

Treball final de grau

GRAU DE MATEMÀTIQUES

Facultat de Matemàtiques
Universitat de Barcelona

MARKOV CHAINS

Autor: Anna Areny Satorra

Director: Dr. David Márquez Carreras
Realitzat a: Departament de probabilitat,
lògica i estadística

Barcelona, June 27, 2016

Abstract

The goal of this project is to analyze and know in detail the Markov chains, a stochastic process characterized by the fact that the result obtained in a given stage only depends on the result obtained in the previous stage. We will study a concrete type of Markov chains, these are the discrete chains. For this, the first part of the thesis is focused on introducing the basic notation used and the basic concepts related with Markov chains, and, in the remaining sections, we will focus on some of its specific characteristics and properties.

Acknowledgements

During the last months it has been highly important the guidance and support of Dr. David Márquez, the director of this thesis. My most sincere words of thanks for his time and the confidence he has shown on me. I would also like to acknowledge the support provided by my family and friends, especially in the most difficult moments. Their unconditional support has made possible the realization of this project.

Contents

1	Introduction	1
2	Discrete time Markov chains	3
2.1	Homogeneous Markov chains	3
2.2	Basic examples	7
3	Chapman-Kolmogorov equation	10
4	Classification of states	13
4.1	Basic concepts	13
4.2	Classification of states	14
4.3	Periodicity and cyclic classes	20
4.4	Example: an analysis of the random walk on $\mathbb{Z}$	22
5	Hitting times	24
6	Distribution and measure	30
6.1	Stationary or invariant distribution	30
6.2	Positive recurrence	38
7	Ergodic theory	42
7.1	Ergodic theorem	42
7.2	Finite state space	47
8	Conclusions	52
A	Basic probability concepts	53
A.1	Probability of countable sets	53
A.2	Conditional probability and independence	54
A.3	Discrete random variables and expected value	56
B	Examples	58

1 Introduction

In the field of probability theory we find the concept of stochastic process, a kind of process that is characterized by a set of random variables that represent the evolution of a system of random values along time.

The simplest case of a stochastic process is that its results depend directly from the other previous results. This type of process is known as Markov chain and will be the area which this bachelor's degree final thesis will focus in.

One of the most important properties of a Markov chain, also called Markov property, is that the values it can take in the future only depend on its present value, not in past values, so it is said that Markov chains have no memory, due to the track incurred till present does not affect its future path.

These chains receive their name from the Russian mathematician Andrey Andreyevich Markov, who dedicated his studies to stochastic process as well as be one of the most important participants in the political and social events that occurred in Europe at the beginning of the 20th Century.

It is essential to know that Markov chains have several and useful applications in the field of the investigation. Though its applications are numerous, its importance is greater when we talk about chemistry, medicine, biology and physics, perhaps the four most important fields in the huge world of science. An easy example of these applications is the Ehrenfest chain, which allows us to parametrize the exchange of gas molecules between two bodies. However, this project is not focused on the practical applications of Markov chains, in it we make a theoretical study of discrete time Markov chains.

During the bachelor's degree we have taken probability and statistics subjects. However, the amount of time dedicated to these topics is not very high, so I found interesting to focus this thesis on these fields. After looking for several options, I decided that Markov chains would be an interesting topic due to its characteristic of lack of memory caught my attention.

Structure of the project

This project analyzes the elemental theory of discrete time Markov chains. The thesis is divided in several chapters, each of them based on a different question related with this Markov chains.

To begin with, in the first chapter we can find a brief introduction that helps us to understand the basic concepts related with Markov chains as well as introducing some of their basic properties. Additionally, the basic notation that we will find thereafter is introduced in this section.

Sometimes, in order to move from one state to another within a Markov chain we can choose among several ways. In chapter 3, we study the probability of moving between two states taking into account all the possible ways, so we will introduce the Chapman-Kolmogorov equation for this purpose.

Once introduced several useful concepts in the prior sections, in chapter 4 we will focus on studying the behavior of each one of the states of a Markov chain. In particular, we will analyze the possibility of returning to every state.

In the previous section we have studied in the hypothetical return to the original state. In chapter 5, we will also focus on arriving at a given state, but now the target is to reach a subset of states of the Markov chain, in which we may not be initially.

Finally, the last two chapters are strongly related between them. Here, we study the long term behavior of a Markov chain and the invariant distributions that the stochastic matrix of the Markov chain presents. Additionally, in chapter 6 we will see some applications of the previous distributions and in chapter 7 we will demonstrate the uniqueness of the invariant distribution.

2 Discrete time Markov chains

We already know that Markov chains are very important within the theory of probabilities. In this chapter, we begin focusing on discrete Markov chains, in particular homogeneous chains, which do not depend on time. This type of chains go with us throughout the chapter where we study one important result, which will allow us to know which chains satisfy the homogeneous property. At the end of this section we also include some basic examples about homogeneous Markov chains.

To begin this section we will give some basic definitions about Markov theory which are a great help to achieve results that are in later sections.

2.1 Homogeneous Markov chains

In this chapter we should take into account a probability space $(\Omega, \mathcal{F}, P)$ and a family of discrete random variables $X_n : \Omega \longrightarrow I$ where I is a numerable set known as state space and every $i \in I$ is a state.

Moreover, recall that every random variable $X : \Omega \longrightarrow I$ will have associated a distribution ν , in other words, ν is its law, so $\nu_i = P(X = i)$.

Definition 2.1.1. *A stochastic process is a collection of random variables which are included by an arbitrary set $\{X_t; t \in T\}$. For every t , we have $X_t : \Omega \longrightarrow \mathbb{R}$, it represents the evolution of some system of random values over time.*

We are going to study the discrete case, that is to say $T = \mathbb{Z}^+ = \{0, 1, 2, \dots\}$; as a result we have the process defined as $\{X_t; t \geq 0\}$. Moreover, we will focus on the stochastic process known as Markov chain. This process has no memory because its future behaviour is only affected by the present one.

Definition 2.1.2. *A stochastic matrix, also called transition (probability) matrix, $\Pi = (p_{i,j}; i, j \in I)$ is a matrix which satisfies the following conditions*

$$(i) \ p_{i,j} \in [0, 1], \forall i, j \in I.$$

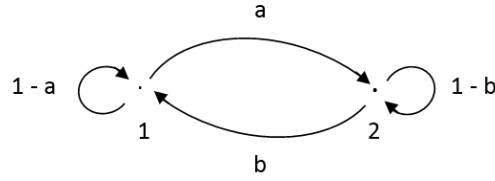
$$(ii) \ \forall i \in I, \sum_{j \in I} p_{i,j} = 1.$$

The element $p_{i,j}$ is called transition probability, that is to say, it is the probability that in the point of time k , the process is in the state j given that in the point of time $k - 1$ it was in i .

Now we will see, with the help of examples, that there is a strong relationship between diagrams and the stochastic matrix described above.

Example 2.1.3.

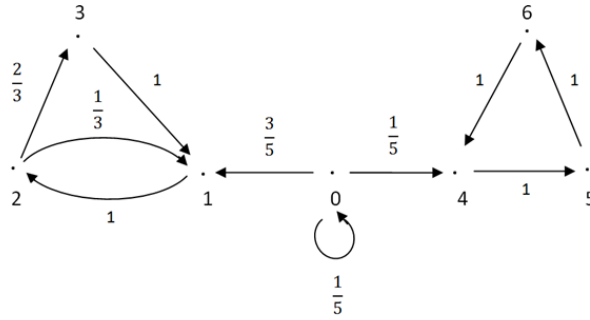
1. We start with a general example in order to clearly see that the stochastic matrix satisfies its own properties. Given $a, b \in [0, 1]$, we have the diagram



and the stochastic matrix is

$$\begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$$

2. Now, we are going to see a numerical example, in this case we have the diagram



and the stochastic matrix is

$$\begin{pmatrix} \frac{1}{5} & \frac{3}{5} & 0 & 0 & \frac{1}{5} & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & \frac{1}{3} & 0 & \frac{2}{3} & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$

Then, using the concepts defined so far we are going to study two important issues for the Markov theory.

Definition 2.1.4. Given a stochastic process $\{X_n; n \geq 0\}$ that takes values in the state space I , we have

- (i) $P(X_0 = i) = \nu_i, \forall i \in I$, which is called initial distribution.
- (ii) $P(X_{n+1} = i_{n+1} | X_0 = i_0, \dots, X_n = i_n) = P(X_{n+1} = i_{n+1} | X_n = i_n), \forall n \geq 0$ and $\forall i_0, \dots, i_n, i_{n+1} \in I$.

So the stochastic process $\{X_n; n \geq 0\}$ is a Markov chain with initial distribution $\nu = \{\nu_i; i \in I\}$ and transition matrix $\Pi = (p_{i,j}; i, j \in I)$.

The last equality is known as Markov property, which tells us that the probability of a future event only depends on the probability of the prior event instead of the whole system evolution. This implies that Markov chains are a non-memory process.

Definition 2.1.5. A homogeneous Markov chain is a Markov chain, such that $\forall n \geq 0, \forall i, j \in I$, the probability

$$P(X_{n+1} = j | X_n = i) = p_{i,j}$$

is independent of n .

Often, we write $HMC(\nu, \Pi)$ in reference to a homogeneous Markov chain with distribution (or law) ν and transition matrix Π .

To continue, based on concepts previously defined, we introduce a theorem related with homogeneous Markov chains, which will be so useful to study some outcomes we will face in next chapters.

Theorem 2.1.6. A stochastic process $\{X_n; n \geq 0\}$ with values in I is a $HMC(\nu, \Pi)$ if and only if, $\forall n \geq 0$ and $\forall i_0, \dots, i_{n-1}, i \in I$ we have

$$P(X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) = \nu_{i_0} p_{i_0, i_1} \cdots p_{i_{n-1}, i}. \quad (2.1)$$

Proof. Firstly, we demonstrate the right implication. Given X_n , a homogeneous Markov chain (ν, Π) , applying the compound probability and the Markov property we get

$$\begin{aligned} P(X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) \\ &= P(X_0 = i_0) P(X_1 = i_1 | X_0 = i_0) \times \cdots \times P(X_n = i | X_0 = i_0, \dots, X_{n-1} = i_{n-1}) \\ &= \nu_{i_0} p_{i_0, i_1} \cdots p_{i_{n-1}, i}. \end{aligned}$$

Now we prove the left implication. Considering (2.1), we have to study that X_n is a Markov chain as well as it is homogenous. Firstly, we analyze the initial distribution, in that way $\forall i_0 \in I$ we obtain

$$P(X_0 = i_0) = \sum_{j \in I} \nu_{i_0} p_{i_0, j} = \nu_{i_0} \sum_{j \in I} p_{i_0, j} = \nu_{i_0}.$$

Secondly, we have to check Markov property, in this case we use the conditional probability, so $\forall i_0, \dots, i_{n-1}, i, j \in I$ we have

$$\begin{aligned} P(X_{n+1} = j | X_0 = i_0, \dots, X_n = i) &= \frac{P(X_0 = i_0, \dots, X_n = i, X_{n+1} = j)}{P(X_0 = i_0, \dots, X_n = i)} \\ &= \frac{\nu_{i_0} p_{i_0, i_1} \cdots p_{i_{n-1}, i} p_{i, j}}{\nu_{i_0} p_{i_0, i_1} \cdots p_{i_{n-1}, i}} = p_{i, j}. \end{aligned}$$

In order to prove the property we are studying, the value of the last equality must be equal to the value of $P(X_{n+1} = j | X_n = i)$, we are going to check if this is certain

$$\begin{aligned}
P(X_{n+1} = j | X_n = i) &= \frac{P(X_{n+1} = j, X_n = i)}{P(X_n = i)} \\
&= \frac{\sum_{i_0, \dots, i_{n-1} \in I} P(X_0 = i_0, \dots, X_n = i, X_{n+1} = j)}{\sum_{i_0, \dots, i_{n-1} \in I} P(X_0 = i_0, \dots, X_n = i)} \\
&= \frac{\sum_{i_0, \dots, i_{n-1} \in I} \nu_{i_0} p_{i_0, i_1} \cdots p_{i_{n-1}, i} p_{i, j}}{\sum_{i_0, \dots, i_{n-1} \in I} \nu_{i_0} p_{i_0, i_1} \cdots p_{i_{n-1}, i}} = \frac{p_{i, j} \sum_{i_0, \dots, i_{n-1} \in I} \nu_{i_0} p_{i_0, i_1} \cdots p_{i_{n-1}, i}}{\sum_{i_0, \dots, i_{n-1} \in I} \nu_{i_0} p_{i_0, i_1} \cdots p_{i_{n-1}, i}} = p_{i, j}.
\end{aligned}$$

As a result, we have $P(X_{n+1} = j | X_0 = i_0, \dots, X_n = i) = p_{i, j} = P(X_{n+1} = j | X_n = i)$, therefore the Markov property is satisfied.

Finally, we have to study the homogeneous property. From the last equality we have that $P(X_{n+1} = j | X_n = i) = p_{i, j}$ where $p_{i, j}$ is independent of n which is the necessary condition for homogeneity.

In conclusion, we have that X_n is a homogeneous Markov chain with distribution (or law) ν and transition (probability) matrix Π .

□

The result we have just obtained is very important, due to it is useful to check if a Markov chain is homogenous or not. We use this result in order to prove that given a homogenous Markov chain, this can be moved, and in that way obtain another chain that is homogenous too.

Proposition 2.1.7. *Given $\{X_n; n \geq 0\}$, a $HMC(\nu, \Pi)$; then $\{X_{m+n}; m \geq 0\}$ is another $HMC(\eta, \Pi)$, that is to say, $\forall m \geq 0, \forall j_0, \dots, j_n \in I$*

$$P(X_{m+0} = j_0, \dots, X_{m+n} = j_n) = P(X_m = j_0) p_{j_0, j_1} \cdots p_{j_{n-1}, j_n}.$$

Proof. Suppose $\{X_n; n \geq 0\}$ is a homogeneous Markov chain and we want to demonstrate that $\{X_{m+n}; m \geq 0\}$ too. Consequently, $\forall m \geq 0$ and $\forall i_0, \dots, i_{m-1}, j_0, \dots, j_n \in I$ we have

$$\begin{aligned}
P(X_{m+0} = j_0, \dots, X_{m+n} = j_n) &= \sum_{i_0, \dots, i_{m-1} \in I} P(X_0 = i_0, \dots, X_{m-1} = i_{m-1}, X_m = j_0, \dots, X_{m+n} = j_n) \\
&= \sum_{i_0, \dots, i_{m-1} \in I} P(X_0 = i_0) p_{i_0, i_1} \cdots p_{i_{m-1}, j_0} p_{j_0, j_1} \cdots p_{j_{n-1}, j_n} \\
&= P(X_m = j_0) p_{j_0, j_1} \cdots p_{j_{n-1}, j_n}.
\end{aligned}$$

In conclusion, as we have seen that equality is satisfied, we have that X_{m+n} is a homogeneous Markov chain.

□

2.2 Basic examples

In this section, we are studying some basic examples about Markov chains, in particular about homogeneous ones. These type of chains are considered as one of the most important for getting into the Markov theory world.

1. Random walks on $\mathbb{Z}$:

Given the stochastic process $\{X_n; n \geq 0\}$, we define X_0 as initial position which is constant ($X_0 = 0$) and the process

$$X_n = X_0 + \xi_1 + \cdots + \xi_n = X_0 + \sum_{i=1}^n \xi_i$$

for $n \geq 1$ and $\{\xi_i; i \geq 1\}$ is a set of random variables.

Let $p \in (0, 1)$, and suppose that $\{\xi_i; i \geq 1\}$ is a sequence of $\{-1, 1\}$ values, which are 1 with probability p and -1 with probability $1-p = q$, this is to say $P(\xi_i = 1) = p$ and $P(\xi_i = -1) = q$ respectively.

In this case, ξ_i are identically distributed Bernoulli random variables and also independents.

To sum up, the process begins at time $i = 0$ with a value of 0 and moves along a straight line for every point of time $i \geq 1$. The probability that step is to the right is p , whereas, the probability that the step is to the left is q ; both are independent of where it has been before time i .

Once exposed this basic example about Markov theory, we are introducing a result in which are reflected some of the concepts studied here.

Proposition 2.2.1. *Given $\{Z_n; n \geq 1\}$ a family of identically distributed random variables and also independents with $Z_n : \Omega \rightarrow I$. Suppose $f : I \times I \rightarrow I$ and X_0 a random variable with values in I , independent of $\{Z_n; n \geq 1\}$. If $X_{n+1} = f(X_n, Z_{n+1})$ where $n \geq 0$, then $\{X_n; n \geq 0\}$ is a homogeneous Markov chain.*

Proof. In this case we have to prove the Markov property and the homogeneous one, to see that X_n is a homogeneous Markov chain. Firstly, let us to see the first property using the conditional probability, so $\forall n \geq 0$ and $\forall i_0, \dots, i_{n-1}, i, j \in I$ we have

$$\begin{aligned} P(X_{n+1} = j | X_n = i) &= \frac{P(f(X_n, Z_{n+1}) = j, X_n = i)}{P(X_n = i)} = \frac{P(f(i, Z_{n+1}) = j, X_n = i)}{P(X_n = i)} \\ &= \frac{P(f(i, Z_{n+1}) = j)P(X_n = i)}{P(X_n = i)} = P(f(i, Z_{n+1}) = j) := q_{i,j}. \end{aligned}$$

In order to satisfy the property we are studying, the value of the last equality must be equal to the value of $P(X_{n+1} = j | X_0 = i_0, \dots, x_{n-1} = i_{n-1}, X_n = i)$, we are going

to check if this is certain

$$\begin{aligned}
& P(X_{n+1} = j | X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) \\
&= \frac{P(X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i, X_{n+1} = j)}{P(X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i)} \\
&= \frac{P(X_0 = i_0, f(X_0, Z_1) = i_1, \dots, f(X_{n-1}, Z_n) = i, f(X_n, Z_{n+1}) = j)}{P(X_0 = i_0, f(X_0, Z_1) = i_1, \dots, f(X_{n-1}, Z_n) = i)} \\
&= \frac{P(X_0 = i_0, f(i_0, Z_1) = i_1, \dots, f(i_{n-1}, Z_n) = i, f(i, Z_{n+1}) = j)}{P(X_0 = i_0, f(i_0, Z_0) = i_1, \dots, f(i_{n-1}, Z_n) = i)} \\
&= \frac{P(X_0 = i_0)P(f(i_0, Z_1) = i_1) \cdots P(f(i_{n-1}, Z_n) = i)P(f(i, Z_{n+1}) = j)}{P(X_0 = i_0)P(f(i_0, Z_1) = i_1) \cdots P(f(i_{n-1}, Z_n) = i)} \\
&= P(f(i, Z_{n+1}) = j) := q_{i,j}.
\end{aligned}$$

With this we have

$$P(X_{n+1} = j | X_0 = i_0, \dots, X_{n-1} = i_{n-1}, X_n = i) = q_{i,j} = P(X_{n+1} = j | X_n = i)$$

therefore the Markov property is correct.

Now let us to see the homogeneous property. From previous results in the first part of the demonstration we get $P(X_{n+1} = j | X_n = i) = q_{i,j}$ where $q_{i,j}$ is independent of n , this is the necessary condition for homogeneity.

So, the family of random variables $\{X_n; n \geq 0\}$ satisfy the Markov property and homogeneity so that is a homogeneous Markov chain.

□

Corollary 2.2.2. *Let $\{Y_n; n \geq 1\}$ be a family of identically distributed random variables and also independents with values in I . If $X_n = \sum_{i=1}^n Y_i$, then $\{X_n; n \geq 0\}$ is a homogeneous Markov chain.*

With that, we have that given a set of identically distributed random variables, which also are independent, we can build different Markov chains, which satisfy the properties defined at the beginning.

To conclude this example we give the transition matrix, where p and $q = 1 - p$ are probabilities defined above ($p \in (0, 1)$)

$$\Pi = \begin{pmatrix} & \vdots & \vdots & & & & \\ \cdots & 0 & p & 0 & 0 & 0 & \cdots \\ \cdots & q & 0 & p & 0 & 0 & \cdots \\ \cdots & 0 & q & 0 & p & 0 & \cdots \\ \cdots & 0 & 0 & q & 0 & p & \cdots \\ & \vdots & \vdots & & & & \end{pmatrix}$$

2. Random walk on $\mathbb{Z}$ with absorbing barriers:

Consider two players A and B who play at coin flipping (or heads or tails) and their capital are a and b respectively. The game ends when one player runs out of money.

This game is the practice of throwing a coin in the air. Consider the player A who wins 1 coin when the result is a head, this happen with probability p ; moreover, the player loses 1 coin if the result is a tail, this happen with probability $q = 1 - p$.

Given the stochastic process $\{X_n; n \geq 0\}$, which we note as the evolution of the capital for the player A . Therefore, we define $X_0 = a$ as initial condition and the process $X_{n+1} = X_n + \xi_{n+1}$ such that $X_n : \Omega \longrightarrow \{0, 1, \dots, a + b\}$ where $\{\xi_i; i \geq 1\}$ is a set of identically distributed random variables and also independents.

To conclude this example we give their transition matrix, where p and $q = 1 - p$ are probabilities defined above

$$\Pi = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ q & 0 & p & 0 & \cdots & 0 \\ 0 & q & 0 & p & \cdots & 0 \\ 0 & 0 & q & 0 & \cdots & 0 \\ \vdots & \vdots & & & & \vdots \\ 0 & 0 & \cdots & q & 0 & p \\ 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}$$

3. Random walk on $\mathbb{Z}$ with reflecting barriers:

This section has a similar argument to the last example, but with a difference; when one player runs out of money not ends the game due to that the other player gives him one coin. In this case the transition matrix is

$$\Pi = \begin{pmatrix} 0 & 1 & 0 & 0 & \cdots & 0 \\ q & 0 & p & 0 & \cdots & 0 \\ 0 & q & 0 & p & \cdots & 0 \\ 0 & 0 & q & 0 & \cdots & 0 \\ \vdots & \vdots & & & & \vdots \\ 0 & 0 & \cdots & q & 0 & p \\ 0 & 0 & \cdots & 0 & 1 & 0 \end{pmatrix}$$

3 Chapman-Kolmogorov equation

One of the most relevant elements when introducing to Markov chains is the Chapman-Kolmogorov equation. In order to study this type of equation, in this chapter, we focus on analyzing the simple way that allows us to discompose the probability of moving from one state i to another state j in m steps, in the sum of the probabilities of the trajectories that go from the state i to the state j and pass through any other state in an intermediate point of time.

In this chapter, we should take into account a probability space $(\Omega, \mathcal{F}, P)$ and a family of discrete random variables $X_n : \Omega \rightarrow I$ where $\{X_n; n \geq 0\}$ is a HMC(ν, Π) and I is a countable set called state space.

Definition 3.0.1. *Given $\{X_n; n \geq 0\}$ a time homogeneous Markov chain, we define the m -step transition probability, with $m > 1$, as*

$$p_{i,j}^{(m)} = P(X_m = j | X_0 = i) = P(X_{n+m} = j | X_n = i), \forall i, j \in I$$

which is the probability of going from i to j in m steps. In this case, the transition matrix is given as $\Pi_m = \Pi^{(m)} := (p_{i,j}^{(m)}; i, j \in I)$.

Observation 3.0.2. Considering the last definition, for $m = 1$ we have the transition probability $p_{i,j}^{(1)} = p_{i,j} = P(X_{n+1} = j | X_n = i)$ which goes from i to j in one step. Now, the transition matrix is $\Pi^{(1)} = \Pi$.

Moreover, we are checking that m -step transition probability is well defined, hence we have to calculate $p_{i,j}^{(2)} = P(X_{n+2} = j | X_n = i)$. In order to compute $X_{n+2} = j$, X_{n+1} must go through some state k .

$$\begin{aligned} p_{i,j}^{(2)} &= P(X_{n+2} = j | X_n = i) = \sum_{k \in I} P(X_{n+2} = j, X_{n+1} = k | X_n = i) \\ &= \sum_{k \in I} \frac{P(X_{n+2} = j, X_{n+1} = k, X_n = i)}{P(X_n = i)} \\ &= \sum_{k \in I} \frac{P(X_{n+2} = j, X_{n+1} = k, X_n = i)}{P(X_n = i)} \cdot \frac{P(X_{n+1} = k, X_n = i)}{P(X_{n+1} = k, X_n = i)} \\ &= \sum_{k \in I} \frac{P(X_{n+2} = j, X_{n+1} = k, X_n = i)}{P(X_{n+1} = k, X_n = i)} \cdot \frac{P(X_{n+1} = k, X_n = i)}{P(X_n = i)} \\ &= \sum_{k \in I} P(X_{n+2} = j | X_{n+1} = k, X_n = i) P(X_{n+1} = k | X_n = i) = \sum_{k \in I} p_{i,k} p_{k,j}, \end{aligned}$$

where the last result is the i, j -th entry of the transition matrix Π^2 .

After that, if we do mathematical induction on the number of steps, we get the following result.

Proposition 3.0.3. *The m -step transition probability is the m -th power of the transition matrix Π , that is to say $\Pi^m = \Pi \cdot \dots \cdot \Pi$, $\forall m > 2$*

Proof. To study this proposition, we just need to prove the following equality

$$p_{i,j}^{(m)} = \sum_{k \in I} p_{i,k}^{(m-1)} p_{k,j}, \forall m > 2$$

which implies that $\Pi^m = \Pi^{m-1}\Pi$, because of the first definition. Using the conditional probability we get

$$\begin{aligned} p_{i,j}^{(m)} &= P(X_{n+m} = j | X_n = i) = \sum_{k \in I} P(X_{n+m} = j, X_{n+m-1} = k | X_n = i) \\ &= \sum_{k \in I} \frac{P(X_{n+m} = j, X_{n+m-1} = k, X_n = i)}{P(X_n = i)} \\ &= \sum_{k \in I} \frac{P(X_{n+m} = j, X_{n+m-1} = k, X_n = i)}{P(X_n = i)} \cdot \frac{P(X_{n+m-1} = k, X_n = i)}{P(X_{n+m-1} = k, X_n = i)} \\ &= \sum_{k \in I} \frac{P(X_{n+m} = j, X_{n+m-1} = k, X_n = i)}{P(X_{n+m-1} = k, X_n = i)} \cdot \frac{P(X_{n+m-1} = k, X_n = i)}{P(X_n = i)} \\ &= \sum_{k \in I} P(X_{n+m} = j | X_{n+m-1} = k, X_n = i) P(X_{n+m-1} = k | X_n = i) \\ &= \sum_{k \in I} p_{k,j} p_{i,k}^{(m-1)}. \end{aligned}$$

Finally we have $p_{i,j}^{(m)} = \sum_{k \in I} p_{k,j} p_{i,k}^{(m-1)} = \sum_{k \in I} p_{i,k}^{(m-1)} p_{k,j}$ as we desired. □

Observation 3.0.4. From the previous proposition we get $p_{i,j}^{(m)} = \sum_{k \in I} p_{i,k}^{(m-1)} p_{k,j}$, now $\forall i_1, \dots, i_m \in I$ we can rewrite the equality as follows

$$p_{i,j}^{(m)} = \sum_{i_{m-1} \in I} p_{i,i_{m-1}}^{(m-1)} p_{i_{m-1},j} = \sum_{i_{m-1}, \dots, i_1 \in I} p_{i,i_1} p_{i_1,i_2} \cdots p_{i_{m-1},j}.$$

Now, considering the defined items and the outcomes obtained in the prior introduction, we are already able to study the Chapman-Kolmogorov equation, which is defined as following.

Definition 3.0.5. Let $m, n \geq 0$, then $\forall i, j \in I$ the probability of going from i to j in $m + n$ steps is

$$p_{i,j}^{(m+n)} = \sum_{k \in I} p_{i,k}^{(m)} p_{k,j}^{(n)}$$

which is called Chapman-Kolmogorov equation.

To continue, we have to see that the Chapman-Kolmogorov equation is correct. We have to take into account that for going from i to j in $m + n$ steps; firstly we are going from i to k in m steps and then, from k to j in n steps. As a result we

have

$$\begin{aligned}
p_{i,j}^{(m+n)} &= P(X_{m+n} = j | X_0 = i) = \sum_{k \in I} P(X_{m+n} = j, X_m = k | X_0 = i) \\
&= \sum_{k \in I} P(X_{m+n} = j | X_m = k, X_0 = i) P(X_m = k | X_0 = i) \\
&= \sum_{k \in I} p_{i,k}^{(m)} p_{k,j}^{(n)}.
\end{aligned}$$

Observation 3.0.6. In this case, the transition matrix is given as $\Pi^{m+n} = \Pi^m \Pi^n$, where $m, n \geq 0$. It can be defined due to the probability distribution is discrete and the Markov chain is homogeneous.

Then we are studying the effects of considering that the first state of the $\text{HMC}(\nu, \Pi)$ is generated randomly, so $\forall j \in I$ we can write

$$P(X_n = j) = \sum_{i \in I} P(X_0 = i, X_n = j) = \sum_{i \in I} P(X_0 = i) P(X_n = j | X_0 = i),$$

where $P(X_n = j | X_0 = i) = p_{i,j}^{(n)}$ and $P(X_0 = i) = \nu_i$. Using this last notation, we can rewrite the previous equality as follows

$$P(X_n = j) = \sum_{i \in I} \nu_i p_{i,j}^{(n)}.$$

With this we obtain the product between the transition matrix and a initial distribution which form a vector, then we get

$$\nu_j^{(n)} = P(X_n = j) = \sum_{i \in I} \nu_i p_{i,j}^{(n)} \text{ where } \nu^{(n)} = \nu \Pi^n \text{ is the law for } X_n.$$

Also using **Observation 3.0.4.**, for all $j \in I$, we can express the equality as follows

$$\nu_j^{(n)} = P(X_n = j) = \sum_{i \in I} \nu_i p_{i,j}^{(n)} = \sum_{i \in I} \sum_{i_1, \dots, i_{n-1} \in I} \nu_i p_{i,i_1} \cdots p_{i_{n-1},j}.$$

To continue, we study how we can determine the law of the random vector $(X_{n_1}, \dots, X_{n_m})$ with $n_1, \dots, n_m \geq 0$. For this, it is necessary using the distribution ν and the transition matrix Π . In general, we use the compound probability to find the law of the vector, so $\forall i_1, \dots, i_m \in I$ we have

$$\begin{aligned}
P(X_{n_1} = i_1, \dots, X_{n_m} = i_m) \\
&= P(X_{n_1} = i_1) P(X_{n_2} = i_2 | X_{n_1} = i_1) \cdots P(X_{n_m} = i_m | X_{n_1} = i_1, \dots, X_{n_{m-1}} = i_{m-1}) \\
&= \sum_{k \in I} \nu_k p_{k,i_1}^{(n_1)} p_{i_1,i_2}^{(n_2-n_1)} \cdots p_{i_{m-1},i_m}^{(n_m-n_{m-1})}.
\end{aligned}$$

With this, the result obtained consists in the product of the initial distribution and the elements of the stochastic matrix, which are raised to a given power, so it presents the same characteristics than the law obtained in the case of the homogeneous Markov chain X_n .

4 Classification of states

In this chapter, we analyze the way in which we can classify the states that compose a Markov chain. This classification may be based in the communication relationship that the different states present, so we study the possibility of moving from one state to another and then come back to the initial state, or weather given a specific state we do not move from this point.

In this section, we should take into account a probability space $(\Omega, \mathcal{F}, P)$ and a family of discrete random variables $X_n : \Omega \rightarrow I$ where I is a countable set and $\{X_n; n \geq 0\}$ is a HMC(ν, Π) where the transition matrix is $\Pi = (p_{i,j}; i, j \in I)$.

4.1 Basic concepts

We begin defining two concepts which show us that Markov chains are divided into parts.

Definition 4.1.1. *Given the states $i, j \in I$, we write $i \rightarrow j$ and say that j is accessible from i if the chain can go from state i to state j in a finite number of transitions with positive probability. This means that, for some $n \geq 0$*

$$p_{i,j}^{(n)} = P(X_n = j | X_0 = i) > 0.$$

Once given this definition, we use it to introduce the following concept, in which we talk about the possible relationship between two states.

Definition 4.1.2. *The state i communicates with j if they are accessible between them, in other words, this is i is accessible from j and j is accessible from i , which we write as $i \leftrightarrow j$.*

To continue, we are going to see that the property of communication is an equivalence relation, so, it has to satisfy the properties of reflexivity, symmetry and transitivity respectively, which are shown below

- (i) $p_{i,i}^{(0)} = P(X_0 = i | X_0 = i) = 1 > 0$, so $i \leftrightarrow i$ in 0 steps.
- (ii) If $i \leftrightarrow j$, exist $n, m \geq 0$ such that $p_{i,j}^{(n)} > 0$ and $p_{j,i}^{(m)} > 0$, then $j \leftrightarrow i$.
- (iii) If $i \leftrightarrow j$, exist $n_1, n_2 \geq 0$ where $p_{i,j}^{(n_1)} > 0$, $p_{j,i}^{(n_2)} > 0$ and also $j \leftrightarrow l$, so exist $m_1, m_2 \geq 0$ where $p_{j,l}^{(m_1)} > 0$, $p_{l,j}^{(m_2)} > 0$, then $i \leftrightarrow l$ because, using the Chapman-Kolmogorov equation, the probabilities satisfy

$$p_{i,l}^{(n_1+m_1)} = \sum_{k \in I} p_{i,k}^{(n_1)} p_{k,l}^{(m_1)} \geq p_{i,j}^{(n_1)} p_{j,l}^{(m_1)} > 0,$$

$$p_{l,i}^{(n_2+m_2)} = \sum_{k \in I} p_{l,k}^{(n_2)} p_{k,i}^{(m_2)} \geq p_{l,j}^{(n_2)} p_{j,i}^{(m_2)} > 0.$$

Therefore, we have that the property $\leftrightarrow$ is an equivalence relation which gives a partition of the state space, I , into equivalence classes, which are called communication classes. Moreover, we have that two states that communicate are in the same class. The concept we have just studied helps us to define the following property.

Definition 4.1.3. *A Markov chain is irreducible if there is a unique equivalence class, in other words, when every state communicates with each other.*

Now, we define the concept of a closed class, which will be useful later for classifying states.

Definition 4.1.4. *A subset $C \subseteq I$ is a closed class if $i \in C$ and $i \rightarrow j$ implies that $j \in C$, that is to say, from a state of a class C we can never have access to a state of $I \setminus C$ (hence, we cannot move from it).*

To continue, once given the last definition, we say that a class C is closed if the elements of the stochastic matrix satisfy the following property

$$\sum_{j \in C} p_{i,j} = 1, \forall i \in C.$$

An alternative way to study whether a class C is closed, is using the n -step transition probability, which must take the following value $p_{i,j}^{(n)} = 0$, for all $i \in C$, $j \in I \setminus C$ and $n \geq 1$.

4.2 Classification of states

Right now, we have defined some concepts, which form an introduction in order to be able to study the behaviour of any state of a Markov chain, which depends on whether we return to the initial state or not. For this, we begin defining two highly important concepts within the classification of the states.

Definition 4.2.1. *For any state $i \in I$, we define*

$$f_i = P(X_n = i \text{ for infinitely many } n | X_0 = i) = P_i(X_n = i \text{ for infinitely many } n)$$

and hence, we say that i is recurrent if $f_i = 1$, and it is transient if $f_i = 0$.

In general, a state is said to be recurrent if we will return to it in the future, if not, it is transient.

Definition 4.2.2. *The first passage time to state $i \in I$ is defined as*

$$T_i = \inf \{n \geq 1; X_n = i\},$$

this is the first time in which the chain visits the state i .

To continue, we define a new concept which will also help us to determine if a state is recurrent or transient.

Definition 4.2.3. The probability that a chain which starts in state i pass through j , denoted $\rho_{i,j}$, is defined as

$$\rho_{i,j} = P_i(T_j < \infty).$$

Once given the last definition, we can rewrite $\rho_{i,j}$ as follows

$$\rho_{i,j} = \sum_{k=1}^{\infty} P_i(T_j = k).$$

Moreover, we have the particular case where the probability that a chain which starts in state i returns to the starting point, i , it is denoted as $\rho_{i,i}$.

Now, we are going to introduce a new concept, for which we will need to consider the indicator function $\mathbb{1}_{\{X=k\}} = \begin{cases} 1 & \text{if } X = k \\ 0 & \text{if } X \neq k \end{cases}$.

Definition 4.2.4. The number of visits that the chain does to the state j , denoted by $N(j)$, is defined as

$$N(j) = \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n=j\}}.$$

Considering the concepts we have just defined, we have an equivalence between both elements $\{N(j) \geq 1\}$ and $\{T_j < \infty\}$, given $X_0 = i$, so the element $\rho_{i,j}$ can be expressed as $\rho_{i,j} = P_i(T_j < \infty) = P_i(N(j) \geq 1)$.

To carry on, given the last property, we are going to prove the following equality

$$P_i(N(j) \geq k) = \rho_{i,j} \rho_{j,j}^{k-1}, \forall k \geq 2.$$

Now, we are going to compute the probability that a Markov chain which starts at i visits the state j for the first time in the point of time k and, the next time it comes back to j , it takes n instants of time, then the probability is

$$\begin{aligned} P(X_{n+k} = j, X_{n+k-1} \neq j, \dots, X_k = j, \dots, X_1 \neq j | X_0 = i) \\ = P(X_{n+k} = j, \dots, X_{k+1} \neq j | X_k = j) P(X_k = j, \dots, X_1 \neq j | X_0 = i) \\ = P_j(T_j = n) P_i(T_j = k). \end{aligned}$$

Using the expression of $\rho_{i,j}$, we have that the probability of visiting the state i at least twice is

$$\begin{aligned} P_i(N(j) \geq 2) &= \sum_{k=1}^{\infty} \sum_{n=1}^{\infty} P_i(T_j = k) P_j(T_j = n) \\ &= \left(\sum_{k=1}^{\infty} P_i(T_j = k) \right) \left(\sum_{n=1}^{\infty} P_j(T_j = n) \right) = \rho_{i,j} \rho_{j,j}. \end{aligned}$$

After that, we do mathematical induction on the number of visits to the state j . Assuming the result is true for $k-1$, we will see that also holds for k . To continue,

we calculate the probability that a chain which starts at i visits the state j for the first time in the point of time l , after this, the chain will visit again the state j $k-1$ times.

$$\begin{aligned} P(X_{n+k+\dots+m+l} = j, \dots, X_{k+\dots+m+l} = j, \dots, X_{m+l} = j, \dots, X_l = j, \dots, X_1 \neq j | X_0 = i) \\ = P_j(T_j = n) \cdot \dots^{k-1} P_j(T_j = m) P_i(T_j = l). \end{aligned}$$

As we have done previously, once computed the last equality, we have the following probability

$$\begin{aligned} P_i(N(j) \geq k) &= \sum_{l=1}^{\infty} \sum_{m=1}^{\infty} \dots \sum_{n=1}^{\infty} P_i(T_j = l) P_j(T_j = m) \dots P_j(T_j = n) \\ &= \left(\sum_{l=1}^{\infty} P_i(T_j = l) \right) \left(\sum_{m=1}^{\infty} P_j(T_j = m) \right) \cdot \dots^{k-1} \left(\sum_{n=1}^{\infty} P_j(T_j = n) \right) \\ &= \rho_{i,j} \rho_{j,j} \cdot \dots^{k-1} \rho_{j,j} = \rho_{i,j} \rho_{j,j}^{k-1}. \end{aligned}$$

This result is very important, due to it will help us to demonstrate some of the theorems, related with the classification of states, that we will find later.

Considering the concepts previously defined and given one state $i \in I$, if the probability satisfies the following equality $\rho_{i,i} = 1$, then

$$P_i(N(i) = \infty) = \lim_{k \rightarrow +\infty} P_i(N(i) \geq k) = \lim_{k \rightarrow +\infty} \rho_{i,i} \rho_{i,i}^{k-1} = \lim_{k \rightarrow +\infty} 1^k = 1,$$

and hence, the state i is recurrent because we can visit this state ∞ times. Moreover, the probability $\rho_{i,i}$ may also be less than 1, $\rho_{i,i} < 1$, and in this case we have that

$$P_i(N(i) = \infty) = \lim_{k \rightarrow +\infty} P_i(N(i) \geq k) = \lim_{k \rightarrow +\infty} \rho_{i,i} \rho_{i,i}^{k-1} = 0$$

and now the state i is transient, because the number of times we visit this state is low.

Once studied the previous equalities, we focus on the definition of the number of times that a chain visits one state, in this context, we will introduce the calculation of the expected value. Therefore, given the expected value $E_i(\mathbb{1}_{\{X_n=j\}}) = P_i(X_n = j) = p_{i,j}^{(n)}$, the expected number of visits to state j , for a chain that begins in i , is the following

$$E_i(N(j)) = \sum_{n=1}^{\infty} p_{i,j}^{(n)}.$$

Now, we study the relationship between different types of states of a Markov chain and the expected value of the number of times that a chain visits them.

Theorem 4.2.5. *Given a state $j \in I$ which can be transient or recurrent, then we have $E_j(N(j)) = \frac{\rho_{j,j}}{1-\rho_{j,j}}$ or $E_j(N(j)) = \infty$ respectively.*

Proof. Firstly, we consider the case in which the probability $\rho_{j,j}$ is less than 1, so $\rho_{j,j} < 1$ and using the property $P_i(N(j) \geq k) = \rho_{i,j}\rho_{j,j}^{k-1}$ we have

$$P_i(N(j) = \infty) = \lim_{k \rightarrow +\infty} P_i(N(j) \geq k) = \lim_{k \rightarrow +\infty} \rho_{i,j}\rho_{j,j}^{k-1} = 0.$$

In particular, $P_j(N(j) = \infty) = 0$ and hence, the state j is transient. Now, we compute the expected value

$$\begin{aligned} E_j(N(j)) &= \sum_{k=1}^{\infty} k P_j(N(j) = k) = \sum_{k=1}^{\infty} k (P_j(N(j) \geq k) - P_j(N(j) \geq k+1)) \\ &= \sum_{k=1}^{\infty} k (\rho_{j,j}\rho_{j,j}^{k-1} - \rho_{j,j}\rho_{j,j}^k) = \rho_{j,j}(1 - \rho_{j,j}) \sum_{k=1}^{\infty} k \rho_{j,j}^{k-1} \\ &= \rho_{j,j}(1 - \rho_{j,j}) \frac{1}{(1 - \rho_{j,j})^2} = \frac{\rho_{j,j}}{1 - \rho_{j,j}} < \infty. \end{aligned}$$

Secondly, suppose the following equality holds $\rho_{j,j} = 1$ and using the property $P_i(N(j) \geq k) = \rho_{i,j}\rho_{j,j}^{k-1}$ we get

$$P_i(N(j) = \infty) = \lim_{k \rightarrow +\infty} P_i(N(j) \geq k) = \lim_{k \rightarrow +\infty} \rho_{i,j} \cdot 1^{k-1} = \rho_{i,j}.$$

In particular, $P_j(N(j) = \infty) = 1$ and hence, the state j is recurrent. From this equality, we obtain that the value for the random variables is ∞ , due to the probability of happening is 1, in this case, the expected value is $E_j(N(j)) = \infty$. □

Once given the prior theorem, the opposite implication is also true. This means that given the expected value of the number of times a chain visits one state, we are able to know if this state is recurrent or transient.

To continue, we are going to study some important properties related with recurrent and transient states, which are necessary, in practice, to classify the states of all classes.

Proposition 4.2.6. *Given one state i , if i is recurrent and $i \rightarrow j$, then $j \rightarrow i$.*

Proof. We begin assuming that the next property does not hold $j \rightarrow i$, so we have $j \not\rightarrow i$. This fact means that once the chain has reached the state j it will never be able to return to the state i again, in which it was located initially. Hence, once we had given up the state i we cannot come back again, so the state i is not recurrent, it is transient, which is a contradiction. □

By the last proposition, we can check that given both states i, j ; the state i is recurrent if once we have given up it we are not able to come back, this means that, $i \rightarrow j$ but $j \not\rightarrow i$.

Proposition 4.2.7. *Consider two states i, j if i is recurrent and $i \rightarrow j$, then j is also recurrent.*

Proof. To begin with, we know that if i is recurrent and $i \rightarrow j$, then also we have that $j \rightarrow i$. Therefore, there exists integers $M, m \geq 1$ such that $p_{j,i}^{(M)} > 0$ and $p_{i,j}^{(m)} > 0$ and then, using the Chapman-Kolmogorov equation, we obtain the following outcome $p_{j,j}^{(M+n+m)} \geq p_{j,i}^{(M)} p_{i,i}^{(n)} p_{i,j}^{(m)}$, which means that to go from state j to the state j in $M + n + m$ steps, we can go from j to i in M steps, then from i to i in n steps and finally from i to j in m steps. With this, we can compute the following expected value

$$\begin{aligned} E_j(N(j)) &= \sum_{k=1}^{\infty} p_{j,j}^{(k)} \geq \sum_{k=M+n+m}^{\infty} p_{j,j}^{(k)} = \sum_{n=1}^{\infty} p_{j,j}^{(M+n+m)} \geq p_{j,i}^{(M)} p_{i,i}^{(m)} \sum_{n=1}^{\infty} p_{i,i}^{(n)} \\ &= p_{j,i}^{(M)} p_{i,i}^{(m)} E_i(N(i)) = \infty, \end{aligned}$$

the last equality is due to the fact that i is a recurrent state, so now we have that j is also recurrent. □

To carry on, as an immediate consequence of both properties we have already studied, we have the following results.

Corollary 4.2.8. *Considering a communication class, C , then all its elements are recurrent or transient.*

Corollary 4.2.9. *Given a class, C , which is finite, irreducible and also closed, then all its states are recurrent.*

Proof. Firstly, we have that a given class C is finite and closed, which implies that there is at least one recurrent state. The class C is also irreducible, this means that all its states are communicated so, considering the results studied before, these states have to be recurrent. □

Now, we introduce an example in which we analyze all the concepts we have just studied.

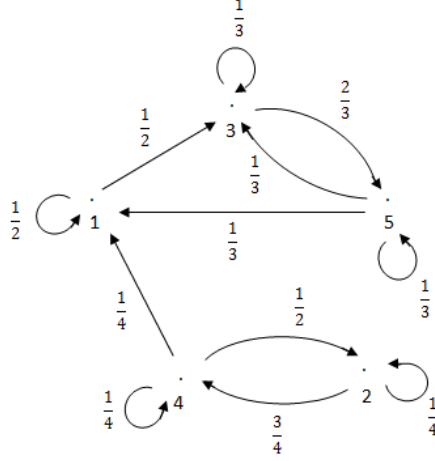
Example 4.2.10. Consider the Markov chain with state space $I = \{1, 2, 3, 4, 5\}$ and transition matrix

$$\Pi = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 \\ 0 & \frac{1}{4} & 0 & \frac{3}{4} & 0 \\ 0 & 0 & \frac{1}{3} & 0 & \frac{2}{3} \\ \frac{1}{4} & \frac{1}{2} & 0 & \frac{1}{4} & 0 \\ \frac{1}{3} & 0 & \frac{1}{3} & 0 & \frac{1}{3} \end{pmatrix}$$

In this case there are two communication classes which are

$$C_1 = \{2, 4\}, C_2 = \{1, 3, 5\}.$$

To continue, we analyze the communication relation that exists between the different states in order to determine weather the previous classes are recurrent or transient. For this, we use the diagram associated to the stochastic matrix, which is



We begin studying the states of the class C_1 . In this case, the state 4 is transient because if we start in this state, we are able to move to state 1 and once there we cannot return back to state 4. We know that if a state is transient, the rest of the states within the same class are transient too. Hence, the whole class C_1 is transient.

Finally, we study the defining characteristics of the class C_2 . This class is closed, because of we can move from the state $4 \in C_1$ to the state $1 \in C_2$, but once we reach the class C_2 we are not able to give up this class. Furthermore, this class is finite and irreducible due to all the states satisfy the property of communication, then, based on **Corollary 4.2.9**. C_2 is a recurrent class.

Once defined the concepts of recurrent state and transient state, and having studied some of their properties, we can define new ways to classify the states of a Markov chain.

Definition 4.2.11. A state i of a Markov chain is called absorbing if $p_{i,i} = 1$, so an absorbing state is necessarily recurrent.

With the last definition, we obtain that if a Markov chain is in an absorbing state i , it will not leave it due to $p_{i,j} = 0, \forall i \neq j$, so the closed class is only make up by state i , $C = \{i\}$.

Definition 4.2.12. An essential state is one that allows exit and return, in other words, given an essential state i for all $j \in I$ such that $i \rightarrow j$ it is also true that $j \rightarrow i$.

Given the prior definition, if the state allows to leave it but not to return, then it is called inessential, hence state i is inessential if for $j \in I$ exist $n \geq 1$ such that $p_{i,j}^{(n)} > 0$ but $p_{j,i}^{(m)} = 0, \forall m \geq 1$.

4.3 Periodicity and cyclic classes

Now, we suppose the state space is made up of disjoint sets and at each point of time we go from one set to another until at the last moment we return to the initial set. These sets are called cyclic subclasses, which we analyze later, but first we will study the period of the essential states which determines the number of cyclic subclasses that each one of them has.

Definition 4.3.1. Given essential state $i \in I$, the period of the state i denoted $d(i)$ is $d(i) = \gcd \left\{ n \geq 1; p_{i,i}^{(n)} > 0 \right\}$, so we say that the state i is periodic if $d(i) > 1$. Moreover i is said to be aperiodic if $d(i) = 1$.

Hence, we have seen that a state $i \in I$ is periodic if the greatest common divisor of the number of steps to return to the starting point is greater than one.

To continue, the following proposition studies the fact that if all states are in the same class, then they have the same period.

Proposition 4.3.2. Consider two states $i, j \in I$ which satisfy the following property $i \leftrightarrow j$, then the period of the states are the same, $d(i)=d(j)$.

Proof. Given $i, j \in I$ a pair of distinct states. Then $i \leftrightarrow j$ and there exists $n, m \geq 1$ such that $p_{i,j}^{(n)} > 0$ and $p_{j,i}^{(m)} > 0$, thereby we have $p_{i,i}^{(n+m)} \geq p_{i,j}^{(n)} p_{j,i}^{(m)} > 0$ and hence, $n + m$ must be divisible by $d(i)$. Let $k \geq 1$ such that $p_{j,j}^{(k)} > 0$, now we have $p_{i,i}^{(n+k+m)} \geq p_{i,j}^{(n)} p_{j,j}^{(k)} p_{j,i}^{(m)} > 0$, in this case $d(i) | n + k + m$, thus k is divisible by $d(i)$. This is true for every k such that $p_{j,j}^{(k)} > 0$, so $d(i) | d(j)$ by the definition of period. Reversing the roles of i and j we have $d(j) | d(i)$, so $d(i) = d(j)$. □

Consequently, the period of a class can be defined as the period of its states.

Now, we consider the congruence equivalence relation, so given $m, n \in \mathbb{Z}$ we have $m \equiv n \pmod{d}$ which means that $m - n$ is divisible by d . For the following results, we suppose that C forms an essential states class and we fix a reference state $i \in C$ with period d .

To continue, given $j \in C$ and considering $r \geq 1$ such that $p_{j,i}^{(r)} > 0$, if the probabilities satisfy $p_{i,j}^{(m)} > 0$ and $p_{i,j}^{(n)} > 0$, then we have that $p_{i,i}^{(m+r)} \geq p_{i,j}^{(m)} p_{j,i}^{(r)} > 0$ and $p_{i,i}^{(n+r)} \geq p_{i,j}^{(n)} p_{j,i}^{(r)} > 0$. Suppose that the period of the state i is d , so it divides

$m+r$ and also $n+r$. Consequently, $m-n$ is divisible by d and hence we can rewrite this fact as $m \equiv n(\text{mod } d)$. Finally, we define s_j as the remainder when n is divided by d for any n with $p_{i,j}^{(n)} > 0$.

Corollary 4.3.3. *For any $j \in C$ corresponds an integer $s_j = 0, 1, \dots, d-1$ such that $p_{i,j}^{(n)} > 0$ which implies that $n \equiv s_j(\text{mod } d)$.*

To carry on, we are going to present the cyclic classes. In order study this concept, we will need the outcomes obtained right now, where we have been studying the congruence relation, so considering $h \in \{0, 1, \dots, d-1\}$, we define

$$C_h = \{j \in C; p_{i,j}^{(n)} > 0 \text{ for } s_j \equiv h(\text{mod } d)\} = \{j \in C; p_{i,j}^{(n)} > 0 \text{ for } n \equiv h(\text{mod } d)\},$$

the last equality is due to the congruence for a fixed module is an equivalence relation, moreover $C_0 = C_d$. Once given this, we can express $C = \bigcup_{h=0}^{d-1} C_h$ so the sets $C_0, \dots, C_{d-1}$, which are disjoint, are called the cyclic subclasses of I .

Now, we study a result that will be helpful when classifying the states that compose a Markov chain into cyclical subclasses. This result takes into account all the non-null elements of the transition matrix.

Proposition 4.3.4. *Given $j \in C_k$ and $p_{j,l} > 0$, by convention $C_d = C_0$ it follows that if $k < d-1$, then $l \in C_{k+1}$ and if $k = d-1$, then $l \in C_0$.*

Proof. Let $n \geq 1$ such that $p_{i,j}^{(n)} > 0$, we have that $p_{i,l}^{(n+1)} \geq p_{i,j}^{(n)} p_{j,l} > 0$. As $j \in C_k$, we get that $n \equiv k(\text{mod } d)$ and, as the congruence equivalence relation is preserved by sums, we obtain $n+1 \equiv k+1(\text{mod } d)$ which implies that $l \in C_{k+1}$. □

To carry on, we introduce an example, of a Markov chain with finite state space, in which we study the period and the cyclic subclasses.

Example 4.3.5. Consider the Markov chain with state space $I = \{1, 2, 3, 4, 5, 6, 7\}$ and stochastic matrix

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

In this case, we have that the Markov chain is irreducible, due to there is an unique equivalence class. Now, we study the cyclic subclasses that the chain has. For this,

we note that all states are essential, thus, the period of this class is the period of any state. Consider for example the state $i = 1$, and then $d(1) = 3$. Therefore, there are three cyclic subclasses and these are the following ones

$$C_0 = \{1, 2\}, \quad C_1 = \{4, 7\}, \quad C_2 = \{3, 5, 6\}$$

4.4 Example: an analysis of the random walk on $\mathbb{Z}$

In this section, we have an example in which we analyze all the concepts that we have studied previously.

As we explain in the first chapter, this example focuses on the real line of integers. We started at state i and each time of point we can move one step forward or one backward with probabilities p and $q = 1 - p$ respectively.

Using the previous explanation, it is easy to check that all the states communicate, thereby there is only one class which is irreducible.

To continue, we know that all the states of a given class should be recurrent or transient. In our case, we will focus on checking if the initial state, which is the state 0 ($X_0 = 0$) to simplify, is recurrent or transient and then we will have that the rest of the states belong to the same type.

Before carrying on, we need to mention that the stochastic process is composed by random variables ζ_i that follow a Bernoulli distribution, then the stochastic process $X_n = \sum_{i=1}^n \zeta_i$ follows a Binomial distribution so we have that

$$P(X_n = 2k - n) = \binom{n}{k} p^k (1 - p)^{n-k} \text{ where } k = 0, 1, \dots, n.$$

Now, to classify the state 0, we use the **Theorem 4.2.5.**, so it is only necessary consider the following sum $\sum_{m=1}^{\infty} p_{0,0}^{(m)}$, which is the expected value of visits to the state 0, for a chain that begins in 0.

Firstly, if we consider m as an odd number we have $p_{0,0}^{(2n+1)} = 0$, so we cannot come back to the initial state 0 with an odd number of movements, due to the number of steps to the right should be equal to the number of steps to the left. Then, if the number of movements is even we have that

$$p_{0,0}^{(2n)} = P(X_{2n} = 0) = \binom{2n}{n} p^n (1 - p)^n.$$

Therefore we get

$$\sum_{m=1}^{\infty} p_{0,0}^{(m)} = \sum_{n=1}^{\infty} p_{0,0}^{(2n)} = \sum_{n=1}^{\infty} \binom{2n}{n} p^n (1 - p)^n = \sum_{n=1}^{\infty} \frac{(2n)!}{n!n!} p^n (1 - p)^n,$$

in this case, Stirling's formula, if n is large, provides a good approximation to $n!$, and using the notation of [3, pag 9] we have

$$n! \approx n^{n+\frac{1}{2}} e^{-n} \sqrt{2\pi} \quad \text{as } n \rightarrow \infty.$$

Hence, we can approximate the probability $p_{0,0}^{(2n)} \approx \frac{(4p(1-p))^n}{\sqrt{\pi n}}$. To continue, we have to consider two cases, the first one is when $p = \frac{1}{2}$ then we have $p_{0,0}^{(2n)} \approx \frac{1}{\sqrt{\pi n}}$ so the sum is

$$\sum_{m=1}^{\infty} p_{0,0}^{(m)} \approx \sum_{n=1}^{\infty} \frac{1}{\sqrt{\pi n}} = \infty$$

in this case the state 0 is recurrent as the last sum is infinite, and hence, all states are recurrent.

The second case is when $p \neq \frac{1}{2}$, then $p_{0,0}^{(2n)} \approx \frac{(4p(1-p))^n}{\sqrt{\pi n}} = \frac{r^n}{\sqrt{\pi n}}$ where $0 < r < 1$ so the sum is

$$\sum_{m=1}^{\infty} p_{0,0}^{(m)} \approx \sum_{n=1}^{\infty} \frac{r^n}{\sqrt{\pi n}} \leq \sum_{n=1}^{\infty} r^n < \infty$$

and now the state 0 is transient as the last sum is finite, and hence, all states are transient.

Finally, we are going to study the cyclic subclasses. We observe that all the states are considered as essentials, due to if we move from anyone of them we are able to come back in a future point of time.

As we said before, the chain is irreducible, then exists an unique class which has a period of 2, so we just have two cyclic subclasses which are

$$C_0 = \{j \in \mathbb{Z}; p_{0,j}^{(n)} > 0 \text{ for } n \equiv 0(\text{mod } 2)\} = \{0, \pm 2, \pm 4, \dots\},$$

$$C_1 = \{j \in \mathbb{Z}; p_{0,j}^{(n)} > 0 \text{ for } n \equiv 1(\text{mod } 2)\} = \{\pm 1, \pm 3, \pm 5, \dots\}.$$

5 Hitting times

In this chapter, we should take into account a probability space $(\Omega, \mathcal{F}, P)$ and a family of discrete random variables $X_n : \Omega \rightarrow I$ where $\{X_n; n \geq 0\}$ is a HMC(ν, Π) and I is a countable set called state space. Let A be a subset of the state space I , so $A \subseteq I$.

In this section, we study which is the probability that a Markov chain reaches a state which is within the subset A . In order to analyze this concept would be necessary to introduce some new concepts.

The random variable $H^A : \Omega \rightarrow \{0, 1, 2, \dots\} \cup \{\infty\}$ defined by

$$H^A = \inf\{n \geq 0; X_n \in A\}$$

is called the hitting time of A , which is the first time the chain hits the subset A . In addition, if we consider the empty set, then its infimum is ∞ .

To continue, we define the probability that, starting from state $i \in I$, the chain hits A as

$$h_i^A = P_i(H^A < \infty).$$

Consider the particular case in which the chain is initially in state i and also $i \in A$, then we have that the hitting time is $H^A = 0$ and the probability h_i^A takes the following value $h_i^A = 1$. We also have to consider the case in which the subset A is a closed class, then h_i^A is called absorption probability.

Now, we study the average time a chain needs to reach the subset A . This fact is called mean hitting time and is defined by

$$\mu_i^A = E_i(H^A) = \sum_{n < \infty} n P_i(H^A = n) + \infty P(H^A = \infty).$$

In this case, we also consider the fact in which the subset A is a closed class, then μ_i^A is called the absorption time.

In practice there are some cases in which we use the following notation

$$h_i^A = P_i(\text{hit } A), \quad \mu_i^A = E_i(\text{time to hit } A)$$

because these expressions are easy to find once we have the stochastic matrix.

To continue, we are going to study all the possible values that the probability h_i^A can take in function of whether the state i is on the subset A or not. To make this analysis, we use the minimal solution expression which means that if $x = (x_i; i \in I)$ is a minimal solution and $y = (y_i; i \in I)$ is another solution such that $y_i \geq 0$, then $y_i \geq x_i \forall i \in I$.

Theorem 5.0.1. *The vector of probabilities $h^A = (h_i^A; i \in I)$ is the minimal non-negative solution of the following system*

$$\begin{cases} h_i^A = 1 & \text{for } i \in A \\ h_i^A = \sum_{j \in I} p_{i,j} h_j^A & \text{for } i \notin A. \end{cases} \quad (5.1)$$

Proof. Initially, we show that the probability h^A is a solution of the system. Firstly, if $X_0 = i \in A$, then $H^A = 0$ and in this case the probability is $h^A = 1$. Secondly, if $X_0 = i \notin A$, then $H^A \geq 1$ and using the Markov property we have

$$P_i(H^A < \infty | X_1 = j) = P_j(H^A < \infty) = h_j^A$$

and hence

$$\begin{aligned} h_i^A &= P_i(H^A < \infty) = \sum_{j \in I} P_i(H^A, X_1 = j) = \sum_{j \in I} P_i(H^A < \infty | X_1 = j) P_i(X_1 = j) \\ &= \sum_{j \in I} h_j^A p_{i,j} \end{aligned}$$

with this we have that $h_i^A = \sum_{j \in I} p_{i,j} h_j^A$, $\forall i \notin A$ and hence h^A is a solution of (5.1).

In general, there exist multiple solutions for the system. In this context, let us to prove that the vector h^A is the minimal solution. Now, suppose $l = (l_i; i \in I)$ is another solution to (5.1) so we have $h_i^A = l_i = 1 \forall i \in A$, if otherwise we have $i \notin A$ we can rewrite the linear equation $l_i = \sum_{j \in I} p_{i,j} l_j$ as follows

$$l_i = \sum_{j \in A} p_{i,j} l_j + \sum_{j \notin A} p_{i,j} l_j.$$

In the last expression we substitute l_j and we obtain

$$\begin{aligned} l_i &= \sum_{j \in A} p_{i,j} l_j + \sum_{j \notin A} p_{i,j} \left(\sum_{k \in A} p_{j,k} l_k + \sum_{k \notin A} p_{j,k} l_k \right) \\ &= P_i(X_1 \in A) + P_i(X_2 \in A, X_1 \notin A) + \sum_{j \notin A} \sum_{k \notin A} p_{i,j} p_{j,k} l_k. \end{aligned}$$

Repeating this argument, which consists in substitute for l_j the final term we have

$$\begin{aligned} l_i &= P_i(X_1 \in A) + \dots + P_i(X_n \in A, X_{n-1} \notin A, \dots, X_1 \notin A) \\ &\quad + \sum_{j_1 \notin A} \dots \sum_{j_n \notin A} p_{i,j_1} p_{j_1,j_2} \dots p_{j_{n-1},j_n} l_{j_n}. \end{aligned}$$

Finally, if the solution l_{j_n} is non-negative and the previous terms $P_i(X_1 \in A), \dots, P_i(X_n \in A, X_{n-1} \notin A, \dots, X_1 \notin A)$ sum to $P_i(H^A \leq \infty)$, then $l_i \geq P_i(H^A \leq n)$ which implies that

$$l_i \geq \lim_{n \rightarrow +\infty} P_i(H^A \leq n) = P_i(H^A < \infty) = h_i^A.$$

With this we get that the vector of probabilities $h^A = (h_i^A; i \in I)$ is the minimal non-negative solution to the linear equations (5.1). □

Once we studied this result, we find a similar fact for the vector of mean hitting times μ^A in the following theorem.

Theorem 5.0.2. *The vector of mean hitting times $\mu^A = (\mu_i^A; i \in I)$ is the minimal non-negative solution of the following system*

$$\begin{cases} \mu_i^A = 0 & \text{for } i \in A \\ \mu_i^A = 1 + \sum_{j \in I} p_{i,j} \mu_j^A & \text{for } i \notin A. \end{cases}$$

Proof. To begin we show that the mean hitting time μ^A is a solution of the system. Firstly, if $X_0 = i \in A$, then $H^A = 0$ which implies that $\mu_i^A = 0$. Secondly, if $X_0 = i \notin A$, then $H^A \geq 1$ and using the Markov property we have

$$E_i(H^A | X_1 = j) = 1 + E_j(H^A) = 1 + \mu_j^A$$

and hence using some properties of the conditional expectation we get

$$\begin{aligned} \mu_i^A &= E_i(H^A) = \sum_{j \in I} E_i(H^A \mathbb{1}_{\{X_1=j\}}) = \sum_{j \in I} E_i(H^A | X_1 = j) P_i(X_1 = j) \\ &= 1 + \sum_{j \notin A} p_{i,j} \mu_j^A. \end{aligned}$$

with this we have that $\mu_i^A = 1 + \sum_{j \notin A} p_{i,j} \mu_j^A$, $\forall i \notin A$ and hence μ^A is a solution for the system.

In general, there exists multiple solutions for the system. In this context, let us to prove that the vector μ^A is the minimal solution. To continue, suppose that $r = (r_i; i \in I)$ is another solution to the system, so we have $k_i^A = r_i = 0$, $\forall i \in A$. Now suppose that $i \notin A$ and then we have

$$\begin{aligned} r_i &= 1 + \sum_{j \notin A} p_{i,j} r_j = 1 + \sum_{j \notin A} p_{i,j} \left(1 + \sum_{k \notin A} p_{j,k} r_k \right) \\ &= P_i(H^A \geq 1) + P_i(H^A \geq 2) + \sum_{j \notin A} \sum_{k \notin A} p_{i,j} p_{j,k} r_k. \end{aligned}$$

Repeating this argument, which consists in substitute for r_j the final term we have

$$r_i = P_i(H^A \geq 1) + \cdots + P_i(H^A \geq n) + \sum_{j_1, \dots, j_n \notin A} p_{i,j_1} \cdots p_{j_{n-1}, j_n} r_{j_n}.$$

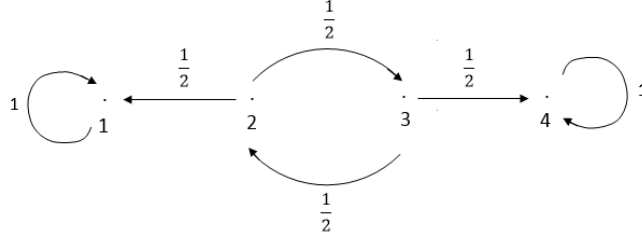
Finally, if r_{j_n} is non-negative we get

$$r_i \geq \lim_{n \rightarrow +\infty} (P_i(H^A \geq 1) + \cdots + P_i(H^A \geq n)) = E_i(H^A) = \mu_i^A$$

and with this we get that the vector of mean hitting times $\mu^A = (\mu_i^A; i \in I)$ is the minimal non-negative solution of the system. □

To continue, we introduce two examples in which we analyze the concepts previously studied. In the first one the state space is finite but in the second one not, in this case, the minimality condition is essential.

Example 5.0.3. Consider the Markov chain with symmetric random walk on the integers 1, 2, 3, 4. We want to study which is the probability of absorption in 1 if we start in the state 2 or in the state 3. Using the result of the first theorem, this is the system (5.1), we have to compute $h_2 = P_2(\text{hit } 1)$ and $h_3 = P_3(\text{hit } 1)$.



First we get $h_1 = P_1(\text{hit } 1) = 1$ this is the probability of hitting 1 if we left 1.

Now, suppose that we start at state 2 and consider the situation after making one step, so we jump to the state 1 or to the state 3 with the same probability which is $\frac{1}{2}$ and hence we have

$$h_2 = P_2(\text{hit } 1) = \frac{1}{2}h_1 + \frac{1}{2}h_3.$$

Similarly, the expression is obtained for the probability h_3 and this is

$$h_3 = P_3(\text{hit } 1) = \frac{1}{2}h_2 + \frac{1}{2}h_4.$$

Both equations we have just found form a system of linear equations and hence, is easier to get the values of h_2 and h_3 so, we have

$$\begin{cases} h_2 = \frac{2}{3} + \frac{1}{3}h_4 \\ h_3 = \frac{1}{3} + \frac{2}{3}h_4. \end{cases}$$

The value of h_4 is not defined by the system (5.1) because being in state 4 we cannot reach the state 1. In this case, we use the minimality condition and hence we get $h_4 = P_4(\text{hit } 1) = 0$. Substituting this value in the previous system we have that h_2 and h_3 take the following values $h_2 = \frac{2}{3}$ and $h_3 = \frac{1}{3}$ respectively. If we had not used the minimality condition the result would be the same because, as the state space is finite, to study the probabilities h_i , is only necessary the prior diagram.

To continue, we compute the time that must take until the chain is absorbed in state 1 or 4 if we start in the state 2 or in the state 3. In this case, we have to compute the value of mean hitting time $\mu_2 = E_2(\text{time to hit } \{1, 4\})$ and $\mu_3 = E_3(\text{time to hit } \{1, 4\})$ using the system of **Theorem 5.0.2.**

First we get $\mu_1 = \mu_4 = 0$ because we are initially in one of the states in which we want to arrive.

Now, suppose that we start at state 2 and consider the situation after making one step so we jump as we defined before. In this case the mean hitting time is

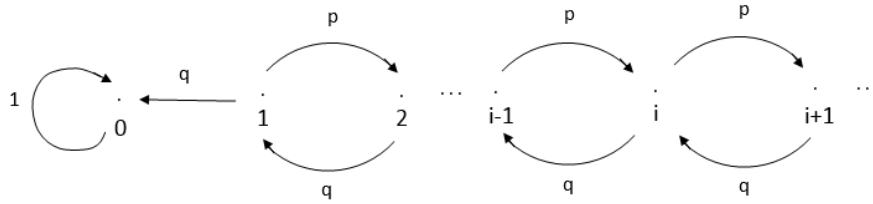
$$\mu_2 = 1 + \frac{1}{2}\mu_1 + \frac{1}{2}\mu_3 = 1 + \frac{1}{2}\mu_3.$$

Similarly, the expression is obtained for the mean hitting time μ_3 and this is

$$\mu_3 = 1 + \frac{1}{2}\mu_2 + \frac{1}{2}\mu_4 = 1 + \frac{1}{2}\mu_2.$$

Using the two equations which we have just found, then, starting in the state 2, the mean time for the chain to be absorbed by the state 1 or 4 is 2. On the other hand, if we are initially in the state 3, the mean time to be absorbed by the state 1 or 4 is also 2.

Example 5.0.4. In the first chapter, we studied the example of a random walk on $\mathbb{Z}$ with absorbing barriers, also called Gambler's ruin on $0, 1, \dots$. Let us consider the homogeneous Markov chain with the following diagram



Suppose that the gambler plays at heads or tails and initially he has i coins, at each point of time he wins or loses one coin with probability p or $q = 1 - p$ respectively. The game ends when the gambler loses everything. We want to study the probability that the player loses everything, we note that the only absorbing state is $\{0\}$.

The transitions probabilities are

$$\begin{aligned} p_{0,0} &= 1 \\ p_{i,i-1} &= 1 - p, \quad p_{i,i+1} = p \quad \text{for } i = 1, 2, \dots \end{aligned}$$

We want to study the probability h_i^A for state 0 when we start at state i , in other words, we want analyze $h_i = P_i(\text{hit } 0)$. Using the previous theorem we have the following system

$$\begin{aligned} h_0 &= 1 \\ h_i &= ph_{i+1} + qh_{i-1} \quad \text{for } i = 1, 2, \dots \end{aligned}$$

In this case, we have a recurrence relation of the form $rh_{i+1} + sh_i + th_{i-1} = 0$ where r and t are non-zero. We compute a solution of the form $h_n = \lambda^n$, so we have

$r\lambda^2 + s\lambda + t = 0$. There are two solutions of the quadratic equation, these are $\lambda_+ = 1$ and $\lambda_- = \frac{q}{p}$, then $h_n = \beta\lambda_+^n + \alpha\lambda_-^n$ is a solution.

Firstly, if $p \neq q$ we have that $\lambda_+ \neq \lambda_-$ and we can solve the following equation $h_0 = \alpha + \beta$ which implies that $\beta = h_0 - \alpha = 1 - \alpha$. Therefore, the previous recurrence has the following general solution

$$h_i = 1 - \alpha + \alpha \left(\frac{q}{p}\right)^i, \text{ for } i \geq 1$$

where $\alpha \in [0, 1]$ and the probability h_i takes different values depending on the values p and q . To continue we study two cases

- (i) If $p < q$, then as $h_i \in [0, 1]$ implies that $\alpha = 0$ and hence the minimal solution is given by $h_i = 1$.
- (ii) If $p > q$ and we have to find a minimal solution, the value of α has to be the biggest one, then $\alpha = 1$ and the minimal solution is given by $h_i = \left(\frac{q}{p}\right)^i$.

Finally, if $p = q = \frac{1}{2}$, then we have that $\lambda_+ = \lambda_- = 1$ and $h_n = (\beta + n\alpha)\lambda_+^n$ is a solution. To continue, we can solve the following equation $h_0 = \beta\lambda_+^n$ which implies that $\beta\lambda_+^n = 1$. Therefore, the previous recurrence has the following general solution

$$h_i = 1 + \alpha i, \text{ for } i \geq 1$$

in this case the restriction $h_i \in [0, 1]$ implies that $\alpha = 0$ and again we have $h_i = 1$ for all i .

In conclusion, for $p \leq \frac{1}{2}$ the player loses everything.

6 Distribution and measure

In this chapter, we shall study the limiting behaviour of Markov chains when at time n approaches infinity. In particular, we will focus on the relationship between this behaviour and probability distributions which are invariant. The previous concept tells us the fraction of times that the Markov chain spends in each state as n becomes large. Moreover, we study other utilities that has this distribution in Markov chains.

In this chapter, we should take into account a probability space $(\Omega, \mathcal{F}, P)$ and a family of discrete random variables $X_n : \Omega \rightarrow I$ where $\{X_n; n \geq 0\}$ is a HMC(ν, Π) and I is a countable set called state space.

6.1 Stationary or invariant distribution

In this section, we study the relationship between the long time behaviour of Markov chain with the invariant distribution or measure. Remember that a measure μ is a row vector $\mu = (\mu_i; i \in I)$ with non-negative entries. Moreover the measure μ is invariant if $\mu\Pi = \mu$.

Definition 6.1.1. *Given $\gamma = (\gamma_i; i \in I)$ a row vector which is a probability distribution such that satisfies*

$$0 \leq \gamma_i \leq 1, \quad \forall i \in I \quad \text{and} \quad \sum_{i \in I} \gamma_i = 1$$

and considering $\Pi = (p_{i,j}; i, j \in I)$ a stochastic matrix, then if

$$\sum_{i \in I} \gamma_i p_{i,j} = \gamma_j, \quad \forall j \in I \tag{6.1}$$

we say that γ is an invariant distribution or stationary distribution for stochastic matrix Π .

For a Markov chain $\{X_n; n \geq 0\}$ with transition matrix Π and invariant distribution γ , we can rewrite in matrix form the condition (6.1) as follows

$$\gamma\Pi = \gamma$$

where γ is a row vector as we said above.

Given γ an invariant distribution, we will study that the distribution (or law) of the HMC(γ, Π), $\{X_n; n \geq 0\}$, is independent of n , for all $n \geq 0$. Therefore, we will see that all random variables X_n have the same law.

For $n = 1$ we have the equality (6.1) which we have analyzed previously. Now we study the case in which n is 2, for this we will use the Chapman-Kolmogorov equation so we have

$$\sum_{i \in I} \gamma_i p_{i,j}^{(2)} = \sum_{i \in I} \gamma_i \sum_{k \in I} p_{i,k} p_{k,j} = \sum_{k \in I} \left(\sum_{i \in I} \gamma_i p_{i,k} \right) p_{k,j} = \sum_{k \in I} \gamma_k p_{k,j} = \gamma_j,$$

and hence we get $\gamma_j^{(2)} = P(X_2 = j) = \sum_{i \in I} \gamma_i p_{i,j}^{(2)} = \gamma_j$ which is the same law as the previous case, where $n = 1$.

After that, we use mathematical induction over the number of steps needed in order to the chain reaches the state j . Assuming the result is true for $n - 1$, we will see that also holds for n . To continue, we want to compute the distribution of the random variable X_n . For this, first of all, we study whether equality (6.1) holds but in this case are needed n steps to go from state i to state j , so we have

$$\begin{aligned} \sum_{i \in I} \gamma_i p_{i,j}^{(n)} &= \sum_{i \in I} \gamma_i \sum_{i_1, \dots, i_{n-1} \in I} p_{i,i_1} \cdots p_{i_{n-1},j} = \sum_{i \in I} \gamma_i \sum_{i_{n-1} \in I} p_{i,i_{n-1}}^{(n-1)} p_{i_{n-1},j} \\ &= \sum_{i_{n-1} \in I} \left(\sum_{i \in I} \gamma_i p_{i,i_{n-1}}^{(n-1)} \right) p_{i_{n-1},j} = \sum_{i_{n-1} \in I} \gamma_{i_{n-1}} p_{i_{n-1},j} = \gamma_j, \end{aligned}$$

in this case we get $\gamma_j^{(n)} = P(X_n = j) = \sum_{i \in I} \gamma_i p_{i,j}^{(n)} = \gamma_j$ and this is the law of the random variable X_n .

Therefore, if the distribution of the initial state X_0 is γ , then the following equality $\sum_{i \in I} \gamma_i p_{i,j}^{(n)} = \gamma_j$ implies that, for all n , $P(X_n = j) = \gamma_j$ so, all random variables have the same law. In addition, we can note that the distribution of all random variables is independent of n .

To continue, once given the definition of invariant distribution and some properties, we study the existence of an invariant distribution for any stochastic matrix. After this, we analyze this fact through some examples.

Proposition 6.1.2. *Given a homogeneous Markov chain $\{X_n; n \geq 0\}$ with finite state space I , then all stochastic matrix Π has an invariant distribution.*

Proof. First we consider a probability over the space I , which we refer as $w = (w_i; i \in I)$, it satisfies $0 \leq w_i \leq 1$ and also $\sum_{i \in I} w_i = 1$.

Before carrying on, we say that the number of elements which the finite space I contains is denoted as $|I|$ and it is called cardinal.

To continue, we define for all $n \geq 1$ the equality $w_n = \frac{1}{n} \sum_{m=0}^{n-1} w \Pi^{(m)}$, and now we analyze that w_n also defines a probability over the space I . On the one hand we have

$$\begin{aligned} \sum_{i \in I} w_{n,i} &= \frac{1}{n} \sum_{i \in I} \sum_{m=0}^{n-1} \sum_{j \in I} w_j p_{j,i}^{(m)} = \frac{1}{n} \sum_{m=0}^{n-1} \sum_{j \in I} w_j \sum_{i \in I} p_{j,i}^{(m)} \\ &= \frac{1}{n} \sum_{m=0}^{n-1} \sum_{j \in I} w_j \cdot 1 = \frac{1}{n} n \sum_{j \in I} w_j = 1, \end{aligned}$$

and on the other hand we get that $w_{n,i} \geq 0$ because $w \in [0, 1]$ and also $p_{j,i} \geq 0$. Then, both necessary conditions are satisfied in order to consider w_n as a probability.

The set of probability over the space I is a closed and bounded set in $[0, 1]^{|I|}$. To continue, suppose that there exists a convergent subsequence whose limit is an element of the set defined above. This element has to be a probability over the space I , which we note as γ . Thus, for any subsection $\{w_{n_j}; j \geq 1\}$ we have that w_{n_j} , when j takes very large values, converges to γ .

Now, we analyze that the probability γ is an invariant distribution. First we have

$$w_{n_j} - w_{n_j}\Pi = \frac{1}{n_j} \sum_{m=0}^{n_j-1} w\Pi^{(m)} - \frac{1}{n_j} \sum_{m=0}^{n_j-1} w\Pi^{(m+1)} = \frac{1}{n_j} (w - w\Pi^{n_j})$$

and hence as $\{w - w\Pi^{n_j}; j \geq 1\}$ is a bounded sequence due to state space I is finite we have

$$\lim_{j \rightarrow +\infty} w_{n_j} - w_{n_j}\Pi = \gamma - \gamma\Pi, \quad \lim_{j \rightarrow +\infty} \frac{1}{n_j} (w - w\Pi^{n_j}) = 0.$$

with this we get that $\gamma - \gamma\Pi = 0$ and hence, γ is an invariant distribution for Π because the following equality is satisfied $\gamma\Pi = \gamma$.

□

Now, to study the previous concepts, we will use two practical examples that differ on the number of invariant distributions for the stochastic matrix.

Example 6.1.3. Consider the Markov chain with state space $I = \{1, 2\}$ and transition matrix

$$\Pi = \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{1}{5} & \frac{4}{5} \end{pmatrix}$$

In this example, we want to study that this chain has an invariant distribution γ . To find this distribution, we have to prove the equality $\gamma\Pi = \gamma$, this is

$$(\gamma_1 \ \gamma_2) \begin{pmatrix} \frac{1}{4} & \frac{3}{4} \\ \frac{1}{5} & \frac{4}{5} \end{pmatrix} = (\gamma_1 \ \gamma_2)$$

with this we obtain the following equations

$$\gamma_1 = \frac{1}{4}\gamma_1 + \frac{1}{5}\gamma_2,$$

$$\gamma_2 = \frac{3}{4}\gamma_1 + \frac{4}{5}\gamma_2.$$

Moreover, we know the vector γ is a probability distribution, so it also has to satisfy the equality $\gamma_1 + \gamma_2 = 1$.

Once resolved the system of linear equations, we find that $\gamma = (\frac{4}{19}, \frac{15}{19})$, all the components are positive and satisfy the three equations above, so they represent an unique invariant distribution for the chain.

With this example, we have seen that the Markov chain with finite state space has an unique invariant distribution. To carry on, we prove the prior result in one of the examples that we have defined in Section 1, the random walk on $\mathbb{Z}$ with absorbing barriers, with finite state space. In this case, we will see that there may be more than one invariant distributions, therefore, the result above only assures its existence.

Example 6.1.4. In this case, the Markov chain with state space $I = \{0, 1, \dots, M\}$ is given by stochastic matrix

$$\Pi = \begin{pmatrix} 1 & 0 & 0 & 0 & \cdots & 0 \\ q & 0 & p & 0 & \cdots & 0 \\ 0 & q & 0 & p & \cdots & 0 \\ 0 & 0 & q & 0 & \cdots & 0 \\ \vdots & \vdots & & & & \vdots \\ 0 & 0 & \cdots & q & 0 & p \\ 0 & 0 & \cdots & 0 & 0 & 1 \end{pmatrix}$$

Now we want to compute the invariant distribution for the stochastic matrix Π . For this, we have to analyze the following equality $\gamma\Pi = \gamma$ where $\gamma = (\gamma_1, \gamma_2, \dots, \gamma_M)$, so we have the following system

$$\begin{cases} \gamma_0 = \gamma_0 + \gamma_1 q, \\ \gamma_1 = \gamma_2 q, \\ \gamma_j = \gamma_{j-1} p + \gamma_{j+1} q \text{ for } j = 2, \dots, M-2, \\ \gamma_{M-1} = \gamma_{M-2} p, \\ \gamma_M = \gamma_{M-1} p + \gamma_M. \end{cases}$$

Moreover, we know that the vector γ is a probability distribution, so it also has to satisfy $\gamma_0 + \dots + \gamma_M = 1$ and $\gamma_i \geq 0$ for all $i \in I$.

Once solved the system of linear equations we get that $\gamma_1 = \dots = \gamma_{M-1} = 0$ and $\gamma_0 + \gamma_M = 1$. Hence the family of invariant distribution is

$$\gamma = (1 - \alpha, 0, \dots, 0, \alpha) \text{ for } \alpha \in [0, 1].$$

In the last example, we observe that sometimes there exists more than one invariant distribution, so this distribution is not always unique. Moreover, in the last two examples, we have seen that there is at least one invariant distribution.

In the following theorem, in which we can note the relationship between invariant distributions and n -step transition probabilities, we study that the invariant distribution is an equilibrium distribution because for all instants of time the invariant distribution is the same.

Theorem 6.1.5. *Given the state space I which is finite, we suppose that for some $i \in I$ is satisfied that*

$$p_{i,j}^{(n)} \rightarrow \gamma_j \text{ as } n \rightarrow \infty \text{ for all } j \in I,$$

then $\gamma = (\gamma_j; j \in I)$ is an invariant distribution.

Proof. Firstly, we know that $0 \leq \gamma_j \leq 1$ for all $j \in I$, this also holds for the probability $p_{i,j}^{(n)}$ so this is $0 \leq p_{i,j}^{(n)} \leq 1$ for all $n \geq 1$ and $i, j \in I$. Now, we analyze the vector γ and see that it is a probability distribution. Using the commutation between the summation and the limit, due to state space is finite, we have

$$\sum_{j \in I} \gamma_j = \sum_{j \in I} \lim_{n \rightarrow +\infty} p_{i,j}^{(n)} = \lim_{n \rightarrow +\infty} \sum_{j \in I} p_{i,j}^{(n)} = \lim_{n \rightarrow +\infty} 1 = 1.$$

Finally, using the Chapman-Kolmogorov equation, we analyze if γ is an invariant distribution, so we get

$$\gamma_j = \lim_{n \rightarrow +\infty} p_{i,j}^{(n)} = \lim_{n \rightarrow +\infty} p_{i,j}^{(n+1)} = \lim_{n \rightarrow +\infty} \sum_{k \in I} p_{i,k}^{(n)} p_{k,j} = \sum_{k \in I} \lim_{n \rightarrow +\infty} p_{i,k}^{(n)} p_{k,j} = \sum_{k \in I} \gamma_k p_{k,j},$$

with this we get that the vector γ is an invariant distribution because the equality holds. □

Once given this result, we consider an example in which the state space I is not finite, this can be the random walk on $\mathbb{Z}$ which we have studied in Section 4.4, then the limit of the probability to go from state i to state j in n steps is invariant because it satisfies $p_{i,j}^{(n)} \rightarrow 0 := \gamma_j$ as $n \rightarrow \infty$ for all $i, j \in I$, but the vector γ is not a probability distribution because $\sum_{i \in I} \gamma_i \neq 1$ as well as it is not an invariant distribution.

To continue, we introduce a practical example in which we study the result we have just studied.

Example 6.1.6. Consider the Markov chain of **Example 6.1.3.**. Our goal is to compute the transition probabilities $p_{1,1}^{(n)}$ and $p_{2,2}^{(n)}$.

First we have to compute the eigenvalues of the transition matrix Π . For this, we calculate the characteristic equation, this is

$$\det(\Pi - \lambda Id) = 0 \Rightarrow \left(\frac{1}{4} - \lambda\right) \left(\frac{4}{5} - \lambda\right) - \frac{3}{20} = 0$$

thus, solving the equation we obtain the following eigenvalues $1, \frac{1}{20}$. Now, we can rewrite the transition matrix Π as a diagonal matrix and hence exists an invertible matrix A such that

$$\Pi = A \begin{pmatrix} 1 & 0 \\ 0 & \frac{1}{20} \end{pmatrix} A^{-1} \Rightarrow \Pi^{(n)} = A \begin{pmatrix} 1^n & 0 \\ 0 & \left(\frac{1}{20}\right)^n \end{pmatrix} A^{-1}.$$

Now, we want to compute the transition probability $p_{1,1}^{(n)}$. Using the n -th power of the stochastic matrix we have

$$p_{1,1}^{(n)} = \alpha \cdot 1^n + \beta \left(\frac{1}{20}\right)^n = \alpha + \beta \left(\frac{1}{20}\right)^n.$$

To continue, we calculate the values of the constants α and β , so we have to solve the following system of linear equations

$$\begin{cases} 1 = p_{1,1}^{(0)} = \alpha + \beta \\ \frac{1}{4} = p_{1,1}^{(1)} = \alpha + \frac{1}{20}\beta. \end{cases}$$

so, the values of the constants are the following $\alpha = \frac{4}{19}$ and $\beta = \frac{15}{19}$.

Now, we use the last theorem applied on the probability $p_{1,1}^{(n)}$ and we have

$$\lim_{n \rightarrow +\infty} p_{1,1}^{(n)} = \lim_{n \rightarrow +\infty} \left(\frac{4}{19} + \frac{15}{19} \left(\frac{1}{20} \right)^n \right) = \frac{4}{19} = \gamma_1$$

and with this we obtain that the distribution γ_1 is invariant. If we had computed the probability $p_{2,1}^{(n)}$, the result would have been the same. Now, we repeat the same idea on the probability $p_{2,2}^{(n)}$, then we get

$$p_{2,2}^{(n)} = \mu \cdot 1^n + \lambda \left(\frac{1}{20} \right)^n = \mu + \lambda \left(\frac{1}{20} \right)^n,$$

and now we want to compute the constants μ and λ , so we have to solve the following system of linear equations

$$\begin{cases} 1 = p_{2,2}^{(0)} = \mu + \lambda \\ \frac{4}{5} = p_{2,2}^{(1)} = \mu + \frac{1}{20}\lambda. \end{cases}$$

In this case, we obtain the values $\mu = \frac{15}{19}$ and $\lambda = \frac{4}{19}$, and hence the limit is the following

$$\lim_{n \rightarrow +\infty} p_{2,2}^{(n)} = \lim_{n \rightarrow +\infty} \left(\frac{15}{19} + \frac{4}{19} \left(\frac{1}{20} \right)^n \right) = \frac{15}{19} = \gamma_2.$$

With this we obtain that the distribution γ_2 is invariant. If we had computed the probability $p_{1,2}^{(n)}$ the result would have been the same. Therefore, by **Theorem 6.1.5**. we get that $\gamma = (\gamma_1, \gamma_2) = \left(\frac{4}{19}, \frac{15}{19} \right)$ and it is an invariant distribution.

In the examples of this chapter, we can observe that the invariant distribution may not exist, be unique or more than one distribution may exist. To continue, we are going to introduce some results that allow us to set conditions in order to guarantee the existence and uniqueness of the invariant measure. Firstly, we remember the concept of measure and invariant measure.

Definition 6.1.7. A measure is any row vector $\mu = (\mu_i; i \in I)$ with $\mu_i \geq 0$ for all $i \in I$. Moreover, we say a measure μ is invariant, if for any transition matrix $\Pi = (p_{i,j}; i, j \in I)$ satisfies

$$\sum_{i \in I} \mu_i p_{i,j} = \mu_j, \quad \forall j \in I,$$

this equality in matrix form is the following $\mu\Pi = \mu$.

Now, in the following result, we analyze the existence of the invariant measure.

Theorem 6.1.8. *Given the stochastic matrix $\Pi = (p_{i,j}; i, j \in I)$ of the Markov chain $\{X_n; n \geq 0\}$, which is irreducible and all states are recurrent. Consider*

$$\lambda_i^r = E_r \sum_{n=0}^{T_r-1} \mathbb{1}_{\{X_n=i\}},$$

then

(i) $\lambda_i^r \in (0, \infty)$ for all $i \in I$.

(ii) $\lambda^r = (\lambda_i^r; i \in I)$ satisfies $\lambda^r \Pi = \lambda^r$.

Proof. Given $n \geq 1$ we have that the element $\{T_r \geq n\}$ depends only on $X_0, \dots, X_{n-1}$ because none of the random variables X_n for all $n \geq 1$ are in state r . Using the Markov property at time $n-1$ we get

$$P_r(X_1 = l, \dots, X_{n-1} = i, X_n = j \text{ and } T_r \geq n) = P_r(X_{n-1} = i \text{ and } T_r \geq n) p_{i,j}.$$

We also know that the states of the Markov chain are recurrent, so it is satisfied $P_r(T_r < \infty) = 1$, which is the same as $P(X_0 = X_{T_r} = r) = 1$. Now, we have

$$\lambda_j^r = E_r \sum_{n=1}^{T_r} \mathbb{1}_{\{X_n=j\}} = E_r \sum_{n=1}^{\infty} \mathbb{1}_{\{X_n=j \text{ and } T_r \geq n\}} = \sum_{n=1}^{\infty} P_r(X_n = j \text{ and } T_r \geq n)$$

To continue, the chain, before visiting the state j at time n , has been in some state $i \in I$ at time $n-1$ so, this is the following

$$\begin{aligned} \lambda_j^r &= \sum_{i \in I} \sum_{n=1}^{\infty} P_r(X_{n-1} = i, X_n = j \text{ and } T_r \geq n) \\ &= \sum_{i \in I} p_{i,j} \sum_{n=1}^{\infty} P_r(X_{n-1} = i \text{ and } T_r \geq n) = \sum_{i \in I} p_{i,j} E_r \sum_{l=0}^{\infty} \mathbb{1}_{\{X_l=i \text{ and } T_{r-1} \geq l\}} \\ &= \sum_{i \in I} p_{i,j} E_r \sum_{l=0}^{T_r-1} \mathbb{1}_{\{X_l=i\}} = \sum_{i \in I} \lambda_i^r p_{i,j}. \end{aligned}$$

and with this we have that the following equality is satisfied $\lambda^r \Pi = \lambda^r$, which shows (ii).

To continue, we have to prove (i). For this purpose, we know that all the states of the Markov chain communicate among them. Then, for each state $i \in I$ there are $n, m \geq 0$ such that $p_{i,r}^{(m)} > 0$ and $p_{r,i}^{(n)} > 0$, with this we have

$$\lambda_i^r = \sum_{k \in I} \lambda_k^r p_{k,i}^{(n)} \geq \lambda_r^r p_{r,i}^{(n)} = p_{r,i}^{(n)} > 0,$$

and, in addition, we obtain that $1 = \lambda_r^r = \sum_{k \in I} \lambda_k^r p_{k,r}^{(m)} \geq \lambda_i^r p_{i,r}^{(m)}$. With this we get $\lambda_i^r \leq \frac{1}{p_{i,r}^{(m)}} < \infty$. Hence, the vector λ^r satisfies $\lambda_i^r > 0$ for all $i \in I$, this shows (i).

□

In the previous theorem, the equality $\lambda_i^r = E_r \sum_{n=0}^{T_r-1} \mathbb{1}_{\{X_n=i\}}$ is defined as: given a fixed state r , λ_i^r is the mean time we have been in the state i before reaching the state r . Moreover, remember that $T_r = \inf\{n \geq 1; X_n = r\}$ is the first time the chain visits the state r .

To continue, we will analyze a concept which we have used to study the result which we have proved above.

Observation 6.1.9. Consider a Markov chain that starts in the state r , then we have that $\sum_{n=0}^{T_r-1} \mathbb{1}_{\{X_n=r\}} = 1$ and this implies that

$$\lambda_r^r = E_r \sum_{n=0}^{T_r-1} \mathbb{1}_{\{X_n=r\}} = E_r(1) = 1.$$

Now, once analyzed the existence of the invariant measure, we are going to study the uniqueness for this invariant measure. In the following result we see that the uniqueness is satisfied except for multiplicative constants.

Theorem 6.1.10. *Given the stochastic matrix Π , which is irreducible and all states are recurrent, then the invariant measure μ for Π is unique up to a multiplication by a constant.*

Proof. Firstly, given the invariant measure μ , we want to prove $\mu \geq \mu_r \lambda^r$. Therefore, considering the measure μ we have that

$$\mu_j = \sum_{i_0 \in I} \mu_{i_0} p_{i_0,j} = \sum_{i_0 \neq r} \mu_{i_0} p_{i_0,j} + \mu_r p_{r,j}.$$

To continue, applying the same procedure to μ_{i_0} we obtain the following equality

$$\begin{aligned} \mu_j &= \sum_{i_0 \neq r} \left(\sum_{i_1 \neq r} \mu_{i_1} p_{i_1,i_0} + \mu_r p_{r,i_0} \right) p_{i_0,j} + \mu_r p_{r,j} \\ &= \sum_{i_0, i_1 \neq r} \mu_{i_1} p_{i_1,i_0} p_{i_0,j} + \left(\sum_{i_0 \neq r} \mu_r p_{r,i_0} p_{i_0,j} + \mu_r p_{r,j} \right), \end{aligned}$$

and now, once calculated the previous equalities, we repeat this fact $n - 1$ times, and then for all $n \in \mathbb{N}$ we obtain that

$$\begin{aligned} \mu_j &= \sum_{i_0, \dots, i_n \neq r} \mu_{i_n} p_{i_n, i_{n-1}} \cdots p_{i_0, j} \\ &\quad + \mu_r \left(\sum_{i_0, \dots, i_{n-1} \neq r} p_{r, i_{n-1}} \cdots p_{i_0, j} + \cdots + \sum_{i_0 \neq r} p_{r, i_0} p_{i_0, j} + p_{r, j} \right) \\ &\geq \mu_r P_r(X_1 = j \text{ and } T_r \geq 1) + \cdots + \mu_r P_r(X_n = j \text{ and } T_r \geq n) \\ &= \mu_r \sum_{k=1}^n P_r(X_k = j \text{ and } T_r \geq k) \rightarrow \mu_r \lambda_j^r \text{ as } n \rightarrow \infty. \end{aligned}$$

In this case, for each $j \in I$, we get the following relation $\mu_j \geq \mu_r \lambda_j^r$ for all $j \in I$, so this implies that $\mu \geq \mu_r \lambda^r$.

In addition, if the irreducible Markov chain has all its states recurrent, then, by **Theorem 6.1.8.**, λ^r is invariant, and hence, $\alpha = \mu - \mu_r \lambda^r$ is also invariant and $\alpha \geq 0$. We have Π is irreducible so, given $i \in I$ we can go from state i to state r in n steps, thus $p_{i,r}^{(n)} > 0$. Using this we have $0 = \mu_r - \mu_r \lambda_r^r = \alpha_r = \sum_{j \in I} \alpha_j p_{j,r}^{(n)} \geq \alpha_i p_{i,r}^{(n)}$, and hence $\alpha_i = 0$. This implies that the vector $\alpha = (\alpha_l; l \in I)$ is null because all its components are null, so we get that $0 = \alpha = \mu - \mu_r \lambda^r$ which is the same as $\mu = \mu_r \lambda^r$ where μ_r is a constant, therefore, the invariant measure is unique. $\square$

With the results we have analyzed, we have been studied the existence and uniqueness of the invariant measure. Now, we will introduce a practical example in which we analyze the concepts we have studied in the last two theorems.

Example 6.1.11. Consider the example of the random walk on $\mathbb{Z}$ which we study in Section 4.4. In this case, we have to consider the symmetric random walk, this is $p = \frac{1}{2} = q$ so, the stochastic matrix is

$$\Pi = \begin{pmatrix} & \vdots & \vdots & & \\ \cdots & \frac{1}{2} & 0 & \frac{1}{2} & 0 & 0 & \cdots \\ \cdots & 0 & \frac{1}{2} & 0 & \frac{1}{2} & 0 & \cdots \\ \cdots & 0 & 0 & \frac{1}{2} & 0 & \frac{1}{2} & \cdots \\ & \vdots & \vdots & & \end{pmatrix}$$

In this case, the Markov chain is irreducible and also all its states are recurrent. Now, we have to prove that the measure $\mu = (\mu_i; i \in I)$ is invariant, so it has to satisfy the following equality $\mu \Pi = \mu$ and this is equivalent to $\mu_i = \frac{1}{2} \mu_{i-1} + \frac{1}{2} \mu_{i+1}$ for all $i \in I$. The previous equality holds if $\mu_i = 1$ for all i and hence, the measure μ is invariant.

6.2 Positive recurrence

Now, we analyze the different types of recurrent states which we can find. Moreover, we also study the relationship between these states and the invariant distribution.

First, remember that a state i is recurrent if it satisfies the equality $P_i(X_m = i \text{ for infinitely many } n) = 1$. In chapter 3 we studied that this is equivalent to the following one $\rho_{i,i} = P_i(T_i < \infty) = 1$. To continue, we introduce some new concepts.

Definition 6.2.1. A state $i \in I$ is said to be positive recurrent if, starting at the state i , the expected time till the chain returns to the state i is finite, this is equivalent to $m_i = E_i(T_i) < \infty$.

In the case in which the state $i \in I$ is recurrent and does not satisfy the previous condition, in other words if $m_i = E_i(T_i) = \infty$, the state i is called null recurrent.

In the last definition, the variable m_i is called mean recurrence time at state i , which is the expected return time to this state.

To continue, we analyze the relationship between recurrent states and null recurrent or positive recurrent states.

Corollary 6.2.2. *Positive recurrent states and null recurrent states are both recurrent.*

Proof. For a positive recurrent state, we have $m_i < \infty$ and this means that T_i cannot be ∞ with strictly positive probability. Hence, state i is recurrent. On the other hand, for null recurrent states, is given by definition that these states are recurrent. □

Now, we analyze that for a Markov chain with stochastic matrix Π , to say that the chain has positive recurrent states is equivalent to say that, the stochastic matrix Π has an invariant distribution. In our case, the study focuses on irreducible Markov chains.

Theorem 6.2.3. *Given an irreducible Markov chain with stochastic matrix $\Pi = (p_{i,j}; i, j \in I)$, then the following properties are equivalent*

- (i) *some state i is positive recurrent,*
- (ii) *all states are positive recurrent,*
- (iii) *the stochastic matrix Π has an invariant distribution γ .*

In particular, when (iii) holds we have $m_i = \frac{1}{\gamma_i}$ for all $i \in I$.

Proof. Firstly, we prove that (ii) implies (i). In this case, we have that all states are positive recurrent, hence, each state i is positive recurrent and this shows (i).

Secondly, we prove that (i) implies (iii). Given $r \in I$ a positive recurrent state, then r is recurrent. We know that the chain is irreducible, hence there exists an unique equivalence class and it contains one recurrent state, so all states are recurrent. **Theorem 6.1.8.** tells that λ^r is an invariant measure, using this, we construct an invariant distribution. First note that

$$\begin{aligned} \sum_{i \in I} \lambda_i^r &= \sum_{i \in I} E_r \sum_{n=0}^{T_r-1} \mathbb{1}_{\{X_n=i\}} = E_r \sum_{n=0}^{T_r-1} \sum_{i \in I} \mathbb{1}_{\{X_n=i\}} \\ &= E_r \sum_{n=0}^{T_r-1} 1 = E_r(T_r) = m_r < \infty. \end{aligned}$$

From this equality we can define a new random variable α_j as $\alpha_j = \frac{\lambda_j^r}{m_r}$ where $j \in I$. In this case we have that, $\sum_{j \in I} \alpha_j = 1$ and hence, $\alpha = (\alpha_j; j \in I)$ is an invariant distribution.

To continue, we prove that (iii) implies (ii). Now, we know the Markov chain is irreducible and also the stochastic matrix Π has an invariant distribution γ so, $\sum_{i \in I} \gamma_i = 1$. Given a state $r \in I$, then we have that $\gamma_r = \sum_{i \in I} \gamma_i p_{i,r}^{(n)} > 0$ for any $n \geq 1$. To continue, we can define an invariant measure $\mu = (\mu_i; i \in I)$ as $\mu_i = \frac{\gamma_i}{\gamma_r}$ and hence we have $\mu_r = 1$. Besides, as we know that the Markov chain is irreducible, by **Theorem 6.1.10.** we obtain $\mu \geq \mu_r \lambda^r = \lambda^r$. Therefore, we have

$$m_r = \sum_{i \in I} \lambda_i^r \leq \sum_{i \in I} \frac{\gamma_i}{\gamma_r} = \frac{1}{\gamma_r} < \infty$$

with this we obtain that the state r is positive recurrent which shows (ii).

Finally, we have to prove the equality $m_r = \frac{1}{\gamma_r}$, for this we assume that the properties (i), (ii), and (iii) hold, so, all states are recurrent. By **Theorem 6.1.10.** we have $\mu = \mu_r \lambda^r$ and, using this, we obtain the following equality

$$m_r = \sum_{i \in I} \lambda_i^r = \sum_{i \in I} \frac{\gamma_i}{\gamma_r} = \frac{1}{\gamma_r},$$

with this we obtain the required formula. □

Now, once we have studied the positive recurrent and null recurrent states, which are both recurrent states, we will analyze in an example the result we have just studied.

Example 6.2.4. Consider the example of the random walk on $\mathbb{Z}$ for the case in which $p = \frac{1}{2} = q$, we know that it is an irreducible Markov chain. In **Example 6.1.11.** we saw that there is an invariant measure μ and by **Theorem 6.1.10.**, any invariant measure is a scalar multiple of μ . Now, we analyze whether the symmetric random walk is null recurrent or positive recurrent. We have $\sum_{i \in I} \mu_i = \infty$ so, there can be no invariant distribution and, by **Theorem 6.2.3.**, all states of the walk are null recurrent.

Finally, we introduce an important result which is based on irreducible Markov chains with finite state space, I . In this result, it is established a relationship with positive recurrent states.

Proposition 6.2.5. *Given an irreducible homogeneous Markov chain with finite state space, then it is positive recurrent.*

Proof. First, we show the recurrence. For this we assume that the Markov chain is transient so, for all $i, j \in I$, we get

$$\sum_{n=1}^{\infty} p_{i,j}^{(n)} = \sum_{n=1}^{\infty} P_i(X_n = j) = \sum_{n=1}^{\infty} E_i(\mathbb{1}_{\{X_n=j\}}) = E_i(N(j)) < \infty,$$

and hence, as the state space is finite we have $\sum_{j \in I} \sum_{n=1}^{\infty} p_{i,j}^{(n)} < \infty$, but the previous sum is equal to

$$\sum_{n=1}^{\infty} \sum_{j \in I} p_{i,j}^{(n)} = \sum_{n=1}^{\infty} 1 = \infty,$$

which is a contradiction. Therefore, this implies that Markov chain is recurrent. Now, by **Theorem 6.1.8.** there exists an invariant measure λ^r and, as the state space is finite, $\sum_{i \in I} \lambda_i^r = E_r(T_r) = m_r < \infty$. Therefore, by **Theorem 6.2.3.** the chain is positive recurrent.

□

7 Ergodic theory

In the previous chapter, we have introduced the concept of invariant distribution where, among other things, we have also studied its possible existence. In this section, we focus on doing an analysis of the uniqueness of the invariant distribution which is related to a new concept, the ergodicity.

In this chapter, we should take into account a probability space $(\Omega, \mathcal{F}, P)$ and a family of discrete random variables $X_n : \Omega \rightarrow I$ where $\{X_n; n \geq 0\}$ is a HMC (ν, Π) and I is a countable set called state space.

7.1 Ergodic theorem

Now, we analyze the problem of the uniqueness of invariant distributions. For this is necessary to know that a Markov chain is called ergodic if it is positive recurrent and also we can go from each state to each other. First, we will introduce some results that help us with the required analysis.

Definition 7.1.1. *The stopping time is a random variable $T : \Omega \rightarrow \mathbb{N} \cup \{\infty\}$ such that, for all $m \geq 0$, the event $\{T = m\}$ can be expressed in terms of $X_0, X_1, \dots, X_m$.*

To continue, we use the concept defined above to introduce an important result, the strong Markov property. In this theorem, given a state $j \in I$, we simply wait that the process hits this state at some random point of time, and then, we study what happens thereafter. For this, it is necessary to define the unit mass at j as, $\delta_j = (\delta_{j,i}; i \in I)$ which takes the following values

$$\delta_{j,i} = \begin{cases} 1 & \text{if } j = i \\ 0 & \text{if } j \neq i \end{cases}$$

Theorem 7.1.2. *Given a HMC (ν, Π) , $\{X_n; n \geq 0\}$, and consider a stopping time T of this Markov chain, then, conditional on $T < \infty$ and $X_T = j$, $\{X_{T+n}; n \geq 0\}$ is a new HMC (δ_j, Π) and also independent of the random variables $X_0, \dots, X_T$.*

Proof. Firstly, let T be a stopping time and consider the random variables $X_0, \dots, X_T$ which determine an event A , where $A \subseteq \Omega$ and then, for all $k \geq 0$, $A \cap \{T = k\}$ is determined by $X_0, \dots, X_k$. With this, we have the following

$$\begin{aligned} & P(\{X_T = i_0, X_{T+1} = i_1, \dots, X_{T+n} = i_n\} \cap A \cap \{T = k\} \cap \{X_T = j\}) \\ &= P(\{X_k = i_0, X_{k+1} = i_1, \dots, X_{k+n} = i_n\} \cap A \cap \{T = k\} \cap \{X_k = j\}). \end{aligned}$$

Now, remember the fact that $A \cap \{X_T = j\}$ is determined by $X_0, \dots, X_k$ so, using Markov property at time k and also considering the conditional probability we get

$$\begin{aligned} & P(X_k = i_0, \dots, X_{k+n} = i_n | A \cap \{T = k\} \cap \{X_k = j\}) P(A \cap \{T = k\} \cap \{X_k = j\}) \\ &= P(X_k = i_0, X_{k+1} = i_1, \dots, X_{k+n} = i_n | X_k = j) P(A \cap \{T = k\} \cap \{X_k = j\}) \\ &= p_{j,i_0} p_{i_0,i_1} \cdots p_{i_{n-1},i_n} P(A \cap \{T = k\} \cap \{X_k = j\}). \end{aligned}$$

The last equality is caused by the fact that $\{X_n; n \geq 0\}$ is a homogeneous Markov chain with transition probabilities $p_{i,j}$. To continue, we sum over $k = 0, 1, \dots$ in the equalities we have previously studied, and we obtain

$$\begin{aligned} &P(\{X_T = i_0, X_{T+1} = i_1, \dots, X_{T+n} = i_n\} \cap A \cap \{T < \infty\} \cap \{X_T = j\}) \\ &= p_{j,i_0} p_{i_0,i_1} \cdots p_{i_{n-1},i_n} P(A \cap \{T < \infty\} \cap \{X_T = j\}). \end{aligned}$$

Finally, we divide by $P(\{T < \infty\} \cap \{X_T = i\})$ the equality which we have just calculated and we get

$$\begin{aligned} &P(\{X_T = i_0, X_{T+1} = i_1, \dots, X_{T+n} = i_n\} \cap A | \{T < \infty\}, X_T = j) \\ &= p_{j,i_0} p_{i_0,i_1} \cdots p_{i_{n-1},i_n} P(A | \{T < \infty\}, X_T = j), \end{aligned}$$

and this proves that $\{X_{T+n}; n \geq 0\}$ is a homogeneous Markov chain and also it is independent of the random variables $X_0, \dots, X_T$, as we want. $\square$

To continue, remember that in chapter 4 we introduced the random variable T_i defined as $T_i = \inf\{n \geq 1; X_n = i\}$, which is the first time in which the chain visits the state i and it is known as the passage time. Now, we define the r -th moment in which the chain visits the state i with the following recurrence

$$T_i^{(0)} = 0, \quad T_i^{(1)} = T_i, \quad T_i^{(r+1)} = \inf\{n \geq T_i^{(r)} + 1; X_n = i\}.$$

Now, we use the concept prior analyzed to introduce a new one, the length of the r -th walk between two passage times, T_i . We define it as follows

$$S_i^{(r)} = \begin{cases} T_i^{(r)} - T_i^{(r-1)} & \text{if } T_i^{(r-1)} < \infty \\ 0 & \text{if } T_i^{(r-1)} = \infty \end{cases}$$

To carry on, we will study an outcome which will be useful to prove the ergodic theorem. Before, we note that, for all $r \geq 1$, random variables $T_i^{(r)}$ and $S_i^{(r)}$ are stopping times. This fact is due to $\{T_i^{(r)} = m\}$ is equivalent to say that $X_m = i$ and previously we have passed $r - 1$ times through the state i . With this, we note that $T_i^{(r)}$ only depends on $X_0, \dots, X_m$. Now, we introduce a new result which links the probability of the first passage time with the r -th walk between two passage times.

Proposition 7.1.3. *Given $r = 2, 3, \dots$, conditional on $T_i^{(r-1)}$, the r -th walk $S_i^{(r)}$ is independent of the random variables $\{X_k; k \leq T_i^{(r-1)}\}$, then we have*

$$P(S_i^{(r)} = m | T_i^{(r-1)} < \infty) = P_i(T_i = m).$$

Proof. Firstly, note that, to prove this result we use the strong Markov property. In this case, we consider the stopping time $T = T_i^{(r-1)} < \infty$ of the homogeneous Markov chain $\{X_m; m \geq 0\}$. Then, $X_T = i$ and conditional on T , $\{X_{T+m}; m \geq 0\}$ is a HMC(δ_i, Π) and also independent of $X_0, \dots, X_T$. Once given these properties,

$S_i^{(r)}$ can be defined as the first passage time to state i for the chain $\{X_{T+m}; m \geq 0\}$ so, this is $S_i^{(r)} = \inf\{m \geq 1; X_{T+m} = i\}$.

Therefore, using the properties we have just analyzed, we have the following equality $P(S_i^{(r)} = m | T_i^{(r-1)} < \infty) = P_i(T_i = m)$.

□

Using the concepts studied before, we can conclude that the non-negative random variables $S_i^{(1)}, S_i^{(2)}, \dots$ are independent and identically distributed. Now, we continue considering the passage time and we will introduce a new result which links this random variables with the Markov chains whose states are all recurrent. Remember that a recurrent state, i , is one which satisfies $P_i(T_i < \infty) = 1$.

Proposition 7.1.4. *Given an irreducible Markov chain $\{X_n; n \geq 0\}$ with stochastic matrix Π and all its states recurrent, then for all $i \in I$ we have $P(T_i < \infty) = 1$.*

Proof. Firstly, we know that the first passage time to state i , T_i , is independent of the initial position of the Markov chain, thus, we can rewrite the probability $P(T_i < \infty)$ as follows $P(T_i < \infty) = \sum_{j \in I} P(X_0 = j) P_j(T_i < \infty)$. In this case, we just have to prove $P_j(T_i < \infty) = 1$ for all $j \in I$. Given $n \geq 1$ such that $p_{i,j}^{(n)} > 0$ and by Section 4.2., as all states are recurrent, we have

$$\begin{aligned} 1 &= P_i(X_m = i \text{ for infinitely many } m) = P_i(X_m = i \text{ for some } m \geq n+1) \\ &= \sum_{k \in I} P_i(X_m = i \text{ for some } m \geq n+1 | X_n = k) P_i(X_n = k) \\ &= \sum_{k \in I} P_k(T_i < \infty) p_{i,k}^{(n)}. \end{aligned}$$

We know that $\sum_{k \in I} p_{i,k}^{(n)} = 1$ and by the prior equation, we have that the probability satisfies $P_j(T_i < \infty) = 1$.

Finally, substituting this value, in the following equality, we obtain

$$P(T_i < \infty) = \sum_{j \in I} P(X_0 = j) P_j(T_i < \infty) = \sum_{j \in I} P(X_0 = j) = 1,$$

the last equality is because the sum of probabilities is always 1.

□

To continue, we introduce an important theorem, this is the strong law of large numbers, which will be useful to prove the ergodic theorem. We are not going to demonstrate this theorem because its proof is highly complex and also requires some previous results which are not closely related with Markov chains. An analysis of this theorem can be found in the following book; Nualart, D.; Sanz, M.: *Curs de probabilitats*.

Theorem 7.1.5. Consider $\{Y_n; n \geq 1\}$ a sequence of independent and identically distributed random variables, and also non-negative, with $E(Y_1) = m$, then

$$P\left(\lim_{n \rightarrow +\infty} \frac{Y_1 + \cdots + Y_n}{n} = m\right) = 1.$$

Now, we know that in chapter 4 we introduced the number of visits that the chain does to the state i , denoted by $N(i)$. We define the number of visits to the state i before the instant n as follows

$$N_n(i) = \sum_{r=0}^{n-1} \mathbb{1}_{\{X_r=i\}},$$

then the coefficient $\frac{N_n(i)}{n}$ determines the time that the chain spends in state i , before arriving at instant n . The following theorem examines the long-run value spent by a Markov chain in each state. Moreover, in it we talk about the uniqueness of the invariant distribution.

Theorem 7.1.6. (Ergodic theorem) Consider $\{X_n; n \geq 0\}$, an irreducible Markov chain with stochastic matrix Π and initial distribution ν , then

$$P\left(\lim_{n \rightarrow +\infty} \frac{N_n(i)}{n} = \frac{1}{m_i}\right) = 1.$$

In addition, if the Markov chain is positive recurrent, then given a bounded function $f : I \rightarrow \mathbb{R}$, we have

$$P\left(\lim_{n \rightarrow +\infty} \frac{1}{n} \sum_{r=0}^{n-1} f(X_r) = \sum_{i \in I} \gamma_i f(i)\right) = 1,$$

in this case, the invariant distribution $\gamma = (\gamma_i; i \in I)$ is unique.

Proof. First suppose the Markov chain is transient or null recurrent, then for all $i \in I$, the number of visits to state i is finite, so $\lim_{n \rightarrow +\infty} \frac{N_n(i)}{n} = 0 = \frac{1}{m_i}$ with probability 1.

Now, suppose the Markov chain is recurrent and fix a state i . Considering $T = T_i$, by **Proposition 7.1.4.**, we get $P(T < \infty) = 1$. Now, by strong Markov property, we have the Markov chain $\{X_{T+n}; n \geq 0\}$ with initial distribution $\delta_i = (\delta_{i,j}; j \in I)$ and stochastic matrix Π and also, independent of $X_0, \dots, X_T$. Moreover, both chains X_{T+n} and X_n spent the same time at state i , for this, we only need to consider that $\mu = \delta_i$.

Previously, we have defined the length of the r -th walk to the state i , denoted by $S_i^{(r)}$. To continue, we use this concept to analyze the time of the visit to the state i depending in which instant, n , of the chain we focus. Firstly, we know that the moment of the last visit to the state i before the instant n is $T_i^{(N_n(i)-1)} = S_i^{(1)} + \cdots + S_i^{(N_n(i)-1)}$ and hence, we have

$$S_i^{(1)} + \cdots + S_i^{(N_n(i)-1)} \leq n - 1.$$

Secondly, we also know that the moment of the first visit to the state i after the instant $n - 1$ is $T_i^{(N_n(i))} = S_i^{(1)} + \dots + S_i^{(N_n(i))}$ and now we have

$$S_i^{(1)} + \dots + S_i^{(N_n(i))} \geq n.$$

Finally, using the inequalities which we have just studied, we obtain the following

$$\frac{S_i^{(1)} + \dots + S_i^{(N_n(i)-1)}}{N_n(i)} \leq \frac{n}{N_n(i)} \leq \frac{S_i^{(1)} + \dots + S_i^{(N_n(i))}}{N_n(i)}.$$

Now, by **Proposition 7.1.3.** we get that $E_i(S_i^{(r)}) = m_i$ and using the strong law of large numbers, we have

$$P\left(\lim_{k \rightarrow +\infty} \frac{S_i^{(1)} + \dots + S_i^{(k)}}{k} = m_i\right) = 1,$$

In addition, we have considered that the Markov chain is recurrent, therefore, by chapter 4 we have the following equality

$$P\left(\lim_{n \rightarrow +\infty} N_n(i) = \infty\right) = 1.$$

Therefore, if we apply what we have just introduced in the inequalities previously analyzed, we get the following

$$P\left(\lim_{n \rightarrow +\infty} \frac{n}{N_n(i)} = m_i\right) = 1,$$

and thereby we obtain what we want to show, so $P\left(\lim_{n \rightarrow +\infty} \frac{N_n(i)}{n} = \frac{1}{m_i}\right) = 1$.

To continue, let $\gamma = (\gamma_i; i \in I)$ be an invariant distribution for $\{X_n; n \geq 0\}$, in this case, suppose that the Markov chain has all its states positive recurrent so, $m_i < \infty$. In addition, given the function $f : I \rightarrow \mathbb{R}$, consider that $|f| \leq 1$. With this, for any $J \subseteq I$, we have

$$\begin{aligned} \left| \frac{1}{n} \sum_{r=0}^{n-1} f(X_r) - \sum_{i \in I} \gamma_i f(i) \right| &= \left| \frac{1}{n} \sum_{i \in I} f(i) \sum_{r=0}^{n-1} (\mathbb{1}_{\{X_r=i\}}) - \sum_{i \in I} \gamma_i f(i) \right| \\ &= \left| \sum_{i \in I} \left(\frac{N_n(i)}{n} - \gamma_i \right) f(i) \right| \leq \sum_{i \in J} \left| \frac{N_n(i)}{n} - \gamma_i \right| + \sum_{i \notin J} \left| \frac{N_n(i)}{n} - \gamma_i \right| \\ &\leq \sum_{i \in J} \left| \frac{N_n(i)}{n} - \gamma_i \right| + \sum_{i \notin J} \left(\frac{N_n(i)}{n} + \gamma_i \right) \end{aligned}$$

To carry on, we know that the probability that a Markov chain is in the state i is γ_i . Then, the time period that we spend in this state, we already know it is $\frac{1}{m_i}$, must be equal to γ_i . So we have that $\gamma_i = \frac{1}{m_i}$ and hence the following equality holds $P\left(\lim_{n \rightarrow +\infty} \frac{N_n(i)}{n} = \gamma_i\right) = 1$, for all $i \in I$. Therefore, using this we have that

$$\sum_{i \in J} \left| \frac{N_n(i)}{n} - \gamma_i \right| + \sum_{i \notin J} \left(\frac{N_n(i)}{n} + \gamma_i \right) \leq 2 \sum_{i \in J} \left| \frac{N_n(i)}{n} - \gamma_i \right| + 2 \sum_{i \notin J} \gamma_i$$

Now, let $\epsilon > 0$ and consider J a finite subset, so $J = \{0, 1, \dots, N\}$. Firstly, we have that $\sum_{i \notin J} \gamma_i < \frac{\epsilon}{4}$. Secondly, for $n \geq N$ we get $\sum_{i \in J} \left| \frac{N_n(i)}{n} - \gamma_i \right| < \frac{\epsilon}{4}$. Then, for $n \geq N$ and using what we have just established we obtain the following

$$\left| \frac{1}{n} \sum_{r=0}^{n-1} f(X_r) - \sum_{i \in I} \gamma_i f(i) \right| < \epsilon,$$

and this shows the equality which we want to prove because the required convergence is satisfied.

□

7.2 Finite state space

In this section, we study the uniqueness of the invariant distribution considering the case in which the state space I is finite, so as $I = \{1, \dots, N\}$. For this, we refer to the non degenerate probability distribution of I , which we denote by $\pi = (\pi_i; i \in I)$. This distribution satisfies $0 < \pi_i \leq 1$ and also $\sum_{i \in I} \pi_i = 1$.

Now, we give a definition which is necessary to study the ergodic theorem with finite state space.

Definition 7.2.1. *A homogeneous Markov chain, with stochastic matrix $\Pi = (p_{i,j}; i, j \in I)$, is ergodic if, for all $i \in I$, it satisfies the following property*

$$\pi_j = \lim_{n \rightarrow +\infty} p_{i,j}^{(n)}, \quad \forall j \in I$$

where π_j is a non degenerate probability distribution.

In the previous definition, we observe that the limit is independent of the state $i \in I$. To continue, the ergodicity helps us to study the problem of the uniqueness of the invariant distribution. In the following result we analyze this fact.

Theorem 7.2.2. *Given a homogeneous Markov chain, $\{X_n; n \geq 0\}$, with finite state space I and stochastic matrix Π . Suppose that there exists $k \geq 1$ such that $\min_{i,j \in I} p_{i,j}^{(k)} > 0$. Then, for all $i \in I$, there exists $\pi = (\pi_j; j \in I)$ which satisfies*

$$\pi_j = \lim_{n \rightarrow +\infty} p_{i,j}^{(n)}$$

where π is a non degenerate probability distribution. Furthermore, for all $j \in I$, it satisfies

$$\sum_{l \in I} \pi_l p_{l,j} = \pi_j,$$

with this we have that π is the unique invariant distribution for Π .

Proof. Firstly, for all $n \geq 1$, we denote by $\alpha_j^{(n)}$ and $\beta_j^{(n)}$ the values $\min_{i \in I} p_{i,j}^{(n)}$ and $\max_{i \in I} p_{i,j}^{(n)}$ respectively. Now, using Chapman-Kolmogorov equation and knowing that $\sum_{r \in I} p_{i,r}^{(n)} = 1$, we get

$$\alpha_j^{(n+1)} = \min_{i \in I} p_{i,j}^{(n+1)} = \min_{i \in I} \sum_{l \in I} p_{i,l} p_{l,j}^{(n)} \geq \min_{i \in I} \sum_{l \in I} p_{i,l} \min_{l \in I} p_{l,j}^{(n)} = \min_{i \in I} \sum_{l \in I} p_{i,l} \alpha_j^{(n)} = \alpha_j^{(n)},$$

with this we get that $\{\alpha_j^{(n)}\}_{n \geq 1}$ is an increasing sequence. Then, as $0 \leq p_{i,j} \leq 1$, the above sequence is bounded, therefore exists a limit which we denote by α_j . Repeating the same argument for the maximum, we obtain

$$\beta_j^{(n+1)} = \max_{i \in I} \sum_{l \in I} p_{i,l} p_{l,j}^{(n)} \leq \max_{i \in I} \sum_{l \in I} p_{i,l} \max_{l \in I} p_{l,j}^{(n)} = \max_{i \in I} \sum_{l \in I} p_{i,l} \beta_j^{(n)} = \beta_j^{(n)},$$

and now, the sequence $\{\beta_j^{(n)}\}_{n \geq 1}$ is decreasing. Then, as $0 \leq p_{i,j} \leq 1$, the above sequence is bounded, therefore exists a limit which we denote by β_j .

For the moment, by definition we know that $\alpha_j^{(n)} \leq p_{i,j}^{(n)} \leq \beta_j^{(n)}$, so if it complies $\alpha_j = \beta_j$, then, for all $j = 1, \dots, N$, we will have that the following equality is satisfied, $\lim_{n \rightarrow +\infty} p_{i,j}^{(n)} = \pi_j$ where π_j is equal to α_j and also β_j . Therefore, to show the desired result, it is necessary to study, for all $j = 1, \dots, N$, the following equality

$$\lim_{n \rightarrow +\infty} (\beta_j^{(n)} - \alpha_j^{(n)}) = 0,$$

and then we have that $\beta_j = \alpha_j$. In the following lines, we will analyze the previous equality.

Firstly, we define θ as $\theta = \min_{i,j \in I} p_{i,j}^{(k)} > 0$, then we have the following probability

$$\begin{aligned} p_{i,j}^{(k+n)} &= \sum_{l \in I} p_{i,l}^{(k)} p_{l,j}^{(n)} = \sum_{l \in I} (p_{i,l}^{(k)} - \theta p_{j,l}^{(n)}) p_{l,j}^{(n)} + \sum_{l \in I} \theta p_{j,l}^{(n)} p_{l,j}^{(n)} \\ &= \sum_{l \in I} (p_{i,l}^{(k)} - \theta p_{j,l}^{(n)}) p_{l,j}^{(n)} + \theta p_{j,j}^{(2n)}. \end{aligned}$$

To continue, as $p_{i,l}^{(k)} \geq \theta$ and we know that $p_{j,l}^{(n)} \leq 1$, then we have the following relationship $p_{i,l}^{(k)} - \theta p_{j,l}^{(n)} \geq 0$. Therefore, using this, we obtain

$$\begin{aligned} p_{i,j}^{(k+n)} &\geq \sum_{l \in I} (p_{i,l}^{(k)} - \theta p_{j,l}^{(n)}) \min_{l \in I} p_{l,j}^{(n)} + \theta p_{j,j}^{(2n)} = \sum_{l \in I} p_{i,l}^{(k)} \alpha_j^{(n)} - \sum_{l \in I} \theta p_{j,l}^{(n)} \alpha_j^{(n)} + \theta p_{j,j}^{(2n)} \\ &= (1 - \theta) \alpha_j^{(n)} + \theta p_{j,j}^{(2n)}, \end{aligned}$$

in this equalities we use that $\sum_{j \in I} p_{i,j}^{(n)} = 1$. Hence, using we have just calculated we have $\alpha_j^{(k+n)} \geq \alpha_j^{(n)}(1 - \theta) + \theta p_{j,j}^{(2n)}$.

Using the same argument, we get $\beta_j^{(k+n)} \leq \beta_j^{(n)}(1 - \theta) + \theta p_{j,j}^{(2n)}$.

Now, combining the two inequalities which we calculate, we have $\beta_j^{(k+n)} - \alpha_j^{(k+n)} \leq (\beta_j^{(n)} - \alpha_j^{(n)})(1 - \theta)$. Then, by induction we obtain, for all $r \geq 1$, the following

$$0 \leq \beta_j^{(rk+n)} - \alpha_j^{(rk+n)} \leq (\beta_j^{(n)} - \alpha_j^{(n)})(1 - \theta)^r,$$

and with this we get that $\lim_{r \rightarrow +\infty} (\beta_j^{(n)} - \alpha_j^{(n)})(1 - \theta)^r = 0$ due to $\theta > 0$. Therefore, there exists a sequence $\{n_r\}_{r \geq 1}$ which satisfies

$$\lim_{r \rightarrow +\infty} \beta_j^{(n_r)} - \alpha_j^{(n_r)} = 0.$$

Hence, due to for $\{\beta_j^{(n)} - \alpha_j^{(n)}\}$ each term is less than or equal to the previous one, then we get $\lim_{n \rightarrow +\infty} \beta_j^{(n)} - \alpha_j^{(n)} = 0$ as we want, and hence this implies that $\beta_j = \alpha_j$, which is the desired equality.

Finally, we will study that π is an invariant distribution. Firstly, we need to see that π is a non degenerate probability distribution, for this suppose that $n \geq k$, so $\alpha_j^{(n)} \geq \alpha_j^{(k)} \geq \min_{i,j \in I} p_{i,j}^{(k)} = \theta > 0$, then we get

$$\lim_{n \rightarrow +\infty} \alpha_j^{(n)} \geq \lim_{n \rightarrow +\infty} \theta = \theta \quad \text{and} \quad \pi_j = \alpha_j = \lim_{n \rightarrow +\infty} \alpha_j^{(n)}$$

and this, implies that $\pi_j > 0$ for all $j = 1, \dots, N$.

Now, knowing that $\sum_{j \in I} p_{i,j}^{(n)} = 1$ and using the hypothesis of the theorem, $\pi_j = \lim_{n \rightarrow +\infty} p_{i,j}^{(n)}$, we have the following equality

$$\sum_{j \in I} \pi_j = \sum_{j \in I} \lim_{n \rightarrow +\infty} p_{i,j}^{(n)} = \lim_{n \rightarrow +\infty} \sum_{j \in I} p_{i,j}^{(n)} = 1, \quad \forall i \in I$$

thus, $\sum_{j \in I} \pi_j = 1$. Moreover, we have previously seen that $\pi_j > 0$, therefore we have obtained that $\pi = (\pi_j; j \in I)$ is a non degenerate probability distribution.

To continue, we have to study that the probability distribution is invariant, this is

$$\sum_{r \in I} \pi_r p_{r,j} = \lim_{n \rightarrow +\infty} \sum_{r \in I} p_{i,r}^{(n)} p_{r,j} = \lim_{n \rightarrow +\infty} p_{i,j}^{(n+1)} = \lim_{n \rightarrow +\infty} p_{i,j}^{(n)} = \pi_j,$$

where we have used the Chapman-Kolmogorov equation. Therefore, using what we have just analyzed, we have that $\pi = (\pi_j; j \in I)$ is an invariant distribution. □

Now, we will introduce a practical example in which we analyze the properties that we have studied in this section.

Example 7.2.3. Consider the homogeneous Markov chain of **Example 6.1.3.** Our goal is to determine the number of invariant distributions that the chain has. Firstly, for all $i = 1, 2$, we look at **Example 6.1.6.** where we study the values of the limits of the probabilities $p_{i,1}^{(n)}$ and $p_{i,2}^{(n)}$, which are $\gamma_1 = \frac{4}{19}$ and $\gamma_2 = \frac{15}{19}$ respectively. In this case, we have $\gamma_1 > 0$ and $\gamma_2 > 0$, and also $\gamma_1 + \gamma_2 = 1$, hence, the Markov chain is ergodic. Now, by **Theorem 7.2.2.** the invariant distribution $\gamma = (\frac{4}{19}, \frac{15}{19})$ is unique.

To continue, once studied the ergodic theorem for finite state space, we will analyze that the left implication is also true.

Corollary 7.2.4. *Given a non degenerate probability distribution $\pi = (\pi_j; j \in I)$ over space I which satisfies $\lim_{n \rightarrow +\infty} p_{i,j}^{(n)} = \pi_j$, then exists an integer number $k \geq 1$ for which $\min_{i,j \in I} p_{i,j}^{(k)} > 0$.*

Proof. Firstly, given the equality $\lim_{n \rightarrow +\infty} p_{i,j}^{(n)} = \pi_j$ we have that, for all $j = 1, \dots, N$, the required number of steps to reach the state j is n_j , so for all $n \geq n_j$, $p_{i,j}^{(n)} > 0$ where $i \in I$. Therefore, this implies that for all $j = 1, \dots, N$, $\min_{i \in I} p_{i,j}^{(n)} > 0$. Now, we define $k = \max(n_1, \dots, n_N)$ and then, for all $j = 1, \dots, N$, we get $\min_{i \in I} p_{i,j}^{(k)} > 0$ and this is what we wanted to prove. $\square$

The ergodicity of Markov chains on an arbitrary finite state space can be characterized by the following notion.

Definition 7.2.5. *A Markov chain with finite state space is regular if there exists a power of the transition matrix, Π , with only positive entries.*

To continue, we introduce a practical example in which we analyze the concept which we have just defined.

Example 7.2.6. Consider the Markov chain with state space $I = \{1, 2, 3\}$ and transition matrix

$$\Pi = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{5} & \frac{1}{5} & \frac{3}{5} \\ 1 & 0 & 0 \end{pmatrix}$$

To continue, we will study if the Markov chain is regular or not. For this, we are going to compute the successive powers of the stochastic matrix in order to check if there exists a matrix where all of its entries are non-null. Firstly, we calculate $\Pi^{(2)}$ and we get

$$\Pi^{(2)} = \begin{pmatrix} \frac{7}{20} & \frac{7}{20} & \frac{3}{10} \\ \frac{37}{50} & \frac{7}{50} & \frac{3}{25} \\ \frac{1}{2} & \frac{1}{2} & 0 \end{pmatrix}$$

with this we have an entry which is null, therefore, we have to calculate $\Pi^{(3)}$, which is the following

$$\Pi^{(3)} = \begin{pmatrix} \frac{109}{200} & \frac{49}{200} & \frac{21}{100} \\ \frac{259}{500} & \frac{199}{500} & \frac{21}{250} \\ \frac{7}{20} & \frac{7}{20} & \frac{3}{10} \end{pmatrix}$$

Now, all entries of the matrix are positive, therefore, this implies that the Markov chain is regular.

To continue, using the concept which we have just introduced, the regularity, we will study the uniqueness of the invariant distribution of **Theorem 7.2.2.**

Proposition 7.2.7. *Consider the stochastic matrix $\Pi = (p_{i,j}; i, j \in I)$ which is regular, then the invariant distribution $\pi = (\pi_j; j \in I)$ is unique.*

Proof. Firstly, due to the stochastic matrix is regular, for all $i, j \in I$, we get that $p_{i,j}^{(n)} > 0$ and hence, by **Theorem 7.2.2.** we have the following equality $\lim_{n \rightarrow +\infty} p_{i,j}^{(n)} = \pi_j$. To continue, given that the column vector $\alpha = (\alpha_k; k \in I)$ has all its components equal to 1, then we can define the matrix $Q = (q_{k,j}; k, j \in I)$ as $Q = \alpha\pi$, in other words, each row of Q is the same and it is the non degenerate probability distribution π . Once defined this, we have that $\lim_{n \rightarrow +\infty} \Pi^{(n)} = Q$.

Now, we will study that the invariant distribution π is unique. For this, we suppose that there exists another invariant distribution $\beta = (\beta_i; i \in I)$, therefore, it has to satisfy the following equalities $\beta\Pi = \beta$ and $\beta\Pi^{(n)} = \beta$ for all $n \geq 1$, but as β is an invariant distribution, then $\beta\alpha = \sum_{i \in I} \beta_i = 1$ which implies that

$$\lim_{n \rightarrow +\infty} \beta\Pi^{(n)} = \beta \lim_{n \rightarrow +\infty} \Pi^{(n)} = \beta Q = \beta\alpha\pi = \pi,$$

and hence, $\beta = \pi$ and this show that the invariant distribution $\pi = (\pi_j; j \in I)$ is unique.

□

8 Conclusions

This project has a theoretical approach so it is difficult to get conclusions or specific results. However, the research and analytical work carried out constitute a good framework in order to develop future studies that look for practical conclusions using the basics of Markov chains.

From my point of view, this thesis has been useful in order to analyze the several applications of the probability concepts we have learnt along the degree. For example, we have used the concept of conditional probability and independence in the study of the Chapman-Kolmogorov equation.

Furthermore, I have been able to know the basics of Markov chains, as long as their different properties and characteristics, an aspect which was unknown for me till I carried out this project. However, I have not only learnt specific mathematical aspects due to this project has been a great challenge for me and have allowed me to improve my skills related with the research of information and presentation of ideas in a clear and accurate manner.

Finally, I want to say that the project has been only focused on discrete time Markov chains. Despite this, there exists another type of chains, continuous time Markov chains, which have not been included in the scope of this thesis due to I preferred to focus the analysis in a specific topic. In this regard, it could be interesting to expand the scope of this study and to carry out a similar analysis focused on continuous time chains.

A Basic probability concepts

In this section we will see the most relevant concepts about probability theory which we will be very useful to study Markov chains.

In this chapter we will focus on the probability of countable sets. For this, we will study the concepts of conditional probability and independence. Moreover, we will focus on random variables, concretely discrete variables, due to they are the ones that help us to define the concept of Markov chain.

A.1 Probability of countable sets

We are starting this section giving the general probability definition. After, we will focus on the case of countable sets. Moreover, we also enumerate the most relevant properties related with this probability.

Definition A.1.1. *The triple $(\Omega, \mathcal{F}, P)$ is a probability space which is defined by*

- (i) Ω is the sample space (set of possible outcomes).
- (ii) $\mathcal{F}$ has σ -algebra structure, is a collection of subsets of Ω , this satisfies that
 - $\Omega \in \mathcal{F}$,
 - if $A \in \mathcal{F}$, then $A^c \in \mathcal{F}$,
 - if the sequence $\{A_n; n \geq 1\} \subseteq \mathcal{F}$, then $\bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.
- (iii) P , called probability of $\mathcal{F}$, is an application $P: \mathcal{F} \rightarrow [0, 1]$ which satisfies
 - $P(\Omega)=1$,
 - if $\{A_n; n \geq 1\} \subseteq \mathcal{F}$ disjoint two to two (this is to say $A_i \cap A_j = \emptyset, \forall i \neq j$), then

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{i=1}^{\infty} P(A_i).$$

The definition of probability we just exposed is quite general, so we focus on the case where the structure of the sample space is usually finite (or countable) and always equiprobable.

Definition A.1.2. *Consider $\Omega = \{w_i; i \in I\}$ where I is finite or numerable (in the numerable case $I = \mathbb{N}$). Now, $\{p_i; i \in I\}$ is a probability if $p_i \in [0, 1], \forall i \in I$ and $\sum_{i \in I} p_i = 1$. In this case, the probability of any event $A \subseteq \Omega$ is defined as*

$$P(A) = \sum_{i, w_i \in A} P(\{w_i\}) = \sum_{i, w_i \in A} p_i.$$

Once given the last definition, we have the following result.

Corollary A.1.3. *Assume that $\Omega = \{w_i; i \in I\}$ where I is finite, so $I = \{1, \dots, n\}$, and we consider that the sample space is equiprobable. Then, probabilities $\{p_1, \dots, p_n\}$ take values $p_i = \frac{1}{n}$ for all $i = 1, \dots, n$. Moreover, the probability of any event $A \subseteq \Omega$ is defined as*

$$P(A) = \sum_{i, w_i \in A} p_i = \sum_{i, w_i \in A} \frac{1}{n} = \frac{1}{\#\Omega} \sum_{i, w_i \in A} 1 = \frac{\#A}{\#\Omega}.$$

Now we list some important properties of the probability.

1. $P(\emptyset) = 0, P(\Omega) = 1$.
2. For all $A \in \mathcal{F}$, $P(A) \in [0, 1]$.
3. Assume that $A, B \in \mathcal{F}$ are disjoint, then $P(A \cup B) = P(A) + P(B)$.

This property may be more general if we consider $A_1, A_2, \dots, A_n$ mutually disjoint, in this case we have

$$P\left(\bigcup_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i).$$

4. $P(A^c) = 1 - P(A)$, for all $A \in \mathcal{F}$.
5. Let $A, B \in \mathcal{F}$ and $A \subseteq B$, then $P(A) \leq P(B)$.
6. If $A, B \in \mathcal{F}$, then $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
7. Consider $A_1, \dots, A_n \in \mathcal{F}$, then $P\left(\bigcup_{i=1}^n A_i\right) \leq \sum_{i=1}^n P(A_i)$.

A.2 Conditional probability and independence

In this section, we study two important concepts. Conditional probability and the independence. Firstly we start to talk about the conditional probability.

Definition A.2.1. *Let us assume that $A, B \in \mathcal{F}$ are events with $P(A) > 0$, we can define the probability of B conditional on A as*

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

The following results are three important formulas related to the probability defined above.

1. **Theorem (Compound probability)** If we consider $A_1, \dots, A_n \in \mathcal{F}$ such that $P(A_1 \cap \dots \cap A_{n-1}) > 0$, then

$$P(A_1 \cap \dots \cap A_n) = P(A_1)P(A_2|A_1)P(A_3|A_1 \cap A_2) \cdots P(A_n|A_1 \cap \dots \cap A_{n-1}).$$

2. **Theorem (Total probabilities)** Given the partition $B_1, \dots, B_n$ of Ω ($\{B_i; 1 \leq i \leq n\} \subseteq \mathcal{F}$) with $P(B_i) > 0, \forall i = 1, \dots, n$. Then, for all $A \in \mathcal{F}$ (event) we have

$$P(A) = \sum_{i=1}^n P(A \cap B_i) = \sum_{i=1}^n P(A|B_i) P(B_i).$$

3. **Bayes' theorem** Given $A_1, \dots, A_n$ a partition of Ω ($\{A_i; 1 \leq i \leq n\}$) and B an event (it means $B \in \mathcal{F}$) where $P(A_i) > 0, \forall i = 1, \dots, n$ and $P(B) > 0$. Then for all $i = 1, \dots, n$

$$P(A_i|B) = \frac{P(B \cap A_i)}{P(B)} = \frac{P(B|A_i)P(A_i)}{\sum_{j=1}^n P(B|A_j)P(A_j)}.$$

After that, we study the concept of independence and some of its properties.

Definition A.2.2. Consider two events $A, B \in \mathcal{F}$ which are independents if

$$P(A \cap B) = P(A)P(B).$$

Once given this specific definition of independence, we can consider the general one.

Definition A.2.3. Let us assume that the events $A_1, \dots, A_n \in \mathcal{F}$ are independents if for all finite subsets $\{A_{i_1}, \dots, A_{i_k}\} \subset \{A_1, \dots, A_n\}$ satisfy

$$P(A_{i_1} \cap \dots \cap A_{i_k}) = \prod_{j=1}^k P(A_{i_j}).$$

To continue, we enumerate the properties of the independence.

1. $\emptyset$ and Ω are independents from the others.
2. An event A is independent from itself $\Leftrightarrow P(A) = 0$ or $P(A) = 1$.
3. Two events, A and B , are independents $\Leftrightarrow A^c$ and B are independents $\Leftrightarrow A^c$ and B^c are independents $\Leftrightarrow A$ and B^c are independents.

Now we want to expose a result that links the two concepts we have studied so far.

Proposition A.2.4. Let $A, B \in \mathcal{F}$ events such that their probability are strictly positive, then

$$A \text{ and } B \text{ are independents} \Leftrightarrow P(B|A) = P(B) \text{ and } P(A|B) = P(A).$$

Finally, we explain the concept of conditionally independent and we study the relation which it has with the concept of Markov family.

Definition A.2.5. Consider the events $A_1, A_2, A_3 \in \mathcal{F}$ such that $P(A_3) > 0$, then A_1 and A_2 are conditionally independent given A_3 if

$$P(A_1 \cap A_2 | A_3) = P(A_1 | A_3)P(A_2 | A_3).$$

Definition A.2.6. Given the events $\{A_1, A_2, A_3\}$, these form a Markov family if

$$P(A_3 | A_1 \cap A_2) = P(A_3 | A_2).$$

Observation A.2.7. If we consider A_1, A_2, A_3 like a chronological sequence where: A_1 is the past, A_2 is the present and A_3 is the future; then $P(A_3 | A_1 \cap A_2) = P(A_3 | A_2)$ shows that the dependence relationship only matters in the present (not in the past).

Proposition A.2.8. Let us assume that A_1, A_2, A_3 are events, so $A_1, A_2, A_3 \in \mathcal{F}$, then A_1 and A_1 and A_3 are conditionally independents respect A_2 if, and only if $\{A_1, A_2, A_3\}$ are a Markov family (in other words, $P(A_3 | A_1 \cap A_2) = P(A_3 | A_2)$).

Observation A.2.9. The last proposition can be summarized with the following

$$P(A_1 \cap A_3 | A_2) = P(A_1 | A_2)P(A_3 | A_2) \Leftrightarrow P(A_3 | A_1 \cap A_2) = P(A_3 | A_2).$$

Proof. Firstly, we show the right implication. Given $A_1, A_2, A_3 \in \mathcal{F}$ and using the equality $P(A_1 \cap A_3 | A_2) = P(A_1 | A_2)P(A_3 | A_2)$ we have

$$\begin{aligned} P(A_3 | A_1 \cap A_2) &= \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_1 \cap A_2)} = \frac{P(A_1 \cap A_3 | A_2)P(A_2)}{P(A_1 \cap A_2)} \\ &= \frac{P(A_1 | A_2)P(A_3 | A_2)P(A_2)}{P(A_1 | A_2)P(A_2)} = P(A_3 | A_2). \end{aligned}$$

Now we prove the left implication, let $A_1, A_2, A_3 \in \mathcal{F}$ and consider $P(A_3 | A_1 \cap A_2) = P(A_3 | A_2)$ we may appreciate

$$\begin{aligned} P(A_1 \cap A_3 | A_2) &= \frac{P(A_1 \cap A_2 \cap A_3)}{P(A_2)} = \frac{P(A_3 | A_1 \cap A_2)P(A_1 \cap A_2)}{P(A_2)} \\ &= \frac{P(A_3 | A_2)P(A_1 | A_2)P(A_2)}{P(A_2)} = P(A_3 | A_2)P(A_1 | A_2). \end{aligned}$$

□

A.3 Discrete random variables and expected value

To complete the chapter, we study the discrete random variables. We begin defining the random variable concept and then we focus on the discrete case, for which we define its distribution too. Finally, we talk about the expected value.

Definition A.3.1. Given a probability space $(\Omega, \mathcal{F}, P)$ and a measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, a random variable is a map which transforms elements of the sample space Ω with reals, this means that $X : \Omega \rightarrow \mathbb{R}$ such that

$$\forall A \in \mathcal{B}(\mathbb{R}), X^{-1}(A) = \{w \in \Omega; X(w) \in A\} \in \mathcal{F},$$

where $\mathcal{B}(\mathbb{R})$ is the Borel σ -algebra.

Observation A.3.2. The Borel σ -algebra is the smallest σ -algebra generated by the collection of open or closed sets in $\mathbb{R}$.

Definition A.3.3. The law (or distribution) of a random variable X is the probability $P \circ X^{-1}$ such that

$$\forall A \in \mathcal{B}(\mathbb{R}), P \circ X^{-1}(A) = P(X^{-1}(A)) = P(\omega \in \Omega; X(\omega) \in A).$$

Observation A.3.4. Usually we write P_X instead of $P \circ X^{-1}$, so we write $(P \circ X^{-1})(A) = P_X(A)$.

To continue, we focus on discrete random variables which we wanted to study.

Definition A.3.5. Given a random variable, this is discrete if the set $X(\Omega)$ is finite or countable, this can be written as $X : \Omega \rightarrow I$ where I is finite or countable. Therefore, we can express the discrete random variable like $\{a_i; i \in I\}$.

Definition A.3.6. The law (or distribution) of a discrete random variable is called probability mass function and it is defined by

$$p(a_i) = P(X = a_i), \text{ for } i \in I \text{ and } \sum_{i \in X(\Omega)} p(a_i) = \sum_{i \in X(\Omega)} P(X = a_i) = 1.$$

Definition A.3.7. The set $X_i : \Omega \rightarrow I, i = 1, \dots, n$ of random variables are independent if, for any $x_1, \dots, x_n \in I$

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n P(X_i = x_i).$$

To conclude this section, we talk about the expected value focusing on discrete random variables.

Definition A.3.8. Given a discrete random variable $X, \{a_i; i \in I\}$ where $I \subseteq \mathbb{N}$ such that

$$\sum_{i \in I} |a_i| p(a_i) = \sum_{i \in I} |a_i| P(X = a_i) < \infty$$

we define the expected value of X as $E(X) = \sum_{i \in I} a_i p(a_i) = \sum_{i \in I} a_i P(X = a_i)$.

B Examples

Example B.0.1. Consider a coin repeatedly tossed (p the probability of heads and $q = 1 - p$ that of tails). Let H_n and T_n the number of heads and tails in the first n releases, proof that $X_n = (-1)^{2T_n + H_n}$ is a homogeneous Markov chain.

We define the set of random variables $\forall k = \{1, \dots, n\}$ as

$$\xi_k = \begin{cases} 0 & \text{if } k \text{ is head} \\ 1 & \text{if } k \text{ is tail} \end{cases}$$

which are independents, so we can express the number of heads and tails as following manner

$$H_n = \sum_{k=1}^n \xi_k \quad \text{and} \quad T_n = n - \sum_{k=1}^n \xi_k,$$

this implies that $X_n = (-1)^{2T_n + H_n} = (-1)^{2(n - \sum_{k=1}^n \xi_k) + \sum_{k=1}^n \xi_k} = (-1)^{2n - \sum_{k=1}^n \xi_k}$ and $X_{n+1} = (-1)^{2(n+1) - \sum_{k=1}^{n+1} \xi_k} = X_n(-1)^{2 - \xi_{n+1}} = X_n(-1)^{-\xi_{n+1}}$. Now, using the conditional probability and independence, we see that X_n is a homogeneous Markov chain, $\forall i, j \in I$ we have

$$\begin{aligned} P(X_{n+1} = j | X_n = i) &= P(X_n(-1)^{-\xi_{n+1}} = j | X_n = i) = \frac{P(X_n(-1)^{-\xi_{n+1}} = j, X_n = i)}{P(X_n = i)} \\ &= \frac{P((-1)^{-\xi_{n+1}} = \frac{j}{i}, X_n = i)}{P(X_n = i)} = \frac{P((-1)^{-\xi_{n+1}} = \frac{j}{i})P(X_n = i)}{P(X_n = i)} \\ &= P\left((-1)^{-\xi_{n+1}} = \frac{j}{i}\right) := p_{i,j}. \end{aligned}$$

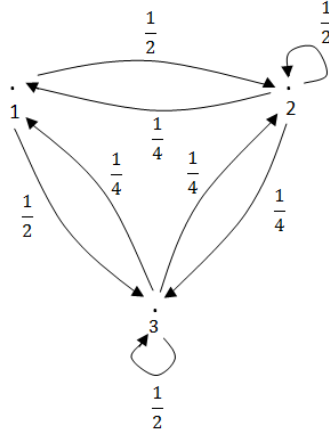
We obtain that $p_{i,j}$ is independent of n so X_n is a homogeneous Markov chain.

Example B.0.2. *Weather in the Land of Oz.* The Land of Oz is wonderful by many reasons, but weather is not one of them. There is never two consecutive days of good weather. If the weather has been good, the next day will be rainy or snowy. If it rains (or snow) there are the same probabilities that the following day the weather changes or remains unchanged, and if it changes, only half the time it would be a good day. Represent the weather in the Land of Oz according to Markov chains.

The stochastic matrix of this problem can be the following one

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

so the associated diagram is

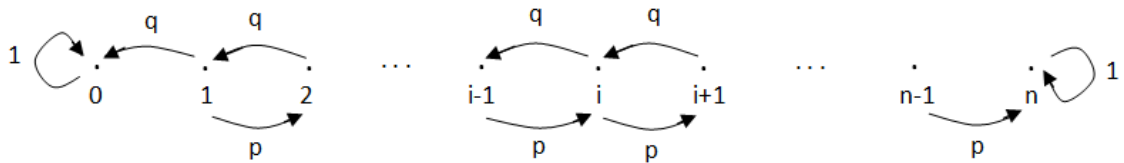


Example B.0.3. *Drunkard's walk.* Consider a completely drunk person who walks between his house and a bench. All the steps along his path can be designated by a whole number between 0 (house) and n (bench). In the spot i there is a bar, which is the starting point of the walk. The drunk person moves leftwards with a probability p , and moves rightwards with probability $q = 1 - p$. When he reaches his house or the bench, he stays there. Describe the drunkard's walk according to Markov chains.

In this case the stochastic matrix is

$$\begin{pmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & q & 0 & p & 0 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & q & 0 & p & 0 & \cdots & 0 \\ 0 & \cdots & 0 & 0 & q & 0 & p & \cdots & 0 \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{pmatrix}$$

and the diagram is represented as the following

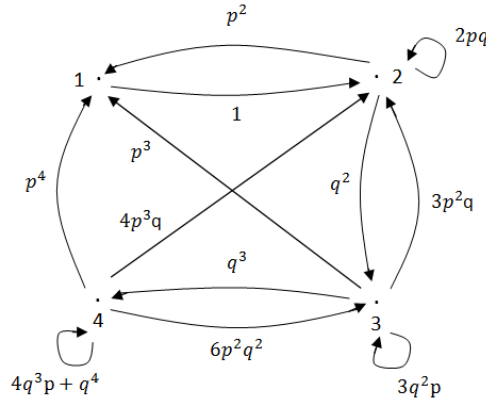


Example B.0.4. A fisherman and his boats. A fisherman owns 4 boats that rents to tourists. Each boat can be damaged with a probability p , independently of the conditions of the other boats. The fisherman can repair his boats at nights, but just one each night. Because of municipal rules, he can not rent any boat if there are not 2 or more boats available. If one day he can rent his boats, he will offer all the boats available. Study the Markov chain associated to this problem.

Given p , the probability that the boat is not damaged is $q = 1 - p$, so the stochastic matrix is

$$\Pi = \begin{pmatrix} 0 & 1 & 0 & 0 \\ p^2 & 2pq & q^2 & 0 \\ p^3 & 3p^2q & 3pq^2 & q^3 \\ p^4 & 4p^3q & 6p^2q^2 & 4q^3p + q^4 \end{pmatrix}$$

and the associated diagram is



Example B.0.5. Consider the following stochastic matrix

$$\Pi = \begin{pmatrix} 1-a & a \\ b & 1-b \end{pmatrix}$$

where $a, b > 0$ and $a + b > 0$. Now, we want to compute the m -step transition probability of every state. For this, is necessary to compute the m -th power of the stochastic matrix, $\Pi^{(m)}$. Firstly, we have to calculate the eigenvalues of Π , for this, we compute the characteristic equation which is

$$\det(\Pi - \lambda Id) = 0 \Rightarrow \lambda^2 - \lambda + (a + b)(\lambda - 1) + 1 = 0$$

thus, solving the equation we obtain the following eigenvalues 1 and $1 - a - b$, whose eigenvalues are $(1, 1)$ and $(a, -b)$, respectively. With this, we can define the following

$$Q = \begin{pmatrix} 1 & a \\ 1 & -b \end{pmatrix}, \quad Q^{-1} = \frac{1}{a+b} \begin{pmatrix} b & a \\ 1 & -1 \end{pmatrix}, \quad D = \begin{pmatrix} 1 & 0 \\ 0 & 1-a-b \end{pmatrix}.$$

Now, we can rewrite the transition matrix Π as follows $\Pi = QDQ^{-1}$, and hence, this implies that $\Pi^{(m)} = QD^{(m)}Q^{-1}$ so, we get

$$\begin{aligned}\Pi^{(m)} &= \frac{1}{a+b} \begin{pmatrix} 1 & a \\ 1 & -b \end{pmatrix} \begin{pmatrix} 1^m & 0 \\ 0 & (1-a-b)^m \end{pmatrix} \begin{pmatrix} b & a \\ 1 & -1 \end{pmatrix} \\ &= \frac{1}{a+b} \begin{pmatrix} b & a \\ b & a \end{pmatrix} + \frac{(1-a-b)^m}{a+b} \begin{pmatrix} a & -a \\ -b & b \end{pmatrix},\end{aligned}$$

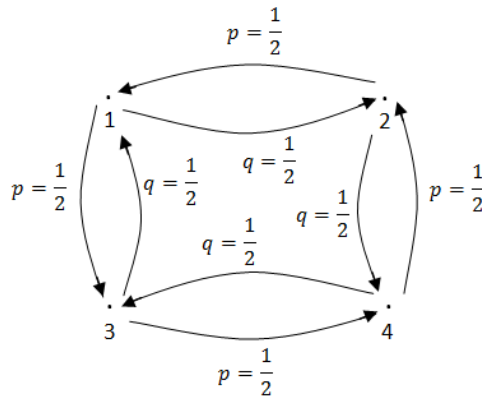
with this we have that each matrix entry is the m -step transition probability of going from one state to another.

Example B.0.6. *Guardian of the tower.* A guardian can watch from the four corners of a tower in the following manner: after staying 5 minutes in a corner he throws a coin to the air to determine if he moves to the left (heads) or he moves to the right (tails). This process is repeated to the infinity. Study the Markov chain associated to the process. If $p = \frac{1}{2}$ (p is the probability of obtaining heads) compute Π^n .

Given $p = \frac{1}{2}$, the probability of obtaining tails is $q = 1 - p = \frac{1}{2}$, so the stochastic matrix is

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

and the associated diagram is



In this case, to obtain the matrix $\Pi^{(n)}$ we have to consider two cases, if n is an even number, then the stochastic matrix is

$$\Pi^{(n)} = \begin{pmatrix} \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix}$$

if n is an odd number we have the stochastic matrix

$$\Pi^{(n)} = \begin{pmatrix} 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & 0 & \frac{1}{2} \\ \frac{1}{2} & 0 & \frac{1}{2} & 0 \end{pmatrix}$$

(a) Suppose that the starting corner is selected randomly. If X_n is the random variable which defines where the guard is located after $5n$ minutes, compute the X_n law.

In this case, as the start is selected randomly we have $X_0 = i$ where $i = \{1, 2, 3, 4\}$ and the probability is $P(X_0 = i) = \frac{1}{4}$, so the vector known as initial distribution is $(\frac{1}{4}, \frac{1}{4}, \frac{1}{4}, \frac{1}{4})$ and the law for X_n is

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

which is the same if we consider n an even number or an odd number

(b) Now suppose that the guard starts in a fixed corner.

In this case, we suppose that the guard starts in the first corner, so the vector is $(1, 0, 0, 0)$ and we have to study the law for two cases. Firstly, we consider n an even number, the law for X_n is

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

Secondly, if n is an odd number, then the law for X_n is

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

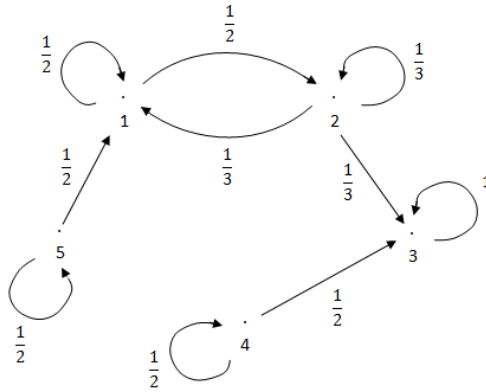
Example B.0.7. Consider the Markov chain with the state space $I = \{1, 2, 3, 4, 5\}$ and transition matrix

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

In this case, there are four communication classes which are

$$C_1 = \{1, 2\}, C_2 = \{3\}, C_3 = \{4\}, C_4 = \{5\}.$$

To continue, we analyze the communication relationship that exists between the different states. For this, we use the following diagram



The class C_1 is recurrent because once we reach the class C_1 we are not able to give up this class, and also it is finite and irreducible. The state of the class C_2 is recurrent, in particular it is an absorbing state, because if we start in state 4, we are able to move to state 3 and if we start in state 3 we are always on it. The state of the class C_3 is transient because if we start in state 4, we are able to move to state 3 and once there we cannot return back to state 4. The state of the class C_4 is transient because if we start in state 5, we are able to move to state 1 and once there we cannot return back to state 5.

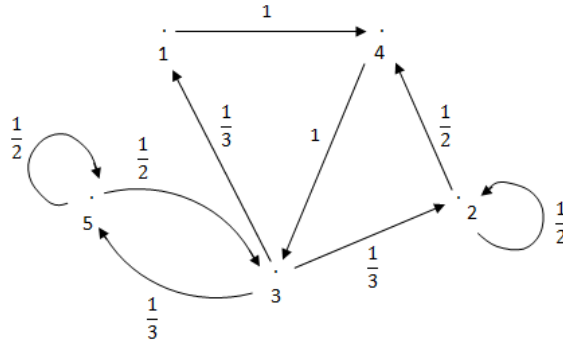
Example B.0.8. Consider the Markov chain with the state space $I = \{1, 2, 3, 4, 5\}$ and transition matrix

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

In this case, there are one communication class which is

$$C_1 = \{1, 2, 3, 4, 5\}.$$

To continue, we analyze the communication relationship that exists between the different states. For this, we use the following diagram

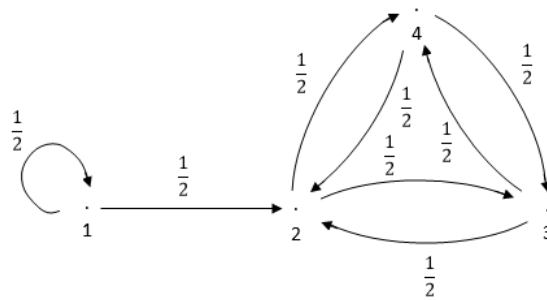


In this case we have an unique class, C_1 , which contains all the states of the Markov chain. Moreover, using the previous diagram it is easy to see that all the states communicate, so the chain is irreducible, in particular the class C_1 is recurrent because we can visit every state infinite times.

Example B.0.9. Consider the Markov chain with the state space $I = \{1, 2, 3, 4\}$ and transition matrix

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

and the associated diagram is



In this case, we have to study if the states of the Markov chain are essential or inessential. First, we note that being in state 1 we can leave but we cannot return to it, so this state is inessential. Now, being in state 2 we can go to state 3 and once there we can return to the initial state or go to state 4. If we go to state 4, from there, we can return to the initial state. With this we have that state 2 is essential. The same argument is valid for states 3 and 4, which are also essentials.

References

- [1] Ash, R.B.: *Basic probability theory*, New York: Wiley, 1970.
- [2] Brémaud, P.: *Markov chains : Gibbs fields and Monte Carlo simulation, and queues*, New York: Springer, 1999.
- [3] Ching, W.; Huang, X.; Ng, M.K.; Siu, T.K.: *Markov chains models, algorithms and applications*, 2nd ed., Boston, MA: Springer, 2013.
- [4] Freedman, D.: *Markov chains*, New York: Springer, 1983, Reprint of 1971 Holden-Day edition.
- [5] Kemeny, J.G.; Snell, J.L.: *Finite Markov chains*, Princeton, N. J: Van Nostrand, 1960.
- [6] Norris, J.R.: *Markov chains*, Cambridge University Press, 1997.
- [7] Sanz, M: *Probabilitats*, Edicions Universitat de Barcelona, 1999.
- [8] Schmidt, V.: *Markov Chains and Monte-Carlo Simulation*,
[http : //www.mathematik.uniulm.de/stochastik/lehre/ss06/markov/skript_engl/node10.html](http://www.mathematik.uniulm.de/stochastik/lehre/ss06/markov/skript_engl/node10.html), May 2016, Ulm University, 2006.
- [9] Stroock, D.W.: *An Introduction to Markov Process*, 2nd ed., Berlin: Springer, 2014.