Eötvös Loránd University

Faculty of Science

László Udvardi

Discrete time Markov chains and a few tools from linear algebra

BSc thesis

Consultant:

Villő Csiszár

Department of Probability and Statistics

Budapest, 2018

Contents

1. Introduction

The evolution of probability theory has come a long way since its early stages. From describing simple gambling models it has grown into a tool applied in various areas of social and natural sciences. The nature of most of the phenomena for which this field of mathematics was applied directed the course of research towards models which can handle the changes by time in our world. One of the most useful tools was introduced by Andrei Andreyevich Markov in 1907[1] with an introduction of a certain type of stochastic process which retains no memory of the previous states. Such process is called a Markov process. The findings related to Markov's original idea, the Markov property, are the basis of the most widely used stochastic models. In physics, especially in particle physics and statistical physics arose the technique of Markov Chain Monte Carlo (MCMC), introduced in the celebrated paper of Metropolis, Rosenbluth, Rosenbluth, Teller and Teller[2]. Later this method was also used in statistics in the work of Gelfand and Smith [3]applying effectively MCMC to Bayesian problems. But it is just one of the many ways of application. In information technology the application of a model called Markov chain is introduced by Howard[4] for problems in dynamic programming.

The reason behind the popularity of these models lies in their relative simplicity. The Markov process, because of its memoryless property we only need information about the current state. It makes it easier to make predictions for the behaviour of the process in the future. For simplicity in my thesis I suppose that the process can only assume countably many states. In this case the Markov Process is called a Markov chain.

In the present study my aim is to introduce the most important concepts regarding Markov chains and, with the help of linear algebra provide a few basic computational tools for the properties I will describe. I have also narrowed the topic down to discrete time Markov chains. These methods however, are still far from the real applications but essential for their understanding and provide a useful basis for further research. I have divided my work to two main sections. The first one, after a brief introduction, aims to give a perspective about the most important properties of Markov chains with a few essential theorems that give us an idea about how to handle them. In the second chapter I will show how to make basic computations regarding different chains and also give a few methods of making claims about the properties described in the first chapter.

2. An introduction to Discrete Time Markov Processes

2.1 Basic Definitions

In this chapter I would like to introduce the basic definitions regarding Markov chains and for the sake of generality, Markov processes. In this work it is not my intention to build everything up from the bases of probability theory but I think the next definition would be a good starting point.

2.1.1. Definition. *Let* $(\Omega, \mathcal{F}, \mathrm{P})$ *be a probability space and* (I,Σ) *some measurable space (often called the state space). Then the set of measurable functions* $X_t : \Omega \mapsto I, t \in T$, *is called a stochastic process with the index set* T.

2.1.2. Remark. If T is some discrete set then it is called a discrete stochastic process but it can also be an interval of real numbers or even the Cartesian plane.

Stochastic processes are easiest to imagine as we are trying to model the outcome of random events changing in time. When T is discrete we are taking the outcomes in given moments.

Markov Processes are types of stochastic processes with some special properties. The most important property used for their description is the following.

2.1.3. Definition. *Let $(X_t)_{t \in T}$ be a stochastic process,where T is an index set, I a countable state space and $\mathcal{F}_m = \sigma(X_i) : i \leq m$ generated $\sigma - algebras$, this is called filtration. In this case we say that the process has the Markov Property if the next equation holds for all* $\mathrm{B} \subseteq I$ *and for all* $m \geq n$.

$$\mathrm{P}(X_m \in \mathrm{B} \mid \mathcal{F}_n) = \mathrm{P}(X_m \in \mathrm{B} \mid X_n) \tag{2.1}$$

[5]

2.1.4. Remark. In my thesis I am only concerned with discrete time Markov processes where the index set T is a discrete set (let us say $T \subseteq \mathbb{N}$). Using that the $\mathcal{F}_n$ $\sigma - algebras$ are atomic we can rephrase the Markov property the following way.

$$\mathrm{P}(X_m = i_m \mid X_1 = i_1, X_2 = i_2, ...X_n = i_n) = \mathrm{P}(X_m = i_m \mid X_n = i_n) \tag{2.2}$$

The above definition describes the 'memoryless' property of Markov processes, which means that the following events depend only the last known outcome and independent from the ones preceding it.

2.1.5. Example. The simplest example for a Markov Chain is a random walk on the line of integers starting from zero. In this case transitions are only possible between neighbouring numbers. Suppose we are currently standing on number k from here we can go to $k+1$ with probability p or go to $k-1$ with probability $1-p$. It is easy to see that this process has the Markov property since the next state depend only the current state we are in.

We can come across various, a bit different definitions for Markov Property depending on which book we are reading. They are all equivalent but in some cases it is easier to prefer one over the others.

2.1.6. Claim. *The following definitions of Markov Property in the discrete time case are equivalent .*

1.

$$\mathrm{P}(X_m \in \mathrm{B} \mid \mathcal{F}_n) = \mathrm{P}(X_m \in \mathrm{B} \mid X_n) \tag{2.3}$$

2.

$$\mathrm{P}(X_{n+1} \in \mathrm{B} \mid \mathcal{F}_n) = \mathrm{P}(X_{n+1} \in \mathrm{B} \mid X_n) \tag{2.4}$$

Proof.

1. $1 \Rightarrow 2$ is trivial since if $m = 1$ the equation still holds.

2. $2 \Rightarrow 1$ For the proof it is enough to see that $\mathrm{P}(X_m = i_m \mid \mathcal{F}_n)$ only depends from X_n and on the preceding parts of $\mathcal{F}_n$. Using the Marginal Probability Theorem we can write the following equation.

   $$\mathrm{P}(X_m = i_m \mid \mathcal{F}_n) = \sum_{j_{m-1} \in I} \mathrm{P}(X_m = i_m \mid (X_{m-1} = j_{m-1}) \cap \mathcal{F}_n)\mathrm{P}(X_{m-1} = j_{m-1} \mid \mathcal{F}_n) \tag{2.5}$$

   In the first part of the summand we can leave $\mathcal{F}_n$ from the equation since, because of the one-step Markov property (2), the mth X_m only depends on the previous state. This leaving us with

   $$\begin{aligned}\mathrm{P}(X_m = i_m \mid X_{m-1} = j_{m-1} \cap \mathcal{F}_n) = \mathrm{P}(X_m = i_m \mid X_{m-1} = j_{m-1}) = \\ = \mathrm{P}(X_m = i_m \mid X_{m-1} = j_{m-1} \mid X_n)\end{aligned} \tag{2.6}$$

   Because of the above mentioned reasons we can even replace $\mathcal{F}_n$ with x_n thus we proved that this part of the summand is realy independent from the states preceding X_n.

   To prove the whole claim we need to prove this to the second part as we. For this we will use induction. Because of (2) we know that for $m = n + 1$ X_m only depends on X_n, for $m = n + 2$ it only depends on X_{n+1} which only depends on X_n, and so on.

□

In the definitions mentioned earlier, the transition probabilities depend on the 'time' described by the index sets. But there are books where it is defined a bit more (for example in the one written by R. Gallager, see in Ref.[6]) specifically in the following way.

2.1.7. Definition. *A stochastic process which has the following property:*

$$\mathrm{P}(X_{n+1} = i_{n+1} \mid X_1 = i_1, X_2 = i_2, ...X_n = i_n) = \mathrm{P}(X_{n+1} = i_{n+1} \mid X_n = i_n) = p_{i_n, i_{n+1}} \quad \forall n \tag{2.7}$$

is called a Homogeneous Markov Chain

Most commonly when people talk about Markov Chains they are really thinking about the homogeneous ones this way with the transition probabilities becoming independent from time allowing us to introduce convenient methods for different calculations (more of this in Chapter 3). For the sake of simplicity in the following part of my work I also postulate homogeneity.

In a homogeneous Markov process using Definition 2.1.7 as a sample to denote the higher order transition probabilities the following way.

$$\mathrm{P}(X_{n+m} = i_{n+m} \mid X_1 = i_1, X_2 = i_2, ...X_n = i_n) = \mathrm{P}(X_{i_{n+m}} \mid X_n = i_n) = p^{(m)}_{i_1 i_{n+m}} \tag{2.8}$$

There is one other important factor which determines a Markov Process is its initial distribution.

2.1.8. Definition. *Let λ be a probability measure on the state space which tells us the distribution of X_0.*

There is a special case of initial distributions called the stationary distribution. Again, for the sake of simplicity I will only introduce the definition concerning discrete Markov processes.

2.1.9. Definition. *Let* $(X_n)_{n\in\mathbb{N}}$ *be a Markov process with the initial distribution* π*. We say that* π *is a stationary distribution (or invariant distribution) if* $\forall m \in \mathbb{N}$ $(X_{m+n})_{n\in\mathbb{N}}$ *also has the Markov property with also the initial distribution* π*.*

It basically the displacement invariant property of the discrete distributions rephrased. In the second chapter of my thesis I will introduce a more convenient way of looking at this distribution.

Markov chains have multiole properties that make them useful tools in dealing with different mathematical and real world problems. To understand how they are applicable, we will introduce a few more definitions in this section.

There are ways to divide the states into smaller partitions in which some statements hold for all states (or for none of them). With this division it is easier to examine the possible outcomes of a process. To clarify exactly what type of statements are the ones in question we first have to introduce the notions below.

2.1.10. Definition.

1. *We say that a state* i *leads to state* j *(written as* $i \to j$*) if* $\exists n, m$ $\mathrm{P}(X_{m+n} = j \mid X_n = i) > 0$
2. *Two states communicate (written as* $\leftrightarrow$*) if* $i \to j$ *and* $j \to i$
3. *A state* i *is significant if* $i \to j \Rightarrow j \to i$

We can see that communication (2.) is an equivalence relation since it is reflexive, transitive and symmetric. The equivalence classes determined by this relation will be called the classes of the Markov chain.

2.1.11. Definition. *A* $C \subset I$ *class is said to be closed if* $\forall i \in C, j \in I : i \to j \Rightarrow j \in C$

2.1.12. Definition. *We say that the state i is recurrent if*

$$\mathrm{P}(X_n = i \; for \; infinitely \; many \; n \;\; | \; X_0 = i) = 1 \tag{2.9}$$

and we say that it is transient if

$$\mathrm{P}(X_n = i \; for \; infinitely \; many \; n \;\; | \; X_0 = i) = 0 \tag{2.10}$$

2.1.13. Remark. There are a few different definitions for this property in various books, the one I am using is in Ref.[7]

In other words, a process reaching a recurrent sate will eventually come back to it with a probability of one while a transient state will eventually be left forever. If we were more concerned with infinite state Markov Chains we should divide the above definition for more cases depending on the expected time of recurrence.

2.1.14. Remark. In the finite state-space case if a state is transient it means it is always possible to go to another state from where there is no return. After this, it is easy to see that if there are only finitely many states there must be at least one recurrent state.

2.1.15. Definition. *The greatest common divisor of the values n for which $p_{ii}^{(n)} > 0$ is called the period of the state i, denoted as $d(i)$. When $d(i) = 1$ the state is called aperiodic, when there is no such n, the period is not defined.*

With the above properties introduced we set the field for the following definition which will give us a useful tool to prove a few interesting theorems.

A property is called a class property, if either all states in the same class have it or none of them has it. For examples see below.

2.1.16. Example. 1. All the significant states are forming a class.

Proof. Suppose i and j are in the same class and i is significant while j is not. In this case $i \to j$, because they are in the same class, and $\exists k$: $j \to k$ but not $k \to j$, because j is not significant. On the other hand $i \to j \to k$ and because of j's significance $k \to i \to j$ which is a contradiction. □

2. Being transient or recurrent are also class properties.

   2.1.17. Remark. This will be proved later on for now, let us just remain with the fact that in the finite state case there is no difference between a state being significant or recurrent.

3. States being in the same class also have the same period (in this case it is called the period of the class).

   Proof. (source: Gallager[6])

   Suppose the states i and j are in the same class C. In this case $\exists r$ for which $p^r_{ij} > 0$ and $\exists p$ for which $p^p_{ji} > 0$ since $i \leftrightarrow j$. In this case $p^{r+p}_{ii} > 0$ so $r + p$ must be divisible by $d(i)$. Let t be an integer for which $p^t_{jj} > 0$ so $r + t + p$ must be divisible by $d(i)$ but in this case t must be also divisible by it. Since it is true for all t then $d(i) \mid d(j)$. With reversing the roles we can see that $d(i) = d(j)$ and it is true for all i and j □

2.2 Recurrence and Transience

Let us take a step back and go into the topic of recurrence and transience a bit further. Previously we proposed the question of recurrence as an interesting property to be examined. Now suppose we successfully decided whether a state (a class) is recurrent here pops up the question that how much time is expected to be passed until a recurrent state comes up again. (Note that this question is strongly related to for example determining a period of a state.) For this, first we have to know when was the first occurrence of the state in question (suppose we did not start from there).

2.2.1. Definition. *$T_i : \Omega \to \mathbb{N}$ is called first passage time and $T_i(\omega) = \inf(n \geq 1 : X_n(\omega) = i)$*

Inductively, with $T_i^{(0)} = 0$ and $T_i^{(1)} = T_i$ we can define rth passage time for arbitrary $r \in \mathbb{N}$.

2.2.2. Definition. $T_i^{(r)}(\omega) = \inf(n \geq T_i^{(r-1)} + 1 : X_n(\omega) = i)$

To be able to deduct any properties of passage time we first have to take a little turnout towards a bit more abstract version of Markov property. But for this first we need to introduce stopping time.

2.2.3. Definition. *A stopping time is a random variable* $\tau : \Omega \to \mathbb{N}$ *for which* $\forall n :$ $(\tau = n) \in \mathcal{F}_n$ *where* $\mathcal{F}_n$*is still the* $\sigma(X_1, X_2, ...X_n)$ *generated* $\sigma - algebra$*. In this case* $X_\tau : \omega \mapsto X_{\tau(\omega)}(\omega)$ *is valid as a random variable.*

For the sake of easier understanding I will provide a simple example for a stopping time. Let us say we are rolling a die a waiting for the outcome six. Now if we denote the first time we throw for the outcome of six with k we can easily see that k is a random variable and $(k = n)$ only depends on the first n rolls. Denote the outcome of the ith roll with the random variable X_i, with this notation we can easily see that k satisfies the above definition.
We have seen the Markov property as a mean of starting a new stochastic process from a given time n and after that there are no effects of what happened before the nth event. But what happens when we do not know exactly the time from when we wish to examine the process? For example we want to start the process from a given state instead of a given time. Are there any similar rules which apply here? With the next definition we state a very similar thing to Markov property for the case described here.

2.2.4. Definition. (Strong Markov property) *A stochastic process is said to has the strong Markov property is for all finite stopping time* τ*,* B *measurable set and* $\forall\ k \geq 0$

$$\mathrm{P}(X_{\tau+k} \in B \mid \mathcal{F}_\tau) = \mathrm{P}(X_{\tau+k} \in B \mid X_\tau) \qquad (2.11)$$

2.2.5. Remark. It is easy to see if $\tau = n$ with probability one we get the Markov property. Since this definition applies for all τ random variables it is a stronger property.

With the strong Markov property introduced we can prove a quite essential property of Markov chains through a few lemmas with passage time. In the first lemma for the sake of simplicity introduce a notation for the time spent between two recurrences of the state i.

$$S_i^{(r)} = \begin{cases} T_i^{(r)} - T_i^{(r-1)} & \text{if } T_i^{(r-1)} < \infty \\ 0 & \text{otherwise} \end{cases} \tag{2.12}$$

2.2.6. Remark. These notation and the corresponding theorems we will use can be found in J. R. Norris' book in section 1.5.[7] The proofs provided by Norris are are the basis of the ones provided in the following.

2.2.7. Lemma. $\forall\ r = 2, 3, ...$ *and* $\forall\ n \in \mathbb{N}^+$

$$\mathrm{P}(S_i^{(r)} = n \mid T_i^{(r-1)} < \infty) = \mathrm{P}(T_i = n \mid X_0 = i) \tag{2.13}$$

Proof. We have seen before that the rth passage time is a stopping time and now we have to prove that an arbitrary discrete Markov chain has the strong Markov property with $\tau = T_i^{(r-1)}$. Shoving $X_{\tau+k}$ is independent from $X_0, X_1, ..., X_\tau$ is the crucial part of proving strong Markov property. If $B \in \sigma(X_0, X_1, ..., X_\tau)$ that means $B \cap \{\tau = m\} \in \sigma(X_0, X_1, ...X_m)$, now using $(X_n)_{n\in\mathbb{N}}$ is Markov we get the following.

$$\mathrm{P}(\{X_{\tau+k} = j_k\} \cap B \cap \{\tau = m\} \cap \{X_\tau = i\}) = \tag{2.14}$$

$$\mathrm{P}(X_k = j_k \mid X_0 = i)\mathrm{P}(B \cap \{\tau = m\} \cap \{X_\tau = i\}) \tag{2.15}$$

Now we sum it over $m = 0, 1, ...$ and divide by $\mathrm{P}(\tau < \infty, X_\tau = i)$ and obtain:

$$\mathrm{P}(\{X_{\tau+k} = j_k\} \cap B \mid \tau < \infty, X_\tau = i) = \tag{2.16}$$

$$\mathrm{P}(X_k = j_k \mid X_0 = i)\mathrm{P}(B \mid \tau < \infty, X_\tau = i) \tag{2.17}$$

Now we can apply the strong Markov property with the above-defined stopping time for our initial problem. On the condition τ is finite $(X_{\tau+k})_k \geq 0$ has the Markov property and independent from $X_0, X_1, ...X_\tau$. So because

$$S_i^{(r)} = \inf\{n \geq 1 : X_{\tau+n} = i\} \tag{2.18}$$

S_i is the first passage time. □

For the second lemma let us introduce the number of visits (V_i) which a process takes in state i. The easiest way to define it is with indicator functions and it follows as

2.2.8. Definition.

$$V_i = \sum_{n=0}^{\infty} I_{\{X_n=i\}} \tag{2.19}$$

For further definitions let us introduce the following notations.

$$f_{ij}^{(n)} = \mathrm{P}(X_n = j, X_k \neq j : k = 1, 2, ...n-1 \mid X_0 = i)\ n \geq 1,\ f_{ij}^{(0)} = 0 \tag{2.20}$$

$$f_{ij}^{*} = \sum_{n=1}^{\infty} f_{ij}^{(n)} \tag{2.21}$$

2.2.9. Lemma. *For* $r = 0, 1, ...$ *we have* $\mathrm{P}(V_i > r \mid X_0 = i) = (f_{ii}^{*})^r$

Proof. We can observe that if $X_0 = i$ then $\{V_i > r\} = \left\{T_i^{(r)} < \infty\right\}$, since if we want the process to return to i more than r times it means the rth return must occur in a finite time. For $r = 0$ the statement is true and for $r \geq 1$ we can prove it by induction. Now suppose for r the statement holds, then

$$\mathrm{P}(V_i > r+1 \mid X_0 = i) = \mathrm{P}(T_i^{(r+1)} < \infty \mid X_0 = i) = \tag{2.22}$$

$$\mathrm{P}(T_i^{(r)} < \infty \textit{ and } S_i^{(r+1)} < \infty \mid X_0 = i) = \tag{2.23}$$

$$\mathrm{P}(S_i^{(r+1)} < \infty \mid T_i^{(r)} < \infty, X_0 = i)\mathrm{P}(T_i^{(r)} < \infty \mid X_0 = i) \tag{2.24}$$

Using Lemma 2.2.7 and the fact that $f_{ii} = \mathrm{P}(T_i < \infty \mid X_0 = i)$, we can continue the equation and get $f_{ii} f_{ii}^{r} = f_{ii}^{r+1}$ □

Using these two lemmas we can prove that in a discrete Markov chain all states are either transient or recurrent which may seem a little obvious but it is a crucial cornerstone of any finding involving these two properties.

2.2.10. Theorem. *1. if* $\mathrm{P}(T_i < \infty \mid X_0 = i) = 1$ *(equivalently if* $f_{ii}^* = 1$*) then* i *is recurrent.*

2. if $\mathrm{P}(T_i < \infty \mid X_0 = i) < 1$ *(*$f_{ii}^* < 1$*) then* i *is transient. As a consequence we have proven the fact that all states are either transient or recurrent.*

Proof. If $\mathrm{P}(T_i < \infty \mid X_0 = i) = 1$ then, if we apply Lemma 2.2.7, we get

$$\forall r \in \mathbb{N}: \ \mathrm{P}(T_i > r \mid X_0 = i) = 1 \tag{2.25}$$

Which means that that state i is going to be recurrent.

For the second part let us us the same lemma again with $\mathrm{P}(T_i < \infty \mid X_0 = i) < 1$. Now from the lemma we have

$$\sum_{r=0}^{\infty} \mathrm{P}(V_i > r \mid X_0 = i) = \sum_{r=0}^{\infty} (f_{ii}^*)^r = \frac{1}{1 - f_{ii}^*} < \infty \tag{2.26}$$

And for this we need $\mathrm{P}(V_i = \infty \mid X_0 = i) = 0$ so the state i must be transient. □

Now we have introduced everything that is necessary to prove 2.1.16/2).

Proof. Now suppose i, j are in the same class and i is recurrent. Because $i \to j$ and because of $f_{ii}^* = 1$ we know that also $f_{ji}^* = 1$ which means starting from j we will eventually reach i. After we have hit i we will return to it infinitely many times with probability one. Now define the following event

$$A_r = \left\{ \exists m : T_i^{(r-1)} < m < T : i^{(r)} : X_m = j \right\} \tag{2.27}$$

And, since the events $\{A_r\}_{r \geq 1}$ are independent because of the strong Markov property, we can apply the Borel-Cantelli lemma. By the lemma we know that A_r will occur infinitely many times since $\mathrm{P}(A_r) > 0 \ \forall r$.

Finally, we got that starting from j, the chain will return to j infinitely many times with probability one, thus j is also recurrent. □

2.2.11. Definition. *We say that a Markov chain is irreducible if it consists of one class (alternatively:* $\forall i, j \in I f_{ij} > 0 \Leftrightarrow \exists n \in \mathbb{N} for\ which P_{ij}^{(n)} > 0$*[8])*

2.2.12. Theorem. *If a Markov chain is irreducible exactly one of the following holds.*

1. *All states are transient*

2. *All states are recurrent*

For this theorem the proof comes from the fact that recurrence and transience are class properties.

3. Linear Algebraic Tools

In the previous chapter we introduced a few important properties of Markov chains. These definitions however, are very general but not quite easy to work with. This piece is only concerned with discrete Markov chains with finite state space and luckily for us, in this case there is a very convenient way of dealing with the various problems. With introducing the matrix representations of Markov chains we could use the findings of linear algebra. For this we have to investigate the connection between matrices and the mathematical objects in question. First let us introduce a few linear algebraic definitions.

3.1 Connection with Matrices - Existence Theorem

3.1.1. Definition. *A matrix is called a stochastic matrix if the summation of values in each row equals to one and all entries are non-negative.*

3.1.2. Definition. *A matrix is called a sub-stochastic matrix if the summation of each row is not greater than one and all entries are non-negative.*

Let Π be a non-negative vector and its values summed be equal to one. We saw that the Markov Chain is entirely determined by the transition probabilities and the initial distribution. If we create a matrix P, where $p_{ij} = \mathrm{P}(X_{n+1} = j \mid X_n = i)$ and Π contains the initial probabilities we can easily see that every Markov chain determines a stochastic matrix. To prove that it also works the other way around we need the following theorem.

3.1.3. Theorem. *For all P stochastic matrix and Π vector with the above stated properties exists a Markov chain with initial distribution Π and with transition probabilities p_{ij}.*

Proof. For the proof we will use Kolmogorov's theorem which states the following. Let χ be a complete, separable metric space and β be the $\sigma-$algebra of Borel sets. Now let T be an index set and for all n and $i_1, i_2...i_n \in T$ define a measure $Q^n_{i_1,i_2,...,i_n}$ on (χ, β) with the following properties

1. $\forall B \in \beta^n$

$$Q^{(n+1)}_{i_1,i_2,...i_n,i_{n+1}}(B \times \chi) = Q^n_{i_1,i_2,...i_n)}(B) \tag{3.1}$$

2. $\forall \pi$ permutations and $\forall B \in \beta$

$$Q^n_{i_1,i_2,...i_n}(B) = Q^n_{i_{\pi(1)},i_{\pi(2)},...i_{\pi(n)}}(\pi(B)) \tag{3.2}$$

For all the sets of measure defined above exists a stochastic process $(X_i)_{i \in T}$, on some suitable probability space, for which the following holds.

$$\mathrm{P}((X_{i_1}, X_{i_2}, ...X_{i_n}) \in B) = Q^n_{i_1,i_2,...i_n}(B) \tag{3.3}$$

To prove our theorem let us define a set of measure from the above type the following way. Let χ be I discrete state space, T be the natural numbers and the measures are the following.

$$Q^n_{1,2,...n}(i_0, i_1, i_2, ...i_n) = \pi_{i_0} p_{i_1 i_2} p_{i_2 i_3} ... p_{i_{n-1} i_n} \tag{3.4}$$

Now we have to prove the defined properties still apply to this. The 2. property can easily be seen since $Q^n_{i_{\pi(1)},i_{\pi(2)},...i_{\pi(n)}}(\pi(i_1), \pi(i_2), ...\pi(i_n)) = p_{\pi(i_1)\pi(i_2)} p_{\pi(i_2)\pi(i_3)} ... p_{\pi(i_{n-1})\pi(i_n)} = p_{i_1 i_2} p_{i_2 i_3} ... p_{i_{n-1} i_n} = Q^n_{1,2,...n}(i_1, i_2, ...i_n)$. The 1. comes from the computation of the product measures $Q^{(n+1)}_{1,2,...n,n+1}((i_1, i_2, ...i_n, i_{n+1}) \times I) = Q^n_{1,2,...n}(i_1, i_2, ...i_n)$ and, since $Q(I) = 1$

because the matrix was stochastic this all equals to $Q^n_{1,2,...n}(i_1, i_2, ...i_n)$. Now, that we assigned a stochastic process for an arbitrary stochastic matrix, our next step is to prove it has the Markov property. For this we use the 2.1.6/2 form of the property.

$$\begin{aligned} \mathrm{P}(X_{n+1} &= i_{n+1} \mid X_1 = i_1, X_2 = i_2, ...X_n = i_n) = \frac{p_{i_1 i_2} p_{i_2 i_3} \cdots p_{i_{n-1} i_n} p_{i_n i_{n+1}}}{p_{i_1 i_2} p_{i_2 i_3} \cdots p_{i_{n-1} i_n}} \\ &= p_{i_n i_{n+1}} = \mathrm{P}(X_{n+1} = i_{n+1} \mid X_n = i_n) \end{aligned} \tag{3.5}$$

□

The matrix corresponding to a Markov Chain also called its transition matrix. Before proceeding here we have to note that only in the homogeneous case exists such matrix.

Using the matrix representations we can show that the irreducubility of a chain, which was defined in the previous chapter is really not different than the similar definition for matrices.

3.1.4. Definition. *An $A \in \mathbb{R}^{n \times n}$ matrix is said to be reducible if there exists a nonempty, disjoint set of indexes defined the following way. $i_1, i_2, ...i_\mu$ and $j_1, j_2, ...j_\gamma$ where $\mu + \gamma = n$ and*

$$a_{i_\alpha j_\beta} = 0 \quad \alpha = 1, 2, ...\mu, \ \beta = 1, 2, ...\gamma \tag{3.6}$$

Being a matrix reducible it is also easy to see that any of its powers will also be reducible, since for any $\alpha = 1, 2, ...\mu$ and $\beta = 1, 2, ...\gamma$ $a^{(n+1)}_{i_\alpha j_\beta} = a^{(n)}_{i_\alpha \cdot} a^{(n)}_{\cdot j_\beta}$. And, because of the first μ elements of $a^{(n)}_{i_\alpha \cdot}$ are non-zero while the non-zero elements of $a^{(n)}_{\cdot j_\beta}$ are the last γ, $a^{(n+1)}_{i_\alpha j_\beta} = 0$ for all $n \in \mathbb{N}$. What it means is that from the states corresponding to the first set of indices it is impossible to get to the sates corresponding to the other set. This way we can see that a Markov Chain or its transition matrix being irreducible means the same and the indices can be partitioned in the same way as the classes of the chain.

Another way of looking at Markov Chains is treating them as weighted directed graphs with their transition matrices as adjacency matrices.

3.2 A Few Computational Methods

With the proven theorem in mind let us take another look at the definitions and theorems introduced in the previous chapter. Here I am going to show a few methods for calculating the defined properties using the tools provided by the findings of linear algebra.

3.2.1 Linear Equations

Let us start with the inspection of the problem of recurrence and transience. The method I will introduce was proposed in the book of P. Billingsley in chapter 8.3[8]. Based on the previous proof we can easily see that a substochastic matrix can be the representation of a part of a Markov chain. Let Q be the matrix representation belonging to $U \in I$ states with q_{ij} transition probabilities.To calculate higher order transitions we can use

$$q_{ij}^{(n+1)} = \sum_{v} q_{iv} q_{vj}^{(n)} \tag{3.7}$$

Where $q_{ij}^{(n)}$ is the (i,j)th entry of $Q^{(n)}$ In the future we use the following notation for the sum of the ith row.

$$\rho_i^{(n)} = \sum_{j} q_{ij}^{(n)} \tag{3.8}$$

Using (3.7) we can introduce the following inductive way of computing for the row sums.

$$\rho_i^{(n+1)} = \sum_{j} q_{ij}^{(n+1)} = \sum_{j} \sum_{v} q_{iv} q_{vj}^{(n)} = \sum_{v} q_{iv} \rho_v^{(n)} \tag{3.9}$$

It can be also written in another way.

$$\rho_i^{(n+1)} = \sum_{j} \sum_{v} q_{iv} q_{vj}^{(n)} = \sum_{v} q_{iv}^{(n)} \rho_v^{(1)} \tag{3.10}$$

With Q being substochastic we know that $\rho_i^{(1)} \leq 1$ for all i keeping it in regard from (3.10) we can deduct that the series $\rho_i^{(n)}$ is monotone descending for all i. Since this series is non-negative the following limit exists.

$$\rho_i = \lim_{n\to\infty} \sum_j q_{ij}^{(n)} \tag{3.11}$$

By this we can deduct $\rho_i = \sum_j q_{ij}\rho_j$ in other words, the vector ρ solves the following linear equation.

$$x_i = \sum_j q_{ij}x_j, \quad i \in U \tag{3.12}$$

$$q \geq x_i \geq 1, \quad i \in U \tag{3.13}$$

For our further investigation now I will introduce a few lemmas for the next theorem which will give us a tool to decide whether a class is transient or not.

3.2.1. Lemma. *For a substochastic matrix Q the limits defined by (3.11) are the maximal solutions of the above described equation.*

Proof. Let x_i be an arbitrary solution for (3.12 and 3.13). In this case

$$x_i = \sum_j q_{ij}x_j \leq \sum_j q_{ij} = \rho_i^{(1)} \tag{3.14}$$

Now we proceed with an induction for all ns. If $\exists\ n$ for which $x_i \leq \rho_i^{(n)}$ then it implies the following.

$$x_i \leq \sum_j q_{ij}\rho_j^{(n)} = \rho_i^{(n+1)} \tag{3.15}$$

Thus $x_i \leq \rho_i^{(n)}$ for all n and so $x_i \leq \rho_i$. And, since $\rho_i^{(n)}$ is a monotone ascending series it gives us that ρ_i is truly the maximal solution. $\square$

As I mentioned earlier, the matrix Q belongs to a U subset of the state space S. With keeping this in mind we can see that the above defined $\rho_i^{(n)}$ can be written as $\sum_{j_1,j_2,\ldots j_n} p_{ij_1}p_{j_1j_2}\ldots p_{j_{n-1}j_n}$ where j ranges over all the possible combinations of the states in U. This leads us to the formulation of the following lemma.

3.2.2. Lemma. *For $U \subseteq I$ the probability of remaining forever in U given that we started in state $i \in U$ is the maximal solution of the following equation.*

$$x_i = \sum_{j \in U} p_{ij} x_j, \quad i \in U \tag{3.16}$$

$$0 \leq x_i \leq 1, \quad i \in U \tag{3.17}$$

Proof. Continuing the from before we formulated the lemma we can say that $\rho_i^{(n)}$ gives us the following probability.

$$\rho_i^{(n)} = \mathrm{P}(X_j \in U \; for \; j \leq n \mid X_0 = i) \tag{3.18}$$

With the limit introduced for $n \to \infty$ we can also formulate the probability for this case.

$$\rho_i = \mathrm{P}(X_j \in U, \; for \; j = 1, 2, ... \mid X_0 = i), \quad i \in U \tag{3.19}$$

Which is exactly the probability we are looking for and applying Lemma 3.2.1 we provided proof for this lemma. □

Suppose the chain is irreducible. Now consider this system with U being $I \setminus \{i_0\}$ where i_0 is an arbitrary state. Here we can write down a linear equation similar to the ones before corresponding to this particular case. Note that for such equation the $x_i = 0$ is always a solution. With the help of this and the lemmas we prove our theorem for recognizing transient chains.

3.2.3. Theorem. *The linear equation corresponding to the $U = I \setminus \{i_0\}$ case has a non-trivial solution if and only if the chain is transient.*

Proof. Consider the probabilities

$$1 - f^*_{ii_0} = \mathrm{P}(X_n \neq i_0, \; for \; n \geq 1 \mid X_1 = i), \quad i \neq i_0 \tag{3.20}$$

By Lemma 3.2.2 these are equal to the maximal solutions to the corresponding linear equation. We can see that it only has a non-trivial solution if $f_{ii_0} < 1$ which, according to theorem 2.2.10 cannot happen if the chain is recurrent.

Now let us start from the other way around and suppose the chain is transient.

$$f_{i_0 i_0} = \mathrm{P}(X_2 = i_0 \mid X_1 = i_0) + \sum_{n=2}^{\infty} \sum_{i \in U, i \neq i_0} p_{i_0 i} f_{i i_0}^{(n)} = \\ = q_{i_0 i_0} + \sum_{i \in U, i \neq i_0} p_{i_0 i} f_{i i_0}^{*} \tag{3.21}$$

So, because of the transience $f_{i_0 i_0}^{*} < 1$, also $f_{i i_0} < 1$ for some $i \neq i_0$. □

3.2.4. Remark. The reasoning proposed in Lemma 3.2.2. can be also applied the other way around. We can declare a corresponding linear equation for finding the probability of ever leaving a class U from $i \in U$.

$$z_i = \sum_{j \in U} p_{ij} z_j + \sum_{j \notin U} p_{ij} \quad i \in U \\ 0 \leq z_i \leq 1 \quad i \in U \tag{3.22}$$

Based on the reasoning of the above-cited lemma it is easy to see that

$$\mathrm{P}(\textit{the process leaving U up to the nth step} \mid X_1 = i) = 1 - \rho_i^{(n)} \tag{3.23}$$

and from this follows

$$\mathrm{P}(\textit{the process leaving U in any steps} \mid X_1 = i) = 1 - \rho_i. \tag{3.24}$$

Now first we can easily show $1 - \rho_i$ is a solution for the equation (2.21).

$$\sum_{j \in U} p_{ij}(1 - \rho_i) + \sum_{j \notin U} p_{ij} = \sum_{j \in U} p_{ij} - \sum_{j \in U} p_{ij} \rho_j + \sum_{j \notin U} p_{ij} = \\ -\rho_i + \sum_{j \in I} p_{ij} = 1 - \rho_i \tag{3.25}$$

And since lemma 3.2.1 stated that for the equation (3.12 and 3.13) ρ_i is a maximal solution so $1 - \rho_i$ will be the minimal solution for (3.22).

3.2.5. Remark. Another observation we shall make is that the second restrictions on the equations, the coordinates being between 0 and 1, in both cases can be omitted.

In the first case we have the equation (3.16) which is homogeneous with the homogenity degree of 1. If we look at it as a function and multiply all of its factors by a real number α the value of the function will also be multiplied by α.

$$if\ x_i = \sum_{j \in U} p_{ij} x_j\ then\ \sum_{j \in U} p_{ij\alpha x_j} = \alpha x_i \tag{3.26}$$

In this case the question is reduced to finding a non-trivial, non-negative and bounded solution for the first part. Then this solution can be rescaled in a way that will satisfie $0 \leq x_i \leq 1,\ i \in U$.

In the second case, explained in the previous note, the answer is quite obvious. For the equation (3.22) $z \equiv 1$ is always a solution so when we are looking for the minimal solution, the restriction on the possible solutions will not have any effect.

Now that we successfully reduced the difficult question whether an irreducible chain is transient to a simple question of resolvability of a linear equation, I will provide an example to show why this is a really convenient method.

3.2.6. Example. Let us look at an unrestricted random walk on the line of integers. Here each site has two neighbouring sites with the transition probabilities p and q. We have seen before that this walk has the Markov property and it is also easy to see that it is irreducible, since all states communicate. The equation, like the one proposed in lemma 2.2.1, corresponding to this chain is the following.

$$\begin{aligned} x_i = p x_{i+1} + q x_{i-1}, \quad i \in \mathbb{Z} \\ 0 \leq x_i \leq 1, \quad i \in \mathbb{Z} \end{aligned} \tag{3.27}$$

If $p = q$ and $p + q = 1$ (so $p = \frac{1}{2}$) we are talking about a symmetric random walk in one dimension. Using the theorem 3.2.3 I will show that only under this term can such walk be a recurrent chain. The restructured version of equation (3.27) belonging to this problem can be written with i_0 being the origin.

$$\begin{aligned} x_i &= px_{i+1} + qx_{i-1}, \quad i \neq 1, -1, 0 \\ x_1 &= px_2 \\ x_{-1} &= qx_{-2} \\ 0 \leq \; x_i &\leq 1, \quad i \neq 0 \end{aligned} \tag{3.28}$$

Using remark 3.2.5 the last restriction on x can be omitted and our only interest is whether it has a nontrivial, non-negative, bounded solution. Now let us separate the equation above into two parts. The first part corresponding to the cases where we are on the positive part of the line and the other to the negative part. With the index of the states in the second equation taken with the multiplication factor -1 here is the rephrased problem.

$$\begin{aligned} x_1 &= px_2 \\ x_i &= px_{i+1} + qx_{i-1} \;\; i \geq 2 \end{aligned} \tag{3.29}$$

$$\begin{aligned} x_1 &= qx_2 \\ x_i &= qx_{i+1} + px_{i-1} \quad -i \leq -2 \end{aligned} \tag{3.30}$$

Now we have to check under what terms have the two separate equations a desired solution, then with the two combined we will get the condition for the recurrence of the chain. Using the tool provided in the appendix we are looking for the solution of (3.29) in the form of $x_k = A + b(q/p)^k$ if $p \neq q$ and in the form of $x_k A + Bk$ if $p = q$. Because of the first part of the equation we can express A in terms of B and will get the following results.

$$x_k = \begin{cases} (\frac{q-q^2}{p^2-p} + (\frac{q}{p})^k)B & \text{if } \; p \neq q \\ 3Bk & \text{p=q} \end{cases} \tag{3.31}$$

Similarly the form of the solution of the equation (3.30) can be computed and the result can be seen below.

$$x_k = \begin{cases} (\frac{p-p^2}{q^2-q} + (\frac{p}{q})^k)B & \text{if } \; p \neq p \\ 3Bk & \text{p=q} \end{cases} \tag{3.32}$$

In the first case there exists a non-trivial bounded solution if $q < p$, otherwise there is no such solution. For the second case it works the other way around there is no bounded solution if $q \leq p$. Putting the two together we will get that only if $p = q$ has neither of the equation a desired solution. Because of theorem 3.2.3 it means that a one dimensional unrestricted random walk is recurrent if and only if it is symmetric.

3.2.7. Remark. The above example was proposed in P. Billingsley's book[8] and I came up with the solution on my own.

3.2.2 Stationary Distribution - Limit of Transition Matrix

Before proceeding to our next part, I will show you through a few important examples why using the matrix representations of Markov chains is a very convenient way of dealing with various problems. In claim 2.1.6 we saw that using only the framework provided by probability theory it is not that trivial to prove that Markov property between the n-th and the $(n+1)$-th time is the same as the Markov property for higher order transitions. But if we look at our chain as a transition matrix we can easily see the second order transitions are the following:

$$P_{ij}^{(2)} = \sum_{k=1}^{N} P_{ik} P_{kj} = P_{ij}^2 \tag{3.33}$$

$$so$$

$$P^{(2)} = P^2 \tag{3.34}$$

3.2.8. Remark. Here I use the capital P-s to distinct from the values p_{ij}^2, which would be the square of the value p_{ij}, from the (i, j)th element of the matrix P^2.

And inductively for higher order transitions

$$P_{ij}^{(m)} = \sum_{k=1}^{N} P_{ik}^{(m-1)} P_{kj} = P_i^m j \quad (3.35)$$

$$so$$

$$P^{(m)} = P^m \quad (3.36)$$

And it also proves the difficult part of the claim from the first chapter.

Since $P^n P^m = P^{n+m}$, using the rules for matrix powers, we come to the equation below

$$P_{ij}^{m+n} = \sum_{k=1}^{N} P_{ik}^m P_{kj}^n \quad (3.37)$$

which is known as the Chapman-Kolmogorov equation.

Now we get to one of the main purposes of such representation with rephrasing the definition 2.1.10 for stationary distributions.

3.2.9. Claim. *The definition 2.1.10 in the case of homogeneous Markov chains is equivalent with the following:*
A probability vector $\Pi = (\pi_1, \pi_2, ..., \pi_N)$ is a stationary distribution (or stationary vector) of a Markov chain with the corresponding transition matrix P if the following holds.

$$\pi_j = \sum_{i=1}^{N} \pi_i P_{ij}; \quad 1 \leq j \leq N \quad (3.38)$$

or for infinite state Markov chains, if the solution exists

$$\pi_j = \sum_{i=1}^{\infty} \pi_i P_{ij}; \quad 1 \leq j \quad (3.39)$$

Although in this case we just looking it as a linear equation which does not always have a desired solution.

3.2.10. Remark. It also means the Π satisfies $\Pi = \Pi P$ which means it is the left eigenvector of P corresponding to the eigenvalue 1.

The importance of such distributions becomes apparent when we wish to study the behavior of chains in the long run. The intuition suggests that homogeneous Markov property holds some kind of memoryless attributes since it states if we stop the process at some integer time then restart it it behaves like we started a completely new process and what happened before the stoppage is irrelevant. The best way of inspecting that what happens as the memory of the past dies out is to examine the powers of the transition matrix P^n as $n \to \infty$. We can see if the powers converge is suggest that the chain has reached a steady state where the states are independent from from both n and the preceding state. Which means if the convergence occurs the values $(P^n_{1,j}, P^n_{2,j}, ...P^n_{Nj})$ tends to the same value. Now if we look at the inductive formula of the transition probabilities, $P^{n+1}_{ij} = \sum_{k=1}^{N} P^n_{ik} P_{kj}$, we can see that the solution of the limit equation $(\pi_j = \sum_{k=1}^{N} \pi_k P_{kj})$ will be the stationary distribution.

3.2.11. Remark. Here we have to note that if we find the solution of $\Pi = \Pi P$ it does not imply that P^n converges.

After all the obvious question what comes to our minds is when the P^n converges and whether and where are there multiple stationary vectors. First, to make thing easier to explain, give a few attributes.

3.2.12. Definition. *A Markov chain is said to be ergodic if it is recurrent and aperiodic. A finite state Markov chain is called an unichain if it has only one recurrent class.*

3.2.13. Claim. *For a finite-state Markov chain there exists a single non trivial solution for the equation $\Pi = \Pi P$ if and only if it is an unichain. If it has multiple recurrent classes it will have a stationary vector for each.*

I will get back to the proof of this claim later when we are discussing the application of a few more advanced concepts of linear algebra.

The question of convergence and the rate of it is not as simple as whether there exists such vector. In the following I will go through the instances when the P^n sequence will converge to the desired state and I will show the requirements as well. To gain a better grasp on the question we will need a lemma which tells us something about the behaviour of the columns where we wish to get the steady state vectors.

3.2.14. Lemma. *Let P be the transition matrix of a finite-state Markov Chain. Then for each j state and each $n \geq 1$ integer*

$$\max_i P_{ij}^{n+1} \leq \max_k P_{kj}^n; \quad \min_i P_{ij}^{n+1} \geq \min_k P_{kj}^n \tag{3.40}$$

Proof. We apply the Chapman-Kolmogorov equation for each i, j and n and also the fact that $\forall k, i \in S$ $P_{ji}^n \leq \max_j P_{ji}^n$ and P is stochastic.

$$P_{ij}^{n+1} = \sum_{k=1}^{N} P_{ik} P_{kj}^n \leq \sum_{k=1}^{N} P_{ik} \max_l P_{lj}^n = \max_l P_{lj}^n \tag{3.41}$$

Similarly, we can also prove it for the minimums as well. □

Now first we prove the convergence for a quite specific matrix and after that we will generalize it for other cases.

3.2.15. Theorem. *Consider the transition matrix P where $P_{ij} > 0$ for all $i, j \in I$ and denote the smallest element with α. Then for all states j and alln ≥ 1 integers:*

$$\max_i P_{ij}^{n+1} - \min_i P_{ij}^{n+1} \leq (\max_k P_{kj}^n - \min_k P_{kj}^n)(1 - 2\alpha) \tag{3.42}$$

$$(\max_k P_{kj}^n - \min_k P_{kj}^n) \leq (1 - 2\alpha)^n \tag{3.43}$$

$$\lim_{n\to\infty} \max_k P_{kj}^n = \lim_{n\to\infty} \min_k P_{kj}^n > 0 \tag{3.44}$$

Proof. Introduce a notation for the l which minimizes $\min_l P_{lj}^n$, let it be l_j^n. We use again the Chapman-Kolmogorov equation the way we used it for the proof of the previous lemma.

$$P_{ij}^{n+1} = \sum_{k=1}^{N} P_{ik} P_{kj}^n \leq \tag{3.45}$$

$$\sum_{k \neq l_j^n} P_{ik} \max_l P_{lj}^n + P_{il_j^n} \min_l P_{lj}^n$$

Now for the next step add and subtract $P_{il_j^n} \max_l P_{ij}^n$ and continue the equation.

$$\max_l P_{lj}^n - P_{il_j^n}(\max_l P_{ij}^n - \min_l P_{lj}^n) \le \max_l P_{lj}^n - \alpha(\max_l P_{ij}^n - \min_l P_{lj}^n)$$

With the same arguments and the roles of max and min reversed we get

$$P_{ij}^{n+1} \ge \min_l P_{lj}^n + \alpha(\max_l P_{lj}^n - \min_l P_{lj}^n) \tag{3.46}$$

Subtracting the lower bound for P_{ij}^{n+1} from the upper bound we get (3.42). Since $\min_l P_{lj}^n \ge \alpha$ and $\max_l P_{lj}^n \le \alpha - 1$ we get $\max_l P_{lj}^n - \min_l P_{lj}^n \le 1 - 2\alpha$. If we raise it to the nth power from the binomial theorem we get $(\max_l P_{lj}^n - \min_l P_{lj}^n)^n - C$, where C is some positive real number, on the left side and $(1-2\alpha)^n$ on the right side which leads us to (3.43). Because $\alpha > 0$ the right side will converge to 0 exponentially which concludes our proof. □

And now we arrived to our most important case when we know exactly that the sequence of powers of the transition matrix converges and we even know at which rate it does.

3.2.16. Theorem. *The transition matrix of an ergodic Markov chain, which is a chain consists only a single ergodic class, converges exponentially.*

Proof. For the proof we will use the previous theorem but first we have to find a way to alter the matrix so that it will satisfy the prequisite of the theorem. When we are checking convergence we can chop down a finite part from the start of the series and prove the convergence for only the remaining part. To apply the previous theorem I will show you that for an N state chain for $n \ge (N-1)^2 + 1$ the transition matrix P^n is strictly positive.

Look at our chain as a directed graph with the adjacency matrix P. It is easy to see the it must contain a directed circle of length $\mu < N$. μ can not be equal to N because in this case the period of the chain would be N so it would not be ergodic on the other hand it must contain a circle because it is a finite recurrent chain so $i \to j \Rightarrow j \to i$. Fix a state l which is a part of a μ state circle and denote the set of states reachable from l

in n steps with $\Gamma(n)$. It is obvious that $\Gamma(n) \subseteq \Gamma(n+\mu)$ since if we can reach j from l in n steps the we can still reach it if we go in circle back to l the go to j. Se we can say,if $\Gamma(0) = \{l\}$,

$$\Gamma(0) \subseteq \Gamma(\mu) \subseteq \Gamma(2\mu) \subseteq ... \tag{3.47}$$

If at some point the inclusion satisfies equality the from then on all inclusions satisfy it. So at maximum only the first $N-1$ inclusions can be satisfied strictly and since the chain consists of a single class $\Gamma((N-1)\mu)$ should contain all of the states (if there was a state i which was not contained in this set it would be unreachable from l so it would not be in the same class). And because of it from any state we can reach any other state in at most $(N-1)^2+1$ steps which gives us $P_{ij}^{(N-1)^2+1} > 0$.

Now we apply theorem 3.2.15 with $P' P^{(N-1)^2+1} =$ and define π_j as

$$\pi_j = \lim_{n\to\infty} \max_k P'^n_{kj} = \lim_{n\to\infty} \min_k P'^n_{kj} > 0 \tag{3.48}$$

Which proves that P^n converges to a matrix with identical rows since the sequence is non-increasing and in this case for the proof of the convergence finding a convergent subsequence is enough. □

We can generalize this case for Markov chains with multiple classes but just one recurrent class (this is what was called a unichain). Denote the set of recurrent states with R and the set of transient states with T. In this case we can organize the elements of the transition matrix so that the recurrent and transient states are forming distinct blocks in the matrix while the transition between these blocks is also forming one. Denote the recurrent block with P_R. the transient blocks with P_T and the transitions between the two with P_{TR} then the matrix has the following form.

$$\begin{pmatrix} P_T & P_{TR} \\ 0 & P_R \end{pmatrix} \tag{3.49}$$

Here the transition from the recurrent block to the transient has all zero elements since if a state can be reached from a recurrent state then it also works the other way around.

From the previous theorem we know that the R_T block of a matrix converges, the question is what happens with the processes starting in a transient state. It is easy to see that after a finite number of steps we have to arrive to a recurrent state form where we will never reach a transient state again. Because of this, being in a transient state after n steps converges to 0 as $n \to \infty$. It means that all the elements of $\lim_{n\to\infty}$ equal 0. Now our only concern is the transitions from a transient state to a recurrent. Let us say j is a recurrent state with the steady-state probability π_j and i is a transient state and let $m = \lfloor n/2 \rfloor$. With the Chapman-Kolmogorov equation we can say

$$\begin{aligned}
\mid P_{ij} - \pi_j \mid &= \mid \sum_{k\in T} P_{ik}^m (P_{kj}^{n-m} - \pi_j) + \sum_{k\in R} P_{ik}^m (P_{kj}^{n-m} - \pi_j) \leq \mid \qquad (3.50) \\
&\leq \mid \sum_{k\in T} P_{ik}^m \mid P_{kj}^{n-m} - \pi_j \mid + \sum_{k\in R} P_{ik}^m \mid P_{kj}^{n-m} - \pi_j \mid \leq \\
&\leq \sum_{k\in T} P_{ik}^m + \sum_{k\in R} P_{ik}^m \mid P_{kj}^{n-m} - \pi_j \mid \qquad (3.51)
\end{aligned}$$

In (3.51) it is easy to see that the two parts converge to 0. The first part is the probability of remaining in a transient state starting from i after m steps. In the second part P_{ij}^{n-m} converges to π_j since the starting state is in the recurrent class as well and $n-m \geq \lfloor n/2 \rfloor$ so it tends to infinity.

The asymptotic behaviour of P^n as $n \to \infty$ for an arbitrary finite state Markov chain in most cases can be deduced from the previous case.

I have noted before that the steady state vector must be an eigenvector belonging to the eigenvalue 1. Because the transition matrix is stochastic this eigenvalue has to be the largest in magnitude. The question is whether such eigen value exists and if it does, does it yield a left eigenvector suitable to be a distribution and how can we find it. If we have found an eigenvector which is non-negative (and non-zero) we can easily scale it down so it would be a stochastic vector and it would still be an eigenvector. So the question is reduced to whether the vector in question is non-negative. To solve these problem I will use a theorem from linear algebra which I will not prove here.

3.2.17. Theorem. *(Perron-Frobenius) Let A be an irreducible, non-negative matrix with* *the spectral radius $\rho(A) = r$. In this case the following are true*

1. *r is a positive real number and it is an eigenvalue of the matrix A (called the Perron-Frobenious eigenvalue)*

2. *The Perron–Frobenius eigenvalue r is simple. Both right and left eigenspaces associated with r are one-dimensional.*

3. *A has both a left and a right eigenvector corresponding to r whose components are all positive.*

As we have seen earlier a chain with one class has an irreducible transition matrix. In the multiple class case we can solve the problem for the different classes separately and just leave the remaining coordinates as zero. The other important factor which needs to be concerned to see why this theorem helps us is to prove that $\rho(P) = 1$ where P is the transition matrix. Because P is stochastic its eigenvalues are less or equal to 1 and also the vector $x = (1, 1, ...1)$ satisfies

$$x = xP$$

This way we found a non-negative, unique solution for the equation and with scaling we can make it a distribution vector thus proving claim 3.2.13.

3.3 An Example from Physics

Finding the stationary distribution of a Markov process is often a crucial part of simulating different real world phenomena. With the tools introduced in this chapter the task becomes fairly easy as I am going to demonstrate it with the following example borrowed from physics. (it was proposed in W. Feller's book [9]) The model was introduced by Bernoulli and Laplace and used to simulate the diffusion of two types of in-compressible liquids between two containers. [1]The model is the following, imagine the particles of the two liquids as small balls of different colour. Suppose we start with $2m$ balls, m is black and m is white. In each container there are equal number of balls and at each independent draw one is drawn from each and placed to the other. This way the density of the particles remain the same representing the in-compressible nature of the liquid. Now define the random variables $X_n(n = 0, 1, ...)$ as the number of black balls in the first container after the nth trial this way the state I_k denotes that there are k black balls int the first container. It is easy to see that the stochastic process $(X_n)_{n\in\mathbb{N}}$ has the Markov property since $\mathrm{P}(X_{n+1} = k)$ depend only on the number of black balls in the first container before the draw. The next step is determining the transition probabilities and the easiest way to do it is to list the possible outcomes of a draw and a change.

1. The first is when $X_{n+1} = X_n + 1$ here we are looking for the transition probability p_{kk+1}. To get this outcome first we have to draw a white ball from the first container, with probability $\frac{m-k}{m}$ and after that we have to draw a black one from the other with the same probability. So we got $p_{kk+1} = (\frac{m-k}{m})^2$.

2. The second case is $X_{n+1} = X_n - 1$ with the transition probability p_{kk-1}. Here we have to draw a black ball from the first with probability $\frac{k}{m}$ and a white one from the second with also the probability $\frac{k}{m}$. So $p_{kk-1} = (\frac{k}{m})^2$.

3. The last case is when the ration of the black and white balls remains the same with the transition probability p_{kk}. Here we have to chose balls with the same color from each container with the probability $2\frac{(m-k)}{m^2}$.

Here we must omit the cases when $k = 0$ or $k = m$ here we have the transition probabilities $p_{01} = 1$ and $p_{mm-1} = 1$. For all the other transitions not mentioned here the probability is 0.

[1]In its earliest form as an urn model it was first described by Daniel Bernoulli in 1769 and later in 1812 it was analyzed and reconstructed by Laplace.

3.3.1. Example. As an example here is the transition matrix for $m = 5$.

$$\begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 0.04 & 0.32 & 0.64 & 0 & 0 & 0 \\ 0 & 0.16 & 0.48 & 0.36 & 0 & 0 \\ 0 & 0 & 0.36 & 0.48 & 0.16 & 0 \\ 0 & 0 & 0 & 0.64 & 0.32 & 0.04 \\ 0 & 0 & 0 & 0 & 1 & 0 \end{bmatrix}$$

In the example provided we will see, which we would expect by intuition, that the liquid tends towards an even distribution of different particles between the two containers. As I have described earlier the stationary distribution of a Markov chain is a distribution of states which does not change by time. It is easy to see that it will give us the distribution of particles in the steady state of the system.

Before we move to presenting the results of the simulations let us make a few preceding calculations. Given we know the transition matrix of a chain it is still not always easy to calculate the stationary distribution (or even decide whether it exists) even after we have seen a few methods for it. However there are a few notable exceptions and one is when $\forall i, j \in S :| i - j |> 1 \Rightarrow p_{ij} = 0$. With the Bernoulli-Laplace model this is exactly the case. We start from the previously proven fact that we can compute the stationary distribution from

$$\pi = \pi P \tag{3.52}$$

If we omit the zero elements from the different components of the equation we get the following linear equation.

$$\pi_1 = p_{11}\pi_1 + p_{21}\pi_2 \tag{3.53}$$

$$\pi_2 = p_{12}\pi_1 + p_{22}\pi_2 + p_{32}\pi_3 \tag{3.54}$$

$$\textit{and so on...} \tag{3.55}$$

These equations can be solved successively with π_1 as a parameter which should be chosen so that $\sum_{i=1}^{m+1} \pi_i = 1$. Here we get

$$\pi_2 = \frac{p_{12}}{p_{21}}\pi_1, \quad \pi_3 = \frac{p_{12}p_{23}}{p_{21}p_{32}}\pi_1 \quad \pi_4 = \frac{p_{12}p_{23}p_{34}}{p_{21}p_{32}p_{43}}\pi_1 \tag{3.56}$$

Using the formulas for the transition probabilities deducted a few pages before (namely $p_{kk+1} = (\frac{m-k}{m})^2$ and $p_k k-1 = (\frac{k}{m})^2$) we can compute the general stationary distribution for these particular types of chains used in the Bernoulli-Laplace model. From the recursive formula from (2.37) we know that $\pi_k = \frac{\prod_{i=1}^{k-1} p_{ii+1}}{\prod_{i=1}^{k-1} p_{i+1i}}$. This general formula look like the following in this particular case.

$$\pi_k = \frac{\prod_{i=1}^{k}(\frac{m-i}{m})^2}{\prod_{i=1}^{k}(\frac{i}{m})^2}\pi_1 = \frac{(\frac{m!}{(m-k)!})^2\frac{1}{m^2}}{m!^2\frac{1}{m^2}}\pi_1 = \binom{m}{k}^2\pi_1 \tag{3.57}$$

Now all we have to do is to choose the parameter π_1 in a way that the π vector defined by the recursion will be a distribution. We know that the the square of the binomial coefficients $\binom{m}{k}$ add to $\binom{2m}{m}$ and hence the stationary distribution will be represented as

$$\pi_k = \frac{\binom{m}{k}^2}{\binom{2m}{m}} \tag{3.58}$$

It comes as no surprise that the stationary distribution is the hypergeometric. It means that at the equilibrium the distribution of two types of particles between the two containers is the same as the white and the black balls had been chosen at random with the same probability from a collection where the number of each colours is the same, denoted here as m.

I have written a matlab code to simulate the Bernuolli-Laplace model for $m = 50$. Here I started from the initial distribution being $I_{(i=1)}$ which gives me a starting scenario where there are two different types of liquids in the two containers. The results are shown in the figures below.

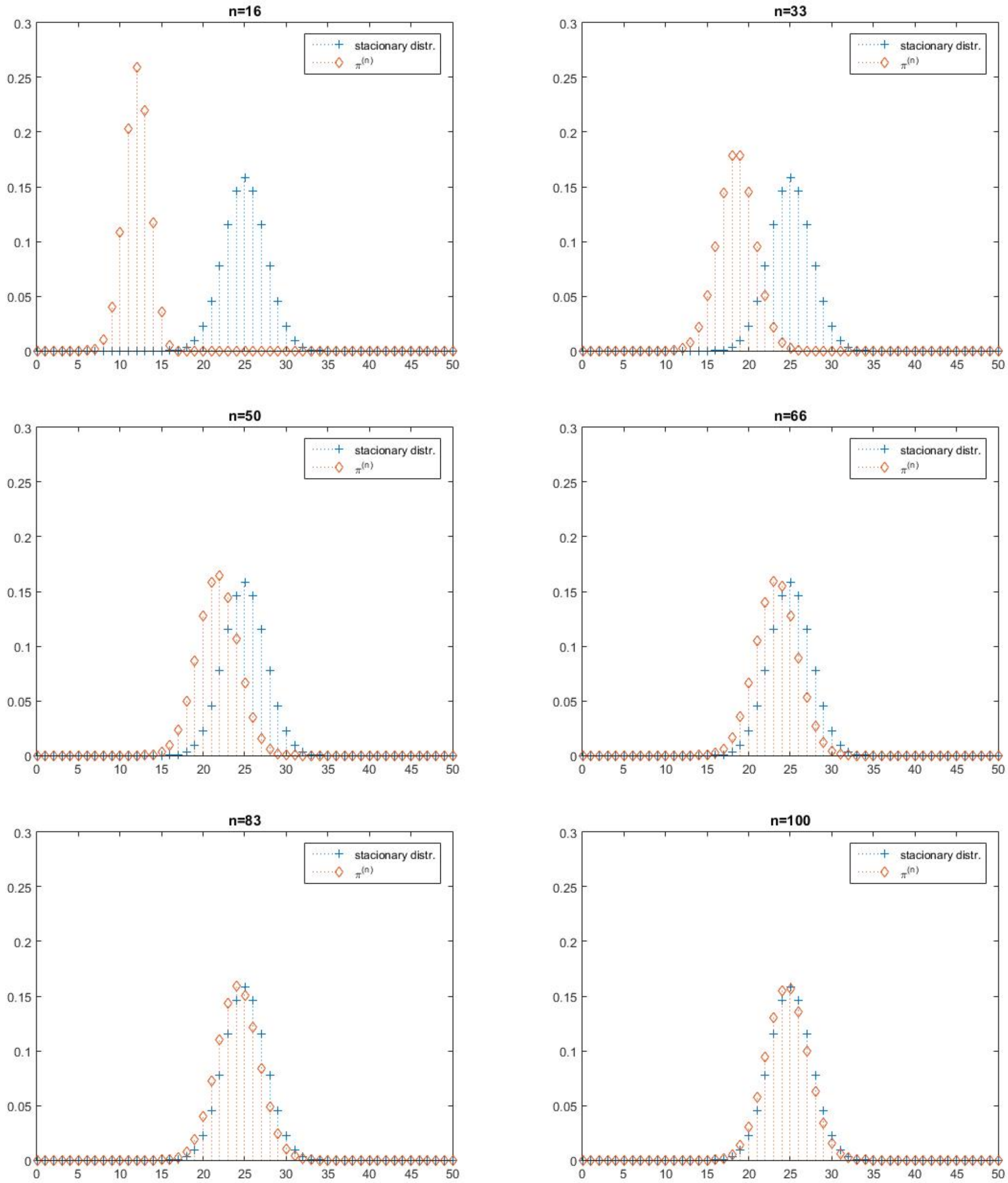

Figure 3.1: The number of iterations (n) passed before displaying the current distributions of states are at the top of each figure. I also highlighted the stationary distribution calculated as the left eigen vector corresponding to the greatest eigen value (divided by the sum of element to get a proper distribution).

4. Summary

In the present thesis my main aim was to provide a throughout introduction to Markov chains with some examples to the main methods of calculation. Although the instances provided in the second section are merely a quick foresight to the various models and interesting problems related to this topic but they provide a foundation for further work. One of the topics, that gives perspective to further investigations are the various MCMC based models. The underlying idea behind these models is that if one wishes to sample randomly from a specific probability distribution he needs Markov chain whose long-time equilibrium is the distribution in question. For this purpose the iteration method discussed in my thesis is a good tool. The other instance where Markov chains are used is in queueing theory where the chains are used with rewards. Here the the important factors are first passage times and hitting times which are related to the recurrent and transient properties.

5. Appendix

1. Suppose that a and b are integers and $a < b$. Now let x_n be defined for every n $a \leq n \leq b$ and let it satisfies the following

$$x_n = px_{n+1} + qx_{n-1} \;\; for \; a < n < b, \tag{5.1}$$

Here p and q are positive real numbers and $p + q = 1$.In this case the general form of the differential equation is the one below.

$$x_k = \begin{cases} A + B(\frac{q}{p})^n & for \; a < n < b, \text{ if } \; p \neq q \\ A + Bn & for \; a < n < b, \text{ p=q} \end{cases} \tag{5.2}$$

Where A and B are positive real numbers.

2.
```
function [ A ] = bldiff( p,n,a )

A=zeros(p+1);%p number of particles
P=zeros(1,p+1);
 if exist('a','var')
      P=a;
 else
      for i=1:p+1
           P(i)=1/(p+1);
      end
```

```
end

for i=1:p-1
    A(i+1,i+1)=2*i*(p-i)/p^2;
    A(i+1,i)=(i/p)^2;
    A(i+1,i+2)=((p-i)/p)^2;
end
A(1,2)=1;
A(p+1,p)=1;
[V,D,W]=eig(A);
[m,i]=max(diag(D));
X=0:p;
Y=transpose(W(:,i)./sum(W(:,i)))*A;
v=P*A;
v=v./sum(v);
 figure
    stem(X,Y,':+');
    axis([0 p 0 0.3]);
    hold on
    stem(X,v,':diamond');
for i=1:6
    v=P*A^(floor(i*n/6));
    v=v./sum(v);
    txt=texlabel('pi^((n))');
    figure
    stem(X,Y,':+');
    axis([0 p 0 0.3]);
```

```
        hold on
        stem(X,v,':diamond');
        title(['n=',num2str(floor(i*n/6))]);
        legend('stacionary distr.',txt);
end
```

6. Acknowledgements

I would like to express my sincere gratidue to my consultant Dr, Villő Csiszár for her support and the helpful discussons which guided me through the writing of this thesis.

Bibliography

[1] A. A. Markov. Untersuchung eines merkwürdigen Falles der abhängigen Versuche. *Leningrad, Bull. Ac. Sc. (6)*, 1:61–80, 1907.

[2] Nicholas Metropolis, Arianna W. Rosenbluth, Marshall N. Rosenbluth, Augusta H. Teller, and Edward Teller. Equation of state calculations by fast computing machines. *The Journal of Chemical Physics*, 21(6):1087–1092, 1953.

[3] Alan E. Gelfand and Adrian F. M. Smith. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, 85(410):398–409, 1990.

[4] R. A. Howard. Dynamic programming and markov processes. 1960.

[5] V. Csiszár. Diszkrét és folytonos paraméterű markov láncok. Available at `http://csvillo.web.elte.hu/ml/ML.pdf`.

[6] Robert G. Gallager. *Stochastic Processes: Theory for Applications.* Cambridge University Press, 2014.

[7] James R. Norris. *Markov chains.* Cambridge series in statistical and probabilistic mathematics. Cambridge University Press, 1998.

[8] Patrick Billingsley. *Probability and Measure.* John Wiley and Sons, second edition, 1986.

[9] W. Feller. *An Introduction to Probability Theory, Vol. 1.* Wiley, New York, NY, third edition, 1968.