Stochastic Modeling
Christel Geiss and Stefan Geiss

Department of Mathematics and Statistics
University of Jyväskylä

September 18, 2020
Contents

1 Introduction .................................................. 7
   1.1 Mathematical modeling ................................... 7
       1.1.1 First example: Meteorology ......................... 7
       1.1.2 Second example: Financial Mathematics .......... 9
       1.1.3 More general: Random and non-random models .... 9
   1.2 Some basic concepts of stochastics ......................... 11
       1.2.1 The probability space \((\Omega, \mathcal{F}, \mathbb{P})\) .... 11
       1.2.2 Conditional probability and independence .......... 13
       1.2.3 Random variables .................................. 13
       1.2.4 Expectation ....................................... 18

2 Random walk .................................................. 21
   2.1 The probability space \((\mathbb{D}_n, \mathcal{F}^{dyad}_n, \mu_n^{(p)})\) ... 21
   2.2 The discrete time geometric Brownian motion .......... 23
   2.3 The Cox-Ross-Rubinstein model for stock prices ....... 24
   2.4 Snoqualmie Falls precipitation .......................... 25

3 Discrete time Markov chains ................................ 27
   3.1 Introduction ............................................. 28
   3.2 The Chapman - Kolmogorov equations ..................... 33
       3.2.1 Modeling of Snoqualmie Falls precipitation with a
             Markov chain ....................................... 37
   3.3 Classification of states .................................. 39
       3.3.1 Absorbing states .................................. 39
       3.3.2 Absorbing states for branching processes ......... 39
       3.3.3 Communicating states ............................... 49
       3.3.4 Reid-Landau model for radiation damage ........... 51
       3.3.5 Periodic and aperiodic states ..................... 56
CONTENTS

3.3.6 Persistent and transient states ........................................ 58
3.3.7 Decomposition of the state space ...................................... 62
3.3.8 Summary of the classification ........................................... 64
3.4 Ergodic Theorem and stationary distribution .......................... 69

4 MCMC methods .................................................................. 77
4.1 The classical Monte Carlo method ........................................ 77
4.2 General idea of MCMC .......................................................... 78
4.3 Basic example for the classical Monte Carlo method ............... 78
4.4 The Gibbs sampler ............................................................... 79
4.5 Burn-in period for MCMC methods ........................................ 81
4.6 The hard-core model ............................................................ 82
Week 1  16.03. – 20.03.2020

What is the aim of the first week, what is the main goal to reach?

- In Section 1.1 some basic ideas of mathematical modeling are explained. This part is - in a sense - not mathematically strict, but important to understand the general approach.

- The main topic for this week consists in Section 1.2. It is absolutely necessary that you go line-by-line carefully through these pages. We have to clarify all open questions by email/skype etc.
Chapter 1

Introduction

1.1 Mathematical modeling

In order to do mathematical modeling we need first some information: knowledge of natural laws, economical or social laws ... as well as scientific data, i.e.

- numerical data describing real phenomena,
- data collected according to certain strict rules.

1.1.1 First example: Meteorology

We start with an example from [2], the observation of the weather at Snoqualmie Falls (Washington, U.S.A.) in the month January between the years 1948-1983:

Rules

- 1 day = 8 am - 8 am (following day)
- day is wet $\iff$ at least 0.01 inches ($\approx 0.0254$ cm) of precipitation (=snow, rain, ...)
- day is dry $\iff$ $< 0.01$ inches of precipitation

Data $(X_{ij})_{i=1, j=1948}^{31,1983}$, where

$$X_{ij} = \begin{cases} 
0 & \text{if day } i \text{ of year } j \text{ is dry} \\
1 & \text{if day } i \text{ of year } j \text{ is wet}
\end{cases}.$$
In the picture below, which shows the data taken from [2], a day with precipitation is a black square:
1.1. MATHEMATICAL MODELING

1.1.2 Second example: Financial Mathematics

The price of one share of NOKIA over the time period of one year:

**Rules**

- The price is taken in Euro.
- The price is taken on all trading days at 12:00.

**Data** \((X_1, \ldots, X_N)\), where \(X_i\) is the price at the \(i\)-th trading day in Euro.

1.1.3 More general: Random and non-random models

\[
\begin{array}{c}
\text{mathematical models} \\
\text{deterministic models} \\
\text{probabilistic models}
\end{array}
\]

Differential equations: Kepler’s laws of planet movement
Stochastic processes: share prices

**Sources of random behavior:**

- Sensitivity to or randomness of initial conditions.
- Systems with many parameters, that are partially unknown (incomplete information) and complex dependency structures.
- Fundamental description: impulse and position in modern quantum theory.
CHAPTER 1. INTRODUCTION

- data
- mathematical model including unknown parameters
- fitting
- 'concrete' mathematical model
- describing the process under consideration
  - predict future events
1.2 Some basic concepts of stochastics

We shortly recall the needed facts from probability theory. For more information see, for example [1] or [4].

1.2.1 The probability space \((\Omega, \mathcal{F}, \mathbb{P})\)

The set \(\Omega\), which we also call set of elementary events, is a set which we sometimes assume to be:

- Finite: \(\Omega = \{\omega_1, \ldots, \omega_n\}\)
- Countable: \(\Omega = \{\omega_1, \omega_2, \ldots\}\)
- Uncountable: \(\Omega = \{\omega : \omega \in [0, 1]\}\)

The \(\sigma\)-algebra (or \(\sigma\)-field) is a basic tool in probability theory. It contains the sets where the probability measure is defined on.

**Definition 1.2.1. [\(\sigma\) algebra]**

Let \(\Omega\) be a non-empty set. A system \(\mathcal{F}\) of subsets \(A \subseteq \Omega\) is a \(\sigma\)-algebra on \(\Omega\) if

1. \(\emptyset, \Omega \in \mathcal{F}\),
2. \(A \in \mathcal{F}\) implies that \(A^c := \Omega \setminus A \in \mathcal{F}\),
3. \(A_1, A_2, \ldots \in \mathcal{F}\) implies that \(\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}\).

A set \(A \in \mathcal{F}\) is called event.

**Remark 1.2.2.** If \(\mathcal{F}\) is a \(\sigma\)-algebra, then \(A_1, A_2, \ldots \in \mathcal{F}\) implies that \(\bigcap_{i=1}^{\infty} A_i \in \mathcal{F}\).

We consider some first examples of \(\sigma\) algebras.

**Example 1.2.3.**

(a) The largest \(\sigma\) algebra on \(\Omega\): if \(\mathcal{F} = 2^\Omega\) is the system of all subsets \(A \subseteq \Omega\), then \(\mathcal{F}\) is a \(\sigma\)-algebra.

(b) The smallest \(\sigma\) algebra: \(\mathcal{F} = \{\Omega, \emptyset\}\).

(c) Let \(A \subseteq \Omega\). Then \(\mathcal{F} = \{\Omega, \emptyset, A, A^c\}\) is a \(\sigma\) algebra.
**Definition 1.2.4. [Borel σ-algebra]**

1. If \( \Omega = \mathbb{R} \), then \( \mathcal{B}(\mathbb{R}) \) denotes the smallest σ-algebra that contains all open intervals \((a, b)\) with \(-\infty < a < b < \infty\).

2. If \( \Omega = [0, 1] \), then \( \mathcal{B}([0, 1]) \) denotes the smallest σ-algebra that contains all open intervals \((a, b)\) with \(0 \leq a < b \leq 1\) and the sets \{0\} and \{1\}.

The σ-algebras \( \mathcal{B}(\mathbb{R}) \) and \( \mathcal{B}([0, 1]) \) are called Borel-σ algebras.

In [1] you find a proof such that such a smallest σ-algebra exists. There is no explicit description of all elements of \( \mathcal{B}(\mathbb{R}) \) or \( \mathcal{B}([0, 1]) \). Note, that these elements are subsets of \( \mathbb{R} \) and \([0, 1]\), respectively.

**Definition 1.2.5. [Probability measure, Probability space]**

Let \((\Omega, \mathcal{F})\) be a measurable space, i.e. \( \Omega \) is a non-empty set and \( \mathcal{F} \) is a σ-algebra on \( \Omega \).

1. A map \( \mathbb{P} : \mathcal{F} \to [0, 1] \) is called probability measure if
   
   - (a) \( \mathbb{P}(\Omega) = 1 \),
   - (b) for all \( A_1, A_2, \ldots \in \mathcal{F} \) with \( A_i \cap A_j = \emptyset \) for \( i \neq j \) one has
     \[
     \mathbb{P}\left( \bigcup_{i=1}^{\infty} A_i \right) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).
     \]

2. The triple \((\Omega, \mathcal{F}, \mathbb{P})\) is called probability space.

**Remark 1.2.6.** It follows from the definition of \( \mathbb{P} \):

- \( \mathbb{P}(\emptyset) = 0 \)
- \( \mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B) \)
- \( \mathbb{P}(A^c) = 1 - \mathbb{P}(A) \)
- ’Continuity of \( \mathbb{P} \) from below’: if \( A_1, A_2, \ldots \in \mathcal{F} \) such that \( A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots \) then
  \[
  \mathbb{P}\left( \bigcup_{n=1}^{\infty} A_n \right) = \lim_{N \to \infty} \mathbb{P}(A_N).
  \]

**Example 1.2.7. [Binomial Distribution]** We choose \( \Omega := \{0, \ldots, n\} \) and \( \mathbb{P}(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k} \) for \( k = 0, \ldots, n \). This is the binomial distribution.
1.2. SOME BASIC CONCEPTS OF STOCHASTICS

1.2.2 Conditional probability and independence

Definition 1.2.8. [Conditional probability] Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space, \(A \in \mathcal{F}\) with \(\mathbb{P}(A) > 0\). Then

\[
\mathbb{P}(B|A) := \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)}, \quad \text{for } B \in \mathcal{F},
\]

is the conditional probability of \(B\) given \(A\).

Problem 1.2.9. Let \((\Omega, \mathcal{F}, \mathbb{P})\) be a probability space and \(A \in \mathcal{F}\) with \(\mathbb{P}(A) > 0\). Prove that \((\Omega, \mathcal{F}, \mathbb{P}_A)\) is a probability space.

Definition 1.2.10. [Independence of events] The sets \(A_1, \ldots, A_n\) \((A_i \in \mathcal{F}, i = 1, \ldots, n)\) are called independent, if for each \(k = 2, 3, \ldots, n\) with \(1 \leq i_1 < i_2 \cdots < i_k \leq n\), it holds

\[
\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2})\cdots\mathbb{P}(A_{i_k}).
\]

Problem 1.2.11. Assume two events \(A, B \in \mathcal{F}\) with \(\mathbb{P}(A) > 0\). Prove that \(A\) and \(B\) are independent if and only if \(\mathbb{P}(B|A) = \mathbb{P}(B)\).

1.2.3 Random variables

Definition 1.2.12. A real function \(f : \Omega \to \mathbb{R}\) is called \(\mathcal{F}\)-measurable or random variable, if for all \((a, b)\) with \(-\infty < a < b < \infty\) the pre-image

\[
f^{-1}((a, b)) := \{\omega \in \Omega : f(\omega) \in (a, b)\} \in \mathcal{F}.
\]

The notation of a random variable is fundamental for the probability theory and for its applications, like stochastic modelling. The importance can be compared with the notation of a differentiable function in the theory of differential equations. The following statement is fundamental and proved in [1]:

Proposition 1.2.13. For an \(\mathcal{F}\)-measurable function \(f\) it holds that

\[
f^{-1}(B) := \{\omega \in \Omega : f(\omega) \in B\} \in \mathcal{F} \quad \forall B \in \mathcal{B}(\mathbb{R}).
\]

The fundamental importance of the above proposition, and therefore of the concept of measurable functions, has its reason in the following concept of law or image measure:
Problem 1.2.14. Assume a probability space \((\Omega, \mathcal{F}, \mathbb{I} P)\) and a measurable function \(f : \Omega \to \mathbb{R}\). Define
\[
\mathbb{I} P_f(B) := \mathbb{I} P \left( \{\omega \in \Omega : f(\omega) \in B \} \right) \quad \text{for} \quad B \in \mathcal{B}(\mathbb{R}).
\]
Verify that \((\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{I} P_f)\) is a probability space. The probability measure \(\mathbb{I} P_f\) is called the law of \(f\).

Example 1.2.15. Let \(A \in \mathcal{F}\). Then
\[
f(\omega) = \mathbb{I} A(\omega) = \begin{cases} 
1 & \text{if } \omega \in A \\
0 & \text{if } \omega \not\in A 
\end{cases}
\]
is a random variable. The function \(\mathbb{I} A(\cdot)\) is called indicator function of \(A\).

Definition 1.2.16. [Independence of random variables] The random variables \(f_i : \Omega \to \mathbb{R}\), \(i = 1, \ldots, n\), are called independent, if it holds
\[
\mathbb{I} P(\{f_1 \in B_1, \ldots, f_n \in B_n\}) = \prod_{i=1}^{n} \mathbb{I} P(\{f_i \in B_i\}), \quad \forall B_i \in \mathcal{B}(\mathbb{R}).
\]

Remark 1.2.17. (1) It is enough to check for independence in the above definition only with any open intervals \(B_i = (a_i, b_i)\) instead of \(B_i \in \mathcal{B}(\mathbb{R})\).

(2) Notation:
\[
\{f_1 \in B_1, \ldots, f_n \in B_n\} = \bigcap_{i=1}^{n} \{\omega \in \Omega : f_i(\omega) \in B_i\} = \bigcap_{i=1}^{n} f_i^{-1}(B_i) \in \mathcal{F}.
\]

(3) In case \(\Omega\) is countable (or finite), the random variables \(f_i\) have a countable range: \(\{f_i(\omega_1), f_i(\omega_2), \ldots\}\). Then \(f_1, \ldots, f_n\) are independent \(\iff\)
\[
\mathbb{I} P(f_1 = x_1, f_2 = x_2, \ldots, f_n = x_n) = \prod_{i=1}^{n} \mathbb{I} P(f_i = x_i)
\]
for any \(x_i\) of the range of the \(f_i\).

Example 1.2.18. Two dice: \(\Omega = \{(\omega_1, \omega_2) : \omega_1, \omega_2 \in \{1, \ldots, 6\}\}\)
\[
\mathbb{I} P((\omega_1, \omega_2)) = \frac{1}{36}
\]
1.2. SOME BASIC CONCEPTS OF STOCHASTICS

\[ f_1, f_2 : \Omega \to \mathbb{R} : \begin{align*}
  f_1((\omega_1, \omega_2)) &= \omega_1 + \omega_2 \\
  f_2((\omega_1, \omega_2)) &= \omega_1 - \omega_2
\end{align*} \]

Does it hold \( \forall x_1, x_2 \in \mathbb{R} \)

\[ \mathbb{P}(f_1 = x_1, f_2 = x_2) = \mathbb{P}(f_1 = x_1)\mathbb{P}(f_2 = x_2)? \]

Choose, for example, \( x_1 = 4 \) and \( x_2 = 2 \). Then

\[ \{ (\omega_1, \omega_2) : f_1((\omega_1, \omega_2)) = 4 \} = \{ (\omega_1, \omega_2) : \omega_1 + \omega_2 = 4 \} = \{ (1, 3), (2, 2), (3, 1) \} \]

and

\[ \{ (\omega_1, \omega_2) : f_1((\omega_1, \omega_2)) = 2 \} = \{ (\omega_1, \omega_2) : \omega_1 - \omega_2 = 2 \} = \{ (6, 4), (5, 3), (4, 2), (3, 1) \}. \]

Hence \( \mathbb{P}(f_1 = x_1, f_2 = x_2) = \frac{1}{36} \) but \( \mathbb{P}(f_1 = x_1)\mathbb{P}(f_2 = x_2) = \frac{1}{129} \), so \( f_1 \) and \( f_2 \) are not independent.
Checkup after reading, the more heuristic part:

- What is - in a way - the difference between deterministic and stochastic models?
- Why stochastic models are used?
- Give an example of a stochastic model.

The mathematical part:

- Explain the notion of a $\sigma$-algebra. Check whether you understood that a $\sigma$-algebra is a system of sub-sets of $\Omega$ and that a $\sigma$-algebra $\mathcal{F}$ on $\Omega$ is not a subset of $\Omega$.
- Why do we need a $\sigma$-algebra to define a probability measure?
- Problem 1.2.14: Check and understand for you the following: Assume a probability space $(\Omega, \mathcal{F}, P)$ and a random variable $f : \Omega \to \mathbb{R}$. You can think that $(\Omega, \mathcal{F}, P)$ is a stochastic model and $f$ is a parameter you see by an observation. Why it is necessary that $f$ is a random variable in order to compute

$$P \left( \{ \omega \in \Omega : 1 < f(\omega) < 2 \} \right)$$

An example might be, that $f(\omega)$ is the current price of a share of NOKIA (today it was at 2.32 Euro).
I would propose a change of the chat and video call times in Teams (or skype if you prefer):

- Wednesday 10-12: Here you can ask all your questions regarding the lecture.
- Thursday 14-16: Here we should discuss the Demo of the week.

In case of questions in between, you can reach by email or you can catch me in Teams if I am around there any time.

What is the aim of the second week, what is the main goal to reach?

- You need to recall the notion of the expectation in Section 1.2.4. Please check the three special cases. In case there is a problem, please let me know this and we have to clarify this in Teams by chat or video call.
- In Section 2.1 a special simple probability space is introduced in 2.1.1, which describes dyadic trees.
- There are two applications of these dyadic trees:
  1. In Section 2.3 we consider the important model from Finance, the Cox-Ross-Rubinstein model. This model works fine!
  2. In Section 2.4 we come back to the Snoqualmie Falls precipitation from Section 1.1.1 and will learn that we cannot proceed as simple as in the Cox-Ross-Rubinstein model. This model does not work fine so far!
- To model the Snoqualmie Falls precipitation we have to use Markov Chains. We start with this topic in Chapter 3.
1.2.4 Expectation

Assume a probability space $(\Omega, \mathcal{F}, \mathbb{P})$. For a general definition of the expectation of a random variable $f : \Omega \to \mathbb{R}$, i.e.

$$
\mathbb{E} f = \int_{\Omega} f(\omega) d\mathbb{P}(\omega),
$$

see [4] or [1].

Special cases to compute the expectation $\mathbb{E} f$ for a random variable $f : \Omega \to \mathbb{R}$

In this lecture we will use the following special cases:

(1) If $f$ is a step function i.e. $f(\omega) = \sum_{i=1}^{n} a_i \mathbb{1}_{A_i}(\omega)$ with $a_i \in \mathbb{R}$ and $A_i \in \mathcal{F}$, then

$$
\mathbb{E} f = \sum_{i=1}^{n} a_i \mathbb{P}(A_i).
$$

(2) If $\mathbb{P}_f(B) := \mathbb{P}(\{\omega : f(\omega) \in B\}) = \sum_{l=0}^{\infty} p_l \delta_{x_l}(B)$ with $p_l \geq 0$, $\sum_{l=0}^{\infty} p_l = 1$, and $x_0, x_1, \ldots \in \mathbb{R}$, then $\mathbb{P}_f$ is a discrete measure, where

$$
\delta_{x_l}(B) = \begin{cases} 
0 & ; \ l \not\in B \\
1 & ; \ l \in B
\end{cases}
$$

is called Dirac measure. It holds

$$
\mathbb{E} f = \sum_{l=0}^{\infty} f(x_l) p_l.
$$

(3) If $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, $f : \Omega \to \{y_0, y_1, \ldots\} \subseteq \mathbb{R}$ and

$$
\{\omega \in \Omega : f(\omega) = y_l\} \in \mathcal{F} \quad \text{for all} \quad l = 0, 1, \ldots
$$

If $\sum_{l=0}^{\infty} \max\{y_l, 0\} \mathbb{P}(f = y_l) < \infty$ or $\sum_{l=0}^{\infty} \max\{-y_l, 0\} \mathbb{P}(f = y_l) < \infty$, then the expectation is defined by

$$
\mathbb{E} f = \sum_{l=0}^{\infty} y_l \mathbb{P}(f = y_l).$$
(4) In case $\mathbb{P}_f(B) = \int_B h(y)dy$ for all $B \in \mathcal{B}({\mathbb{R}})$, i.e. if $\mathbb{P}_f$ has a density $h(x)$, we have

$$\mathbb{E}f = \int_{\mathbb{R}} yh(y)dy$$

if the expression on the right-hand side is well defined (that means if $\int_{\mathbb{R}} \max\{y, 0\}h(y)dy < \infty$ or $\int_{\mathbb{R}} \max\{-y, 0\}h(y)dy < \infty$).

Please note that this example would require further explanations, we skip here: Namely, the integral $\int_{\mathbb{R}} yh(y)dy$ is a Lebesgue integral and the function $h$ measurable and Lebesgue-integrable. So far, it is sufficient to assume that $h : \mathbb{R} \to [0, \infty)$ is continuous and to consider the Riemann-integral with the condition

$$\lim_{N \to \infty} \int_{-N}^{N} |y|h(y)|dy < \infty.$$
Chapter 2

Random walk

2.1 The probability space \((\mathbb{I}_d, \mathcal{F}_{\text{dyad}}^n, \mu_n^{(p)})\)

Definition 2.1.1 (Dyadic tree). Let \(0 < p < 1\) and \(n \geq 1\). Then

\[
\mathbb{I}_n := \{ (\varepsilon_1, \ldots, \varepsilon_n) : \varepsilon_1 = \pm 1, \ldots, \varepsilon_n = \pm 1 \},
\]

\[
\mathcal{F}_{\text{dyad}}^n := 2^{\mathbb{I}_n} \quad \text{(system of all subsets)},
\]

\[
\mu_n^{(p)} \left( \{ (\varepsilon_1, \ldots, \varepsilon_n) \} \right) := (1 - p)^k p^l,
\]

where

\[
n = k + l,
\]

\[
k = \# \{ i : \varepsilon_i = 1 \},
\]

\[
l = \# \{ i : \varepsilon_i = -1 \},
\]

and \(\#\) denotes the cardinality of the set.

In the next proposition we show that \((\mathbb{I}_n, \mathcal{F}_{\text{dyad}}^n, \mu_n^{(p)})\) is a finite probability space and then we explain why it is called dyadic tree.

Proposition 2.1.2. \((\mathbb{I}_n, \mathcal{F}_{\text{dyad}}^n, \mu_n^{(p)})\) is a finite probability space.

Proof. We check the defining properties of a probability space:

(1) \(\mathbb{I}_n\) is finite, has \(2^n\) elements,

(2) \(\mathcal{F}_{\text{dyad}}^n\) is a \(\sigma\)-algebra,
We have only to check whether
\[ \mu_n^{(p)}(\mathbb{I}_n) = \mathbb{P}(\Omega) = 1. \]

This follows from
\[ \mathbb{P}(\Omega) = \sum_{\omega \in \Omega} \mathbb{P}(\{\omega\}) \]
\[ = \sum_{\epsilon_1 = \pm 1} \ldots \sum_{\epsilon_n = \pm 1} \mu_n^{(p)}(\{\{\epsilon_1, \ldots, \epsilon_n\}\}) \]
\[ = \sum_{k=0}^{n} \sum_{(\epsilon_1, \ldots, \epsilon_n)} \mu_n^{(p)}(\{\{\epsilon_1, \ldots, \epsilon_n\}\}) \]
\[ \quad \#\{i : \epsilon_i = 1\} = k \]
\[ = \sum_{k=0}^{n} \sum_{(\epsilon_1, \ldots, \epsilon_n)} (1 - p)^k p^{n-k} \]
\[ \quad \#\{i : \epsilon_i = 1\} = k \]
\[ = \sum_{k=0}^{n} \binom{n}{k} (1 - p)^k p^{n-k} \]
\[ = (1 - p + p)^n = 1. \]

\[ \square \]

**Where the name dyadic tree comes from: interpretation of** \((\mathbb{I}_n, \mathcal{F}_n^{\text{dyad}}, \mu_n^{(p)})\)

Each \(\omega = (\epsilon_1, \ldots, \epsilon_n) \in \mathbb{I}_n\) corresponds to a path in the dyadic tree from the left to the right with:
\[ \epsilon_i = 1 \approx \text{in the i-th step up} \ (\uparrow) \]
\[ \epsilon_i = -1 \approx \text{in the i-th step down} \ (\downarrow) \]

Now consider a **random walk** such that in each step:
(RW_1) One goes up with the probability \((1 - p)\)
(RW_2) One goes down with the probability \(p\)
(RW_3) The behavior in the i-th step does not depend on the steps 1, \ldots, \(i - 1\).
Then one has:
Probability, that a path has k up-moves (↑) and l down-moves (↓) = $(1 - p)^k p^l$

$$\mu_n^{(p)}(\{\varepsilon_1, \ldots, \varepsilon_n\}), \text{ if } k = \#\{i : \varepsilon_i = 1\} \text{ and } l = \#\{i : \varepsilon_i = -1\}.$$  

Hence $\mu_n^{(p)}$ can be interpreted as the probability measure on the paths on a dyadic tree, if one assumes properties (RW$_1$), (RW$_2$) and (RW$_3$).

### 2.2 The discrete time geometric Brownian motion

Now we introduce a process that describes a random walk with up-and down moves that are proportional to the previous value. This process is discrete time analog of the so-called geometric Brownian motion. For this reason we choose the name discrete time geometric Brownian motion. This process has a central role in models for share prices as we indicate later.

**Definition 2.2.1 (Geometric Brownian motion, dyadic setting).**

1. We define the function $\Delta_i : \mathbb{D}_n \to \mathbb{R}$ for $i = 1, \ldots, n$ by
   $$\Delta_i(\varepsilon_1, \ldots, \varepsilon_n) := \varepsilon_i.$$

2. Given $\sigma > 0$ and $b \in \mathbb{R}$ we define $S_i : \mathbb{D}_n \to \mathbb{R}$ for $i = 1, \ldots, n$ by
   $$S_i(\omega) := e^{\sigma(\Delta_1(\omega) + \cdots + \Delta_i(\omega)) - bi} \quad \text{and} \quad S_0(\omega) := 1.$$

The process $(S_i)_{i=0}^n$ is called (discrete time) geometric Brownian motion.

Now we collect the main properties of our geometric Brownian motion:

**Proposition 2.2.2.**

1. If $1 \leq i \leq n$, then
   $$\mu_n^{(p)} \left( \frac{S_i}{S_{i-1}} = e^{-\sigma - b} \right) = p,$$
   $$\mu_n^{(p)} \left( \frac{S_i}{S_{i-1}} = e^{\sigma - b} \right) = 1 - p.$$

2. The random variables
   $$\frac{S_1}{S_0}, \frac{S_2}{S_1}, \ldots, \frac{S_n}{S_{n-1}}$$

are independent.
CHAPTER 2. RANDOM WALK

Proof. (1) We have \( \frac{S_i}{S_{i-1}} = \frac{e^{\sigma(\Delta_1 + \cdots + \Delta_i) - bi}}{e^{\sigma(\Delta_1 + \cdots + \Delta_{i-1}) - bi_{i-1}}} = e^{\sigma \Delta_i - b} \), so that

\[
\mu_n^{(p)} \left( \frac{S_i}{S_{i-1}} = e^{-\sigma - b} \right) = \mu_n^{(p)} (\Delta_i = -1) = p.
\]

\[
\mu_n^{(p)} \left( \frac{S_i}{S_{i-1}} = e^{\sigma - b} \right) = \mu_n^{(p)} (\Delta_i = 1) = 1 - p.
\]

(2) The random variables \( \Delta_1, \ldots, \Delta_n \) are independent, that means

\[
\mu_n^{(p)}(\Delta_1 = x_1, \ldots, \Delta_n = x_n) = \mu_n^{(p)}(\Delta_1 = x_1) \cdots \mu_n^{(p)}(\Delta_n = x_n)
\]

for all \( x_1, \ldots, x_n = \pm 1 \) as we check in the demo. Hence \( \frac{S_1}{S_0}, \frac{S_2}{S_1}, \ldots, \frac{S_n}{S_{n-1}} \) are independent because \( \frac{S_i}{S_{i-1}} = e^{\sigma \Delta_i - b} \) and the independence of \( \Delta_1, \ldots, \Delta_n \) implies the independence of \( e^{\sigma \Delta_1 - b}, \ldots, e^{\sigma \Delta_n - b} \) which is intuitively clear and we check this in the demos.

Remark 2.2.3. The process \( (\log S_i)_{i=1}^n = (\sigma \sum_{k=1}^i \Delta_k - bi)_{i=1}^n \) is a process with independent and identically distributed increments, where the increments are

\[
\log S_1 - \log S_0 = \sigma \Delta_1 - b,
\]

\[
\log S_2 - \log S_1 = \sigma \Delta_2 - b, \ldots.
\]

This follows from the fact that \( \Delta_1, \ldots, \Delta_n \) have the same distribution and are independent.

2.3 The Cox-Ross-Rubinstein model for stock prices

**Assumptions:** The price of one share changes from one day to the next according to these rules:

(A1) The price goes up by the factor \( c_u > 1 \) with probability \( 1 - p \),

(A2) the price goes down by the factor \( 0 < c_d < 1 \) with probability \( p \),

(A3) the price changes are independent from the foregoing days.

**Model:** \( X_i = \) price in Euro for one share of NOKIA at the \( i \)-th day \( (i = 0, \ldots, n) \).

We let

\[
X_i := X_0 e^{\sigma(\Delta_1 + \cdots + \Delta_i) - bi}.
\]
2.4. SNOQUALMIE FALLS PRECIPITATION

The parameter $b$ is called drift and the parameter $\sigma$ is called volatility and is responsible for the fluctuation around the mean.

**In practice** we have:

| Data | Estimates for $p, c_u, c_d$ | Computation of $b, \sigma$ |

Please note that the parameters $c_d$ and $c_u$ in

$$\begin{align*}
e^{-\sigma - b} &= c_d \\
e^{\sigma - b} &= c_u
\end{align*}$$

imply the values of the parameters $\sigma, b$.

In certain cases it is enough to estimate $\sigma$. For example, the formula of the price of an European Call Option does not use $p$ and $b$. There are 2 ways to estimate $\sigma$:

1. **way:** In the **Historical Method** one fits $\sigma$ to the historical data of the prices of the share.

2. **way:** Using the method **Implicit Volatility** one fits $\sigma$ to already known prices for European call options.

### 2.4 Snoqualmie Falls precipitation

**Assumptions:** Suppose that there is some $p \in (0, 1)$ such that

(A1) $\Pr$(day is wet) = $p$,

(A2) $\Pr$(day is dry) = $1 - p$,

(A3) the weather of one day does not depend on the weather of the other days (in particular, does not depend on the foregoing days).

**Data** (January 1948 - January 1983):

325 dry days

791 wet days

**Estimate for $p$:** Use a **maximum likelihood estimate** for $p$. We have to find that $p$ which explains our data best possible, where

$$L(p) := \Pr(325 \text{ dry days}, 791 \text{ wet days}) = p^{791}(1 - p)^{325}.$$ 

Then the maximum likelihood estimator $\hat{p}$ of $p$ is defined by
\[ L(\hat{p}) := \max \{ L(p) : 0 \leq p \leq 1 \}. \]

**Computation of \( \hat{p} \) :**

- \( L(0) = L(1) = 0 \)
- We have \( L'(p) = 791p^{790}(1 - p)^{325} + p^{791}325(1 - p)^{324}(-1) \) so that
  \[ L'(p) = 0 \iff 791(1 - p) - 325p = 0 \]
  \[ \iff 791 = p(791 + 325) \]
  \[ \iff p = \frac{791}{325 + 791} = 0.709\ldots \]

Hence: \( \hat{p} \approx 0.709 \).

**Now we test of our model:** We the values above we would get that

\[
\mathbb{P}( \text{weather changes from one day to the other} )
= \mathbb{P}( \text{wet day }\rightarrow\text{ dry day} ) + \mathbb{P}( \text{dry day }\rightarrow\text{ wet day} )
= (1 - p)p + p(1 - p) = 2(1 - p)p.
\]

Using \( \hat{p} \) this yields

\[
\mathbb{P}( \text{weather changes from one day to the other} ) = 2(1 - \hat{p})\hat{p} \approx 0.412638
\]

and

Expected weather changes changes in January
\[
= \text{number of years } \times \text{ possible changes } \times \mathbb{P}( \text{weather changes from one day to the other} )
= 36 \times 30 \times 0.412638
\approx 446.
\]

However, from the data we observe 251 weather changes. Therefore, this model is **not the right one.** This is the reason to introduce Markov chains in the next chapter.
Chapter 3

Discrete time Markov chains

Andrey A. Markov (1856-1922) was a Russian mathematician, who studied and worked as a professor in St. Petersburg and developed the theory of Markov chains.

With this model we get

$$\Pr(f_1 = 2) = \frac{1}{2} \quad \text{and} \quad \Pr(f_1 = 4) = \frac{1}{2}.$$  

How to compute $\Pr(f_n = 1)$? For this we use conditional probabilities, that can be explained as follows: Assume the walker was at the time $n$ in corner 2, then

$$\Pr(f_{n+1} = 1|f_n = 2) = \frac{1}{2} \quad \text{and} \quad \Pr(f_{n+1} = 3|f_n = 2) = \frac{1}{2}.$$ 

It holds

$$\Pr(f_{n+1} = 1|f_0 = k_0, f_1 = k_1, \ldots, f_{n-1} = k_{n-1}, f_n = 2) = \frac{1}{2},$$
\[ \mathbb{P}(f_{n+1} = 3 | f_0 = k_0, f_1 = k_1, \ldots, f_{n-1} = k_{n-1}, f_n = 2) = \frac{1}{2} \]

for all \( k_i \in \{1, 2, 3, 4\} \), since coin flipping at time \( n + 1 \) is independent of all previous coin flippings and therefore independent of \( f_0, \ldots, f_n \).

### 3.1 Introduction

We assume the following:

- A set \( X = \{x_1, \ldots, x_M\} \), that is interpreted as state space.
- \( p_0 : \Omega \to [0, \infty) \) with \( \sum_{l=1}^{M} p_0(x_l) = 1 \) which will be interpreted as initial distribution.
- We assume matrices
  \[
  T_i = \begin{pmatrix}
  p_i(x_1, x_1) & \ldots & p_i(x_1, x_M) \\
  \vdots & \ddots & \vdots \\
  p_i(x_M, x_1) & \ldots & p_i(x_M, x_M)
  \end{pmatrix}
  \]
  for \( i = 1, 2, \ldots, n \) with \( p_i(x_l, x_m) \geq 0 \) for all \( 1 \leq l, m \leq M \) and
  \[
  \sum_{m=1}^{M} p_i(x_l, x_m) = 1 \quad \text{for all} \quad l = 1, \ldots, M.
  \]
  These matrices will be interpreted as transition kernels.

**Theorem 3.1.1 (Existence of a Markov chain).** For a given \( p_0 \) and matrices \( T_i \), \( i = 1, \ldots, n \) there exists a finite probability space \( (\Omega, 2^\Omega, \mathbb{P}) \) and a family of random variables \( (f_i)_{i=0}^{n} \), where \( f_i : \Omega \to X \), such that the following holds:

1. \( p_0(x_l) = \mathbb{P}(f_0 = x_l) \) for \( l = 1, \ldots, M \).
2. For \( i = 1, \ldots, n \) and \( l, m = 1, \ldots, M \) one has
   \[
   p_i(x_l, x_m) = \mathbb{P}(f_i = x_m | f_{i-1} = x_l) \quad \text{if} \quad \mathbb{P}(f_{i-1} = x_l) > 0.
   \]
3.1. INTRODUCTION

(3) For \(y_0, \ldots, y_i \in X\) with \(i \in \{1, \ldots, n\}\) one has

\[
\mathbb{P}(f_i = y_i | f_{i-1} = y_{i-1}, f_{i-2} = y_{i-2}, \ldots, f_0 = y_0) = \mathbb{P}(f_i = y_i | f_{i-1} = y_{i-1}) \tag{3.1}
\]

if \(\mathbb{P}(f_{i-1} = y_{i-1}, f_{i-2} = y_{i-2}, \ldots, f_0 = y_0) > 0\).

The interpretation of Theorem 3.1.1 is as follows:

– The initial distribution is given by \(p_0\) and this is the distribution of the random variable \(f_0\), see Item (1).

– The probabilities for the transition from time \(i - 1\) to time \(i\) is given in Item (2).

– In Item (3) it is said that the transition from time \(i - 1\) to time \(i\) is independent from the ‘history’ of the process, but it only depends on the of \(f_{i-1}\). This is called Markov property.
Checkup after reading:

- Check the 4 examples of integration in 1.2.4. What is the connection to the Riemann integral?

- Geometric Brownian motion: choose simple $\sigma$ and $b$ in Definition 2.2.1 and draw on a paper all possible paths, i.e. $i \mapsto S_i(\omega)$ as a function $\{0, 1, 2, 3\} \rightarrow \mathbb{R}$? You should get 8 functions. Check these 8 functions to see that they meet each other sometimes, this is called recombinining.

- Why is the Cox-Ross-Rubenstein model invariant against a change of the currency: How the model differs if the same share is expressed in Euro or in Japanese Yen? **Hint:** It changes only by a factor and the relativ changes of the prices stay the same.

- Check whether you fully understood that the model in Section 2.4 does not work.

- Check whether the geometric random walk satisfies the assumptions of Theorem 3.1.1.
I received many solved exercises from you and this is very good! Thank you, please continue like that. I have an internal list and think about, how to deliver this list to you so that you know, how many exercises you solved successfully.

I would like to discuss with you your solution(s) in Teams on Wednesday 10-12 or latest on Thursday 14-16 to point out things that can be improved or/and have not been correct. Does this work for you?

So, I am in Teams (or skype if you prefer):

- Wednesday 10-12: Here you can ask all your questions regarding the lecture and Demo 1.
- Thursday 14-16: Here we should discuss the Demo 2.

What is the aim of the third week, what is the main goal to reach?

- Notion of a Markov chain (Definition 3.1.2).
- Chapman-Kolmogorov equations and their intuitive understanding (Proposition 3.2.1).
- Computation of the marginal distribution (Corollary 3.2.3).
- Example of a Markov chain in Section 3.2.1.
- What is an absorbing state (Section 3.3.1).
- Setting of Branching process in Section 3.3.2.
We will not prove Theorem 3.1.1, but formalize the assertion of this theorem as a definition. The following definition of a Markov chain is one of the main definitions of this course:

**Definition 3.1.2 (Markov chain).**

1. The triplet \((f_i^n)_{i=0}^n, p_0, (T_i^n)_{i=1}^n\) is called **Markov chain**.

2. The relation (3.1) is called **Markov property**.

3. The distribution \(p_0\) is called the **initial distribution** of the Markov chain.

4. The matrices \((T_i^n)_{i=1}^n\) are called **transition matrices** of the Markov chain. If \(T_1 = T_2 = \cdots = T\), then the Markov chain is called **homogeneous**.

**Remark 3.1.3.** The relation \(\sum_{m=1}^M p_i(x_i, x_m) = 1\) corresponds to

\[
\sum_{m=1}^M \mathbb{P}(f_i = x_m | f_{i-1} = x_l) = \mathbb{P}(\bigcup_{m=1}^M \{f_i = x_m\} | f_{i-1} = x_l)
\]

\[
= \mathbb{P} \left( f_i \in \{x_1, \ldots, x_M\} \big| f_{i-1} = x_l \right) = 1
\]

if \(\mathbb{P}(f_{i-1} = x_l) > 0\).

**Example 3.1.4.** The geometric Brownian motion \((S_i^n)_{i=0}^n\) with \(S_i = e^{\sigma(\Delta_1 + \cdots + \Delta_i) - b_i}\) has the Markov property.

**Proof.** By definition we have \(S_0(\omega) := 1\), so that for the initial distribution \(p_0\) we get \(p_0(1) = \mathbb{P}(S_0 = 1) = 1\). Now we verify the Markov property. For \(y_i > 0\) we have

\[
\mathbb{P}(S_{i+1} = y_{i+1} | S_i = y_i, \ldots, S_0 = y_0)
\]

\[
= \frac{\mathbb{P}(S_{i+1} = y_{i+1}, S_i = y_i, \ldots, S_0 = y_0)}{\mathbb{P}(S_i = y_i, \ldots, S_0 = y_0)}
\]

\[
= \frac{\mathbb{P} \left( \frac{S_{i+1}}{S_i} = \frac{y_{i+1}}{y_i}, S_i = y_i, \ldots, S_0 = y_0 \right)}{\mathbb{P}(S_i = y_i, \ldots, S_0 = y_0)}
\]

\[
= \mathbb{P} \left( \frac{S_{i+1}}{S_i} = \frac{y_{i+1}}{y_i} \right)
\]
3.2. THE CHAPMAN - KOLMOGOROV EQUATIONS

\[
\begin{align*}
&= \frac{\mathbb{P}(S_{i+1} = \frac{y_{i+1}}{y_i}, S_i = y_i)}{\mathbb{P}(S_i = y_i)} \\
&= \frac{\mathbb{P}(S_{i+1} = \frac{y_{i+1}}{y_i}, S_i = y_i)}{\mathbb{P}(S_i = y_i)} \\
&= \frac{\mathbb{P}(S_{i+1} = y_{i+1}, S_i = y_i)}{\mathbb{P}(S_i = y_i)} \\
&= \mathbb{P}(S_{i+1} = y_{i+1} | S_i = y_i).
\end{align*}
\]

Here we used that the \(\Delta_k\)'s are independent. The set \(\{S_i = y_i, \ldots, S_0 = y_0\}\) can be expressed using only \(\Delta_1, \ldots, \Delta_i\), while \(\{S_i = y_i, \ldots, S_0 = y_0\}\) is independent. In the same way one can see that \(\{S_i = y_i, \ldots, S_0 = y_0\}\) and \(\{S_i = y_i\}\) are independent.

3.2 The Chapman - Kolmogorov equations

Let \(((f_i)_{i=0}^n, p_0, T)\) be a Markov chain with state space \(X = \{x_1, \ldots, x_M\}\). Often one is interested in the so-called **marginal distributions**

\[
\Pi^{(i)} := \left( p^{(i)}(x_k) \right)_{k=1,\ldots,M} \quad \text{where} \quad p^{(i)}(x_k) := \mathbb{P}(f_i = x_k).
\]

In other words, the marginal distribution \(\Pi^{(i)}\) describes the distribution of the Markov chain at time \(i\), i.e. the distribution of \(f_i\). We also need the matrices

\[
P^{(i)} := \left( p^{(i)}(x_k, x_l) \right)_{k,l=1,\ldots,M} \quad \text{where} \quad p^{(i)}(x_k, x_l) := \mathbb{P}(f_i = x_k | f_0 = x_l).
\]

Similar to the transition matrices \((T_i)_{i=1}^n\), the matrices \(P^{(i)}\) describe the transition probabilities from time 0 to time \(i\). Now we obtain the famous Chapman-Kolmogorov equations:

**Proposition 3.2.1 (Chapman-Kolmogorov equations).** For a homogeneous Markov chain \(((f_i)_{i=0}^n, p_0, T)\) the following holds

1. \(P^{(i+j)} = P^{(i)} \circ P^{(j)}\) for \(1 \leq i, j, i + j \leq n\)
2. \(\Pi^{(i+j)} = \Pi^{(i)} \circ P^{(j)}\) for \(1 \leq j, i + j \leq n\) and \(0 \leq i \leq n\).
**Proof.** One has to show that

\[ p^{(i+j)}(x_k, x_l) = \sum_{s=1}^{M} p^{(i)}(x_k, x_s) p^{(j)}(x_s, x_l) \]

and

\[ p^{(i+j)}(x_l) = \sum_{s=1}^{M} p^{(i)}(x_s) p^{(j)}(x_s, x_l). \]

Because of Ω = \( \bigcup_{s=1}^{M} \{ f_i = x_s \} \) and \( \{ f_i = x_{s_1} \} \cap \{ f_i = x_{s_2} \} = \emptyset \) for \( s_1 \neq s_2 \) it holds

\[
p^{(i+j)}(x_k, x_l) = \mathbb{P}(f_{i+j} = x_l | f_0 = x_k) = \sum_{s=1}^{M} \mathbb{P}(f_{i+j} = x_l, f_i = x_s | f_0 = x_k) = \sum_{s=1}^{M} \frac{\mathbb{P}(f_{i+j} = x_l, f_i = x_s, f_0 = x_k)}{\mathbb{P}(f_0 = x_k)}
\]

by the definition of conditional probability. Later we will see that the Markov property implies

\[ \mathbb{P}(f_{i+j} = x_l | f_i = x_s, f_0 = x_k) = \mathbb{P}(f_{i+j} = x_l | f_i = x_s). \quad (3.2) \]

Using this relation and that the Markov chain is homogeneous we may continue with

\[
\sum_{s=1}^{M} \frac{\mathbb{P}(f_{i+j} = x_l, f_i = x_s, f_0 = x_k)}{\mathbb{P}(f_0 = x_k)} = \sum_{s=1}^{M} \mathbb{P}(f_{i+j} = x_l | f_i = x_s, f_0 = x_k) \mathbb{P}(f_i = x_s | f_0 = x_k) = \sum_{s=1}^{M} \mathbb{P}(f_{i+j} = x_l | f_i = x_s) \mathbb{P}(f_i = x_s | f_0 = x_k) = \sum_{s=1}^{M} \mathbb{P}(f_j = x_l | f_0 = x_s) \mathbb{P}(f_i = x_s | f_0 = x_k)
\]
This implies $P^{(i+j)} = P^{(i)} \circ P^{(j)}$. Similarly, one gets

$$p^{(i+j)}(x_t) = \sum_{s=1}^{M} p^{(j)}(x_s, x_t) p^{(i)}(x_k, x_s).$$

**Example 3.2.2.** We illustrate why (3.2) should be true by the help of the following example. Let $(f_i)_{i=0}^{n}$ be a Markov chain. We will check whether

$$\mathbb{P}(f_3 = y_3 | f_2 = y_2, f_0 = y_0) = \mathbb{P}(f_3 = y_3 | f_2 = y_2)$$

is true. The Markov property states that

$$\mathbb{P}(f_3 = y_3, f_2 = y_2, f_1 = y_1, f_0 = y_0) = \mathbb{P}(f_3 = y_3 | f_2 = y_2).$$

Hence

$$\mathbb{P}(f_3 = y_3 | f_2 = y_2, f_0 = y_0) = \frac{\mathbb{P}(f_3 = y_3, f_2 = y_2, f_0 = y_0)}{\mathbb{P}(f_2 = y_2, f_0 = y_0)} = \sum_{s=1}^{M} \frac{\mathbb{P}(f_3 = y_3, f_2 = y_2, f_1 = x_s, f_0 = y_0)}{\mathbb{P}(f_2 = y_2, f_0 = y_0)} = \sum_{s=1}^{M} \frac{\mathbb{P}(f_3 = y_3, f_2 = y_2, f_1 = x_s, f_0 = y_0)}{\mathbb{P}(f_2 = y_2, f_0 = y_0)} \frac{\mathbb{P}(f_2 = y_2, f_1 = x_s, f_0 = y_0)}{\mathbb{P}(f_2 = y_2, f_0 = y_0)}.$$
\[ P(f_3 = y_3 | f_2 = y_2) \sum_{s=1}^{M} P(f_2 = y_2, f_1 = x_s, f_0 = y_0) \]
\[ = P(f_3 = y_3 | f_2 = y_2). \]

**Corollary 3.2.3 (Marginal distribution).** Let \((f_i)_{i=0}^{n}\) be a homogeneous Markov chain and let \(j \in \{1, \ldots, n\}\). Then we have

(1) \[ P^{(j)} = T \circ \cdots \circ T, \]
\[ j \text{ times} \]

(2) \[ \Pi^{(j)} = \Pi^{(0)} \circ T \circ \cdots \circ T. \]
\[ j \text{ times} \]

**Proof.** By definition we have that \(P^{(1)} = T\), so that, by induction,

\[ P^{(j)} = P^{(1)} \circ P^{(j-1)} = P^{(1)} \circ (P^{(1)} \circ P^{(j-2)}) = \cdots = T \circ \cdots \circ T, \]
\[ j \text{ times} \]

and

\[ \Pi^{(j)} = \Pi \circ P^{(j)} = \Pi \circ T \circ \cdots \circ T. \]
\[ j \text{ times} \]

\[ \square \]

**Theorem 3.2.4 ('Step-by-Step formula').** Let \((f_i)_{i=0}^{n}\) be a homogeneous Markov chain. Then for \(0 \leq k < m \leq n\)

\[ P(f_m = y_m, \ldots, f_{k+1} = y_{k+1} | f_k = y_k) \]
\[ = P(f_m = y_m | f_{m-1} = y_{m-1})P(f_{m-1} = y_{m-1} | f_{m-2} = y_{m-2}) \times \]
\[ \cdots \times P(f_{k+1} = y_{k+1} | f_k = y_k), \]

if \(P(f_{m-1} = y_{m-1}, \ldots, f_k = y_k) > 0\) and \(y_i \in X, i = k, k+1, \ldots, m\).

We will not prove this formula here.
3.2. THE CHAPMAN - KOLMOGOROV EQUATIONS

Example 3.2.5 (Independent random variables). Let $f_0, f_1, \ldots, f_n$ be independent random variables. Is $f_0, f_1, \ldots, f_n$ also a Markov chain? Let us check the Markov property:

$$
P(f_{k+1} = y_{k+1} | f_k = y_k, \ldots, f_0 = y_0) = \frac{\mathbb{P}(f_{k+1} = y_{k+1}, f_k = y_k, \ldots, f_0 = y_0)}{\mathbb{P}(f_k = y_k, \ldots, f_0 = y_0)} = \frac{\mathbb{P}(f_{k+1} = y_{k+1}) \mathbb{P}(f_k = y_k) \cdots \mathbb{P}(f_0 = y_0)}{\mathbb{P}(f_k = y_k) \cdots \mathbb{P}(f_0 = y_0)} = \frac{\mathbb{P}(f_{k+1} = y_{k+1})}{\mathbb{P}(f_k = y_k)} = \frac{\mathbb{P}(f_{k+1} = y_{k+1}, f_k = y_k)}{\mathbb{P}(f_k = y_k)} = \mathbb{P}(f_{k+1} = y_{k+1} | f_k = y_k).
$$

So, we see that the Markov property is satisfied.

3.2.1 Modeling of Snoqualmie Falls precipitation with a Markov chain

Let $X = \{0, 1\}$, where 0 stands for a dry day and 1 for a rainy day. We assume that the weather in different years behaves independently.

We shall use the following parameters:

- Initial distribution $p_0 = (p, 1 - p)$ with the interpretation
  $$
  \mathbb{P}(\text{1st of January is dry}) = p, \quad \mathbb{P}(\text{1st of January is rainy}) = 1 - p.
  $$

- The Transition matrix
  $$
  T := \begin{pmatrix}
  p_{00} & p_{01} \\
  p_{10} & p_{11}
  \end{pmatrix}.
  $$

- As time scale we take $1948 \leq j \leq 1983$, which is mathematically the same as $i = 0, \ldots, n$, so that please do not worry about this.
CHAPTER 3. DISCRETE TIME MARKOV CHAINS

Then \((f^{(j)}_i)_{i=1}^{31} ; p_0, T)\) is a Markov chain which describes the weather in January of year \(j\), i.e. for each year \(j\) we consider a Markov chain.

**Maximum Likelihood estimates for** \(T\): For each January in year \(j\) we let

\[
A_j = \text{number of changes } 0 \rightarrow 0, \\
B_j = \text{number of changes } 0 \rightarrow 1, \\
C_j = \text{number of changes } 1 \rightarrow 0, \\
D_j = \text{number of changes } 1 \rightarrow 1 \text{ in year } j.
\]

We observe that \(A_j + B_j + C_j + D_j = 30\) as we have 31 days in January and observe therefore 30 changes.

The probability of the observed data we can express, since the years are assumed to be independent and one has the 'step-by-step-formula' (Theorem 3.2.4), by

\[
\prod_{j=1948}^{1983} \mathbb{P}(f^{(j)}_1 = x_{1j}) p_{00}^{A_j} p_{01}^{B_j} p_{10}^{C_j} p_{11}^{D_j} = \prod_{j=1948}^{1983} \mathbb{P}(f^{(j)}_1 = x_{1j}) p_{00}^{186} p_{01}^{123} p_{10}^{128} p_{11}^{643}.
\]

According to the Maximum Likelihood principle we look for

\[
L(\hat{p}_{00}, \hat{p}_{01}, \hat{p}_{10}, \hat{p}_{11}) = \max_{p_{ij}} L(p_{00}, p_{01}, p_{10}, p_{11}).
\]

Because \(p_{00} + p_{01} = 1\) and \(p_{10} + p_{11} = 1\) we can maximize \(p_{00}^{186} p_{01}^{123} p_{10}^{128} p_{11}^{643}\) separately. This yields to the matrix

\[
T = \begin{pmatrix}
0.602 & 0.398 \\
0.166 & 0.834
\end{pmatrix}.
\]

**Example 3.2.6.** What is the probability of wet/dry at the 6-th of January if it is raining at the 1-st of January? Computation yields to

\[
T^5 = \begin{pmatrix}
0.305 & 0.695 \\
0.290 & 0.710
\end{pmatrix}.
\]

Hence

\[
\mathbb{P}(f^{(j)}_5 = 1 | f^{(j)}_0 = 1) = 0.710, \\
\mathbb{P}(f^{(j)}_5 = 0 | f^{(j)}_0 = 1) = 0.290.
\]
3.3 Classification of states for a homogeneous Markov chain

Let

\[ T = (p_{ij})_{i,j=0}^{K} = \begin{pmatrix} p_{00} & p_{01} & \cdots & p_{0K} \\ \vdots & & \ddots & \vdots \\ p_{K0} & p_{K1} & \cdots & p_{KK} \end{pmatrix} \]

be the transition matrix and \( X := \{0, \ldots, K\} \) the state space. (Also an infinite state space \( X = \{0, 1, \ldots\} \) is possible here.)

3.3.1 Absorbing states

**Definition 3.3.1.** A state \( k \in X \) of a homogeneous Markov chain \((f_i)_{i=0}^{n}\) is called **absorbing** if and only if

\[ p_{kk} = \mathbb{P}(f_{i+1} = k | f_i = k) = 1. \]

In other words, if a Markov chain reaches an absorbing state \( k \), then it stays there forever. How to compute the probability, that a Markov chain reaches an absorbing state after some time? To discuss this problem, let us start with an important example.

3.3.2 Absorbing states for branching processes

In this example we assume an infinite Markov chain \((f_i)_{i=0}^{\infty}\) with a finite state space \( X = \{0, \ldots, K\} \). In a sense, all statements discussed so far stay valid. Only the proof of Theorem 3.1.1 would be more involved, however we did not prove Theorem 3.1.1 here in the course, so that it does not make a difference for us.

To start with, we assume we have a process \((f_i)_{i=0}^{\infty}\) with the following properties:

- \( f_0 = 1 \) particle
- \( f_i = \) number of particles at time \( i \)
• At each step each particle will be independently replaced by

0 particles with probability \( p_0 \in (0, 1) \),

1 particle . . . \( p_1 \in [0, 1] \),

... \( p_K \in [0, 1] \),

Note that \( p_0 + p_1 + \cdots + p_K = 1 \). The only absorbing state here is \( \{0\} \). We want to compute

\[
P(\{\omega \in \Omega : \text{there exists an } i \geq 1 \text{ such that } f_i(\omega) = 0\}).
\]

For this we use the method of probability generating functions.

**Definition 3.3.2.** Let \( F, F_i : [0, 1] \rightarrow [0, 1] \) be given by

\[
F(\theta) := \sum_{l=0}^{K} \theta^l p_l,
\]

\[
F_i(\theta) := \sum_{l=0}^{K} \theta^l \mathbb{P}(f_i = l), \quad i = 1, 2, \ldots
\]

**Remark 3.3.3.** (1) It holds that \( p_l = \frac{1}{l} F^{(l)}(0) \).

(2) One has \( F = F_1 \) because of \( p_l = \mathbb{P}(f_1 = l) \) and \( f_0 = 1 \).

The following proposition is crucial to investigate branching processes by generating functions:

**Proposition 3.3.4.** For \( i = 1, 2, \ldots \) one has

\[
F_{i+1}(\theta) = F_i(F(\theta)).
\]

Consequently,

\[
F_i = \underbrace{F \circ \cdots \circ F}_{i \text{ times}} \quad \text{and} \quad F_{i+1}(\theta) = F(F_i(\theta)).
\]

We will use the following
Lemma 3.3.5. Assume $g_1, \ldots, g_n$ are independent random variables, with $\mathbb{E}|g_i| < \infty$ for $i = 1, \ldots, n$. Then

$$\mathbb{E}|\Pi_{i=1}^n g_i| < \infty \quad \text{and} \quad \mathbb{E}\Pi_{i=1}^n g_i = \Pi_{i=1}^n \mathbb{E}g_i.$$ 

One can prove this Lemma by using Lemma 3.7.4 of [4] and induction.

Proof of Proposition 3.3.4. We let $0 \leq \theta \leq 1$, and $i \in \{1, 2, \ldots\}$

$$F_{i+1} (\theta) = \sum_{l=0}^{\infty} \theta^l \mathbb{P}(f_{i+1} = l)$$

$$= \sum_{l=0}^{\infty} \theta^l \mathbb{E} \mathbb{I}_{\{f_{i+1} = l\}}$$

$$= \mathbb{E} \sum_{l=0}^{\infty} \theta^l \mathbb{I}_{\{f_{i+1} = l\}} = \mathbb{E}\theta^{f_{i+1}}.$$ 

We will assume that $\xi_1, \xi_2, \ldots$ are independent and identically distributed like $f_1$. That means that $\forall m \geq 1$ we have $\mathbb{P}(\xi_m = k) = p_k$ for $k = 0, 1, 2, \ldots$ independently from $f_i$. Hence, by Lemma 3.3.5,

$$\mathbb{E}\theta^{f_{i+1}} = \mathbb{E} \sum_{l=0}^{\infty} \mathbb{I}_{\{f_{i+1} = l\}} \theta^{\xi_1 + \cdots + \xi_l}$$

$$= \sum_{l=0}^{\infty} \mathbb{E} \mathbb{I}_{\{f_{i+1} = l\}} \theta^{\xi_1} \times \cdots \times \theta^{\xi_l}$$

$$= \sum_{l=0}^{\infty} \mathbb{P}(f_i = l) \mathbb{E}\theta^{\xi_1} \times \cdots \times \mathbb{E}\theta^{\xi_l}$$

$$= \sum_{l=0}^{\infty} \mathbb{P}(f_i = l) (\mathbb{E}\theta^{f_1})^l$$

$$= F_i(\mathbb{E}\theta^{f_1}) = F_i(F(\theta)).$$

This implies the first assertion. For the second we continue with

$$F_i = F_{i-1} \circ F = (F_{i-2} \circ F) \circ F = \cdots = F \circ \cdots \circ F.$$ 

\[\square\]
Checkup after reading:

- What is the difference between a general Markov chain and a homogeneous Markov chain?

- Go to Proposition 3.2.1: Did you understood the matrix multiplication $P^{(i)} \circ P^{(j)}$ and the multiplication of a vector with a matrix $\Pi^{(i)} \circ P^{(j)}$?

  Compare the approaches in Sections 2.4 and 3.2.1. for modeling the Snoqualmie Falls precipitation.

- What other real phenomena would you model with a branching process introduced in this week? Think about recent developments.
3.3. CLASSIFICATION OF STATES

Week 4  13.04. – 17.04.2020

Again I received many solved exercises from you and this is good! Thank you, please continue like that.

I am in Teams (or skype if you prefer):

– April 15, Wednesday 10-12: Here you can ask all your questions regarding the lecture and Demo 1.

– April 16, Thursday 14-16: Here we should discuss the Demo 3.

What is the aim of the 4th week, what is the main goal to reach?

• Notion of an absorbing state and applications
  – computation of the probability to reach an absorbing state with the probability generating function (Propositions 3.3.6 and 3.3.7).
  – When a system is sub-critical, critical, and super-critical (Definition 3.3.8).
  – Examples 3.3.9, 3.3.10, and 3.3.11, where the Example 3.3.11 contains some heuristics and approximations typical in applications (and which you can accept as there are presented).

• Communicating states
  – Definition 3.3.12
  – Decomposition of the state space and notion of an irreducible Markov chain (Proposition 3.3.14 and Definition 3.3.15).

• Reid-Landau model in Section 3.3.4.
The next proposition is crucial to compute the probability \( q \) to reach an absorbing state in a branching process. In Item (4) we obtain for \( q \) a theoretical fix-point equation \( q = F(q) \), which can be solved in different ways, also by numerical methods:

**Proposition 3.3.6.** For \( q_i = \mathbb{P}(f_i = 0) \) where \( i = 0, 1, \ldots \) one has the following:

1. \( 0 = q_0 < p_0 = q_1 \leq q_2 \leq \cdots \leq 1 \).
2. \( q_{i+1} = F(q_i) \) for \( i = 0, 1, \ldots \)
3. For \( q := \lim_{i \to \infty} q_i \) one has
   \[
   \mathbb{P}(\{\omega \in \Omega : \exists i \text{ such that } f_i(\omega) = 0\}) = q.
   \]
4. \( q = F(q) \).
5. \( F \) is continuous, increasing, convex and satisfies \( F(0) = p_0 \) and \( F(1) = 1 \).

**Proof.** (1) We have that \( f_i(\omega) = 0 \) implies \( f_{i+1}(\omega) = 0 \) since 0 is absorbing. Hence
   \[ q_i = \mathbb{P}(f_i = 0) \leq \mathbb{P}(f_{i+1} = 0) = q_{i+1}. \]
Moreover, we have \( q_0 = 0 \) since \( f_0 = 1 \).
(2) follows from \( q_{i+1} = F_{i+1}(0) = F(F_i(0)) = F(q_i) \).
(3) By Remark 1.2.6 (continuity from below) and \( \{f_n = 0\} \subseteq \{f_{n+1} = 0\} \) we get
   \[
   \mathbb{P}(\{\omega \in \Omega : \exists n \text{ such that } f_n(\omega) = 0\}) = \mathbb{P}\left( \bigcup_{n=1}^{\infty} \{f_n = 0\} \right) = \lim_{N \to \infty} \mathbb{P}(f_N = 0) = \lim_{N \to \infty} q_N.
   \]
(4) The function \( F \) is continuous (polynomial). Hence
   \[ q = \lim_{i \to \infty} q_{i+1} = \lim_{i \to \infty} F(q_i) = F(\lim_{i \to \infty} q_i) = F(q). \]
(5) For \( 0 \leq \theta < 1 \) one has that
   \[
   F'(\theta) = \sum_{k=1}^{\infty} k \theta^{k-1} p_k \geq 0,
   \]
so that \( F \) is increasing. From \( F''(\theta) \geq 0 \) it follows that \( F \) is convex. \( \square \)
3.3. CLASSIFICATION OF STATES

By the next proposition we obtain more necessary information about \( q \):

**Proposition 3.3.7.** Assume that \( \lim_{\theta \uparrow 1} F'(\theta) = F'(1) < \infty \). Then one has the following:

1. If \( F'(1) \leq 1 \), then the only solution to \( F(q) = q \) is \( q = 1 \).
2. If \( F'(1) > 1 \), then there are two solutions \( 0 < z_1 < z_2 = 1 \) to \( F(q) = q \). The solution of our problem is given by \( q = z_1 \).

The interpretation is as follows: It holds

\[
F'(1) = \sum_{k=1}^{\infty} kp_k = \mathbb{E}f_1.
\]

If \( \mathbb{E}f_1 \leq 1 \), then the system does not grow fast enough and dies after some time with probability 1.

**Definition 3.3.8.** We call the system

| Sub-critical if \( \mathbb{E}f_1 < 1 \), |
| Critical if \( \mathbb{E}f_1 = 1 \), |
| Super-critical if \( \mathbb{E}f_1 > 1 \). |

**Example 3.3.9 (Branching into at most 2 particles).** Let \( p_0 \in (0, 1), \ p_1 = 0, \ p_2 = 1 - p_0 \). This implies

\[
F(\theta) = p_0 + (1 - p_0)\theta^2,
F'(\theta) = 2(1 - p_0)\theta,
F'(1) = 2(1 - p_0).
\]

Hence we have the following cases:

- Sub-critical if \( 2(1 - p_0) < 1 \) if \( \frac{1}{2} < p_0 < 1 \).
- Critical if \( 2(1 - p_0) = 1 \) if \( p_0 = \frac{1}{2} \).
- Super-critical if \( 2(1 - p_0) > 1 \) if \( 0 < p_0 < \frac{1}{2} \).
The computation of $q$ in the super-critical case yields to:

$$z = F(z) \iff z = p_0 + (1 - p_0)z^2 \iff z = \begin{cases} 1 \\ \frac{1 - \sqrt{1 - 4p_0(1 - p_0)}}{2(1 - p_0)} \end{cases}$$

So we get that $q = \frac{1 - \sqrt{1 - 4p_0(1 - p_0)}}{2(1 - p_0)}$.

**Example 3.3.10** (Preservation of family names (Bienaymé, 1845)). We consider the following model:

<table>
<thead>
<tr>
<th>generation</th>
<th>0</th>
<th>1 father</th>
<th>$f_0 = 1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>number of sons of this father</td>
<td>$f_1$</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>number of sons of the first generation</td>
<td>$f_2$</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td>number of sons of the second generation</td>
<td>$f_3$</td>
</tr>
<tr>
<td>...</td>
<td></td>
<td>...</td>
<td>...</td>
</tr>
</tbody>
</table>

We assume

- $\mathbb{P}(\text{a father has no son}) = p \in (0, 1)$,
- $\mathbb{P}(\text{a father has 1 son}) = pa$,
- $\mathbb{P}(\text{a father has 2 sons}) = pa^2$,
- $\mathbb{P}(\text{a father has } k \text{ sons}) = pa^k$

with $a = 1 - p$, because of $p + pa + pa^2 + \cdots = p\frac{1}{1-a} = 1$.

Now the **Question** is how to compute

$$q = \mathbb{P}(\text{a family name dies}) = \mathbb{P}(\{\omega : \exists i : f_i(\omega) = 0\})?$$

We use the **probability generating function** $F : [0, 1] \to [0, 1]$,

$$F(\theta) = p + \theta(pa) + \theta^2(pa^2) + \cdots + \theta^k(pa^k) + \cdots = p\frac{1}{1 - \theta a},$$

and obtain

$$F'(\theta) = \frac{ap}{(1 - \theta a)^2} \quad \text{and} \quad F'(1) = \frac{ap}{(1 - a)^2} = \frac{a}{p}.$$

We have the following cases:

- $a < p$ $\iff$ $\frac{1}{2} < p < 1$ $\iff$ sub-critical
3.3. **CLASSIFICATION OF STATES**

\[
a = p \iff p = \frac{1}{2} \iff \text{critical}
\]
\[
a > p \iff 0 < p < \frac{1}{2} \iff \text{super-critical}
\]

By the help of Proposition 3.3.7 we may compute \( q \):

- For \( a \leq p \) one has \( q = 1 \).
- For \( a > p \) we get

\[
F(q) = q \iff \frac{p}{1 - qa} = q \iff q^2 - \frac{1}{a}q + \frac{p}{a} = 0.
\]

Hence

\[
q = \frac{1}{2a} - \sqrt{\frac{1}{4a^2} - \frac{p}{a}} = \frac{p}{1-p}.
\]

**Example 3.3.11 (Survival of mutant genes).** This example is from [5]. We need the following facts from genetics:

- In diploid creatures chromosomes are grouped into a fixed number of pairs of homologous chromosomes (for example 23 pairs in a human being), where always one comes from the father and the other one from the mother.

- A gene is a segment of the DNA which carries the information how to build a certain protein.
• On one **locus** (this is a certain location of a gene in the DNA) can be different versions of a gene, which are called **alleles**.

We will consider only one locus and assume that in the past was only the **AA** allele combination (called **genotype**) existent. Let the population consist of **$N_0$** individuals. In generation 0 happened a mutation **$A \rightarrow a$**. So 1 individual has now **$Aa$** while **$N_0 - 1$** have **$AA$**.

Let **$\mu > 0$** be the **relative fitness** (i.e., the average number of children or progenies per generation) of the genotype **$AA$** and **$\mu(1+s)$** the relative fitness of the genotype **$Aa$**.

Then we get the frequency of the **$a$** allele in the first generation:

$$\frac{\# \{ a \text{ alleles} \}}{\# \{ \text{all alleles} \}} = \frac{\mu(1+s)}{2(N-1)\mu + 2\mu(1+s)}$$

Note that the number of alleles in generation 1 is double of the number of species in generation 1. This leads to the following model (but we will not justify this in detail):

**Model:** If one has a generation with **$N$** individuals, then - assuming random mating and that **$N$** is large - yields to the Ansatz:

$$p_j = \text{IP}\left( j \text{ } a \text{ alleles in the next generation} \right) = \binom{2N}{j} \left( \frac{1+s}{2N} \right)^j \left( 1-\frac{1+s}{2N} \right)^{2N-j}.$$ 

By Poisson’s Theorem (see, for example, [4]) we can approximate

$$p_j \approx e^{-\left(1+s\right)}\frac{(1+s)^j}{j!}, \quad j = 0, 1, 2, \ldots$$

i.e. use the Poisson distribution and we continue **with this approximation**. If we assume that there is no further mutation we are in the situation of a branching process. The probability generating function is

$$F(\theta) = \sum_{j=0}^{\infty} \theta^j e^{-\left(1+s\right)}\frac{(1+s)^j}{j!} = e^{-\left(1+s\right)}e^{\theta(1+s)} = e^{(\theta-1)(1+s)}$$

so that

$$F'(\theta) = (1+s)e^{(\theta-1)(1+s)},$$
3.3. CLASSIFICATION OF STATES

\[
F'(1) = 1 + s = \begin{cases} 
\text{sub-critical} & \iff s < 0 \\
\text{critical} & \iff s = 0 \\
\text{super-critical} & \iff s > 0
\end{cases}
\]

Now we estimate \( q \) for the case \( s > 0 \) with some approximations we do not justify here: Since it holds by Taylor expansion

\[
q = e^{(1+s)(q-1)} = \sum_{l=0}^{\infty} \frac{F^{(l)}(1)}{l!} (q - 1)^l
\]

\[
= 1 + (1 + s)(q - 1) + \frac{(1 + s)^2}{2} (q - 1)^2 + \ldots
\]

\[
\approx 1 + (q - 1) + s(q - 1) + \frac{1 + 2s + s^2}{2} (q - 1)^2,
\]

we get for small \( s \) if we consider only the 'first order terms' that

\[
q \approx 1 + (q - 1) + s(q - 1) + \frac{(q - 1)^2}{2}
\]

which implies

\[
q \approx 1 - 2s.
\]

3.3.3 Communicating states

The communicating states in a Markov chain allow a clustering of the chain into sub-groups to understand the structure of the Markov chain better. As described later in an example communicating states might describe different degrees of a disease where the patient can move from any state to any other one from one day to the following day (either he recovers step by step, or there is an exacerbation of the disease).

**Definition 3.3.12.** Let \((f_i)_{i=0}^{\infty}\) be a homogeneous Markov chain with state space \(X\) and let \(k, l \in X\).

1. We say that **k reaches l** \((k \rightarrow l) :\iff\)

   \[
   \exists i \geq 0 \text{ such that } \mathbb{P}(f_i = l | f_0 = k) > 0.
   \]

2. We say that **k and l communicate** with each other \((k \leftrightarrow l) :\iff\)

   \[
   k \rightarrow l \text{ and } l \rightarrow k.
   \]
CHAPTER 3. DISCRETE TIME MARKOV CHAINS

Proposition 3.3.13. The relation $\leftrightarrow$ is an equivalence relation on $X$, that means:

1. Reflexivity: $k \leftrightarrow k$
2. Symmetry: $k \leftrightarrow l \implies l \leftrightarrow k$
3. Transitivity: $k \leftrightarrow l$ and $l \leftrightarrow m \implies k \leftrightarrow m$

Proof. (1) follows from $\mathbb{P}(f_0 = k | f_0 = k) = 1 > 0$ as the definition allows $i = 0$.

(2) follows directly from the definition as well as because $k \to l$ and $l \to k$ is the same as $l \to k$ and $k \to l$.

(3) Let us see why it is transitive. It is enough to show

$$(k \to l \text{ and } l \to m) \implies k \to m.$$ 

We choose $i, j \geq 0$ such that $\mathbb{P}(f_i = l | f_0 = k)\mathbb{P}(f_j = m | f_0 = l) > 0$. By homogeneity, Markov property and the definition of conditional probability we can continue with

$$
\begin{align*}
\mathbb{P}(f_i = l | f_0 = k) & \mathbb{P}(f_j = m | f_0 = l) \\
& = \mathbb{P}(f_i = l | f_0 = k) \mathbb{P}(f_{j+i} = m | f_i = l) \\
& = \mathbb{P}(f_i = l | f_0 = k) \mathbb{P}(f_{j+i} = m | f_i = l, f_0 = k) \\
& = \mathbb{P}(f_{j+i} = m, f_i = l | f_0 = k) \\
& \leq \mathbb{P}(f_{j+i} = m | f_0 = k)
\end{align*}
$$

which implies $k \to m$. \hfill \Box

The next proposition is crucial to separate a Markov chain into different parts that do not interact with each other.

Proposition 3.3.14. The state space of a homogeneous Markov chain $(f_i)_{i=0}^{\infty}$ can be portioned into equivalence classes with respect to $\leftrightarrow$, that means

$$X = \bigcup_{m=1}^{L} X_m \text{ or } X = \bigcup_{m=1}^{\infty} X_m$$

such that
3.3. CLASSIFICATION OF STATES

(1) \( X_{m_1} \cap X_{m_2} = \emptyset \) for \( m_1 \neq m_2 \).

(2) if \( k, l \in X_m \), then \( k \leftrightarrow l \),

(3) if \( k \leftrightarrow l \), then \( k \) and \( l \) belong to the same \( X_m \).

We will not prove this proposition formally, but the statement should be intuitive obvious. The next definition describes the extreme case, that one has only one component:

**Definition 3.3.15.** A homogeneous Markov chain \((f_i)_{i=0}^{\infty}\) is called **irreducible** \(\iff\) There is only one equivalence class with respect to \(\leftrightarrow\).

### 3.3.4 Reid-Landau model for radiation damage

In this example we demonstrate how to use the concept of absorbing states and communicating states. Let \( X = \{0, 1, \ldots, K\} \). Then \( l \in X \) stands for the degree of radiation damage of one fixed person as follows:

\[
\begin{align*}
0 & = \text{healthy} \\
1 & = \text{initial damage} \\
2 & \quad \vdots \\
K - 1 & = \text{non-visible stages} \\
K & = \text{visible stage}
\end{align*}
\]

The Markov chain \((f_i)_{i=0}^{\infty}\) starts at \( f_0 = 1 \) as initial damage. The behavior is as follows:

\[
\begin{align*}
k \to k & = \text{staying at the same stage of damage} \\
k \to k - 1 & = \text{recovery} \\
k \to k + 1 & = \text{amplification}
\end{align*}
\]

We assume the following transition matrix:

\[
T = \begin{pmatrix}
1 & 0 & 0 & 0 & \ldots & 0 & 0 & 0 \\
q_1 & r_1 & p_1 & 0 & \ldots & 0 & 0 & 0 \\
0 & q_2 & r_2 & p_2 & 0 & \ldots & 0 & 0 \\
& \vdots \\
0 & 0 & \ldots & 0 & q_{K-1} & r_{K-1} & p_{K-1} \\
0 & 0 & \ldots & 0 & 0 & 0 & 1
\end{pmatrix}
\]
Proposition 3.3.16 (Reid-Landau model). We assume that
\[ q_l = 1 - \frac{l}{K}, \quad \text{recovery probability}, \]
\[ r_l = 0, \quad \text{and} \]
\[ p_l = \frac{l}{K}, \quad \text{amplification probability} \]
for \( l = 1, \ldots, K - 1 \). Then the following holds:

1. The absorbing states consist of \( \{0, K\} \).
2. The chain consists of the 3 equivalence classes \( \{0\}, \{1, \ldots, K - 1\}, \{K\} \).
3. The recovery probability is
\[ \mathbb{P}(\{\omega \in \Omega : \exists i : f_i(\omega) = 0\}) = 1 - \frac{1}{2^{K-1}}. \]

Proof. (1) follows because these are the only states that cannot be left by the chain as \( p(0, 0) = p(K, K) = 1 \).

(2) From (1) it follows that \( \{0\} \) and \( \{K\} \) are clusters in the sense of Proposition 3.3.14. Checking the remain set set \( \{1, \ldots, K - 1\} \) one sees that with positive probability the chain can move from any state to any other state.

(3) Choose \( K = 3 \). Then the transition matrix has the form
\[ T = \begin{pmatrix}
1 & 0 & 0 & 0 \\
2 & 0 & \frac{1}{3} & 0 \\
0 & \frac{1}{3} & 0 & \frac{2}{3} \\
0 & 0 & 0 & 1
\end{pmatrix}. \]

We have
\[ \mathbb{P}(\{\omega : \exists i : f_i(\omega) = 0\}) = p_{10} \sum_{n=0}^{\infty} \mathbb{P}(f_n = 1|f_0 = 1) \]
\[ = p_{10} \sum_{n=0}^{\infty} \mathbb{P}(f_n = 1|f_0 = 1) \text{ for even} \]
\[ = p_{10} \sum_{n=0}^{\infty} \left(\frac{1}{3}\right)^{2n} = \frac{2}{3} \frac{1}{1 - \frac{1}{9}} = 1 - \frac{1}{2^2}. \]
3.3. CLASSIFICATION OF STATES

For $K = 4$ we obtain

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

and

$$\mathbb{P}(\{\omega : \exists i : f_i(\omega) = 0\}) = p_{10} \left( 1 + \sum_{m=1}^{\infty} \mathbb{P}\left( \text{way from } 1 \to 1 \text{ which passes } 1 \to 2 \text{ exactly } m \text{ times} \right) \right)$$

$$= p_{10} \left( 1 + \sum_{m=1}^{\infty} \mathbb{P}\left( \text{way from } 1 \to 1 \text{ which passes } 1 \to 2 \text{ exactly one times} \right)^m \right)$$

$$= p_{10} \frac{1}{1 - \mathbb{P}\left( \text{way from } 1 \to 1 \text{ which passes } 1 \to 2 \text{ exactly one times} \right)}$$

$$= \frac{p_{10}}{1 - p_{12}p_{21} \left( 1 + \mathbb{P}\left( \text{way from } 2 \to 2 \text{ which does not pass } 1 \right) \right)}$$,

where

$$\mathbb{P}\left( \text{way from } 2 \to 2 \text{ which does not pass } 1 \right) = p_{23}p_{32} + (p_{23}p_{32})^2 + \cdots = \frac{p_{23}p_{32}}{1 - p_{23}p_{32}}.$$

Hence

$$\mathbb{P}(\{\omega : \exists i : f_i(\omega) = 0\}) = \frac{p_{10}}{1 - p_{12}p_{21} \left( \frac{1}{1 - p_{23}p_{32}} \right)} = 1 - 2^{-3}.$$

For $K = 5$ we have to replace $p_{23}p_{32}$ by $p_{23}p_{32} \left( \frac{1}{1 - p_{34}p_{43}} \right)$ and so on.
Checkup after reading:

- You should be able to give without notes the definition of absorbing states and communicating states.
- Can an absorbing state be part of a pair of communicating states?
- Try to sketch a proof for Proposition 3.3.14.
- For which $s$ in Example 3.3.11 a mutation will disappear over time?
Week 5  20.04. – 24.04.2020
Again I received many solved exercises from you and this is good!
Thank you, please continue like that.

I am in Teams (or skype if you prefer):

– April 22, Wednesday 10-12: Here you can ask all your questions regarding the lecture.

– April 23, Thursday 14-16: Here we should discuss the Demo.

What is the aim of the 5th week, what is the main goal to reach?

• Knowledge of periodic and aperiodic states of a Markov chain.
• Knowledge of persistent and transient states of a Markov chain.
• Irreducible and closed subsets of the state space of a Markov chain.
• Decomposition of the state space as described in Proposition 3.3.28.
3.3.5 Periodic and aperiodic states

In this section we investigate states where a homogeneous Markov chain returns with a positive probability after a certain positive time. In the Reid-Landau model these states are the damages of degrees $1, \ldots, K - 1$. In contrast to that, the states $0, K$ (healthy, visible damage) are a-periodic as these states are absorbing.

Let us start by recalling that we defined for a homogeneous Markov chain $(f_i)_{i=0}^{\infty}$ the probabilities

$$p^{(i)}(k, l) := \mathbb{P}(f_i = l|f_0 = k) \quad \text{for} \quad k, l \in X.$$ 

**Definition 3.3.17.** Let $(f_i)_{i=0}^{\infty}$ be a homogeneous Markov chain with state space $X$.

1. A state $k \in X$ has **period** $d \in \{1, 2, \ldots\}$ if and only if
   
   (a) $p^{(i)}(k, k) = 0$ if $i = jd + r$ where $j = 0, 1, \ldots$ and $r = 1, \ldots, d - 1$

   (b) $d$ is the largest number satisfying this property

   We denote the period of $k$ by $d(k)$.

2. A state $k \in X$ is called **aperiodic** if and only if $d(k) = 1$.

3. If $p^{(i)}(k, k) = 0$ for all $i = 1, 2, \ldots$, then we let $d(k) := \infty$.

Next we give an interpretation of the period of a state and show that the period is exactly defined. Regarding the definition we have to check that if $p^{(i)}(k, k) = 0$ for all $i = 1, 2, \ldots$ then there does not exists a finite $d$ as in Definition 3.3.17(1).

**Remark 3.3.18.**

1. If $d(k) = d < \infty$ and if $f_i = k$, then the Markov chain can (but does not necessarily have to) only return to the state $k$ at times $i + d, i + 2d, i + 3d, \ldots$

2. If there exists some $i_0 \in \{1, 2, \ldots\}$ such that $p^{(i_0)}(k, k) > 0$, then

   $$d(k) \leq i_0 < \infty.$$ 

This one can be verified as follows: We fix $k \in X$ and let

$$\mathcal{M} := \{d \in \{1, 2, \ldots\} : p^{(i)}(k, k) = 0 \text{ for all } i = jd + r\}$$
In other words, we collect in \( \mathcal{M} \) all finite \( d \in \{1, 2, \ldots \} \) such that the Markov chain can only return to the state \( k \) at times \( i + d, i + 2d, i + 3d, \ldots \) if \( f_i = k \). Then we have the following:

- \( 1 \in \mathcal{M} \) because the chain can return at times \( i + 1, i + 2, i + 3, \ldots \)
- Assume \( d \in \mathcal{M} \) with \( d > i_0 \). Then

\[
0 = p^{(1)}(k, k) = \cdots = p^{(i_0)}(k, k) = \cdots = p^{(d-1)}(k, k),
\]

which is a contradiction to \( p^{(i_0)}(k, k) > 0 \).

(3) From (2) it follows that for all \( k \in X \) the period \( d(k) \in \{1, 2, \ldots \} \cup \{\infty\} \) is correctly defined.

In the next proposition we show that communicating states have the same period. This is not unexpected, but requires a proof.

**Proposition 3.3.19.** Let \( k, l \in X \) such that \( k \leftrightarrow l \), i.e. the states \( k \) and \( l \) communicate. Then

\[
d(k) = d(l).
\]

**Proof.** As \( k \) and \( l \) are communicating states, we find \( m, n \geq 0 \) such that \( p^{(m)}(k, l) > 0 \) and \( p^{(n)}(l, k) > 0 \). Assume that \( d(l) < \infty \) and let \( s_1, s_2 \geq 1 \) be such that \( p^{(s_1)}(l, l) > 0 \) and \( p^{(s_2)}(l, l) > 0 \). Then \( s \in \{d(l), 2d(l), 3d(l), \ldots \} \).

Without rigorous proof we use the observation, that \( 1 \leq s_1 < s_2 < \infty \) can be chosen such that \( s_2 - s_1 = d(l) \) (assuming that this is not true, then the period would be at least \( 2d(l) \) which is a contradiction. So we get

\[
p^{(m+n+s_1)}(k, k) = p^{(m)}(k, l)p^{(s_1)}(l, l)p^{(n)}(l, k) > 0,
\]

\[
p^{(m+n+s_2)}(k, k) = p^{(m)}(k, l)p^{(s_2)}(l, l)p^{(n)}(l, k) > 0.
\]

Hence \( d(k) \leq m + n + s_1 \), and \( m + n + s_1 \) and \( m + n + s_2 \) are multiples of \( d(k) \). Therefore \( d(l) = s_2 - s_1 \) is a multiple of \( d(k) \) as well.

(b) We can interchange in (a) the states \( k \) and \( l \) and obtain that \( d(k) \) is divisible by \( d(l) \).

(c) The steps (a) and (b) imply that \( d(l) = d(k) \). \( \square \)
CHAPTER 3. DISCRETE TIME MARKOV CHAINS

Example 3.3.20. For the model about radiation damage from Reid/Landau (Section 3.3.4) we have
\[ d(0) = 1, \quad d(K) = 1, \quad d(l) = 2 \quad \text{for} \quad l \in \{1, \ldots, K - 1\}. \]

Example 3.3.21. Consider a Markov chain with transition matrix
\[ T = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1/2 & 0 & 0 & 1/2 \end{pmatrix} \]
Then \( d(k) = 1 \) for \( k = 0, \ldots, 3 \).

3.3.6 Persistent and transient states

In this section of classify the states according to possible returns to the state the Markov chain was starting from.

Definition 3.3.22. Let \((f_i)_{i=0}^{\infty}\) be a homogeneous Markov chain with state space \( X \).

(1) A state \( k \in X \) is called **persistent** (or **recurrent**) if and only if
\[ \mathbb{P}(\exists i \geq 1 \text{ with } f_i = k | f_0 = k) = 1. \]

(2) A state \( k \) is called **transient** if and only if \( k \) is not persistent.

(3) The expression
\[ f_{kj}^{(n)} := \mathbb{P}(f_n = j, f_{n-1} \neq j, \ldots, f_1 \neq j | f_0 = k) \]
is called the **first passage distribution from state** \( k \) **to state** \( j \).

(4) Finally we define
\[ f_{kj} := \sum_{n=1}^{\infty} f_{kj}^{(n)}. \]

The above definition has the following **Interpretation:**

(a) \( k \) is persistent \(\iff\) with probability 1 the Markov chain returns to \( k \).
(b) $k$ is transient $\iff$ with positive probability there is no return to the state $k$.

The important characterization of a persistent state is as follows:

**Proposition 3.3.23.** If for $k \in X$ we let

$$p^{(n)}(k, k) := \mathbb{P}(f_n = k | f_0 = k),$$

then the following assertions are equivalent:

1. $k$ is persistent.
2. $\sum_{n=1}^{\infty} p^{(n)}(k, k) = \infty.$
3. $f_{kk} = 1.$

For the proof of this proposition we need the following lemma:

**Lemma 3.3.24.** Let $N(k) \geq 1$ be the number of returns of the Markov chain $(f_i)_{i=0}^{\infty}$ to $k \in X$ after step 0 and let $s \in \{1, 2, 3, \ldots \}$. Then one has

$$\mathbb{P}(N(k) \geq s | f_0 = k) = (f_{kk})^s.$$

**Proof.** We prove the statement only for the case $s = 2$. Using the Markov property and the homogeneity we get

$$\mathbb{P}(N(k) \geq 2 | f_0 = k)$$

$$= \mathbb{P}(\exists i > j \geq 1 : f_i = f_j = k | f_0 = k)$$

$$= \sum_{i > j \geq 1} \mathbb{P}(f_i = k, f_{i-1} \neq k, \ldots, f_{j+1} \neq k, f_j = k, f_{j-1} \neq k, \ldots, f_1 \neq k | f_0 = k)$$

$$= \sum_{i > j \geq 1} \mathbb{P}(f_i = k, f_{i-1} \neq k, \ldots, f_{j+1} \neq k | f_j = k)$$

$$\times \mathbb{P}(f_j = k, f_{j-1} \neq k, \ldots, f_1 \neq k | f_0 = k)$$

$$= \sum_{i > j \geq 1} \mathbb{P}(f_{i-j} = k, f_{i-j-1} \neq k, \ldots, f_1 \neq k | f_0 = k)$$

$$\times \mathbb{P}(f_j = k, f_{j-1} \neq k, \ldots, f_1 \neq k | f_0 = k)$$

$$= \sum_{i,j \geq 1} \mathbb{P}(f_i = k, f_{i-1} \neq k, \ldots, f_1 \neq k | f_0 = k)$$

$$\times \mathbb{P}(f_j = k, f_{j-1} \neq k, \ldots, f_1 \neq k | f_0 = k).$$
Now we can prove Proposition 3.3.23:

**Proof of Proposition 3.3.23.** It holds that

\[ P(\exists i \geq 1 \text{ with } f_i(\omega) = k | f_0(\omega) = k) \]

\[ = \sum_{n=1}^{\infty} P(f_n = k, f_{n-1} \neq k, \ldots, f_1 \neq k | f_0 = k) = f_{kk}. \]

Hence that \( k \) is persistent is equivalent to \( f_{kk} = 1 \).

Assume now that \( P(f_0 = k) = 1 \). The expression

\[ N(k)(\omega) = \sum_{n=1}^{\infty} \mathbb{1}_{\{f_n = k\}}(\omega) \]

is a random variable. We have

\[ E N(k) = \sum_{n=1}^{\infty} E \mathbb{1}_{\{f_n = k\}} = \sum_{n=1}^{\infty} P(f_n = k | f_0 = k) = \sum_{n=1}^{\infty} p^{(n)}(k, k). \quad (3.3) \]

We compute

\[ \sum_{s=1}^{\infty} P(N(k) \geq s | f_0 = k) \]

\[ = P(N(k) \geq 1 | f_0 = k) + P(N(k) \geq 2 | f_0 = k) + P(N(k) \geq 3 | f_0 = k) + \ldots \]

\[ = P(N(k) = 1 \text{ or } N(k) \geq 2 | f_0 = k) + P(N(k) \geq 3 | f_0 = k) + \ldots \]

\[ = P(N(k) = 1 | f_0 = k) + 2P(N(k) = 2 | f_0 = k) + 3P(N(k) = 3 | f_0 = k) + \ldots \]

\[ = \sum_{m=1}^{\infty} m P(N(k) = m | f_0 = k) = E N(k). \quad (3.4) \]
3.3. **CLASSIFICATION OF STATES**

Hence, by Lemma 3.3.24,

\[ \mathbb{E} N(k) = \sum_{s=1}^{\infty} \mathbb{P}(N(k) \geq s | f_0 = k) = \sum_{s=1}^{\infty} (f_{kk})^s. \]

Comparing this with (3.3) gives

\[ \sum_{n=1}^{\infty} p^{(n)}(k, k) = \sum_{s=1}^{\infty} (f_{kk})^s, \]

which implies

\[ 0 \leq f_{kk} < 1 \iff \sum_{n=1}^{\infty} p^{(n)}(k, k) < \infty. \]

\[ \square \]

**Example 3.3.25 (Tossing a coin).** Consider the Markov chain \((f_i)_{i=0}^{\infty}\) with \(f_0 = 0\) and state space \(X := \{ -3, -2, -1, 0, 1, 2, 3, \ldots \}\) and for a fixed \(p \in (0, 1)\) the transition probabilities

\[ \mathbb{P}(f_{i+1} = k + 1 | f_i = k) = 1 - p, \]
\[ \mathbb{P}(f_{i+1} = k - 1 | f_i = k) = p. \]

The ** Interpretation** is as follows:

(a) Tossing a coin gives 'heads' or 'tails'

(b) We start with \(f_0 = 0\).

(c) If one tosses the coin, then in case of

'heads' \(\rightarrow f_{i+1}(\omega) = f_i(\omega) + 1 \) (gain),

'tails' \(\rightarrow f_{i+1}(\omega) = f_i(\omega) - 1 \) (loss).

The coin is 'fair' if and only if 0 is a persistent state: By Proposition 3.3.23 the state 0 is persistent if and only if

\[ \sum_{n=1}^{\infty} p^{(n)}(0, 0) = \infty. \]
So, we will compute \( p^{(n)}(0, 0) \):
First we observe that \( p^{(n)}(0, 0) = 0 \) for all odd \( n \). For even \( n \) it holds
\[
p^{(n)}(0, 0) = \left( \frac{n}{n} \right) p^{\frac{n}{2}} (1 - p)^{\frac{n}{2}} = \left( \frac{n}{n} \right) (p(1 - p))^{\frac{n}{2}}.
\]

Stirling’s formula
\[
\lim_{m \to \infty} \frac{m!}{\sqrt{2\pi m}} e^{-m} m^m = 1
\]
gives
\[
p^{(2m)}(0, 0) = \frac{(2m)!}{(ml)^2} (p(1 - p))^m
\]
\[
\sim \frac{\sqrt{4\pi m} e^{-2m} (2m)^2}{(2\pi m e^{-m} m^m)^2} (p(1 - p))^m
\]
\[
= \frac{2^{2m}}{\sqrt{\pi m}} (p(1 - p))^m.
\]

Therefore 0 is persistent if and only if \( \sum_{n=1}^{\infty} p^{(n)}(0, 0) = \infty \) which is equivalent for any \( m_0 \geq 1 \) to
\[
\sum_{m=m_0}^{\infty} \frac{2^{2m}}{\sqrt{\pi m}} (p(1 - p))^m = \frac{1}{\sqrt{\pi}} \sum_{m=m_0}^{\infty} \frac{(4p(1 - p))^m}{\sqrt{m}} = \infty \iff 4p(1 - p) \geq 1
\]
\[
\iff p = \frac{1}{2}.
\]

Finally, recall that the coin is ‘fair’ if and only if \( p = \frac{1}{2} \).

### 3.3.7 Decomposition of the state space

**Definition 3.3.26.** Let \( ((f_i)_{i=0}^{\infty}, p_o, T) \) be a homogeneous Markov chain with state space \( X \).

1. A subset \( C \subseteq X \) is called **irreducible** \( : \iff k \leftrightarrow l \) for all \( k, l \in C \).
2. A subset \( C \subseteq X \) is called **closed** \( : \iff \forall k \in C \ \forall l \in X \setminus C \ \forall n = 1, 2, \ldots \ \ p^{(n)}(k, l) = 0 \).
3.3. CLASSIFICATION OF STATES

Remark 3.3.27. (1) If the Markov chain arrives in a closed subset $C$, then the chain will stay there forever.

(2) If $k \in X$ is an absorbing state, then $C = \{k\} \subseteq X$ is closed.

Now we formulate and prove a fundamental decomposition of the state space of a Markov chain:

Proposition 3.3.28. Let $((f_i)_{i=0}^{\infty}, p_\alpha, T)$ be a homogeneous Markov chain with state space $X$ Let

$$X_T := \{k \in X : k \text{ transient}\},$$

$$X_P := \{k \in X : k \text{ persistent}\}.$$

Then the following holds:

(1) $X = X_T \cup X_P$ with $X_T \cap X_P = \emptyset$

(2) $X_P = \bigcup_m C_m$, where the $(C_m)_m$ are pair-wise disjoint, irreducible and closed.

Proof. We only need to show

$$X_P = \bigcup_m C_m.$$

Step 1: Fix $k \in X_P$ and let

$$C(k) := \{l \in X : k \to l\}.$$

Then it holds

- $k \in C(k)$: it always holds that $k \leftrightarrow k$.
- $C(k)$ is closed: $m \in C(k), l \in X \setminus C(k)$

Assume that there is an $n \in \{1, 2, \ldots\}$ such that

$$p^{(n)}(m, l) > 0.$$

Then $m \to l$. Because of $k \to m$ one gets $k \to m \to l$ and $k \to l$, so that $l \in C(k)$. This is a contradiction.
• \( C(k) \) is irreducible: Let \( l, m \in C(k) \).

\[
\begin{align*}
k \to l & \text{ by definition } \\
l \to k & \text{ by persistence}
\end{align*}
\]

\( \implies k \leftrightarrow l. \)

In the same way we get \( k \leftrightarrow m \). Hence \( l \leftrightarrow m \).

**Step 2:** For \( k, l \in X \) we show that either \( C(k) = C(l) \) or \( C(k) \cap C(l) = \emptyset \). Let us assume \( m \in C(k) \cap C(l) \). Then for all \( \tilde{m} \in C(l) \) we have \( k \to m \leftrightarrow \tilde{m} \) because \( C(l) \) is irreducible. This implies \( k \to \tilde{m} \) for all \( \tilde{m} \in C(l) \). Hence

\[
C(l) \subseteq C(k).
\]

The other inclusion \( C(k) \subseteq C(l) \) can be proved in the same way.

3.3.8 **Summary of the classification**

In the next two subsections we summarize facts about the classification of the states of a Markov chain.
3.3. CLASSIFICATION OF STATES

(A) Classification according to arithmetic properties of the transition probabilities $p^{(n)}(k, l)$

(A1) $k$ is an absorbing state if

$$\mathbb{P}(f_{i+1} = k | f_i = k) = p(k, k) = 1 \quad \text{if} \quad \mathbb{P}(f_i = k) > 0.$$  

(A2) $k \rightarrow l$ (k reaches l) if

$$\exists n \geq 0 \text{ such that } \mathbb{P}(f_n = l | f_0 = k) = p^{(n)}(k, l) > 0.$$  

(A3) $k \leftrightarrow l$ (k and l are communicating) if

$$\exists m, n \geq 0 \text{ such that } p^{(n)}(k, l) > 0 \text{ and } p^{(m)}(l, k) > 0.$$  

(A4) $k$ has period $d(k)$ if $d(k)$ is the largest number $m$ for which it holds

$$p^{(n)}(k, k) = 0 \text{ for all } n \text{ which are not divisible by } m.$$  

– This means that the Markov chain may only return after a time that is a multiple of $d(k)$.

– If $d(k) = 1$, then we call the state $k$ aperiodic.

(A5) Decomposition of the state space $X$ with respect to $\leftrightarrow$ one gets equivalence classes

$$X = \bigcup_{m=1}^{L} X_m$$  

such that

(a) $X_l \cap X_k = \emptyset$ for $l \neq k$,

(b) $k \leftrightarrow l \iff k$ and $l$ belong to the same $X_m$.  

(B) Classification according to the asymptotic properties of the transition probabilities $p^{(n)}(k, l)$

(B1) **Recurrence time:** Assume $\mathbb{P}(f_0 = k) = 1$. Then the (extended) random variable $T_k : \Omega \to \{1, 2, \ldots \} \cup \{+\infty\}$ defined by

$$T_k := \inf\{n \geq 1; f_n = k\}$$

is called *recurrence time*.

(B2) The state $k \in X$ is **persistent** (recurrent) $\iff f_{kk} = 1 = \sum_{n=1}^{\infty} f_{kk}^{(n)}$.

(B3) **Distribution of recurrence time:** Because

$$\mathbb{P}(T_k = n) = \mathbb{P}(f_n = k, f_{n-1} \neq k, \ldots, f_1 \neq k | f_0 = k) = f_{kk}^{(n)}$$

we get that the distribution of $T_k$ is given by the discrete probability on $\mathbb{N}$

$$\mathbb{P}_{T_k} = \sum_{n=1}^{\infty} f_{kk}^{(n)} \delta_n.$$  

(B4) The **mean recurrence time** is given by $\mu_k := \mathbb{E} T_k = \sum_{n=1}^{\infty} n f_{kk}^{(n)}$.

Classification the states of a Markov chain:

- **Set $X$ of all states**
  - **Persistent states** $k \in X$
    - $\sum_{n=1}^{\infty} p^{(n)}(k, k) = \infty$
  - **Transient states** $k \in X$
    - $\sum_{n=1}^{\infty} p^{(n)}(k, k) < \infty$
  - **Positive persistent states** $k \in X$
    - $\mu_k^{-1} > 0$
  - **Null persistent states** $k \in X$
    - $\mu_k^{-1} = 0$

Decomposition of the states of a Markov chain: We have

$$X = X_T \cup \bigcup_m C_m$$

where the $C_m$’s are pair-wise disjoint, irreducible and closed.
Checkup after reading:

- What is the relation between absorbing states and persistent states?
- Give an intuitive interpretation for Proposition 3.3.23(2).
- Verify Example 3.3.20.
- Verify Example 3.3.21.
- Compare all statement from the overview in Section 3.3.8 to the material before.
Week 6  27.04. – 01.05.2020

Again I received many solved exercises from you and this is good! Thank you, please continue like that.

I am in Teams (or skype if you prefer):

- April 29, Wednesday 10-12: Here you can ask all your questions regarding the lecture.

- April 30, Thursday 14-16: Here we should discuss the Demo.

- I will inform you about the exam in the beginning of the week by email.

What is the aim of the 6th (and therefore last) week, what is the main goal to reach?

- The Ergodic Theorem 3.4.5 is one of the main results of the course.

- This ergodic Theorem 3.4.5 is applied to the Gibbs-sampler in Proposition 4.4.5. The Gibbs sampler is a numerical method that can be tried when the standard Monte-Carlo method from Section 4.1 fails as the method is not feasible from the viewpoint of a numerical implementation.

- The Hard-Core-Model is described in Section 4.6 and is a slight modification of the Gibbs-sampler. Here the mathematics is not presented.
3.4 Ergodic Theorem and stationary distribution

If we have a look at the example of precipitation from Section 3.2.1, then we see for the transition matrices that

\[
T = \begin{pmatrix}
0.602 & 0.398 \\
0.166 & 0.834
\end{pmatrix}
\]

and

\[
T^5 = \begin{pmatrix}
0.305 & 0.695 \\
0.290 & 0.710
\end{pmatrix}.
\]

Observation: The rows of \(T^5\) are more similar to each other than the rows of \(T\). This observation is the starting point of the notion of ergodicity:

Definition 3.4.1. Let \(((f_i)_{i=0}^\infty, p, T)\) be a homogeneous Markov chain with state space \(X = \{0, \ldots, K\}\) and \(p(k) > 0\) for all \(k \in X\). Let

\[
T = \begin{pmatrix}
p_{00} & \cdots & p_{0K} \\
\vdots & \ddots & \vdots \\
p_{K0} & \cdots & p_{KK}
\end{pmatrix}
\]

be the transition matrix, i.e. \(p_{kl} = \mathbb{P}(f_1 = l | f_0 = k)\), and let

\[
\underbrace{T \circ \cdots \circ T}_{n \text{ times}} =: \left( p^{(n)}_{kl} \right)_{k,l=0}^K.
\]

(1) \(((f_i)_{i=0}^\infty, p, T)\) is called ergodic if and only if there exist \(s_l > 0, l = 0, \ldots, K\), such that

(a) \(s_0 + \cdots + s_K = 1\),
(b) \(\lim_{n \to \infty} p^{(n)}_{kl} = s_l\) for all \(k, l \in X\).

(2) The vector \(s = (s_0, \ldots, s_K)\) is called stationary distribution if and only if

(a) \(s_0 + \cdots + s_K = 1\) with \(0 \leq s_l \leq 1\),
(b) \(s \circ T = s\) (i.e. \(\sum_{m=0}^K s_m p_{ml} = s_l\)).

The notion ergodic means that we observe some kind of stabilization of the behaviour of the Markov chain. The name stationary distribution takes it name from the fact that the Markov chain preserves a stationary distribution in each time-step.
Chapter 3. Discrete Time Markov Chains

Proposition 3.4.2 (Ergodicity implies a unique stationary distribution). An ergodic Markov chain has a unique stationary distribution.

Proof. (a) Existence of stationary distribution: Let

\[ A := \begin{pmatrix} s_0 & \cdots & s_K \\ \vdots & \ddots & \vdots \\ s_0 & \cdots & s_K \end{pmatrix} \]

where the \( s_0, \ldots, s_K \) are taken from Definition 3.4.1. Then, by this definition,

\[ A = \lim_{n \to \infty} T^n = \left( \lim_{n \to \infty} T^n \right) \circ T = A \circ T. \]

By the special form of the matrix \( A \) this implies \( s \circ T = s \), whence \( s \) is another stationary distribution.

(b) Uniqueness of stationary distribution: Assume that \( b = (b_0, \ldots, b_K) \) is a stationary distribution. Then

\[ b \circ T = b \quad \text{and} \quad b \circ T^n = b. \]

By \( n \to \infty \) this implies \( b \circ A = b \). But that means

\[ (b_0 + \cdots + b_K)s_l = b_l \quad \text{and} \quad s_l = b_l. \]

The next example illustrates that a unique stationary distribution does not imply ergodicity:

Example 3.4.3 (Unique stationary distribution does not imply ergodicity). Let \( X := \{0, 1\} \) and \( T := \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \). Then

\[ s \circ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = s \iff s_2 = s_1 \iff s_2 = s_1 = \frac{1}{2}. \]

Hence we have a unique stationary distribution. But the Markov chain is not ergodic since

\[ T^2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \]

and consequently

\[ T^n = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ for even } n \quad \text{and} \quad T^n = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ for odd } n, \]
which means that \( \lim_{n \to \infty} P_{kl}^{(n)} \) does not exist.

Now we illustrate that a stationary distribution does not need to be unique:

**Example 3.4.4 (Stationary distribution does not need to be unique).** Let

\[
T := \begin{pmatrix}
1 & 0 & 0 \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
0 & 0 & 1
\end{pmatrix}.
\]

We compute

\[
(s_0, s_1, s_2) \circ T = s \iff s_0 + \frac{1}{3}s_1 = s_0 \iff s_1 = 0.
\]

This implies that one gets the set of stationary distributions

\[
(s_0, s_1, s_2) = (s_0, 0, 1 - s_0) \quad \text{if} \quad s_0 \in [0, 1],
\]

hence no unique one.

The main result of this section and one of the main results of this course is the following Ergodic Theorem:

**Theorem 3.4.5 (Ergodic Theorem).** Assume a homogeneous Markov chain \( ((f_i)_{i=0}^{\infty}, p, T) \) with state space \( X = \{0, \ldots, K\} \) and transition matrices

\[
T \circ \cdots \circ T \quad \text{n-times} = \left( P_{kl}^{(n)} \right)_{k,l=0}^{K}.
\]

If there exists some \( n_0 \geq 1 \) such that

\[
\inf_{k,l} P_{kl}^{(n_0)} > 0,
\]

then one has the following:

1. \( ((f_i)_{i=0}^{\infty}, p, T) \) is ergodic.
2. The stationary distribution \( (s_0, \ldots, s_K) \) satisfies \( s_l > 0 \) for all \( l \in X \).
Proof. We will prove the result for the special case \( n_0 = 1 \). It holds

\[
m^{(n)}_l := \min_k p^{(n)}_{kl} \leq \max_k p^{(n)}_{kl} =: M^{(n)}_l,
\]

where

\[
\begin{pmatrix}
p^{(n)}_{00} & \ldots & p^{(n)}_{0l} & \ldots & p^{(n)}_{0K} \\
\vdots & \ddots & \vdots & \ddots & \vdots \\
p^{(n)}_{K0} & \ldots & p^{(n)}_{Kl} & \ldots & p^{(n)}_{KK}
\end{pmatrix}.
\]

(a) It holds \( m^{(n)}_l \leq m^{(n+1)}_l \): This can be seen from

\[
m^{(n+1)}_l = \min_k p^{(n+1)}_{kl} = \min_k \sum_{s=0}^K p_{ks} p^{(n)}_{sl} \geq \min_k \sum_{s=0}^K p_{ks} \left( \min_r p^{(n)}_{rl} \right) = \min_k p^{(n)}_{kl} = m^{(n)}_l.
\]

(b) Let \( \varepsilon := \min_{k,l} p^{(1)}_{kl} > 0 \). Then we show

\[
M^{(n+1)}_l - m^{(n+1)}_l \leq (1 - \varepsilon) \left( M^{(n)}_l - m^{(n)}_l \right).
\]

For this we compute

\[
p^{(n+1)}_{kl} = \sum_{s=0}^K p_{ks} p^{(n)}_{sl} = \sum_{s=0}^K \left( p_{ks} - \varepsilon p_{ls} \right) p^{(n)}_{sl} + \varepsilon \sum_{s=0}^K p_{ls} p^{(n)}_{sl}.
\]

Because of \( p_{ks} - \varepsilon p_{ls} \geq 0 \) and \( p^{(n)}_{sl} \geq m^{(n)}_l \) we get

\[
p^{(n+1)}_{kl} \geq m^{(n)}_l \sum_{s=0}^K \left( p_{ks} - \varepsilon p_{ls} \right) + \varepsilon \sum_{s=0}^K p_{ls} p^{(n)}_{sl} = m^{(n)}_l \sum_{s=0}^K \left( p_{ks} - \varepsilon p_{ls} \right) + \varepsilon p^{(2n)}_{ll}.
\]
3.4. ERGODIC THEOREM AND STATIONARY DISTRIBUTION

\[ m^{(n)}_l (1 - \varepsilon) + \varepsilon p^{(2n)}_{ll}, \]

which implies

\[ m^{(n+1)}_l \geq m^{(n)}_l (1 - \varepsilon) + \varepsilon p^{(2n)}_{ll}. \]

In the same way we obtain that

\[ M^{(n+1)}_l \leq M^{(n)}_l (1 - \varepsilon) + \varepsilon p^{(2n)}_{ll}. \]

(c) By iteration we get

\[ M^{(n+1)}_l - m^{(n+1)}_l \leq (1 - \varepsilon)^n (M^{(1)}_l - m^{(1)}_l). \]

Hence

\[ M^{(n+1)}_l - m^{(n+1)}_l \rightarrow 0 \]

which implies that the limit \( \lim_{n \rightarrow \infty} p^{(n)}_{kl} \) exists. Finally, from

\[ \varepsilon \leq m^{(n)}_l \leq m^{(n+1)}_l \leq p^{(n+1)}_{kl} \]

it follows that \( \lim_{n \rightarrow \infty} p^{(n)}_{kl} > 0 \).

We close this section with an example from genetics:

**Example 3.4.6 (Frequency of genes in a population).** Our assumptions are:

- We have a large population.
- We are concerned with diploids, i.e. organisms having DNA (double helix).
- There is no selection, no mutation, no immigration, and no migration.

We consider two scenarios, self-fertilization and random mating.

**a) Self-fertilization** (rice, wheat, ...): We will consider 1 locus with two alleles \( A \) and \( a \). Then the possible genotypes are \( \{AA, aa, Aa\} \), which will be our state space. Self-fertilization implies that mother and father have the same genotype:

<table>
<thead>
<tr>
<th>mother</th>
<th>father</th>
<th>result</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>AA</td>
<td>AA</td>
</tr>
<tr>
<td>aa</td>
<td>aa</td>
<td>aa</td>
</tr>
<tr>
<td>Aa</td>
<td>Aa</td>
<td>( \frac{1}{4} AA, \frac{1}{2} Aa, \frac{1}{4} aa )</td>
</tr>
</tbody>
</table>
We get the transition matrix (assume $AA = 0$, $aa = 1$, $Aa = 2$)

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

and the flow diagram:

The stationary distribution $(s_0, s_1, s_2) = (s, 1 - s, 0)$, where $s \in [0, 1]$, is not unique. The state $\{Aa\}$ is transient. Moreover, one can interpret the probability

\[P(f_n = AA) \approx \frac{\#\{\text{individuals of the nth generation having genotype } AA\}}{\#\{\text{all individuals of the nth generation}\}}\]

b) Random mating: We keep the assumptions like above but, instead of self-fertilization, we assume the random mating. Again we consider 1 locus with 2 alleles, $A$ and $a$, and the state space $\{AA, aa, Aa\}$. Let

\[p = \text{frequency of allele } A \quad \text{and} \quad q = 1 - p.\]

Since parent and child must have one gene in common and the other gene will be $A$ with probability $p$ and $a$ with probability $q$ we get:

- $1$ parent $\rightarrow$ child
  - $AA \rightarrow pAA, \quad qAa$
  - $aa \rightarrow pAa, \quad qaa$
  - $Aa \rightarrow \frac{1}{2}pAA, \quad \frac{1}{2}(q + p)Aa, \quad \frac{1}{2}qaa$

we get the transition matrix (assume $AA = 0$, $aa = 1$, $Aa = 2$)

\[
T = \begin{pmatrix}
p & 0 & q \\
0 & q & p \\
\frac{1}{2}p & \frac{1}{2}q & \frac{1}{2}
\end{pmatrix}
\]
and the flow diagram:

The Markov chain is ergodic with stationary distribution

\[(s_0, s_1, s_2) = (p^2, (1 - p)^2, 2p(1 - p)).\]

This result is well known as the **Hardy-Weinberg law** in population genetics: 'If there is random mating with no selection, no mutation, no immigration and no migration then the population will reach a steady state condition under which the frequencies of genes and genotypes will not change from generation to generation.'
Chapter 4

Markov chain Monte Carlo methods (MCMC methods)

4.1 The classical Monte Carlo method

Assume a probability space \((X, \mathcal{G}, \mu)\), and a Borel measurable function \(F : X \to \mathbb{R}\) such that \(\int_X |F(x)|d\mu(x) < \infty\). If one wants to compute

\[
\mathbb{E}_\mu F = \int_X F(x)d\mu(x),
\]

but a direct computation is not possible, then one can use the SLLN (strong law of large numbers): One generates independent identically distributed (i.i.d.) random variables \(\xi_1, \xi_2, \ldots\) with \(\xi_i : \Omega \to X\). Then

\[
\frac{1}{n} \sum_{i=1}^n F(\xi_i) \xrightarrow{n \to \infty} \mathbb{E}_\mu F \quad \text{a.s.}
\]

which means by definition that

\[
\mathbb{P} \left( \left\{ \omega \in \Omega : \frac{1}{n} \sum_{i=1}^n F(\xi_i(\omega)) \xrightarrow{n \to \infty} \mathbb{E}_\mu F \right\} \right) = 1.
\]

But there are situations where the computational complexity of this method is so large, such that it cannot be done in practice (even with fast computers). A way out consists in the usage of MCMC methods.
4.2 General idea of MCMC

We are looking for a Markov chain \((f_i)_{i=0}^{\infty}\) such that

- the state space is \(X\)
- the stationary distribution is \(\mu\), and

\[
\mathbb{P}\left( \left\{ \omega \in \Omega : \frac{1}{n} \sum_{i=1}^{n} F(f_i(\omega)) \stackrel{n \to \infty}{\longrightarrow} \mathbb{E}_\mu F \right\} \right) = 1.
\]

So instead of independent random values \(\xi_1, \xi_2, \ldots\) here dependent \(f_1, f_2, \ldots\) are used. One hopes that it is much easier to generate \(F(f_1), F(f_2), \ldots\) than \(F(\xi_1), F(\xi_2), \ldots\)

4.3 Basic example for the classical Monte Carlo method

Let \(F : [0, 1] \to [0, 1]\) be a continuous function. We want to compute

\[
\int_0^1 F(x) dx.
\]

Let \(X_1, Y_1, X_2, Y_2, X_3, Y_3, \ldots\) be a sequence of uniformly on \([0, 1]\) distributed independent random variables. Define

\[
Z_i := \mathbb{1}_{\{F(X_i) > Y_i\}}.
\]

We know that

\[
\frac{1}{n} \sum_{i=1}^{n} Z_i \stackrel{n \to \infty}{\longrightarrow} \mathbb{E} Z_1 \quad \text{a.s.}
\]

It holds, since \(X_1\) and \(Y_1\) are independent and uniformly on \([0, 1]\) distributed, that

\[
\mathbb{E} Z_1 = \mathbb{E} \mathbb{1}_{\{F(X_1) > Y_1\}} = \int_0^1 \int_0^1 \mathbb{1}_{\{F(x) > y\}} dy dx = \int_0^1 \int_0^{F(x)} dy dx = \int_0^1 F(x) dx.
\]
4.4 The Gibbs sampler: an example of a MCMC method

The development of the Gibbs sampler goes back (among others) to Pentti Suomela (around 1970) and Geman & Geman (Statistics, 1984). In Statistical Physics the Gibbs sampler is called Heat Bath Method.

**Stating the problem:** Given random variables $X_1, \ldots, X_m : \Omega \to \mathbb{R}$ and a measurable function $F : \mathbb{R}^m \to \mathbb{R}$. How to compute
\[
\mathbb{E} F(X_1, \ldots, X_m)
\]
umerically?

In the following we assume a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ such that $\Omega$ has only finitely many elements, $\mathcal{F} = 2^\Omega$ ($\mathcal{F}$ is the system of all subsets), and that $\mathbb{P}(\{\omega\}) > 0$ for all $\omega \in \Omega$. Furthermore, we will use the following set of assumptions:

**Assumption 4.4.1 (Assumptions on the state space $X$).**

(A) $S_1 \times \cdots \times S_m = \{(X_1(\omega), \ldots, X_m(\omega)) : \omega \in \Omega\}$ with $S_i := \{X_i(\omega) : \omega \in \Omega\}$.

**Example 4.4.2.** Let $m = 2$, $X_1, X_2 : \Omega \to \{-1, 1\}$ be independent and
\[
\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}.
\]
This implies $S_1 = S_2 = \{-1, 1\}$ and
\[
\{(X_1(\omega), X_2(\omega)) : \omega \in \Omega\} = \{(-1, -1), (-1, 1), (1, -1), (1, 1)\} = S_1 \times S_2.
\]
Hence assumptions (A) is satisfied.

**Example 4.4.3.** Let $m = 2$, $X_1 = X_2 : \Omega \to \{-1, 1\}$ and
\[
\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}.
\]
This implies $S_1 = S_2 = \{-1, 1\}$ and
\[
\{(X_1(\omega), X_2(\omega)) : \omega \in \Omega\} = \{(-1, -1), (1, 1)\} \neq S_1 \times S_2.
\]
Hence assumption (A) is not satisfied.
Description 4.4.4 (Description of the Gibbs sampler).

[Step 0] We set \((X_1^0, \ldots, X_m^0) := (X_1, \ldots, X_m)\). Then we simulate on the computer a realization \((\xi_1^0, \ldots, \xi_m^0)\) of the random vector \((X_1^0, \ldots, X_m^0)\). In other words, we assume that there is some \(\omega \in \Omega\) such that
\[
(\xi_1^0, \ldots, \xi_m^0) = (X_1^0(\omega), \ldots, X_m^0(\omega)).
\]
numbers in the computer

[Step \(n \rightarrow n + 1\)] Assume we have the random vector \((X_1^n, \ldots, X_m^n)\). Then we generate the random vector \((X_1^{n+1}, \ldots, X_m^{n+1})\) as follows:
If \((\xi_1^n, \ldots, \xi_m^n) = (X_1^n(\omega), \ldots, X_m^n(\omega))\) is the realization of step \(n\), then we proceed as follows:

- We simulate \((X_1, \ldots, X_m)\) under the condition that
  \[X_2 = \xi_2^n, \ldots, X_m = \xi_m^n\]
  and take \(\xi_1^{n+1} = X_1^{n+1}(\omega)\).

- We simulate \((X_1, \ldots, X_m)\) under the condition that
  \[X_1 = \xi_1^{n+1}, X_3 = \xi_3^n, \ldots, X_m = \xi_m^n\]
  and take \(\xi_2^{n+1} = X_2^{n+1}(\omega)\).

- \ldots

- We simulate \((X_1, \ldots, X_m)\) under the condition that
  \[X_1 = \xi_1^{n+1}, \ldots, X_{m-1} = \xi_{m-1}^{n+1}\]
  and take \(\xi_m^{n+1} = X_m^{n+1}(\omega)\).

Proposition 4.4.5. Let \(f_i := (X_1^{(i)}, \ldots, X_m^{(i)})\). Then the following holds for the Gibbs sampler.

1. \((f_i)_{i=0}^{\infty}\) is a homogeneous Markov chain with state space \(S_1 \times \cdots \times S_m\).
4.5. BURN-IN PERIOD FOR MCMC METHODS

(2) \((f_i)_{i=0}^{\infty}\) is an ergodic Markov chain with stationary distribution \((s_k)_{k \in S_1 \times \ldots \times S_m}\) where

\[ s_k = \Pr((X_1, \ldots, X_m) = k). \]

**Proof.** (1) follows by construction.

(2) Because of (A2) the transition matrix has only positive values. By Ergodic Theorem 3.4.5 it follows that the Markov chain is ergodic. We check the stationary distribution for \(m = 2\). It holds for the transition matrix that

\[
p_{(\xi_1, \eta_1), (\xi_2, \eta_2)} = \Pr(X_1 = \eta_1 | X_2 = \xi_2) \Pr(X_2 = \eta_2 | X_1 = \xi_1).
\]

We have the desired stationary distribution if

\[
\sum_{\xi_1, \xi_2} \Pr(X_1 = \xi_1, X_2 = \xi_2)p_{(\xi_1, (\eta_1), (\xi_2, \eta_2)} = \Pr(X_1 = \eta_1, X_2 = \eta_2).
\]

It holds that

\[
\sum_{\xi_1, \xi_2} \Pr(X_1 = \xi_1, X_2 = \xi_2)p_{(\xi_1, (\eta_1), (\xi_2, \eta_2)} \\
= \sum_{\xi_1, \xi_2} \Pr(X_1 = \xi_1, X_2 = \xi_2)\Pr(X_1 = \eta_1 | X_2 = \xi_2)\Pr(X_2 = \eta_2 | X_1 = \xi_1) \\
= \sum_{\xi_2} \left( \sum_{\xi_1} \Pr(X_1 = \xi_1, X_2 = \xi_2) \right) \Pr(X_1 = \eta_1 | X_2 = \xi_2) \Pr(X_2 = \eta_2 | X_1 = \xi_1) \\
= \sum_{\xi_2} \Pr(X_2 = \xi_2)\Pr(X_1 = \eta_1 | X_2 = \xi_2)\Pr(X_2 = \eta_2 | X_1 = \eta_1) \\
= \Pr(X_1 = \eta_1)\Pr(X_2 = \eta_2 | X_1 = \eta_1) \\
= \Pr(X_2 = \eta_2, X_1 = \eta_1).
\]

\[ \square \]

4.5  Burn-in period for MCMC methods

One does not start from the very beginning of the Gibbs sampler to collect observations. First one waits for a 'burn-in period' such that the starting distribution is close enough to the stationary distribution. The basis for this is the following
Theorem 4.5.1. Let \((f_i)_{i=0}^{\infty}\) be a homogeneous Markov chain with state space \(X = \{0, \ldots, K\}\) and transition matrix \(T = (p_{ij})_{i,j=0}^{K}\). Assume that

\[ \epsilon := \min_{i,j} p_{ij} > 0. \]

If \((s_0, \ldots, s_K)\) is the unique stationary distribution (see Theorem 3.4.5) and \(T^n = (p_{kl}^{(n)})_{k,l=0}^{K}\) then

\[ |p_{kl}^{(n+1)} - s_l| \leq (1 - \epsilon)^n \sup_{i,j} p_{ij}. \]

Proof. This follows immediately from the estimate

\[ M_{t}^{(n+1)} - m_{t}^{(n+1)} \leq (1 - \epsilon)^n (M_{t}^{(1)} - m_{t}^{(1)}) \]

in the proof of Theorem 3.4.5.

4.6 The hard-core model as an example for a modified Gibbs sampler

The following example (if considered in 3 dimensions) is related to statistical physics (see [3]). We think of a model of a gas whose particles have a radius and can not 'overlap'. A feasible (=acceptable) configuration is one where each particle (=black circle) has no direct neighbor.
As state space we take $X = \{ \text{ feasible configurations } \}$. This means, for example, if we have a $10 \times 10$ lattice, then $\#X \geq 2^{50} \approx 1.1 \times 10^{15}$.

**Algorithm**

1. Pick a vertex $v$ at random (uniformly).

2. Toss a fair coin.

3. $X_{n+1}(v) := \left\{ \begin{array}{ll} \text{‘black’} & \text{if coin=heads} \\ \text{and all direct neighbors} & \text{of } v \text{ are ‘white’ in } X_n, \\ \text{‘white’} & \text{otherwise} \end{array} \right.$

4. $X_{n+1}(w) := X_n(w)$ for all other vertices $w \neq v$ of the lattice.

If one compares the above algorithm with the description of the Gibbs sampler, it turns out that here the coordinates (vertices) are chosen at random whereas in
the Gibbs sampler they are chosen one after the other. So strictly speaking, the
hard core model is a 'randomized' Gibbs sampler.
Checkup after reading:

- Write up the connections between the notion of a stationary distribution of a Markov chain and the ergodicity of a Markov chain.

- Verify the computation of stationary distributions in both parts of Example 3.4.6.

- Did you understand the difference of the method for the Hard-Core model in Section 4.6 to the Gibbs sampler?
Bibliography


