Stochastic Modeling

Christel Geiss

February 3, 2015

Contents

Chapter 1

Introduction

1.1 Mathematical modeling

In order to make 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]: Observation of the weather at Snoqualmie Falls (Washington, U.S.A.) in the month January between the years 1948-1983

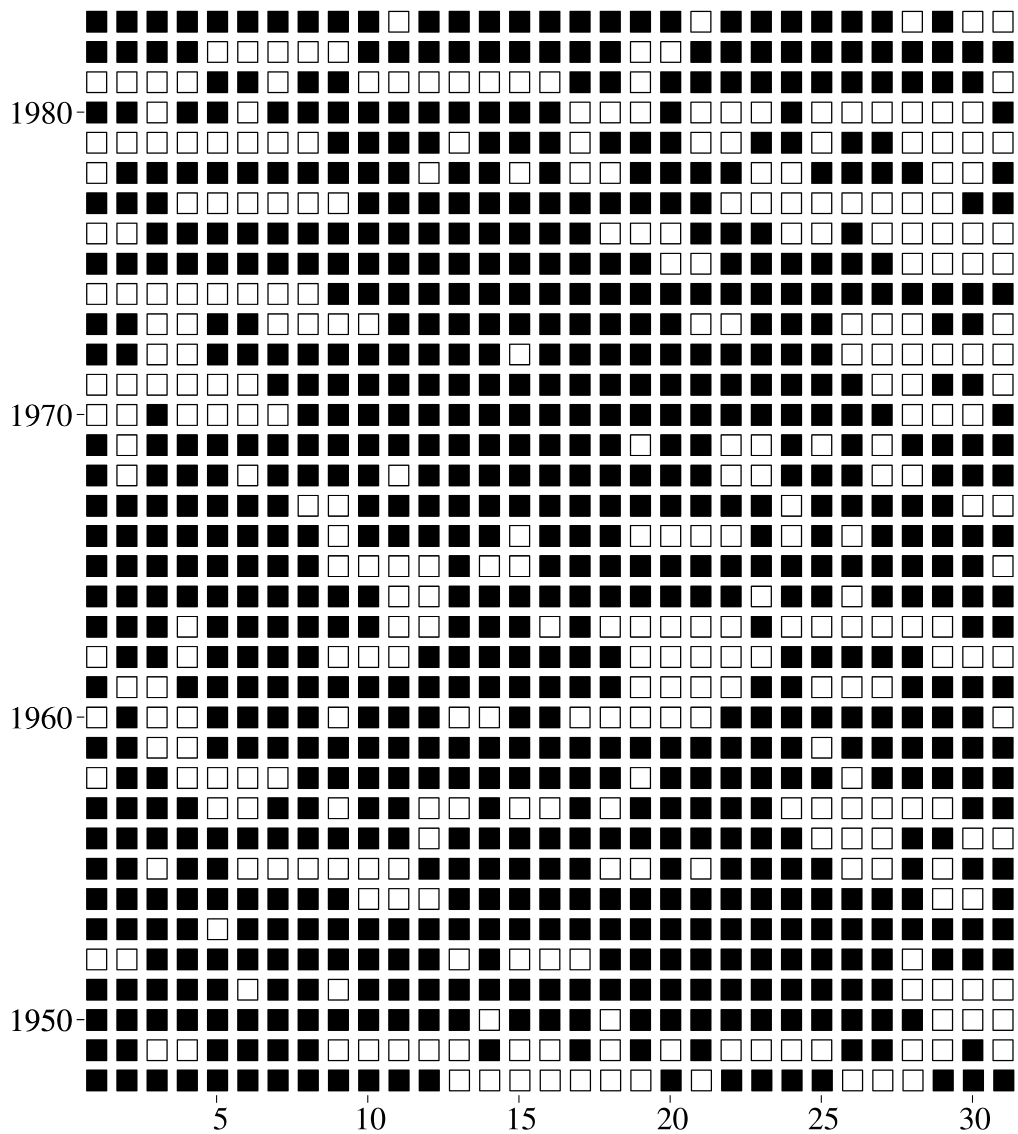
Rules

- 1 day = 8 am - 8 am (following day)
- day is wet $\Longleftrightarrow$ at least 0.01 inches (≈ 0.0254cm) of precipitation (=snow, rain, $\ldots$)
- day is dry $\Longleftrightarrow$ < 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 a day with precipitation is a black square. It shows the data taken from [2].



weather at Snoqualmie Falls

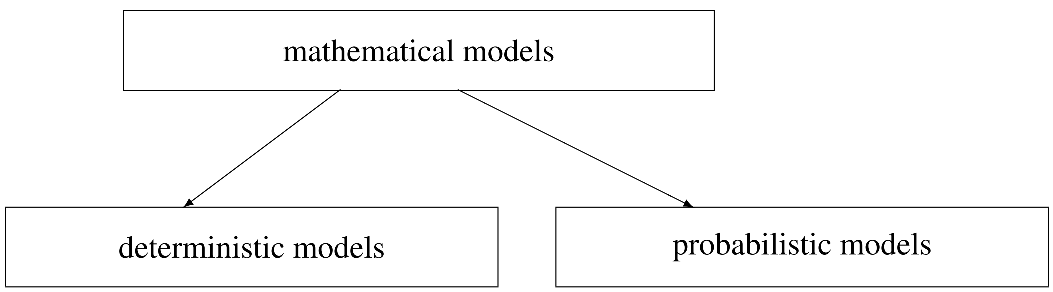
1.1.2 Second example: Financial Mathematics

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 Random and non-random models

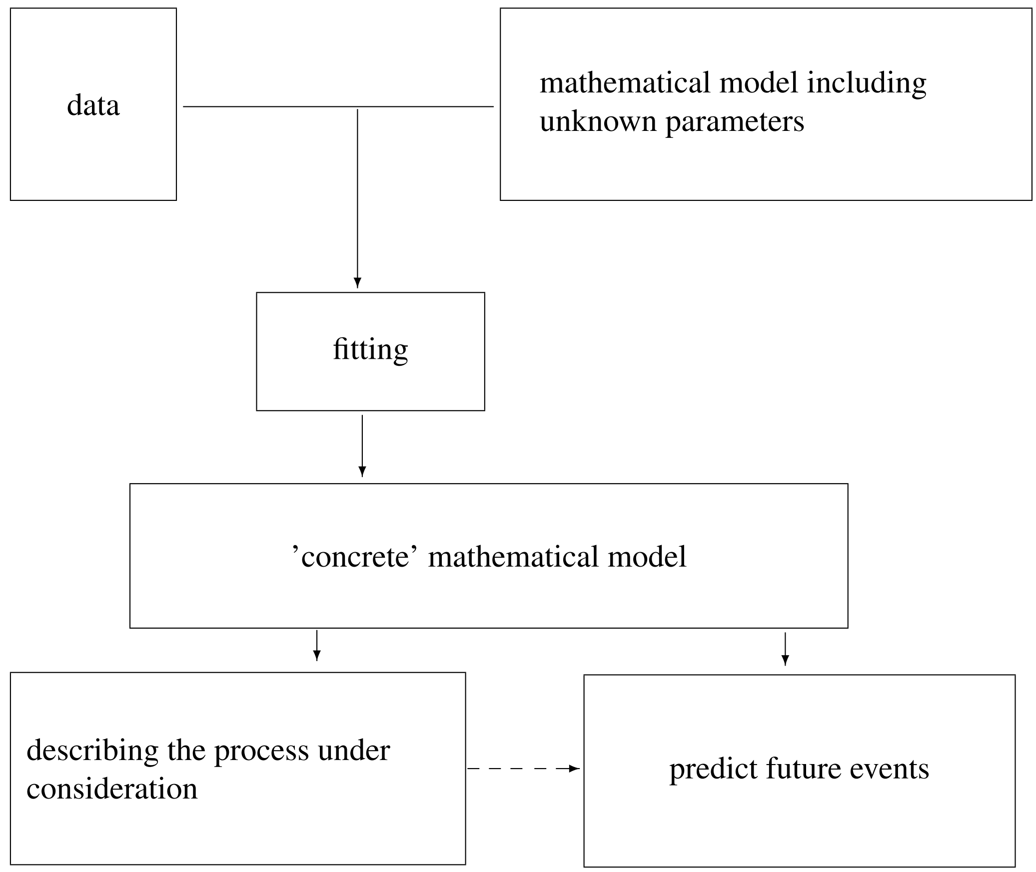


Differential equations: Kepler's laws of planet movement

Stochastic processes: share prices

Sources of random behavior:

- sensitivity to or randomness of initial conditions: weather forecast
- incomplete information: economic modeling
- fundamental description: impulse and position in modern quantum theory



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 $(\Omega, \mathcal{F}, \mathbb{P})$

Ω is a set which we sometimes assume to be *finite*: $\Omega = \{\omega_1, \dots, \omega_n\}$,
countable: $\Omega = \{\omega_1, \omega_2, \dots\}$,
or *uncountable*: $\Omega = \{\omega : \omega \in [0, 1]\}$.
The σ algebra (or σ field) is a basic tool in probability theory. It contains the sets where the probability measure is defined on.

Definition 1.2.1. [σ algebra]
Let Ω be a non-empty set. A system $\mathcal{F}$ of subsets $A \subseteq \Omega$ is a σ **algebra** on Ω if

(1) $\emptyset, \Omega \in \mathcal{F}$,

(2) $A \in \mathcal{F}$ implies that $A^c := \Omega \backslash A \in \mathcal{F}$,

(3) $A_1, A_2, ... \in \mathcal{F}$ implies that $\bigcup_{i=1}^{\infty} A_i \in \mathcal{F}$.

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

We consider some first examples of σ algebras.

Example 1.2.3. [σ algebras]

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

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

(c) Let $A \subseteq \Omega$. Then $\mathcal{F} = \{\Omega, \emptyset, A, A^c\}$ is a σ algebra.

(d) $\mathcal{B}(\mathbb{R})$ denotes the Borel σ algebra on $\mathbb{R}$
$\mathcal{B}([0,1])$ denotes the Borel σ algebra on $[0,1]$

Definition 1.2.4. [probability measure, probability space]
Let $(\Omega, \mathcal{F})$ be a measurable space, i.e. Ω is a non-empty set and $\mathcal{F}$ a σ algebra on it.

(1) A map $\mathbb{P} : \mathcal{F} \to [0,1]$ is called **probability measure** if

  (a) $\mathbb{P}(\Omega) = 1$,
  (b) for all $A_1, A_2, ... \in \mathcal{F}$ with $A_i \cap A_j = \emptyset$ for $i \neq j$ one has

$$\mathbb{P}\Big(\bigcup_{i=1}^{\infty} A_i\Big) = \sum_{i=1}^{\infty} \mathbb{P}(A_i).$$

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

Remark 1.2.5. 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 $A_1, A_2, ... \in \mathcal{F}$ such that $A_1 \subseteq A_2 \subseteq A_3 \subseteq \ldots$ then
$$\mathbb{P}(\bigcup_{n=1}^{\infty} A_n) = \lim_{N\to\infty} \mathbb{P}(A_N).$$

Example: 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**.

1.2.2 Conditional probability and independence

Definition 1.2.6. [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**.
A and B are called **independent,** if

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B), \quad \text{for } A, B \in \mathcal{F}.$$

Note: $\mathbb{P}(B|A) = \mathbb{P}(B)$ if A and B are independent, $\mathbb{P}(A) > 0$.

Definition 1.2.7. 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 $1 \le i_1 < i_2 \cdots < i_k \le n$, it holds

$$\mathbb{P}(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = \mathbb{P}(A_{i_1})\mathbb{P}(A_{i_2}) \ldots \mathbb{P}(A_{i_k}).$$

1.2.3 Random variables

Definition 1.2.8. A real function $f : \Omega \to \mathbb{R}$ is called $\mathcal{F}-$ **measurable** or **random variable,** if $\forall(a, b)$ with $-\infty < a < b < \infty$ the **pre-image**

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

Remark 1.2.9. One can prove that for an $\mathcal{F}-$ **measurable** function f it holds

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

Example 1.2.10. Let $A \in \mathcal{F}$. Then

$$f(\omega) = \mathbb{1}_A(\omega) = \begin{cases} 1 & ; \quad \omega \in A \\ 0 & ; \quad \omega \notin A \end{cases}$$

is a random variable. $\mathbb{1}_A(\omega)$ is called **indicator function.**

Definition 1.2.11. The random variables $f_i : \Omega \to \mathbb{R}, i = 1, \ldots, n$ are called **independent** , if it holds

$$\mathbb{P}(\{f_1 \in B_1, \ldots, f_n \in B_n\}) = \Pi_{i=1}^n \mathbb{P}(\{f_i \in B_i\}), \quad \forall B_i \in \mathcal{B}(\mathbb{R}).$$

Remark 1.2.12.

(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 Ω 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 $\Longleftrightarrow$.

$$\mathbb{P}(f_1 = x_1, f_2 = x_2, \ldots, f_n = x_n) = \Pi_{i=1}^n \mathbb{P}(f_i = x_i)$$

for any x_i of the range of the f_i.

Example 1.2.13. Two dice: $\Omega = \{(\omega_1, \omega_2) : \omega_1, \omega_2 \in \{1, \ldots, 6\}\}$

$$\mathbb{P}((\omega_1, \omega_2)) = \frac{1}{36}$$

$$f_1, f_2 : \Omega \to \mathbb{R} : \quad \begin{aligned} f_1((\omega_1, \omega_2)) &= \omega_1 + \omega_2 \\ f_2((\omega_1, \omega_2)) &= \omega_1 - \omega_2 \end{aligned}$$

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

$$\begin{aligned} \{(\omega_1, \omega_2) : f_1((\omega_1, \omega_2)) = 4\} &= \{(\omega_1, \omega_2) : \omega_1 + \omega_2 = 4\} \\ &= \{(1,3), (2,2), (3,1)\} \end{aligned}$$

and

$$\begin{aligned} \{(\omega_1, \omega_2) : f_1((\omega_1, \omega_2)) = 2\} &= \{(\omega_1, \omega_2) : \omega_1 - \omega_2 = 2\} \\ &= \{(6,4), (5,3), (4,2), (3,1)\}. \end{aligned}$$

Hence $\mathbb{P}(f_1 = x_1, f_2 = x_2) = \frac{1}{36}$ but $\mathbb{P}(f_1 = x_1)\mathbb{P}(f_2 = x_2) = \frac{1}{12}\frac{1}{9}$, so f_1 and f_2 are not independent.

1.2.4 Expectation

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

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

see [4] or [1].

Special cases to compute $\mathbb{E}f$

(1) 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) $\mathbb{P}_f(B) := \mathbb{P}(\{\omega : f(\omega) \in B\}) = \sum_{l=0}^\infty p_l \delta_l(B)$ with $p_l \geq 0$ and $\sum_{l=0}^\infty p_l = 1$. Then $\mathbb{P}_f$ is a **discrete measure,** where

$$\delta_l(B) = \begin{cases} 0 & ; \quad l \notin B \\ 1 & ; \quad l \in B \end{cases}$$

is called **Dirac measure.** It holds

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

(3) More general, assume that $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space, $f : \Omega \to \{x_1, x_2, ...\} \subseteq \mathbb{R}$ and $\{\omega \in \Omega : f(\omega) = x_l\} \in \mathcal{F}$ for all l. If $\sum_{l=1}^{\infty} \max\{x_l, 0\}\mathbb{P}(f = x_l) < \infty$ or $\sum_{l=1}^{\infty} \max\{-x_l, 0\}\mathbb{P}(f = x_l) < \infty$ then the expectation is defined by

$$\mathbb{E}f = \sum_{l=1}^{\infty} x_l \mathbb{P}(f = x_l).$$

(4) In case $\mathbb{P}_f(B) = \int_B h(x)dx \quad \forall B \in \mathcal{B}(\mathbb{R})$, i.e. if $\mathbb{P}_f$ has a density $h(x)$, we have

$$\mathbb{E}f = \int_{\mathbb{R}} x h(x) dx$$

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

Chapter 2

Random walk

2.1 Dyadic trees

Definition 2.1.1. (Dyadic tree, $\mu_n^{(p)}$)

Let $0 < p < 1$ and n ≥ 1. Then

$$\mathbb{D}_n := \{(\varepsilon_1, \ldots, \varepsilon_n) : \varepsilon_1 = \pm 1, \ldots, \varepsilon_n = \pm 1\}$$
$$\mathcal{F}_n^{dyad} := 2^{\mathbb{D}_n} \quad \text{(system of all subsets)}$$
$$\mu_n^{(p)}(\{(\varepsilon_1, \ldots, \varepsilon_n)\}) := (1-p)^k p^l,$$

where

$$k = \#\{i : \varepsilon_i = 1\}$$
$$l = \#\{i : \varepsilon_i = -1\},$$

and $k + l = n$.

Proposition 2.1.2. $(\mathbb{D}_n, \mathcal{F}_n^{dyad}, \mu_n^{(p)})$ is a finite probability space.
Proof

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

(2) $\mathcal{F}_n^{dyad}$ is a σ-algebra,

(3) We have only to check whether

$$\mu_n^{(p)}(\mathbb{D}_n) = \mathbb{P}(\Omega) = 1.$$

$$
\begin{aligned}
\mathbb{P}(\Omega) &= \sum_{\omega\in\Omega} \mathbb{P}(\{\omega\}) \\
&= \sum_{\varepsilon_1=\pm1} \cdots \sum_{\varepsilon_n=\pm1} \mu_n^{(p)}(\{(\varepsilon_1,\ldots,\varepsilon_n)\}) \\
&= \sum_{k=0}^{n} \sum_{\substack{(\varepsilon_1,\ldots,\varepsilon_n) \\ \#\{i:\varepsilon_i=1\}=k}} \mu_n^{(p)}(\{(\varepsilon_1,\ldots,\varepsilon_n)\}) \\
&= \sum_{k=0}^{n} \sum_{\substack{(\varepsilon_1,\ldots,\varepsilon_n) \\ \#\{i:\varepsilon_i=1\}=k}} (1-p)^k p^{n-k} \\
&= \sum_{k=0}^{n} \binom{n}{k} (1-p)^k p^{n-k} \\
&= (1-p+p)^n = 1.
\end{aligned}
$$

□

Interpretation of $(\mathbb{D}_n, \mathcal{F}_n^{dyad}, \mu_n^{(p)})$

Each $\omega = (\varepsilon_1, \ldots, \varepsilon_n) \in \mathbb{D}_n$ corresponds to a path in the dyadic tree from the left to the right with:

$\varepsilon_i \; = 1 \cong$ in the i-th step up $(\uparrow)$
$\varepsilon_i \; = -1 \cong$ 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 $(\uparrow)$ and l down-moves $(\downarrow) = (1-p)^k p^l$
$= \mu_n^{(p)}(\{\varepsilon_1, \ldots, \varepsilon_n\}), \quad$ if $\quad k = \#\{i : \varepsilon_i = 1\}$ 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).

Definition 2.1.3. (geometric Brownian motion, dyadic setting)

(1) Let $\Delta_i : \mathbb{D}_n \to \mathbb{R}$ 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

$$S_i(\omega) := e^{\sigma(\Delta_1(\omega)+\cdots+\Delta_i(\omega))-bi}$$

and $S_0(\omega) := 1$.

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

Proposition 2.1.4. (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.

Proof

(1) $\frac{S_i}{S_{i-1}} = \frac{e^{\sigma(\Delta_1+\cdots+\Delta_i)-bi}}{e^{\sigma(\Delta_1+\cdots+\Delta_{i-1})-b(i-1)}} = e^{\sigma\Delta_i-b}$

Hence

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

(2) $\Delta_1, \ldots, \Delta_n$ are independent, that means

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

$$\text{for all} \quad x_1, \ldots, x_n = \pm 1.$$

Hence $\frac{S_1}{S_0}, \frac{S_2}{S_1}, \ldots, \frac{S_n}{S_{n-1}}$ are independent. □

Remark 2.1.5. 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** (the increments are

$$\begin{aligned} \log S_1 - \log S_0 &= \sigma\Delta_1 - b, \\ \log S_2 - \log S_1 &= \sigma\Delta_2 - b, \ldots). \end{aligned}$$

2.2 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

$$X_i := X_0 e^{\sigma(\Delta_1 + \cdots + \Delta_i) - bi}$$

b is responsible for the drift
σ is responsible for the fluctuation around the mean.

In practice

$\boxed{\text{Data}} \rightarrow \boxed{\text{Estimates for } p, c_u, c_d} \rightarrow \boxed{\text{Computation of } b, \sigma}$

Note that:

$$\left.\begin{aligned} e^{-\sigma-b} &= c_d \\ e^{\sigma-b} &= c_u \end{aligned}\right\} \text{implies } \sigma, b$$

Remark 2.2.1. In certain cases it is enough to estimate σ. For example, the formula of the price of an European Call Option does not use p and b. There are 2 ways to estimate σ :

1.way	Historical method: One fits σ to the historical data of the prices of the share.
2.way	Implicit volatility: One fits σ to already known prices for European call options.

2.3 Snoqualmie Falls precipitation

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

(A1) $\mathbb{P}$ (day is wet) = p,
(A2) $\mathbb{P}$ (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.

$$\mathbb{P}(\text{325 dry days, 791 wet days}) = p^{791}(1-p)^{325}.$$

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 $\hat{p}$:
•$L(0) = L(1) = 0$
•$L'(p) = 791p^{790}(1-p)^{325} + p^{791}325(1-p)^{324}(-1)$

$$\begin{aligned} L'(p) = 0 \quad &\Leftrightarrow \quad 791(1-p) - 325p = 0 \\ &\Leftrightarrow \quad 791 = p(791+325) \\ &\Leftrightarrow \quad p = \frac{791}{325+791} = 0.709\ldots \end{aligned}$$

Hence: $\hat{p} \approx 0.709$.

Test of our model

$\mathbb{P}$(weather changes from one day to the other)

$$\begin{aligned} &= \mathbb{P}(\text{wet day} \to \text{dry day}) + \mathbb{P}(\text{dry day} \to \text{wet day}) \\ &= (1-p)p + p(1-p) = 2(1-p)p. \end{aligned}$$

Using $\hat{p}$ one gets:
$\mathbb{P}$(weather changes from one day to the other)

$$= 2(1-\hat{p})\hat{p} \approx 0.412638$$

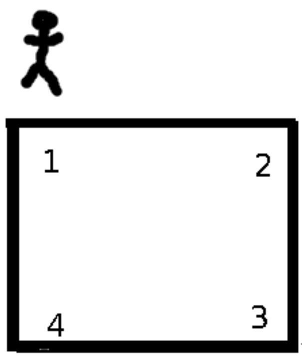
Expected weather changes =(possible) changes in January $\times$ number of years $\times$ $\mathbb{P}$(weather changes from one day to the other) $= 30 \times 36 \times 0.412638 \approx 446.$
From the data observed weather changes: 251

$$\Rightarrow \boxed{\textbf{The model is not the right one.}}$$

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.



<u>time 0</u>	walker is at corner 1: "$f_0 = 1$"
<u>time 1</u>	he flips a coin:
	"heads" $\to$ corner 2: "$f_1 = 2$"
	"tails" $\to$ corner 4: "$f_1 = 4$"
<u>time 2</u>	he flips a coin :
	"heads" $\to$ clockwise
	"tails" $\to$ counterclockwise
$\ldots$	

$\mathbb{P}(f_1 = 2) = \frac{1}{2}$,
$\mathbb{P}(f_1 = 4) = \frac{1}{2}$.
How to compute: $\mathbb{P}(f_n = 1)$?

We use **conditional probabilities:**
Assume the walker was a the time n in corner 2, then

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

It holds :

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

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

Assume $X = \{x_1, \ldots, x_M\}$ and a function p_0 with $p_0(x_l) \geq 0$, for $l = 1, \ldots, M$ and $\sum_{l=1}^{M} p_0(x_l) = 1$.

Moreover, assume matrices

$$T_k = \begin{pmatrix} p_k(x_1, x_1) & \ldots & p_k(x_1, x_M) \\ \vdots & & \vdots \\ p_k(x_M, x_1) & \ldots & p_k(x_M, x_M) \end{pmatrix}$$

for $k = 1, 2, \ldots, n$ with $p_k(x_l, x_m) \geq 0$ for all $1 \leq l, m \leq M$ and

$$\sum_{m=1}^{M} p_k(x_l, x_m) = 1, \quad \forall l = 1, \ldots, M.$$

Theorem 3.1.1. [Existence of a Markov chain]
For a given p_0 and given matrices T_k, $k = 1, \ldots, n$ there exists always a finite probability space $(\Omega, 2^{\Omega}, \mathbb{P})$ and a family of random variables $(f_k)_{k=0}^{n}$, where $f_k : \Omega \to X$, with the properties

$$p_0(x_l) = \mathbb{P}(f_0 = x_l), \quad l = 1, \ldots, M$$

and

$$p_k(x_l, x_m) = \mathbb{P}(f_k = x_m | f_{k-1} = x_l), \quad l, m = 1, \ldots, M,$$

whenever $\mathbb{P}(f_{k-1} = x_l) > 0, \quad k = 1, \ldots, n$
and for all $y_0, \ldots, y_{k+1} \in X$

$$\begin{aligned} \mathbb{P}(f_{k+1} = y_{k+1} | f_k = y_k, f_{k-1} = y_{k-1}, \ldots, f_0 = y_0) & \\ = \ \mathbb{P}(f_{k+1} = y_{k+1} | f_k = y_k) & \qquad (3.1) \end{aligned}$$

whenever $\mathbb{P}(f_k = y_k, f_{k-1} = y_{k-1}, \ldots, f_0 = y_0) > 0.$

Definition 3.1.2. The triplet $((f_i)_{i=0}^n, p_0, (T_k)_{k=1}^n)$ we call a **Markov chain.** (3.1) is the **Markov property.** p_0 is called the **initial distribution** of the Markov chain. T_k is a **transition matrix** of the Markov chain. If $T_1 = T_2 = \cdots = T$ then the Markov chain is called **homogeneous.**

Remark 3.1.3. $\sum_{m=1}^M p(x_l, x_m) = 1$ corresponds to

$$\begin{aligned}\sum_{m=1}^{M} \mathbb{P}(f_k = x_m | f_{k-1} = x_l) &= \mathbb{P}(\bigcup_{m=1}^{M} \{f_k = x_m\} | f_{k-1} = x_l) \\ &= \mathbb{P}\left(\underbrace{f_k \in \{x_1 \dots, x_M\}}_{\Omega} | f_{k-1} = x_l\right) = 1\end{aligned}$$

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

Example: The geometric Brownian motion $(S_i)_{i=0}^n$ with $S_i = e^{\sigma(\Delta_1 + \cdots + \Delta_i) - bi}$ has the Markov property.
Proof. $S_0(\omega) := 1$, so for the initial distribution p_0 we have $p_0(1) = \mathbb{P}(S_0 = 1) = 1$.

We check the Markov property. For $y_i > 0$ we have

$$\begin{aligned}&\mathbb{P}(S_{i+1} = y_{i+1} | S_i = y_i, \dots, S_0 = y_0) \\ &\quad = \frac{\mathbb{P}(S_{i+1} = y_{i+1}, S_i = y_i, \dots, S_0 = y_0)}{\mathbb{P}(S_i = y_i, \dots, S_0 = y_0)} \\ &\quad = \frac{\mathbb{P}\left(\frac{S_{i+1}}{S_i} = \frac{y_{i+1}}{y_i}, S_i = y_i, \dots, S_0 = y_0\right)}{\mathbb{P}(S_i = y_i, \dots, S_0 = y_0)} \\ &\quad = \mathbb{P}\left(\frac{S_{i+1}}{S_i} = \frac{y_{i+1}}{y_i}\right) \\ &\quad = \frac{\mathbb{P}\left(\frac{S_{i+1}}{S_i} = \frac{y_{i+1}}{y_i}\right)\mathbb{P}(S_i = y_i)}{\mathbb{P}(S_i = y_i)} \\ &\quad = \frac{\mathbb{P}\left(\frac{S_{i+1}}{S_i} = \frac{y_{i+1}}{y_i}, S_i = y_i\right)}{\mathbb{P}(S_i = y_i)} \\ &\quad = \frac{\mathbb{P}(S_{i+1} = y_{i+1}, S_i = y_i,)}{\mathbb{P}(S_i = y_i)} \\ &\quad = \mathbb{P}(S_{i+1} = y_{i+1} | S_i = y_i).\end{aligned}$$

Here we used that the Δ_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 $\{\frac{S_{i+1}}{S_i} = \frac{y_{i+1}}{y_i}\} = \{e^{\sigma\Delta_{i+1}-b} = \frac{y_{i+1}}{y_i}\}$. This implies that the sets $\{\frac{S_{i+1}}{S_i} = \frac{y_{i+1}}{y_i}\}$ and $\{S_i = y_i, \ldots, S_0 = y_0\}$ are independent. In the same way one can see that $\{\frac{S_{i+1}}{S_i} = \frac{y_{i+1}}{y_i}\}$ 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 so-called **marginal distributions:**

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

Let us also define the matrices

$$\mathbf{P}^{(i)} := \left(p^{(i)}(x_k, x_l)\right)_{k,l=1,\ldots,M}$$

by

$$p^{(i)}(x_k, x_l) := \mathbb{P}(f_i = x_l | f_0 = x_k).$$

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

1. $\mathbf{P}^{(i+j)} = \mathbf{P}^{(i)} \circ \mathbf{P}^{(j)}$ *for* $1 \leq i, j, i+j, \leq n$
2. $\Pi^{(i+j)} = \Pi^{(i)} \circ \mathbf{P}^{(j)}$ *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 $\Omega = \bigcup_{s=1}^{M} \{f_i = x_s\}$ and $\{f_i = x_{s_1}\} \cap \{f_i = x_{s_2}\} = \emptyset$ for $s_1 \neq s_2$ it holds

$$\begin{aligned}
p^{(i+j)}(x_k, x_l) &= \mathbb{P}(f_{i+j} = x_l | f_0 = x_k) \\
&= \sum_{s=1}^{M} \mathbb{P}(f_{i+j} = x_l, f_i = x_s | f_0 = x_k) \\
&= \sum_{s=1}^{M} \frac{\mathbb{P}(f_{i+j} = x_l, f_i = x_s, f_0 = x_k)}{\mathbb{P}(f_0 = x_k)}
\end{aligned}$$

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

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

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

$$\begin{aligned}
&\sum_{s=1}^{M} \frac{\mathbb{P}(f_{i+j} = x_l, f_i = x_s, f_0 = x_k)}{\mathbb{P}(f_0 = x_k)} \\
&= \sum_{s=1}^{M} \mathbb{P}(f_{i+j} = x_l | f_i = x_s, f_0 = x_k)\mathbb{P}(f_i = x_s | f_0 = x_k) \\
&= \sum_{s=1}^{M} \mathbb{P}(f_{i+j} = x_l | f_i = x_s)\mathbb{P}(f_i = x_s | f_0 = x_k) \\
&= \sum_{s=1}^{M} \mathbb{P}(f_j = x_l | f_0 = x_s)\mathbb{P}(f_i = x_s | f_0 = x_k) \\
&= \sum_{s=1}^{M} p^{(j)}(x_s, x_l)p^{(i)}(x_k, x_s).
\end{aligned}$$

This implies $\mathbf{P}^{(i+j)} = \mathbf{P}^{(i)} \circ \mathbf{P}^{(j)}$. Similarly, one gets

$$\begin{aligned}
p^{(i+j)}(x_l) &= \mathbb{P}(f_{i+j} = x_l) \\
&= \sum_{s=1}^{M} \mathbb{P}(f_{i+j} = x_l, f_i = x_s) \\
&= \sum_{s=1}^{M} \mathbb{P}(f_{i+j} = x_l | f_i = x_s)\mathbb{P}(f_i = x_s) \\
&= \sum_{s=1}^{M} \mathbb{P}(f_j = x_l | f_0 = x_s)\mathbb{P}(f_i = x_s) \\
&= \sum_{s=1}^{M} p^{(j)}(x_s, x_l) p^{(i)}(x_s).
\end{aligned}$$

□

Example We will 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

$$\begin{aligned}
&\mathbb{P}(f_3 = y_3 | f_2 = y_2, f_0 = y_0) \\
&\quad = \frac{\mathbb{P}(f_3 = y_3, f_2 = y_2, f_0 = y_0)}{\mathbb{P}(f_2 = y_2, f_0 = y_0)} \\
&\quad = \frac{\sum_{s=1}^{M} \mathbb{P}(f_3 = y_3, f_2 = y_2, f_1 = y_s, f_0 = y_0)}{\mathbb{P}(f_2 = y_2, f_0 = y_0)} \\
&\quad = \sum_{s=1}^{M} \frac{\mathbb{P}(f_3 = y_3, f_2 = y_2, f_1 = y_s, f_0 = y_0)}{\mathbb{P}(f_2 = y_2, f_1 = y_s, f_0 = y_0)} \, \frac{\mathbb{P}(f_2 = y_2, f_1 = y_s, f_0 = y_0)}{\mathbb{P}(f_2 = y_2, f_0 = y_0)} \\
&\quad = \mathbb{P}(f_3 = y_3 | f_2 = y_2) \frac{\sum_{s=1}^{M} \mathbb{P}(f_2 = y_2, f_1 = y_s, f_0 = y_0)}{\mathbb{P}(f_2 = y_2, f_0 = y_0)} \\
&\quad = \mathbb{P}(f_3 = y_3 | f_2 = y_2).
\end{aligned}$$

Corollary 3.2.2. (Marginal distribution)

1. $\mathbf{P}^{(j)} = \underbrace{T \circ \cdots \circ T}_{j \text{ times}}$,

2. $\Pi^{(j)} = \Pi \circ \underbrace{T \circ \cdots \circ T}_{j \text{ times}}$.

We will use the notation $T^j := \underbrace{T \circ \cdots \circ T}_{j \text{ times}}$.

Proof. By induction, since $\mathbf{P}^{(1)} = T$,

$$\mathbf{P}^{(j)} = \mathbf{P}^{(1)} \circ \mathbf{P}^{(j-1)} = \mathbf{P}^{(1)} \circ (\mathbf{P}^{(1)} \circ \mathbf{P}^{(j-2)}) = \cdots = \underbrace{T \circ \cdots \circ T}_{j \text{ times}}.$$

and

$$\Pi^{(j)} = \Pi \circ \mathbf{P}^{(j)} = \Pi \circ \underbrace{T \circ \cdots \circ T}_{j \text{ times}}.$$

□

Theorem 3.2.3. 'step-by-step formula'
Let $(f_i)_{i=0}^n$ *be a homogeneous Markov chain. Then for* $0 \le k < m \le n$

$$\begin{aligned} \mathbb{P}(& f_m = y_m, \ldots, f_{k+1} = y_{k+1} | f_k = y_k) \\ &= \mathbb{P}(f_m = y_m, | f_{m-1} = y_{m-1}) \times \\ & \qquad \mathbb{P}(f_{m-1} = y_{m-1} | f_{m-2} = y_{m-2}) \times \\ & \qquad\quad \times \cdots \times \mathbb{P}(f_{k+1} = y_{k+1} | f_k = y_k), \end{aligned}$$

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

Example: 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:

$$\begin{aligned}
\mathbb{P}(f_{k+1} &= y_{k+1}|f_k = y_k, \dots, f_0 = y_0) \\
&= \frac{\mathbb{P}(f_{k+1} = y_{k+1}, f_k = y_k, \dots, f_0 = y_0)}{\mathbb{P}(f_k = y_k, \dots, f_0 = y_0)} \\
&= \frac{\mathbb{P}(f_{k+1} = y_{k+1})\mathbb{P}(f_k = y_k) \cdots \mathbb{P}(f_0 = y_0)}{\mathbb{P}(f_k = y_k) \cdots \mathbb{P}(f_0 = y_0)} \\
&= \mathbb{P}(f_{k+1} = y_{k+1}) \\
&= \frac{\mathbb{P}(f_{k+1} = y_{k+1})\mathbb{P}(f_k = y_k)}{\mathbb{P}(f_k = y_k)} \\
&= \frac{\mathbb{P}(f_{k+1} = y_{k+1}, f_k = y_k)}{\mathbb{P}(f_k = y_k)} \\
&= \mathbb{P}(f_{k+1} = y_{k+1}|f_k = y_k).
\end{aligned}$$

Answer: yes.

3.2.1 Example: Snoqualmie Falls precipitation

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.
Parameters:

- initial distribution $p_0 = (p, 1 - p)$

  $\mathbb{P}($ 1st of January is dry$) = p$
  $\mathbb{P}($ 1st of January is rainy$) = 1 - p$

- transition matrix
$$T = \begin{pmatrix} p_{00} & p_{01} \\ p_{10} & p_{11} \end{pmatrix}$$

- Fix $1948 \leq j \leq 1983$.

Then $[(f_i^{(j)})_{i=1}^{31}, p_0, T]$ is a Markov chain which describes the weather in January of year j.

Maximum Likelihood estimates for T

A_j = number of changes $0 \to 0$,
$B_j = \dots$ $0 \to 1$,
$C_j = \dots$ $1 \to 0$,
$D_j = \dots$ $1 \to 1$ in year j.

Observe: $A_j + B_j + C_j + D_j = 30$.

The probability of the observed data we can express, since the years are independent and the "step-by-step-formula" holds, by

$$\begin{aligned}
&\Pi_{j=1948}^{1983}\left(\mathbb{P}(f_1^{(j)} = x_{1j})\, p_{00}^{A_j}\, p_{01}^{B_j}\, p_{10}^{C_j}\, p_{11}^{D_j}\right) \\
&= \left(\Pi_{j=1948}^{1983}\mathbb{P}(f_1^{(j)} = x_{1j})\right)\, \underbrace{p_{00}^{186}\, p_{01}^{123}\, p_{10}^{128}\, p_{11}^{643}}_{L(p_{00},p_{01},p_{10},p_{11})}
\end{aligned}$$

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 $p_{00}^{186}\, p_{01}^{123}$ and $p_{10}^{128}\, p_{11}^{643}$ separately. This yields to the matrix

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

Problem: What is the probability of wet/dry at the 6-th of January if it is raining at the 1st of January? Computation yields to

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

Hence

$$\begin{aligned}
\mathbb{P}(f_5^{(j)} = 1 | f_0^{(j)} = 1) &= 0.710, \\
\mathbb{P}(f_5^{(j)} = 0 | f_0^{(j)} = 1) &= 0.290.
\end{aligned}$$

3.3 Equivalent formulations of the Markov property

Theorem 3.3.1. *Let $(f_i)_{i=0}^n$ be a sequence of random variables on $(\Omega, \mathcal{F}, \mathbb{P})$, $f_i : \Omega \to X$. Then the following assertions are equivalent:*

1. *It holds*

$$\begin{aligned}\mathbb{P}(f_{k+1} = y_{k+1} | f_k = y_k, f_{k-1} = y_{k-1}, \ \ldots, f_0 = y_0) \\ = \ \mathbb{P}(f_{k+1} = y_{k+1} | f_k = y_k),\end{aligned}$$

   whenever $\mathbb{P}(f_k = y_k, f_{k-1} = y_{k-1}, \ldots, f_0 = y_0) > 0.$

2. *For all* $0 \le n_0 < n$, *for all* $S \subseteq \{0, \ldots, n_0\}$, *for all* $T \subseteq \{n_0 + 1, \ldots, n\}$ *one has*

$$\mathbb{P}(\{f_j = y_j : j \in T\} | \{f_i = y_i : i \in S\}) = \mathbb{P}(\{f_j = y_j : j \in T\} | f_{i_0} = y_{i_0})$$

   for $\mathbb{P}(\{f_i = y_i : i \in S\}) > 0$, *where* $i_0 := \max\{i \in S\}$.

3. *For all* $y_i \in X$, *where* $i = 0, ..., k+1$ *it holds*

$$\begin{aligned}\mathbb{P} \ & (f_{k+1} = y_{k+1}, f_{k-1} = y_{k-1}, ..., f_0 = y_0 | f_k = y_k) \\ = \ & \mathbb{P}(f_{k+1} = y_{k+1} | f_k = y_k) \mathbb{P}(f_{k-1} = y_{k-1}, ..., f_0 = y_0 | f_k = y_k),\end{aligned}$$

   if $\mathbb{P}(f_k = y_k) > 0.$
   'If one knows the present state, past and future are independent.'

Question: Is the "step-by-step-formula" also equivalent to the Markov property?

3.4 Classification of states for a homogeneous Markov chain

Let

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

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

3.4.1 Absorbing states

Definition 3.4.1. *A state* $k \in X$ *of a homogeneous Markov chain* $(f_i)_{i=0}^{n}$ *is called* **absorbing** $:\Longleftrightarrow$

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

(If the Markov chain reaches k, *then it stays there forever.)*

Problem: How to compute the probability, that a Markov chain reaches an absorbing state after some time?

Example: Branching process
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 ...	$p_1 \in [0,1]$,
2	particles ...	$p_2 \in [0,1], \ldots$

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

$$\mathbb{P}(\{\omega \in \Omega : \text{ there exists an } i \text{ such that } f_i(\omega) = 0\}).$$

For this we use the method of **probability generating functions.**

Definition 3.4.2. *Let* $F, F_i : [0,1] \to [0,1]$ *be given by*

$$F(\theta) := \sum_{l=0}^{\infty} \theta^l p_l,$$

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

Remark

1. It holds $p_l = \frac{1}{l!} F^{(l)}(0)$.
2. $F = F_1$ because of $p_l = \mathbb{P}(f_1 = l)$, note that $f_0 = 1$.

Proposition 3.4.3. *For* $i = 1, 2, \ldots$ *one has*

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

Consequently,

$$F_i = \underbrace{F \circ \cdots \circ F}_{i \text{ times}},$$

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

Proof. We will use the following

Lemma 3.4.4. *Assume $g_1, \dots, g_n$ are independent random variables, with $\mathbb{E}|g_i| < \infty$ for $i = 1, \dots, n$. Then*

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

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

Now let $0 \le \theta \le 1$, and $i \in \{1, 2, \dots\}$

$$\begin{aligned} F_{i+1}(\theta) &= \sum_{l=0}^{\infty} \theta^l \mathbb{P}(f_{i+1} = l) \\ &= \sum_{l=0}^{\infty} \theta^l \mathbb{E} \mathbb{1}_{\{f_{i+1}=l\}} \\ &= \mathbb{E} \sum_{l=0}^{\infty} \theta^l \mathbb{1}_{\{f_{i+1}=l\}} = \mathbb{E}\theta^{f_{i+1}}. \end{aligned}$$

We will assume that $\xi_1, \xi_2, \dots$ are independent and identically distributed like f_1. That means that $\forall m \ge 1$ we have $\mathbb{P}(\xi_m = k) = p_k$ for $k = 0, 1, 2, \dots$ independently from f_i. Hence, by Lemma 3.4.4,

$$\begin{aligned} \mathbb{E}\theta^{f_{i+1}} &= \mathbb{E} \sum_{l=0}^{\infty} \mathbb{1}_{\{f_i=l\}} \theta^{\xi_1 + \cdots + \xi_l} \\ &= \sum_{l=0}^{\infty} \mathbb{E} \mathbb{1}_{\{f_i=l\}} \theta^{\xi_1} \times \cdots \times \theta^{\xi_l} \\ &= \sum_{l=0}^{\infty} \mathbb{P}(f_i = l) \mathbb{E}\theta^{\xi_1} \times \cdots \times \mathbb{E}\theta^{\xi_l} \\ &= \sum_{l=0}^{\infty} \mathbb{P}(f_i = l) \left(\mathbb{E}\theta^{f_1}\right)^l \\ &= F_i(\mathbb{E}\theta^{f_1}) = F_i(F(\theta)). \end{aligned}$$

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

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

$\square$

Proposition 3.4.5. *For $q_i = \mathbb{P}(f_i = 0)$ where $i = 0, 1, \ldots$ one has the following:*

1. $0 = q_0 < p_0 = q_1 \le q_2 \le q_3 \le \cdots \le 1.$
2. $q_{i+1} = F(q_i) \quad$ *for* $\quad i = 0, 1, \ldots$
3. *For $q := \lim_{i\to\infty} q_i$ one has*
$$\mathbb{P}(\{\omega \in \Omega : \exists i \quad \textit{such that} \quad f_i(\omega) = 0\}) = q.$$
4. $q = F(q).$
5. *F is continuous, increasing, convex and satisfies $F(0) = p_0$ and $F(1) = 1$.*

Proof.

1. $f_i(\omega) = 0 \Longrightarrow f_{i+1}(\omega) = 0 \quad$ since 0 is **absorbing.** Hence
$$q_i = \mathbb{P}(f_i = 0) \le \mathbb{P}(f_{i+1} = 0) = q_{i+1}.$$
Moreover, we have $q_0 = 0$ since $f_0 = 1$.
2. $q_{i+1} = F_{i+1}(0) = F(F_i(0)) = F(q_i).$
3. By Remark 1.2.5 (continuity from below), because of $\{f_n = 0\} \subseteq \{f_{n+1} = 0\}$, we have
$$\begin{aligned}\mathbb{P}(\{\omega \in \Omega : \exists n \quad \text{such that} \quad f_n(\omega) = 0\}) &= \mathbb{P}(\bigcup_{n=1}^{\infty}\{f_n = 0\}) \\ &= \lim_{N\to\infty} \mathbb{P}(f_N = 0) \\ &= \lim_{N\to\infty} q_N.\end{aligned}$$
4. F is continuous. 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 $0 \le \theta < 1$ one has that
$$F'(\theta) = \sum_{k=1}^{\infty} k\theta^{k-1} p_k \ge 0,$$
so F is increasing. From $F''(\theta) \ge 0$ it follows that F is convex.

Proposition 3.4.6. *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. *For* $F'(1) > 1$ *there are two solutions,* $0 < z_1 < z_2 = 1$, *to* $F(q) = q$. *The solution of our problem is given by* $q = z_1$.

Interpretation: It holds

$$F'(1) = \sum_{k=1}^{\infty} k p_k = \mathbb{E} f_1.$$

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

Notation:
We call the system, if

- $\mathbb{E} f_1 < 1$, **sub-critical,**
- $\mathbb{E} f_1 = 1$, **critical,**
- $\mathbb{E} f_1 > 1$, **super-critical.**

Example
Let $p_0 \in (0,1)$, $\quad p_1 = 0$, $\quad p_2 = 1 - p_0$. 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

$$2(1-p_0) < 1 \iff \frac{1}{2} < p_0 < 1 \iff \text{sub-critical}$$

$$2(1-p_0) = 1 \iff p_0 = \frac{1}{2} \iff \text{critical}$$

$$2(1-p_0) > 1 \iff 0 < p_0 < \frac{1}{2} \iff \text{super-critical}$$

Computation of q in the super-critical case:

$$z = F(z) \iff z = p_0 + (1-p_0)z^2$$

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

Consequently, $q = \frac{1-\sqrt{1-4p_0(1-p_0)}}{2(1-p_0)}$.

Preservation of family names

(Bienaymé, 1845). We consider the following model.

generation	0	1 father	$f_0 = 1$
	1	number of sons of this father	f_1
	2	number of sons of the first generation	f_2
	3	number of sons of the second generation	f_3
	$\vdots$	$\vdots$	$\vdots$

We assume

$$\mathbb{P}(\text{ a father has no son }) = p \in (0,1).$$

$$\mathbb{P}(\text{ a father has 1 son }) = pa,$$

$$\mathbb{P}(\text{ a father has 2 sons }) = pa^2,$$

$$\mathbb{P}(\text{ a father has k sons }) = pa^k$$

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

Question
How to compute

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

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

$$\begin{aligned} F(\theta) &= p + \theta(pa) + \theta^2(pa^2) + \cdots + \theta^k(pa^k) + \ldots \\ &= p\frac{1}{1-\theta a}. \end{aligned}$$

Then

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

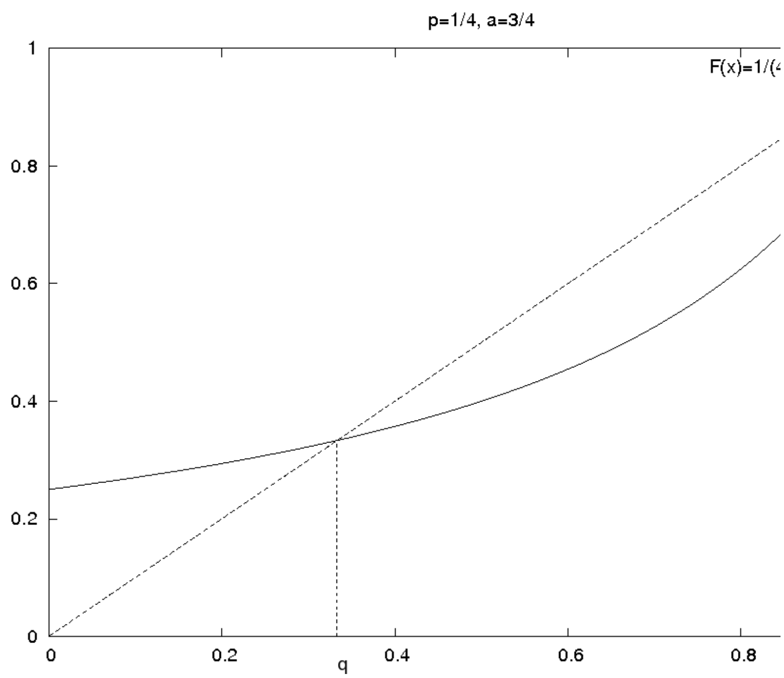
We have

$$a < p \iff \frac{1}{2} < p < 1 \iff \text{sub-critical}$$

$$a = p \iff p = \frac{1}{2} \iff \text{critical}$$

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

Again, by the help of Proposition 3.4.6 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 $q = \frac{1}{2a} - \sqrt{\frac{1}{4a^2} - \frac{p}{a}} = \frac{p}{1-p}.$



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** (=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 (= **genotype**) existent. Let the population consist of N individuals. In generation 0 happened a mutation $A \to a$. So 1 individual has now Aa while $N-1$ have AA.

Let $\mu > 0$ be the **relative fitness** (=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:

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

Assuming random mating and that N is large yields to

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

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

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

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

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

$$F'(\theta) = (1+s)e^{(\theta-1)(1+s)}$$

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

We compute q for the case $s > 0$. Since it holds by Taylor expansion

$$\begin{aligned} q = e^{(1+s)(q-1)} &= \sum_{l=0}^{\infty} \frac{F^{(l)}(1)}{l!}(q-1)^l \\ &= 1 + (1+s)(q-1) + \frac{(1+s)^2}{2}(q-1)^2 + \dots \\ &\approx 1 + (q-1) + s(q-1) + \frac{1+2s+s^2}{2}(q-1)^2, \end{aligned}$$

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

$$q \approx 1 - 2s.$$

3.4.2 Communicating states

Definition 3.4.7. *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** *($k \to l$) :$\Longleftrightarrow$*

$$\exists i \geq 0 \quad \textit{such that} \quad \mathbb{P}(f_i = l | f_0 = k) > 0.$$

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

$$k \to l \quad \textit{and} \quad l \to k.$$

Proposition 3.4.8. *$\leftrightarrow$ is an* **equivalence relation** *on X, that means:*

1. *$k \leftrightarrow k$ reflexivity*
2. *$k \leftrightarrow l \implies l \leftrightarrow k$ symmetry*
3. *($k \leftrightarrow l$ and $l \leftrightarrow m$) $\implies k \leftrightarrow m$ transitivity*

Proof. Obviously, $\leftrightarrow$ is reflexive and symmetric. Let us see why it is transitive. It is enough to show

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

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

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

which implies $k \to m$. □

Proposition 3.4.9. *One can partition all states of a homogeneous Markov chain $(f_i)_{i=0}^{\infty}$ into equivalence classes with respect to $\leftrightarrow$, that means*

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

with

1. $X_{m_1} \cap X_{m_2} = \emptyset$ *for* $m_1 \neq m_2$,
2. $k, l \in X_m \quad \Longrightarrow \quad k \leftrightarrow l$,
3. $k \leftrightarrow l \quad \Longrightarrow \quad$ *k and l belong to the same X_m.*

Definition 3.4.10. A homogeneous Markov chain $(f_i)_{i=0}^{\infty}$ is called **irreducible** $:\Longleftrightarrow$ There is only one equivalence class with respect to $\leftrightarrow$.

A model for radiation damage

Let $X = \{0, 1, \ldots, K\}$. Then $l \in X$ stands for the degree of radiation damage of one fixed person as follows:

$$\begin{array}{lcl} 0 & = & \text{healthy} \\ 1 & = & \text{initial damage} \\ \left.\begin{array}{l} 2 \\ \vdots \\ K-1 \end{array}\right\} & = & \text{non-visible stages} \\ K & = & \text{visible stage} \end{array}$$

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

$$\begin{array}{rcl} k \to k & = & \text{staying at the same stage of damage} \\ k \to k-1 & = & \text{recovery} \\ k \to k+1 & = & \text{amplification} \end{array}$$

Assume the transition matrix

$$T = \begin{pmatrix} 1 & 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ q_1 & r_1 & p_1 & 0 & \dots & 0 & 0 & 0 \\ 0 & q_2 & r_2 & p_2 & 0 & \dots & 0 & 0 \\ \vdots & & & & & & & \\ 0 & 0 & & \dots & 0 & q_{K-1} & r_{K-1} & p_{K-1} \\ 0 & 0 & & & \dots & 0 & 0 & 1 \end{pmatrix}$$

Example 3.4.11. *The model to according to Reid/Landau:*
$q_l = 1 - \frac{l}{K}$ *recovery probability*
$r_l = 0$
$p_l = \frac{l}{K}$ *amplification probability*
for $l = 1, \dots, K-1$.

Proposition 3.4.12. *1. The absorbing states consist of* $\{0, K\}$.

2. *The chain consists of the 3 equivalence classes* $\{0\}, \{1, \dots, K-1\}, \{K\}$.

3. *The recovery probability is*

$$\mathbb{P}(\{\omega \in \Omega : \exists i : f_i(\omega) = 0\}) = 1 - \frac{1}{2^{K-1}}.$$

Proof. The assertions 1. and 2. are clear. We will consider examples for the third assertion. Choose $\boxed{\text{K=3}}$

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

We have

$$\begin{aligned}
\mathbb{P}(\{\omega : \exists i : f_i(\omega) = 0\}) &= p_{10} \sum_{n=0}^{\infty} \mathbb{P}(f_n = 1 | f_0 = 1) \\
&= p_{10} \sum_{n=0,\ n \text{ even}}^{\infty} \mathbb{P}(f_n = 1 | f_0 = 1) \\
&= p_{10} \sum_{n=0}^{\infty} \left(\frac{1}{3}\right)^{2n} = \frac{2}{3} \frac{1}{1 - \frac{1}{9}} = 1 - \frac{1}{2^2}
\end{aligned}$$

Choose $\boxed{\text{K=4}}$

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

$$\begin{aligned}
\mathbb{P}(\{\omega : \exists i : f_i(\omega) = 0\}) &= p_{10} \left(1 + \sum_{m=1}^{\infty} \mathbb{P}\left(\begin{array}{l} \text{way from } 1 \to 1 \text{ which passes} \\ 1 \to 2 \text{ exactly } m \text{ times} \end{array} \right) \right) \\
&= p_{10} \left(1 + \sum_{m=1}^{\infty} \mathbb{P}\left(\begin{array}{l} \text{way from } 1 \to 1 \text{ which passes} \\ 1 \to 2 \text{ exactly one times} \end{array} \right)^m \right) \\
&= p_{10} \frac{1}{1 - \mathbb{P}\left(\begin{array}{l} \text{way from } 1 \to 1 \text{ which passes} \\ 1 \to 2 \text{ exactly one times} \end{array} \right)} \\
&= \frac{p_{10}}{1 - p_{12} p_{21} \left(1 + \mathbb{P}\left(\begin{array}{l} \text{way from } 2 \to 2 \text{ which does} \\ \text{not pass } 1 \end{array} \right) \right)},
\end{aligned}$$

where

$$\begin{aligned}\mathbb{P}\left(\begin{array}{l}\text{way from } 2 \to 2 \text{ which does} \\ \text{not pass } 1\end{array}\right) &= p_{23}p_{32} + (p_{23}p_{32})^2 + \ldots \\ &= \frac{p_{23}p_{32}}{1 - p_{23}p_{32}}.\end{aligned}$$

Hence

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

If we choose $\boxed{\text{K=5}}$
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. □

3.4.3 Periodic and aperiodic states

We recall that we defined for a homogeneous Markov chain $(f_i)_{i=0}^{\infty}$

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

Definition 3.4.13. *Let $(f_i)_{i=0}^{\infty}$ be a homogeneous Markov chain with state space X.*

1. *A state k has* **period** $d \in \{1, 2, \ldots\}$ $:\Longleftrightarrow$
   - $p^{(i)}(k,k) = 0$ *if* $i = jd + r$ *where* $j = 0, 1, \ldots$ *and* $r = 1, \ldots, d-1$
   - d *is the largest number satisfying this property*

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

2. *k is called* **aperiodic** $:\Longleftrightarrow d(k) = 1$.

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

Remark

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 $i \in \{1, 2, \dots\}$ such that $p^{(i)}(k,k) > 0$, then $d(k) \leq i < \infty$. This one can see as follows:
   Fix $k \in X$ and let

   $$\mathcal{M} := \{d \in \{1,2,\dots\} : \ p^{(i)}(k,k) = 0 \text{ for all } i = jd + r \\ \text{and } j = 0, 1, \dots, \quad r = 1, \dots d-1\}$$

   Then

   - $\mathcal{M} \neq \emptyset$ since $1 \in \mathcal{M}$.
   - Assume $d \in \mathcal{M}$ with $d > i$. Then

     $$0 = p^{(1)}(k,k) = \dots = p^{(i)}(k,k) = \dots = p^{(d-1)}(k,k),$$

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

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

Proposition 3.4.14. *Let $k, l \in X$. Then*

$$k \leftrightarrow l \quad \Longrightarrow \quad d(k) = d(l),$$

that means periodicity is an equivalence property.

Proof. Let m, n be such that $p^{(m)}(k,l) > 0$, and $p^{(n)}(l,k) > 0$ and s such that $p^{(s)}(l,l) > 0$. Then

$$p^{(m+n)}(k,k) \geq p^{(m)}(k,l)p^{(n)}(l,k) > 0$$

and

$$p^{(m+n+s)}(k,k) \geq p^{(m)}(k,l)p^{(s)}(l,l)p^{(n)}(l,k) > 0.$$

Hence $m+n$ is divisible by $d(k)$, and so is $m+n+s$. This yields s is divisible by $d(k)$. Since $s \in \{d(l), 2d(l), 3d(l), \dots\}$ it follows that $d(l)$ is divisible by $d(k)$. That $d(k)$ is divisible by $d(l)$ one can get in the same way. This implies $d(l) = d(k)$. ☐

Example 3.4.15. For the model about radiation damage from Reid/Landau (Example 3.4.11) we have
$d(0) = 1, \quad d(K) = 1, \quad d(l) = 2 \quad \text{for} \quad l \in \{1, \dots, K-1\}$.

Example 3.4.16. Consider a Markov chain with transition matrix

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

Then $d(k) = 1$ for $k = 0, \ldots, 3$.

3.4.4 Persistent and transient states

Definition 3.4.17. Let $(f_i)_{i=0}^{\infty}$ be a homogeneous Markov chain with state space X and let $k \in X$.

1. k is called **persistent** (or **recurrent**) $:\Longleftrightarrow$

$$\mathbb{P}(\{\omega : \exists i \geq 1 \text{ with } f_i(\omega) = k\} | \{\omega : f_0(\omega) = k\}) = 1.$$

2. k is called **transient** $:\Longleftrightarrow$ k is not persistent.

3. The expression

$$\mathbf{f}_{kj}^{(n)} := \mathbb{P}(f_n = j, f_{n-1} \neq j, \ldots, f_1 \neq j | f_0 = k)$$

   is called the **first passage distribution from state k to state j.**

4. We define

$$\mathbf{f}_{kj} := \sum_{n=1}^{\infty} \mathbf{f}_{kj}^{(n)}.$$

Interpretation:

1. k is persistent $:\Longleftrightarrow$ with probability 1 the Markov chain returns to k.

2. k is transient $:\Longleftrightarrow$ with positive probability there is no return to the state k.

Proposition 3.4.18. *If $p^{(n)}(k,k) = \mathbb{P}(f_n = k | f_0 = k)$, $k \in X$, then the following assertions are equivalent:*

1. k is persistent.

2. $\sum_{n=1}^{\infty} p^{(n)}(k,k) = \infty$.

3. $\mathbf{f}_{kk} = 1$.

Lemma 3.4.19. *Let $N(k)$ be the number of returns of the Markov chain $(f_i)_{i=0}^{\infty}$ to $k \in X$ after step 0. Let $s \in \{1, 2, 3, \dots\}$. Then one has*

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

Proof of Lemma 3.4.19 for $s = 2$.
Using the Markov property and the homogeneity we get

$$\begin{aligned}
& \mathbb{P}(N(k) \geq 2 | f_0 = k) \\
= \; & \mathbb{P}(\exists i > j \geq 1 : f_i = f_j = k | f_0 = k) \\
= \; & \sum_{i > j \geq 1} \mathbb{P}(f_i = k, f_{i-1} \neq k, \dots, f_{j+1} \neq k, f_j = k, f_{j-1} \neq k, \dots, f_1 \neq k | f_0 = k) \\
= \; & \sum_{i > j \geq 1} \mathbb{P}(f_i = k, f_{i-1} \neq k, \dots, f_{j+1} \neq k | f_j = k) \\
& \qquad\qquad \times \mathbb{P}(f_j = k, f_{j-1} \neq k, \dots, f_1 \neq k | f_0 = k) \\
= \; & \sum_{i > j \geq 1} \mathbb{P}(f_{i-j} = k, f_{i-j-1} \neq k, \dots, f_1 \neq k | f_0 = k) \\
& \qquad\qquad \times \mathbb{P}(f_j = k, f_{j-1} \neq k, \dots, f_1 \neq k | f_0 = k) \\
= \; & \sum_{\tilde{i}, j \geq 1} \mathbb{P}(f_{\tilde{i}} = k, f_{\tilde{i}-1} \neq k, \dots, f_1 \neq k | f_0 = k) \\
& \qquad\qquad \times \mathbb{P}(f_j = k, f_{j-1} \neq k, \dots, f_1 \neq k | f_0 = k) \\
= \; & \left(\sum_{j=1}^{\infty} \mathbb{P}(f_j = k, f_{j-1} \neq k, \dots, f_1 \neq k | f_0 = k) \right)^2 \\
= \; & \mathbf{f}_{kk}^2 .
\end{aligned}$$

□

Proof of Proposition 3.4.18.
It holds

$$\begin{aligned}
& \mathbb{P}(\{\exists i \geq 1 \text{ with } f_i(\omega) = k\} | \{f_0(\omega) = k\}) \\
& \qquad = \sum_{n=1}^{\infty} \mathbb{P}(f_n = k, f_{n-1} \neq k, \dots, f_1 \neq k | f_0 = k) = \mathbf{f}_{kk} .
\end{aligned}$$

Hence that k is persistent is equivalent to $\mathbf{f}_{kk} = 1$.

Assume now that $\mathbb{P}(f_0 = k) = 1$. The expression

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

is a random variable. We have

$$\begin{aligned} \mathbb{E}N(k) &= \sum_{n=1}^{\infty} \mathbb{E}\mathbb{I}_{\{f_n=k\}} \\ &= \sum_{n=1}^{\infty} \mathbb{P}(f_n = k | f_0 = k) = \sum_{n=1}^{\infty} p^{(n)}(k,k). \end{aligned} \tag{3.3}$$

We compute

$$\begin{aligned} & \sum_{s=1}^{\infty} \mathbb{P}(N(k) \geq s | f_0 = k) \\ = \; & \mathbb{P}(N(k) \geq 1 | f_0 = k) + \mathbb{P}(N(k) \geq 2 | f_0 = k) + \mathbb{P}(N(k) \geq 3 | f_0 = k) + \ldots \\ = \; & \mathbb{P}(N(k) = 1 \text{ or } N(k) \geq 2 | f_0 = k) + \mathbb{P}(N(k) \geq 2 | f_0 = k) \\ & + \mathbb{P}(N(k) \geq 3 | f_0 = k) + \ldots \\ = \; & \mathbb{P}(N(k) = 1 | f_0 = k) + 2\mathbb{P}(N(k) \geq 2 | f_0 = k) + \mathbb{P}(N(k) \geq 3 | f_0 = k) + \ldots \\ = \; & \mathbb{P}(N(k) = 1 | f_0 = k) + 2\mathbb{P}(N(k) = 2 | f_0 = k) + 3\mathbb{P}(N(k) \geq 3 | f_0 = k) + \ldots \\ = \; & \sum_{m=1}^{\infty} m\mathbb{P}(N(k) = m | f_0 = k) = \mathbb{E}N(k). \end{aligned} \tag{3.4}$$

Hence, by Lemma 3.4.19,

$$\mathbb{E}N(k) = \sum_{s=1}^{\infty} \mathbb{P}(N(k) \geq s | f_0 = k) = \sum_{s=1}^{\infty} (\mathbf{f}_{kk})^s .$$

Comparing this with (3.3) gives

$$\sum_{n=1}^{\infty} p^{(n)}(k,k) = \sum_{s=1}^{\infty} (\mathbf{f}_{kk})^s ,$$

which implies

$$0 \leq \mathbf{f}_{kk} < 1 \Longleftrightarrow \sum_{n=1}^{\infty} p^{(n)}(k,k) < \infty.$$

□

Example 3.4.20. (tossing a coin)
Consider the Markov chain $(f_i)_{i=0}^{\infty}$ with $f_0 = 0$ and state space $X := \{\dots, -3, -2, -1, 0, 1, 2, 3, \dots\}$ and for a fixed $p \in (0,1)$ the transition probabilities

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

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

Interpretation:

1. tossing a coin → 'heads' or 'tails'
2. start with $f_0 = 0$.
3. if one tosses the coin, then in case of
   'heads' → $f_{i+1}(\omega) = f_i(\omega) + 1$ (gain),
   'tails' → $f_{i+1}(\omega) = f_i(\omega) - 1$ (loss).

The coin is 'fair' iff 0 is a persistent state. By Proposition 3.4.18 the state 0 is persistent $\Leftrightarrow \sum_{n=1}^{\infty} p^{(n)}(0,0) = \infty$.

We will compute $p^{(n)}(0,0)$:

Clearly, $p^{(n)}(0,0) = 0$ for all odd n. For even n it holds

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

Stirling's formula

$$\lim_{m \to \infty} \frac{m!}{\sqrt{2\pi m} \quad e^{-m} m^m} = 1$$

gives

$$\begin{aligned} p^{(2m)}(0,0) &= \frac{(2m)!}{(m!)^2}(p(1-p))^m \\ &\sim \frac{\sqrt{4\pi m}\, e^{-2m}(2m)^{2m}}{(\sqrt{2\pi m}e^{-m}m^m)^2}(p(1-p))^m \\ &= \frac{2^{2m}}{\sqrt{\pi m}}(p(1-p))^m. \end{aligned}$$

Hence 0 is persistent $\Longleftrightarrow \sum_{n=1}^{\infty} p^{(n)}(0,0) = \infty$

$\Longleftrightarrow \sum_{m=m_0}^{\infty} \frac{2^{2m}}{\sqrt{\pi m}}(p(1-p))^m = \frac{1}{\sqrt{\pi}} \sum_{m=m_0}^{\infty} \frac{(4p(1-p))^m}{\sqrt{m}} = \infty$ for m_0 large

$\Longleftrightarrow \quad 4p(1-p) \geq 1 \quad \Longleftrightarrow \quad p = \frac{1}{2}$.

Hence the coin is 'fair' $\Longleftrightarrow \quad p = \frac{1}{2}$.

3.4.5 Decomposition of the state space

Definition 3.4.21. *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** $:\Longleftrightarrow k \leftrightarrow l$ *for all $k, l \in C$.*
2. *A subset $C \subseteq X$ is called* **closed** $:\Longleftrightarrow \forall k \in C \quad \forall l \in X \setminus C \quad \forall n = 1, 2, \ldots$
$$p^{(n)}(k,l) = 0.$$

Remark 3.4.22. 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.

Proposition 3.4.23. *Let $[(f_i)_{i=0}^{\infty}, p_o, T]$ 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 $X = X_T + X_P$ (i.e. $X = X_T \cup X_P$ and $X_T \cap X_P = \emptyset$) and $X_P = \bigcup_m C_m$, 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

- $\underline{k \in C(k)}$: it always holds that $k \leftrightarrow k$.
- $\underline{C(k) \text{ is closed:}}$ $m \in C(k),\ l \in X \setminus C(k)$
  Assume that there is an $n \in \{1, 2, \dots\}$ 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.
- $\underline{C(k) \text{ is irreducible:}}$ Let $l, m \in C(k)$.

  $$\left.\begin{array}{l} k \to l \text{ by definition} \\ l \to k \text{ by persistence} \end{array}\right\} \Rightarrow k \leftrightarrow l.$$

  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. □

3.4.6 Summary of the classification

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

$\boxed{k \text{ is an } \textbf{absorbing} \text{ state:}}$

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

$\boxed{k \to l \text{ (} k \textbf{ reaches } l\text{):}}$

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

$\boxed{k \leftrightarrow l \text{ (} k \text{ and } l \text{ are } \textbf{communicating}\text{):}}$

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

$\boxed{k \text{ has } \textbf{period } d(k)\text{:}}$
if $d(k)$ is the largest number m for which it holds

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

that means the Markov chain may only return after time $d(k)$.
If $d(k) = 1$ we call the state k **aperiodic.**

$\boxed{\textbf{decomposition} \text{ of the state space } X\text{:}}$
with respect to $\leftrightarrow$ one gets equivalence classes

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

such that

1. $X_l \cap X_k = \emptyset, \quad l \neq k$
2. $k, l \in X_m \Rightarrow \quad k \leftrightarrow l$
3. $k \leftrightarrow l \Rightarrow k$ and l belong to the same X_m

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

Assume $\mathbb{P}(f_0 = k) = 1$. Then the (extended) random variable

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

is the **recurrence time.**
Recall, k is persistent (=recurrent) $\iff \quad \mathbf{f}_{kk} = 1 = \sum_{n=1}^{\infty} \mathbf{f}_{kk}^{(n)}$.

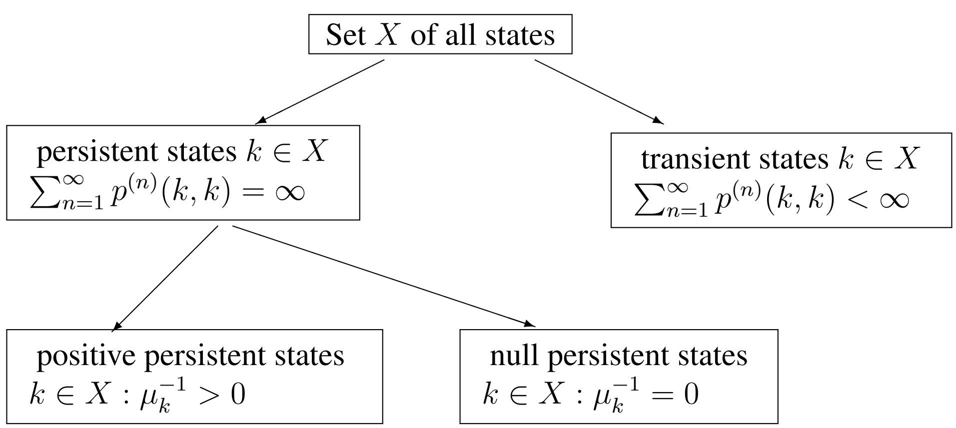
$$\mathbb{P}(T_k = n) = \mathbb{P}(f_n = k, f_{n-1} \neq k, \dots, f_1 \neq k | f_0 = k) = \mathbf{f}_{kk}^{(n)}.$$

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

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

We define the **mean recurrence time**

$$\mu_k := \mathbb{E}T_k = \sum_{n=1}^{\infty} n\mathbf{f}_{kk}^{(n)}.$$



We have the disjoint union

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

where the C_m's are pair-wise disjoint, irreducible and closed.

3.4.7 Ergodic Theorem (first version) and stationary distribution

Example of precipitation:

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

$$T^5 = \begin{pmatrix} 0.305 & 0.695 \\ 0.290 & 0.710 \end{pmatrix}$$

Observation: The rows of T^5 are more similar to each other than the rows of T.

Definition 3.4.24. Let $[(f_i)_{i=0}^{\infty}, p, T]$ be a Markov chain with state space $X = \{0, \dots, K\}$ and $p(k) > 0$ for all $k \in X$. Let

$$T = \begin{pmatrix} p_{00} & \dots & p_{0K} \\ \vdots & & \vdots \\ p_{K0} & \dots & p_{KK} \end{pmatrix}$$

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

$$\underbrace{T \circ \dots \circ T}_{n \text{ times}} =: \left(p_{kl}^{(n)}\right)_{k,l=0}^{K}.$$

1. $[(f_i)_{i=0}^{\infty}, p, T]$ is called **ergodic** $:\Longleftrightarrow \exists s_l > 0$, $l = 0, \dots, K$ such that
   (a) $s_0 + \dots + s_K = 1$,
   (b) $\lim_{n\to\infty} p_{kl}^{(n)} = s_l \quad$ for all $k, l \in X$.
2. The vector $s = (s_0, \dots, s_K)$ is called **stationary distribution** $:\Longleftrightarrow$
   (a) $s_0 + \dots + s_K = 1$ with $0 \le s_l \le 1$,
   (b) $s \circ T = s \quad \left(\text{i.e. } \sum_{m=0}^{K} s_m p_{ml} = s_l\right)$.

Proposition 3.4.25. *(ergodic $\Longrightarrow$ unique stationary distribution)*
An ergodic Markov chain has a unique stationary distribution.

Proof. STEP 1: Existence
Let

$$A = \begin{pmatrix} s_0 & \dots & s_K \\ \vdots & & \vdots \\ s_0 & \dots & s_K \end{pmatrix}.$$

Then

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

Hence $s \circ T = s$ is a stationary distribution.
STEP 2: Uniqueness
Assume that $b = (b_0, \dots, b_K)$ is a stationary distribution. Then

$$b \circ T = b, \quad b \circ T^n = b,$$

and by $n \to \infty$, $b \circ A = b$. But that means

$$(b_0 + \dots + b_K)s_l = b_l \quad \text{and } s_l = b_l.$$

$\square$

Example (stationary distribution $\not\Longrightarrow$ ergodic)
Let $X = \{0,1\}$, $\quad T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.
Then
$$s \circ \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = s \Longleftrightarrow \begin{array}{c} s_2 = s_1 \\ s_1 = s_2 \end{array} \Longleftrightarrow s_1 = s_2 = \frac{1}{2}.$$
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
$$T^n = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \text{ for even } n \text{ and } T^n = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \text{ for odd } n,$$
which means $\lim_{n\to\infty} p_{kl}^{(n)}$ does not exist.

Example (stationary distribution which is not unique)
$$T = \begin{pmatrix} 1 & 0 & 0 \\ \frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\ 0 & 0 & 1 \end{pmatrix}$$
We compute
$$(s_0, s_1, s_2) \circ T = s \Longleftrightarrow \begin{array}{rcl} s_0 + \frac{1}{3}s_1 & = & s_0 \\ \frac{1}{3}s_1 & = & s_1 \\ \frac{1}{3}s_1 + s_2 & = & s_2 \end{array} \Longleftrightarrow s_1 = 0.$$
This implies one gets the set of stationary distributions
$$(s_0, s_1, s_2) = (s_0, 0, 1 - s_0), \quad \text{if } s_0 \in [0,1],$$
hence no unique one.

Theorem 3.4.26. *(Ergodic Theorem, first version)*
Assume a homogeneous Markov chain $[(f_i)_{i=0}^{\infty}, p, T]$ with state space $X = \{0, \ldots, K\}$, and transition matrices
$$\underbrace{T \circ \cdots \circ T}_{n-times} = \left(p_{kl}^{(n)}\right)_{k,l=0}^{K}.$$

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

$$m_l^{(n)} := \min_k p_{kl}^{(n)} \leq \max_k p_{kl}^{(n)} =: M_l^{(n)},$$

where

$$\begin{pmatrix} p_{00}^{(n)} & \ldots & p_{0l}^{(n)} & \ldots & p_{0K}^{(n)} \\ \vdots & & \vdots & & \vdots \\ p_{K0}^{(n)} & \ldots & p_{Kl}^{(n)} & \ldots & p_{KK}^{(n)} \end{pmatrix}.$$

STEP 1 It holds $m_l^{(n)} \leq m_l^{(n+1)}$:

$$\begin{aligned} m_l^{(n+1)} &= \min_k p_{kl}^{(n+1)} \\ &= \min_k \sum_{s=0}^{K} p_{ks}\, p_{sl}^{(n)} \\ &\geq \min \sum_{s=0}^{K} p_{ks} \left(\min_r p_{rl}^{(n)} \right) \\ &= \min_r p_{rl}^{(n)} = m_l^{(n)}. \end{aligned}$$

STEP 2 Let $\varepsilon := \min_{k,l} p_{kl}^{(1)} > 0$. We show

$$M_l^{(n+1)} - m_l^{(n+1)} \leq (1 - \varepsilon) \left(M_l^{(n)} - m_l^{(n)} \right).$$

For this we compute

$$\begin{aligned} p_{kl}^{(n+1)} &= \sum_{s=0}^{K} p_{ks} p_{sl}^{(n)} \\ &= \sum_{s=0}^{K} \left(p_{ks} - \varepsilon p_{ls}^{(n)} \right) p_{sl}^{(n)} + \varepsilon \sum_{s=0}^{K} p_{ls}^{(n)} p_{sl}^{(n)}. \end{aligned}$$

Because of $p_{ks} - \varepsilon p_{ls}^{(n)} \geq 0$ and $p_{sl}^{(n)} \geq m_l^{(n)}$ we get

$$
\begin{aligned}
p_{kl}^{(n+1)} &\geq m_l^{(n)} \sum_{s=0}^{K} \left(p_{ks} - \varepsilon p_{ls}^{(n)}\right) + \varepsilon \sum_{s=0}^{K} p_{ls}^{(n)} p_{sl}^{(n)} \\
&= m_l^{(n)} \sum_{s=0}^{K} \left(p_{ks} - \varepsilon p_{ls}^{(n)}\right) + \varepsilon p_{ll}^{(2n)} \\
&= m_l^{(n)} \left(1 - \varepsilon\right) + \varepsilon p_{ll}^{(2n)},
\end{aligned}
$$

which implies

$$m_l^{(n+1)} \geq m_l^{(n)}(1-\varepsilon) + \varepsilon p_{ll}^{(2n)}.$$

In the same way:

$$M_l^{(n+1)} \leq M_l^{(n)}(1-\varepsilon) + \varepsilon p_{ll}^{(2n)}.$$

STEP 3 By iteration we get

$$M_l^{(n+1)} - m_l^{(n+1)} \leq (1-\varepsilon)^n \left(M_l^{(1)} - m_l^{(1)}.\right)$$

Hence

$$M_l^{(n+1)} - m_l^{(n+1)} \to 0$$

which implies that the limit $\lim_{n\to\infty} p_{kl}^{(n)}$ exists. From

$$\varepsilon \leq m_l^{(n)} \leq m_l^{(n+1)} \leq p_{kl}^{(n+1)}$$

it follows $\lim_{n\to\infty} p_{kl}^{(n)} > 0$. □

Example: frequency of genes in a population

Assumptions

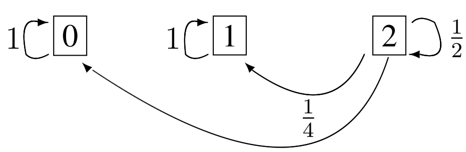
- a large population
- diploid
- no selection, no mutation, no immigration, no migration

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.

$$\begin{array}{ccccl} \text{mother} & & \text{father} & & \\ AA & \times & AA & \rightarrow & AA \\ aa & \times & aa & \rightarrow & aa \\ Aa & \times & Aa & \rightarrow & \frac{1}{4}AA, \quad \frac{1}{2}Aa, \quad \frac{1}{4}aa \end{array}$$

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

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



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

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

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

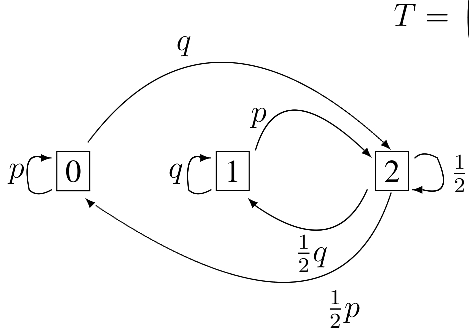
$$p = \text{ frequency of allele A} \qquad \text{and} \quad q = 1 - p.$$

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

$$\begin{array}{ccl} \text{1 parent} & & \text{child} \\ AA & \rightarrow & pAA, \quad qAa \\ aa & \rightarrow & pAa, \quad qaa \\ Aa & \rightarrow & \frac{1}{2}pAA, \quad \frac{1}{2}\underbrace{(q+p)}_{=1}Aa, \quad \frac{1}{2}qaa \end{array}$$

Hence the transition matrix is (assume $AA = 0$, $aa = 1$, $Aa = 2$)

$$T = \begin{pmatrix} p & 0 & q \\ 0 & q & p \\ \frac{1}{2}p & \frac{1}{2}q & \frac{1}{2} \end{pmatrix}.$$



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.'

3.5 Some more tools from probability theory

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

3.5.1 Stopping times

Definition 3.5.1. *A sequence of* $\sigma-$*algebras* $(\mathcal{F}_n)_{n=0}^{\infty}$ *with* $\mathcal{F}_0 \subseteq \mathcal{F}_1 \subseteq \mathcal{F}_2 \subseteq \cdots \subseteq \mathcal{F}$ *is called a* **filtration.**

Example 3.5.2. Let $(f_i)_{i=0}^{\infty}$ be a stochastic process (i.e. a sequence of random variables $f_0, f_1, \ldots$).Define

$$\mathcal{F}_n := \sigma(f_0, f_1, \ldots, f_n),$$

which is the smallest $\sigma-$algebra such that $f_0, f_1, \ldots, f_n$ are measurable.
Special case:
Let us take $A_1, A_2, \cdots \in \mathcal{F}$ and set

$$\begin{aligned} f_0 &:= \mathbb{1}_{\Omega}, \\ f_i &:= \mathbb{1}_{A_i}, \quad i = 1, 2, \ldots \end{aligned}$$

Then

$$\begin{aligned} \mathcal{F}_0 &:= \{\emptyset, \Omega\}, \\ \mathcal{F}_1 &:= \{A_1, A_1^c, \emptyset, \Omega\}, \\ \mathcal{F}_2 &:= \{A_1, A_2, A_1^c, A_2^c, A_1 \cup A_2, A_1 \cup A_2^c, \ldots\}, \\ \mathcal{F}_3 &:= \sigma(\mathbb{1}_{\Omega}, \mathbb{1}_{A_1}, \mathbb{1}_{A_2}, \mathbb{1}_{A_3}), \\ &\vdots \end{aligned}$$

Definition 3.5.3. *A random variable* $T : \Omega \to \{0, 1, \ldots, +\infty\}$ *is called a* **stopping time** *(with respect to* $(\mathcal{F}_n)$*) if*

$$\{T = n\} \in \mathcal{F}_n \quad \textit{for all} \quad n = 0, 1, 2, \ldots$$

Stopping time: yes or no?

Let $(f_n)_{n=0}^{\infty}$ be a stochastic process on $(\Omega, \mathcal{F}, \mathbb{P})$. We assume $\mathcal{F}_n = \sigma\{f_0, \ldots, f_n\}$. Then we define $T : \Omega \to \{0, 1, \ldots, \infty\}$ by

1. $T(\omega) := \inf\{n : f_n(\omega) > 1\}$,
2. $T(\omega) := \max\{n : f_n(\omega) \leq 1\}$,
3. $T(\omega) := \inf\{n : f_n(\omega) + f_{n+1}(\omega) > 1\}$.

Now let $(f_n)_{n=0}^{\infty}, (g_n)_{n=0}^{\infty}$ be stochastic processes and define

$$\mathcal{F}_n := \sigma\{f_0, \ldots, f_n, g_0, \ldots, g_n\}.$$

4. $T(\omega) := \inf\{n : f_n(\omega) > g_n(\omega)\}$,
5. $T(\omega) := \max\{n : f_n(\omega) > g_{n+1}(\omega)\}$.

To find out which $T's$ are stopping times we compute

1.
$$\begin{aligned} \{\omega : T(\omega) = i\} &= \{\omega : \inf\{n : f_n(\omega) > 1\} = i\} \\ &= \bigcap_{n=0}^{i-1} \underbrace{\{f_n(\omega) \leq 1\}}_{\in \mathcal{F}_n \subseteq \mathcal{F}_{i-1}} \cap \underbrace{\{f_i(\omega) > 1\}}_{\in \mathcal{F}_i} \end{aligned}$$

   Hence $\{\omega : T(\omega) = i\} \in \mathcal{F}_i$ and T is a stopping time.

2.
$$\begin{aligned} \{\omega : T(\omega) = i\} &= \{\omega : \max\{n : f_n(\omega) \leq 1\} = i\} \\ &= \{f_i(\omega) \leq 1\} \cap \bigcap_{n=i+1}^{\infty} \{f_n(\omega) > 1\} \end{aligned}$$

   This means T is not a stopping time.

3.
$$\{T = i\} = \{f_i + f_{i+1} > 1\} \cap \bigcap_{k=0}^{i-1} \{f_k + f_{k+1} \leq 1\}$$

   So T is not a stopping time.

The cases 4. and 5. are left as an exercise.

3.5.2 About convergence of sequences

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space.

Theorem 3.5.4. *(dominated convergence)*
Let $g, f, f_1, f_2, \ldots$ be random variables with

$$|f_n| \leq g,$$

$$\mathbb{E}g < \infty,$$

$$\text{and} \quad \lim_{n\to\infty} f_n(\omega) = f(\omega), \quad \forall \omega \in \Omega.$$

Then $\mathbb{E}|f| < \infty$ and $\mathbb{E}f = \lim_{n\to\infty} \mathbb{E}f_n$.

Proof: see Proposition 3.2.7 in [1].

Theorem 3.5.5. *(The strong law of large numbers for i.i.d. random variables (Kolmogorov))*
Let $f_1, f_2, \ldots$ be i.i.d.random variables with $\mathbb{E}|f_1| < \infty$. Then

$$\mathbb{P}\left(\left\{\omega : \frac{\sum_{i=1}^n f_i(\omega)}{n} \stackrel{n\to\infty}{\to} \mathbb{E}f_1\right\}\right) = 1.$$

Proof*: see [4] (Thm3, p.391) or [1] section 4.2.*

3.6 The strong Markov property

Proposition 3.6.1. *Let $(f_i)_{i=0}^\infty$ be a homogeneous Markov chain,*

$$\mathcal{F}_i := \sigma(f_0, \ldots, f_i),$$

and $T : \Omega \to \{0, 1, 2, \ldots\}$ (which implies that $\mathbb{P}(T < \infty) = 1$) a stopping time. Then

1. *it holds for all $x_0, x_1 \in X$*

$$\mathbb{P}(f_{T+1} = x_1 | \{f_T = x_0\} \cap B) = \mathbb{P}(f_{T+1} = x_1 | \{f_T = x_0\})$$

if $\mathbb{P}(\{f_T = x_0\} \cap B) > 0$ and

$$B \cap \{T = k\} \in \mathcal{F}_k \quad \text{for } k = 0, 1, \ldots$$

2. *it holds for all* $x_0, \dots, x_k \in X$

$$\begin{aligned} &\mathbb{P}(f_{T+k} = x_k, \dots, f_{T+1} = x_1 | \{f_T = x_0\} \cap B) \\ &\qquad = \mathbb{P}(f_{T+k} = x_k, \dots, f_{T+1} = x_1 | \{f_T = x_0\}) \end{aligned}$$

if $\mathbb{P}(\{f_T = x_0\} \cap B) > 0$ *and*

$$B \cap \{T = k\} \in \mathcal{F}_k \quad \textit{for } k = 0, 1, \dots$$

Remarks:

1. $\mathcal{F}_T := \{A \subseteq \Omega : A \cap \{T = k\} \in \mathcal{F}_k, \forall k \geq 0\}$ is the $\sigma-$ algebra generated by T.

2. A special case of Proposition 3.6.1:
Set $T \equiv i$. Then $B \cap \underbrace{\{\omega : T(\omega) = i\}}_{\Omega} \in \mathcal{F}_i \Longleftrightarrow B \in \mathcal{F}_i$.
For $k \neq i$ one has $B \cap \underbrace{\{\omega : T(\omega) = k\}}_{\emptyset} \in \mathcal{F}_k$.
Hence $B \cap \{\omega : T(\omega) = k\} \in \mathcal{F}_k$ for all $k \geq 0 \Longleftrightarrow B \in \mathcal{F}_i$. So

$$\mathbb{P}(f_{i+1} = x_1 | \{f_i = x_0\} \cap B) = \mathbb{P}(f_{i+1} = x_1 | f_i = x_0) \quad \forall B \in \mathcal{F}_i$$

is equivalent to the Markov property (compare Theorem 3.3.1)

Proof. (of Proposition 3.6.1)
$(f_i)_{i=0}^{\infty}$ is assumed to be homogeneous. Then

$$\begin{aligned} &\mathbb{P}(f_{T+1} = x_1 | \{f_T = x_0\} \cap B) \\ &\quad = \frac{\mathbb{P}(\{f_{T+1} = x_1\} \cap \{f_T = x_0\} \cap B)}{\mathbb{P}(\{f_T = x_0\} \cap B)} \\ &\quad = \frac{1}{\mathbb{P}(f_T = x_0, B)} \sum_{i=0}^{\infty} \mathbb{P}(f_{i+1} = x_1, f_T = x_0, B, T = i) \\ &\quad = \frac{1}{\mathbb{P}(f_T = x_0, B)} \sum_{\mathbb{P}(f_i = x_0, B, T = i) > 0} \mathbb{P}(f_{i+1} = x_1 | f_i = x_0, B, T = i) \\ &\qquad\qquad \times \mathbb{P}(f_i = x_0, B, T = i) \end{aligned}$$

It holds by assumption $B \cap \{T = i\} \in \mathcal{F}_i$. This implies

$$\{f_i = x_0\} \cap B \cap \{T = i\} = \{f_i = x_0\} \cap \underbrace{\{\ \dots\ \}}_{\in \mathcal{F}_{i-1}},$$

Using the Markov property one gets

$$\begin{aligned}\mathbb{P}(f_{i+1} = x_1 | f_i = x_0, B, T = i) &= \mathbb{P}(f_{i+1} = x_1 | \{f_i = x_0\} \cap \underbrace{\{\ \dots\ \}}_{\in \mathcal{F}_{i-1}}) \\ &= \mathbb{P}(f_{i+1} = x_1 | f_i = x_0) \\ &= \mathbb{P}(f_1 = x_1 | f_0 = x_0),\end{aligned}$$

since the Markov chain is homogeneous. Hence we can continue the above computation with

$$\begin{aligned}&\mathbb{P}(f_{T+1} = x_1 | \{f_T = x_0\} \cap B) \\ &= \frac{1}{\mathbb{P}(f_T = x_0, B)} \sum_{i=0}^{\infty} \mathbb{P}(f_1 = x_1 | f_0 = x_0) \mathbb{P}(f_i = x_0, B, T = i) \\ &= \frac{\mathbb{P}(f_1 = x_1 | f_0 = x_0)}{\mathbb{P}(f_T = x_0, B)} \mathbb{P}(f_T = x_0, B) \\ &= \mathbb{P}(f_1 = x_1 | f_0 = x_0).\end{aligned}$$

Now the same computation with $B := \Omega$ yields

$$\mathbb{P}(f_{T+1} = x_1 | f_T = x_0) = \mathbb{P}(f_1 = x_1 | f_0 = x_0).$$

$\square$

3.7 Ergodic Theorem (second version)

Let $X = \{0, 1, \dots\}$.

Lemma 3.7.1. *Assume that the Markov chain $[(f_i)_{i=0}^{\infty}, p, T]$ is irreducible and that*

$$\exists m \in X \quad \textit{such that} \quad \lim_{n \to \infty} p_{mm}^{(n)} = 0.$$

Then

$$\forall k, l \quad \lim_{n \to \infty} p_{kl}^{(n)} = 0.$$

Proof. The Markov chain is irreducible $\Longrightarrow$ there exist $i, j \geq 1$ such that

$$p_{lm}^{(i)} \geq \delta_1 > 0 \quad \text{and} \quad p_{mk}^{(j)} \geq \delta_2 > 0.$$

Hence

$$p_{mm}^{(j+n+i)} \geq p_{mk}^{(j)} p_{kl}^{(n)} p_{lm}^{(i)} \geq \delta_1 \delta_2 p_{kl}^{(n)}.$$

Now $p_{mm}^{(j+n+i)} \to 0$ for $n \to \infty$ implies $\lim_{n\to\infty} p_{kl}^{(n)} = 0$. $\square$

Proposition 3.7.2. *Assume that the Markov chain $[(f_i)_{i=0}^{\infty}, p, T]$ is irreducible and that*

$$\exists m \in X \quad \textit{such that} \quad \lim_{n\to\infty} p_{mm}^{(n)} = 0.$$

Then $(f_i)_{i=0}^{\infty}$ does not have a stationary distribution.

Proof. Assume that $(s_0, s_1, \ldots)$ is a stationary distribution. Then the relation

$$(s_0, s_1, \ldots) T^n = (s_0, s_1, \ldots)$$

gives $\sum_{k=0}^{\infty} s_k p_{kl}^{(n)} = s_l$. But

$$s_l = \lim_{n\to\infty} \sum_{k=0}^{\infty} s_k p_{kl}^{(n)} = \sum_{k=0}^{\infty} s_k \left(\lim_{n\to\infty} p_{kl}^{(n)} \right) = 0 \quad \forall l$$

(notice that we may interchange summation and limit because of dominated convergence: take $f_n(k) := p_{kl}^{(n)} \Longrightarrow 0 \leq f_n(k) \leq 1$ and $\mathbb{E}_s f_n = \sum_k f_n(k) s_k$). This is a contradiction to $\sum_{l=0}^{\infty} s_l = 1$. $\square$

Definition 3.7.3. *A Markov chain is* **positive persistent** $:\Longleftrightarrow$

1. *all states are persistent,*
2. *all (persistent) states are positive.*

Remark Assume $\mathbb{P}(f_0 = k) = 1$. Then the state k is

$$\text{persistent} \iff \mathbb{P}(T_k < \infty) = 1$$

and

$$\text{positive} \iff \mu_k = \mathbb{E} T_k < \infty \iff \frac{1}{\mu_k} > 0.$$

Hence the Markov chain is positive persistent $\iff \frac{1}{\mu_k} > 0$ for all $k \in X$.

Theorem 3.7.4. *Let $(f_i)_{i=0}^{\infty}$ be a homogeneous irreducible Markov chain with state space*

$$X = \{0, \ldots, K\} \quad \textit{or} \quad X = \{0, 1, 2, \ldots, \}.$$

Then the following assertions are equivalent:

1. *$(f_i)_{i=0}^{\infty}$ has a stationary distribution $s = (s_0, s_1, \ldots)$ with $s_k > 0$ for all $k \in X$.*

2. *$(f_i)_{i=0}^{\infty}$ is positive persistent.*

Moreover, the stationary distribution is unique and given by

$$s_k = \frac{1}{\mu_k}, \quad k \in X.$$

Example 3.7.5. Let $X = \{0, 1\}$ and $T = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$. Then

- all states are persistent: $\mathbb{P}(f_2 = k | f_0 = k) = 1$, for $k = 0, 1$,

- all states are positive:

$$\begin{aligned} \mu_k &= \mathbb{E}T_k = 1\mathbb{P}(f_1 = k | f_0 = k) + 2\mathbb{P}(f_2 = k, f_1 \neq k | f_0 = k) \\ &\quad +3\mathbb{P}(f_3 = k, f_2 \neq k, f_1 \neq k | f_0 = k) + \ldots \\ &= 0 + 2 + 0 = 2. \end{aligned}$$

Hence $s_k = \frac{1}{\mu_k} = \frac{1}{2}$.

Example 3.7.6. (all states are persistent, but not positive)
Let $X = \{\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots\}$ and

$$\mathbb{P}(f_{i+1} = k + 1 | f_i = k) = \mathbb{P}(f_{i+1} = k - 1 | f_i = k) = \frac{1}{2}.$$

Then

- all states are persistent (see the Example 3.4.20 : tossing a coin)

- but there is no stationary distribution:
  Assume that $s = (\dots, s_{-2}, s_{-1}, s_0, s_1, s_2, \dots)$ is a stationary distribution. Because of

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

  one has that $\frac{1}{2}s_{k-1} + \frac{1}{2}s_{k+1} = s_k$ for all $k \in X$. Hence $k \mapsto s_k$ is a linear function. Because of $s_k \geq 0$, it follows that there is a constant c such that $s_k \equiv c$. But this is not possible because of $1 = \sum_{k=-\infty}^{\infty} s_k$. Hence for the mean recurrence time one gets $\mu_k = \infty$.

This example is called a **diffusion with infinite state space**.

Example 3.7.7. (diffusion with finite state space)
Let $X = \{0, \dots, K\}$ and

$$\begin{aligned} \mathbb{P}(f_{i+1} = l+1 | f_i = l) &= \mathbb{P}(f_{i+1} = l-1 | f_i = l) = \frac{1}{2} \quad \text{for } l = 1, \dots, K-1, \\ \mathbb{P}(f_{i+1} = K | f_i = K) &= \mathbb{P}(f_{i+1} = K-1 | f_i = K) = \frac{1}{2}, \\ \mathbb{P}(f_{i+1} = 0 | f_i = 0) &= \mathbb{P}(f_{i+1} = 1 | f_i = 0) = \frac{1}{2}. \end{aligned}$$

Then

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

which implies

$$(s_0, \dots, s_K) = \left(\frac{1}{K+1}, \dots, \frac{1}{K+1} \right),$$

and $\mu_k = K + 1$.

Proof. (of Theorem 3.7.4: $(1) \Longrightarrow (2)$)

- Assume that a state $k \in X$ is not persistent. Applying Proposition 3.4.18 gives

  $$\sum_{n=1}^{\infty} p^{(n)}(k,k) < \infty \quad \text{and therefore } \lim_{n\to\infty} p^{(n)}(k,k) = 0.$$

  Applying Proposition 3.7.2 gives that there is no stationary distribution.

- from Erdös, Feller & Pollard (1949): If $k \in X$ is null persistent with $\mu_k = \infty$ (i.e. not positive persistent) then $\lim_{n\to\infty} p^{(n)}(k,k) = 0$. Again by Proposition 3.7.2 it follows that there is no stationary distribution. Hence $(1) \Longrightarrow (2)$. □

Preparation for the proof of Theorem 3.7.4: $(2) \Longrightarrow (1)$

Lemma 3.7.8. *Assume a Markov chain $[(f_i)_{i=0}^{\infty}, s, T]$, where s is a stationary distribution. Then it holds for each n and for all $k_1, \ldots, k_n \in X$ and $m = 0, 1, 2, \ldots$*

$$\mathbb{P}(f_n = k_n, \ldots, f_0 = k_0) = \mathbb{P}(f_{n+m} = k_n, \ldots, f_m = k_0),$$

i.e. the finite-dimensional distributions are time-invariant if the initial distribution is a stationary distribution.

Proof is left as an exercise: use the 'step-by-step formula' and the Chapman-Kolmogorov equations.

Lemma 3.7.9. *Assume a homogeneous, irreducible Markov chain $(f_i)_{i=0}^{\infty}$ with $\mathbb{P}(f_0 = k) = 1$. If k is persistent, then*

$$\mathbb{P}(T_l < \infty) = 1 \quad \text{for all} \quad l \in X,$$

where

$$T_l := \inf\{n > 0 : f_n = l\}.$$

The *Proof* can be done as follows: Define

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

and

$$T_k^{i+1} := \inf\{n > T_k^i : f_n = k\}.$$

One can show that $\mathbb{P}(T_k^i < \infty) = 1$ for all $i = 1, 2, \ldots$ by induction in i. This implies $\{\omega : T_l = \infty\} \subseteq \{\omega : T_l > T_k^i\}$ for all i and hence

$$\mathbb{P}(T_l = \infty) \leq \mathbb{P}(T_l > T_k^i).$$

Since the Markov chain is irreducible we have $p := \mathbb{P}(T_l < T_k^1) > 0$. Because the Markov chain is homogeneous and the strong Markov property holds one can show

$$\mathbb{P}(T_l < T_k^{i+1} | T_l > T_k^i) = \mathbb{P}(T_l < T_k^1) = p,$$

which can be used to get

$$\mathbb{P}(T_l > T_k^i) = (1-p)^i \to 0 \quad \text{if } i \to \infty.$$

□

Lemma 3.7.10. *Let $(f_i)_{i=0}^\infty$ be a homogeneous, positive persistent Markov chain starting in k_0. Define*

$$T_{k_0} := \inf\{n \geq 1 : f_n = k_0\}$$

and

$$v_k := \mathbb{E} \sum_{n=0}^{T_{k_0}-1} 1\!\!1_{\{f_n = k\}},$$

which is the expected number of visits of k if we are starting in k_0 before we reach k_0 again. Then

$$v_k = \sum_{l \in X} v_l p_{lk}.$$

Proof. It holds

$$\begin{aligned} v_k &= \mathbb{E} \sum_{n=0}^{T_{k_0}-1} 1\!\!1_{\{f_n = k\}} \\ &= \mathbb{E} \sum_{n=0}^{\infty} 1\!\!1_{\{f_n = k, n < T_{k_0}\}} \\ &= \sum_{n=0}^{\infty} \mathbb{P}(f_n = k, n < T_{k_0}). \end{aligned}$$

We consider first the case $k \neq k_0$. Then

$$\mathbb{P}(f_0 = k_0) = 1 \quad \text{implies} \quad \mathbb{P}(f_0 = k, 0 \leq T_{k_0}) = 0,$$

and so we can continue with

$$\begin{aligned}
v_k &= \sum_{n=0}^{\infty} \mathbb{P}(f_n = k, n < T_{k_0}) \\
&= \sum_{n=1}^{\infty} \mathbb{P}(f_n = k, n \leq T_{k_0}) \\
&= \sum_{l \in X} \sum_{n=1}^{\infty} \mathbb{P}(f_n = k, f_{n-1} = l, n-1 < T_{k_0}) \\
&= \sum_{l \in X} \sum_{n=1}^{\infty} \mathbb{P}(f_n = k | f_{n-1} = l, n-1 < T_{k_0}) \mathbb{P}(f_{n-1} = l, n-1 < T_{k_0})
\end{aligned}$$

It holds

$$\begin{aligned}
\mathbb{P}(f_n = k | f_{n-1} = l, n-1 < T_{k_0}) &= \mathbb{P}(f_n = k | f_{n-1} = l, f_{n-2} \neq k_0, \ldots, f_1 \neq k_0) \\
&= \mathbb{P}(f_n = k | f_{n-1} = l) = p_{lk}
\end{aligned}$$

because of the Markov property. Hence it follows

$$\begin{aligned}
v_k &= \sum_{l \in X} p_{lk} \sum_{n=1}^{\infty} \mathbb{P}(f_{n-1} = l, n-1 < T_{k_0}) \\
&= \sum_{l \in X} p_{lk} v_l.
\end{aligned}$$

In the case $k = k_0$ we proceed as follows

$$\begin{aligned}
v_{k_0} &= \sum_{n=0}^{\infty} \mathbb{P}(f_n = k_0, n < T_{k_0}) \\
&= \mathbb{P}(f_0 = k_0, 0 < T_{k_0}) + 0 + 0 \cdots \\
&= 1.
\end{aligned}$$

Because k_0 is persistent we have

$$
\begin{aligned}
1 = \mathbb{P}(T_{k_0} < \infty) &= \sum_{n=1}^{\infty} \mathbb{P}\left(T_{k_0} = n\right) \\
&= \sum_{n=1}^{\infty} \mathbb{P}\left(T_{k_0} > n-1, f_n = k_0\right) \\
&= \sum_{n=1}^{\infty} \sum_{l \in X} \mathbb{P}(T_{k_0} > n-1, f_n = k_0, f_{n-1} = l) \\
&= \ldots \text{ as in case } k \neq k_0 \ldots \\
&= \sum_{l \in X} p_{lk} v_l. \qquad \square
\end{aligned}
$$

Proof of Theorem 3.7.4: $(2) \Longrightarrow (1)$
What we have to show is

$$\frac{1}{\mu_k} > 0 \quad \forall k \Longrightarrow \exists s = (s_0, s_1, \ldots) \quad \text{with} \quad s_k > 0.$$

We start with (see the proof of Proposition 3.4.18)

$$
\begin{aligned}
\mu_{k_0} = \mathbb{E} T_{k_0} &= \sum_{n=0}^{\infty} \mathbb{P}(T_{k_0} > n) \\
&= \sum_{n=0}^{\infty} \sum_{k \in X} \mathbb{P}(T_{k_0} > n, f_n = k) \\
&= \sum_{k \in X} v_k.
\end{aligned}
$$

Set

$$s_k := \frac{v_k}{\mu_{k_0}} \quad \text{for} \quad k \in X. \tag{3.5}$$

Then

$$\sum_{k \in X} s_k = 1.$$

By Lemma 3.7.10 it holds $v_k = \sum_{l \in X} p_{lk} v_l$, hence

$$s_k = \sum_{l \in X} p_{lk} s_l$$

which implies $\left(\frac{v_0}{\mu_{k_0}}, \frac{v_1}{\mu_{k_0}}, \dots\right)$ is a stationary distribution.

Next we show that $s_k = \frac{1}{\mu_k}$, for all $k \in X$. This implies $s_k > 0, \quad \forall k \in X$ ($\mu_k < \infty$ because the Markov chain was assumed to be positive persistent) and the uniqueness of the stationary distribution. So let us assume $b = (b_0, b_1, \dots)$ is a stationary distribution. We define the fist passage time

$$T_l := \inf\{n \geq 1 : f_n = l\}.$$

We take b as initial distribution and let $k \neq l$. Then, by Lemma 3.7.8, we get

$$\begin{aligned}
\mathbb{P}(f_n = k, T_l \geq n+1) &= \mathbb{P}(f_n = k, f_{n-1} \neq l, \dots, f_1 \neq l) \\
&= \mathbb{P}(\underbrace{f_{n-1} = k, f_{n-2} \neq l, \dots, f_1 \neq l}_{\{f_{n-1}=k, T_l \geq n\}}, f_0 \neq l)) \\
&= \mathbb{P}(f_{n-1} = k, T_l \geq n) - \mathbb{P}(f_{n-1} = k, T_l \geq n, f_0 = l) \\
&= \mathbb{P}(f_{n-1} = k, T_l \geq n) - \mathbb{P}(f_{n-1} = k, T_l \geq n | f_0 = l) b_l.
\end{aligned}$$

Summation over n yields

$$\begin{aligned}
&\sum_{n=1}^{N} \mathbb{P}(f_{n-1} = k, T_l \geq n | f_0 = l) b_l \\
&\qquad = \sum_{n=1}^{N} \{\mathbb{P}(f_{n-1} = k, T_l \geq n) - \mathbb{P}(f_n = k, T_l \geq n+1)\} \\
&\qquad = \mathbb{P}(f_0 = k, T_l \geq 1) - \mathbb{P}(f_N = k, T_l \geq N+1) \\
&\qquad = b_k - \mathbb{P}(f_N = k, T_l \geq N+1).
\end{aligned}$$

Summation over $k \in X$ and taking the limit $N \to \infty$ leads to

$$\begin{aligned}
\sum_{k \in X} \sum_{n=1}^{\infty} \mathbb{P}(f_{n-1} = k, T_l \geq n | f_0 = l) b_l &= \sum_{k \in X} b_k - \lim_{N \to \infty} \sum_{k \in X} \mathbb{P}(f_N = k, T_l \geq N+1) \\
&= 1 - \lim_{N \to \infty} \mathbb{P}(T_l \geq N+1).
\end{aligned}$$

For the left hand side of the equation we get

$$\begin{aligned}
\sum_{k \in X} \sum_{n=1}^{\infty} \mathbb{P}(f_{n-1} = k, T_l \geq n | f_0 = l) b_l &= \sum_{n=1}^{\infty} \mathbb{P}(T_l \geq n | f_0 = l) b_l \\
&= (\mathbb{E} T_l) b_l = \mu_l \, b_l.
\end{aligned}$$

If we can show that $\lim_{N\to\infty} \mathbb{P}(T_l \geq N+1) = 0$ this would imply

$$\mu_l\, b_l = 1, \quad \forall l \in X$$

and complete the proof. It holds by Lemma 3.7.9

$$1 = \mathbb{P}(T_l < \infty | f_0 = k) = \sum_{n=1}^{\infty} \mathbb{P}(T_l = n | f_0 = k).$$

Hence

$$\mathbb{P}(T_l \geq N+1 | f_0 = k) = \sum_{n=N+1}^{\infty} \mathbb{P}(T_l = n | f_0 = k) \to 0 \quad \text{as} \quad N \to \infty.$$

By the Theorem about dominated convergence we may interchange summation and limit and therefore

$$\begin{aligned}
\lim_{N\to\infty} \mathbb{P}(T_l \geq N+1) &= \lim_{N\to\infty} \sum_{k\in X} \mathbb{P}(T_l \geq N+1 | f_0 = k) b_k \\
&= \sum_{k\in X} \lim_{N\to\infty} \mathbb{P}(T_l \geq N+1 | f_0 = k) b_k = 0.
\end{aligned}$$

□

Proposition 3.7.11. *If X is finite, then*

1. *there is at least one persistent state,*
2. *all persistent states are positive*

Example X is infinite and all states are transient:
Choose $X = \{0, 1, 2, \dots\}$ and $p_{k,k+1} = 1$.

Proof. (of Proposition 3.7.11)

1. Assume there is no persistent state and X is irreducible. Then $\sum_{n=1}^{\infty} p^{(n)}(k,k) < \infty$ and hence

$$\lim_{n\to\infty} p^{(n)}(k,k) = 0 \quad \forall k \in X.$$

By Lemma 3.7.1 we have that $p_{kl}^{(n)} \to 0$ for all $k, l \in X$ which yields

$$1 = \lim_{n\to\infty} \sum_{l\in X} p_{kl}^{(n)} = \sum_{l\in X} \lim_{n\to\infty} p_{kl}^{(n)} = 0$$

which is a contradiction, so at least one state has to be persistent. If X is not irreducible, then one considers the disjoint and irreducible subclasses $X_1, X_2, \ldots, X_L$ with $X = \bigcup_{M=1}^{L} X_m$. Since at least one of them has to be closed, let us assume it is X_L, we can find a k with $\mathbb{P}(f_n \in X_L | f_0 = k) = 1$ for all n. But again

$$1 = \lim_{n\to\infty} \mathbb{P}(f_n \in X_L | f_0 = k) = \sum_{l\in X_L} \lim_{n\to\infty} p_{kl}^{(n)} = 0$$

is a contradiction, so a persistent state exists.

2. Because of Proposition 3.4.23 (decomposition of the state space) we can assume that X is irreducible. Assume $k_0 \in X$ is persistent and $\mathbb{P}(f_0 = k_0) = 1$. By Lemma 3.7.10,

$$v_k = \sum_{l\in X} v_l p_{lk} \quad \Longrightarrow \quad v_k = \sum_{l\in X} v_l p_{lk}^{(n)} \quad \forall k \in X,$$

and, since the Markov chain is starting in k_0 we have $1 = v_{k_0} \geq v_l p_{lk_0}^{(n)}$. Irreducibility implies that there exists an n with $p_{lk_0}^{(n)} > 0$. Hence $v_l < \infty$ for all $l \in X$. From the proof of Theorem 3.7.4 we know that

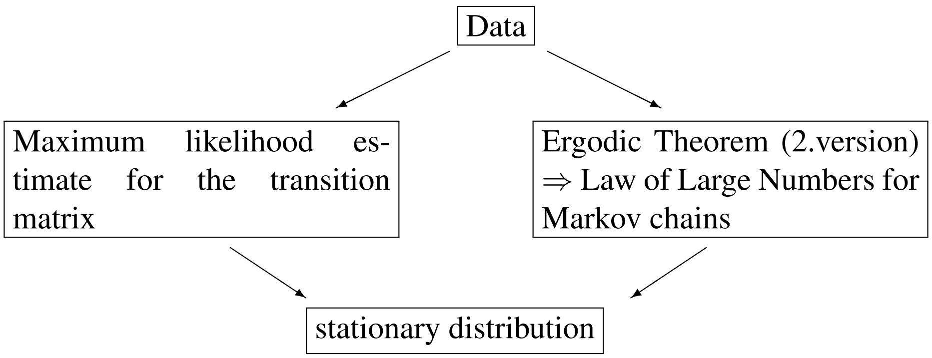
$$\mu_{k_0} = \sum_{l\in X} v_l.$$

From this it follows $\mu_{k_0} < \infty$ since X is finite.

□

Corollary 3.7.12. *Let $(f_i)_{i=0}^{\infty}$ be a homogeneous irreducible Markov chain with state space $X = \{0, \ldots, K\}$. Then there is a unique stationary distribution.*

Problem: For given data one would like to get estimates for the stationary distribution:



Theorem 3.7.13. *(Ergodic theorem, second version) Let $(f_i)_{i=0}^{\infty}$ be a homogeneous, irreducible Markov chain which is positive persistent. Let $s = (s_0, s_1, \dots)$ be the stationary distribution and*

$$F : X \to \mathbb{R} \quad \textit{with} \quad \mathbb{E}_s|F| := \sum_{k \in X} s_k |F(k)| < \infty.$$

Then

$$\mathbb{P}\left(\left\{\omega : \frac{1}{n}\sum_{j=1}^{n} F(f_j(\omega)) \overset{n\to\infty}{\to} \mathbb{E}_s F\right\}\right) = 1. \tag{3.6}$$

Remarks

1. The convergence in (3.6) is called 'almost surely'. We will only prove a weaker assertion - the convergence in probability- i.e. we will prove:
$$\forall \varepsilon > 0 \;\; \mathbb{P}\left(\left\{\omega : \left|\frac{1}{n}\sum_{j=1}^{n} F(f_j(\omega)) - \mathbb{E}_s F\right| > \varepsilon\right\}\right) \overset{n\to\infty}{\to} 0.$$
2. the assumption 'irreducible' is missing in [2] (p.49), but it is needed. Example: $X = \{0, 1\}$ and $T = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$ Then
$$\frac{1}{n}\sum_{j=1}^{n} F(f_j(\omega)) = \begin{cases} F(0) & \forall \omega \text{ with } f_0(\omega) = 0, \\ F(1) & \forall \omega \text{ with } f_0(\omega) = 1. \end{cases}$$

If $\mathbb{P}(f_0 = 0) > 0$ and $\mathbb{P}(f_0 = 1) > 0$ then (3.6) is not possible.

Application of Theorem 3.7.13
From our data we can get the stationary distribution as follows: Choose $k_0 \in X$ and put $F(k) := \mathbb{I}_{\{k_0\}}(k)$. Then

$$\mathbb{P}\left(\frac{1}{n}\sum_{j=1}^{n} \mathbb{I}_{\{k_0\}}(f_j) \stackrel{n\to\infty}{\to} s_{k_0}\right) = 1,$$

because

$$\mathbb{E}_s F = \sum_{k\in X} \mathbb{I}_{\{k_0\}}(k) s_k = s_{k_0}.$$

Example (Theorem 3.7.13, independent case)
Assume $f_0, f_1, \ldots \Omega \to X$ i.i.d. and

$$\mathbb{P}(f_i = k) = s_k > 0, \quad \text{for } k \in \{0, \ldots, K\}, \quad i = 0, 1, \ldots$$

Then $(f_i)_{i=0}^{\infty}$ is a positive persistent Markov chain with

$$T = \begin{pmatrix} s_0 & s_1 & \ldots & s_K \\ \vdots & \vdots & & \vdots \\ s_0 & s_1 & \ldots & s_K \end{pmatrix}.$$

The stationary distribution is $s = (s_0, s_1, \ldots, s_K)$. By Theorem 3.7.13 for all function $F : X \to \mathbb{R}$ such that $\mathbb{E}|F(f_1)| < \infty$ it holds.

$$\mathbb{P}\left(\frac{1}{n}\sum_{j=1}^{n} F(f_j) \stackrel{n\to\infty}{\to} \mathbb{E}F(f_1)\right) = 1,$$

This is the **Strong Law of Large Numbers** for i.i.d. random variables.

Corollary 3.7.14. *Let $(f_i)_{i=0}^{\infty}$ be a homogeneous irreducible Markov chain and let $k_0 \in X$ be persistent. Define the stopping times*

$$\begin{aligned} T_0 &\equiv 0, \\ T_1 &:= \inf\{n > T_0 : f_n = k_0\}, \\ T_{i+1} &:= \inf\{n > T_i : f_n = k_0\}. \end{aligned}$$

Let

$$Z_i := \sum_{j=T_i+1}^{T_{(i+1)}} F(f_j(\omega)), \quad \textit{for} \quad i = 0, 1, \dots$$

where $F : X \to \mathbb{R}$. *Then* $Z_0, Z_1, \dots$ *are independent.*

Proof. (sketch)
Letting $0 \le r < s < \infty$, we will only prove that Z_r and Z_s are independent. Let A, B be Borel-sets. One has to show that

$$\mathbb{P}(Z_s \in B, Z_r \in A) = \mathbb{P}(Z_s \in B)\mathbb{P}(Z_r \in A).$$

We may assume that $\mathbb{P}(Z_r \in A) > 0$.

$$\mathbb{P}(Z_s \in B, Z_r \in A) = \mathbb{P}(Z_s \in B | Z_r \in A)\mathbb{P}(Z_r \in A).$$

Hence we have to show that

$$\mathbb{P}(Z_s \in B | Z_r \in A) = \mathbb{P}(Z_s \in B).$$

It holds

$$\mathbb{P}(Z_s \in B | Z_r \in A) = \mathbb{P}(Z_s \in B | f_{T_{r+1}} = k_0, Z_r \in A),$$

where $\{Z_s \in B\} = \{\sum_{i=T_s+1}^{T_{(s+1)}} F(f_i(\omega)) \in B\}$ can be written as a disjoint countable union of sets H_l, i.e. $\{Z_s \in B\} = \bigcup_{l=1}^{\infty} H_l$ with

$$H_l := \{f_{T_{(r+1)}+1} = y_1^{(l)}, f_{T_{(r+1)}+2} = y_2^{(l)}, \dots\}$$

for certain $y_i^{(l)} \in X$. So by an approximation argument it is sufficient to show that

$$\begin{aligned} &\mathbb{P}(f_{T_{(r+1)}+m} = y_m, \dots, f_{T_{(r+1)}+1} = y_1 | f_{T_{(r+1)}} = k_0, Z_r \in A) \\ &\quad = \mathbb{P}(f_{T_{(r+1)}+m} = y_m, \dots, f_{T_{(r+1)}+1} = y_1 | f_{T_{(r+1)}} = k_0). \end{aligned} \tag{3.7}$$

It holds

$$\{Z_r \in A\} \cap \{T_{r+1} = i\} \in \mathcal{F}_i := \sigma\{f_0, \dots, f_i\}, \quad \forall i = 0, 1, 2, \dots$$

So we can apply Proposition 3.6.1 (the strong Markov property) to get (3.7).

$\square$

Limiting occupation probability

Definition 3.7.15. *Let $(f_i)_{i=0}^{\infty}$ be a homogeneous Markov chain and $k \in X$. Then*

$$N_k(n) := \sum_{i=1}^{n} \mathbb{1}_{\{f_i=k\}}(\omega)$$

Interpretation $N_k(n)$ is the time the Markov chain spent in k till time n. $N_k(n)$ is a random variable.
Example If k is absorbing and $f_1(\omega) = k$, then $N_k(n) = n$ for all $n = 1, 2, \ldots$

Proposition 3.7.16. *Let $(f_i)_{i=0}^{\infty}$ be a homogeneous irreducible Markov chain and $k \in X$ be positive persistent. Then*

$$\lim_{n\to\infty} \frac{N_k(n)}{n} = \frac{1}{\mu_k} \quad a.s.$$

where the limit is called **limiting occupation probability.**

Proof: Exercise.

Example: Diffusion with boundaries (see Example 3.7.7)

Let $X = \{-K, -K+1, \ldots, 0, \ldots, K-1, K\}$. Then one has

$$\lim_{n\to\infty} \frac{N_l(n)}{n} = \frac{1}{2K+1}, \quad \forall \;\; l \in X.$$

Proof. We know that the stationary distribution is

$$(s_{-K}, \ldots, s_0, \ldots, s_K) = (\frac{1}{2K+1}, \ldots, \frac{1}{2K+1}).$$

By Theorem 3.7.4 it holds $\frac{1}{\mu_l} = \frac{1}{2K+1}$. Consequently, by Proposition 3.7.16,

$$\lim_{n\to\infty} \frac{N_k(n)}{n} = \frac{1}{2K+1} \quad a.s.$$

$\square$

Proof of Theorem 3.7.13

Fix $k_0 \in X$. Define the stopping times

$$\begin{aligned} T_0 &:= 0 \\ T_1 &:= \inf\{n > 0 : f_n = k_0\}, \\ T_2 &:= \inf\{n > T_1 : f_n = k_0\}, \dots \end{aligned}$$

We define

$$Z_i := \sum_{j=T_i+1}^{T_{(i+1)}} F(f_j(\omega))$$

and notice that

$$T_{N_{k_0}(n)} \leq n < T_{(N_{k_0}(n)+1)}.$$

So we can write

$$\frac{\sum_{j=1}^{n} F(f_j)}{n} = \frac{Z_0}{n} + \frac{\sum_{i=1}^{N_{k_0}(n)} Z_i}{n} - \frac{\sum_{j=n+1}^{T_{(N_{k_0}(n)+1)}} F(f_j)}{n}.$$

If we can show that

$$\frac{Z_0}{n} \to 0, \quad a.s. \tag{3.8}$$

$$\frac{\sum_{j=1}^{N_{k_0}(n)} Z_i}{n} \to \frac{1}{\mu_{k_0}} \mathbb{E} Z_1, \quad a.s. \tag{3.9}$$

and

$$\frac{\sum_{j=n+1}^{T_{(N_{k_0}(n)+1)}} F(f_j)}{n} \to 0 \quad \text{in probability} \tag{3.10}$$

we are done. This is true because $\frac{1}{\mu_{k_0}} \mathbb{E} Z_1 = \mathbb{E}_s F$, which one can see as follows.

$$\begin{aligned} \mathbb{E}|Z_1| &= \mathbb{E}|\sum_{i=T_1+1}^{T_2} F(f_i)| \\ &= \mathbb{E}\sum_{k \in X} |\sum_{i=T_1+1}^{T_2} F(f_i)| \mathbb{I}_{\{f_i = k\}} \\ &= \mathbb{E}\sum_{k \in X} |F(k)| \sum_{i=T_1+1}^{T_2} \mathbb{I}_{\{f_i = k\}} \\ &= \sum_{k \in X} v_k |F(k)| \end{aligned}$$

since $v_k = \mathbb{E}\sum_{i=T_1+1}^{T_2} \mathbb{I}_{\{f_i=k\}}$. From the proof of Theorem 3.7.4 (formula (3.5)) we know that $v_k = s_k\mu_{k_0}$. So we continue

$$\begin{aligned}\sum_{k\in X} v_k|F(k)| &= \sum_{k\in X} s_k\mu_{k_0}|F(k)| \\ &= \mu_{k_0}\mathbb{E}_s|F| < \infty\end{aligned}$$

by assumption.This implies that Z_1 is integrable and a similar computation like the one above gives $\mathbb{E}Z_1 = \mu_{k_0}\mathbb{E}_s F$. Consequently, the convergences (3.8), (3.9) and (3.10) will imply the assertion of the Theorem. The convergence (3.8) holds because Z_0 does not depend on n, so $\frac{Z_0(\omega)}{n} \to 0$ as $n \to \infty$.
To get the convergence (3.9) we write

$$\frac{1}{n}\sum_{j=1}^{N_{k_0}(n)} Z_j = \left(\frac{1}{N_{k_0}(n)}\sum_{j=1}^{N_{k_0}(n)} Z_j\right)\left(\frac{N_{k_0}(n)}{n}\right).$$

The random variables $(Z_j)_{j=1}^{\infty}$ are independent and identically distributed according to Corollary 3.7.14, hence by Theorem 3.5.5 , the Strong Law of Large Numbers, we have

$$\frac{1}{N_{k_0}(n)}\sum_{j=1}^{N_{k_0}(n)} Z_j \to \mathbb{E}Z_1, \quad a.s.$$

On the other hand, Proposition 3.7.16 implies

$$\frac{N_{k_0}(n)}{n} \to \frac{1}{\mu_{k_0}}, \quad a.s.$$

The convergence (3.10) will not be shown here. □

Remark The proof of Lemma 3.7.9 is now easier: Assume we are given an irreducible, homogeneous Markov chain, such that $\mathbb{P}(f_0 = k) = 1$, where k is persistent and we want to show that this implies

$$\mathbb{P}(T_l < \infty) = 1.$$

By defining $T_k^0 := 0$ and $T_k^{i+1} := \inf\{n > T_k^i : f_n = k\}$ and setting

$$Z_i := \sum_{j=T_{k+1}^i}^{T_k^{i+1}} \mathbb{I}_{\{l\}}(f_j),$$

we get, because the (Z_i) are i.i.d.

$$\begin{aligned} \mathbb{P}(T_l = \infty) &= \mathbb{P}(\bigcap_{i=0}^{\infty}\{Z_i = 0\}) \\ &= \Pi_{i=0}^{\infty}\mathbb{P}(Z_i = 0) = \left\{ \begin{array}{l} 0 \\ 1 \end{array} \right. \end{aligned} \tag{3.11}$$

But $k \leftrightarrow l$ implies $\exists m_0 \geq 1$ such that $\mathbb{P}(f_{m_0} = l) > 0$. Consequently,

$$\mathbb{P}(T_l < \infty) \geq \mathbb{P}(T_l \leq m_0) \geq \mathbb{P}(f_{m_0} = l) > 0,$$

which yields for the probability of the complement set $\mathbb{P}(T_l = \infty) < 1$. In view of (3.11) this implies $\mathbb{P}(T_l = \infty) = 0$. $\square$

Chapter 4

Markov chain Monte Carlo methods (MCMC methods)

4.1 The classical Monte Carlo method

Assume a probability space $(X, \mathcal{G}, \mu)$, for example $X \subseteq \mathbb{R}^n$, and a Borel measurable function $F : X \to \mathbb{R}$. If one wants to compute

$$\mathbb{E}_\mu F = \int_X F(x) d\mu(x)$$

(and if one knows that $\mathbb{E}_\mu |F| < \infty$) but a direct computation is not possible then one can use the SLLN (strong law of large numbers): One generates independent identically distributed ($\mathbb{P}_{\xi_i} = \mu$) random variables $\xi_1, \xi_2, \ldots \quad \xi_i : \Omega \to X$. Then

$$\frac{1}{n} \sum_{i=1}^n F(\xi_i(\omega)) \overset{n\to\infty}{\longrightarrow} \mathbb{E}_\mu F \quad a.s.$$

But if, for example, $X = \mathbb{R}^n$ and n is large, then ξ depends on many coordinates and it may be difficult to generate independent copies of ξ. Here one uses better 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 μ, and

$$\frac{1}{n}\sum_{i=1}^{n} F(f_i(\omega)) \overset{n\to\infty}{\longrightarrow} \mathbb{E}_\mu F \quad a.s.$$

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

$$\frac{1}{n}\sum_{i=1}^{n} Z_i \overset{n\to\infty}{\longrightarrow} \mathbb{E}Z_1 \quad a.s.$$

It holds, since X_1 and Y_1 are independent and uniformly on $[0,1]$ distributed

$$\begin{aligned}
\mathbb{E}Z_1 &= \mathbb{E}\mathbb{1}_{\{F(X_1)>Y_1\}} \\
&= \int_0^1 \int_0^1 \mathbb{1}_{\{F(x)>y\}} dydx \\
&= \int_0^1 \int_0^{F(x)} dydx \\
&= \int_0^1 F(x)dx.
\end{aligned}$$

4.4 The Gibbs sampler: the first example of a MCMC

$\approx$ 1970 by Pentti Suomela
Statistics, 1984 by Geman & Geman
In Statistical Physics the Gibbs sampler is called 'heat bath method'.

The problem:
Given random variables $X_1, \ldots, X_m : \Omega \to \mathbb{R}$ and a function $F : \mathbb{R}^m \to \mathbb{R}$. How to compute

$$\mathbb{E}F(X_1, \ldots, X_m)$$

numerically ?

Assumptions on the state space X

(A1) $S_i = \{X_i(\omega) : \mathbb{P}(\omega) > 0\}$ is finite for $i = 1, \ldots, m$

(A2) $S_1 \times \cdots \times S_m = \{(X_1(\omega), \ldots, X_m(\omega)) : \mathbb{P}(\omega) > 0\}$

Example 4.4.1. Let $m = 2$, X_1, X_2 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

$$\{(X_1(\omega), X_2(\omega)) : \mathbb{P}(\omega) > 0\} = \{(-1,-1),\ (-1,1),\ (1,-1),\ (1,1)\} = S_1 \times S_2.$$

Hence (A1) and (A2) are satisfied.

Example 4.4.2. Let $m = 2$, $X_1 = X_2$ and

$$\mathbb{P}(X_i = 1) = \mathbb{P}(X_i = -1) = \frac{1}{2}.$$

This implies $S_1 = S_2 = \{-1, 1\}$ and

$$\{(X_1(\omega), X_2(\omega)) : \mathbb{P}(\omega) > 0\} = \{(-1,-1),\ (1,1)\} \neq S_1 \times S_2.$$

Hence (A2) is not satisfied.

4.4.1 Description of the Gibbs sampler

$\boxed{\text{Step } 0}$ 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 one assumes that there is some $\omega \in \Omega$ such that

$$\underbrace{(\xi_1^0, \ldots, \xi_m^0)}_{\text{numbers in the computer}} = (X_1^0(\omega), \ldots, X_m^0(\omega))$$

$\boxed{\text{Step } n \quad \to \text{Step } n+1}$ Assume we have the random vector $(X_1^n, \ldots, X_m^n)$. We produce the random vector $(X_1^{n+1}, \ldots, X_m^{n+1})$ in the following way:
If $(\xi_1^n, \ldots, \xi_m^n) = (X_1^n(\omega), \ldots, X_m^n(\omega))$ is the realization of step n, then

$$\begin{aligned}
X_1^{n+1} &= \text{independent copy of } (X_1, \ldots, X_m) \text{ under the condition that} \\
& \quad X_2 = \xi_2^n, \ldots, X_m = \xi_m^n \\
\Longrightarrow & \quad \xi_1^{n+1} = X_1^{n+1}(\omega).
\end{aligned}$$

$$\begin{aligned}
X_2^{n+1} &= \text{independent copy of } (X_1, \ldots, X_m) \text{ under the condition that} \\
& \quad X_1 = \xi_1^{n+1}, X_3 = \xi_3^n, \ldots, X_m = \xi_m^n \\
\Longrightarrow & \quad \xi_2^{n+1} = X_2^{n+1}(\omega).
\end{aligned}$$

$$\vdots$$

$$\begin{aligned}
X_m^{n+1} &= \text{independent copy of } (X_1, \ldots, X_m) \text{ under the condition that} \\
& \quad X_1 = \xi_1^{n+1}, \ldots, X_{m-1} = \xi_{m-1}^{n+1} \\
\Longrightarrow & \quad \xi_m^{n+1} = X_m^{n+1}(\omega).
\end{aligned}$$

Proposition 4.4.3. *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 $S_1 \times \cdots \times S_m$.*

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

$$s_k = \mathbb{P}((X_1, \ldots, X_m) = k).$$

Proof.

1. follows by construction.

2. Because of (A2) the transition matrix has only positive values. By Theorem 3.4.26 (the first version of the ergodic Theorem) it follows that the Markov chain is ergodic. We check the stationary distribution for $m = 2$. It holds for the transition matrix:

$$p_{\binom{\xi_1}{\xi_2},\binom{\eta_1}{\eta_2}} = \mathbb{P}(X_1 = \eta_1 | X_2 = \xi_2)\mathbb{P}(X_2 = \eta_2 | X_1 = \eta_1)$$

We have the desired stationary distribution if

$$\sum_{\xi_1,\xi_2} \mathbb{P}(X_1 = \xi_1, X_2 = \xi_2) p_{\binom{\xi_1}{\xi_2},\binom{\eta_1}{\eta_2}} = \mathbb{P}(X_1 = \eta_1, X_2 = \eta_2).$$

It holds

$$\begin{aligned}
& \sum_{\xi_1,\xi_2} \mathbb{P}(X_1 = \xi_1, X_2 = \xi_2) p_{\binom{\xi_1}{\xi_2},\binom{\eta_1}{\eta_2}} \\
= & \sum_{\xi_1,\xi_2} \mathbb{P}(X_1 = \xi_1, X_2 = \xi_2)\mathbb{P}(X_1 = \eta_1 | X_2 = \xi_2)\mathbb{P}(X_2 = \eta_2 | X_1 = \eta_1) \\
= & \sum_{\xi_2} \left(\sum_{\xi_1} \mathbb{P}(X_1 = \xi_1, X_2 = \xi_2) \right) \mathbb{P}(X_1 = \eta_1 | X_2 = \xi_2) \\
& \qquad \times \mathbb{P}(X_2 = \eta_2 | X_1 = \eta_1) \\
= & \sum_{\xi_2} \mathbb{P}(X_2 = \xi_2)\mathbb{P}(X_1 = \eta_1 | X_2 = \xi_2)\mathbb{P}(X_2 = \eta_2 | X_1 = \eta_1) \\
= & \mathbb{P}(X_1 = \eta_1)\mathbb{P}(X_2 = \eta_2 | X_1 = \eta_1) \\
= & \mathbb{P}(X_2 = \eta_2, X_1 = \eta_1)
\end{aligned}$$

□

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 $X = \{0, \dots, K\}$ and transition matrix $T = (p_{ij})_{i,j=0}^{K}$. Assume that*

$$\varepsilon := \min_{i,j} p_{ij} > 0.$$

If $(s_0, \dots, s_K)$ is the unique stationary distribution (see Theorem 3.4.26) and $T^n = (p_{kl}^{(n)})_{k,l=0}^{K}$ then

$$|p_{k,l}^{(n+1)} - s_l| \leq (1-\varepsilon)^n \sup_{i,j}\ p_{ij}.$$

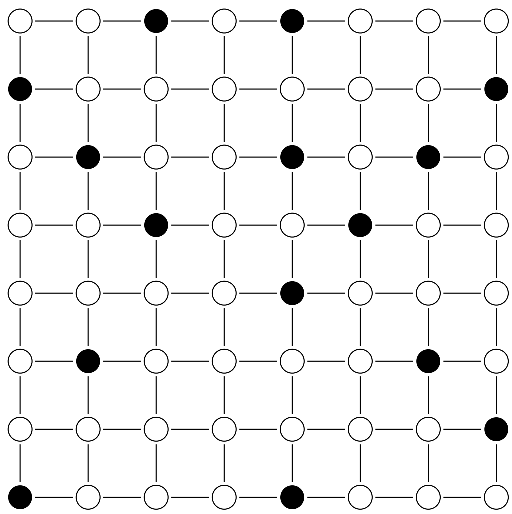
Proof. This follows immediately from the estimate

$$M_l^{(n+1)} - m_l^{(n+1)} \leq (1-\varepsilon)^n (M_l^{(1)} - m_l^{(1)})$$

in the proof of Theorem 3.4.26. □

4.6 The hard-core model as an example for the 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.



a feasible configuration

As state space we take $X = \{$ feasible configurations $\}$ This means, for example, if we have a 10×10 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. $X_{n+1}(v) = \begin{cases} \text{'black'} & \text{if coin=heads} \\ & \text{and all direct neighbors} \\ & \text{of } v \text{ are 'white' in } X_n, \\ \text{'white'} & \text{otherwise} \end{cases}$
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.

4.7 The Metropolis algorithm

is historically the first Monte Carlo method (1953), it was developed by Metropolis, Rosenbluth & Rosenbluth, Teller & Teller
One is given $X = \{0, \dots, K\}$ and a distribution $s = (s_0, \dots, s_K)$ with $s_l > 0$ for $l = 0, \dots, K$.

How to produce a Markov chain having the stationary distribution s?

We take an auxiliary symmetric transition matrix $\widetilde{T} = (q_{kl})_{k,l=0}^{K}$ (symmetric means $q_{kl} = q_{lk}$) such that $q_{kl} > 0$.

$\boxed{\text{Step } n \quad \to \text{Step } n+1}$

Assume we had generated in the last step $\xi_n \in X$. Now generate η from the distribution $(q_{\xi_n l})_{l=0}^{K}$ (or in other words we use $\mathbb{P}(\eta = l) = q_{\xi_n l}, \quad l = 0, \dots, K$). Calculate

$$r := \frac{s_\eta}{s_{\xi_n}}.$$

Put

$$\xi_{n+1} := \begin{cases} \eta & \text{if } r \geq 1 \\ \eta & \text{with probability } r \quad (0 < r < 1) \\ \xi_n & \text{with probability } 1 - r \quad (0 < r < 1) \end{cases}$$

Let $(f_i)_{i=0}^{\infty}$ be the corresponding sequence of random variables.

Theorem 4.7.1. *The sequence $(f_i)_{i=0}^{\infty}$ is an ergodic Markov chain with stationary distribution $s = (s_0, \dots, s_K)$.*

Proof.

1. Let us assume that $0 < s_0 \leq s_1 \leq \cdots \leq s_K \leq 1$ and let us denote the transition matrix of $(f_i)_{i=0}^{\infty}$ by $T = (p_{kl})_{k,l=0}^{K}$. Then

$$T = \begin{pmatrix} q_{00} & q_{01} & q_{02} & q_{03} & \cdots & q_{0K} \\ \frac{s_0}{s_1} q_{10} & \begin{matrix} q_{11} \\ +(1 - \frac{s_0}{s_1}) q_{10} \end{matrix} & q_{12} & q_{13} & \cdots & q_{1K} \\ \frac{s_0}{s_2} q_{20} & \frac{s_1}{s_2} q_{21} & \begin{matrix} q_{22} \\ +(1 - \frac{s_0}{s_2}) q_{20} \\ +(1 - \frac{s_1}{s_2}) q_{21} \end{matrix} & q_{23} & \cdots & q_{2K} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Hence

$$p_{kl} = \begin{cases} q_{kl} & \text{if } k < l, \\ \frac{s_l}{s_K} q_{kl} & \text{if } k > l. \end{cases}$$

In particular, $p_{kl} > 0$ for $k, l \in X$, such that $(f_i)_{i=0}^{\infty}$ is an ergodic Markov chain (compare Theorem 3.4.26).

2. Now we check whether $s = (s_0, \ldots, s_K)$ is a stationary distribution. If so, the equation

$$\begin{aligned} s_0 q_{0l} + \cdots + s_{l-1} q_{l-1,l} + s_l p_{ll} & \\ + s_{l+1} \left[\frac{s_l}{s_{l+1}} q_{l+1,l} \right] + \cdots + s_K \left[\frac{s_l}{s_K} q_{K,l} \right] &= s_l \end{aligned} \tag{4.1}$$

where $p_{ll} = q_{ll} + (1 - \frac{s_0}{s_l}) q_{l0} + \cdots + (1 - \frac{s_{l-1}}{s_l}) q_{l,l-1}$, must be true for each $l = 0, \ldots, K$. Calculating yields that the left hand side of (4.1) is equal to

$$\begin{aligned} s_0 q_{0l} + & \cdots + s_{l-1} q_{l-1,l} \\ & + s_l \left[q_{ll} + (1 - \frac{s_0}{s_l}) q_{l0} + \cdots + (1 - \frac{s_{l-1}}{s_l}) q_{l,l-1} \right] \\ & + s_l q_{l+1,l} + \cdots + s_l q_{K,l} \\ = \quad & s_l \left[q_{ll} + q_{l0} + \cdots + q_{l,l-1} \right] + s_l [q_{l+1,l} + \cdots + q_{K,l}] = s_l. \end{aligned}$$

$\square$

Remark:
In particular, for $q_{kl} = \frac{1}{K+1}$ one gets

$$T = \frac{1}{K+1} \begin{pmatrix} 1 & 1 & 1 & 1 & \ldots & 1 \\ \frac{s_0}{s_1} & (2 - \frac{s_0}{s_1}) & 1 & 1 & \ldots & 1 \\ \frac{s_0}{s_2} & \frac{s_1}{s_2} & (3 - \frac{s_0}{s_2} - \frac{s_1}{s_2}) & 1 & \ldots & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}.$$

This matrix corresponds to the following procedure:

$\boxed{\text{Step } n \quad \to \text{Step } n+1}$

Given ξ_n, one chooses randomly (i.e. uniformly) a state η and then

$$\xi_{n+1} = \begin{cases} \eta & \text{with probability } \frac{s_\eta}{s_{\xi_n}} \wedge 1 \\ \xi_n & \text{with probability } 1 - \left(\frac{s_\eta}{s_{\xi_n}} \wedge 1 \right) \end{cases}$$

where $a \wedge b := \min\{a, b\}$. But in the even more special case where also $s_0 = \cdots = s_K$ this means that given ξ_n one chooses randomly (and independently from ξ_n) a state η and put $\xi_{n+1} = \eta$.

Generalized Metropolis algorithm

Hastings, 1970.
Under the assumptions of the Metropolis algorithm one takes a transition Matrix $T = (q_{kl})_{k,l=0}^{K}$ with $q_{kl} > 0$ which is not assumed to be symmetric. The probability r will in this case be computed as

$$r := \frac{s_\eta}{s_{\xi_n}} \frac{q_{\eta,\xi_n}}{q_{\xi_n,\eta}}.$$

Then one continues in the same way.

Application of the Metropolis algorithm

Assume that we have an irreducible homogeneous Markov chain with

- a state space $X = \{0, \ldots, K\}$,
- a stationary distribution $s = \left(\frac{1}{K+1}, \ldots, \frac{1}{K+1}\right)$,
- the number K is not known

Problem: Estimate K.
We will use the function $F : X \to \mathbb{R}$, where

$$F(k) = 1\!\!1_{\{k_0\}} = \begin{cases} 1 & \text{for } k = k_0 \\ 0 & \text{for } k \neq k_0 \end{cases}$$

with some fixed $k_0 \in X$. Now we get by

$$\frac{1}{K+1} = s_{k_0} = \mathbb{E}_s 1\!\!1_{\{k_0\}} = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} 1\!\!1_{\{k_0\}}(f_i(\omega)) \quad a.s.$$

an estimate for K.

4.7.1 An application: the Rasch model of item analysis

Assume we are given

- R persons, $R \geq 1$
- C test items, $C \geq 1$.

The answers of the test persons are coded as follows:

$$X_{ij} = \begin{cases} 1 & \text{person } i \text{ answers item } j \text{ correctly} \\ 0 & \text{otherwise} \end{cases}$$

Assume $\alpha_1, \ldots, \alpha_R \in \mathbb{R}$ and $\beta_1, \ldots, \beta_C \in \mathbb{R}$ are such that

$$\sum_{i=1}^{R} \alpha_i = \sum_{j=1}^{C} \beta_j = 0,$$

(where β_j stands for the difficulty of question j while α_i gives information about the intelligence or knowledge of person i). Suppose the $(X_{ij})_{i,j}$ are independent and, moreover,

$$\mathbb{P}(X_{ij} = 1) = \frac{e^{\alpha_i + \beta_j}}{1 + e^{\alpha_i + \beta_j}}.$$

Let us consider a realization $(\xi_{ij})_{i,j=1}^{R,C} = (X_{ij}(\omega))_{i,j=1}^{R,C}$

$$\begin{array}{c|cccc||c}
 & & & & & \sum_j \\
 & \xi_{11} & \xi_{12} & \cdots & \xi_{1C} & \xi_{1,\cdot} \\
 & \xi_{21} & \xi_{22} & \cdots & \xi_{2C} & \xi_{2,\cdot} \\
 & \vdots & & \cdots & & \vdots \\
 & \xi_{R1} & \xi_{R2} & \cdots & \xi_{RC} & \xi_{R,\cdot} \\
\hline\hline
\sum_i & \xi_{\cdot,1} & \xi_{\cdot,2} & \cdots & \xi_{\cdot,C} &
\end{array}$$

The question is now how many different $(\xi_{ij})_{i,j=1}^{R,C}$ one can find (=unknown number $K+1$) such that row and column sums are $\xi_{1,\cdot}, \ldots, \xi_{R,\cdot}$ and $\xi_{\cdot,1}, \ldots, \xi_{\cdot,C}$, respectively? We will use the following

Lemma 4.7.2. *Given* $\xi_{1,\cdot}, \dots, \xi_{R,\cdot} \in \{0, 1, \dots\}$ *and* $\xi_{\cdot,1}, \dots, \xi_{\cdot,C} \in \{0, 1, \dots\}$, *all events such that*

$$\begin{aligned} X_{11}(\omega) + \cdots + X_{1C}(\omega) &= \xi_{1,\cdot} \\ &\vdots \\ X_{R1}(\omega) + \cdots + X_{RC}(\omega) &= \xi_{R,\cdot} \\ X_{11}(\omega) + \cdots + X_{R1}(\omega) &= \xi_{\cdot,1} \\ &\vdots \\ X_{1C}(\omega) + \cdots + X_{RC}(\omega) &= \xi_{\cdot,C} \end{aligned}$$

have the same probability.

Proof. By independence and our previous assumptions

$$\begin{aligned} \mathbb{P}(X_{ij} = \xi_{ij}, \quad \forall i, j) &= \Pi_{i,j} \mathbb{P}(X_{ij} = \xi_{ij}) \\ &= \Pi_{i,j} \frac{e^{\xi_{ij}(\alpha_i + \beta_j)}}{1 + e^{\alpha_i + \beta_j}} \\ &= \frac{e^{\sum_{i,j} \xi_{ij}(\alpha_i + \beta_j)}}{\Pi_{i,j}(1 + e^{\alpha_i + \beta_j})} \\ &= \frac{e^{\sum_i \xi_{i,\cdot} \alpha_i + \sum_j \xi_{\cdot,j} \beta_j}}{\Pi_{i,j}(1 + e^{\alpha_i + \beta_j})} \end{aligned}$$

□

Fix now $\xi_{1,\cdot}, \dots, \xi_{R,\cdot}$ and $\xi_{\cdot,1}, \dots, \xi_{\cdot,C}$.
Question What is the cardinality of

$$X := \left\{ (\xi_{ij})_{i,j=1}^{R,C},\ \xi_{ij} \in \{0, 1\},\ \sum_i \xi_{ij} = \xi_{\cdot,j},\ \sum_j \xi_{ij} = \xi_{i,\cdot} \right\}?$$

Apply the Metropolis algorithm:

- state space X as above
- auxiliary transition matrix $\widetilde{T} = (q_{(\xi_{ij}),(\eta_{ij})})$

We consider sub-rectangles $\mathcal{R}$ as follows:

$$\begin{array}{cccc} \xi_{11} & & & \xi_{1C} \\ & \circ \cdots \circ & & \\ & \vdots \;\; \mathcal{R} \;\; \vdots & & \\ & \circ \cdots \circ & & \\ \xi_{R1} & & & \xi_{RC} \end{array}$$

We call a sub-rectangle 'switchable' (row- and column-sums of the matrix do not change if we 'switch' the corners) $:\Longleftrightarrow$

$$\begin{array}{ccc} \circ & \cdots & \circ \\ \vdots & \mathcal{R} & \vdots \\ \circ & \cdots & \circ \end{array} = \begin{array}{ccc} 0 & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & 0 \end{array} \quad \text{or} \quad \begin{array}{ccc} 1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & 1 \end{array}$$

If we 'switch' a rectangle only the 4 corners will be changed:

$$\begin{array}{ccc} \underline{0} & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & \underline{0} \end{array} \longrightarrow \begin{array}{ccc} 1 & \cdots & \underline{0} \\ \vdots & & \vdots \\ \underline{0} & \cdots & 1 \end{array}$$

or

$$\begin{array}{ccc} 1 & \cdots & \underline{0} \\ \vdots & & \vdots \\ \underline{0} & \cdots & 1 \end{array} \longrightarrow \begin{array}{ccc} \underline{0} & \cdots & 1 \\ \vdots & & \vdots \\ 1 & \cdots & \underline{0} \end{array}$$

The procedure (an implicit definition of $\widehat{T}$) is as follows. Choose per random (random numbers i_1, i_2, j_1, j_2 which are uniformly distributed) a sub-rectangle $\mathcal{R}$ in $f_n = (\xi_{ij})_{i,j=1}^{R,C}$. Then

$$f_{n+1}(\omega) = \begin{cases} \xi_{ij} & \text{if } \mathcal{R} \text{ is not switchable} \\ \eta_{ij} = \xi_{ij} \text{ with switched } \mathcal{R} & \text{if } \mathcal{R} \text{ is switchable} \end{cases}$$

Now one has to prove

- $\widehat{T}$ is symmetric

- the Markov chain is irreducible (in other words by switching one can reach any element of X)

Then we get for $F = \mathbb{1}_{\{(\xi_{ij})\}}$

$$\frac{1}{\#X} = \mathbb{E}_s F = \lim_{N\to\infty} \frac{1}{N} \sum_{k=1}^{N} F(f_k(\omega)) \quad a.s.$$

$\square$

Index

Bibliography

[1] Ch. Geiss and S. Geiss. *An introduction to probability theory*, volume 60 of *Lecture Notes*. University of Jyväskylä, 2009.

[2] P. Guttorp. *Stochastic Modeling of Scientific Data*. Oxford Science Publications. Chapman & Hall, 1995.

[3] 0. Häggström. *Finite Markov Chains and Algorithmic Applications*. Cambridge, 2002.

[4] A.N. Širjaev. *Probability*, volume 95 of *Graduate Texts in Mathematics*. Springer, 1984.

[5] T. Wai-Yuan. *Stochastic Models with Applications to Genetics, Cancers, AIDS and Other Biomedical Systems*. World Scientific, 2002.