Simulation of two ergodic Markov chains

Mikael Nyman, 33409

Thesis Pro Gradu

Faculty of Science and Engineering
Mathematics
Åbo Akademi

2021

Contents

Chapter 1

Introduction

In 2017 I participated in the course "Markov chains" at Åbo Akademi University, Finland, and found the material, as well as the applications, interesting. Therefore I knew that this was a subject that I wanted to write my master's thesis about. According to Lateef (2019), Markov chains are applied in many areas to solve real-world problems, such as Google page ranking, predicting typing of words, and generating texts. Another area is weather predictions, which we will see in Chapter 3 with empirical observations from the Finnish meteorological institute (FMI).

The purpose of this thesis is to present the theory of discrete Markov chains and simulate two different ergodic and irreducible Markov chains. Moreover, we will see how these chains behave in the long run. Hence, the main core of this thesis is the stationary distribution. The reader is supposed to have basic knowledge of mathematics and programming to understand the content.

The name "Markov" is from the russian mathematician Andrei Andreevich Markov, who lived between the years 1856-1922. According to [5, Chapter 2], Markov systematically studied a certain property. Informally the property states that the probability of being at a certain state at a future time point, given any set of time points up to present time, only depends on the last of the time points up to present time. This property is called the Markov property. The conditional probabilities, which correspond to transitions from one state to another, can be connected as a chain. Hence, the whole designation "Markov chain". Also, the conditional probabilites can be represented in a matrix, which we call a transition matrix. A transition matrix has different properties depending on its entries. One example of Markov chains, with

certain properties, is irreducible and ergodic Markov chains. These are the type of Markov chains I have chosen to simulate in Chapter 4.

Chapter 2 covers some introductory theory that mainly is used in Chapter 4, where we go through the theory of discrete Markov chains, provided with examples. The open source software Geogebra is used to plot and visualize some of the theory, and Microsoft Excel to read in and manipulate the empirical data provided by FMI. The empirical data, which is used in examples, originate from measurements made in Kilpisjärvi (in the north of Finland) during 30 years. After manipulation, the empirical data is transformed into a transition matrix. Most of the theory for Markov chains is obtained from [5].

In Chapter 4 we simulate two ergodic and irreducible Markov chains, where the first chain consists of the empirical data from FMI. The second chain we simulate is larger, in terms of dimension, consisting of 7 rows and 7 columns. Chapter 4 starts with a presentation of a coding algorithm to simulate Markov chains, where the algorithm is mainly based on an internet publication from Bonakdarpour M. (2016). All code for tables and charts is added to the appendices at the end of the thesis.

I have put much of effort in the visualizations, since I believe that they are vital. Vital especially for showing content to the reader for the purpose of understanding. One aspect of that can be seen in the simulations in Chapter 4, where I found a way to visualize convergence for Markov chains for smaller and larger numbers of transitions.

Chapter 2

Introductory theory

In this section, we review some of the basic theory of probability needed in this thesis. For further details, see [7] and [9].

2.1 Distribution functions and probability measures

Definition 2.1. The function $F : \mathbb{R} \to [0,1]$ is said to be a probability distribution function if

i) F is non-decreasing and right-continuous,

ii) $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to+\infty} F(x) = 1$

Example 2.2. Let for $a < b$

$$F(x) := \begin{cases} 0, & x \le a \\ \frac{x-a}{b-a}, & a \le x \le b \\ 1, & x \ge b \end{cases}.$$

Then F is the probability distribution function of a uniform probability distribution on $[0,1]$.

Definition 2.3. $(\Omega, \mathcal{F}, \mathbb{P})$ is said to be a probability space if it consists of

the following three parts:

$$\begin{cases} 1) & \text{The sample space } \Omega \text{ consisting all possible outcomes.} \\ 2) & \text{The } \sigma\text{-algebra } \mathcal{F}, \text{ where } \mathcal{F} \text{ is a set of subsets } E \subset \Omega. \\ 3) & \text{The probability measure } \mathbb{P} \text{ assigning probabilities to events.} \end{cases}$$

Remark 2.4. The subsets $E \in \mathcal{F}$ are called events.

Definition 2.5. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and $B \in \mathcal{F}$, where $\mathcal{F}$ is a σ-algebra, be such that $\mathbb{P}(B) > 0$. The conditional probability for event A with respect to event B is defined as

$$\mathbb{P}(A|B) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \tag{2.1}$$

Theorem 2.6. For $n \in \mathbb{N} = \{1, 2, ...\}$, let $\{A_i\}_{i=1}^n$ be events such that

$$\begin{cases} A_i \cap A_j = \varnothing, i \neq j \\ \cup_{i=1}^n A_i = \Omega. \end{cases}$$

Furthermore let $\mathbb{P}(A_i) > 0$ for $i = 1, 2, ..., n$. For $B \in \mathcal{F}$ it holds that

$$\mathbb{P}(B) = \sum_{i=1}^{n} \mathbb{P}(A_i)\, \mathbb{P}(B|A_i). \tag{2.2}$$

Theorem 2.7. Bayes Formula. Let $\{A_i\}_{i=1}^n$ be events as in Theorem 2.6 and $B \in \mathcal{F}$ be such that $\mathbb{P}(B) > 0$. Then for $k = 1, ..., n$, it holds that

$$\mathbb{P}(A_k|B) = \frac{\mathbb{P}(A_k)\, \mathbb{P}(B|A_k)}{\sum_{i=1}^{n} \mathbb{P}(A_i)\, \mathbb{P}(B|A_i)} \tag{2.3}$$

Definition 2.8. A function $X : \Omega \to \mathbb{N}$ is a geometrically distributed random variable, with parameter $p \in (0, 1)$ if

$$\mathbb{P}(X = k) := \mathbb{P}(\{\omega : X(\omega) = k\}) = p^k(1 - p) =: p^k q, \forall k \geq 0. \tag{2.4}$$

Definition 2.9. The expectation value of a geometrically distributed random variable (with parameter p) is

$$\mathbb{E}[X] = \sum_{k=0}^{\infty} k\, \mathbb{P}(X = k) = \sum_{k=0}^{\infty} k p^k q. \tag{2.5}$$

Theorem 2.10. Let X be a discrete random variable with the geometric distribution with paramater p. Then

$$\mathbb{E}[X] = \frac{p}{1-p} \tag{2.6}$$

Proof. From the definition of expectation and by the definition of the geometric distribution we get that

$$\begin{aligned} \mathbb{E}[X] &= (1-p)\sum_{k=0}^{n} kp^k = q\sum_{k=0}^{n} kp^k \\ &= qp\sum_{k=1}^{n} kp^{k-1} = qp\frac{1}{(1-p)^2} = \frac{p}{1-p}, \end{aligned} \tag{2.7}$$

where we used the well known identity: For $0 < p < 1$

$$\sum_{k=1}^{\infty} kp^{k-1} = \frac{1}{(1-p)^2}.$$

□

Theorem 2.11. (Markov's Inequality) Let $X : \Omega \to \mathbb{N}$ be a function, $\mathbb{P}(X \geq n) = \mathbb{P}(\omega : X(\omega) \geq n)$, and $\mathbb{E}[X] = \sum_{n=0}^{\infty} n\,\mathbb{P}(\omega : X(\omega) = n)$, then

$$\mathbb{P}\left(X \geq n\right) \leq \frac{\mathbb{E}[X]}{n}. \tag{2.8}$$

Proof. By (2.5) we get that

$$\begin{aligned} \mathbb{E}[X] &= \sum_{x=0}^{\infty} x\,\mathbb{P}(X = x) \\ &= \sum_{x=0}^{n} x\,\mathbb{P}(X = x) + \sum_{x=n}^{\infty} x\,\mathbb{P}(X = x) \\ &\geq \sum_{x=n}^{\infty} x\,\mathbb{P}(X = x). \end{aligned} \tag{2.9}$$

Since $x \geq n$, it follows that

$$\begin{aligned} \sum_{x=n}^{\infty} x\,\mathbb{P}(X=x) &\geq \sum_{x=n}^{\infty} n\,\mathbb{P}(X=x) \\ &= n\sum_{x=n}^{\infty} \mathbb{P}(X=x) \\ &= n\,\mathbb{P}(X \geq n). \end{aligned} \tag{2.10}$$

□

Theorem 2.12. Let $\{X_i\}_{i=1}^{n}$ be a sequence of independent and identical random variables, and let $\mathbb{E}[X_i] = \mu$. Then, as $n \to \infty$, we have that

$$\frac{X_1 + X_2 + ... + X_n}{n} \to \mu \tag{2.11}$$

with probability one, or equivalently

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{X_1 + X_2 + ... + X_n}{n} = \mu\right) = 1. \tag{2.12}$$

Equation (2.11) is called the strong law of large numbers. For a simplified proof, see [8]. The somewhat simplified proof assumes that $\mathbb{E}[X_i^2] < \infty$ and $\mathbb{E}[X_i^4] < \infty$, in comparison to the formal proof, which only requires that $\mathbb{E}[X_i] < \infty$.

Chapter 3

Discrete time Markov chains

3.1 Snowfall at Kilpijärvi, Finland

The Finnish Meteorological Institute (FMI) provides a large amount of open data for the public use. The amount of precipitation throughout Finland is one measurement that is constantly monitored and has been monitored for many years back. One weather station that monitors the amount of precipitation is located in the far north of Finland, namely in Kilpisjärvi. We will use observed data in January between the time period 01.01.1989-31.12.2018 to explain some of the theory of Markov chains. In the first simple model called the Bernoulli model we only look at the pattern of snowy and dry days. A day is defined as "snowy" if the precipitation amount is larger than $0,1$ millimeters (mm) and "dry" otherwise. By using the data we get 626 snowy and 304 dry days in total. Let $X_{i,j} = 1(A)$, where A=(day i of year j snowy). $1(A) = 1$ if the event A occurs and 0 otherwise.

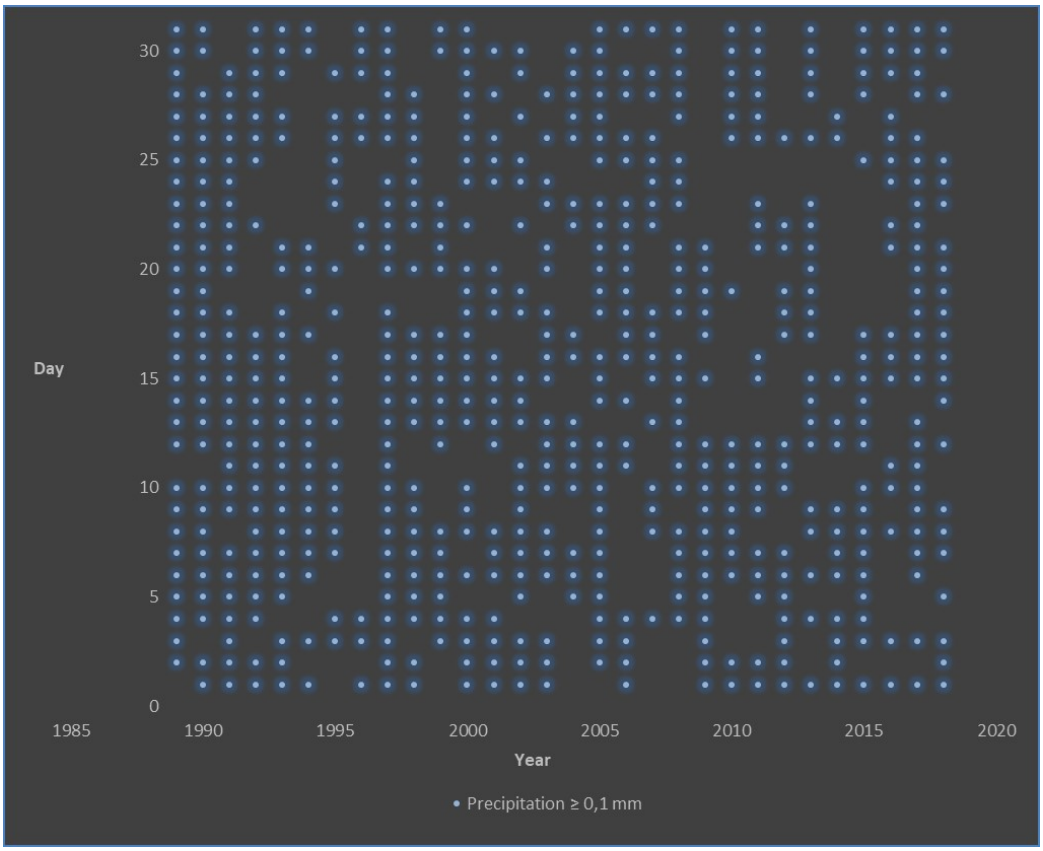

Figure 3.1: The pattern of January precipitation at Kilpisjärvi, Finland, between the time period 01.01.1989-31.12.2018. (30 years in total) Rows correspond to days, columns to years, and the blue dots to matching data whenever we have precipitation $\geq 0,1$ mm.

	Today dry	Today wet	Total
Yesterday dry	82	87	169
Yesterday wet	88	313	401
Total	170	400	570

Table 3.1: Observed precipitation data in January at Kilpisjärvi, Finland, between the time period 01.01.1989-31.12.2018.

Definition 3.1. A random process $\{X_0, X_1, ...\}$ with finite and discrete state space $S = \{0, 1, ..., N\}$, for simpilicity, composes a Markov chain (abbreviation M.C.) if for all $n \in \mathbb{N}$ and all $i_0, i_1, ..., j \in S$ it holds

$$\mathbb{P}(X_n = j | X_0 = i_0, X_1 = i_1, \dots, X_{n-1} = i_{n-1}) = \mathbb{P}(X_n = j | X_{n-1} = i_{n-1}). \tag{3.1}$$

A M.C. is said to be time homogenous if the probability in (3.1) does not depend on n, i.e. the transition probability from one state to another stays the same no matter the time point we observe the chain. (3.1) is called the Markov property.

Definition 3.2. An $(N+1) \times (N+1)$ matrix P with elements $P = (p_{i,j})_{i,j=0}^{N}$, is called a transition matrix of the time homogenous M.C. $\{X_n\}_{n=0}^{\infty}$.

The matrix representation of the elements $(p_{i,j})_{i,j=0}^{N}$ is

$$P = \begin{pmatrix} p_{0,0} & p_{0,1} & \cdots & p_{0,N} \\ p_{1,0} & p_{1,1} & \cdots & p_{1,N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N,0} & p_{N,1} & \cdots & p_{N,N} \end{pmatrix} \tag{3.2}$$

where all rows in P add up to 1, such that

$$\begin{aligned} p_{0,0} + p_{0,1} + \cdots + p_{0,N} &= p_{1,0} + p_{1,1} + \cdots + p_{1,N} \\ &\vdots \\ &= p_{N,0} + p_{N,1} + \cdots + p_{N,N} = 1. \end{aligned}$$

3.2 The marginal distribution

Marginal probabilities play an important role in Markov chains and are not very hard to compute. If we consider a $0-1$ Markov chain, i.e. $S = \{0, 1\}$,

we have that

$$\begin{aligned}\mathbb{P}(X_{n+1}=1) &= \mathbb{P}(X_{n+1}=1, X_n=0) + \mathbb{P}(X_{n+1}=1, X_n=1)\\ &= \mathbb{P}(X_n=0)p_{0,1} + \mathbb{P}(X_n=1)p_{1,1}\\ &= (1-P(X_n=1))p_{0,1} + \mathbb{P}(X_n=1)p_{1,1}\\ &= p_{0,1} - \mathbb{P}(X_n=1)p_{0,1} + \mathbb{P}(X_n=1)p_{1,1}\\ &= \mathbb{P}(X_n=1)(p_{1,1}-p_{0,1}) + p_{0,1}\end{aligned} \tag{3.3}$$

Let $p_1 := \mathbb{P}(X_0=1)$. Then $p_0 = \mathbb{P}(X_0=0) = 1-p_1$, and from (3.3) we get a general recursive scheme that can be calculated for each n:

$$\begin{aligned}\mathbb{P}(X_1=1) &= p_1(p_{1,1}-p_{0,1}) + p_{0,1}\\ \mathbb{P}(X_2=1) &= \mathbb{P}(X_1=1)(p_{1,1}-p_{0,1}) + p_{0,1}\\ &= (p_1(p_{1,1}-p_{0,1}) + p_{0,1})(p_{1,1}-p_{0,1}) + p_{0,1}\\ &= p_1(p_{1,1}-p_{0,1})^2 + p_{0,1}(p_{1,1}-p_{0,1}) + p_{0,1}\\ &= p_1(p_{1,1}-p_{0,1})^2 + p_{0,1}(1+(p_{1,1}-p_{0,1}))\\ \mathbb{P}(X_3=1) &= p_1(p_{1,1}-p_{0,1})^3 + p_{0,1}(1+(p_{1,1}-p_{0,1})+(p_{1,1}-p_{0,1})^2)\\ &= p_1(p_{1,1}-p_{0,1})^3 + p_{0,1}\sum_{r=0}^{2}(p_{1,1}-p_{0,1})^r\\ &\vdots\\ \mathbb{P}(X_n=1) &= p_1(p_{1,1}-p_{0,1})^n + p_{0,1}\sum_{r=0}^{n-1}(p_{1,1}-p_{0,1})^r\end{aligned} \tag{3.4}$$

In case $p_{0,0} = p_{1,1} = 1$ in equation (3.4) we see that $\mathbb{P}(X_n=1) = p_1$, since $p_{0,1} = 0$. If $p_{0,1} \neq p_{1,1}$ we can write

$$\begin{aligned}p_{0,1}\sum_{r=0}^{n-1}(p_{1,1}-p_{0,1})^r &= \frac{p_{0,1}(1-(p_{1,1}-p_{0,1})^n)}{1-(p_{1,1}-p_{0,1})}\\ &= \frac{p_{0,1}}{1-(p_{1,1}-p_{0,1})} - \frac{p_{0,1}(p_{1,1}-p_{0,1})^n}{1-(p_{1,1}-p_{0,1})}\end{aligned} \tag{3.5}$$

by using geometric series. Therefore equation (3.4) can be written

$$\mathbb{P}(X_n=1) = \frac{p_{0,1}}{1-(p_{1,1}-p_{0,1})} + \left[p_1 - \frac{p_{0,1}}{1-(p_{1,1}-p_{0,1})}\right](p_{1,1}-p_{0,1})^n, \tag{3.6}$$

when $p_{0,1} \neq p_{1,1}$. If we take a look at equation (3.6) we see that the choice of the initial distribution p_1 is crucial. The effect of p_1 is dampened exponentially and when $p_1 = p_{0,1}/1 - (p_{1,1} - p_{0,1})$ it has no effect at all. In that case $\mathbb{P}(X_n = 1)$ is the same for each n. This choice of p_1 is called the stationary initial distribution, which we will return to later on.

The following lemma is an important computation, called the Chapman-Kolmogorov equation. We let the state space $S = \{0, ..., N\}$, and define

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

Lemma 3.3. Chapman-Kolmogorov equation. It holds that

$$p_{i,j}^{(n)} = \sum_{k=0}^{N} p_{i,k}^{(m)} p_{k,j}^{(n-m)}, 1 \leq m \leq n-1. \tag{3.7}$$

Proof. For $1 \leq m \leq n-1$

$$\begin{aligned} p_{i,j}^{(n)} &= \mathbb{P}(\{X_n = j\} \cap (\cup_{k=0}^{N} \{X_m = k\}) | X_0 = i) \\ &= \sum_{k=0}^{N} \mathbb{P}(\{X_n = j\} \cap \{X_m = k\} | X_0 = i) \\ &= \sum_{k=0}^{N} \mathbb{P}(X_n = j, X_m = k | X_0 = i) \end{aligned}$$

By using Theorem 2.7, we know that

$$\begin{aligned} \frac{\mathbb{P}(X_n = j, X_m = k | X_0 = i)}{\mathbb{P}(X_m = k | X_0 = i)} &= \frac{\mathbb{P}(X_n = j, X_m = k, X_0 = i)\, \mathbb{P}(X_0 = i)}{\mathbb{P}(X_0 = i)\, \mathbb{P}(X_m = k, X_0 = i)} \\ &= \frac{\mathbb{P}(X_n = j, X_m = k, X_0 = i)}{\mathbb{P}(X_m = k, X_0 = i)} \\ &= \mathbb{P}(X_n = j | X_m = k, X_0 = i). \end{aligned}$$

Finally by using Markov property in the following equation we get that

$$\begin{aligned} p_{i,j}^{(n)} &= \sum_{k=0}^{N} \mathbb{P}(X_n = j | X_m = k, X_0 = i)\, \mathbb{P}(X_m = k | X_0 = i) \\ &= \sum_{k=0}^{N} \mathbb{P}(X_n = j | X_m = k)\, \mathbb{P}(X_m = k | X_0 = i) \\ &= \sum_{k=0}^{N} p_{i,k}^{(m)} p_{k,j}^{(n-m)} \end{aligned}$$

□

Equation (3.7) can be rewritten in matrix notation as

$$P^n \equiv (p_{i,j}^{(n)}) = P^{n-m} P^m.$$

Let $p_n = (p_n(0), ..., p_n(N)) = (\mathbb{P}(X_n = 0), ..., \mathbb{P}(X_n = N))$ denote the probability distribution of X_n. By recalling the computation of $\mathbb{P}(X_n = 1)$ in equation (3.4), we see that

$$p_n = p_{n-1} P \tag{3.8}$$

can be calculated with the initial distribution p_0. Therefore

$$p_n = p_0 P^n. \tag{3.9}$$

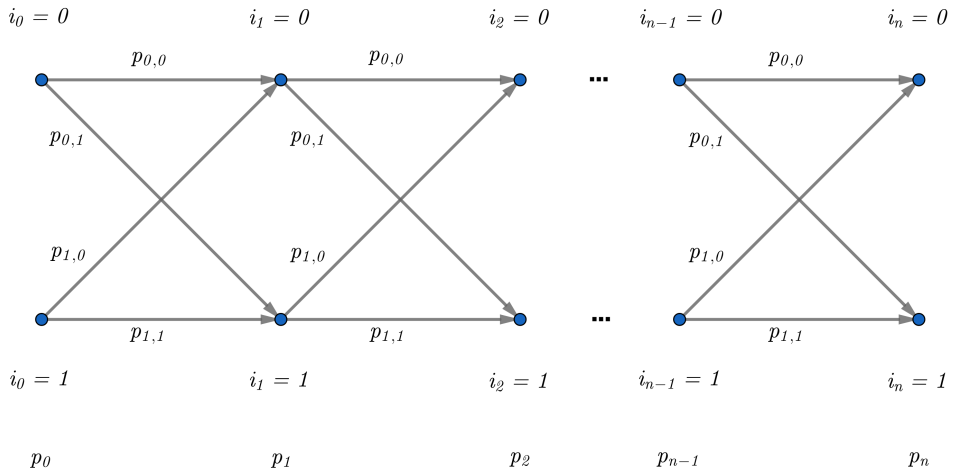

Figure 3.2: Visualization of the transitions for S = {0,1}.

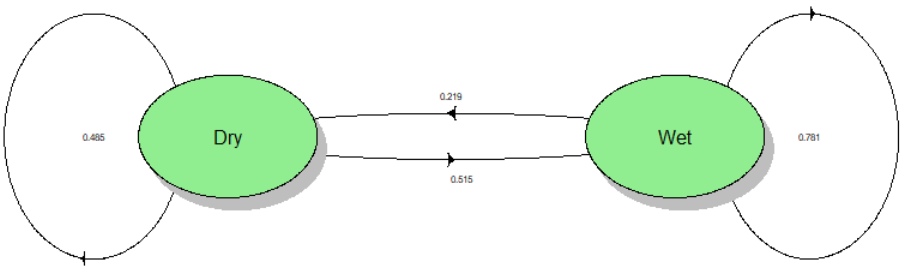

Figure 3.3: Kilpisjärvi transition diagram

Example 3.4. (Kilpisjärvi precipitation) The observed data in table (3.1) can be used as a Markov chain model. Let us say that it rains on January 1 this year. In our model that is considered as "wet", i.e. the initial distribution $p_0 = (0, 1)$. Now, what would the probability of rain be 6 days hence on January 7? By using (3.9) we see that we need to determine $P^{(6)}$, where

$$P = \begin{pmatrix} 0.485 & 0.515 \\ 0.219 & 0.781 \end{pmatrix} \tag{3.10}$$

and after matrix-multiplication of P we get that

$$P^{(6)} = \begin{pmatrix} 0.299 & 0.701 \\ 0.298 & 0.702 \end{pmatrix}. \tag{3.11}$$

Since we want to know the probability distribution (p_n) 6 days hence, given that we obeserve "rain" the first day we get from (3.9)

$$p_6 = (0, 1)P^{(6)} = (0.298, 0.702). \tag{3.12}$$

In the first column in (3.12) we have the probability of a "dry" day on January 7, given "wet" day on January 1. In the second column we have the probability we want to compute. The probability of a "wet" day in Kilpisjärvi on January 7, given a "wet" day on January 1, is 0.702.

3.3 Classification of states

Definition 3.5. For a Markov chain $\{X_n\}_{n=0}^{\infty}$ in the state space S, the hitting time of a given state $j \in S$ is defined as

$$T_j := \inf\{n > 0 : X_n = j\}. \tag{3.13}$$

Definition 3.6. A random variable $T \in \mathbb{Z}_+ \cup \{\infty\}$ is called a stopping time if the event $\{T = m\}$ can be expressed in terms of $X_1, X_2, ..., X_m$.

This means that a stopping time is a specific random time such that we at time m know weather an event has occured or not. The hitting time T_j is a stopping time, since the event $\{T_j = m\} = \{X_1 \neq j, X_2 \neq j, ..., X_m = j\}$.

Theorem 3.7. (The strong Markov property) Let $\{X_n\}$ be an M.C. with the transitions probabilities $p_{i,j}$ and statespace S. Also, let T_j be a stopping time with respect to $\{X_n\}$. Then for any integer m, we need to show that

$$\begin{aligned} &\mathbb{P}(X_{T+m} = j | X_0 = i_0, X_1 = i_1, ..., X_T = i) \\ &= \mathbb{P}(X_m = j | X_0 = i) = p_{i,j}^{(m)} \end{aligned} \tag{3.14}$$

and

$$\mathbb{P}(X_{T+m} = j | X_T = i) = \mathbb{P}(X_m = j | X_0 = i) = p_{i,j}^{(m)}. \tag{3.15}$$

Proof. First we prove equality (3.14):

$$\begin{aligned} &\mathbb{P}(X_{T+m} = j | X_0 = i_0, X_1 = i_1, ..., X_T = i) \\ =&\frac{\mathbb{P}(X_{T+m} = j, X_0 = i_0, X_1 = i_1, ..., X_T = i)}{\mathbb{P}(X_0 = i_0, X_1 = i_1, ..., X_T = i)} \\ =&\frac{\sum_{\tau=1}^{\infty} \mathbb{P}(X_{T+m} = j, X_0 = i_0, X_1 = i_1, ..., X_T = i, T = \tau)}{\mathbb{P}(X_0 = i_0, X_1 = i_1, ..., X_T = i)} \end{aligned} \tag{3.16}$$

Next, since T is a stopping time, the event $\{T = \tau\}$ can be expressed as the process up to time τ, i.e. as $X_0, X_1, ..., X_\tau$. The Markov property now gives us that

$$\begin{aligned}&\mathbb{P}(X_{T+m} = j | X_0 = i_0, ..., X_T = i, T = \tau) = \mathbb{P}(X_{T+m} = j | X_T = i) \\ &= p_{i,j}^{(m)}\end{aligned} \tag{3.17}$$

Hence, equation (3.16) becomes

$$\begin{aligned}&\frac{\sum_{\tau=1}^{\infty} \mathbb{P}(X_{\tau+m} = j, X_0 = i_0, ..., X_T = i, T = \tau)}{\mathbb{P}(X_0 = i_0, ..., X_T = i)} \\ =&\frac{\sum_{\tau=1}^{\infty} \mathbb{P}(X_{\tau+m} = j, X_0 = i_0, ..., X_T = i, T = \tau)\, \mathbb{P}(X_0 = i_0, ..., X_T = i, T = \tau)}{\mathbb{P}(X_0 = i_0, ..., X_T = i)} \\ =&\frac{\sum_{\tau=1}^{\infty} p_{i,j}^{(m)}\, \mathbb{P}(X_0 = i_0, ..., X_T = i, T = \tau)}{\mathbb{P}(X_0 = i_0, ..., X_T = i)} \\ =&p_{i,j}^{(m)} \frac{\sum_{\tau=1}^{\infty} \mathbb{P}(X_0 = i_0, ..., X_T = i, T = \tau)}{\mathbb{P}(X_0 = i_0, ..., X_T = i)} \\ =&p_{i,j}^{(m)} = \mathbb{P}(X_m = j | X_0 = i)\end{aligned} \tag{3.18}$$

Finally we get equality (3.15) from (3.14):

$$\begin{aligned}&\mathbb{P}(X_{T+m} = j | X_T = i) = \frac{\mathbb{P}(X_{T+m} = j, X_T = i)}{\mathbb{P}(X_T = i)} \\ =&\frac{\sum_{\tau=1}^{\infty} \mathbb{P}(X_{T+m} = j, X_T = i, T = \tau)}{\mathbb{P}(X_T = i)} \\ =&\frac{\sum_{\tau=1}^{\infty} \mathbb{P}(X_{T+m} = j | X_T = i, T = \tau)\, \mathbb{P}(X_T = i, T = \tau)}{\mathbb{P}(X_T = i)} \\ =&p_{i,j}^{(m)} \frac{\sum_{\tau=1}^{\infty} \mathbb{P}(X_T = i, T = \tau)}{\mathbb{P}(X_T = i)} \\ =&p_{i,j}^{(m)}\end{aligned} \tag{3.19}$$

□

Definition 3.8. State i reaches state j, denoted as $i \to j$, if

$$\exists n \geq 0 : p_{i,j}^{(n)} > 0.$$

Definition 3.9. Two states i and j that reach each other are said to communicate, denoted as $\leftrightarrow$, i.e.

$$i \leftrightarrow j \text{ if } i \to j \text{ and } \ j \to i. \tag{3.20}$$

Definition 3.10. The probability of an event associated with the M.C. when the chain starts at state i is defined as

$$\mathbb{P}_i(...) := \mathbb{P}(...|X_0 = i), \tag{3.21}$$

and i is called the initial state.

Proposition 3.11. For all $i, j \in \{0, ..., N\}$ it holds

$$p_{i,j}^{(n)} = \sum_{m=1}^{n} \mathbb{P}_i(T_j = m) p_{j,j}^{(n-m)}$$

Proof. Write $\{X_n = j\} = \sum_{m=1}^{n} \{T_j = m, X_n = j\}$. Starting from initial state i by using conditional probability and finally taking advantage of the Markov property we get that

$$\begin{aligned}
p_{i,j}^{(n)} &= \mathbb{P}_i(X_n = j) \\
&= \sum_{m=1}^{n} \mathbb{P}_i((T_j = m) \cap (X_n = j)) \\
&= \sum_{m=1}^{n} \mathbb{P}_i(X_n = j | T_j = m) \mathbb{P}_i(T_j = m) \\
&= \sum_{m=1}^{n} \mathbb{P}(X_n = j | X_0 = i, X_1 \neq j, ..., X_{m-1} \neq j, X_m = j) \mathbb{P}_i(T_j = m) \\
&= \sum_{m=1}^{n} \mathbb{P}(X_n = j | X_m = j) \mathbb{P}_i(T_j = m) \\
&= \sum_{m=1}^{n} \mathbb{P}_i(T_j = m) p_{j,j}^{(n-m)}.
\end{aligned}$$

□

Definition 3.12. A state j is called absorbing if

$$p_{j,j} = 1,$$

which means that if the chain reaches state j it will stay there with probability one.

Corollary 3.13. For an absorbing state j it holds that

$$p_{i,j}^{(n)} = \mathbb{P}_i(T_j \leq n). \tag{3.22}$$

Proof. Equation (3.22) says that starting from the state i, the chain has to hit j before or at time n. From Proposition 3.11 we know that

$$p_{i,j}^{(n)} = \sum_{m=1}^{n} \mathbb{P}_i(T_j = m) p_{j,j}^{(n-m)}.$$

Since j is an absorbing state we have that $p_{j,j}^{(n-m)} = 1$ for all $m \leq n$. Therefore $p_{i,j}^{(n)} = \mathbb{P}_i(T_j \leq n)$ for an absorbing state j. □

Theorem 3.14. $\leftrightarrow$ is an equivalence relation.

Proof. $P_{i,i}^{(n)} = \mathbb{P}(X_n = i | X_0 = i)$ and $P_{i,i}^{(0)} = 1$ which means that $i \leftrightarrow i$. Next if i communicates with k and k communicates with j we can find integers m and n such that $p_{i,k}^{(m)} > 0$ and $p_{k,j}^{(n)} > 0$. Then by summing over all possible states t in the state space S we have that

$$p_{i,j}^{(m+n)} = \sum_{t=0}^{N} p_{i,t}^{(m)} p_{t,j}^{(n)} \geq p_{i,k}^{(m)} p_{k,j}^{(n)} > 0 \tag{3.23}$$

and $i \to j$. Similarly we can show that $j \to i$ since

$$p_{j,i}^{(r+s)} = \sum_{t=0}^{N} p_{j,t}^{(r)} p_{t,i}^{(s)} \geq p_{j,u}^{(r)} p_{u,i}^{(s)} > 0 \tag{3.24}$$

for integers r and s such that $p_{j,u}^{(r)} > 0$ and $p_{u,i}^{(s)} > 0$. □

Definition 3.15. A Markov chain $(X_0, X_1, ...)$ with state space $S = \{0, ..., N\}$ is irreducible if for all $i, j \in S$ there exist nonnegative integers k and l such that

$$\mathbb{P}(X_k = j | X_0 = i) > 0 \text{ and } \mathbb{P}(X_l = i | X_0 = j) > 0.$$

This means that a chain is irreducible if, given any two states, the states with a positive probability can reach one another. Any two states of the chain are hereby communicating and will always be reachable.

Definition 3.16. The period $d(s_i)$ of a certain state s_i of a given Markov chain with state space S is defined by

$$d(s_i) = gcd\{n \geq 1 : (p_{i,i}^{(n)} > 0)\}. \tag{3.25}$$

Here the abbreviation "gcd" stands for the greatest common divisor. A state s_i is said to be aperiodic if $d(s_i) = 1$ in (3.25). If $d(s_i) = 1$ holds for all $i \in S$ of a Markov chain, it means that the chain itself is aperiodic, otherways the chain is called periodic.
A chain that is both irreducible and aperiodic (Definition 3.15 and 3.16) is called ergodic.

Theorem 3.17. Periodicity is an equivalence class property, i.e., if i communicates with j then $d(i) = d(j)$.

Proof. Let r, s be such that $p_{i,j}^{(r)} > 0, p_{j,i}^{(s)} > 0$ and assume that $p_{j,j}^{(t)} > 0$. Then

$$0 < p_{i,j}^{(r)} p_{j,i}^{(s)} \leq p_{i,i}^{(r+s)} \tag{3.26}$$

and

$$0 < p_{i,j}^{(r)} p_{j,i}^{(s)} p_{i,i}^{(t)} \leq p_{i,i}^{(r+s+t)}. \tag{3.27}$$

By definition $d(i)$ must be a fraction of $r + s$ and $r + s + t$, since state i can not have multiple periods. Hence $d(i)$ must divide their difference t for any t such that $p_{j,j}^{(t)} > 0$. Therefore $d(i)$ divides $d(j)$. By similar arguments $d(j)$ divides $d(i)$, so the two numbers have to be equal. □

Let $f_{i,j}^{(n)} = \mathbb{P}_i(T_j = n)$ be the first passage distribution from starting state i, to ending state j. In zero steps, i.e. when $n = 0$, we have that $f_{i,j}^{(0)} = 0$ and in general for $n \geq 1$

$$f_{i,j}^{(n)} = \mathbb{P}(X_n = j, X_k \neq j, k = 1, ..., n-1 | X_0 = i). \tag{3.28}$$

Lemma 3.18. Let n_1 and n_2 be two positive integers that are relatively prime (greatest common divider = 1). Then any integer $n > n_1 n_2$ can be written as $n = sn_1 + tn_2$, for non-negative integers s and t.

Proof. Consider the modulo n_2 congruence classes of the n_2 distinct positive integers $n - 0n_1, n - 1n_1, n - 2n_1, ..., n - (n_2 - 1)n_1$. We have two possibilities with these congruence classes. The first possibility is that all the congruence classes are different and the second one is that at least two congruence

classes are the same. If all congruence classes are different it means that one congruence class must be 0. Then $n - sn_1$ is divisible by n_2 such that

$$\frac{n - sn_1}{n_2} = t \leftrightarrow n = sn_1 + tn_2. \tag{3.29}$$

Equation (3.29) holds if all congruence classes are different. In case two congruence classes are the same, we are able to write $n - sn_1 = kn_2 + l$ and $n - tn_1 = mn_2 + l$ for $0 \leq t < s \leq n_2 - 1, k < m$, and $0 < l \leq n_2 - 1$. For these two congruence classes to be equal we get that

$$\begin{aligned} & n - sn_1 - (n - tn_1) = kn_2 + l - (mn_2 + l) \\ \leftrightarrow \quad & -sn_1 + tn_1 = kn_2 - mn_2 \\ \leftrightarrow \quad & (s - t)n_1 = (m - k)n_2. \end{aligned} \tag{3.30}$$

Since n_1 and n_2 are relatively prime, for equation (3.30) to be true, $(s - t)$ must contain all prime factors of n_2. That is a contradiction, since then $s - t > n_2 - 1$. □

Proposition 3.19. Let X be an irreducible and aperiodic chain and $i, j \in S$. Then there is an integer $N = N(i, j)$ such that $p_{i,j}^{(n)} > 0$ for all $n \geq N$.

Proof. Since X is aperiodic, we know that $d(j) = 1$. Therefore we can find two integers n_1, n_2 that are relatively prime such that $p_{j,j}^{(n_1)} > 0$ and $p_{j,j}^{(n_2)} > 0$. From Lemma 3.18 any large enough n can be written as $n = sn_1 + tn_2$. Hence

$$p_{j,j}^{(n)} = p_{j,j}^{(sn_1+tn_2)} \geq \left(p_{j,j}^{(n_1)}\right)^s \left(p_{j,j}^{(n_2)}\right)^t > 0, \tag{3.31}$$

since $p_{j,j}^{(n_1)}, p_{j,j}^{(n_2)} > 0$ and $s, t \geq 0$. In other words, starting from state j, the probability of returning to state j in n steps is greater than zero. Now, for each pair i, j there is an n_0 such that $p_{i,j}^{(n_0)} > 0$. Therefore

$$p_{i,j}^{(n+n_0)} \geq p_{j,j}^{(n)} p_{i,j}^{(n_0)} > 0, \tag{3.32}$$

since $p_{j,j}^{(n)}, p_{i,j}^{(n_0)} > 0$. □

Corollary 3.20. Assume that X and Y are identically and independently distributed (iid) irreducible aperiodic Markov chains. Furthermore let $Z = (X, Y)$. Then Z is an irreducible Markov chain.

Proof. The Markov property holds for Z since

$$\begin{aligned}
&\mathbb{P}(Z_n = (j,l)|Z_0 = (i_0,k_0), Z_1 = (i_1,k_1), ..., Z_{n-1} = (i_{n-1},k_{n-1})) \\
&= \mathbb{P}(X_n = j, Y_n = l|X_0 = i_0, Y_0 = k_0; X_1 = i_1, Y_1 = k_1; ...; X_{n-1} = i_{n-1}, \\
&\quad Y_{n-1} = k_{n-1}) \\
&= \mathbb{P}(X_n = j, Y_n = l|X_{n-1} = i_{n-1}, Y_{n-1} = k_{n-1}) \\
&= \mathbb{P}(Z_n = (j,l)|Z_{n-1} = (i_{n-1},k_{n-1})).
\end{aligned} \tag{3.33}$$

By using Theorem 2.7 we get the transition probabilities for Z as following:

$$\begin{aligned}
p_{i,k;j,l} &= \mathbb{P}(Z_n = (j,l)|Z_{n-1} = (i,k)) \\
&= \mathbb{P}(X_n = j, Y_n = l|X_{n-1} = i, Y_{n-1} = k) \\
&= \frac{\mathbb{P}\left(\{X_n = j\} \cap \{X_{n-1} = i\}\right)}{\mathbb{P}(X_{n-1} = i)} \frac{\mathbb{P}\left(\{Y_n = l\} \cap \{Y_{n-1} = k\}\right)}{\mathbb{P}(Y_{n-1} = k)} \\
&= \mathbb{P}(X_n = j|X_{n-1} = i)\, \mathbb{P}(Y_n = l|Y_{n-1} = k) \\
&= p_{i,j} p_{k,l},
\end{aligned} \tag{3.34}$$

since X and Y are iid. By Proposition 3.19 there is an $N = (i,k,j,l)$ such that $p_{i,j}^{(n)} > 0$ and $p_{k,l}^{(n)} > 0$ for $n > N$. Hence $p_{i,k;j,l}^{(n)} > 0$ which means that Z is irreducible. □

Definition 3.21.

$$f_{i,j} := \sum_{n=0}^{\infty} f_{i,j}^{(n)} = \mathbb{P}_i(T_j < \infty) \tag{3.35}$$

The state i in equation (3.35) is persistent, or by another name recurrent, if $f_{i,i} = 1$. Otherwise state i is called transient. We can think of transient states as if there are one or more absorbing states in the chain, there is a chance that we will never return to starting state again. It is also possible that the chain never enters the starting state after running for a while. In both of these cases we have a positive probability of no return with $f_{i,i} \neq 1$. On the other hand a persistent state is one that the process will eventually with certainty return to, in a finite amount of steps.

Theorem 3.22. A state i is persistent if and only if $\sum_{n=1}^{\infty} p_{i,i}^{(n)} = \infty$.

Proof. Suppose i is transient and let M be the number of returns to i. Then

$$M(i) = \sum_{n=1}^{\infty} 1(X_n = i). \tag{3.36}$$

Since

$$f_{i,i} = \mathbb{P}_i(T_i < \infty) = \mathbb{P}_i(M \geq 1)$$

and

$$\begin{aligned} \mathbb{P}_i(M \geq 2) &= \mathbb{P}_i(T_i < \infty, T_i^{(2)} < \infty) \\ &= \mathbb{P}_i(T_i < \infty)\mathbb{P}_i(T_i < \infty) \\ &= f_{i,i}^2, \end{aligned}$$

where $T_i^{(2)}$ is the time we return to i the second time. Because the chain starts over once we return to i and reach state X_{T_i} by strong Markov property. Hence by the strong Markov property

$$\mathbb{P}_i(M \geq k) = (\mathbb{P}_i(T_i < \infty))^k = f_{i,i}^k,$$

which is the probability that we will return to i, k times or more. The expected value of returns to i is hence

$$\mathbb{E}_i[M] = \sum_{k=1}^{\infty} \mathbb{P}_i(M \geq k). \tag{3.37}$$

We are interested in how many times we return to state i and only have two options for each step, i.e. we return to state i or we do not return to state i. We also need to know how many times we hit state i, until i can not be reached anymore (i transient). Hence $M - 1$ has a geometric distribution (Definition 2.8) with parameter $f_{i,i}$. Theorem 2.10 states that

$$\mathbb{E}_i[M] = \frac{f_{i,i}}{(1 - f_{i,i})}. \tag{3.38}$$

Since state i is transient we know that $f_{i,i} < 1$ and $\mathbb{E}_i[M] < \infty$, and consequentely

$$\sum_{n=1}^{\infty} p_{i,i}^{(n)} = \sum_{n=1}^{\infty} \mathbb{E}_i[1(X_n = i)] = \mathbb{E}_i[M] < \infty. \tag{3.39}$$

On the other hand, if i is persistent, the chain returns to i with probability 1. Then the process starts over and again the chain will return to i with probability 1. This happens an infinite number of times. By the strong Markov property $\mathbb{P}_i(M = \infty) = 1$ and $\mathbb{E}_i[M] = \infty$, i.e.

$$\sum_{n=1}^{\infty} p_{i,i}^{(n)} = \mathbb{E}_i[M] = \infty. \tag{3.40}$$

□

Definition 3.23. The mean recurrence time of a persistent state i with probability distribution $f_{i,i}^{(n)}$ is defined as

$$\mu_i := \sum_{n=1}^{\infty} f_{i,i}^{(n)}. \tag{3.41}$$

State i is called null persistent if $\mu_i = \infty$ and positive persistent if $\mu_i < \infty$.

Corollary 3.24. For a transient state i it holds that $\lim_{n\to\infty} p_{i,i}^{(n)} \to 0$.

Proof. Since i is transient, $f_{i,i} < 1$ and thus $\sum_{n=1}^{\infty} p_{i,i}^{(n)} < \infty$. Thus $\lim_{n\to\infty} p_{i,i}^{n} = 0$. □

3.4 The long time behaviour and stationary distribution of Markov chains

Definition 3.25. A row vector $\pi = (\pi_0, \pi_1, ..., \pi_N)$ is a stationary distribution for a Markov chain if $\pi_i \geq 0$ for $i = 0, ..., N$ and $\sum_{i=0}^{N} \pi_i = 1$ and $\pi P = \pi$, which is equivalent of solving the equation

$$\pi(P - I) = 0 \tag{3.42}$$

where I is the identity matrix for P. I has ones on the main diagonal and zeros elsewhere.

Remark 3.26. A Markov chain need not to have a stationary distribution, but if it has a limiting distribution, then the limit is stationary, as one readily checks.

Example 3.27. Consider the Markov chain with the transition matrix

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Since the chain, with a periodicity of 2, will jump between the two states with probability 1, the chain never will converge to any certain distribution π. However, Definition 3.25 states that P has a stationary distribution, since

$$\left(\frac{1}{2}, \frac{1}{2}\right) \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \left(\frac{1}{2}, \frac{1}{2}\right).$$

While we manually calculate the stationary distribution of a Markov chain, it can be convenient to take the transpose of equation (3.42), depending on how the transition matrix we work with looks like. This will be done in the following example where we calculate the stationary distribution using Gaussian elimination.

Example 3.28. Consider a Markov chain with the state space $S = (0, 1, 2, 3, 4)$ and transition matrix

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

and we want to find a stationary distribution of P. Recall from equation (3.42) that we want to solve the equation $\pi(P - I) = 0$ that is equivalent of solving the transpose of the equation, i.e. $(P^T - I)\pi^T = 0$. We start by calculating the matrix $P^T - I$ by using Gaussian elimination.

$$P^T - I = \begin{pmatrix} -1 & \frac{1}{2} & \frac{1}{3} & 0 & \frac{1}{2} \\ \frac{1}{2} & -1 & \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & -1 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} & -1 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & -1 \end{pmatrix}$$
$$\sim \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{3} & 0 & -\frac{1}{2} \\ \frac{1}{2} & -1 & \frac{1}{3} & 0 & 0 \\ \frac{1}{2} & \frac{1}{2} & -1 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} & -1 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & -1 \end{pmatrix}$$

$$
\begin{aligned}
&\sim \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{3} & 0 & -\frac{1}{2} \\ 0 & -\frac{3}{4} & \frac{1}{2} & 0 & \frac{1}{4} \\ 0 & \frac{3}{4} & -\frac{5}{6} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & \frac{1}{3} & -1 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & -1 \end{pmatrix} \\
&\sim \begin{pmatrix} 1 & -\frac{1}{2} & -\frac{1}{3} & 0 & -\frac{1}{2} \\ 0 & 1 & -\frac{2}{3} & 0 & -\frac{1}{3} \\ 0 & \frac{3}{4} & -\frac{5}{6} & \frac{1}{2} & \frac{1}{4} \\ 0 & 0 & \frac{1}{3} & -1 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & -1 \end{pmatrix} \\
&\sim \begin{pmatrix} 1 & 0 & -\frac{2}{3} & 0 & -\frac{2}{3} \\ 0 & 1 & -\frac{2}{3} & 0 & -\frac{1}{3} \\ 0 & 0 & -\frac{1}{3} & \frac{1}{2} & \frac{1}{2} \\ 0 & 0 & \frac{1}{3} & -1 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & -1 \end{pmatrix} \\
&\sim \begin{pmatrix} 1 & 0 & -\frac{2}{3} & 0 & -\frac{2}{3} \\ 0 & 1 & -\frac{2}{3} & 0 & -\frac{1}{3} \\ 0 & 0 & 1 & -\frac{3}{2} & -\frac{3}{2} \\ 0 & 0 & \frac{1}{3} & -1 & \frac{1}{2} \\ 0 & 0 & 0 & \frac{1}{2} & -1 \end{pmatrix} \\
&\sim \begin{pmatrix} 1 & 0 & 0 & -1 & -\frac{5}{3} \\ 0 & 1 & 0 & -1 & -\frac{4}{3} \\ 0 & 0 & 1 & -\frac{3}{2} & -\frac{3}{2} \\ 0 & 0 & 0 & -\frac{1}{2} & 1 \\ 0 & 0 & 0 & \frac{1}{2} & -1 \end{pmatrix} \\
&\sim \begin{pmatrix} 1 & 0 & 0 & -1 & -\frac{5}{3} \\ 0 & 1 & 0 & -1 & -\frac{4}{3} \\ 0 & 0 & 1 & -\frac{3}{2} & -\frac{3}{2} \\ 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & \frac{1}{2} & -1 \end{pmatrix} \\
&\sim \begin{pmatrix} 1 & 0 & 0 & 0 & -\frac{11}{3} \\ 0 & 1 & 0 & 0 & -\frac{10}{3} \\ 0 & 0 & 1 & 0 & -\frac{9}{2} \\ 0 & 0 & 0 & 1 & -2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}
\end{aligned}
$$

Here the notation "$\sim$" means that row operations have been calculated according to Gaussian elimination. Since there are only zeros in the last row, π_5 will be a free variable. We now have an equation system that consists of 5 unknown variables and all equations are functions of π_5. The stationary distribution can be calculated as following. We have that

$$\begin{cases} \pi_1 = \frac{11}{3}x_5 \\ \pi_2 = \frac{10}{3}x_5 \\ \pi_3 = \frac{9}{2}x_5 \\ \pi_4 = 2x_5 \\ \pi_5 = \text{free variable} \end{cases} \qquad \text{where } \pi^T = \pi_5 \begin{bmatrix} \frac{11}{3} \\ \frac{10}{3} \\ \frac{9}{2} \\ 2 \\ 1 \end{bmatrix}$$

Definition 3.25 states that π_5 should be such that

$$\sum_{i=1}^{5} \pi_i^T = \pi_5 \left(\frac{87}{6} \right) = 1$$

that gives us $\pi_5 = \left(\frac{2}{29}\right)$ and thus $\pi_1 = \left(\frac{22}{87}\right), \pi_2 = \left(\frac{20}{87}\right), \pi_3 = \left(\frac{9}{29}\right), \pi_4 = \left(\frac{4}{29}\right)$. The stationary distribution as a row vector is thereby

$$\pi = \left(\frac{22}{87}, \frac{20}{87}, \frac{9}{29}, \frac{4}{29}, \frac{2}{29} \right).$$

Lemma 3.29. Let μ_k be as in (3.41) and $v_{i,k}$ as in (3.44). For an irreducible and positive persistent Markov chain there exists a stationary distribution

$$\pi_i = \frac{v_{i,k}}{\mu_k} \tag{3.43}$$

for a fixed state k and $i, k \in S$. For a proof of this statement, see [5].

Theorem 3.30. The expected number of visits to state i between successive visits to state j is given by

$$v_{i,j} = \mathbb{E}_j \left[\sum_{n=0}^{T_j - 1} 1(X_n = i) \right] = \frac{\mu_j}{\mu_i} = \frac{\pi_i}{\pi_j} \tag{3.44}$$

for $i, j \in S$. For a proof of this statement, see [5].

Theorem 3.31. An irreducible Markov chain has a stationary distribution if and only if the the chain is positive persistent. This implies that the stationary distribution is unique and given by

$$\pi_i = \frac{1}{\mu_i}. \tag{3.45}$$

Theorem 3.32. For an irreducible Markov chain with mean recurrence time $\mu_j \leq \infty$, if j is an aperiodic state, then

$$\lim_{n\to\infty} p_{j,j}^{(n)} = \frac{1}{\mu_j}. \tag{3.46}$$

Corollary 3.33. If a Markov chain is irreducible and aperiodic, then

$$\lim_{n\to\infty} p_{i,j}^{(n)} = \frac{f_{i,j}}{\mu_j}. \tag{3.47}$$

For a proof of these three statements above, see [5].

Definition 3.34. The time spent in state j is defined as

$$N_j(n) := \sum_{i=1}^{n} 1(X_i = j), \tag{3.48}$$

where $N_j(n)$ counts the number of times we hit state j.

Corollary 3.35. Let j be a persistent and an aperiodic state. Then

$$\lim_{n\to\infty} \mathbb{E}\left[\frac{N_j(n)}{n}\right] = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} p_{j,j}^{(i)} = \frac{1}{\mu_j} \tag{3.49}$$

for all starting states communicating with j. Equation (3.49) is called the Cesáro-limit of the transition probabilities $p_{j,j}^{(i)}$.

Proof. Theorem 3.32 states that $p_{j,j}^{(i)} \to \frac{1}{\mu_j}$ as $n \to \infty$. We make use of the basic fact that if a limit exists, it equals the cesáro-limit. Next we have that

$$\mathbb{E}_j[N_j(n)] = \sum_{i=1}^{n} \mathbb{P}_j(X_i = j) = \sum_{i=1}^{n} p_{j,j}^{(i)}. \tag{3.50}$$

Because of persistence in the equation above, starting from state j, we know that we will return to j with probability one. The same holds if we change the starting state from j to k, if j and k communicate with each other. We get that

$$\mathbb{E}_k[N_j(n)] = \sum_{i=1}^{n} \mathbb{P}_k(X_i = j) = \sum_{i=1}^{n} p_{k,j}^{(i)}. \tag{3.51}$$

When we let $n \to \infty$ in (3.51), Corollary 3.33 states that

$$\sum_{i=1}^{n} p_{k,j}^{(i)} \to \frac{f_{k,j}}{\mu_j} = \frac{1}{\mu_j}, \tag{3.52}$$

since k and j communicate and therefore $f_{k,j} = 1$. □

Theorem 3.36. Given an ergodic state j, the limiting occupation probability of j is $\frac{1}{\mu_j}$ with probability one, i.e.

$$\mathbb{P}\left(\lim_{n\to\infty} \frac{N_j(n)}{n} \to \frac{1}{\mu_j}\right) = 1. \tag{3.53}$$

Proof. Suppose that the Markov chain starts in state j. Let $T_j(1), T_j(2), ...$ be the successive times when the chain hit state j. When we hit state j, at specific random time, we can by the strong Markov property consider it as restarting the chain. Hence $T_j(1), T_j(2) - T_j(1), T_j(3) - T_j(2), ...$ are independent and identically distributed random variables with probability generating function $F_{j,j}(s)$ and mean $\mu_j < \infty$. By the Strong law of large numbers (2.12) we have that

$$\begin{aligned} &\lim_{l\to\infty} \frac{T_j(1) + (T_j(2) - T_j(1)) + ... + (T_j(l) - T_j(l-1))}{l} \\ &= \lim_{l\to\infty} \frac{T_j(l)}{l} = \mu_j \end{aligned} \tag{3.54}$$

with probability 1 for the specific random times $T_j(l)$. We have that $\frac{N_j(n)}{n}$ is the proportion of time spent in state j up to time n. Also, we reach state j before, at, or after n steps. Hence

$$T_j(N_j(n)) \leq n \leq T_j(N_j(n) + 1). \tag{3.55}$$

Furthermore, $N_j(n) \to \infty$ as $n \to \infty$ with probability one, since we hit state j an infinite amount of times. Thus, as $n \to \infty$, with probability one

$$\frac{N_j(n)}{n} \leq \frac{N_j(n)}{T_j(N_j(n))} \to \frac{1}{\mu} \tag{3.56}$$

and

$$\frac{N_j(n)+1}{n} \geq \frac{N_j(n)+1}{T_j(N_j(n)+1)} \to \frac{1}{\mu} \tag{3.57}$$

which means that

$$\frac{N_j(n)}{n} \to \frac{1}{\mu}. \tag{3.58}$$

□

Theorem 3.37. (Ergodic theorem for Markov chains) For a positive persistent chain X, if $f : S \to \mathbb{R}$ satisfies $\mathbb{E}_\pi[|f(X_1)|] < \infty$, where π is the stationary distribution, we have that

$$\frac{1}{n}\sum_{k=1}^{n} f(X_k) \to \mathbb{E}_\pi[f(X_1)] \tag{3.59}$$

in probability, regardless of the initial distribution.

Proof. As in Theorem 3.36, we let $T_j(l)$ stand for specific random times of successive hits of state j. By splitting the random times into intervals, each interval represents one successive return to state j. Then we let Y_i represent the intervals with corresponding transitions so that

$$Y_i = \sum_{T_j(l)+1}^{T_j(l+1)} f(X_k). \tag{3.60}$$

$T_j(0) \equiv 0$ since even though we start at state j, it does not count as a successive return to j. By the strong Markov property are $Y_0, Y_1, ...$ independent and furthermore $Y_1, Y_2, ...$ identical.Y_0 starts from an initial distribution at time 0, hence the distribution differs from $Y_1, Y_2, ...$ By decomposing we get

$$\begin{aligned}\sum_{k=1}^{n} f(X_k) &= Y_0 + \sum_{k=1}^{N_j(n)} Y_k - \sum_{k=n+1}^{T_j(N_j(n))} f(X_k) \\ &\equiv Y_0 + S_{N_j(n)} - R_n.\end{aligned} \tag{3.61}$$

Since we have that

$$\begin{cases} Y_0 = f(X_1) + ... + f(X_{T_j(1)}) \\ Y_1 = f(X_{T_j(1)+1}) + ... + f(X_{T_j(2)}) \\ \vdots \\ Y_n = f(X_{T_j(n)+1}) + ... + f(X_{T_j(n+1)}) \end{cases} \tag{3.62}$$

we know that Y_i, for any $i \in S$, is a sum of a finite number of random variables. Therefore $\lim_{n\to\infty} \frac{Y_n}{n} = 0$. Next, persistence ensures us that, as $n \to \infty, \mathbb{P}(N_j(n)) \to \infty) = 1$. Provided that $\mathbb{E}[|Y_1|] < \infty$, the law of large numbers gives us that

$$\begin{aligned} &\frac{f\left(X_{T_j(1)+1}\right) + ... + f\left(X_{T_j(2)}\right) + ... + f\left(X_{T_j(N_j(n)+1)}\right) + ... + f\left(X_{T_j(N_j(n+1))}\right)}{N_j(n)} \\ &= \frac{S_{N_j(n)}}{N_j(n)} \to \mathbb{E}[Z_1] \end{aligned} \tag{3.63}$$

in probability. Also, according to (3.36) $\frac{N_j(n)}{n} \to \frac{1}{\mu_j} = \pi_j$. Hence since

$$\frac{S_{N_j(n)}}{n} = \frac{S_{N_j(n)}}{N_j(n)} \frac{N_j(n)}{n} \tag{3.64}$$

we get that

$$\frac{S_{N_j(n)}}{n} \to \pi_j \, \mathbb{E}[Y_1] \tag{3.65}$$

in probability. Next we know that the earliest time point we can hit state j the next time, after n amount of steps, is at time $n+1$. Also, we know that the time points $T_j(N_j(n)) + 1 \leq n + 1 \leq T_j(N_j(n) + 1)$. Hence

$$|R_n| \leq \sum_{k=n+1}^{T_j(N_j(n)+1)} |f(X_k)| \leq \sum_{k=T_j(N_j(n))+1}^{T_j(N_j(n)+1)} |f(X_k)| \equiv \eta_n. \tag{3.66}$$

By the strong Markov property $\eta_1, \eta_2, ...$ are independent and identically distributed. Hence

$$\mathbb{P}(|R_n| \geq n) \leq \mathbb{P}(\eta_n \geq n) \leq \frac{\mathbb{E}[\eta_n]}{n} \to 0, \tag{3.67}$$

by Markov's inequality (2.11)Next using Theorem 3.30, if $\mathbb{E}[\eta_1 < \infty]$ then $\mathbb{E}[|Y_1| < \infty]$ since

$$\mathbb{E}[\eta_1] = \mathbb{E}\left[\sum_{k=T_j(1)+1}^{T_j(2)} |f(X_k)|\right] = \sum_{i=1}^{m} |f(i)| v_{i,j} = \sum_{i=1}^{m} |f(i)| \frac{\pi_i}{\pi_j} \tag{3.68}$$

and

$$\mathbb{E}[Y_1] = \sum_{i=1}^{m} f(i) v_{i,j} = \frac{1}{\pi_j} \sum_{i=1}^{m} f(i) \pi_i \tag{3.69}$$

for $m \in S$. Then finally we have that

$$\frac{S_{N_j(n)}}{n} \to \pi_j \, \mathbb{E}[Y_1] = \pi_j \frac{1}{\pi_j} \sum_{i=1}^{m} f(i) \pi_i = \mathbb{E}_\pi f(X_1). \tag{3.70}$$

□

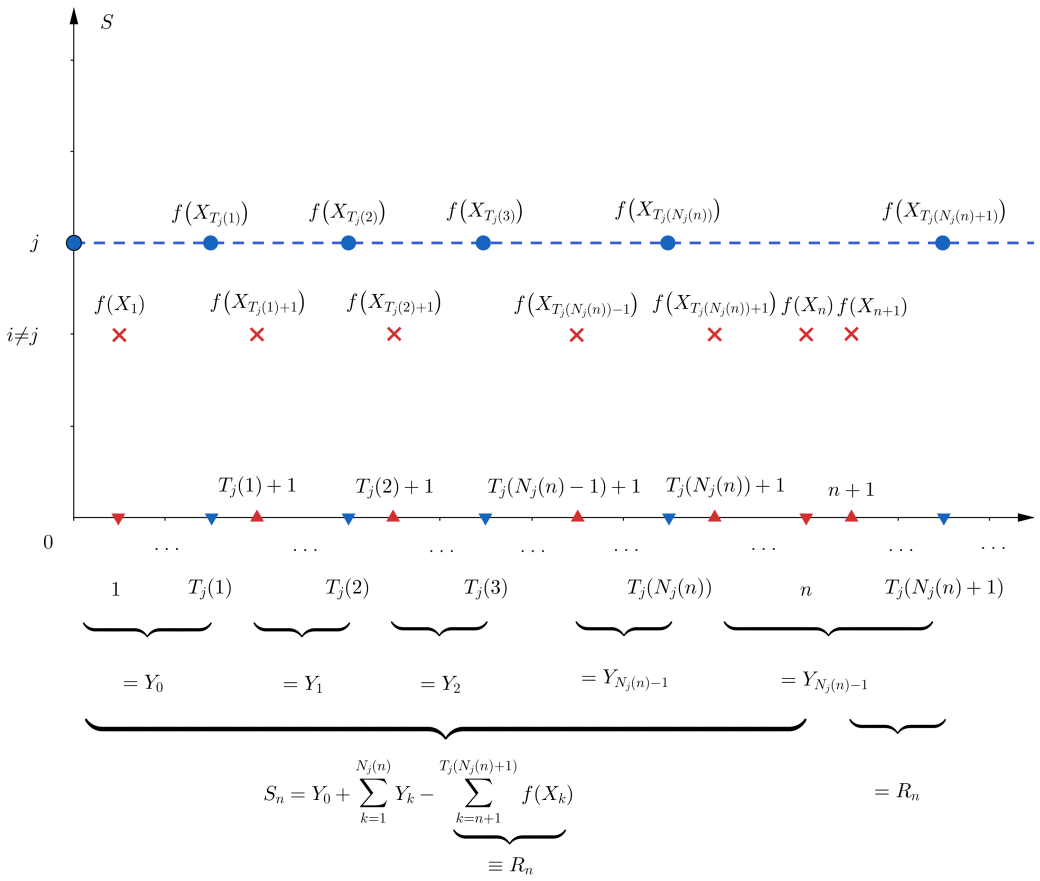

Figure 3.4: Visualization of the intervals in equation 3.60

Chapter 4

Simulation

The purpose of this chapter is to simulate the long time behaviour of Markov chains and compare it to theoretical results. The simulations are done with the programming language R, which is an environment for statistical computing and graphics. For each simulation in R the function *rmultinom*() will be used. That is because we want to include randomness between each transition from one transition to another. The inclusion of randomness reflects well real life situations, where random events usually occur and have an effect.

R has many built-in packages to manipulate data and new ones are continuously created and developed by users around the world. If you can not find a certain package in the library of R, it is usually possible to install new packages manually by downloading the packages to your computer. A more convenient way though is to use the built-in respository (CRAN) to install the package directly in R. In the coding process I have used RStudio that is a user environment to R with many built-in tools. One useful integrated tool is RMarkdown, which is a file format that can convert the code to many other formats. RMarkdown displays both the source code (input) and the result (output) in a convenient way, and hence I have used it for some of my created tables.

I have chosen to simulate two different ergodic Markov chains, where the first matrix is based on my manipulated data from FMI. The second matrix is made up and created by me, with the properties of an ergodic Markov chain. In Section 4.1 we go through the simulation algorithm, used for both matrices, followed by the coding process and the results. The coding process is similar for both matrices, but the data manipulation differs, since

the created matrix is larger in terms of dimension. The results from the simulations are presented in tables as estimates, where

$$\frac{N_j(n)}{n} = \hat{\pi}_j(n)$$

and

$$\frac{n}{N_j(n)} = \hat{\mu}_j(n).$$

In my results tables I also calculate

$$v_j = \frac{\pi_j}{\sum\limits_{i \in S: i \neq j} \pi_i},$$

which is similar to $v_{i,j}$ from (3.44). Estimates of v_j are presented in the result tables as $\hat{v}_j(n)$.

4.1 Algorithm

In order to reproduce the same results in a simulation R has a built-in function $set.seed(seed, kind = Null, normal.kind = Null)$, where $seed$ takes on a single value interpreted as an integer , or Null. Hence if we keep the same $seed$-value for a random number generator, $set.seed()$ function will produce the same random numbers. As random number generator we will use the built-in multinomial distribution $rmultinom(n, size, prob)$, where n is the number of random vectors to draw, $size$ the number of objects going into K boxes, and $prob$ a numeric positive vector of length K. $prob$ specifies the probability for the K classes. Next we use the built-in function

$$function(arglist)\{\{expr\}return(value)\}.$$

$arglist$ can consist of zero, one, or many expression terms as we call for the function. $expr$ and $value$ consist of an expression, where the expression itself can consist of several expressions. Before running the algorithm we set the seed to some number, which in this case is 148:

```
set.seed(148)
```

Again, the number inside $set.seed()$ could be another one, but in this case the number 148 will produce the same randomly generated vectors for this simulation. Next, we have the code for the algorithm:

```
set.seed(148)

sim.mark.chain <- function( P, num.iters=50, num.chains=5 ) {

num.states    <- nrow(P)
S             <- matrix(NA, ncol=num.chains, nrow=num.iters)
probabilities <- matrix(NA, nrow=num.iters, ncol=num.states)
pi_0          <- c(1, rep(0, num.states-1))
P_n           <- P
init.states   <- sample(1:num.states, num.chains, replace = T)

probabilities[1,] <- pi_0
S[1,]             <- init.states

  for(i in 2:num.iters) {


    pi_n              <- pi_0 %*% P_n

    probabilities[i,] <- pi_n

    for(k in seq_len(num.chains)) {

      p               <- P[ S[i-1,k], ]

      S[i,k] <- which(rmultinom(1, 1, p) == 1)

    }

    P_n <- P_n %*% P
  }
  return(list(probabilities, S, num.iters))
}
```

Figure 4.1: Code for simulation algorithm

sim.mark.chain is the name of the function that we call for when we simulate a transition matrix. P is in our case a matrix, *num.iters* a variable name for the number of iterations, and *num.chains* a variable name for the

number of simulations. The first step is that we save the number of rows in the input in the variable *num.states*. Then we create two new empty matrices named S and *probabilities*. S saves the number of columns according to *num.chains* and *probabilities* saves the number of rows according to *num.iters*. Above in picture (4.1) *num.iters* has the value of 50 that can be changed according to how many iterations, i.e. transitions, we want to do. *num.iters* is changed to greater values in the simulations. Similarly, *num.chains* can be changed, according to the number of times we simulate the Markov chain. Next we have the variables pi_0 and P_n, where P_n saves the input matrix and updates its values for each transition, as we see later on. pi_0 is a vector of the same length as the number of rows in *num.states*. The first number in the vector is 1 followed by one or more zeros depending on the number of rows in *num.states*. As an example, if the input matrix has 5 rows (= 5 states), then the output of pi_0 for one transition would be the following:

```
pi_0          <- c(1, rep(0, 5-1))
pi_0
```

```
## [1] 1 0 0 0 0
```

pi_0 is as following saved in the first row of the matrix *probabilities*. We then pick a random starting state, for each simulation, which is saved in the first row of the matrix S. The variable *init.states* saves all random starting states, which depend on the value of *num.chains*:

```
init.states   <- sample(1:num.states, num.chains, replace = T)

probabilities[1,] <- pi_0
S[1,]             <- init.states
```

We loop through the numbers of transitions from the second number, until *num.iters* reaches its last value:

```
for(i in 2:num.iters)
```

Similarly we loop through the number of simulated chains $k \in (1, num.chains)$ for each $i \in (2, num.iters)$:

```
for(k in seq_len(num.chains))
```

In the first for-loop a new variable pi_n obtains the value of $pi\_0 \times P\_n$, where P_n initially has the value of the input matrix P. For each transition P_n is updated and therefore also pi_n. pi_n is then saved in the matrix $probabilities$ in row i for $i \in (2, num.iters)$:

```
pi_n              <- pi_0 %*% P_n

probabilities[i,] <- pi_n
```

A new variable p gets the value from a row of P, in form of a vector, where the vector is taken from the $[i-1, k]$:th matrix position of S, storing all states for the transitions. This means that the state number returned from $S[i-1, k]$ determines the row number for P. Then the vector p storing probabilities, i.e. probability vector, is passed into the $rmultinom()$ function, where the function draws a sample from the probability vector. The sample number, determining the next state, is stored in S on row i, and column k:

```
p              <- P[ S[i-1,k], ]
S[i,k] <- which(rmultinom(1, 1, p) == 1)
```

Finally the matrix P_n, initially set as P, is updated by multiplying it to P for each iteration. The algorithm continues until it reaches its last value of $num.iters$ and $num.chains$. For each iteration i, the values in $probabilities$, S, and $num.iters$ are stored in a list:

```
  P_n <- P_n %*% P
}
return(list(probabilities, S, num.iters))
```

4.2 Simulating the data from Kilpisjärvi precipitation

The first simulation we do is with the 0-1-chain with the obtained data from Table 3.1. Firstly we create the matrix in R with the corresponding states "Dry" and "Wet" and set the name of the matrix to *transMatrix*:

```
states <- c("Dry","Wet")
transMatrix <- matrix(c(0.485, 0.515,0.219, 0.781),
                      byrow = TRUE,
                      nrow = 2,
                      dimnames = list(states, states))
transMatrix
```

```
##       Dry   Wet
## Dry 0.485 0.515
## Wet 0.219 0.781
```

Then we call for the function in figure (4.1) and run the chain:

```
sim2 <- sim.mark.chain(transMatrix)
```

For given number of iterations and number of chains declared in the function in Figure (4.1), the variable *sim*2 stores all values for the simulation of the inputmatrix *transMatrix*. All values in *sim*2 are returned by the function *sim.mark.chain*. The returned data in *sim*2 is devided into three variables storing the two matrices, *probabilities* and S, and the integer number returned from *num.iters*:

```
probabilities <- sim2[[1]]
states2       <- sim2[[2]]
n             <- sim2[[3]]
```

Next we manipulate the data so that it can be used in tables and charts:

```
library(data.table)

T_i             <- as.data.frame(states2[,1:5])
T_j             <- as.data.frame(lapply(T_i, table))


T_j_freq            <- T_j[,c(2,4,6,8,10)]
T_j_freqT           <- transpose(T_j_freq)

limOccProbSt1       <- transform(T_j_freqT[[1]], new = T_j_freqT[[1]] / n)

limOccProbSt2       <- transform(T_j_freqT[[2]], new = T_j_freqT[[2]] / n)

df                  <- cbind.data.frame(limOccProbSt1, limOccProbSt2)

df                  <- transform(df, mu1 = n / df[[1]], mu2 = n / df[[3]])

colnames(df)        <- c("$N_{1}(n)$","$N_{1}(n) / n$", "$N_{2}(n)$",
                         "$N_{2}(n) / n$", "$n/N_{1}(n)$", "$n/N_{2}(n)$")

rownames(df)        <- c("$M.c. 5$", "$M.c. 4$", "$M.c. 3$",
                         "$M.c. 2$", "$M.c. 1$")

df[,1:6]            <- round(df[,1:6], digit =3)

df                  <- df[, c(1,3,2,4,5,6)]
```

T_i stores the columns consisting five simulated markov chains as a dataframe data type. The library *data.table* is added to access the *table* function, for counting state occupancy. Frequencies are added to the dataframe T_j and then selected columns consisting only of the frequencies in the variable T_j_freq. Finally T_j_freq consisting of all the frequencies in each state, for each Markov chain, is manipulated to fit the measurements, and rounded up to three digits.

4.3 Results

	$N_1(n)$	$N_2(n)$	$N_1(n)/n$	$N_2(n)/n$	$n/N_1(n)$	$n/N_2(n)$
$M.c.1$	17	33	0.34	0.66	2.9412	1.5152
$M.c.2$	14	36	0.28	0.72	3.5714	1.3889
$M.c.3$	18	32	0.36	0.64	2.7778	1.5625
$M.c.4$	18	32	0.36	0.64	2.7778	1.5625
$M.c.5$	14	36	0.28	0.72	3.5714	1.3889

Figure 4.2: Table of measurements, $n = 50$.

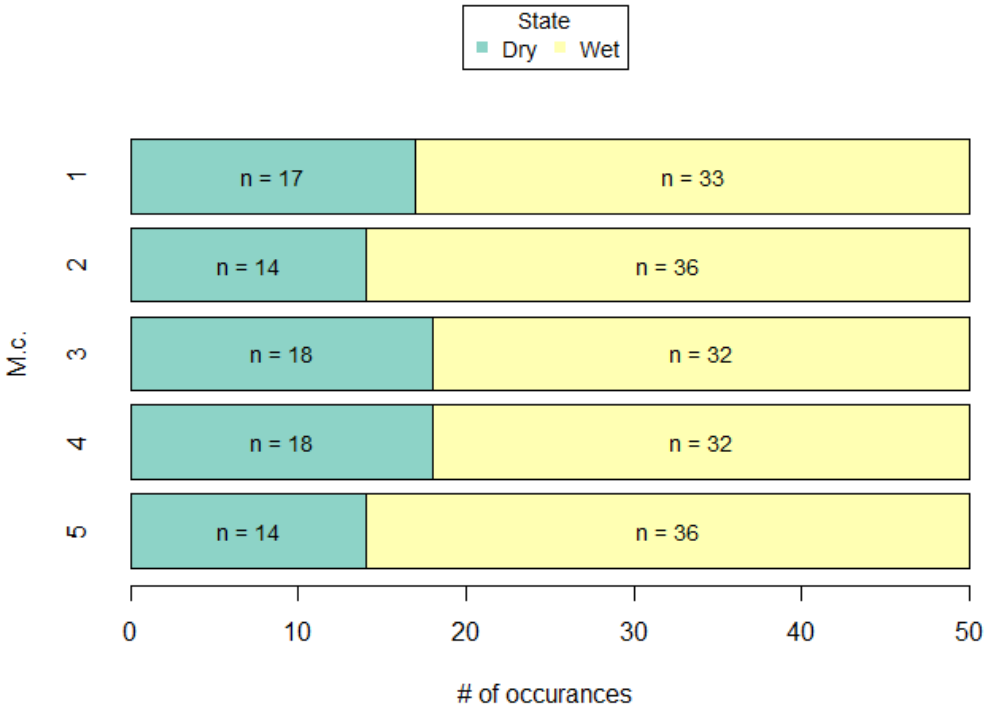

Figure 4.3: Occupation in each state for $n = 50$.

	$N_1(n)$	$N_2(n)$	$N_1(n)/n$	$N_2(n)/n$	$n/N_1(n)$	$n/N_2(n)$
$M.c.1$	173	327	0.346	0.654	2.8902	1.5291
$M.c.2$	141	359	0.282	0.718	3.5461	1.3928
$M.c.3$	139	361	0.278	0.722	3.5971	1.3850
$M.c.4$	144	356	0.288	0.712	3.4722	1.4045
$M.c.5$	151	349	0.302	0.698	3.3113	1.4327

Figure 4.4: Table of measurements, $n = 500$.

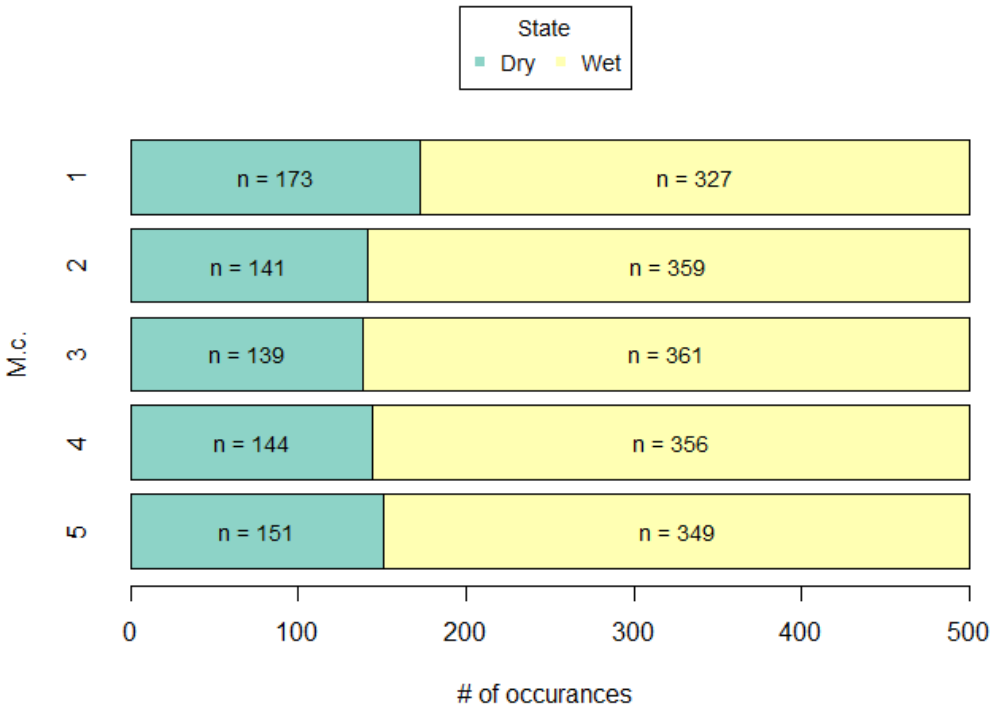

Figure 4.5: Occupation in each state for $n = 500$.

	$N_1(n)$	$N_2(n)$	$N_1(n)/n$	$N_2(n)/n$	$n/N_1(n)$	$n/N_2(n)$
$M.c.1$	1482	3518	0.2964	0.7036	3.3738	1.4213
$M.c.2$	1475	3525	0.2950	0.7050	3.3898	1.4184
$M.c.3$	1529	3471	0.3058	0.6942	3.2701	1.4405
$M.c.4$	1477	3523	0.2954	0.7046	3.3852	1.4192
$M.c.5$	1565	3435	0.3130	0.6870	3.1949	1.4556

Figure 4.6: Table of measurements, $n = 5000$.

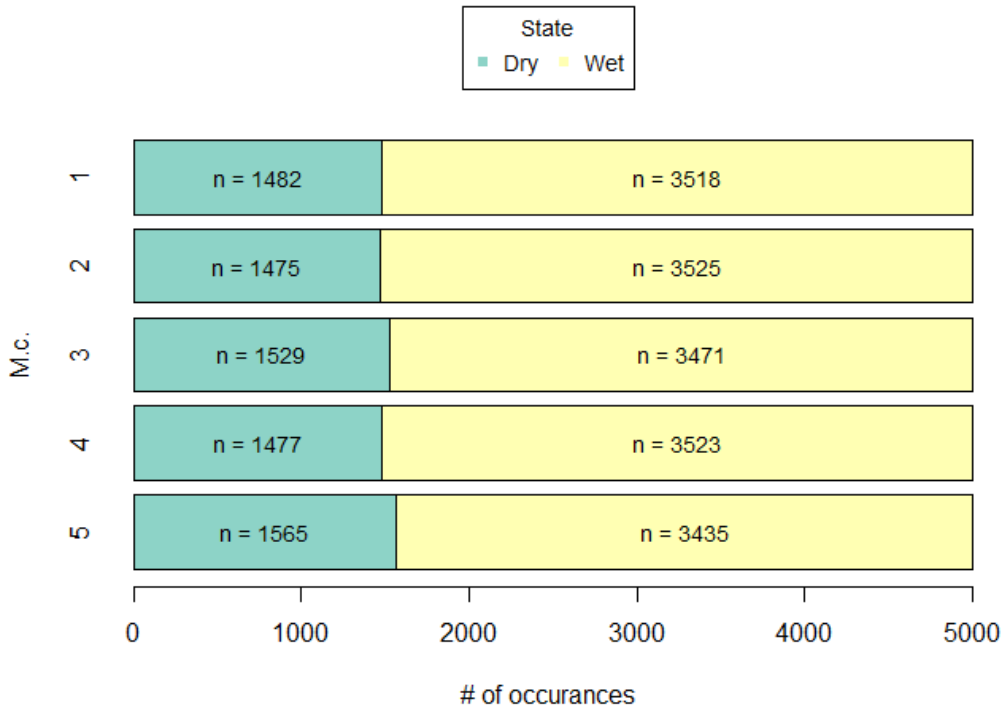

Figure 4.7: Occupation in each state for $n = 5000$.

	$N_1(n)$	$N_2(n)$	$N_1(n)/n$	$N_2(n)/n$	$n/N_1(n)$	$n/N_2(n)$
$M.c.1$	14934	35066	0.2987	0.7013	3.3481	1.4259
$M.c.2$	14773	35227	0.2955	0.7045	3.3846	1.4194
$M.c.3$	14978	35022	0.2996	0.7004	3.3382	1.4277
$M.c.4$	14901	35099	0.2980	0.7020	3.3555	1.4245
$M.c.5$	15172	34828	0.3034	0.6966	3.2955	1.4356

Figure 4.8: Table of measurements, $n = 50000$.

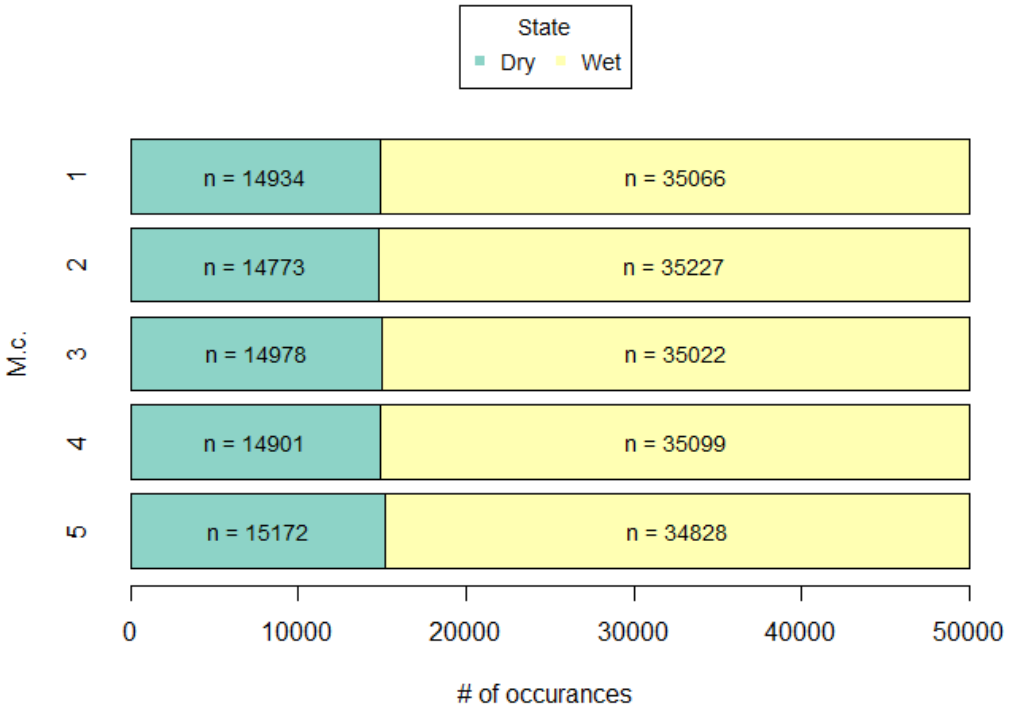

Figure 4.9: Occupation in each state for $n = 50000$.

	$\hat{\pi}_j(50)$	$\hat{\pi}_j(500)$	$\hat{\pi}_j(5.000)$	$\hat{\pi}_j(50.000)$	π_j
j = Dry	0.34	0.346	0.2964	0.2987	0.2984
j = Wet	0.66	0.654	0.7036	0.7013	0.7016

Figure 4.10: Simulated estimates from "M.c. 1" and the real value of the stationary distribution.

	$\hat{\mu}_j(50)$	$\hat{\mu}_j(500)$	$\hat{\mu}_j(5.000)$	$\hat{\mu}_j(50.000)$	μ_j
j = Dry	2.9412	2.8902	3.3738	3.3481	3.3516
j = Wet	1.5152	1.5291	1.4213	1.4259	1.4252

Figure 4.11: Simulated estimates from "M.c. 1" and the real value of the mean occurance time.

	$\hat{v}_j(50)$	$\hat{v}_j(500)$	$\hat{v}_j(5.000)$	$\hat{v}_j(50.000)$	v_j
j = Dry	0.5152	0.5291	0.4213	0.4259	0.4252
j = Wet	1.9412	1.8902	2.3738	2.3481	2.3516

Figure 4.12: Simulated estimates from "M.c. 1" and the real value of v_j.

4.4 Simulation of the 7×7 matrix P

The algorithm in the following simulations is the same as in Figure (4.1) and the coding process in R follows a similar methodology as in Section (4.2). The code for the methodology and the code for the tables and charts is added to appendices. Suppose we have the 7×7 matrix

$$P = \begin{pmatrix} 0.25 & 0.5 & 0.05 & 0.05 & 0.05 & 0.05 & 0.05 \\ 0.2 & 0.25 & 0.35 & 0.05 & 0.05 & 0.05 & 0.05 \\ 0.05 & 0.2 & 0.25 & 0.35 & 0.05 & 0.05 & 0.05 \\ 0 & 0 & 0.25 & 0.5 & 0.25 & 0 & 0 \\ 0.05 & 0.05 & 0.05 & 0.2 & 0.25 & 0.35 & 0.05 \\ 0.05 & 0.05 & 0.05 & 0.05 & 0.2 & 0.25 & 0.35 \\ 0.05 & 0.05 & 0.05 & 0.05 & 0.05 & 0.5 & 0.25 \end{pmatrix},$$

and simulate the matrix P. We simulate matrix P five times, with $50, 5000$, and 500000 numbers of transitions, n. The five simulated chains are named "M.c. 1", "M.c. 2', ... , "M.c. 5". The main goal is to see how the number of occurances, $N_j(n)$, for each $j \in S$, changes through time.

4.5 Results

	$N_1(n)$	$N_2(n)$	$N_3(n)$	$N_4(n)$	$N_5(n)$	$N_6(n)$	$N_7(n)$
$M.c.1$	3	5	14	17	7	3	1
$M.c.2$	4	10	8	7	5	9	7
$M.c.3$	1	2	7	18	10	9	3
$M.c.4$	4	2	2	9	8	14	11
$M.c.5$	3	3	12	8	7	10	7

	$N_1(n)/n$	$N_2(n)/n$	$N_3(n)/n$	$N_4(n)/n$	$N_5(n)/n$	$N_6(n)/n$	$N_7(n)/n$
$M.c.1$	0.06	0.10	0.28	0.34	0.14	0.06	0.02
$M.c.2$	0.08	0.20	0.16	0.14	0.10	0.18	0.14
$M.c.3$	0.02	0.04	0.14	0.36	0.20	0.18	0.06
$M.c.4$	0.08	0.04	0.04	0.18	0.16	0.28	0.22
$M.c.5$	0.06	0.06	0.24	0.16	0.14	0.20	0.14

	$n/N_1(n)$	$n/N_2(n)$	$n/N_3(n)$	$n/N_4(n)$	$n/N_5(n)$	$n/N_6(n)$	$n/N_7(n)$
$M.c.1$	16.6667	10.0000	3.5714	2.9412	7.1429	16.6667	50.0000
$M.c.2$	12.5000	5.0000	6.2500	7.1429	10.0000	5.5556	7.1429
$M.c.3$	50.0000	25.0000	7.1429	2.7778	5.0000	5.5556	16.6667
$M.c.4$	12.5000	25.0000	25.0000	5.5556	6.2500	3.5714	4.5455
$M.c.5$	16.6667	16.6667	4.1667	6.2500	7.1429	5.0000	7.1429

Figure 4.13: Table of measurements, $n = 50$.

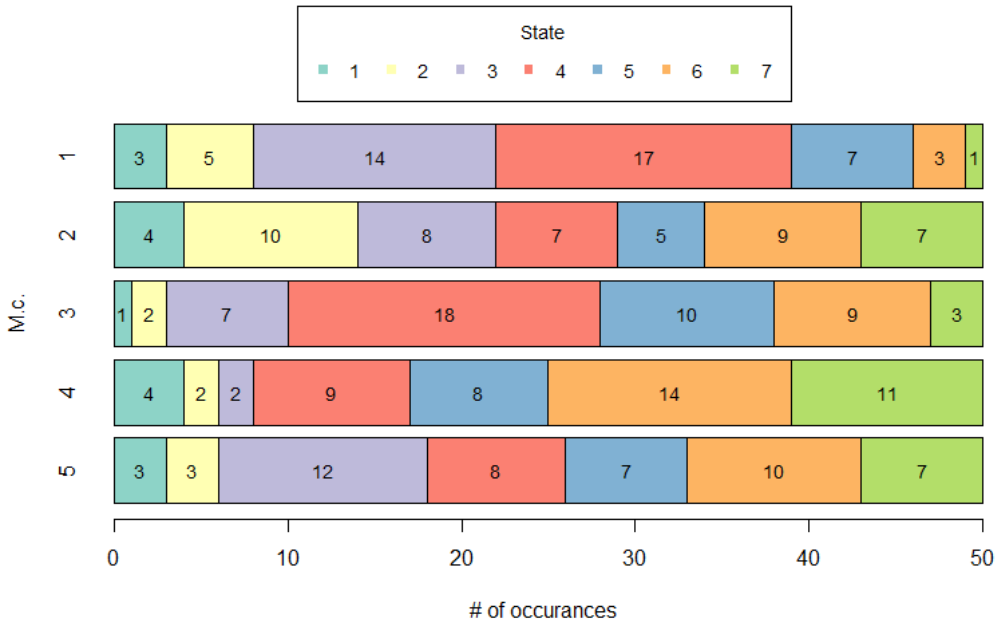

Figure 4.14: Occupation in each state for $n = 50$.

	$N_1(n)$	$N_2(n)$	$N_3(n)$	$N_4(n)$	$N_5(n)$	$N_6(n)$	$N_7(n)$
$M.c.1$	309	525	761	1141	810	866	588
$M.c.2$	365	618	791	1080	729	859	558
$M.c.3$	371	628	847	1108	686	821	539
$M.c.4$	361	607	805	1084	750	826	567
$M.c.5$	331	653	797	1025	711	891	592

	$N_1(n)/n$	$N_2(n)/n$	$N_3(n)/n$	$N_4(n)/n$	$N_5(n)/n$	$N_6(n)/n$	$N_7(n)/n$
$M.c.1$	0.0618	0.1050	0.1522	0.2282	0.1620	0.1732	0.1176
$M.c.2$	0.0730	0.1236	0.1582	0.2160	0.1458	0.1718	0.1116
$M.c.3$	0.0742	0.1256	0.1694	0.2216	0.1372	0.1642	0.1078
$M.c.4$	0.0722	0.1214	0.1610	0.2168	0.1500	0.1652	0.1134
$M.c.5$	0.0662	0.1306	0.1594	0.2050	0.1422	0.1782	0.1184

	$n/N_1(n)$	$n/N_2(n)$	$n/N_3(n)$	$n/N_4(n)$	$n/N_5(n)$	$n/N_6(n)$	$n/N_7(n)$
$M.c.1$	16.1812	9.5238	6.5703	4.3821	6.1728	5.7737	8.5034
$M.c.2$	13.6986	8.0906	6.3211	4.6296	6.8587	5.8207	8.9606
$M.c.3$	13.4771	7.9618	5.9032	4.5126	7.2886	6.0901	9.2764
$M.c.4$	13.8504	8.2372	6.2112	4.6125	6.6667	6.0533	8.8183
$M.c.5$	15.1057	7.6570	6.2735	4.8780	7.0323	5.6117	8.4459

Figure 4.15: Table of measurements, $n = 5000$.

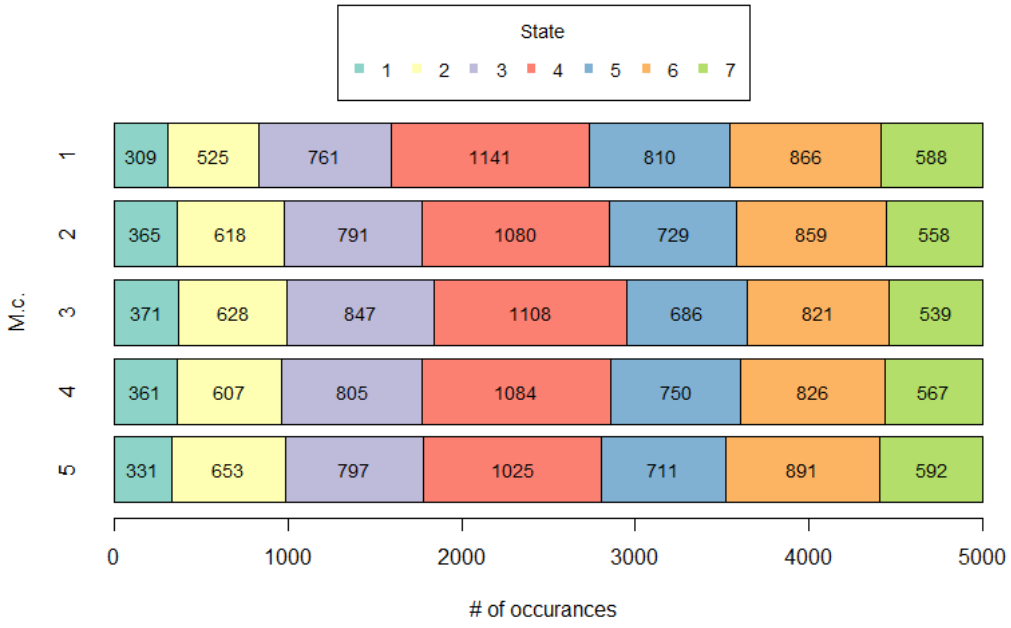

Figure 4.16: Occupation in each state for $n = 5000$.

	$N_1(n)$	$N_2(n)$	$N_3(n)$	$N_4(n)$	$N_5(n)$	$N_6(n)$	$N_7(n)$
$M.c.1$	35437	59185	81324	110104	74489	83797	55664
$M.c.2$	35243	59540	81292	110362	73876	83917	55770
$M.c.3$	35800	60091	80760	109239	74547	83740	55823
$M.c.4$	35536	59930	81329	110290	74272	83257	55386
$M.c.5$	35710	59777	81313	109876	74143	83470	55711

	$N_1(n)/n$	$N_2(n)/n$	$N_3(n)/n$	$N_4(n)/n$	$N_5(n)/n$	$N_6(n)/n$	$N_7(n)/n$
$M.c.1$	0.0709	0.1184	0.1626	0.2202	0.1490	0.1676	0.1113
$M.c.2$	0.0705	0.1191	0.1626	0.2207	0.1478	0.1678	0.1115
$M.c.3$	0.0716	0.1202	0.1615	0.2185	0.1491	0.1675	0.1116
$M.c.4$	0.0711	0.1199	0.1627	0.2206	0.1485	0.1665	0.1108
$M.c.5$	0.0714	0.1196	0.1626	0.2198	0.1483	0.1669	0.1114

	$n/N_1(n)$	$n/N_2(n)$	$n/N_3(n)$	$n/N_4(n)$	$n/N_5(n)$	$n/N_6(n)$	$n/N_7(n)$
$M.c.1$	14.1095	8.4481	6.1482	4.5412	6.7124	5.9668	8.9825
$M.c.2$	14.1872	8.3977	6.1507	4.5305	6.7681	5.9583	8.9654
$M.c.3$	13.9665	8.3207	6.1912	4.5771	6.7072	5.9709	8.9569
$M.c.4$	14.0702	8.3431	6.1479	4.5335	6.7320	6.0055	9.0276
$M.c.5$	14.0017	8.3644	6.1491	4.5506	6.7437	5.9902	8.9749

Figure 4.17: Table of measurements, $n = 500000$.

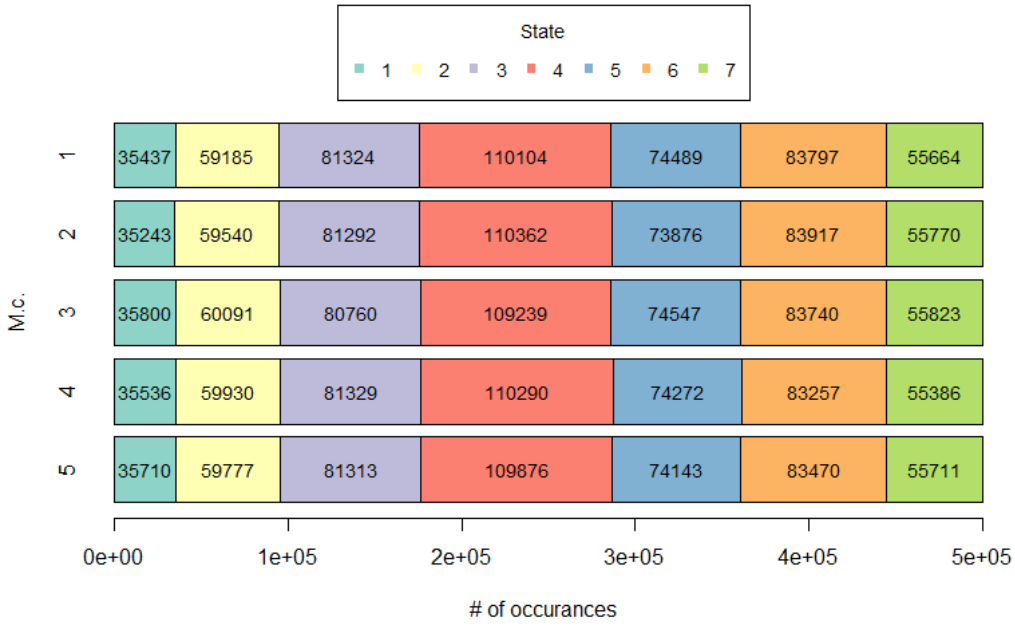

Figure 4.18: Occupation in each state for $n = 500000$.

	$\hat{\pi}_j(50)$	$\hat{\pi}_j(5.000)$	$\hat{\pi}_j(500.000)$	π_j
j = 1	0.06	0.0618	0.0709	0.0711
j = 2	0.10	0.1050	0.1184	0.1191
j = 3	0.28	0.1522	0.1626	0.1622
j = 4	0.34	0.2282	0.2202	0.2200
j = 5	0.14	0.1620	0.1490	0.1489
j = 6	0.06	0.1732	0.1676	0.1673
j = 7	0.02	0.1176	0.1113	0.1115

Figure 4.19: Simulated estimates from "M.c. 1" and the real value of the stationary distribution.

	$\hat{\mu}_j(50)$	$\hat{\mu}_j(5.000)$	$\hat{\mu}_j(500.000)$	μ_j
j = 1	16.6667	16.1812	14.1095	14.0661
j = 2	10.0000	9.5238	8.4481	8.3928
j = 3	3.5714	6.5703	6.1482	6.1663
j = 4	2.9412	4.3821	4.5412	4.5462
j = 5	7.1429	6.1728	6.7124	6.7179
j = 6	16.6667	5.7737	5.9668	5.9779
j = 7	50.0000	8.5034	8.9825	8.9700

Figure 4.20: Simulated estimates from "M.c. 1" and the real value of the mean occurance time.

	$\hat{v}_j(50)$	$\hat{v}_j(5.000)$	$\hat{v}_j(500.000)$	v_j
j = 1	0.0638	0.0659	0.0763	0.0765
j = 2	0.1111	0.1173	0.1343	0.1353
j = 3	0.3889	0.1795	0.1942	0.1936
j = 4	0.5152	0.2957	0.2824	0.2820
j = 5	0.1628	0.1933	0.1751	0.1749
j = 6	0.0638	0.2095	0.2013	0.2009
j = 7	0.0204	0.1333	0.1253	0.1255

Figure 4.21: Simulated estimates from "M.c. 1" and the real value of v_j.

Chapter 5

Conclusions

5.1 Convergence of the Kilpisjärvi simulation

The results in Figure 4.10 show how the estimate of π for 50000 transitions is the best one of the estimates, since it is closest to the real value. 50 and 500 numbers of transitions seem to be too few to get a good estimate, since 50 transitions actually is closer to the real value of π. As I mentioned before, these values are from the first simulation (M.c. 1), so the estimates from $\hat{\pi}_j(50)$ and $\hat{\pi}_j(500)$ may be extreme. Nevertheless, due to randomness, we would need more than 500 transitions to ensure a good estimate. Even though the simulated chain goes in wrong direction, in terms of convergence comparing $n = 50$ and $n = 500$, the local divergences vanish in the long run. That explains why the same simulated chain for $n = 50000$ only deviates 0.0003 from the real value of π.

As $\hat{\pi}_j(n)$ converges to π_j, $\hat{\mu}_j(n)$ converges to μ_j, and $\hat{v}_j(n)$ converges to v_j, since π, μ, and v are depending on each other. For $\mu_j(50000)$, we can expect a dry day on average every 3.3481:st time, and a wet day every 1.4259:th time. The estimate is close to the real mean occurance time, which has a value of $(3.3516, 1.4252)$. According to $\hat{v}_j(50000)$, we can expect a dry day every 0.4259:th time, between two successive wet days. Similarly we can expect 2.3481 wet days between two succesive dry days. The estimate is close to the real value of v_j , which is $(0.4252, 2.3516)$.

5.2 Convergence of the matrix P

Each of the chosen discrete time points (n) for the simulation of matrix P, generated results closer to the stationary distribution, except for $j = 5$. I interpret these results so that if the time points are far enough from each other, each larger time point is closer to the value of the stationary distribution. The magnitude between each of the time point is

In Figure 4.19 we see that for 50 transitions ($n = 50$), we are in most of the states far from the stationary distribution, i.e. it would not be a good estimate for the stationary distribution. For $n = 5000$, we get a correct distribution of two decimals in all states, except for $j = 4$. For $n = 500000$ we get a correct distribution of two decimals in all states and a correct distribution of three decimals for $j = 1, 2, 4, 5$. I interpret the results of the estimates, in relation to the stationary distribution, that an estimate with a lower amount of transitions can be accepted, if we allow some variations. A question that arises is how large the number of transitions should be to be a valid estimate for the stationary distribution. I think, according to the simulation results in Figure 4.19, that an n-value somewhere between 5000 and 500000 ensures a convergence, with two decimals, to the stationary distribution. Nevertheless, since the thesis is limited, I have not studied how large n should be to ensure convergence of a certain decimal accuracy.

5.3 Visualization of convergence

I found the results of the horizontal barcharts very intriguing, since they show convergence in a way that is quite easy to understand. Both of my two different matrix-simulations follow a similar pattern visually. A smaller number of transitions results in large variation between the five simulations. When n grows bigger, the boxes of $M.c.1, ..., M.c.5$ are stacked more similarly. Finally, when we take the highest simulated value of n, $M.c.1, ..., M.c.5$ are stacked almost identically, which is an indication that they all converge.

5.4 Summary

Based on the results I got, it seems as if there are many things that impact how accurate results we get. Firstly, if we have fewer states, the chain converges faster to the stationary distribution. Secondly, the number of

transitions impacts how much variation we have for different simulations. A smaller number of transitions can result in fluctuations, due to randomness, in a way that the chain does not seem to converge. However, for both of the two simulated matrices, for a large enough n, they both converge.

Chapter 6

Summary in Swedish - Svensk sammanfattning

Målet med denna avhandling är att presentera den grundläggande teorin för diskreta Markovkedjor och jämföra teoretiska resultat med egna simuleringar. Tanken är att visa hur Markovkedjor, med vissa egenskaper, beter sig under en lång tidsperiod och att visualisera konvergensen mot Markovkedjans stationära fördelningen. Kapitel 4 är avhandlingens viktigaste del, där jag simulerar två olika ergodiska Markovkedjor och visualiserar dessa i tabeller och diagram. För en djupare förståelse av innehållet i avhandlingen antas läsaren ha grundläggande kunskap i matematik och programmering.

Avhandlingens första helhet är en introduktion om dess innehåll och delmoment och där berättar jag även om vad en Markovkedja är. Jag blev intresserad av att skriva om Markovkedjor efter att ha deltagit i en kurs om ämnet vid Åbo Akademi. Tillämpning av matematik intresserar mig mer än teoretisk matematik, varför valet av Markovkedjor föll naturligt, eftersom de har stora praktiska tillämpningsområden.

Kapitel 2 handlar om introducerande teori till Markovkedjor. Till att börja med tas det upp grundläggande teori om matriser och gränsvärdessatser samt teori från sannolikhetsläran. Denna grundläggande teori i kapitel 2 fungerar som byggsten för att introducera och utveckla innehållet när det gäller Markovkedjor. I den andra helheten behandlar jag även en metod som heter Gauss eliminering, för numerisk beräkning av Markovkedjans stationära fördelning.

Kapitel 3 börjar med en allmän definition av en Markovkedja, samt tillämpning av Markovkedjan i form av empiriskt data från det finska me-

teorologiska institutet (FMI). Teorin för Markovkedjor byggs på för att föra läsaren mot avhandlingens kärna, den stationära fördelningen. Avhandlingens empiriska data härstammar från mätningar i Kilpisjärvi i norra Finland och är tagen under en tidsperiod av 30 år. Data har jag manipulerat i mjukvaroprogrammet Microsoft Excel, där jag använder min manipulerade data för att räkna ut en övergångsmatris, som används för teoriexempel och simuleringar.

I kapitel 4 utförs simuleringar, i programmeringsspråket R, av två olika matriser med flera olika tidsintervall. Den ena matrisen innehåller empirisk data från FMI och den andra matrisen, som är större i dimension, har jag skapat med de egenskaper som krävs för en ergodisk Markovkedja. I kapitel 4 presenteras även den algoritm som utförs för varje simulering och visualiseras i såväl tabeller som diagram. All kod för tabeller och diagram presenteras i slutet av avhandlingen som appendix.

I kapitel 5 diskuteras resultaten från simuleringarna i kapitel 4. Vid låga värden på n uppstod det stora variationer i båda simuleringarna, vilket gav upphov till att ett större n-värde, vid ett förhållandevis lågt n-värde, kunde resultera i en sämre approximation av den stationära fördelningen. Det visade sig ändå i mina tabeller, när n-värdet blev tillräckligt stort, att de simulerade kedjorna konvergerade mot den stationära fördelningen. De tillfälliga variationerna, på grund av slumpmässighet, hade således mindre betydelse för stort n-värde. En annan sak som jag tar fasta på utgående från resultaten är att storleken på en matris, eller närmare bestämt antalet tillstånd, verkar ha stor betydelse för hur snabbt Markovkedjan konvergerar mot sin stationära fördelning. När jag simulerade data från Kilpisjärvi gav ett övergångsvärde på 50000 ett resultat som var väldigt nära den stationära fördelningen. För den andra simuleringen, av matrisen P, gav ett värde på 500000 ungefär samma precision. Jag kan konstatera att båda mina simuleringar, innehållande slumpmässighet, konvergerar mot deras stationära fördelning och motsvarar därmed väl teorin för diskreta och ergodiska Markovkedjor.

Jag funderade över hur jag kunde visualisera konvergensen mot den stationära fördelningen och kom på ett sätt som jag inte har stött på tidigare. Jag kodade horisontella stapeldiagram över antalet träffar i olika tillstånd, för olika antalet övergångar. Därmed visualiseras antalet träffar i relation till antalet övergångar, vilket i min mening ger ett snyggt intryck. Stapeldiagramens värden över antalet träffar motsvarar de värden som finns i mina tabeller, men dessa visualiseringar är betydelsefulla för att enklare se hur

simulationer beter sig under olika och längre tidsperioder. För lågt värde på n var det stora variationer i stapeldiagrammen. När n växte verkade alla fem simulationer se mer och mer lika ut visuellt och till sist för det största värdet, var alla fem simulationer nästan visuellt identiska. Precis som i mina tabeller över simulerade resultat, är detta en indikation på konvergens.

Största delen av teorin för Markovkedjor har hämtats från boken *stochastic modelling of scientific data*, Guttorp, P. (1995), som innehåller uttömmande teoritiska resultat. Algoritmen, som används i simuleringarna i kapitel 4, bygger på en internetpublikation av Bonakdarpour M. (2016). För flera intressanta resultat, kunde simuleringarna utvecklas till att omfatta icke ergodiska och icke reducerbara Markovkedjor. Resultaten från simuleringarna kunde även verifieras, ur ett statistiskt perspektiv, genom att öka antalet gånger kedjorna simuleras. Eftersom avhandlingen är begränsad, har jag valt att avgränsa simulationerna till två olika ergodiska Markovkedjor.

Appendix A

R code for chapter 4

A.1 Code for the plot of probability through time for the Kilpisjärvi simulation for $n = 50$.

```
matplot(probabilities, type='l', col=1:2, lty=1, ylab='probability', xlab='time')
legend('topright', c('Dry', 'Wet'), lty=1, col=1:2, cex=0.8)
```

A.2 Code for the barplot from the Kilpisjärvi simulation

```
library(RColorBrewer)

state.color <- brewer.pal(7, "Set3")

freq_matrix <- as.matrix(T_j_freq)

#get the horizontal position in the boxes for the number of occurances

horPos <- apply(freq_matrix, 2L, cumsum)
horPos <- horPos - freq_matrix / 2

#stacked bar plot

freq.plot <- barplot(freq_matrix,
                     horiz = TRUE,
                     xlab = "# of occurances",
                     ylab = "M.c.",
                     names.arg = c("5", "4", "3", "2", "1"),
                     col=state.color,
                     legend = rownames(freq_matrix))

text(horPos, rep(freq.plot, each = nrow(horPos)), labels = paste("n = ", freq_matrix, sep="") ,cex=0.9)
```

By changing the value of *num.iters* in the algorithm in Figure (4.1), we get the barplot for $n = 500, n = 5000$, and $n = 50000$.

A.3 Code for the 7×7 matrix.

```
P <- matrix(c(0.25, 0.5 , 0.05, 0.05, 0.05, 0.05, 0.05,
              0.2 , 0.25, 0.35, 0.05, 0.05, 0.05, 0.05,
              0.05, 0.2 , 0.25, 0.35, 0.05, 0.05, 0.05,
              0   , 0   , 0.25, 0.5 , 0.25, 0   , 0   ,
              0.05, 0.05, 0.05, 0.2 , 0.25, 0.35, 0.05,
              0.05, 0.05, 0.05, 0.05, 0.2 , 0.25, 0.35,
              0.05, 0.05, 0.05, 0.05, 0.05, 0.5 , 0.25
              ),
              byrow = TRUE,
              nrow = 7, )

sim.seven     <- sim.mark.chain(P)
probabilities <- sim.seven[[1]]
states.seven  <- sim.seven[[2]]
n             <- sim.seven[[3]]
```

sim.seven saves all values returned from the algorithm in Figure (4.1).

A.4 Datamanipulation and code for table of measurements in Figure (4.13).

```
T_i_7             <- as.data.frame(states.seven[,1:5])
Tj.seven          <- as.data.frame(lapply(T_i_7, table))

library(data.table)

Tj.seven_freq     <- Tj.seven[,c(2,4,6,8,10)]
Tj.seven_freqT    <- transpose(Tj.seven_freq)

############## Make table of valuable measurements ######################

limOccProbSt          <- cbind.data.frame(Tj.seven_freq[[1]],
                          new1 = Tj.seven_freq[[1]] / n,
                          mu1 = n / Tj.seven_freq[[1]],
                          Tj.seven_freq[[2]],
                          new2 = Tj.seven_freq[[2]] / n,
                          mu2 = n / Tj.seven_freq[[2]],
                          Tj.seven_freq[[3]],
                          new3 = Tj.seven_freq[[3]] / n,
                          mu3 = n / Tj.seven_freq[[3]],
                          Tj.seven_freq[[4]],
                          new4 = Tj.seven_freq[[4]] / n,
                          mu4 = n / Tj.seven_freq[[4]],
                          Tj.seven_freq[[5]],
                          new5 = Tj.seven_freq[[5]] / n,
                          mu5 = n / Tj.seven_freq[[5]])
```

```
r1                      <- c(t(limOccProbSt[[1]]),t(limOccProbSt[[2]]),t(limOccProbSt[[3]]))
r2                      <- c(t(limOccProbSt[[4]]),t(limOccProbSt[[5]]),t(limOccProbSt[[6]]))
r3                      <- c(t(limOccProbSt[[7]]),t(limOccProbSt[[8]]),t(limOccProbSt[[9]]))
r4                      <- c(t(limOccProbSt[[10]]),t(limOccProbSt[[11]]),t(limOccProbSt[[12]]))
r5                      <- c(t(limOccProbSt[[13]]),t(limOccProbSt[[14]]),t(limOccProbSt[[15]]))

limOccProbSt            <- rbind.data.frame(r1,r2,r3,r4,r5)

colnames(limOccProbSt) <- c("$N_{1}(n)$", "$N_{2}(n)$", "$N_{3}(n)$", "$N_{4}(n)$",
                            "$N_{5}(n)$", "$N_{6}(n)$", "$N_{7}(n)$",

                            "$N_{1}(n) / n$", "$N_{2}(n) / n$", "$N_{3}(n) / n$",
                            "$N_{4}(n) / n$", "$N_{5}(n) / n$", "$N_{6}(n) / n$",
                            "$N_{7}(n) / n$",

                            "$n/N_{1}(n)$", "$n/N_{2}(n)$", "$n/N_{3}(n)$", "$n/N_{4}(n)$",
                            "$n/N_{5}(n)$", "$n/N_{6}(n)$", "$n/N_{7}(n)$")

rownames(limOccProbSt) <- c("$M.c. 5$", "$M.c. 4$", "$M.c. 3$",
                            "$M.c. 2$", "$M.c. 1$")
limOccProbSt[,1:21]     <- round(limOccProbSt[,1:21], digit =4)

#reorder rows to be in same order as plot

limOccProbSt            <- limOccProbSt[c(5,4,3,2,1),]
Occ_7                   <- limOccProbSt[,1:7]
numb_hits_ave           <- limOccProbSt[,8:14]
visits                  <- limOccProbSt[,(15:21)]

library(knitr)
knitr::kable(Occ_7, escape = FALSE)
```

The first argument passed into *kable*(), from library *knitr*, declares what variable we plot. The variables $Occ\_7, numb\_hits\_ave$, and *visits*, consist of values in Figure (4.13).

A.5 Barplot code for the 7×7 matrix.

```
library(RColorBrewer)

state.color <- brewer.pal(7, "Set3")
y.mat       <- as.matrix(Tj.seven_freq)

#get the horizontal position in the boxes for the number of occurances

H           <- apply(y.mat, 2L, cumsum)
H           <- H - y.mat /2

y.plot      <- barplot(y.mat,
                       horiz = TRUE,
                       xlab = "# of occurances",
                       ylab = "M.c.",
                       names.arg = c("5", "4", "3", "2", "1"),
                       col=state.color,
                       legend = rownames(y.mat))

text(H, rep(y.plot, each = nrow(H)), labels = paste(y.mat) ,cex=0.9)
```

Bibliography

[1] Bhattacharya, Rabi N.; Waymire, Edward C.: *Stochastic Processes with Applications*, Society for Industrial and Applied Mathematics 2009

[2] Björkfelt, I.,Bjon, S.: *Matriser*, Matematik och statistik, Åbo Akademi 2003

[3] Bonakdarpour, M.: *Simulating Discrete Markov Chains: Limiting Distri butions*, Github 2016, Last accessed 2:nd Mar 2020: https://stephens999.github.io/fiveMinuteStats/simulating_discrete_chains_2.html

[4] Chen, Y.: *Discrete-Time Markov Chain – Part I*, Washington edu 2018, Last accessed 14:th May 2021: http://faculty.washington.edu/yenchic/18A_stat516/Lec3_DTMC_p1.pdf

[5] Guttorp, P.: *STOCHASTIC MODELLING OF SCIENTIFIC DATA*, Chapman&Hall/CRC 1995

[6] Ross, Sheldon M.: *Introduction to PROBABILITY MODELS - 4th edition*, Academic Press, Inc. 1989

[7] Ross, Sheldon M.: *Introduction to Probability models - 10th edition*, Elsevier 2010

[8] Rothman, A.: *Proof of the Law of Large Numbers Part 2: The Strong Law*, Towardsdatascience 2020, Last accessed 14:th May 2021: https://towardsdatascience.com/proof-of-the-law-of-large-numbers-part-2-the-strong-law-356aa608ca5d

[9] Salminen, P.: *Sannolikhetslära*, Matematik och statistik, Åbo Akademi 2010

[10] The Finnish meteorological institute (FMI) - Enontekiö Kilpisjärvi kyläkeskus: https://cdn.fmi.fi/fmiodata-convert-api/preview/21f718f7-97a6-440a-b6e6-c9ca89973716/?locale=en

[11] Zulaikha, L.: *Introduction To Markov Chains With Examples-Markov Chains With Python*, Edureka 2019, Last accessed 16:th Apr 2020: https://www.edureka.co/blog/introduction-to-markov-chains/