0}^{\infty} 0.5 = \infty,$$

so by Theorem 3.2.5, states 2 and 3 are recurrent.

**Theorem 3.2.7.** *Given a communicating class $C$, either all states are recurrent, or all states are transient.*

*Proof.* We use the proof from [Nor97] pg 26-27. Let $i, j \in C$. Then $i \to j$ and $j \to i$, so there exists $n, m \in \mathbb{N}$ such that $P_{i,j}^n > 0$ and $P_{j,i}^m > 0$.

Let $r \in \mathbb{N}$. Then $P_{i,i}^{n+r+m} = (P^n P^r P^m)_{i,i} \geq P_{i,j}^n P_{j,j}^r P_{j,i}^m$ (since the entries of $P$ are probabilities and so are non-negative).

Therefore

$$\begin{aligned}
\sum_{r=0}^{\infty} P_{j,j}^r &\leq \sum_{r=0}^{\infty} \frac{P_{i,i}^{n+r+m}}{P_{i,j}^n P_{j,i}^m} \\
&= \frac{1}{P_{i,j}^n P_{j,i}^m} \sum_{r=0}^{\infty} P_{i,i}^{n+r+m} \\
&= \frac{1}{P_{i,j}^n P_{j,i}^m} \sum_{s=n+m}^{\infty} P_{i,i}^s \\
&\leq \frac{1}{P_{i,j}^n P_{j,i}^m} \sum_{s=0}^{\infty} P_{i,i}^s.
\end{aligned}$$

By Theorem 3.2.5, if $i$ is transient, $\sum_{s=0}^{\infty} P_{i,i}^s < \infty$, and so $\sum_{r=0}^{\infty} P_{j,j}^r < \infty$, and so $j$ is transient.

Therefore if one state in $C$ is transient, all states in $C$ must be transient – so either all states are transient or all states are recurrent. □

This means we may refer to entire classes as either recurrent or transient. We now explore the relationship between closed and recurrent classes, with proofs of the following two theorems from [Nor97] pg 27.

**Theorem 3.2.8.** *Every recurrent class is closed.*

*Proof.* Let $C$ be an open communicating class. Then there exists $i \in C$, $j \in I \backslash C$ such that $i \to j$ – this means there exists $m \in \mathbb{N}$ such that $\mathbb{P}(X_m = j | X_0 = i) > 0$.

If the MC transitions to state $j$, it can never transition back to state $i$, since $j \notin C$. This means

$$\mathbb{P}(\{X_m = j\} \cap \{X_n = i \text{ for infinitely many } n\} | X_0 = i) = 0.$$

Therefore $\mathbb{P}(X_n = i \text{ for infinitely many } n | X_0 = i) < 1$, and so by Remark 3.2.2, state $i$ is not recurrent.

Therefore by contraposition we have that every recurrent class is closed. □

**Theorem 3.2.9.** *Every closed class with a finite number of states is recurrent.*

*Proof.* Let $C$ be a closed, finite class, and let $X_0 \in C$. Since $C$ has a finite number of states, there must exist some $i \in C$ such that

$$\mathbb{P}(X_n = i \text{ for infinitely many } n) > 0.$$

(Or else the MC would be at a finite number of states a finite number of times, which means the MC has finitely many terms, a contradiction.)

By the Strong Markov Property (Theorem 3.1.2), we have

$$\mathbb{P}(X_n = i \text{ for infinitely many } n) = \mathbb{P}(X_n = i \text{ for some } n)\mathbb{P}(X_n = i \text{ for infinitely many } n | X_0 = i),$$

which means

$$\mathbb{P}(X_n = i \text{ for infinitely many } n | X_0 = i) > 0.$$

By Remark 3.2.3, this means $i$ is recurrent, and therefore $C$ is recurrent. □

We see the results of these theorems in previous examples:

*Example* 3.2.10. We see that in Example 2.2.6 class $\{1, 2\}$ is both open and transient, and classes $\{3, 4\}$ and $\{5\}$ are both closed and recurrent. In Example 3.2.6, we clearly have the open class $\{1\}$ – which is transient – and the closed class $\{2, 3\}$ – which is closed.

Note that Theorem 3.2.9 does not hold for infinite closed classes – we will see an infinite closed class that is transient in Section 6.1.

## 3.3 Periodic and Aperiodic States

We may also classify states in a Markov Chain by the greatest common denominator of all possible times it may take the MC to return to the state after leaving.

**Definition 3.3.1.** The *period* of a state $i \in I$ is

$$d_i = \gcd \left\{ n \geq 1 \middle| P_{i,i}^{(n)} > 0 \right\}.$$

**Definition 3.3.2.** A state is *aperiodic* if it has period 1, and *periodic* otherwise.

**Definition 3.3.3.** A Markov Chain is *aperiodic* if all states have period 1.

As with recurrent and transient states, the periodicity of states in the same communicating class are the same:

**Theorem 3.3.4.** *All states in the same communicating class have the same period.*

*Proof.* We use the proof from [GGDW14] pg 229. Let $i, j \in I$ be in some communicating class $C$. This means $i \to j$ and $j \to i$, so there exists $n, m \in \mathbb{N}$ such that $P_{i,j}^n > 0$ and $P_{j,i}^m > 0$. Let $\alpha = P_{i,j}^n P_{j,i}^m > 0$.

Let $r \in \mathbb{N}_0$. Then

$$P_{i,i}^{n+r+m} = (P^n P^r P^m)_{i,i} \geq P_{i,j}^n P_{j,j}^r P_{j,i}^m = \alpha P_{j,j}^r,$$

because the entries of $P$ are probabilities, and so are non-negative.

If $r = 0$, we have $P_{i,i}^{n+m} \geq \alpha$. This means the MC can start from $i$ and then return to $i$ after $n+m$ steps – so $d_i \mid (n+m)$.

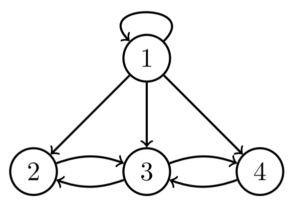
This means if $d_i \nmid r$ then $d_i \nmid (r+n+m)$, and so the MC cannot start from $i$ and then return to $i$ after $n+r+m$ steps, so $P_{i,i}^{n+r+m} = 0$. Therefore $P_{j,j}^r = 0$, so $d_j \nmid r$.

Similarly, if $d_j \nmid r$ then $d_i \nmid r$ – for any $r \in \mathbb{N}_0$, $d_i \mid r$ if and only if $d_j \mid r$. Letting $r = d_i$ means $d_i \mid d_i$ if and only if $d_i \mid d_j$, and so we know $d_i \mid d_j$. Similarly, we know $d_j \mid d_i$, and so $d_i = d_j$, as required. □

*Example* 3.3.5. Consider a Markov Chain with $I = \{1, 2, 3, 4\}$, TPM

$$P = \begin{pmatrix} 0.4 & 0.2 & 0.2 & 0.2 \\ 0 & 0 & 1 & 0 \\ 0 & 0.6 & 0 & 0.4 \\ 0 & 0 & 1 & 0 \end{pmatrix},$$

and graph:




The MC can transition from state 1 directly back to state 1, so state 1 is clearly aperiodic. States 2, 3, and 4 are periodic with period 2 because they can only return to themselves after an even number of steps (as seen from the graph). The communicating classes are $\{1\}$ and $\{2, 3, 4\}$, and we see that all states in $\{2, 3, 4\}$ have a period of 2.

# 4 Hitting Times

## 4.1 Hitting Times and Probabilities

We now discuss the number of transitions (the time taken) between states.

**Definition 4.1.1.** The *hitting time* of $A \subset I$ is

$$H^A = \inf\{n \in \mathbb{N}_0 | X_n \in A\},$$

where $\inf\{\emptyset\} = \infty$.

This means $H^A$ is the time number of steps it takes to reach a state in $A$. The event $\{H^A = n\}$ (for some $n \in \mathbb{N}_0$) is dependent only on the values of $X_0, \ldots, X_n$, and so by Definition 3.1.1, $H^A$ is a stopping time.

From our definition of $H^A$ we can make the following definitions:

**Definition 4.1.2.** The *hitting probability* of $A \subset I$ starting from the state $i$ is

$$h_i^A = \mathbb{P}(H^A < \infty | X_0 = i).$$

We denote $h^A = (h_i^A : i \in I)$.

**Definition 4.1.3.** The *expected hitting time* of a $A \subset I$ starting from the state $i$ is

$$k_i^A = \mathbb{E}(H^A | X_0 = i).$$

We denote $k^A = (k_i^A : i \in I)$.

We calculate hitting probabilities and expected hitting times using the following theorems from [Nor97], pg 13-14 and 17-18 respectively:

**Theorem 4.1.4.** *The vector of hitting probabilities $h^A$, calculated element-wise for some $i \in I$, is the minimum (non-negative) solution to*

$$h_i^A = \begin{cases} 1 & \text{if } i \in A, \\ \sum_{j \in I} P_{i,j} h_j^A & \text{if } i \notin A. \end{cases} \tag{4.1.1}$$

*Proof.* If $X_0 = i \in A$, then $H^A = 0$, so $h_i^A = \mathbb{P}(0 < \infty) = 1$.

If $X_0 = i \notin A$, then $H^A \geq 1$, so we have

$$\begin{aligned} \mathbb{P}(H^A < \infty | X_0 = i) &= \sum_{j \in I} \mathbb{P}(H^A < \infty, X_1 = j | X_0 = i) \\ &= \sum_{j \in I} \mathbb{P}(H^A < \infty | X_1 = j, X_0 = i)\mathbb{P}(X_1 = j | X_0 = i) \\ &= \sum_{j \in I} \mathbb{P}(H^A < \infty | X_1 = j)\mathbb{P}(X_1 = j | X_0 = i) \\ &= \sum_{j \in I} P_{i,j} \mathbb{P}(H^A < \infty | X_0 = j), \end{aligned}$$

which means

$$h_i^A = \sum_{j \in I} P_{i,j} h_j^A,$$

as required.

Let $x$ be any solution of (4.1.1). If $i \in A$, then $x_i = 1 = h_i^A$. If $i \notin A$, then for any $n \in \mathbb{N}$ we have

$$\begin{aligned} x_i &= \sum_{j \in I} P_{i,j} x_j \\ &= \sum_{j \in A} P_{i,j} x_j + \sum_{j \in I \setminus A} P_{i,j} x_j. \end{aligned}$$

When $j \in A$, $x_j = 1$, which means

$$\begin{aligned} x_i &= \sum_{j \in A} P_{i,j} + \sum_{j \in I \setminus A} P_{i,j} \left( \sum_{k \in I} P_{j,k} x_k \right) \\ &= \sum_{j \in A} P_{i,j} + \sum_{j \in I \setminus A} P_{i,j} \left( \sum_{k \in A} P_{j,k} x_k + \sum_{k \in I \setminus A} P_{j,k} x_k \right) \\ &= \sum_{j \in A} P_{i,j} + \sum_{j \in I \setminus A} P_{i,j} \sum_{k \in A} P_{j,k} + \sum_{j \in I \setminus A} P_{i,j} \sum_{k \in I \setminus A} P_{j,k} x_k \\ &= \mathbb{P}(X_1 \in A | X_0 = i) + \mathbb{P}(X_1 \notin A, X_2 \in A | X_0 = i) + \sum_{j \in I \setminus A} P_{i,j} \sum_{k \in I \setminus A} P_{j,k} x_k. \end{aligned}$$

After $n$ steps, we have

$$x_i = \mathbb{P}(X_1 \in A | X_0 = i) + \cdots + \mathbb{P}(X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A | X_0 = i)$$

$$
\begin{aligned}
&+ \sum_{j_1 \in I \setminus A} P_{i,j_1} \cdots \sum_{j_{n-1} \in I \setminus A} P_{j_{n-2},j_{n-1}} \sum_{j_n \in I \setminus A} P_{j_{n-1},j_n} x_{j_n} \\
&\geq \mathbb{P}(X_1 \in A | X_0 = i) + \cdots + \mathbb{P}(X_1 \notin A, \ldots, X_{n-1} \notin A, X_n \in A | X_0 = i) \\
&= \mathbb{P}(H^A = 1 | X_0 = i) + \cdots + \mathbb{P}(H_i^A = n | X_0 = i) \\
&= \mathbb{P}(H^A \leq n | X_0 = i).
\end{aligned}
$$

Taking limits, we have

$$
x_i \geq \lim_{n \to \infty} \mathbb{P}(H^A \leq n | X_0 = i) = \mathbb{P}(H^A < \infty | X_0 = i) = h_i^A,
$$

and so $x_i \geq h_i^A$ for all $i \in I$, as required. □

*Example* 4.1.5. We continue from Example 2.2.6, letting $A = \{3, 4\} \subset I$.

From the graph we see that $h_5^A = 0$, since $5 \nrightarrow 3$ and $5 \nrightarrow 4$. From Theorem 4.1.4, we have $h_3^A = 1$, $h_4^A = 1$, and so

$$
\begin{aligned}
h_1^A &= 0.8 h_2^A + 0.1, \\
h_2^A &= 0.5 h_1^A + 0.4.
\end{aligned}
$$

Solving these equations gives $h_1^A = 0.7$ and $h_2^A = 0.75$, so

$$
h^A = (0.7 \quad 0.75 \quad 1 \quad 1 \quad 0).
$$

**Theorem 4.1.6.** *The vector of expected hitting times $k^A$, calculated element-wise for some $i \in I$, is the minimum (non-negative) solution to*

$$
k_i^A = \begin{cases} 0 & \text{if } i \in A, \\ 1 + \sum_{j \in I \setminus A} P_{i,j} k_j^A & \text{if } i \notin A. \end{cases} \tag{4.1.2}
$$

*Proof.* If $X_0 = i \in A$, then $H^A = 0$, so $k_i^A = \mathbb{E}(0) = 0$.

If $X_0 = i \notin A$, then $H^A \geq 1$, so we have

$$
\begin{aligned}
\mathbb{E}(H^A | X_0 = i) &= \sum_{j \in I} \mathbb{E}(H^A \cdot \mathbf{1}_{X_1 = j} | X_0 = i) \\
&= \sum_{j \in I} \mathbb{E}(H^A | X_1 = j, X_0 = i) \mathbb{P}(X_1 = j | X_0 = i) \\
&= \sum_{j \in A} \mathbb{E}(H^A | X_1 = j, X_0 = i) \mathbb{P}(X_1 = j | X_0 = i) + \sum_{j \in I \setminus A} \mathbb{E}(H^A | X_1 = j, X_0 = i) \mathbb{P}(X_1 = j | X_0 = i).
\end{aligned}
$$

If $X_1 = j \in A$ and $X_0 \notin A$, then $H^A = 1$, so the expected hitting time is 1. If $X_1 = j \notin A$ and $X_0 \notin A$, then by the Strong Markov Property we can consider the MC starting from $X_0 = j$, and add 1 to the new expected hitting time. Therefore

$$
\begin{aligned}
\mathbb{E}(H^A | X_0 = i) &= \sum_{j \in A} 1 \cdot P_{i,j} + \sum_{j \in I \setminus A} (\mathbb{E}(H^A | X_0 = j) + 1) P_{i,j} \\
&= \sum_{j \in A} P_{i,j} + \sum_{j \in I \setminus A} P_{i,j} + \sum_{j \in I \setminus A} \mathbb{E}(H^A | X_0 = j) P_{i,j}
\end{aligned}
$$

$$= 1 + \sum_{j \in I \setminus A} \mathbb{E}(H^A | X_0 = i) P_{i,j},$$

which means

$$k_i^A = 1 + \sum_{j \in I \setminus A} P_{i,j} k_j^A,$$

as required.

Let $y$ be any solution of (4.1.2). If $i \in A$, then $y_i = 0 = k_i^A$. If $i \notin A$, then for any $n \in \mathbb{N}$ we have

$$\begin{aligned}
y_i &= 1 + \sum_{j \in I \setminus A} P_{i,j} y_j \\
&= 1 + \sum_{j \in I \setminus A} P_{i,j} \left( 1 + \sum_{k \in I \setminus A} P_{j,k} y_k \right) \\
&= 1 + \sum_{j \in I \setminus A} P_{i,j} + \sum_{j \in I \setminus A} P_{i,j} \sum_{k \in I \setminus A} P_{j,k} y_k \\
&= \mathbb{P}(H^A \geq 1 | X_0 = i) + \mathbb{P}(H^A \geq 2 | X_0 = i) + \sum_{j \in I \setminus A} P_{i,j} \sum_{k \in I \setminus A} P_{j,k} y_k.
\end{aligned}$$

After $n$ steps, we have

$$\begin{aligned}
y_i &= \mathbb{P}(H^A \geq 1 | X_0 = i) + \mathbb{P}(H^A \geq 2 | X_0 = i) + \cdots + \mathbb{P}(H^A \geq n | X_0 = i) \\
&\quad + \sum_{j_1 \in I \setminus A} P_{i,j_1} \cdots \sum_{j_{n-1} \in I \setminus A} P_{j_{n-2}, j_{n-1}} \sum_{j_n \in I \setminus A} P_{j_{n-1}, j_n} y_i \\
&\geq \mathbb{P}(H^A \geq 1 | X_0 = i) + \mathbb{P}(H^A \geq 2 | X_0 = i) + \cdots + \mathbb{P}(H^A \geq n | X_0 = i) \\
&= \sum_{k=1}^{n} \mathbb{P}(H^A \geq k | X_0 = i).
\end{aligned}$$

Taking limits, we have

$$y_i \geq \lim_{n \to \infty} \sum_{k=1}^{n} \mathbb{P}(H^A \geq k | X_0 = i) = \mathbb{E}(H^A | X_0 = i) = k_i^A,$$

and so $y_i \geq k_i^A$ for all $i \in I$, as required. □

*Example* 4.1.7. We continue from Example 3.3.5, letting $A = \{4\} \subset I$.

From Theorem 4.1.6, we have $k_4^A = 0$, and so

$$\begin{aligned}
k_1^A &= 1 + 0.4 k_1^A + 0.2 k_2^A + 0.2 k_3^A, \\
k_2^A &= 1 + k_3^A, \\
k_3^A &= 1 + 0.6 k_2^A.
\end{aligned}$$

Solving these equations gives $k_1^A = 4\frac{2}{3}$, $k_2^A = 5$, and $h_3^A = 4$, so

$$k^A = \begin{pmatrix} \frac{14}{3} & 5 & 4 & 0 \end{pmatrix}.$$

## 4.2 The Gambler's Ruin Problem

Consider a game where £1 is repeatedly wagered on a (possibly biased) coin flip. What is the probability that the player runs out of money? If the player commits to stop playing after making a certain amount of money, what is the probability of this happening? This is the Gambler's Ruin Problem, which we will now explore.

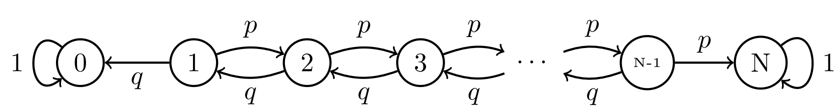
We let $(X_n)$ be the Markov Chain representing the total amount of money the player has at time $n$. For now, we assume the player will happily leave the game when they acquire a wealth of £$N$, and that they must leave when they have £0, so let $I = \{0, 1, \ldots, N\}$.

If $p$ is the probability of flipping a heads (and making £1), $0 < p < 1$, then $(X_n)$ has transition probability matrix $P$, where

$$P_{i,j} = \begin{cases} 1 & \text{if } i = j = 0 \text{ or } i = j = N \\ p & \text{if } j = i + 1 \text{ and } i \neq 0 \\ q = 1 - p & \text{if } j = i - 1 \text{ and } i \neq N \\ 0 & \text{otherwise} \end{cases}$$

for each $i, j \in I$.

This gives the following graph:




First, we calculate the probability of the MC eventually reaching state $N$ or state 0 (from [GGDW14] pg 173-174).

**Theorem 4.2.1.** *Given the MC $(X_n)$ above, we have*

$$\mathbb{P}(X_n = N \text{ for some } n \in \mathbb{N}_0 | X_0 = a) = \begin{cases} \frac{a}{N} & \text{if } p = q = \frac{1}{2} \\ \frac{(q/p)^a - 1}{(q/p)^N - 1} & \text{otherwise.} \end{cases}$$

*Proof.* Let $v(a) = \mathbb{P}(X_n = N \text{ for some } n \in \mathbb{N}_0 | X_0 = a)$ for all $a \in I$.

Since 0 and N are both clearly absorbing states, $v(0) = 0$ and $v(N) = 1$.

Otherwise, by Theorem 4.1.4 we have

$$\begin{aligned} v(a) &= \sum_{j=0}^{N} P_{i,j} v(j) \\ &= v(a+1)p + v(a-1)q. \end{aligned}$$

This is a recurrence relation – rearranging we have $p \cdot v(a+1) - v(a) + q \cdot v(a-1) = 0$, which has characteristic equation $px^2 - x + q = 0$.

By the quadratic formula,

$$\begin{aligned} x &= \frac{1 \pm \sqrt{1 - 4p(1-p)}}{2p} \\ &= \frac{1 \pm \sqrt{(2p-1)^2}}{2p} \end{aligned}$$

$$= 1 \text{ or } \frac{1-p}{p}.$$

If $p = q = \frac{1}{2}$, $\frac{1-p}{p} = \frac{p}{p} = 1$, so we have one solution. This means

$$v(a) = A \cdot 1^a + aB \cdot 1^a = A + aB$$

for some $A, B$.

The initial condition $v(0) = 0$ means that $A = 0$ and $v(N) = 1$ means that $B = \frac{1}{N}$, so we have

$$v(a) = \frac{a}{N},$$

as required.

If $p \neq q$, there are 2 solutions for $x$, so we have

$$v(a) = C \cdot 1^a + D \cdot \left(\frac{1-p}{p}\right)^a = C + D\left(\frac{q}{p}\right)^a$$

for some $C, D$.

The initial conditions $v(0) = 0$ and $v(N) = 1$ mean that

$$\begin{aligned} 0 &= C + D, \\ 1 &= C + D\left(\frac{q}{p}\right)^N, \end{aligned}$$

so $D = \frac{1}{(q/p)^N - 1}$ and $C = -\frac{1}{(q/p)^N - 1}$.

Therefore

$$\begin{aligned} v(a) &= -\frac{1}{(q/p)^N - 1} + \frac{1}{(q/p)^N - 1}\left(\frac{q}{p}\right)^a \\ &= \frac{(q/p)^a - 1}{(q/p)^N - 1}, \end{aligned}$$

as required. □

**Corollary 4.2.2.** *For the same MC we have*

$$\mathbb{P}(X_n = 0 \text{ for some } n \in \mathbb{N}_0 | X_0 = a) = \begin{cases} \frac{N-a}{N} & \text{if } p = q = \frac{1}{2} \\ \frac{(p/q)^{N-a} - 1}{(p/q)^N - 1} & \text{otherwise.} \end{cases}$$

*Proof.* Reaching $N$ starting from $a$ is equivalent to reaching 0 starting from $N - a$ if we swap $p$ and $q$ – which gives the required result. □

**Corollary 4.2.3.** *The MC above will eventually be in either state 0 or $N$ with probability 1.*

*Proof.* For any state $a$, we have

$$\begin{aligned} &\mathbb{P}(X_n = 0 \text{ or } X_n = N \text{ for some } n \in \mathbb{N}_0 | X_0 = a) \\ &= \mathbb{P}(X_n = 0 \text{ for some } n \in \mathbb{N}_0 | X_0 = a) + \mathbb{P}(X_n = N \text{ for some } n \in \mathbb{N}_0 | X_0 = a) \end{aligned}$$

$$= \begin{cases} \frac{N-a}{N} + \frac{a}{N} & \text{if } p = q = \frac{1}{2} \\ \frac{(p/q)^{N-a}-1}{(p/q)^N-1} + \frac{(q/p)^a-1}{(q/p)^N-1} & \text{otherwise} \end{cases}$$
$$= \begin{cases} \frac{N-a+a}{N} & \text{if } p = q = \frac{1}{2} \\ \frac{((p/q)^{N-a}-1)((p/q)^{-N}-1)+((p/q)^{-a}-1)(p/q)^N-1)}{((p/q)^{-N}-1)(p/q)^N-1)} & \text{otherwise} \end{cases}$$
$$= 1,$$

as required. □

Since we know we are guaranteed to eventually reach either 0 or $N$, we calculate the expected time taken to reach one of the absorbing states (from [GGDW14] pg 174-175).

**Theorem 4.2.4.** *Let $T$ be the number of steps until the MC reaches either 0 or $N$. Then*

$$\mathbb{E}(T|X_0 = a) = \begin{cases} a(N-a) & \text{if } p = q = \frac{1}{2} \\ \frac{1}{p-q}\left(N\frac{(q/p)^a-1}{(q/p)^N-1} - a\right) & \text{otherwise.} \end{cases}$$

*Proof.* Let $e(a) = \mathbb{E}(T|X_0 = a)$ for all $a \in I$. Clearly $e(0) = e(N) = 0$.

Otherwise, by Theorem 4.1.6,

$$\begin{aligned} e(a) &= 1 + \sum_{j=1}^{N-1} P_{a,j}e(j) \\ &= 1 + e(a-1)q + e(a+1)p, \end{aligned}$$

so we have $p \cdot e(a+1) - e(a) + q \cdot e(a-1) + 1 = 0$.

This is an inhomogeneous recurrence relation – first we solve the homogeneous form,

$$p \cdot e(a+1) - e(a) + q \cdot e(a-1) = 0.$$

From the proof of Theorem 4.2.1, we have

$$e(a) = \begin{cases} A + aB & \text{if } p = q = \frac{1}{2} \\ C + D(q/p)^a & \text{otherwise} \end{cases}$$

for some constants $A, B, C, D$.

For the case when $p = q = \frac{1}{2}$, we let $e(a) = i_2a^2 + i_1a + i_0$ be a particular solution, for some constants $i_0, i_1, i_2$. Then

$$\begin{aligned} 0 &= p \cdot e(a+1) - e(a) + q \cdot e(a-1) + 1 \\ &= \frac{i_2(a+1)^2 + i_1(a+1) + i_0}{2} - i_2a^2 - i_1a - i_0 + \frac{i_2(a-1)^2 + i_1(a-1) + i_0}{2} + 1 \\ &= i_2a^2 + i_2 + i_1a + i_0 - i_2a^2 - i_1a - i_0 + 1 \\ &= i_2 + 1, \end{aligned}$$

so we have $i_0 = 0, i_1 = 0, i_2 = -1$.

This means $e(a) = A + aB - a^2$. From $e(0) = 0$ we have $A = 0$, and then from $e(N) = 0$ we have $B = N$.

Therefore

$$e(a) = 0 + aN - a^2 = a(N-a),$$

as required.

If $p \neq q$, we let $e(a) = k_1 a + k_0$ be a particular solution, for some constants $k_0, k_1$. Then

$$\begin{aligned}
0 &= p \cdot e(a+1) - e(a) + q \cdot e(a-1) + 1 \\
&= p(k_1(a+1) + k_0) - k_1 a - k_0 + q(k_1(a-1) + k_0) + 1 \\
&= (p-q)k_1 + 1,
\end{aligned}$$

so we have $k_0 = 0, k_1 = \frac{-1}{p-q}$.

This means $e(a) = C + D\left(\frac{q}{p}\right)^a - \frac{a}{p-q}$.

From $e(0) = 0$ and $e(N) = 0$ we have

$$\begin{aligned}
0 &= C + D, \\
0 &= C + D\left(\frac{q}{p}\right)^N - \frac{N}{p-q},
\end{aligned}$$

which means

$$\begin{aligned}
D &= \frac{N}{(p-q)} \cdot \frac{1}{(q/p)^N - 1}, \\
C &= -\frac{N}{(p-q)} \cdot \frac{1}{(q/p)^N - 1}.
\end{aligned}$$

Therefore

$$\begin{aligned}
e(a) &= -\frac{N}{(p-q)} \cdot \frac{1}{(q/p)^N - 1} + \frac{N}{(p-q)} \cdot \frac{1}{(q/p)^N - 1}\left(\frac{q}{p}\right)^a - \frac{a}{p-q} \\
&= \frac{1}{p-q}\left(\frac{-N + N(q/p)^a}{(q/p)^N - 1} - a\right) \\
&= \frac{1}{p-q}\left(N\frac{(q/p)^a - 1}{(q/p)^N - 1} - a\right),
\end{aligned}$$

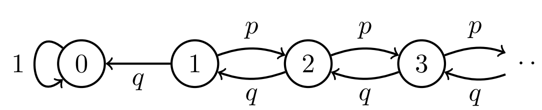
as required. □

Instead of playing until reaching a certain amount, what if they played for as long as possible without going bankrupt? We now calculate the probability the player goes bankrupt in this scenario (from [GGDW14] pg 223). We let $I = \mathbb{N}_0$ and $(X_n)$ be the MC with TPM $P$, where

$$P_{i,j} = \begin{cases} 1 & \text{if } i = j = 0 \\ p & \text{if } j = i+1 \text{ and } i \neq 0 \\ q = 1-p & \text{if } j = i-1 \\ 0 & \text{otherwise} \end{cases}$$

for some $0 < p < 1$ and for each $i, j \in I$.

This gives the following graph:



**Theorem 4.2.5.** *We have that*

$$\mathbb{P}(X_n = 0 \textit{ for some } n \in \mathbb{N}_0 | X_0 = a) = \begin{cases} (q/p)^a & \textit{if } p > q \\ 1 & \textit{if } p \leq q. \end{cases}$$

*Proof.* Let $h(a) = \mathbb{P}(X_n = 0 \text{ for some } n \in \mathbb{N}_0 | X_0 = a)$ for all $a \in \mathbb{N}_0$.

Similarly to the proof of Theorem 4.2.1 (using Theorem 4.1.4), we have $h(0) = 1$ and $h(a) = h(a+1)p + h(a-1)q$, with

$$h(a) = \begin{cases} A + aB & \text{if } p = q = \frac{1}{2} \\ C + (q/p)^a D & \text{otherwise} \end{cases}$$

for some constants $A, B, C, D$.

First consider $p < q$. This means $\frac{q}{p} > 1$, so if $D > 0$ we have that $h(a) > 1$ for some $a$, and if $D < 0$ we have that $h(a) < -1$ for some $a$. So $D = 0$.

From $h(0) = 1$ we have $C = 1$. Then $h(a) = 1$ for all $a \in \mathbb{N}_0$, as required.

Similarly, if $p = q = \frac{1}{2}$, then we must have $B = 0$, and so $A = 1$, and so $h(a) = 1$ for all $a \in \mathbb{N}_0$, as required.

Finally consider $p > q$. Since $h(0) = 1$, we have $0 = C + D$, so

$$h(a) = \left(\frac{q}{p}\right)^a + C\left(1 - \left(\frac{q}{p}\right)^a\right).$$

By Theorem 4.1.4 we choose the value of $C$ that minimises $h(a)$, so we have $C = 0$.

Then $h(a) = (q/p)^a$, as required. □

Note that even when a fair coin is tossed ($p = q$), the player is guaranteed to eventually lose all their money.

## 4.3 Return Times

We now consider the time taken between leaving a state and returning to it, with the following definitions analogous to the definitions in Section 4.1.

**Definition 4.3.1.** The *passage time* to a state $i \in I$ is

$$T_i = \inf\{n \in \mathbb{N} | X_n = i\},$$

where $\inf\{\emptyset\} = \infty$. If $X_0 = i$, then $T_i$ is called the *return time* of $i$ (the time between leaving a state and returning to it).

The main difference between hitting time and passage time is that hitting time from a given state to itself is 0, but the passage time is some other value (at least 1). Comparing Definition 4.3.1 to Definition 4.1.1, we take the infimum of $n$ in $\mathbb{N}$, rather than in $\mathbb{N}_0$.

The event $\{T_i = n\}$ (for some $i \in I$, $n \in \mathbb{N}$) is dependent only on the values of $X_0, \ldots, X_n$, and so by Definition 3.1.1, $T_i$ is a stopping time. From our definition of $T_i$ we can make the following definitions:

**Definition 4.3.2.** The *return probability* to a state $i \in I$ is

$$f_i = \mathbb{P}(T_i < \infty | X_0 = i).$$

We denote $f = (f_i : i \in I)$.

**Definition 4.3.3.** The *expected return time* to a state $i \in I$ is

$$m_i = \mathbb{E}(T_i | X_0 = i).$$

We denote $m = (m_i : i \in I)$.

Clearly these terms relate to the definition of recurrent states (Definition 3.2.1), which leads to the following corollary and stronger form of recurrence:

**Corollary 4.3.4.** *Let $i \in I$. If $i$ is recurrent, then $f_i = 1$, and if $i$ is transient, then $f_i < 1$.*

*Proof.* The result follows from Definitions 3.2.1 and 4.3.2. □

**Definition 4.3.5.** A state $i \in I$ is *positive recurrent* is it is both recurrent and $m_i < \infty$.

*Remark* 4.3.6. If the state space is finite, then clearly any recurrent state must be positive recurrent.

We will determine a method of calculating the expected return time in Section 5.3.

# 5 Invariant and Asymptotic Distributions

## 5.1 Distribution Definitions

Along with the transition probability matrix, our other main tool to analyse Markov Chains are distributions:

**Definition 5.1.1.** Let $\pi = (\pi_i : i \in I)$, where each $\pi_i \geq 0$ and $\sum_{i \in I} \pi_i = 1$. Then $\pi$ is a *distribution* on $I$.

If $I$ is finite we clearly have a well-defined vector. If $I$ is countably infinite we cannot write out the vector $\pi$ in full, but we can define $\pi$ element-wise. In this case we still have for any $j \in I$ that

$$(\pi P)_j = \sum_{i \in I} \pi_i P_{i,j},$$

where $\sum_{i \in I} \pi_i P_{i,j}$ is a convergent series.

The distribution after $n$ steps is $\pi^{(n)}$, where $\pi_i^{(n)} = \mathbb{P}(X_n = i)$ for each $i \in I$. Often with the value of $\pi^{(0)}$ is known and we use the TPM to calculate further values by the following theorem:

**Theorem 5.1.2.** *For a given value of $\pi^{(m)}$, we have that $\pi^{(n+m)} = \pi^{(m)} P^{(n)}$.*

*Proof.* Considering $\pi^{(m)} P^{(n)}$ element-wise we have

$$\begin{aligned}(\pi^{(m)} P^{(n)})_i &= \sum_{j \in I} \pi_j^{(m)} P_{j,i}^{(n)} \\ &= \sum_{j \in I} \mathbb{P}(X_m = j) \mathbb{P}(X_n = i | X_0 = j).\end{aligned}$$

We have assumed all MCs are homogeneous (Remark 2.1.3), and so

$$\begin{aligned}(\pi^{(m)} P^{(n)})_i &= \sum_{j \in I} \mathbb{P}(X_m = j) \mathbb{P}(X_{n+m} = i | X_m = j) \\ &= \mathbb{P}(X_{n+m} = i) \\ &= \pi_i^{(n+m)},\end{aligned}$$

so $\pi^{(n+m)} = \pi^{(m)} P^{(n)}$, as required. □

We now consider two specific types of distribution: invariant distributions and asymptotic distributions.

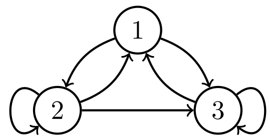
**Definition 5.1.3.** A distribution $\pi$ is an *invariant* distribution for a MC if $\pi P = \pi$.

If $\pi$ is invariant, then $\pi^{(0)} = \pi^{(1)} = \ldots$, so the superscript is often omitted. When the state space is finite (with $|I| = n$ for some $n \in \mathbb{N}$) we can often calculate $\pi$ directly by solving the $n$ simultaneous equations from $\pi P = \pi$ along with $\sum_{i \in I} \pi_i = 1$. We now see an example of this method.

*Example* 5.1.4. We will obtain the invariant distribution of a MC with $I = \{1, 2, 3\}$, TPM

$$P = \begin{pmatrix} 0 & 0.5 & 0.5 \\ 0.5 & 0.5 & 0 \\ 0.3 & 0.3 & 0.4 \end{pmatrix},$$

and graph:




Let $\pi = (\pi_1 \ \pi_2 \ \pi_3)$ be the invariant distribution. Then $\pi = \pi P$, so

$$\pi_1 = 0.5\pi_2 + 0.3\pi_3, \tag{5.1.1}$$
$$\pi_2 = 0.5\pi_1 + 0.5\pi_2 + 0.3\pi_3, \tag{5.1.2}$$
$$\pi_3 = 0.5\pi_1 + 0.4\pi_3. \tag{5.1.3}$$

From (5.1.1) we have $\pi_3 = \frac{5}{6}\pi_1$, and so from (5.1.3) we see that $\pi_2 = \frac{3}{2}\pi_1$.

Since $\pi_1 + \pi_2 + \pi_3 = 1$, we have that $\pi_1 + \frac{3}{2}\pi_1 + \frac{5}{6}\pi_1 = 1$, and so $\pi_1 = 0.3$. This means $\pi_2 = 0.45$ and $\pi_3 = 0.25$ (which aligns with (5.1.2)).

Therefore the invariant distribution is (0.3 0.45 0.25).

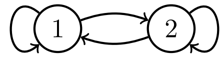
**Definition 5.1.5.** Consider the sequence of distributions $\pi^{(0)}, \pi^{(1)}, \ldots$. If $\pi^{(n)} \to \pi$ as $n \to \infty$ whatever the initial distribution $\pi^{(0)}$, then $\pi$ is an *asymptotic* distribution.

Asymptotic distributions are often difficult to calculate directly. We now look an example when the MC has only two states; we will later see a different method of calculation (Theorem 5.1.7).

*Example* 5.1.6. We will obtain the asymptotic distribution of a MC with $I = \{1, 2\}$, TPM

$$P = \begin{pmatrix} 0.5 & 0.5 \\ 0.25 & 0.75 \end{pmatrix},$$

and graph:




Let $\pi^{(n)} = (\pi_1^{(n)} \quad 1 - \pi_1^{(n)})$. From $\pi^{(n+1)} = \pi^{(n)} P$, we have

$$\begin{aligned} \pi_1^{(n+1)} &= 0.5\pi_1^{(n)} + 0.25(1 - \pi_1^{(n)}) \\ &= 0.25 + 0.25\pi_1^{(n)}. \end{aligned}$$

Letting $\pi^{(n)} = x_n$, we have the inhomogeneous difference equation

$$4x_{n+1} - x_n = 1. \tag{$*$}$$

The homogeneous form is $4x_{n+1} - x_n = 0$, which has auxiliary equation $4y - 1 = 0$, which has solution $y = 0.25$. Therefore the general solution is $x_n = A(0.25)^n$ for some constant $A$.

We now wish to find a particular solution to $4x_{n+1} - x_n = 1$. Consider $x_n = B$, for some constant $B$. Then $4B - B = 1$, so $B = \frac{1}{3}$. Therefore

$$x_n = A\left(\frac{1}{4}\right)^n + \frac{1}{3}$$

is the solution to $(*)$ for some $A$.

This means

$$\pi^{(n)} = \begin{pmatrix} \frac{1}{3} + A\left(\frac{1}{4}\right)^n & \frac{2}{3} - A\left(\frac{1}{4}\right)^n \end{pmatrix},$$

and so as $n \to \infty$ we have asymptotic distribution

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

We relate invariant and asymptotic distributions by the following theorem, with proof from [Kor20] pg 53.

**Theorem 5.1.7.** *If a MC has an asymptotic distribution $\pi$, then $\pi$ is the unique invariant distribution.*

*Proof.* From Theorems 2.1.7 and 5.1.2, we have that $\pi^{(n+1)} = \pi^{(0)}P^{n+1}$ and $\pi^{(n)} = \pi^{(0)}P^n$. This means $\pi^{(n+1)} = \pi^{(n)}P$. As $n \to \infty$, both $\pi^{(n)} \to \pi$ and $\pi^{(n+1)} \to \pi$, so we have $\pi = \pi P$. Thus $\pi$ is an invariant distribution.

To show $\pi$ is unique, let $\pi'$ also be an invariant distribution. This means if $\pi^{(0)} = \pi'$, then $\pi^{(n)} = \pi'$ for all $n$, and so $\pi^{(n)} \to \pi'$ as $n \to \infty$. This means $\pi'$ is also the asymptotic distribution, and so $\pi' = \pi$. □

Invariant distributions are significantly easier to calculate than asymptotic distributions (either by directly solving $\pi = \pi P$ or by using the detailed balance equations in Section 5.2), so this theorem allows us to more easily calculate asymptotic distributions. However, we can only use it if we know the asymptotic distribution exists, which we will determine in Section 5.4.

*Example* 5.1.8. We continue from Example 5.1.6. Given asymptotic distribution $\pi = (\frac{1}{3}\ \frac{2}{3})$, we have

$$\begin{aligned}
\pi P &= \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{6} + \frac{2}{12} & \frac{1}{6} + \frac{6}{12} \end{pmatrix} \\
&= \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \end{pmatrix} \\
&= \pi,
\end{aligned}$$

so $\pi$ is the invariant distribution, as expected.

## 5.2 Detailed Balance

In Example 5.1.4 we calculated an invariant distribution directly, by solving $\pi = \pi P$ for $\pi$. In this section we will use a different method to calculate invariant distributions: detailed balance.

**Definition 5.2.1.** Let $(X_n)$ be a MC with TPM $P$. $(X_n)$ is said to be in *detailed balance* with a distribution $\pi$ if

$$\pi_i P_{i,j} = \pi_j P_{j,i}$$

for all $i, j \in I$.

**Theorem 5.2.2.** *If detailed balance holds with respect to a distribution $\pi$, then $\pi$ is an invariant distribution.*

*Proof.* We use the proof from [GGDW14] pg 241. For any $j \in I$ we have

$$(\pi P)_j = \sum_{i \in I} \pi_i P_{i,j}.$$

Then since detailed balance holds,

$$\begin{aligned}(\pi P)_j &= \sum_{i \in I} \pi_j P_{j,i} \\ &= \pi_j \sum_{i \in I} P_{j,i} \\ &= \pi_j,\end{aligned}$$

and so $\pi P = \pi$. Therefore $\pi$ is invariant, as required. □

We will now use detailed balance to calculate the invariant distribution of a Markov Chain with a countably infinite state space.

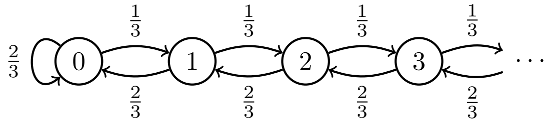
*Example* 5.2.3 (Reflected Random Walk). Let $(X_n)$ be a MC on state space $I = \mathbb{N}_0$. We consider a random walk where for any $i \in I \backslash \{0\}$, there is a 1/3 chance of moving to state $i + 1$ and a 2/3 chance of moving to state $i - 1$. (At state 0, there is a 1/3 chance of moving to state 1 and a 2/3 chance of remaining at state 0.)

This means $(X_n)$ has transition probability matrix $P$, where

$$P_{i,j} = \begin{cases} 1/3 & \text{if } j = i + 1 \\ 2/3 & \text{if } j = i - 1 \text{ or } i = j = 0 \\ 0 & \text{otherwise} \end{cases}$$

for each $i, j \in I$.

This gives the following graph:




We wish to calculate an invariant distribution $\pi = (\pi_0 \ \pi_1 \ \pi_2 \ \cdots)$. From the detailed balance equation we have for any $i \in I$

$$\pi_i P_{i,i+1} = \pi_{i+1} P_{i+1,i},$$

and so

$$\pi_{i+1} = \frac{P_{i,i+1}}{P_{i+1,i}}\pi_i = \frac{1/3}{2/3}\pi_i = \frac{1}{2}\pi_i.$$

Therefore (by iteration) we have

$$\pi_i = \frac{\pi_0}{2^i}.$$

This means

$$1 = \sum_{i=0}^{\infty} \frac{\pi_0}{2^i} = \pi_0 \sum_{i=0}^{\infty} \frac{1}{2^i} = 2\pi_0,$$

and so $\pi_0 = 1/2$.

Therefore the invariant distribution is

$$\pi = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{8} & \cdots \end{pmatrix}.$$

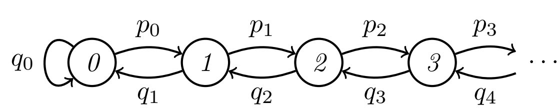
We now consider a more general case from [GGDW14] pg 242, where the probability of moving from a given state is potentially different for different states.

**Theorem 5.2.4** (Reflected Birth-Death Chain)**.** *Let $(X_n)$ be a MC on state space $I = \mathbb{N}_0$, where for each state $i \in I\backslash\{0\}$ there is a $p_i$ chance of moving to state $i+1$ and a $q_i = 1 - p_i$ chance of moving to state $i-1$. (At state 0, there is a $p_0$ chance of moving to state 1 and a $q_0 = 1 - p_0$ chance of remaining at state 0.) We assume that $0 < p_i < 1$ for all $i \in \mathbb{N}_0$.*

*This means $(X_n)$ has transition probability matrix $P$, where*

$$P_{i,j} = \begin{cases} p_i & \text{if } j = i+1 \\ q_i & \text{if } j = i-1 \text{ or } i = j = 0 \\ 0 & \text{otherwise} \end{cases}$$

*for each $i, j \in I$, and graph:*




*Then $(X_n)$ has an invariant distribution if and only if*

$$\sum_{i=1}^{\infty} \frac{p_0 p_1 \ldots p_{i-1}}{q_1 q_2 \ldots q_i} < \infty.$$

*Proof.* We wish to calculate an invariant distribution $\pi = (\pi_0\ \pi_1\ \pi_2\ \cdots)$. From the detailed balance equation we have for any $i \in I$

$$\pi_i P_{i,i+1} = \pi_{i+1} P_{i+1,i},$$

and so

$$\pi_{i+1} = \frac{P_{i,i+1}}{P_{i+1,i}}\pi_i = \frac{p_i}{q_{i+1}}\pi_i.$$

Therefore we have

$$\pi_1 = \frac{p_0}{q_1}\pi_0,$$
$$\pi_2 = \frac{p_1}{q_2}\pi_1 = \frac{p_0 p_1}{q_1 q_2}\pi_0,$$
$$\pi_3 = \frac{p_2}{q_3}\pi_2 = \frac{p_0 p_1 p_2}{q_1 q_2 q_3}\pi_0,$$

and so for any $i \in \mathbb{N}$,

$$\pi_i = \frac{p_0 p_1 \dots p_{i-1}}{q_1 q_2 \dots q_i}\pi_0.$$

We must have $\sum_{i\in I} \pi_i = 1$, so

$$\pi_0 + \sum_{i=1}^{\infty} \frac{p_0 p_1 \dots p_{i-1}}{q_1 q_2 \dots q_i}\pi_0 = 1.$$

Therefore we have an invariant distribution if

$$\sum_{i=1}^{\infty} \frac{p_0 p_1 \dots p_{i-1}}{q_1 q_2 \dots q_i} < \infty,$$

as required. □

## 5.3 Existence of Invariant Distributions

In this section we wish to know under what conditions invariant distributions exist, and how to use invariant distributions to calculate expected return times – the relationship shown in Theorems 5.3.6 and 5.3.7. The proofs of these theorems (from [Nor97] across pg 37-38) use the vector $\gamma^k$, which is defined for some state $k \in I$ by

$$\gamma_i^k = \mathbb{E}\left(\sum_{n=0}^{T_k-1} \mathbf{1}_{X_n=i} \middle| X_k = 0\right)$$

for all $i \in I$. This means $\gamma_i^k$ is the expected time spent in $i$ between the MC leaving $k$ and first returning to $k$.

*Remark* 5.3.1. We clearly see that $\gamma_k^k = 1$.

*Remark* 5.3.2. For some $i \in I$, $\gamma_i^k$ is the expected amount of time spent in $i$ between the MC leaving $k$ and first returning to $k$ – this means that

$$\sum_{i\in I} \gamma_i^k = \mathbb{E}(T_k | k = 0) = m_k.$$

Before we can prove Theorems 5.3.6 and 5.3.7, we prove the following three lemmas (from [Nor97] across pg 35-37):

**Lemma 5.3.3.** *Let $(X_n)$ be an irreducible MC where every state is recurrent and let $k \in I$. Then $\gamma^k$ satisfies $\gamma^k = \gamma^k P$.*

*Proof.* Let $n \in \mathbb{N}$. Then the event $\{n \leq T_k\}$ is dependent only on $X_0, \ldots, X_{n-1}$, so by the Markov Property we have

$$\mathbb{P}(X_n = j | X_{n-1} = i, n \leq T_k, X_0 = k) = \mathbb{P}(X_n = j | X_{n-1} = i) \tag{5.3.1}$$

for some $i, j \in I$.

By the law of total probability we have

$$\mathbb{P}(X_n = j, n \leq T_k | X_0 = k) = \sum_{i \in I} \mathbb{P}(X_n = j, X_{n-1} = i, n \leq T | X_0 = k),$$

and then by Bayes' theorem

$$\mathbb{P}(X_n = j, n \leq T_k | X_0 = k) = \sum_{i \in I} \mathbb{P}(X_n = j | X_{n-1} = i, X_0 = k, n \leq T_k) \mathbb{P}(X_{n-1} = i, n \leq T_k | X_0 = k).$$

This means by (5.3.1) we have

$$\begin{aligned}
\mathbb{P}(X_n = j, n \leq T_k | X_0 = k) &= \sum_{i \in I} \mathbb{P}(X_n = j | X_{n-1} = i) \mathbb{P}(X_{n-1} = i, n \leq T_k | X_0 = k) \\
&= \sum_{i \in I} P_{i,j} \mathbb{P}(X_{n-1} = i, n \leq T_k | X_0 = k) \qquad (5.3.2)
\end{aligned}$$

for some $j \in I$.

Therefore, using the fact that $X_0 = X_{T_k}$ to change summation, for any $j \in I$,

$$\begin{aligned}
\gamma_j^k &= \mathbb{E}\left( \sum_{n=0}^{T_k - 1} \mathbf{1}_{X_n = j} \middle| X_k = 0 \right) \\
&= \mathbb{E}\left( \sum_{n=1}^{T_k} \mathbf{1}_{X_n = j} \middle| X_k = 0 \right) \\
&= \mathbb{E}\left( \sum_{n=1}^{\infty} \mathbf{1}_{X_n = j, n \leq T_k} \middle| X_k = 0 \right) \\
&= \sum_{n=1}^{\infty} \mathbb{P}(X_n = j, n \leq T_k | X_k = 0),
\end{aligned}$$

and then by (5.3.2),

$$\begin{aligned}
\gamma_j^k &= \sum_{n=1}^{\infty} \sum_{i \in I} P_{i,j} \mathbb{P}(X_{n-1} = i, n \leq T_k | X_0 = k) \\
&= \sum_{i \in I} P_{i,j} \sum_{n=1}^{\infty} \mathbb{P}(X_{n-1} = i, n \leq T_k | X_0 = k).
\end{aligned}$$

By changing the lower bound of summation, we have

$$\begin{aligned}
\gamma_j^k &= \sum_{i \in I} P_{i,j} \sum_{m=0}^{\infty} \mathbb{P}(X_m = i, m \leq T_k - 1 | X_0 = k) \\
&= \sum_{i \in I} P_{i,j} \mathbb{E}\left( \sum_{m=0}^{\infty} \mathbf{1}_{X_m = i, m \leq T_k - 1} \middle| X_k = 0 \right)
\end{aligned}$$

$$
\begin{aligned}
&= \sum_{i \in I} P_{i,j} \mathbb{E}\left( \sum_{m=0}^{T_k - 1} \mathbf{1}_{X_m = i} \middle| X_k = 0 \right) \\
&= \sum_{i \in I} P_{i,j} \gamma_i^k \\
&= (\gamma^k P)_j,
\end{aligned}
$$

which means $\gamma^k = \gamma^k P$, as required. □

**Lemma 5.3.4.** *Let $(X_n)$ be an irreducible MC, let $k \in I$, and let $\lambda$ be a vector where $\lambda_k = 1$ and $\lambda = \lambda P$. Then $\lambda \geq \gamma^k$.*

*Proof.* Since $\lambda = \lambda P$, for some $j \in I \backslash \{k\}$ and $n \in \mathbb{N}$ we have

$$
\begin{aligned}
\lambda_j &= \sum_{i_1 \in I} \lambda_{i_1} P_{i_1,j} \\
&= P_{k,j} + \sum_{i_1 \in I \backslash \{k\}} \lambda_{i_1} P_{i_1,j},
\end{aligned}
$$

since $\lambda_k = 1$. Then

$$
\begin{aligned}
\lambda_j &= P_{k,j} + \sum_{i_1 \in I \backslash \{k\}} \left( \sum_{i_2 \in I} \lambda_{i_2} P_{i_2,i_1} \right) P_{i_1,j} \\
&= P_{k,j} + \sum_{i_1 \in I \backslash \{k\}, i_2 \in I} \lambda_{i_2} P_{i_2,i_1} P_{i_1,j} \\
&= P_{k,j} + \sum_{i_1 \in I \backslash \{k\}} P_{k,i_1} P_{i_1,j} + \sum_{i_1 \in I \backslash \{k\}, i_2 \in I \backslash \{k\}} \lambda_{i_2} P_{i_2,i_1} P_{i_1,j} \\
&= \mathbb{P}(X_1 = j, T_k \geq 1 | X_0 = k) + \mathbb{P}(X_2 = j, T_k \geq 2 | X_0 = k) + \cdots + \sum_{i_1 \in I \backslash \{k\}, i_2 \in I \backslash \{k\}} \lambda_{i_2} P_{i_2,i_1} P_{i_1,j}.
\end{aligned}
$$

After $n$ steps, we have

$$
\begin{aligned}
\lambda_j &= \mathbb{P}(X_1 = j, T_k \geq 1 | X_0 = k) + \mathbb{P}(X_2 = j, T_k \geq 2 | X_0 = k) + \cdots + \mathbb{P}(X_n = j, T_k \geq n | X_0 = k) \\
&\quad + \sum_{i_1, i_2, \ldots, i_n \in I \backslash \{k\}} \lambda_{i_n} P_{i_n, i_{n-1}} \cdots P_{i_2,i_1} P_{i_1,j} \\
&\geq \mathbb{P}(X_1 = j, T_k \geq 1 | X_0 = k) + \mathbb{P}(X_2 = j, T_k \geq 2 | X_0 = k) + \cdots + \mathbb{P}(X_n = j, T_k \geq n | X_0 = k) \\
&= \mathbb{E}(\mathbf{1}_{X_1 = j, T_k \geq 1} | X_0 = k) + \mathbb{E}(\mathbf{1}_{X_2 = j, T_k \geq 2} | X_0 = k) + \cdots + \mathbb{E}(\mathbf{1}_{X_n = j, T_k \geq n} | X_0 = k).
\end{aligned}
$$

As $n \to \infty$,

$$
\mathbb{E}(\mathbf{1}_{X_1 = j, T_k \geq 1} | X_0 = k) + \mathbb{E}(\mathbf{1}_{X_2 = j, T_k \geq 2} | X_0 = k) + \cdots + \mathbb{E}(\mathbf{1}_{X_n = j, T_k \geq n} | X_0 = k) \to \gamma_j^k,
$$

so $\lambda_j \geq \gamma_j^k$ for any $j \neq k$.

Clearly $\lambda_k = 1 = \gamma_k^k$. Therefore $\lambda \geq \gamma^k$, as required. □

**Lemma 5.3.5.** *Let $(X_n)$ be an irreducible MC where every state is recurrent, let $k \in I$, and let $\lambda$ be a vector where $\lambda_k = 1$ and $\lambda = \lambda P$. Then $\lambda = \gamma^k$.*

*Proof.* From Lemma 5.3.4 we see $\lambda \geq \gamma^k$, so let $\mu = \lambda - \gamma^k \geq 0$. Note that

$$\mu_k = \lambda_k - \gamma_k^k = 1 - 1 = 0.$$

Since every state is recurrent, $\gamma^k = \gamma^k P$ (by Lemma 5.3.3). This means $\mu = \mu P$.

$(X_n)$ is irreducible, so for any $j \in I$, $j \to k$. This means there exists $n \in \mathbb{N}$ such that $P_{j,k}^{(n)} > 0$. Therefore

$$0 = \mu_k = \sum_{i \in I} \mu_i P_{i,k}^{(n)} \geq \mu_j P_{j,k}^{(n)}.$$

Since $P_{j,k}^{(n)} > 0$ and $\mu_j \geq 0$, we have $\mu = 0$, which means $\lambda = \gamma^k$, as required. □

This means we can now prove the following two theorems:

**Theorem 5.3.6.** *Let $(X_n)$ be an irreducible MC. If some state is positive recurrent, then $(X_n)$ has an invariant distribution.*

*Proof.* Let $k$ be a positive recurrent state. Since $k$ is recurrent and $(X_n)$ is irreducible, every state is recurrent by Theorem 3.2.7. Then by Lemma 5.3.3, $\gamma^k = \gamma^k P$. Note that by Remark 5.3.2, we have $\sum_{i \in I} \gamma_i^k = m_k$

Consider $\pi$, where $\pi_i = \gamma_i^k / m_k$ for each $i \in I$. (We may divide by $m_k$ because we know it is finite, since $k$ is positive recurrent.) Then

$$\sum_{i \in I} \pi_i = \frac{1}{m_k} \sum_{i \in I} \gamma_i^k = \frac{m_k}{m_k} = 1,$$

and so $\pi$ is a distribution.

Since $\gamma_i^k = \gamma_i^k P$ for all $i \in I$, dividing by $m_k$ we have that $\pi_i = \pi_i P$ for all $i \in I$. This means $\pi$ is an invariant distribution. □

**Theorem 5.3.7.** *Let $(X_n)$ be an irreducible MC. If the invariant distribution $\pi$ exists, then every state is positive recurrent, and*

$$\pi_i = \frac{1}{m_i}$$

*for all $i \in I$.*

*Proof.* Let $k \in I$ be any state, and let $\pi$ be an invariant distribution.

We know that $\sum_{i \in I} \pi_i = 1$, so $\pi_j > 0$ for some $j \in I$. Since $(X_n)$ is irreducible, $j \to k$, so $P_{j,k}^{(n)} > 0$ for some $n \in \mathbb{N}$. Therefore

$$\pi_k = \sum_{i \in I} \pi_i P_{i,k}^{(n)} \geq \pi_j P_{j,k}^{(n)} > 0.$$

This means we may define the vector $\lambda$ by $\lambda = \pi / \pi_k$. This means $\lambda_k = \pi_k / \pi_k = 1$ and

$$\lambda P = \frac{\pi P}{\pi_k} = \frac{\pi}{\pi_k} = \lambda,$$

so by Lemma 5.3.4, $\lambda \geq \gamma^k$.

By Remark 5.3.2, we have

$$m_k = \sum_{i \in I} \gamma_i^k \leq \sum_{i \in I} \lambda_i = \sum_{i \in I} \frac{\pi_i}{\pi_k} = \frac{\sum_{i \in I} \pi_i}{\pi_k} = \frac{1}{\pi_k}.$$

We know $\pi_k > 0$, so $m_k$ is finite. This means every state $k$ is positive recurrent, and so by Lemma 5.3.5 $\lambda = \gamma^k$. Therefore

$$m_k = \sum_{i \in I} \gamma_i^k = \sum_{i \in I} \lambda_i = \sum_{i \in I} \frac{\pi_i}{\pi_k} = \frac{\sum_{i \in I} \pi_i}{\pi_k} = \frac{1}{\pi_k}.$$

This means the invariant distribution $\pi$ is defined by $\pi_k = 1/m_k$ for all $k \in I$, as required. □

Therefore we can calculate expected return times if we know the invariant distribution, and leads to the following corollary:

**Corollary 5.3.8.** *Let $(X_n)$ be an irreducible MC. Then $(X_n)$ has an invariant distribution ($\pi$ where $\pi_i = \frac{1}{m_i}$ for all $i \in I$) if and only if some state is positive recurrent (if and only if every state is positive recurrent).*

*Proof.* The result follows from Theorems 5.3.6 and 5.3.7. □

*Example* 5.3.9. We continue from Example 5.1.4 – we previously found the invariant distribution to be $\pi = (0.3\ 0.45\ 0.25)$. Clearly the MC is irreducible, so by Theorem 5.3.7 we have

$$m = \begin{pmatrix} \frac{10}{3} & \frac{20}{9} & 4 \end{pmatrix}.$$

## 5.4 Existence of Asymptotic Distributions

Similarly to the previous section, we will now investigate under what conditions asymptotic distributions exist. This will be explored in two theorems: 5.4.4 and 5.4.7. Both theorems require the Markov Chain to be irreducible, every state to be aperiodic, and every state to be positive recurrent. Theorem 5.4.4 includes Markov Chains with countably infinite state spaces, while Theorem 5.4.7 requires $I$ to be finite, but shows convergence to the asymptotic distribution geometrically.

Before the theorems can proved, we need the following three lemmas:

**Lemma 5.4.1.** *Let $i \in I$ be an aperiodic state. Then there exists $N \in \mathbb{N}$ such that $P_{i,i}^n > 0$ for all $n \geq N$.*

*Proof.* We know that $\gcd\{n \geq 1 | P_{i,i}^n > 0\} = 1$, so consider $n_1, n_2, \ldots$, where $P_{i,i}^{n_j} > 0$ for each $n_j$. Then there exists $P \in \mathbb{N}$ such that $\gcd\{n_1, \ldots, n_P\} = 1$ (or else $\gcd\{n_1, n_2, \ldots\} > 1$). So from number theory, there exists $l_1, \ldots, l_P \in \mathbb{N}_0$ and $N \in \mathbb{N}$ such that for any $n \geq N$,

$$n = l_1 n_1 + l_2 n_2 + \cdots + l_P n_P.$$

This means we have

$$P_{i,i}^n \geq (P_{i,i}^{n_1})^{l_1} (P_{i,i}^{n_2})^{l_2} \ldots (P_{i,i}^{n_P})^{l_P} > 0$$

for all $n \geq N$. □

**Lemma 5.4.2.** *Let $(X_n)$ be an irreducible MC where every state is aperiodic. Then for any $j, k \in I$, there exists $M \in \mathbb{N}$ such that $P_{j,k}^m > 0$ for all $m \geq M$.*

*Proof.* We use the proof from [Nor97] pg 41. From Lemma 5.4.1, we have $i \in I$ and $N \in \mathbb{N}$ such that $P^n_{i,i} > 0$ for all $n \geq N$. Also, since $(X_n)$ is irreducible, we have $r, s \in \mathbb{N}$ such that $P^r_{j,i}, P^s_{i,k} > 0$. Then

$$P^{r+n+s}_{j,k} \geq P^r_{j,i} P^n_{i,i} P^s_{i,k} > 0$$

for all $n \geq N$. So $P^m_{j,k} > 0$ for all $m \geq M = r + N + s$. □

**Lemma 5.4.3.** *Let $(X_n)$ be an irreducible MC where every state is recurrent. Then*

$$\mathbb{P}(T_j < \infty) = 1$$

*for all $j \in I$.*

*Proof.* We use the proof from [Nor97] pg 28. Since $(X_n)$ is irreducible, for any $i, j \in I$ we have some $m \in \mathbb{N}$ such that $P^{(m)}_{j,i} > 0$. Also, since every state is recurrent, by Remark 3.2.2 we have that

$$\mathbb{P}(X_n = j \text{ for infinitely many } n | X_0 = j) = 1.$$

This means

$$\begin{aligned} 1 &= \mathbb{P}(X_n = j \text{ for some } n \geq m+1 | X_0 = j) \\ &= \sum_{i \in I} \mathbb{P}(X_n = j \text{ for some } n \geq m+1 | X_m = i, X_0 = j) \mathbb{P}(X_m = i | X_0 = j) \\ &= \sum_{i \in I} \mathbb{P}(T_j < \infty | X_0 = i) P^{(m)}_{j,i}. \end{aligned}$$

But we know $\sum_{i \in I} P^{(m)}_{j,i} = 1$, so we must have $\mathbb{P}(T_j < \infty | X_0 = i) = 1$. Since this is true for all $i \in I$ we have that $\mathbb{P}(T_j < \infty) = 1$, as required. □

Now we can prove the first of the theorems, with proof based on [Nor97] pg 41-42:

**Theorem 5.4.4.** *Let $(X_n)$ be an irreducible MC where every state is aperiodic and positive recurrent. Then $(X_n)$ has an asymptotic distribution.*

*Proof.* Let $(X_n)$ have initial distribution $\lambda^{(0)}$ and TPM $P$. By Theorem 5.3.6 $(X_n)$ has an invariant distribution $\pi$. Let $(Y_n)$ be a MC with initial distribution $\pi$ and TPM $P$, independent of $(X_n)$.

We now construct a sequence of ordered pairs of random variables $(W_n)$, where $W_n = (X_n, Y_n)$ for all $n \in \mathbb{N}_0$. Clearly $(W_n)$ has state space $I \times I$ For any $h_0 = (i_0, j_0), h_1 = (i_1, j_1), \ldots, h_n = (i_n, j_n) \in I \times I$, $n \in \mathbb{N}$, we have

$$\begin{aligned} &\mathbb{P}(W_n = h_n | W_{n-1} = h_{n-1}, \ldots, W_0 = h_0) \\ &= \mathbb{P}(X_n = i_n, Y_n = j_n | X_{n-1} = i_{n-1}, Y_{n-1} = j_{n-1}, \ldots, X_0 = i_0, Y_0 = j_0) \\ &= \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) \mathbb{P}(Y_n = j_n | Y_{n-1} = j_{n-1}, \ldots, Y_0 = j_0), \end{aligned}$$

and so by the Markov Property (Definition 2.1.1),

$$\begin{aligned} \mathbb{P}(W_n = h_n | W_{n-1} = h_{n-1}, \ldots, W_0 = h_0) &= \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}) \mathbb{P}(Y_n = j_n | Y_{n-1} = j_{n-1}) \\ &= \mathbb{P}(X_n = i_n, Y_n = j_n | X_{n-1} = i_{n-1}, Y_{n-1} = j_{n-1}) \\ &= \mathbb{P}(W_n = h_n | W_{n-1} = h_{n-1}), \end{aligned}$$

and so $(W_n)$ is a Markov Chain.

From the above equations we see that $(W_n)$ has transition probability matrix $Q$, where

$$Q_{(i,k),(j,l)} = P_{i,j}P_{k,l}$$

for all $i, j, k, l \in I$. In addition, $(W_n)$ has initial distribution $\mu^{(0)}$, where

$$\mu^{(0)}_{i,k} = \lambda^{(0)}_i \pi_k$$

for all $i, k \in I$.

Let $i, j, k, l \in I$. Since $(X_n)$ is irreducible and every state is aperiodic, by Lemma 5.4.2 there exists $N, M \in \mathbb{N}$ such that $P^n_{i,j}, P^m_{k,l} > 0$ for all $n \geq N$, $m \geq M$. This means $P^r_{i,j}, P^r_{k,l} > 0$ for all $r \geq R = \max\{N, M\}$. Therefore

$$Q^r_{(i,k),(j,l)} = P^r_{i,j}P^r_{k,l} > 0$$

for all $r \geq R$. This means $(i, k) \to (j, l)$, so $(W_n)$ is irreducible.

Clearly $(W_n)$ has an invariant distribution $\nu$, where $\nu_{i,k} = \pi_i \pi_k$ for all $i, k \in I$. Therefore by Theorem 5.3.7 every state of $(W_n)$ is positive recurrent.

Let $b \in I$ and consider $T_{(b,b)}$ w.r.t. $(W_n)$, the first passage time to $(b, b)$. This means

$$T_{(b,b)} = \inf\{n \in \mathbb{N} | W_n = (b, b)\} = \inf\{n \in \mathbb{N} | X_n = Y_n = b\}.$$

Now consider the sequences of random variables $(A_n)$, $(B_n)$, $(C_n)$ where for all $n \in \mathbb{N}_0$

$$\begin{aligned} A_n &= \begin{cases} X_n & \text{if } n < T_{(b,b)}, \\ Y_n & \text{if } n \geq T_{(b,b)}, \end{cases} \\ B_n &= \begin{cases} Y_n & \text{if } n < T_{(b,b)}, \\ X_n & \text{if } n \geq T_{(b,b)}, \end{cases} \\ C_n &= (A_n, B_n). \end{aligned}$$

Applying the Strong Markov Property – Theorem 3.1.2 – to $(W_n)$ at time $T_{(b,b)}$, we see that $((X_{T_{(b,b)}+n}, Y_{T_{(b,b)}+n}))$ is a MC with TPM $Q$, initial distribution $\mu^{(T_{(b,b)})}$, and is independent of $(X_0, Y_0), (X_1, Y_1), \ldots, (X_{T-1}, Y_{T-1})$.

By symmetry, $((Y_{T_{(b,b)}+n}, X_{T_{(b,b)}+n}))$ is also a MC with TPM $Q$, initial distribution $\mu^{(T_{(b,b)})}$, and is independent of $(X_0, Y_0), (X_1, Y_1), \ldots, (X_{T-1}, Y_{T-1})$. Therefore $(C_n)$ is a MC with TPM $Q$ and initial distribution $\mu^{(0)}$, and so $(A_n)$ is a MC with TPM $P$ and initial distribution $\lambda^{(0)}$.

We then have that for some $n \in \mathbb{N}$, $i \in I$,

$$\begin{aligned} |\lambda^{(n)}_i - \pi_i| &= |\mathbb{P}(A_n = i) - \mathbb{P}(Y_n = i)| \\ &= |\mathbb{P}(A_n = i, n < T_{(b,b)}) + \mathbb{P}(A_n = i, n \geq T_{(b,b)}) - \mathbb{P}(Y_n = i)| \\ &= |\mathbb{P}(X_n = i, n < T_{(b,b)}) + \mathbb{P}(Y_n = i, n \geq T_{(b,b)}) - \mathbb{P}(Y_n = i)| \\ &= |\mathbb{P}(X_n = i, n < T_{(b,b)}) - \mathbb{P}(Y_n = i, n < T_{(b,b)})|, \end{aligned}$$

and then by Bayes' theorem,

$$\begin{aligned} |\lambda^{(n)}_i - \pi_i| &= |\mathbb{P}(X_n = i | n < T_{(b,b)})\mathbb{P}(n < T_{(b,b)}) - \mathbb{P}(Y_n = i | n < T_{(b,b)})\mathbb{P}(n < T_{(b,b)})| \\ &= \mathbb{P}(n < T_{(b,b)})|\mathbb{P}(X_n = i | n < T_{(b,b)}) - \mathbb{P}(Y_n = i | n < T_{(b,b)})| \\ &\leq \mathbb{P}(n < T_{(b,b)}) \\ &= 1 - \mathbb{P}(T_{(b,b)} \leq n). \end{aligned}$$

Taking limits, we have

$$\begin{aligned}\lim_{n\to\infty}|\lambda_i^{(n)}-\pi_i| &\le \lim_{n\to\infty}(1-\mathbb{P}(T_{(b,b)}\le n))\\ &= 1-\mathbb{P}(T_{(b,b)}<\infty).\end{aligned}$$

Since every state of $(W_n)$ is recurrent, by Lemma 5.4.3 we have that $\mathbb{P}(T_{(b,b)}<\infty)=1$, and so

$$\lim_{n\to\infty}|\lambda_i^{(n)}-\pi_i|\le 0.$$

Therefore regardless of the initial distribution $\lambda^{(0)}$, $\lambda^{(n)}\to\pi$ as $n\to\infty$, and so $\pi$ is the asymptotic distribution of $(X_n)$, as required. □

*Example* 5.4.5. Consider the MC from Example 5.2.3. Clearly every state can communicate with every other state, so the MC is irreducible. State 0 is aperiodic, so by Theorem 3.3.4 every state is aperiodic. Since we also found an invariant distribution $\pi$, by Theorem 5.4.4 $\pi$ is also the asymptotic distribution of the MC.

This example shows that if an invariant distribution exists (when all states are positive recurrent), there exists an asymptotic distribution, even when the state space is infinite. We will now look at a different proof (based on [JGKS76] pg 70-71) that only applies when the state space is finite (after the following lemma, which is from [JGKS76] pg 69-70).

**Lemma 5.4.6.** *Let $(X_n)$ be a MC with a finite number of states and TPM $P$, let $\varepsilon>0$, and let $P_{i,j}\ge\varepsilon$ for all $i,j\in I$. Let $x$ be a column vector of length $|I|$, let $M_0$ and $m_0$ be the maximum and minimum elements of $x$ respectively, and let $M_1$ and $m_1$ be the maximum and minimum elements of $Px$ respectively.*

*Then $M_1\le M_0$, $m_1\ge m_0$, and $M_1-m_1\le(1-2\varepsilon)(M_0-m_0)$.*

*Proof.* Let $y$ be the vector formed by replacing every element of $x$ by with $M_0$ (apart from one element $m_0$). Clearly $x\le y$.

Let $k\in I$ such that $y_k=m_0$. Then for any $i\in I$ we have

$$\begin{aligned}(Px)_i &\le (Py)_i\\ &= \sum_{j\in I}P_{i,j}y_j\\ &= P_{i,k}m_0+\sum_{j\in I\setminus\{k\}}P_{i,j}M_0\\ &= P_{i,k}m_0+(1-P_{i,k})M_0\\ &= M_0-P_{i,k}(M_0-m_0).\end{aligned}$$

Since $P_{i,k}\ge\varepsilon$ for all $i\in I$, we have

$$(Px)_i\le M_0-\varepsilon(M_0-m_0).$$

Choosing the $i\in I$ to give the maximum value of $(Px)_i$, we have

$$M_1\le M_0-\varepsilon(M_0-m_0), \tag{5.4.1}$$

and so $M_1\le M_0$.

Let $z$ be the vector formed by replacing every element of $x$ by with $m_0$ (apart from one element $M_0$). Clearly $x\ge z$, so $-x\le -z$. Therefore using a similar method to above we have

$$-m_1\le -m_0-\varepsilon(-m_0+M_0), \tag{5.4.2}$$

which means

$$m_1 \geq m_0 + \varepsilon(M_0 - m_0),$$

and so $m_1 \geq m_0$.

By adding (5.4.1) and (5.4.2) we have

$$M_1 - m_1 \leq M_0 - m_0 - 2\varepsilon(M_0 - m_0) = (1 - 2\varepsilon)(M_0 - m_0),$$

as required. □

**Theorem 5.4.7.** *Let $(X_n)$ be an irreducible MC with a finite number of states, where every state is aperiodic. Then $(X_n)$ has an asymptotic distribution (that it converges to geometrically).*

*Proof.* By Lemma 5.4.2, for each $j, k \in I$ there exists $n_{j,k} \in \mathbb{N}$ such that $P^n_{j,k} > 0$ for all $n \geq n_{j,k}$. Letting $N = \max_{j,k} n_{j,k}$, we have that for $P^n_{j,k} > 0$ for all $n \geq N, j, k \in I$.

Choose some $m \geq N$ and consider a new MC $(Y_n)$ with transition probability matrix $Q = P^m$ and $Y_0 = X_0$. This means $Y_1 = X_m, Y_2 = X_{2m}, Y_3 = X_{3m}, \ldots$.

Since $P^m_{j,k} > 0$ for all $j, k \in I$, we have $Q_{j,k} > 0$ for all $j, k \in I$ – so $(Y_n)$ can move to any other state at any transition. $I$ is a finite set, so there exists $\varepsilon > 0$ such that $Q_{j,k} \geq \varepsilon$ for all $j, k \in I$.

Let $\rho$ be a column vector of length $|I|$, where $\rho_j = 1$ for some $j \in I$ and $\rho_i = 0$ for all $i \in I \backslash \{j\}$. For any $n \in \mathbb{N}_0$, let the maximum and minimum elements of $Q^n \rho$ be $M_n$ and $m_n$ respectively.

Since $Q^n \rho = Q(Q^{n-1} \rho)$, by Lemma 5.4.6 we have that $M_n \leq M_{n-1}$, $m_n \geq m_{n-1}$, and

$$M_n - m_n \leq (1 - 2\varepsilon)(M_{n-1} - m_{n-1}).$$

Therefore by iteration we have

$$M_n - m_n \leq (1 - 2\varepsilon)^n (M_0 - m_0).$$

We must have $\varepsilon < 0.5$, or else each row of $Q$ would sum to more than 1. This means $1 - 2\varepsilon < 1$, and so $M_n - m_n \to 0$ as $n \to \infty$. This means $M_n$ and $m_n$ approach the same value – let this value be $a_j$; we see that $0 < m_n \leq a_j \leq M_n < 1$ for all $n \in \mathbb{N}$.

Since $Q^n \rho$ is the $j$-th column of $Q^n$, we see that every entry in the $j$-th column of $Q^n$ converges to $a_j$; $Q^n_{i,j} \to a_j$ as $n \to \infty$ for all $i \in I$. Also, we have that $\sum_{j \in I} a_j = 1$ (this holds for rows of $Q^n$, so must hold for the limit).

Therefore if $(Y_n)$ has initial distribution $\pi^{(0)}$ – the same as $(X_n)$, since $X_0 = Y_0$ – we have

$$\begin{aligned} \lim_{n\to\infty} \pi^{(n)}_j &= \lim_{n\to\infty} (\pi^{(0)} Q^n)_j \\ &= \lim_{n\to\infty} \sum_{i \in I} \pi^{(0)}_i Q^n_{i,j} \\ &= \sum_{i \in I} \pi^{(0)}_i a_j \\ &= a_j \sum_{i \in I} \pi^{(0)}_i \\ &= a_j, \end{aligned}$$

which means $\pi^{(n)} \to (a_j : j \in I)$ as $n \to \infty$, regardless of the choice of $\pi^{(0)}$. Therefore $(Y_n)$ has an asymptotic distribution (which we call $\pi$). Since the convergence is bounded by $(1 - 2\varepsilon)^n$, the convergence is geometric.

We know that $(Y_n)$ is a subsequence of $(X_n)$ with TPM $P^m$. We make take any $m \geq N$, so all points of $(X_n)$ after $N$ are in a MC of the form $(Y_n)$ with asymptotic distribution $\pi$. Therefore $(X_n)$ must also have an asymptotic distribution, $\pi$, which it converges to geometrically. $\square$

*Example* 5.4.8. We continue from Example 5.1.6, in which we found that the distribution after $n$ transitions is

$$\pi^{(n)} = \begin{pmatrix} \frac{1}{3} + A\left(\frac{1}{4}\right)^n & \frac{2}{3} - A\left(\frac{1}{4}\right)^n \end{pmatrix},$$

where the constant $A$ is determined by the initial distribution $\pi^{(0)}$.

We see that the distribution converges to the asymptotic distribution geometrically, as expected from Theorem 5.4.7.

# 6 Applications of Markov Chains

## 6.1 Simple Random Walks

We now look at wider applications of Markov Chains. Consider a random walk on $\mathbb{Z}^d$, considering separately the cases $d = 1$, $d = 2$, and $d \geq 3$. At any point $(i_1, \ldots, i_d) \in \mathbb{Z}^d$, the chain may transition to one of $2d$ points

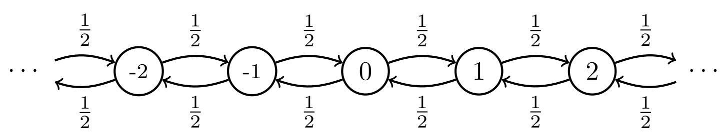
$$(i_1 - 1, i_2, \ldots, i_d), (i_1 + 1, i_2, \ldots, i_d), \ldots, (i_1, i_2, \ldots, i_d - 1), (i_1, i_2, \ldots, i_d + 1),$$

each with probability $\frac{1}{2d}$. This means we have transition probability matrix $P$, where

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

for all $i, j \in \mathbb{Z}^d$.

We will determine whether states in $\mathbb{Z}^d$ are recurrent or transient (the full result known as Pólya's theorem). Firstly, we consider the $d = 1$ case, with the proof from [Nor97] pg 29-30.




**Theorem 6.1.1.** *Let $(X_n)$ be the simple, symmetric random walk on $\mathbb{Z}$, with TPM $P$ as given above. Then every state is recurrent.*

*Proof.* In this case we have

$$\mathbb{P}(X_n = X_{n-1} - 1) = \frac{1}{2},$$
$$\mathbb{P}(X_n = X_{n-1} + 1) = \frac{1}{2},$$

at each step $n \in \mathbb{N}$.

We first consider the state 0 (starting from $X_0 = 0$).

After an odd number of steps, say $2n + 1$, we must be at an odd state, so we cannot be at 0. So we have $P^{(2n+1)}_{0,0} = 0$.

If after an even number of steps, say $2n$, we are back at 0, we must have taken $n$ steps up and $n$ steps down. The number of possible routes back to 0 is the number of possible ways to choose $n$ steps up from a total of $2n$ steps, $\binom{2n}{n}$, and each step has probability $\frac{1}{2}$.

Therefore

$$P_{0,0}^{(2n)} = \binom{2n}{n}\left(\frac{1}{2}\right)^{2n} = \frac{(2n)!}{(n!)^2 \cdot 4^n}.$$

By Stirling's approximation (Lemma A.1.1), $n! \sim A\sqrt{n}(\frac{n}{e})^n$, so we have

$$\begin{aligned} P_{0,0}^{(2n)} &\sim \frac{A\sqrt{2n}\cdot(2n)^{2n}}{e^{2n}}\left(\frac{e^n}{A\sqrt{n}\cdot n^n}\right)^2\frac{1}{4^n} \\ &= \frac{A\sqrt{2}\sqrt{n}\cdot 4^n n^{2n}}{e^{2n}}\frac{e^{2n}}{A^2 n\cdot n^{2n}}\frac{1}{4^n} \\ &= \frac{\sqrt{2}}{A}\frac{1}{\sqrt{n}}. \end{aligned}$$

Since $\sum_{n=1}^{\infty}\frac{1}{\sqrt{n}} = \infty$, we have

$$\sum_{n=0}^{\infty} P_{0,0}^{(n)} = \sum_{n=0}^{\infty} P_{0,0}^{(2n)} = \infty.$$

So by Theorem 3.2.5, 0 is a recurrent state.

Therefore by Theorem 3.2.7, every state is recurrent. □

Now we consider the case where $d = 2$, with the proof from [Nor97] pg 31-32 (following directly from the proof of Theorem 6.1.1).

**Theorem 6.1.2.** *Let $(X_n)$ be the simple, symmetric random walk on $\mathbb{Z}^2$, with TPM $P$ as given above. Then every state is recurrent.*

*Proof.* In this case we have

$$\begin{aligned} \mathbb{P}(X_n = X_{n-1} + (1,0)) &= \frac{1}{4}, \\ \mathbb{P}(X_n = X_{n-1} + (-1,0)) &= \frac{1}{4}, \\ \mathbb{P}(X_n = X_{n-1} + (0,1)) &= \frac{1}{4}, \\ \mathbb{P}(X_n = X_{n-1} + (0,-1)) &= \frac{1}{4}. \end{aligned}$$

at each step $n \in \mathbb{N}$. Similarly to the walk in $\mathbb{Z}$, we first consider the state $0 = (0,0)$ (starting from $X_0 = 0$).

Clearly $P_{0,0}^{(2n+1)} = 0$, so we only need to consider $P_{0,0}^{(2n)} = 0$.

Let $(X_n^+)$ and $(X_n^-)$ be the orthogonal projections of $(X_n)$ on the lines (in $\mathbb{Z}^2$) $y = x$ and $y = -x$ respectively. Clearly $(X_0^+) = 0$ and $(X_0^-) = 0$.

So we have that

$$X_n = X_{n-1} + (1,0) \implies X_n^+ = X_{n-1}^+ + \frac{1}{\sqrt{2}} \text{ and } X_n^- = X_{n-1}^- + \frac{1}{\sqrt{2}},$$

$$X_n = X_{n-1} + (-1,0) \implies X_n^+ = X_{n-1}^+ - \frac{1}{\sqrt{2}} \text{ and } X_n^- = X_{n-1}^- - \frac{1}{\sqrt{2}},$$
$$X_n = X_{n-1} + (0,1) \implies X_n^+ = X_{n-1}^+ + \frac{1}{\sqrt{2}} \text{ and } X_n^- = X_{n-1}^- - \frac{1}{\sqrt{2}},$$
$$X_n = X_{n-1} + (0,-1) \implies X_n^+ = X_{n-1}^+ - \frac{1}{\sqrt{2}} \text{ and } X_n^- = X_{n-1}^+ + \frac{1}{\sqrt{2}},$$

which means

$$\mathbb{P}\left(X_n^+ = X_{n-1}^+ + \frac{1}{\sqrt{2}}\right) = \frac{1}{2}$$
$$\mathbb{P}\left(X_n^+ = X_{n-1}^+ - \frac{1}{\sqrt{2}}\right) = \frac{1}{2}$$

and

$$\mathbb{P}\left(X_n^- = X_{n-1}^- + \frac{1}{\sqrt{2}}\right) = \frac{1}{2}$$
$$\mathbb{P}\left(X_n^- = X_{n-1}^- - \frac{1}{\sqrt{2}}\right) = \frac{1}{2}.$$

This means $(X_n^+)$ and $(X_n^-)$ are (independent) symmetric simple random walks on $\frac{1}{\sqrt{2}}\mathbb{Z}$.

$X_n = 0$ if and only if both $X_n^+ = 0$ and $X_n^- = 0$, and so

$$\begin{aligned} P_{0,0}^{(2n)} &= \mathbb{P}(X_n^+ = 0)\mathbb{P}(X_n^- = 0) \\ &= \left(\binom{2n}{n}\left(\frac{1}{2}\right)^{2n}\right)^2. \end{aligned}$$

Therefore from the proof of Theorem 6.1.1,

$$\begin{aligned} P_{0,0}^{(2n)} &\sim \left(\frac{\sqrt{2}}{A}\frac{1}{\sqrt{n}}\right)^2 \\ &= \frac{2}{A^2}\frac{1}{n}. \end{aligned}$$

Since $\sum_{n=1}^{\infty}\frac{1}{n} = \infty$, we have

$$\sum_{n=0}^{\infty} P_{0,0}^{(n)} = \sum_{n=0}^{\infty} P_{0,0}^{(2n)} = \infty.$$

So by Theorem 3.2.5, 0 is a recurrent state, and then by Theorem 3.2.7, every state is recurrent. □

Finally, we consider the case where $d \geq 3$. The method used in the proof of Theorem 6.1.2 does not work in more than two dimensions, so we use a different proof – adapted from [Nor97] pg 32, which considers the $d = 3$ case only.

**Theorem 6.1.3.** *Let $(X_n)$ be the simple, symmetric random walk on $\mathbb{Z}^d$, $d \geq 3$, with TPM $P$ as given above. Then every state is transient.*

*Proof.* Again, we first consider the state $0 = (0, 0, \dots, 0)$ (starting from $X_0 = 0$). Clearly $P_{0,0}^{(2n+1)} = 0$, so we only need to consider $P_{0,0}^{(2n)}$.

If we take $2n$ steps from 0, returning to 0, consider the number of steps taken along each axis: if $i$ steps were taken in the positive direction, $i$ steps must also have been taken in the negative direction. So if $i_q$ is the number of steps taken in the $q$-direction, we have $i_1, \dots, i_m \geq 0$ such that $i_1 + \cdots + i_d = n$.

Over the $2n$ steps, we have $i_1$ steps along the first axis in the positive direction, $i_1$ steps along the first axis in the negative direction, $i_2$ steps along the first axis in the positive direction, and so on. This means the number of possible paths is

$$\sum_{i_1,\dots,i_d} \frac{(2n)!}{(i_1!)^2 \dots (i_d!)^2},$$

where each path has probability $\left(\frac{1}{2d}\right)^{2n}$.

Therefore

$$\begin{aligned}
P_{0,0}^{(2n)} &= \left(\frac{1}{2d}\right)^{2n} \sum_{i_1,\dots,i_d} \frac{(2n)!}{(i_1!)^2 \dots (i_d!)^2} \\
&= \frac{1}{2^{2n}} \frac{1}{d^{2n}} \frac{(2n)!}{(n!)^2} \sum_{i_1,\dots,i_d} \frac{(n!)^2}{(i_1!)^2 \dots (i_d!)^2} \\
&= \frac{1}{4^n} \frac{1}{d^{2n}} \binom{2n}{n} \sum_{i_1,\dots,i_d} \left(\frac{n!}{i_1! \dots i_d!}\right)^2 \\
&= \frac{1}{4^n} \binom{2n}{n} \sum_{i_1,\dots,i_d} \binom{n}{i_1, \dots, i_d}^2 \left(\frac{1}{d^n}\right)^2.
\end{aligned}$$

The probability of placing $n$ balls into $d$ boxes any one specific way is $\binom{n}{i_1,\dots,i_d}\frac{1}{d^n}$ for some $i_1, \dots, i_d \geq 0$. This means $0 \leq \binom{n}{i_1,\dots,i_d}\frac{1}{d^n} \leq 1$, and so

$$\left(\binom{n}{i_1, \dots, i_d}\frac{1}{d^n}\right)^2 \leq \binom{n}{i_1, \dots, i_d}\frac{1}{d^n}.$$

.

Therefore

$$\begin{aligned}
P_{0,0}^{(2n)} &= \binom{2n}{n}\frac{1}{4^n} \sum_{i_1,\dots,i_d} \left(\binom{n}{i_1, \dots, i_d}\frac{1}{d^n}\right)^2 \\
&\leq \binom{2n}{n}\frac{1}{4^n} \sum_{i_1,\dots,i_d} \binom{n}{i_1, \dots, i_d}\frac{1}{d^n}.
\end{aligned}$$

Also, we have that for any $i_1, \dots, i_d \geq 0$, if $n = md$, then

$$\binom{n}{i_1, \dots, i_d} = \frac{n!}{i_1! \dots i_d!} \leq \frac{n!}{m! \dots m!} = \binom{n}{m, \dots, m}.$$

Therefore if $n = md$,

$$P_{0,0}^{(2n)} \leq \binom{2n}{n}\frac{1}{4^n} \sum_{i_1,\dots,i_d} \binom{n}{m, \dots, m}\frac{1}{d^n}$$

$$
\begin{aligned}
&\leq \binom{2n}{n}\frac{1}{4^n}\binom{n}{m,\ldots,m}\frac{1}{d^n}\\
&= \frac{(2n)!}{(n!)^2}\frac{1}{4^n}\frac{n!}{(m!)^d}\frac{1}{d^n}\\
&= \frac{1}{4^n d^n}\frac{(2n)!}{n!(m!)^d}.
\end{aligned}
$$

By Stirling's approximation (Lemma A.1.1), $n! \sim A\sqrt{n}(\frac{n}{e})^n$, so we have

$$
\begin{aligned}
P_{0,0}^{(2n)} &\sim \frac{1}{4^n d^n}\frac{A\sqrt{2n}\cdot(2n)^{2n}}{e^{2n}}\frac{e^n}{A\sqrt{n}\cdot n^n}\left(\frac{e^m}{A\sqrt{m}\cdot m^m}\right)^d\\
&= \frac{1}{4^n d^n}\frac{A\sqrt{2}\sqrt{n}\cdot 4^n n^{2n}}{e^{2n}}\frac{e^n}{A\sqrt{n}\cdot n^n}\frac{e^n}{A^d\sqrt{m}^d\cdot m^n}\\
&= \frac{\sqrt{2}}{A^d\sqrt{m}^d}\\
&= \frac{\sqrt{2}\sqrt{d}^d}{A^d}\frac{1}{n^{d/2}}.
\end{aligned}
$$

Since $\sum_{n=1}^{\infty}\frac{1}{n^{-a}} < \infty$ for $a > 1$, we have

$$
\sum_{m=0}^{\infty} P_{0,0}^{(2md)} < \infty.
$$

We see that

$$
\begin{aligned}
P_{0,0}^{(2md)} &\geq P_{0,0}^{(2)}P_{0,0}^{(2md-2)} \geq \left(\frac{1}{2d}\right)^2 P_{0,0}^{(2md-2)},\\
P_{0,0}^{(2md)} &\geq P_{0,0}^{(4)}P_{0,0}^{(2md-4)} \geq \left(\frac{1}{2d}\right)^4 P_{0,0}^{(2md-4)},\\
&\vdots\\
P_{0,0}^{(2md)} &\geq P_{0,0}^{(2d-2)}P_{0,0}^{(2md-2d+2)} \geq \left(\frac{1}{2d}\right)^{2d-2} P_{0,0}^{(2md-2d+2)},
\end{aligned}
$$

and so

$$
\begin{aligned}
\sum_{m=1}^{\infty} P_{0,0}^{(2md-2)} &\leq (2d)^2\sum_{m=1}^{\infty} P_{0,0}^{(2md)} < \infty,\\
\sum_{m=1}^{\infty} P_{0,0}^{(2md-4)} &\leq (2d)^4\sum_{m=1}^{\infty} P_{0,0}^{(2md)} < \infty,\\
&\vdots\\
\sum_{m=1}^{\infty} P_{0,0}^{(2md-2d+2)} &\leq (2d)^{2d-2}\sum_{m=1}^{\infty} P_{0,0}^{(2md)} < \infty.
\end{aligned}
$$

Therefore

$$
\sum_{n=0}^{\infty} P_{0,0}^{(n)} = \sum_{n=0}^{\infty} P_{0,0}^{(2n)}
$$

$$= \sum_{m=0}^{\infty} P_{0,0}^{(2md)} + \sum_{m=1}^{\infty} P_{0,0}^{(2md-2)} + \sum_{m=1}^{\infty} P_{0,0}^{(2md-4)} + \cdots + \sum_{m=1}^{\infty} P_{0,0}^{(2md-2d+2)}$$
$$< \infty,$$

and so by Theorem 3.2.5, 0 is a transient state.

Therefore by Theorem 3.2.7, every state is transient. □

Since the MC is irreducible, the single communicating class must be closed, and we see that Theorem 3.2.9 cannot be extended to infinite closed classes.

We can also extend Pólya's theorem to other random walks in $\mathbb{Z}^d$. Let $\zeta_n = X_n - X_{n-1}$ and $B$ be the covariance matrix of $\zeta_1$. If $\mathbb{E}(\zeta_n) = 0$, and all entries of $B$ are finite, then by the *local central limit theorem*

$$P_{0,0}^{(n)} = \frac{1}{(2\pi n)^{d/2}\sqrt{\det(B)}} + o\left(\frac{1}{n^{d/2}}\right).$$

Since $\sum_{n=0}^{\infty} n^{-d/2}$ is convergent if and only if $d = 1$ or $d = 2$, by Theorems 3.2.5 and 3.2.7 all states are recurrent when $d = 1$ or $d = 2$, and all states are transient when $d \geq 3$ (the same result as Pólya's theorem). For example, consider a random walk in $\mathbb{Z}$ where from any given state there is a $\frac{1}{3}$ chance of increasing by 2 and a $\frac{2}{3}$ chance of decreasing by 1 – the expected move is 0, so all states are recurrent. The proof of the local central limit theorem is beyond the scope of this report – see [GL19] for a proof of this specific case.

## 6.2 Branching Processes

In another application of Markov Chains we consider a class of MCs $(X_n)$ called branching processes, which model the size of a population over time.

**Definition 6.2.1.** A MC $(X_n)$ is a *branching process* if for each $i \in I = \mathbb{N}_0$, $X_{i+1}$ is the total offspring of the population $X_i$, each of which produces $j$ offspring ($j \geq 0$) with probability $p_j$.

*Remark* 6.2.2. Let $Z_i$ be the offspring of a single individual $i$. Then

$$\mathbb{E}(Z_i) = \mu = \sum_{j=0}^{\infty} j p_j,$$
$$\mathrm{Var}(Z_i) = \sigma^2 = \sum_{j=0}^{\infty} (j-\mu)^2 p_j,$$

and

$$X_n | X_{n-1} = \sum_{i=1}^{X_{n-1}} Z_i,$$

where $Z_1, \ldots, Z_{X_{n-1}}$ are independently and identically distributed.

From this definition we will calculate the expected population (and variance in population size) after $n$ generations, with the proof from [Ros14] pg 235.

**Theorem 6.2.3.** *Given a branching process $(X_n)$ with $X_0 = 1$, we have*

$$\mathbb{E}(X_n) = \mu^n \text{ and } Var(X_n) = \begin{cases} \sigma^2 \mu \frac{1-\mu^n}{1-\mu} & \text{if } \mu \neq 1 \\ n\sigma^2 & \text{if } \mu = 1 \end{cases}.$$

*Proof.* We will use proof by induction.

When $n = 0$, we have

$$\mathbb{E}(X_0) = \mathbb{E}(1) = 1 = \mu^0$$

and

$$\mathrm{Var}(X_0) = \mathrm{Var}(1) = 0 = \begin{cases} \sigma^2 \mu \frac{1-\mu^0}{1-\mu} & \text{if } \mu \neq 1 \\ 0 \cdot \sigma^2 & \text{if } \mu = 1 \end{cases},$$

so the base case is satisfied.

We will assume that the $k$-th case is true, and so

$$\mathbb{E}(X_k) = \mu^k \text{ and } \mathrm{Var}(X_k) = \begin{cases} \sigma^2 \mu \frac{1-\mu^k}{1-\mu} & \text{if } \mu \neq 1 \\ k\sigma^2 & \text{if } \mu = 1 \end{cases}.$$

For the case when $n = k + 1$ we use the tower property, and so

$$\begin{aligned} \mathbb{E}(X_{k+1}) &= \mathbb{E}(\mathbb{E}(X_{k+1}|X_k)) \\ &= \mathbb{E}\left(\mathbb{E}\left(\sum_{i=1}^{X_k} Z_i\right)\right). \end{aligned}$$

By the linearity of expectation, we have

$$\begin{aligned} \mathbb{E}(X_{k+1}) &= \mathbb{E}\left(\sum_{i=1}^{X_k} \mathbb{E}(Z_i)\right) \\ &= \mathbb{E}\left(\sum_{i=1}^{X_k} \mu\right) \\ &= \mathbb{E}(\mu X_k) \\ &= \mu \mathbb{E}(X_k), \end{aligned}$$

and then by our assumption

$$\mathbb{E}(X_{k+1}) = \mu\mu^k = \mu^{k+1},$$

as required.

Similarly we use the conditional variance formula, and so

$$\begin{aligned} \mathrm{Var}(X_{k+1}) &= \mathbb{E}(\mathrm{Var}(X_{k+1}|X_k)) + \mathrm{Var}(\mathbb{E}(X_{k+1}|X_k)) \\ &= \mathbb{E}\left(\mathrm{Var}\left(\sum_{i=1}^{X_k} Z_i\right)\right) + \mathrm{Var}\left(\mathbb{E}\left(\sum_{i=1}^{X_k} Z_i\right)\right), \end{aligned}$$

and then by independence of $Z_i$ we have

$$\begin{aligned} \mathrm{Var}(X_{k+1}) &= \mathbb{E}\left(\sum_{i=1}^{X_k} \mathrm{Var}(Z_i)\right) + \mathrm{Var}\left(\sum_{i=1}^{X_k} \mathbb{E}(Z_i)\right) \\ &= \mathbb{E}\left(\sum_{i=1}^{X_k} \sigma^2\right) + \mathrm{Var}\left(\sum_{i=1}^{X_k} \mu\right) \end{aligned}$$

$$
\begin{aligned}
&= \mathbb{E}(\sigma^2 X_k) + \mathrm{Var}(\mu X_k) \\
&= \sigma^2 \mathbb{E}(X_k) + \mu^2 \mathrm{Var}(X_k).
\end{aligned}
$$

Therefore by our assumptions

$$
\begin{aligned}
\mathrm{Var}(X_{k+1}) &= \sigma^2\mu^k + \mu^2 \begin{cases} \sigma^2 \mu \frac{1-\mu^k}{1-\mu} & \text{if } \mu \neq 1 \\ k\sigma^2 & \text{if } \mu = 1 \end{cases} \\
&= \begin{cases} \sigma^2\mu^k + \sigma^2\mu^3 \frac{1-\mu^k}{1-\mu} & \text{if } \mu \neq 1 \\ \sigma^2 + k\sigma^2 & \text{if } \mu = 1 \end{cases} \\
&= \begin{cases} \sigma^2 \mu \frac{1-\mu^{k+1}}{1-\mu} & \text{if } \mu \neq 1 \\ (k+1)\sigma^2 & \text{if } \mu = 1 \end{cases},
\end{aligned}
$$

as required. □

This leads to the following corollary:

**Corollary 6.2.4.** *Starting from $X_0 = 1$, if $\mu < 1$ then the expected number of individuals to ever exist is $\frac{1}{1-\mu}$.*

*Proof.* We see that

$$
\mathbb{E}(\text{number of individuals to exist}|X_0 = 1) = \sum_{i=0}^{\infty} \mathbb{E}(X_i) = \sum_{i=0}^{\infty} \mu^i = \frac{1}{1-\mu}.
$$

□

These results can be extended to the case where $X_0 = m$ for some $m \in \mathbb{N}$ by considering the initial $m$ families independently – by the linearity of expectation and variance (since the families are independent) we simply multiply the results of Theorem 6.2.3 and Corollary 6.2.4 by $m$.

Now we consider the likelihood that a given branching process continues for an infinite amount of time, or eventually reaches the absorbing state of 0 ("dying out"). This is called the *extinction probability*. The following proof is expanded from [Ros14] pg 236, and we do not consider the case where $\mu \neq 1$.

**Definition 6.2.5.** Given a branching process $(X_n)$ with $X_0 = 1$, let $\pi_0$ denote the *extinction probability*, which is the probability that the population will eventually reach 0 and then remain at 0:

$$
\pi_0 = \lim_{n\to\infty} \mathbb{P}(X_n = 0).
$$

**Theorem 6.2.6.** *If $\mu < 1$, $\pi_0 = 1$, and if $\mu > 1$, $\pi_0$ is the smallest $\pi > 0$ such that $\pi = \sum_{j=0}^{\infty} \pi^j p_j$.*

*Proof.* By the law of total probability,

$$
\pi_0 = \mathbb{P}(\text{population eventually dies out}|X_0 = 1) = \sum_{j=0}^{\infty} \mathbb{P}(\text{population eventually dies out}|X_1 = j) p_j.
$$

But $\mathbb{P}(\text{population eventually dies out}|X_0 = j) = \pi_0^j$ (considering the $j$ families independently).

Therefore $\pi_0 = \sum_{j=0}^{\infty} \pi_0^j p_j$, as required.

Let $\mu < 1$. From Theorem 6.2.3, we have

$$\begin{aligned}\mu^n &= \mathbb{E}(X_n) \\ &= \sum_{j=0}^{\infty} j\mathbb{P}(X_n = j) \\ &= \sum_{j=1}^{\infty} j\mathbb{P}(X_n = j) \\ &\geq \sum_{j=1}^{\infty} \mathbb{P}(X_n = j) \\ &= \mathbb{P}(X_n \neq 0).\end{aligned}$$

As $n \to \infty$, $\mu^n \to 0$, and so have that $\mathbb{P}(X_n \neq 0) \to 0$. This means $\mathbb{P}(X_n = 0) \to 1$ as $n \to \infty$, and so $\pi_0 = 1$.

Now let $\mu > 1$. First, we wish to show by induction that if $\pi = \sum_{j=0}^{\infty} \pi^j p_j$, then $\pi \geq \mathbb{P}(X_n = 0)$ for all $n \in \mathbb{N}_0$.

We know $X_0 = 1$, so $\mathbb{P}(X_0 = 0) = 0$. Since $\pi$ is a probability, $\pi \geq 0 = \mathbb{P}(X_0 = 0)$, as required.

Assume that for some $k \in \mathbb{N}_0$, $\pi \geq \mathbb{P}(X_k = 0)$. Then

$$\begin{aligned}\mathbb{P}(X_{k+1} = 0) &= \sum_{j=0}^{\infty} \mathbb{P}(X_k = 0)^j p_j \\ &\leq \sum_{j=0}^{\infty} \pi^j p_j \\ &= \pi,\end{aligned}$$

so $\pi \geq \mathbb{P}(X_{k+1} = 0)$.

Therefore by induction we have that $\pi \geq \mathbb{P}(X_n = 0)$ for all $n \in \mathbb{N}_0$.

This means $\lim_{n\to\infty} \pi \geq \lim_{n\to\infty} \mathbb{P}(X_n = 0)$, which means $\pi \geq \pi_0$. Therefore $\pi_0$ is the smallest $\pi > 0$ such that $\pi = \sum_{j=0}^{\infty} \pi^j p_j$. ☐

## 6.3 Markov Chain Monte Carlo

Consider a discrete random variable $X$ on state space $I$, and consider

$$\theta = \mathbb{E}(h(X)) = \sum_{i \in I} h(x_i)\mathbb{P}(X = x_i),$$

which we wish to calculate for some function $h$. Many properties of $X$ can be expressed as the expectation of some function, including variance, where $\mathrm{Var}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2$, or the probability of some event $A$, where $\mathbb{P}(X \in A) = \mathbb{E}(\mathbf{1}_A)$.

If we can evaluate $h$ at each point $x_i$, this is simple; however often this is not the case. If we know $\theta$ has distribution $\pi(\theta)$, we use *Monte Carlo Methods*: we generate $n$ samples from this distribution, $\theta_1, \ldots, \theta_n$ (for some $n \in \mathbb{N}$), which are used to calculate an estimate

$$\hat{\theta} = \frac{1}{n} \sum_{i=1}^{n} h(\theta_i).$$

**Theorem 6.3.1.** *The estimate $\hat{\theta}$ is unbiased (which is equivalent to $\mathbb{E}(\hat{\theta}) = \theta$).*

*Proof.* By the linearity of expectation

$$\begin{aligned}\mathbb{E}(\hat{\theta}) &= \mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} h(\theta_i)\right) \\ &= \frac{1}{n}\sum_{i=1}^{n} \mathbb{E}(h(\theta_i)) \\ &= \frac{1}{n}\sum_{i=1}^{n} \theta \\ &= \theta.\end{aligned}$$

□

The preceding method only works if we can easily sample points from the distribution $\pi$. If we cannot do this we instead use *Markov Chain Monte Carlo*, using the method presented in [GRGDRS01], pg 292-293. We will assume that $I$ is a finite set, and that $\pi$ has no zero entries. We construct a Markov Chain $X_0, X_1, \dots$ with asymptotic distribution $\pi$ that is easy to sample from. To determine the value of this Markov Chain at a given time $n+1$, from $X_n$, we use the *Metropolis-Hastings Algorithm*:

**Theorem 6.3.2** (Metropolis-Hastings Algorithm)**.** *We determine the MC using two matrices: $A$, the acceptance matrix, and $H$, the proposal matrix. The elements $H_{i,j}$ are chosen such that they are easy to sample from – for a random variable $Y$, we let*

$$H_{i,j} = \mathbb{P}(Y = j | X_n = i),$$

*and then*

$$A_{i,j} = \min\left\{1, \frac{\pi_j H_{j,i}}{\pi_i H_{i,j}}\right\}.$$

*Then given $X_n = i$ and $Y = j$, we determine $X_{n+1}$ by*

$$X_{n+1} = \begin{cases} Y & \text{with probability } A_{i,j}, \\ X_n & \text{with probability } 1 - A_{i,j}. \end{cases}$$

*Proof.* To use this algorithm we prove that the MC has an asymptotic distribution, and that the asymptotic distribution is our target distribution $\pi$.

If the MC moves from a state $i$ to a different state $j$, it must be sampled from $H$, probability $H_{i,j}$, and then (independently) accepted, probability $A_{i,j}$. This means the TPM $P$ of $(X_n)$ is

$$P_{i,j} = \begin{cases} H_{i,j}A_{i,j} & \text{if } i \neq j, \\ 1 - \sum_{k \in I \setminus \{i\}} H_{i,k}A_{k,i} & \text{if } i = j, \end{cases}$$

for any $i, j \in I$. Since $H$ is chosen arbitrarily, we may choose the elements of $H$ such that the MC is irreducible and aperiodic. By Theorem 5.4.7 (since the MC has a finite number of states), the MC has an asymptotic distribution.

Let $i, j \in I$. If $i = j$ then it is trivial that $\pi_i P_{i,j} = \pi_j P_{j,i}$ holds. If $i \neq j$, we have

$$\pi_i P_{i,j} = \pi_i H_{i,j} A_{i,j}$$

$$
\begin{aligned}
&= \pi_i H_{i,j} \min \left\{ 1, \frac{\pi_j H_{j,i}}{\pi_i H_{i,j}} \right\} \\
&= \min \left\{ \pi_i H_{i,j}, \pi_j H_{j,i} \right\}.
\end{aligned}
$$

We now have a term symmetric in $i$ and $j$, so by reversing our previous steps (swapping $i$ and $j$) we have

$$
\begin{aligned}
\pi_i P_{i,j} &= \min \left\{ \pi_j H_{j,i}, \pi_i H_{i,j} \right\} \\
&= \pi_j H_{j,i} \min \left\{ 1, \frac{\pi_i H_{i,j}}{\pi_j H_{j,i}} \right\} \\
&= \pi_j H_{j,i} A_{j,i} \\
&= \pi_j P_{j,i}.
\end{aligned}
$$

Therefore detailed balance holds for any $i, j \in I$, and so by Theorem Theorem 5.2.2 $\pi$ is an invariant distribution. We know the asymptotic distribution exists, so by Theorem 5.1.7 $\pi$ is the asymptotic distribution of the MC. □

In practise the Metropolis-Hastings Algorithm is often generalised to continuous (unaccountably infinite) state spaces – instead of transition probability matrices and distribution vectors we have transition probability kernels and distribution densities. Many definitions from countable state-space MCs have analogous definitions to continuous state spaces (see [MT05] from pg 66 onwards) which are not covered in this report.

# A Appendix

## A.1 Stirling's Approximation

We now provide a proof of Stirling's Approximation based on [Nor97] pg 58-59, as used in Section 6.1. It is known that $A = \sqrt{2\pi}$, but we will not prove this (as it requires a significantly longer proof).

**Lemma A.1.1** (Stirling's Approximation)**.** *For any $n \in \mathbb{N}$ we have that*

$$n! \sim A\sqrt{n}\left(\frac{n}{e}\right)^n,$$

*where $A$ is some positive constant.*

*Proof.* Let $x \in (-1, 1)$. Then note the following Maclaurin series:

$$\begin{aligned}\log(1+x) &= x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \cdots,\\ \log(1-x) &= -x - \frac{1}{2}x^2 - \frac{1}{3}x^3 - \cdots.\end{aligned}$$

This means

$$\begin{aligned}\frac{1}{2}\log\left(\frac{1+x}{1-x}\right) &= \frac{1}{2}(\log(1+x) - \log(1-x))\\ &= \frac{1}{2}\left(x - \frac{1}{2}x^2 + \frac{1}{3}x^3 - \cdots + x + \frac{1}{2}x^2 + \frac{1}{3}x^3 + \cdots\right)\\ &= \frac{1}{2}\left(2x + \frac{2}{3}x^3 + \frac{2}{5}x^5 + \cdots\right)\\ &= x + \frac{1}{3}x^3 + \frac{1}{5}x^5 + \cdots. \qquad \text{(A.1.1)}\end{aligned}$$

Let $(A_n)$ and $(a_n)$ be sequences in $\mathbb{R}$, where

$$A_n = \frac{n!}{n^{n+1/2}e^{-n}}$$

and $a_n = \log(A_n)$ for all $n \in \mathbb{N}$.

Then for any $n \in \mathbb{N}$,

$$\begin{aligned}a_n - a_{n+1} &= \log\left(\frac{n!}{n^{n+1/2}e^{-n}}\right) - \log\left(\frac{(n+1)!}{(n+1)^{n+3/2}e^{-(n+1)}}\right)\\ &= \log\left(\left(\frac{n+1}{n}\right)^{n+1/2}\right) - 1\\ &= \left(n + \frac{1}{2}\right)\log\left(\frac{n+1}{n}\right) - 1\\ &= (2n+1)\frac{1}{2}\log\left(\frac{(2n+1)+1}{(2n+1)-1}\right) - 1\\ &= (2n+1)\frac{1}{2}\log\left(\frac{1+(2n+1)^{-1}}{1-(2n+1)^{-1}}\right) - 1.\end{aligned}$$

This means by (A.1.1), we have

$$a_n - a_{n+1} = (2n+1)\left(\frac{1}{2n+1} + \frac{1}{3(2n+1)^3} + \frac{1}{5(2n+1)^5} + \cdots\right) - 1$$

$$
\begin{aligned}
&= \frac{1}{3(2n+1)^2} + \frac{1}{5(2n+1)^4} + \frac{1}{7(2n+1)^6} + \cdots \\
&> 0
\end{aligned}
$$

for any $n \in \mathbb{N}$.

Therefore $(a_n)$ is decreasing and $a_n \geq 0$ for all $n \in \mathbb{N}$ – $a_n \to a$ as $n \to \infty$ for some $a$, and so $A_n \to A = \exp(a)$ as $n \to \infty$.

Therefore

$$\frac{n!}{An^{n+1/2}e^{-n}} \to 1$$

as $n \to \infty$, and so

$$n! \sim A\sqrt{n}\left(\frac{n}{e}\right)^n,$$

as required. □

# B Bibliography


[GGDW14] Geoffrey Grimmett and Dominic Welsh. *Probability: An Introduction, Second Edition*. Oxford University Press, 2014.

[GL19] Giuseppe Genovese and Renato Lucà. Local central limit theorem for a random walk perturbed in one point. *Mathematical Physics, Analysis and Geometry*, 2019.

[GRGDRS01] Geoffrey R. Grimmett and David R. Stirzaker. *Probability and Random Processes, Third Edition*. Oxford University Press, 2001.

[JGKS76] John G. Kemeny and J. Laurie Snell. *Finite Markov Chains*. Springer-Verlag, 1976.

[Kor20] Dmitry Korshunov. MATH332 Lecture Notes. *Lancaster University*, 2020.

[MT05] S.P. Meyn and R.L. Tweedie. *Markov Chains and Stochastic Stability*. Springer-Verlag, 2005.

[Nor97] J.R. Norris. *Markov Chains*. Cambridge University Press, 1997.

[Ros14] Sheldon M. Ross. *An Introduction to Probability Models, Eleventh Edition*. Academic Press, 2014.