Stationary Distribution of Markov Chain

Eleazar Neamat
Contents

Acknowledgments 3
Abstract 4
Public Summary 5
1 Introduction 6
2 Markov Chains 8
  2.1 Markov Property ........................................... 8
  2.2 Transition Probability Matrix ................................ 9
  2.3 Higher Order Transition Probabilities ......................... 10
  2.4 Markov Chains as Graphs .................................... 11
  2.5 Classification of States .................................... 12
  2.6 Ergodic chain .............................................. 15
3 Stationary Distribution and its Significance 17
  3.1 The stationary distribution .................................. 17
  3.2 Existing and uniqueness of the stationary distribution ........ 20
4 Simulation 24
  4.1 Simulation Algorithm ....................................... 25
Conclusions 28
Bibliography 29
Acknowledgments

I am profoundly grateful to God for being my guiding light and providing strength throughout this academic journey. His grace has been a constant source of inspiration and resilience. I want to express my heartfelt gratitude to everyone who supported me in completing this thesis. This was my first thesis in Sweden, in the beginning, it was difficult for me, but I am very thankful to supervisor Lauri Viitasaaari for his guidance, feedback, and encouragement throughout the thesis. I am also very thankful to the examiner, Julian Külshammer, for his time, review, and constructive feedback on this thesis.

I would like to dedicate a special note of appreciation to the memory of my beloved mother. Tragically, during my studies, I faced the profound loss of her passing. The grief was overwhelming, and for a time, I found it difficult to continue my studies. During this challenging period, my wife Sabahat Eleazar encouraged me to continue my studies. Without her, the completion of this thesis would not have been possible.

I extend my heartfelt gratitude to my father, Neamat Masih. Beyond the financial support, his love and motivation have been constant sources of inspiration. I dedicate this accomplishment to my father, and I hope to make him proud.

I extend my deepest appreciation to my brothers, Abiezer and Justin Walter, and my sisters. Their encouragement and love have been a pillar of strength. I also want to dedicate this achievement to my daughters Aliza, Atara, and Eliana and I am fortunate to have such a supportive family. Special thanks go to my friend, Boez Kamran Khan, for his guiding friendship. During the challenging period following my mother’s passing, he became a beacon of hope. His words of encouragement and guidance helped me to find strength during difficult times. He has been a source of inspiration for my journey.

In conclusion, I wish to express my sincere gratitude to Uppsala University for giving me the opportunity of admission. Studying at a top-tier university is a remarkable privilege that I might not have had the opportunity to experience in my homeland. The academic environment, resources, and support provided by Uppsala University have played a pivotal role in shaping my educational journey.
Abstract

Markov chain is a mathematical tool for modeling systems that evolve over time and has been used in many fields such as physics, chemistry, economics, biology, and data science. This thesis contains an introduction to the theory and the applications of Markov chains, focusing on those with finite state spaces. Starting with basic concepts and techniques, the theory of Markov chains is comprehensively studied. The basic concepts covered include the Markov property, transition matrix, higher order transition probabilities, classification of states, and Markov chains as graphs. The stationary distribution, its importance in probability theory, existence, and uniqueness of stationary distribution are then discussed, while the final part of the thesis deals with the simulations of Markov chains. Two examples are presented to illustrate the technique of Markov chain simulation, including a weather prediction model and a DNA sequence model.
Public Scientific Summary

Markov chains are a mathematical tool used to model phenomena that change over time. They have applications in various fields like physics, chemistry, economics, biology, and computer science. This thesis aims to introduce the theory and applications of Markov chains comprehensively. It covers the basic concepts, such as the Markov property and transition matrix, which help calculate probabilities of transitioning between states. The thesis also discusses representing Markov chains as graphs and classifying states to understand the long-term behavior of Markov chains. The importance of the stationary distribution, which represents long-term behavior, is explored. Additionally, the thesis delves into simulating Markov chains with multiple states, showcasing practical examples like weather forecasting and genetic analysis. Overall, this thesis provides a thorough understanding of Markov chains and their practical applications.
Chapter 1

Introduction

Markov chain is a powerful mathematical tool for modeling a variety of phenomena that evolve over time. These chains are named after Andrey Markov, a Russian mathematician who developed the theory of stochastic processes in the early 1900s. In [Sen96], the origin of Markov chains is presented in a wonderful way, and in [Gag17], historical notes are also presented.

The purpose of this thesis is to provide an introduction to the theory and the application of Markov chains. Markov chains have been widely used in various fields, including physics, chemistry, economics, biology, and computer science, to name a few.

One of our initial goals is to conduct an extensive review of the bibliographic references [TK98], [CN06], [Spe17], [SC97], [Ros07], [Ros95], [Kul16], [Pri13], [Gal11], [Sen96] and [Gag17], to present a comprehensive study of Markov chains with finite state spaces. We also intend to provide a detailed analysis of the various concepts and techniques involved in the theory and the application of Markov chains with finite state spaces. Through this review, we intend to provide a comprehensive understanding of the subject matter.

The first chapter of this thesis consists of the introduction, and the second chapter shows a comprehensive overview of the basic concepts of Markov chains. In particular, we will begin by introducing the Markov property. We will then look at the transition probability matrix, which allows us to calculate the probabilities of moving from one state to another state in a Markov chain. We will also introduce higher order transition probability, which plays a crucial role in predicting the behavior of complex Markov chains.

Next, we will focus on the concept of Markov chains as graphs, which provide a graphical representation of the behavior of the chain. This concept facilitates the visualization and analysis of Markov chains, making it an essential tool in the field. In addition, we will be discussing how to classify the states in a Markov chain, which is a fundamental concept in the analysis of the Markov chains. The classification of states allows us to understand the long-term behavior of the Markov chain and gives us insight into the properties of the chain, thus, the concepts introduced in this chapter are critical to understanding the
theory and applications of Markov chains and their importance cannot be overstated. A full understanding of these concepts is essential for anyone studying the subject, and we will aim to provide a thorough analysis of these concepts in this chapter.

The third chapter of this thesis is dedicated to the discussion of the stationary distribution and its importance in probability theory. We begin by introducing the concept of stationary distribution, which is one of the fundamental ideas in the context of Markov chains. The stationary distribution represents the long-term behavior of the Markov chain and provides us with valuable insights into the properties of the chain.

We will also look at the uniqueness and the existence of stationary distribution. The uniqueness and existence of the stationary are crucial concepts during the analysis of Markov chains. We aim to provide a comprehensive analysis of these concepts, highlighting their importance and discussing their implications for the study and application of Markov chains.

Finally, in the fourth chapter, we look at the simulation of Markov chains, which is a crucial concept in the domain of Markov chains. Markov chains are widely used for modeling a wide range of systems, from financial applications to social network analysis. Simulation of Markov chains is a powerful tool for predicting future states based on historical data.

To illustrate this technique, we present two examples that will be used throughout the chapter. In the first example, we simulate a Markov chain representing a weather forecasting model for a city. This example gives us valuable insights into how Markov chain simulation can be used to model and predict weather patterns based on historical data.

In the second example, we simulate a Markov chain that models the sequence of nucleotides in the DNA chain. This example gives us valuable insights into how the simulation of Markov chains can be used to analyze and predict genetic mutations and patterns.

By presenting these examples, we aim to deepen the reader’s understanding of Markov chain simulation and its importance in modelling and prediction. The concepts introduced in this chapter are critical to understanding the practical applications of Markov chains.
Chapter 2

Markov Chains

In this chapter we begin with the study of discrete-time Markov chains, emphasizing the Markov property and transition probability matrices. We will also introduce the concepts of communication and irreducibility in a Markov chain. Finally, such chains will be examined in terms of periodicity and aperiodicity. These basic concepts will be essential for analyzing the long-term behavior of a Markov chain in the following chapter.

2.1 Markov Property

Stochastic processes can be described using Markov chains, mathematical models in which future states are predicted only by their present state, not by their past state.

We consider \( Y_1, Y_2, \ldots, Y_n, \ldots \) to be a sequence of random variables that take values in a discrete-time finite-state \( S \). That is, \( S \) consists of a finite set of states \( 1, 2, \ldots, M \).

**Definition 1.** ([Gal11],[Sen96],[Kul16]) A sequence of random variables \( Y_1, Y_2, \ldots, Y_n, \ldots \) in a discrete state space \( S \) referred to as Markov chain, or to have a Markov property if, for all \( n \geq 1 \), the probability distribution of \( X_{n+1} \) depends only on the value of \( X_n \). Thus, for all \( n \geq 1 \) and all \( i, j \in S \), the following equality holds:

\[
P(Y_{n+1} = j | Y_n = i, Y_{n-1} = y_{n-1}, Y_{n-2} = y_{n-2}, \ldots, Y_0 = y_0) = P(Y_{n+1} = j | Y_n = i).
\]

Alternatively, one can use the first-order transition probabilities for the computation of the joint probability distribution:

\[
P(Y_n = i_n, Y_{n-1} = i_{n-1}, \ldots, Y_0 = i_0) = P(Y_n = i_n | Y_{n-1} = i_{n-1}) \cdots P(Y_1 = i_1 | Y_0 = i_0) P(Y_0 = i_0),
\]

where \( i_0, i_1, \ldots, i_n \in S \).
Also by using the law of total probability, we will have

\[
P(Y_n = i) = \sum_{j \in S} P(Y_n = i, Y_0 = j)
\]

\[
= \sum_{j \in S} P(Y_n = i \mid X_0 = j) P(Y_0 = j), \quad i \in S.
\]

2.2 Transition Probability Matrix

As we have seen in the last section, according to the Markov property, the evolution of a Markov chain can be defined by its first-order transition probabilities

\[
P_{ij}(n) = P(Y_n = j \mid Y_{n-1} = i).
\]

\[
P_{ij}(n) = \text{probability of moving from state } i \text{ to } j, \text{ at time } n.
\]

The notation shows that transition probabilities are not only functions of the first and last states but also functions of the time of transition.

However, if the one-step transition probabilities are independent of the time variable \( n \), we can say that the Markov chain is time-homogeneous, thus if a time homogeneous Markov chain is in the state \( i \) at time \( n \), it goes to state \( j \) at time \( n + 1 \) with the probability \( P_{ij} \), for all values of \( n \). We limit our discussion to this case. In this situation, the probabilities \( P_{ij} \) can be represented by a matrix, which is called the transition matrix of the Markov chain, which can be written as follows ([Kul16])

\[
P = \begin{pmatrix}
P_{11} & P_{12} & P_{13} & \ldots & P_{1M} \\
P_{21} & P_{22} & P_{23} & \ldots & P_{2M} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
P_{M1} & P_{M2} & \ldots & P_{MM}
\end{pmatrix}.
\]

Note the rows of the transition matrix satisfies the condition

\[
\sum_{j \in S} P_{ij} = \sum_{j \in S} P(Y_n = j \mid Y_{n-1} = i)
\]

\[
= \sum_{j \in S} \frac{P(Y_n = j, Y_{n-1} = i)}{P(Y_{n-1} = i)}
\]

\[
= \frac{P(Y_{n-1} = i)}{P(Y_{n-1} = i)} = 1
\]

for all \( i \in S \).

Remark. In this thesis, the Markov chains we will study are homogeneous, and we will not mention this property from now on.


2.3 Higher Order Transition Probabilities

In the previous section, we introduced the first-order transition probabilities, defined as

\[ P_{ij} = \mathbb{P}(Y_n = j \mid Y_{n-1} = i), \]

which represents the probability of transitioning from state \( i \) to state \( j \) in one step.

Now, in this section, we will consider higher order transition probabilities, defined as

\[ P_{ij}^{(n)} = \mathbb{P}(Y_n = j \mid Y_0 = i), \]

which represent the probability of transitioning from state \( i \) to \( j \) in \( n \) steps, starting from time 0.

Higher order transition probabilities are helpful in analyzing the long term behavior of a Markov chain.

Now, if I consider our state space to be \( S = \{1, 2, \ldots, M\} \), we can introduce a convenient notation to represent the \( n \)-step transition probabilities in matrix form. We define the \( n \)-step transition matrix as \( P^{(n)} \), where the entry \((i, j)\) represents the probability of transitioning to state \( j \) after \( n \) steps starting from state \( i \). Formally, in matrix form:

\[
P^{(n)} = \begin{bmatrix}
P^{(n)}_{11} & P^{(n)}_{12} & \cdots & P^{(n)}_{1M} \\
P^{(n)}_{21} & P^{(n)}_{22} & \cdots & P^{(n)}_{2M} \\
\vdots & \vdots & \ddots & \vdots \\
P^{(n)}_{M1} & P^{(n)}_{M2} & \cdots & P^{(n)}_{MM}
\end{bmatrix}.
\]

We can express \( P^{(n)} \) in terms of the one step transition matrix, \( P \), [Kul16].

**Theorem 1.** ([Ros07]) We have

\[
P^{(n)} = P^n.
\]

The introduction of the \( n \)-step transition matrix allows us to analyze the long-term transition probability of the Markov chain.

Now, from the first section, we know that:

\[
\mathbb{P}(Y_n = i) = \sum_{j \in S} \mathbb{P}(Y_n = i, Y_0 = j) \\
= \sum_{j \in S} \mathbb{P}(Y_n = i \mid Y_0 = j) \mathbb{P}(Y_0 = j)
\]

So that, we obtain
\[
(\mathbb{P}(Y_n = 1), \ldots, \mathbb{P}(Y_n = M)) = (\mathbb{P}(Y_0 = 1), \ldots, \mathbb{P}(Y_0 = M)) = \begin{bmatrix}
P^{(n)}_{11} & P^{(n)}_{12} & \cdots & P^{(n)}_{1M} \\
P^{(n)}_{21} & P^{(n)}_{22} & \cdots & P^{(n)}_{2M} \\
& \ddots & \ddots & \ddots \\
& & P^{(n)}_{M1} & P^{(n)}_{M2} & \cdots & P^{(n)}_{MM}
\end{bmatrix}.
\]

Note that, the vector on the left-hand represents a distribution of the \( X_n \), and the vector next to \( n \)-step transition matrix is the initial distribution, which is the distribution of \( X_0 \).

To simplify, we will denote these vectors by \( \pi^{(n)} \) and \( \pi^{(0)} \), respectively, for \( n = 1, 2, \ldots \).

Using this notation, we can write the previous formula as follows:

\[
\pi^{(n)} = \pi^{(0)} P^n.
\]  

In this work, we will study the conditions under which the convergence of \( \pi^{(n)} \) does not depend on an initial distribution and converges to a unique limit. Specifically, we will investigate the properties of the \( n \)-step transition matrix and its relationship to the long-term behavior of the Markov chain. That is, if the \( n \)-step transition matrix satisfies certain conditions, then the distribution of \( X_n \) converges to a unique limit as \( n \to \infty \), regardless of the initial distribution. This limit is known as the stationary distribution of the Markov chain, and it plays an important role in characterizing the long-term behavior of the chain.

Overall, our goal is to develop a deeper understanding of the convergence properties of Markov chains and the role played by the \( n \)-step transition matrix and the stationary distribution.

### 2.4 Markov Chains as Graphs

By a transition matrix of the Markov chain, a directed graph can be constructed that visually represents the state transitions. Each node in this graph represents a state and in this graph, the directional arrows from one node to another represent the transition probabilities between states.

Specifically, for each state \( i \) in the Markov chain, a node \( i \) is drawn in the graph. Then, for each pair of states \( i \) and \( j \) where the transition probability from \( i \) to \( j \) is greater than zero, a directed arrow is drawn from node \( i \) to node \( j \), labeled with the corresponding transition probability. If the transition probability from \( i \) to \( j \) is zero, no arrow is drawn between nodes \( i \) and \( j \).

An example of how a transition matrix can be associated with a graph is the following transition probability matrix:

\[
P = \begin{pmatrix}
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0.3 & 0.4 & 0 & 0.3 \\
0 & 0 & 1 & 0
\end{pmatrix}.
\]

The corresponding graph for this transition matrix is shown below:
In this graph, each node represents a state in a Markov chain, and a directed arrow represents the transition probabilities between states. For example, the arrow going from state 1 to 2 indicates that the transition probability from one to two is 1. The arrow going from state 3 to state 2 indicates the transition probability from state 3 to state 2, which is 0.4.

2.5 Classification of States

Definition 2. ([Gal11],[Ros07],[Pri13]) A state \( m \in S \) is said to be accessible from \( l \in S \) denoted as \( l \rightarrow m \), if there exists a finite integer \( n \geq 0 \) such that the probability of moving from \( l \) to \( m \) in \( n \) steps is positive:

\[
P^n_{lm} = \mathbb{P}(Y_n = m \mid Y_0 = l) > 0.
\]

We can say that, it is possible to go from \( l \) to \( m \) with non-zero probability in a certain random number of steps.

Additionally, since \( P^0 = I_d \), the definition of accessibility says that \( l \) is accessible from \( l \), for every state \( l \in S \).

Definition 3. We can say that \( l \) and \( m \) communicate with each other and we can write \( l \leftrightarrow m \) if \( l \rightarrow m \) and \( m \rightarrow l \).

This binary relation " \( \leftrightarrow \) " satisfies the following attributes:

(i) Reflexive: We have \( l \leftrightarrow l \) for all \( l \in S \).

(ii) Symmetric: We have that \( l \rightarrow m \) is equivalent to \( m \rightarrow l \) for all \( l, m \in S \).

(iii) Transitive: For all \( l, m, k \in S \) such that \( l \rightarrow m \) and \( m \rightarrow k \), we have \( l \rightarrow k \).

Thus the relation \( \leftrightarrow \) is called equivalence relation and it induces a partition of the state space \( S \) into disjoint subsets \( A_1, \ldots, A_m \) which can be written as \( S = A_1 \cup \cdots \cup A_m \), and

- we have \( l \leftrightarrow m \) for all \( l, m \in A_q \), and
- we have \( l \leftrightarrow m \) whenever \( l \in A_p \) and \( m \in A_q \) with \( p \neq q \).
The sets denoted by $A_1, \ldots, A_m$ are referred to as the communicating classes of the Markov chain.

**Irreducible Markov chain**

**Definition 4.** ([Ros95],[Pri13]) If in a Markov chain, a state space has only one communication class, it is called irreducible. On the other hand, if the chain has multiple communicating classes, it is said to be reducible.

**Example 1.** Suppose we have a Markov chain with four states that can be represented by the following transition probability matrix:

$$
P = \begin{pmatrix}
0 & 0.3 & 0.7 & 0 \\
0 & 0.2 & 0.2 & 0.6 \\
0.5 & 0 & 0 & 0.5 \\
0 & 0 & 0 & 1
\end{pmatrix}.
$$

(2.3)

We need to represent the transition probabilities using the following graph:

Now we can conclude by observing the graph that the state space $S = \{1, 2, 3, 4\}$ can be partitioned into two classes, namely $A = \{1, 2, 3\}$ and $B = \{4\}$, and consequently, the Markov chain is not irreducible.

**Recurrent and transient states of Markov chains**

**Definition 5.** ([Gal11],[Ros07]) **Recurrent state**: In a finite-state Markov chain, if a state $i$ is accessible from all states then it is called a recurrent state. In other words, if $i$ is recurrent, it means it can be accessed from any state.

**Transient state**([Ros07]): In a Markov chain, a state is either recurrent or transient. In other words, if a state in a Markov chain is not recurrent then we can say that the state is a Transient state.

Note that in a finite-state Markov chain, a state is considered recurrent if there is always a possibility of going to a state from which it can be returned and a state is transient if there exists a state $j$ that can be accessible from $i$ but from which there is no possibility to return.
For example 1, \( \{1, 2, 3\} \) and \( \{4\} \) are recurrent classes.

We know from [Gal11] the following result:

**Theorem 2.** For finite-state Markov chains, either all states in a class are recurrent or all are transient. Moreover, in every finite-state Markov chain, there exists at least one recurrent class of states.

### Periodicity and Aperiodicity

In this section, we introduce the concepts of periodicity and aperiodicity of Markov chains. These are fundamental properties of Markov chains that determine their long-term behavior and are essential for their analysis. The understanding of these concepts and their implications for the behavior of finite state Markov chains will be studied in the upcoming chapter. This knowledge will be invaluable in analyzing the behavior of Markov chains and making predictions about their future evolution.

**Definition 6.** For a state \( i \in S \) the period state \( i \) is equal to the greatest common divisor of \( \{n \geq 1 : P^n_{ii} > 0\} \). A state of period equal to 1 is said to be aperiodic, which is the case in particular if \( P_{ii} > 0 \).

**Definition 7.** An aperiodic Markov chain is a chain where every state is aperiodic.

**Example 2.** Suppose we have a finite state Markov chain of four states, denoted as 1, 2, 3 and 4, which can be represented by the following transition probability matrix:

\[
P = \begin{pmatrix}
0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 \\
0.5 & 0 & 0 & 0.5 \\
0 & 1 & 0 & 0
\end{pmatrix}.
\tag{2.4}
\]

To determine the sets \( \{n \geq 1 : P^n_{ii} > 0\} \), we need to represent the transition probabilities using the following graph:

![Graph](image)

In the above graph, the curve that leaves the circle indicates the movement of the initial state to make the transition.

Now, it can be observed from the graph that:
2.6. ERGODIC CHAIN

\[
\{ n \geq 1 : P^n_{11} > 0 \} = \{2, 4, 6, 8, \cdots \}
\]
\[
\{ n \geq 1 : P^n_{22} > 0 \} = \{4, 6, 8, \cdots \}
\]
\[
\{ n \geq 1 : P^n_{33} > 0 \} = \{2, 4, 6, 8, \cdots \}
\]
\[
\{ n \geq 1 : P^n_{44} > 0 \} = \{4, 6, 8, \cdots \}
\]

hence all states have period 2.

**Example 3.** Suppose we have a Markov chain of states, 1, 2, 3 and 4, and the following is the transition probability matrix:

\[
P = \begin{pmatrix}
0.5 & 0.5 & 0 \\
0 & 0.5 & 0.5 \\
0.5 & 0 & 0.5
\end{pmatrix}
\]  

(2.5)

In this case, by observing the diagonal of the transition matrix, we conclude that the Markov chain is aperiodic, and from the following graph, we conclude that it is irreducible.

![Graph of a Markov chain](image)

We know from ([Gal11]) the following result:

**Theorem 3.** In a Markov chain (with either a countably infinite or finite number of states), in the same class all states have the same period.

This result gives us a criterion to determine the period of the states in a class, for which it is sufficient to elaborate the periodicity of the state in the class.

### 2.6 Ergodic chain

Note that for a finite-state chain, every class of states can be categorized by its periodicity and in terms of whether it is recurrent or not recurrent. There is a significant case in which a class has both properties recurrence and aperiodicity.

**Definition 8.** In a finite-state Markov chain, an ergodic state class combines both recurrence and aperiodicity. If a finite-state Markov chain consists of entirely one ergodic class is known as an ergodic chain, [Gal11],[Ros07].

Example 3, above, provides us with an ergodic Markov chain.
Remark. We remember that every finite-state Markov chain has at least one recurrent state class. Furthermore, if the Markov chain is irreducible, then the chain as a whole is recurrent. Therefore, a Markov chain which is irreducible and aperiodic is equivalent to an ergodic Markov chain.
Chapter 3

Stationary Distribution and its Significance

In this chapter, we focus on the stationary distribution for a finite-state Markov chain. We will present the formal definition and study their properties ([TK98], [Ros07], [Pri13], [Gal11]). Next, we will present some applications.

3.1 The stationary distribution

The stationary distribution is important in probability theory because it provides a way of describing the long-term behavior of a stochastic process that evolves over time. In particular, for a Markov chain, the stationary distribution represents the distribution of states that the chain will tend to visit repeatedly over time, regardless of the initial state. Another aspect of the importance of stationary distribution is that it allows us to make predictions about the behaviour of a system in the long run. For example, if a Markov chain represents a physical system, the stationary distribution can tell us what the long-term behavior of the system will be, such as the proportion of time it will spend in each state. In addition, the stationary distribution can be used to solve optimisation problems by providing a way to find the optimal allocation of resources or decision strategies that maximise or minimise certain criteria in the long run.

We know from the previous chapter, the formula:

\[ \pi_n = \pi_0 P^n = \pi_0 P^n, \quad n = 1, 2, \ldots \]

With

\[ \pi_n = (P(Y_n = 1), P(Y_n = 2), \ldots, P(Y_n = M)) \],

\[ \pi_0 = (P(Y_0 = 1), P(Y_0 = 2), \ldots, P(Y_0 = M)) \].
And

\[ P^n = \begin{bmatrix} P^{(n)}_{11} & P^{(n)}_{12} & \cdots & P^{(n)}_{1M} \\ P^{(n)}_{21} & P^{(n)}_{22} & \cdots & P^{(n)}_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ P^{(n)}_{M1} & P^{(n)}_{M2} & \cdots & P^{(n)}_{MM} \end{bmatrix}. \]

The purpose of this section is to demonstrate now that if there exists a probability distribution \( \pi \) such that \( \pi_n \to \pi \) (as row vector), then

\[ \pi_0 P^{n+1} = \pi_0 P^n P \to \pi P. \]

From which we can conclude that \( \pi P = \pi \).

With this observation in mind, we will introduce the definition of a stationary distribution of a Markov chain ([TK98], [Ros07], [Pri13], [Gal11]):

**Definition 9.** Let \( \{Y_n\}_{n \geq 0} \) be a Markov chain with a state space \( S = \{1, 2, \cdots, M\} \), with transition probability matrix \( P \). A probability distribution \( \pi \) on \( S \) is said to be a stationary distribution of the Markov chain if it satisfies the following conditions:

- \( \pi(j) \geq 0 \) \( \forall j \in S \).
- \( \sum_{j \in S} \pi(j) = 1 \).
- \( \pi \) satisfies

\[ \pi P = \pi, \quad (3.1) \]

where \( \pi P \) is the vector obtained by multiplying \( \pi \) and \( P \).

Previously, we showed that if \( P^n \) converges to a matrix where all rows are \( \pi \), then \( \pi \) is a stationary distribution for a Markov chain. Now we consider the case where \( \pi = (1/2, 1/2) \) and \( P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \).

So \( \pi P = \pi \). However, we know from [TK98] that

\[ P^n = \frac{1}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} + \frac{(-1)^n}{2} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} \]

and \( P^n \) does not converge as \( n \to \infty \).

This example leads to the following question: Under what conditions does each row of \( P^n \) converge to a probability vector that satisfies \( \pi = P\pi \)?

We will answer this question in the following section. Let us consider some examples that show how to obtain the stationary distribution.
Example 4. Consider a simple example of a two-state Markov chain, where the transition probabilities are defined by the subsequent matrix:

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

This Markov chain models a system that can be in one of two possible states, labeled as state 1 and state 2. The transition probabilities indicate the probability of transitioning from one state to another within one-time step.

We represent the transition probabilities using the following graph:

```
1 ---- 0.3 0.7 ---- 2
    |     |     |
    |     | 0.6  |
    |     |     |
    v     v
  0.4   0.5
```

Note that \( P_{11} > 0 \) and \( P_{22} > 0 \), so that the chain is aperiodic. Also, we can conclude by observing the graph that the state space \( S = \{1, 2\} \) is comprised of a single class and consequently, the Markov chain is irreducible.

To find the stationary distribution, we need to solve the equation \( \pi P = \pi \) for the unknown vector \( \pi \). This is equivalent to finding the eigenvector of \( P \) corresponding to the eigenvalue of 1.

By solving this equation, we found that the unique stationary distribution of this Markov chain is given by:

\[ \pi = (0.5714 \ 0.4286) \]

That shows in the long run, the Markov chain will spend approximately 57.14% of the time in state one and 42.86% of the time in state two, regardless of the starting state.

Example 5. Suppose we have a production line consisting of three machines, labeled as M1, M2, and M3. Each machine can be in one of two states: "operating" or "idle". The production line functions as a Markov chain, where the current state depends only on the previous state.

Suppose that the transition matrix of the Markov chain is as follows:

\[ P = \begin{pmatrix} 0.7 & 0.2 & 0.1 \\ 0.3 & 0.5 & 0.2 \\ 0.2 & 0.3 & 0.5 \end{pmatrix}. \]

In this matrix, the rows represent the current state of the production line, and the columns represent the next state. For example, the entry \( P_{1,2} \) (the second entry of the first row)
represents the probability of the production line transitioning from state 1 (M1 operating, M2 idle, M3 idle) to state 2 (M1 idle, M2 operating, M3 idle) in a single step.

We represent the transition probabilities using the following graph:

![Graph showing transition probabilities between states 1, 2, and 3.]

We can conclude by observing the graph that the state $S = \{1, 2, 3\}$ is comprised of a single class and consequently, this Markov chain is aperiodic and irreducible.

To determine the stationary distribution of this Markov chain, we must solve the equation $\pi P = \pi$, where $\pi$ represents the stationary distribution. We solve numerically and find the stationary distribution of a Markov chain.

The result will be a vector with the values of the stationary distribution. In this case, the result obtained is the vector $v = (0.4634146 \ 0.3170732 \ 0.2195122)$.

This means that in the long run, the production line will spend approximately 46.34146% of the time in each of the three possible states, regardless of the initial state.

This stationary distribution can be useful in decision-making in the manufacturing sector. For example, if maximizing production is desired, the production line configuration could be adjusted so that idle times are as short as possible, thereby reducing the probability of the production line being in an idle state.

### 3.2 Existing and uniqueness of the stationary distribution

In this section, we will study the properties that a Markov chain must have in order for a stationary distribution to exist and be unique.

We return to the question: Under what conditions does each row $P^n$ converge to a probability vector that satisfies $\pi = \pi P$? To answer this question, we will follow [Gal11] and introduce some preliminary results that are needed.
3.2. EXISTING AND UNIQUENESS OF THE STATIONARY DISTRIBUTION

The property of ergodicity for Markov chains given in the previous chapter will be the key to answering this question.

The most simple approach to answer the question, concerning each column \( j \) of \( P^n \), is to examine the variation between the smallest and largest element of that column and how this evolves with \( n \). This study is initiated by the following theorem which is true for all finite-state Markov chains.

**Theorem 4.** ([Gal11]) Suppose \( P \) is the transition probability matrix of a finite-state Markov chain. Let \( P^n \) is the \( n \)-th power of \( P \). Then for every state \( j \) and every integer \( n \geq 1 \) we have

\[
\max_l P^n_{lm} + 1 \leq \max_l P^n_{lm} \quad \text{and} \quad \min_l P^n_{lm} + 1 \geq \min_l P^n_{lm}
\]

The theorem states that for each column denoted as \( j \), the smallest of the elements is non-decreasing with \( n \) and the largest element is non-increasing with \( n \). Although the specific elements within a column that form the minimum and maximum can change as \( n \) changes, the range that is covered by those elements is nested in \( n \), either staying the same or shrinking as the \( n \to \infty \). The proof of this theorem can be seen in [Gal11].

**Theorem 5.** ([Gal11]) Suppose the transition matrix of a finite-state Markov chain satisfies \( P > 0 \) (i.e., \( P_{lm} > 0 \) for all \( l, m \)), and let \( \alpha = \min_{l,m} P_{lm} \). Then for all states \( m \) and all \( n \geq 1 \):

\[
\max_l P^n_{lm} - \min_l P^n_{lm} \leq \left( \max_l P^n_{lm} - \min_l P^n_{lm} \right) (1 - 2\alpha) \leq (1 - 2\alpha)^n.
\]

\[
\lim_{n \to \infty} \max_l P^n_{lm} = \lim_{n \to \infty} \min_l P^n_{lm} > 0
\]

Since \( P_{lm} > 0 \) for all \( l, m \), we have \( \alpha > 0 \). Thus for every \( m \), the elements \( P^n_{lm} \) in column represented by \( m \) of \( [P^n] \) approach equality over both \( n \) and \( l \) as \( n \to \infty \), i.e., at time step \( n \), the state becomes independent of the state at time 0 as \( n \to \infty \). The approach is exponential in \( n \).

Next, we need the following theorem.

**Theorem 6.** For a finite-state ergodic Markov chain with space state \( S = \{1, 2, \cdots, K\} \), \( P_{lm}^n > 0 \) for all \( l, m \), and all \( m \geq (K - 1)^2 + 1 \).

The demonstration of Theorems 5 and 6 above is also in [Gal11].

The upcoming theorem is the most important result of this thesis.

**Theorem 7.** ([Gal11]) If \( P \) represents the transition matrix of a finite-state ergodic Markov chain. Then there exist a unique stationary distribution \( \pi \), such that

\[
\lim_{n \to \infty} P^n_{ij} = \pi_j \quad \text{for each} \ i, j.
\]
Proof:

First, note that Theorem 5 extends easily to any finite-state ergodic Markov chains. This key concept comes from Theorem 6, which represents that if $P$ is the probability matrix for an ergodic $K$ state Markov chain, then the matrix $P^q$ is positive for any $q \geq (K - 1)^2 + 1$. Thus, by choosing $q = (K - 1)^2 + 1$, we apply Theorem 5 to $P^q > 0$. For every integer $\nu \geq 1$,

$$\max_i P^{(\nu+1)}_{ij} - \min_i P^{(\nu+1)}_{ij} \leq \left( \max_m P^{\nu}_{mj} - \min_m P^{\nu}_{mj} \right) (1 - 2\beta)$$

$$\lim_{\nu \to \infty} \max_m P^{\nu}_{mj} = \lim_{\nu \to \infty} \min_m P^{\nu}_{mj} > 0$$

where $\beta = \min_{i,j} P^q_{ij}$. Theorem 4 tells that $\max_i P^{n+1}_{ij}$ which is not decreasing in $n$, so the previous inequality and formula can be replaced by a limit in $n$, yielding

$$\left( \max_m P^n_{mj} - \min_m P^n_{mj} \right) \leq (1 - 2\beta)^{\lfloor n/q \rfloor}$$

$$\lim_{n \to \infty} \max_m P^n_{mj} = \lim_{n \to \infty} \min_m P^n_{mj} > 0.$$ 

Now let $\pi > 0$ by

$$\pi_j = \lim_{n \to \infty} \max_m P^n_{mj} = \lim_{n \to \infty} \min_m P^n_{mj} > 0. \quad (3.2)$$

Since $\pi_j$ falls within the minimum and maximum $P^n_{ij}$ for each $n$,

$$|P^n_{ij} - \pi_j| \leq (1 - 2\beta)^{\lfloor n/q \rfloor}.$$

Then in the limit,

$$\lim_{n \to \infty} P^n_{ij} = \pi_j \quad \text{for each } i, j. \quad (3.3)$$

This tells that $P^n$ has a limit as $n \to \infty$ and the $i, j$ term of that matrix is $\pi_j$ for all $i, j$. In matrix notation, this means that:

$$\lim_{n \to \infty} P^n = \begin{pmatrix} \pi_1 & \cdots & \pi_M \\ \vdots & \ddots & \vdots \\ \pi_1 & \cdots & \pi_M \end{pmatrix}.$$ 

Each row of this limiting matrix is identical and is the vector $\pi$. This can be represented by

$$\lim_{n \to \infty} P^n = e\pi \quad \text{where } e = (1, 1, \ldots, 1)^t.$$
We will show, that $\pi$ is a unique stationary distribution. Consider $\mu$ is any stationary distribution, i.e., any probability vector solution to $\mu P = \mu$. Then $\mu$ must be satisfy $\mu = \mu P^n$ for all $n > 1$. The following limit,

$$\mu = \mu \lim_{n \to \infty} P^n = \mu e \pi = \pi,$$

proves the required result.
Chapter 4

Simulation

In this chapter, we will introduce the simulation of finite-state Markov chains. Markov chains are an essential tool in modeling a wide range of systems, from financial applications to social network analysis. To illustrate this technique, we will first present two examples that will be used throughout the chapter ([CN06], [Gag17], [Ros07]).

In the first example, we will simulate a Markov chain that represents a weather prediction model for a city. This example will allow us to see how the simulation of Markov chains can be used for modeling and predicting weather patterns based on historical data.

In the second example, we will use a Markov chain simulation to create a model of the sequence of nucleotides in the DNA chain. This example will allow us to see how the simulation of Markov chains can be used to analyze and predict genetic mutations and patterns.

**Example 6. Weather prediction model for a city**

As a first example, we consider the modelling of the climate in city A. For simplicity, we assume that there are only two states: $E = \{s_1, s_2\}$, $(s_1 = \text{rainy}, s_2 = \text{sunny})$, and that the sequence of weather states can be modeled with a Markov chain. The weather of the city can be modelled with a Markov chain $(Y_0, Y_1, \ldots)$, with a transition probability matrix

$$P = \begin{pmatrix} 0.5 & 0.5 \\ 0.1 & 0.9 \end{pmatrix}.$$  

This means that $P(Y_{n+1} = \text{sunny}|Y_n = \text{rainy}) = 0.5$ and $P(Y_{n+1} = \text{rainy}|Y_n = \text{sunny}) = 0.1$, for all $n$.

Starting from an initial uniform distribution in the state space, we propose to simulate a sequence of $n = 1000$ states under the assumed model.

**Example 7. Model for the sequence of nucleotides in the DNA chain.** A Markov chain can be used for modeling the sequence of nucleotides in DNA. The state space is then $E = \{A, C, G, T\}$, where $(A = \text{adenine}, C = \text{cytosine}, G = \text{guanine}, T = \text{thymine})$, and as
4.1 SIMULATION ALGORITHM

an example, the transition matrix is used

\[
P = \begin{pmatrix}
0.2 & 0.25 & 0.4 & 0.15 \\
0.1 & 0.6 & 0.1 & 0.2 \\
0.2 & 0.1 & 0.35 & 0.35 \\
0.1 & 0.45 & 0.2 & 0.25
\end{pmatrix}.
\]

Using a uniform initial distribution, \( n = 1000 \) observations of this chain are simulated.

4.1 Simulation Algorithm

In the following we show how to simulate the chain with a finite number of states ([CN06], [Gag17], [Ros07]). We start by simulating the initial distribution \( \mu \) with the function defined by

\[
\phi_{init}(\mu, u) = \begin{cases}
  s_1 & \text{for } u \in [0, \mu_1] \\
  s_2 & \text{for } u \in [\mu_1, \mu_1 + \mu_2] \\
  \vdots & \\
  s_j & \text{for } u \in [\sum_{l=1}^{j-1} \mu_l, \sum_{l=1}^{j} \mu_l] \\
  \vdots & \\
  s_m & \text{for } u \in [\sum_{l=1}^{m-1} \mu_l, 1]
\end{cases}
\]

where \( u \sim U[0, 1] \). This algorithm calculates the inverse of the distribution function for a discrete random variable.

In the following procedure, we can see how the new step is updated, when the chain is in a state \( s_i \), we define the function for each \( s_i \in S \) by

\[
\phi_{update}(s_i, P, u) = \begin{cases}
  s_1 & \text{for } u \in [0, P_{i,1}] \\
  s_2 & \text{for } u \in [P_{i,1}, P_{i,1} + P_{i,2}] \\
  \vdots & \\
  s_j & \text{for } u \in [\sum_{l=1}^{j-1} P_{i,l}, \sum_{l=1}^{j} P_{i,l}] \\
  \vdots & \\
  s_m & \text{for } u \in [\sum_{l=1}^{m-1} P_{i,l}, 1]
\end{cases}
\]

where \( u \sim U[0, 1] \).

Therefore, we have a complete procedure for simulating a Markov chain starting with initial distribution \( \mu \) and transition probability matrix \( P \):

1. make \( X_0 = \phi_{init}(\mu, u_0) \).
2. make \( X_1 = \phi_{update}(X_0, P, u_1) \), \( X_2 = \phi_{update}(X_1, P, u_2) \), \ldots

We write an algorithm in Phyton, for each of the examples, using the procedure described above. The figures that follow illustrate the results obtained.
1. **Weather prediction model for a city.** In the Figure 4.1 we can see the relative frequencies of visits of each state. We note that from this model with 50 days rainy days, we have 16% sunny days and 85% rainy days. With this we conjecture that the invariant measure is $(1/6,5/6)$.

![Figure 4.1: Weather prediction model for a city](image)

2. **Sequence of nucleotides in the DNA chain.** In the Figure 4.2 and 4.3 we can see the Relative frequencies of the number of visits to each state and the trajectory of the chain.

![Figure 4.2: Relative frequencies of nucleotides in the DNA chain](image)
Figure 4.3: Model for the sequence of nucleotides in the DNA chain
Conclusions

In this thesis, we have provided a comprehensive study of discrete-time finite-state Markov chains, covering basic concepts such as the Markov property, the transition matrix, higher-order transition probabilities, and the stationary distribution. We have also studied the convergence properties of Markov chains, with particular emphasis on the role of the n-step transition probability matrix and the stationary distribution in characterizing the long-term behavior of the Markov chain.

We have also presented simulation examples of M-state Markov chains, demonstrating the practical applications of this technique in weather forecasting and DNA sequence modelling. Through these examples, we have shown how Markov chain simulation can be used to analyse and predict patterns in complex systems.

In terms of future work, one possible avenue is to explore more advanced topics in Markov chain theory, such as continuous-time Markov chains and Markov decision processes. In addition, further research could be carried out on the simulation of Markov chains, with a focus on improving the accuracy and efficiency of simulation algorithms.

There is also an opportunity to explore the use of Markov chains in machine learning applications, including reinforcement learning and natural language processing. Finally, exploring the use of Markov chains in other fields such as finance and economics could also be an interesting area of research.
Bibliography


